Toolset for Incorporation of Unnatural Amino Acids into Proteins Expressed in Mammalian Cells

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

SHUBHAM SINGH

aus Indien

2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

First Supervisor: Prof. Dr. Timm Maier
Second Supervisor: Prof. Dr. Markus Affolter
External Expert: Prof. Dr. Ben Schuler

Basel, 20.09.2022

Prof. Dr. Marcel Mayor

The Dean of Faculty

I. ABSTRACT

Proteins are biological nanomachines that perform almost all the biological processes. The survival of a living organism depends heavily on proper coordination and functioning of these biomolecules. Naturally, malfunctioning proteins may stunt the growth and proper development of an organism and might even lead to death. Because of their impact on human health, proteins are among the most studied biomolecules. For such studies, proteins are usually produced using protein expression systems (*e.g.*, mammalian cells, insect cells, yeast cells or bacterial cells). For studying human proteins, mammalian systems or human cell-based systems are usually the best because they can most accurately mimic the production and maturation processes of the target-protein.

To understand the biophysical properties and working mechanisms of proteins, many biophysical methods are available. Such methods often rely on molecules known as probes, that can sense a change in their local environment and generate measurable signals corresponding to these changes. To study a protein molecule, these probes, must be covalently attached to the protein via one of its chemically reactive sites. Attachment of biophysical probe to a specific site on the protein requires a unique reactive site on the protein molecule. However, site-specific attachment of a biophysical probe is either not feasible or not practical when many reactive sites are present on the protein or if the reactive sites are functionally relevant for the protein. In such a scenario, presence of a bioorthogonal reactive handle on the protein of interest may allow the introduction of the biophysical probe without interfering with the reactive sites or functionally relevant amino acids of proteins. Such bioorthogonal reactive handles can be introduced in proteins via unnatural amino acids.

Using protein expression systems, unnatural amino acids can be genetically (or cotranslationally) incorporated into the proteins during the polypeptide biosynthesis. The most widely used strategy for doing this is genetic code expansion (G.C.E.) via stop codon suppression, where, by introducing an orthogonal tRNA and orthogonal aminoacyl-tRNA synthetase pair in the protein expression system, one of the three stop-codons (Opal, Amber, or Ochre) is repurposed as a signal-codon to incorporate unnatural amino acid(s) in proteins. Several genetic code expansion systems have been developed to incorporate more than 100 different unnatural amino acids into the proteins expressed in mammalian systems. These unnatural amino acids provide a gamut of unique biophysical and biochemical characteristics for characterizing proteins and polypeptides. Among other applications, unnatural amino acids can introduce bioorthogonal reaction handles, cross-linking handles, post-translational modifications or fluorescent side chains into proteins.

This work is dedicated for site-specific incorporation of unnatural amino acids in proteins expressed in mammalian cells. For doing so, we have developed mammalian expression vectors and, in this thesis, have demonstrated their potential for assimilating the existing G.C.E. systems.

We have also developed a cell-based screening assay for quantification of the UAA incorporation efficiency of the different G.C.E. systems. We have demonstrated that this screening assay provides a holistic picture about UAA incorporation conditions. We have also demonstrated that this assay is compatible with high-throughput screening experiments.

Finally, we have demonstrated a systematic strategy for efficiently incorporating two UAAs in proteins expressed in mammalian cells. Using this strategy, we have incorporated a fluorescent amino acid as well as a bioorthogonal handle in our test protein eGFP. The introduced biorthogonal handle can be used for site-specific incorporation of a fluorophore to perform FRET based experiments.

II. TABLE OF CONTENTS

III. LIST OF FIGURES

IV. LIST OF TABLES

V. ABBREVIATIONS

µg	microgram
µL	microliter
µm	micrometer
µM	micromolar
°C	degree centigrade
5-HTP	5-Hydroxy Tryptophan
a.k.a.	also known as
AA	Amino Acid
aaRS	Aminoacyl-tRNA Synthetase
aatRNA	Aminoacyl-tRNA
AFM	Atomic Force Microscopy
AmpR	Ampicillin Resistance
ANAP	3-(6-Acetylnaphthalen-2-ylamino)-2-Aminopropionic Acid
AO	Acridine Orange
ATCC	American Type Cell Culture
AzPhe	Azidophenylalanine
AzW	Azidotryptophan
BFP	Blue Fluorescence Protein
bGH	Bovine Growth Hormone
bp	Base Pair
BRET	Bioluminescence Resonance Energy Transfer
C-DMEM	Complete Dulbecco's Modified Eagle Medium
$CaCl_2$	Calcium Chloride
CMV	Cytomegalo Virus
CMV-IE	CMV Immediate Early
CO_2	Carbon Dioxide
csv	Comma Separated Values
DanAla	Dansyl Alanine
DAPI	4',6-Diamidino-2-Phenylindole Dihydrochloride
DEER	Double-electron electron-resonance
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside Triphosphate
DpaTyr	bis((dipicolylamino)methyl)tyrosine)
e.g.	*exempli gratia*
Ec	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid

eGFP	Enhanced Green Fluorescent Protein
eGFP	Enhanced Green Fluorescence Protein
$eGFP^{UAA}$	eGFP with the UAA present in the polypeptide chain
ELISA	Enzyme-Linked Immunosorbent Assay
EPR	Electron Paramagnetic Resonance
EPS	Expressed protein ligation
FACS	Fluorescence Aided Cell Sorting
FBS	Fetal Bovine Serum
FITC	Fluorescein isothiocyanate
fL	Femtoliters
Flp	Flippase
FP	Fluorescent Protein
FRET	Förster or Fluorescence Resonance Energy Transfer
FRT	Flp Recognition Target
FTIR	Fourier Transform Infrared
G.C.E.	Genetic Code Expansion
G.E.M.S.	G.C.E. Efficiency Measurement and Screening
GA	Gibson Assembly
H_3PO_4	Phosphoric Acid
HCl	Hydrochloric Acid
HEK	Human Embryonic Kidney
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HF	High Fidelity
His	Histidine
HMF	10xHis-Myc-FLAG
HPC	High Performance Computing
HPCC	High Performance Computing Cluster
HPLC	High Performance Liquid Chromatography
HS-AFM	High Speed Atomic Force Microscopy
i.e.	*id est*
ITC	Isothermal Calorimetry
KCl	Potassium Chloride
K_d	Dissociation constant
kDa	Kilo Dalton
LB	Luria-Bertani
LC-MS	Liquid Chromatography Tandem Mass Spectrometry
LED	Light Emitting Diode
Leu	Leucine
LoxP	Locus of Crossover in P1
Mb	*Methanosarcina barkeri*
mL	milliliter

Mm	*Methanosarcina mazei*
mM	millimolar
$MnCl_2$	Manganese (II) chloride
MQ	Milli-Q
mRSP	mRaspberry
ms	milli second
N	normality
N.E.R.	Normalized Expression Ratio
N_3Lys	Azidolysine
NaCl	Sodium Hydrochloride
NaOH	Sodium Hydroxide
NCL	Native chemical ligation
NEB	New England Biolabs
ng	nanogram
nm	nanometer
nm	Nanometers
NMR	Nuclear Magnetic Resonance
NP-40	Nonidet P-40
NTA	Nitrilotriacetic acid
OD_{600}	Optical Density at 600 nm
OMeY	Ortho-methyl Tyrosine
oRS	Orthogonal aminoacyl-tRNA Synthetase
otRNA	Orthogonal tRNA
OTS	Orthogonal Translation System
PBS	Phosphate Buffer Saline
PC	Polycarbonate
PCR	Polymerase Chain Reaction
PEI	Polyethyleneimine
PET	Photoinduced Electron Transfer
pg	picogram
PIPES	Piperazine-N, N′-bis(2-Ethanesulfonic Acid)
pK_a	Acid dissociation constant
POI	Protein of Interest
POI^{UAA}	Protein of interest having an unnatural amino acid in its polypeptide
polH	Polyhedrin
poly(A)	poly(adenylic acid)
PrK	Propargyloxylysine
PrPhe	Propargyloxyphenylalanine
PrW	Propargyloxytryptophan
PTS	Protein-*trans* splicing
Pyl	Pyrrolysine

RFU	Relative Fluorescence Units
RNA	Ribonucleic Acid
rpm	revolutions per minute
RS	Aminoacyl-tRNA Synthetase
SBTI	Soybean Trypsin Inhibitor
SDS	Sodium Dodecyl Sulfate
smFRET	Single Molecule FRET
SOB	Super Optimal Broth
SOC	SOB Medium with Glucose
SPR	Surface Plasmon Resonance
SV40	Simian Vacuolating Virus 40
TAA	Ochre Stop Codon
TAG	Amber Stop Codon
TB	Terrific Broth
TCC	Treated Cell Culture
TCEP	tris(2-carboxyethyl) phosphine
TCO	Trans-Cyclooctene
TE	Tris-EDTA
TGA	Opal Stop Codon
tif / tiff	Tag Image File
TIRF	Total Internal Reflection Fluorescence
TO	Tetracycline Operator
TRE	Tetracycline Repressor Element
Tris	tris(hydroxymethyl) aminomethane
tRNA	Transfer RNA
Trp	Tryptophan
UAA	Unnatural Amino Acid
UV	Ultraviolet
UV-Vis	Ultraviolet and Visible spectrum of light
v/v	volume by volume
ver.	Version
VSV	Vesicular Stomatitis Virus
WPRE	Woodchuck hepatitis virus Post-transcriptional Regulatory Element
WT	Wild Type
xg	g-force or Relative Centrifugal Force

VI. DEFINITIONS AND CONVENTIONS

$_{EcTrp}tRNA$	tRNA molecule(s) derived from Tryptophanyl-tRNA (Trp-tRNA) molecules of the *Escherichia coli* (*Ec*) organism.
$tRNA_{CUA}$	tRNA molecule(s) with CUA anticodon loop.
$^{UAA}tRNA$	tRNA molecule(s) evolved to recognize UAA amino acid.
$tRNA^{AA}$	tRNA molecule(s) aminoacylated with AA amino acid.
$tRNA^{AA}_{CUA}$	tRNA molecule(s) with CUA anticodon loop and amino acylated with AA amino acid.
$^{UAA}_{EcTrp}tRNA^{AA}_{CUA}$	tRNA molecule(s) with CUA anticodon loop, derived from Tryptophanyl-tRNA (Trp-tRNA) molecules from *Escherichia coli* (Ec) organism, evolved to recognize UAA amino acids and currently charged with AA amino acid.
$_{Trp}RS$	aaRS protein(s) derived from Tryptophanyl-tRNA synthetase protein.
$_{EcTrp}RS$	aaRS protein(s) derived from Tryptophanyl-tRNA synthetase protein from *Escherichia coli* organism.
^{UAA}RS	aaRS protein(s) evolved to recognize UAA amino acid.
$^{C123D}RS$	aaRS protein(s) with C123D mutation.
$^{UAA}_{Trp}RS$	aaRS protein(s), derived from Tryptophanyl-tRNA synthetase protein, evolved to recognize UAA amino acid.
$^{UAA}_{EcTrp}RS$	aaRS protein(s), derived from Tryptophanyl-tRNA synthetase protein from *Escherichia coli* organism, evolved to recognize UAA amino acid.
$RS\ /\ tRNA$	aaRS and tRNA pair (the aaRS and tRNA are individually defined).
$^{EcTrp-tRNA}_{EcTrp-RS}\mathbf{GCE}^{UAA}_{TAG}$	A genetic code expansion system comprising of $_{EcTrp}RS$ and $_{EcTrp}tRNA$, currently incorporating UAA amino acid in response to TAG codon.
eGFP^{39TAG}	Mutant of wild type eGFP protein with TAG stop codon at position 39.
eGFP$^{39TAG\|151TGA}$	Mutant of wild type eGFP protein with TAG stop codon at position 39 and TGA stop codon at position 151.
G.C.E.	Genetic code expansion: genetic incorporation of unnatural amino acids in response to a specific codon.
OTS	Orthogonal Translation System: orthogonal aminoacylated tRNA synthetase (oRS) and orthogonal tRNA (otRNA) pair.
G.C.E. system	Orthogonal aminoacylated tRNA synthetase (oRS), orthogonal tRNA (otRNA) and unnatural amino acid (UAA) considered together.
G.C.E. machinery	The biomolecules or macromolecules involved in genetic code expansion.
G.C.E. Plasmid	The expression vector used for delivering the OTS in the host cell.
G.E.M.S. Plasmid	The expression vector used for delivering the 'G.C.E. efficiency measurement' genes (*i.e.*, mRaspberry-eGFP tandem with stop-codon in eGFP gene) in the host cells.

1. INTRODUCTION

All living organisms share seven traits: organic nature, high degree of organization, pre-programming, interaction, adaptation, reproduction and evolution* [2]. To exhibit these traits, living organisms undergo chemical and mechanical reactions that can be unified under the term *biological processes*. Living organisms may contain hundreds of thousands of organized components that help execute the biological processes. In higher life forms, such an organization comes from organs, tissues, and cells. At cellular levels, the organization is a result of biochemical and biophysical interaction between inorganic compounds (e.g., water, ions), organic compounds (e.g., carbohydrates, amino acids, nucleotides) and complex biopolymers (e.g., DNA, RNA, polysaccharides, proteins). Over billions of years, the different components of biological life have evolved to work in unison to sustain life by driving the biological processes. For example, DNA stores the instructions for biosynthesis and reproduction, lipids form compartments to control biological process, polypeptides form biological nanomachines to perform biological functions and small molecules such as ions and nucleotides provide material and energy to drive the biological processes.

For functional and organizational analogy, biological life can be compared with a complex machine such as a car or an airplane. Like biological systems, such machines too are composed of numerous components, albeit man-made, that fit perfectly and work together flawlessly. However, when these components do not perform within specifications, the entire machine may breakdown or may cease to function properly. The same is true for living organisms as well – misbehaving biological components impede normal biological processes and are liable to cause diseases or disorders. Among the numerous examples of this phenomenon, perhaps the most well-known is that of sickle cell disease (SCD). In this disease, due the change in a single nucleotide in hemoglobin (β-globin) gene, the sixth amino acid of this protein becomes valine instead of glutamic acid, which causes crystallization of hemoglobin molecules and leads to the rigidity of red blood cells, which assume sickle shape [3]. Eventually, this may lead to anemia, repeated infections, and episodes of pain. On one hand, there are many disorders such as SCD, that are cause by malfunctioning macromolecules, on the other hand, there are numerous diseases caused by external agents such as pathogens. In both the cases, a thorough understanding of the relationship between the pathogenesis and the involved biomolecules and bioprocesses helps to alleviate the disease or disorder.

1.1. POLYPEPTIDES AND PROTEINS

Of all the biological macromolecules, proteins are the most diverse and the most ubiquitous, accounting for about 50% of the dry mass of cells [4]. Chemically, proteins are biopolymers of amino acids. Amino acids are organic compounds where a carbon atom is surrounded by four functional groups - an amino group ($-NH_2$ or $-NH_3^+$)†, a carboxylate group ($-COOH$ or $-CO_2^-$), a hydrogen atom ($-H$) and a sidechain (denoted as $-R$). In this configuration, since the carboxylate carbon (being the most oxidized carbon) is called as C^1 (IUPAC nomenclature), the aforementioned central carbon is called as the α-carbon (C^α). The sidechain ($-R$ group) of the amino acid determines its properties, such as total charge, polarity, or interaction with water.

Two amino acids can be chemically joined together by a peptide bond. Peptide bond ($-CO-NH-$) formation is a condensation reaction where the carboxylate group of one amino acid forms a covalent bond with the amino group of another amino acid, while releasing water as the by-product [5]. When two amino acids are joined in this way, the product is known as a dipeptide‡. After the formation of a dipeptide, the resultant compound still has a free amino group and a free carboxylate group. These free functional groups can form more peptide bonds with other amino acids or other dipeptides. Each condensation reaction increases the number of covalently joined amino acid residues, but the end-product always has a free amino group at one end and a free carboxylate group at the other end (unless the amino group of one compound joins with its own carboxylate group and a cyclic compound is formed). This polymeric product, containing several amino acids covalently joined by peptide bonds is known as a polypeptide. In living organisms, such

* Reproduction and evolution are facultative as not all living organisms display these traits.
† Proline amino acid has $-NH-$, a secondary amino group, rather than a primary amino group.
‡ For example, Aspartame, the famous artificial sweetener, is a dipeptide.

polymers can contain as less as 21 amino acids (e.g., in the chain-A of insulin) [6] or as many as over 34,000 amino acids (e.g., in the titin protein in muscles) [7].

1.1.1. Primary structure of the polypeptides

In a polypeptide, there exists a chain of atoms with repeating $(-N^{\alpha}-C^{\alpha}-C^{1}-)^{*}$ atoms, extending from the α-nitrogen (N^{α}) of the free amino group to the carboxylate carbon (C^{1}) of the free carboxylate group. This chain is known as the polypeptide backbone. Essentially, the polypeptide backbone contains alternating α-carbon (C^{α}) atoms and peptide bonds (–CO–NH–). Conventionally, the amino acid containing the free amino-group is considered the first amino acid of the polypeptide, and this end is known as the N-terminus of the polypeptide [8]. Similarly, the amino acid with free carboxyl-group is considered as the last amino acid or the C-terminus. The direction of the polypeptide backbone is from the N-terminus end to the C-terminus end. The primary structure of polypeptides is the sequence of amino acids in the polypeptide backbone.

1.1.2. Secondary structure of polypeptides

As we move along the polypeptide backbone $(-N^{\alpha}-C^{\alpha}-C^{1}-)_{n}$, torsion angles† between $N^{\alpha}-C^{\alpha}$ and $C^{\alpha}-C^{1}$ provide conformational flexibility to the backbone. While astronomically large number of conformations are possible for the polypeptide backbone [9], depending on their sequence, the polypeptides tend to adopt unique thermodynamically stable conformations through a stochastic search of accessible conformations [10]. In doing so, structured and unstructured segments can arise throughout the polypeptide chain. The secondary structure of polypeptides is the local organization of the amino acids that creates structured segments along the polypeptide backbone [11]. The most common of such secondary structures are alpha-helices and beta sheets, although other secondary structures such as 3_{10} helix, π helix, β-turn are also frequent and structurally significant. Most of the known secondary structure elements are stabilized by hydrogen bonds (H-bonds). Amino acid side chains can stabilize or destabilize polypeptide secondary structure through electrostatic interactions and/or steric effects.

The alpha-helix is a rod-shaped right-handed helical coil which is stabilized by H-bond between every fifth amino acid (i^{th} and $i+4^{th}$ amino acid). It contains a core of the polypeptide backbone with sidechains protruding in the outward direction. Beta sheets involve two different regions of the polypeptide chain that are in stretched conformation and lie side-by-side. If the two chains have the same direction, then the beta-sheet is known as parallel beta-sheet, otherwise it is known as anti-parallel beta-sheet.

1.1.3. Tertiary structure of polypeptides

The secondary structure elements of a polypeptide can interact with each other and with the environment through electrostatic, hydrophobic and van der Waals interactions, to adopt a conformation with low free-energy, in accordance with the laws of thermodynamics [12-14]. The tertiary structure of a polypeptide is a stable low-energy conformation of the polypeptide chain. The process of achieving this conformation is known as polypeptide folding, and in the stable conformation, the polypeptide is known as folded polypeptide. The folded polypeptide assumes a three-dimensional shape with characteristic mechanical properties which allow the polypeptide to interact with other biomolecules (such as ions, small-molecules and macromolecules) and drive biological processes. In this way, a folded peptide can act as a molecular nanomachine or as a component of a larger molecular nanomachine. This is analogous to bending a long metal wire to form paper clips, springs, ropes, meshes or sculptures (figure here). Folded polypeptides that can perform biological functions are known as proteins.

The tertiary structure of a polypeptide is determined by its amino acid sequence. Consequently, by changing the number and order of the amino acids, countless number of folded polypeptides with unique

---

* N^{α} is the nitrogen joined with C^{α}.

† For a molecular structure containing four covalently linked atoms, torsion angle is the angle between the two planes formed by the two sets of three atoms.

mechanical and functional properties can be generated. By exploiting this property of polypeptides, biological systems are able to create tens of thousands of different kinds of unique proteins from 22 proteinogenic amino acids [15].

1.1.4. Quaternary structure of polypeptides

The quaternary structure of polypeptides is the interaction of two or more folded polypeptides [16]. In a quaternary structure, each interacting polypeptide is known as a subunit, and the resulting polypeptide complex is known as a multimer (or dimer, trimer, tetramer etc. depending on the number of interacting subunits). Oligomerization, or formation of a multimeric complex, has several advantages. For example, the modular nature of multimeric complexes conserves energy by recycling of individual subunits. In some instances, association and dissociation of the complex is coupled with the regulation of biological functions. Most importantly, the overall thermodynamic stability of polypeptide complex can be improved by using several smaller polypeptide subunits rather than one large polypeptide to perform the same biological function.

1.1.5. Ubiquity and functions of proteins

Polypeptide folding and tertiary association provides proteins with precisely engineered moving parts where mechanical actions are coupled with biochemical reactions and biophysical responses [17]. This coupling forms the basis of biological processes for the cells and the organisms. Based on their function, seven different kinds of proteins are present in eukaryotic organisms. These are antibodies (responsible for defense against antigens, e.g., omalizumab, crizanlizumab), contractile proteins (responsible for muscle movement, e.g., actin, myosin), enzymes (responsible for catalyzing biochemical reactions, e.g., amylase, lysozyme), hormonal proteins (act as signaling molecules to regulate biological processes, e.g., insulin, glucagon), structural proteins (provide strength and support, e.g., collagen, keratin), storage proteins (store small molecules for future use, e.g., ferritin, casein) and transport proteins (assist in the movement of molecules, e.g., hemoglobin, ion-channels) [18].

1.2. PROTEIN BIOSYNTHESIS

The process of formation of proteins in biological cells is known as protein biosynthesis. For eukaryotic cells, this happens in five steps – transcription, post-transcriptional modification of RNA, translation of mRNA, polypeptide folding and post-translational modifications. The first two steps occur in the nucleus of the cell, translation occurs in the cytoplasm and the last two may occur either in the cytoplasm, within the cell membrane or outside the cells.

1.2.1. Transcription and post-transcriptional modifications

The 'synthesis instruction' for proteins and RNA is present in a region of DNA known as the gene. For protein biosynthesis, this instruction must be executed by the cellular machinery. In living cells, protein biosynthesis begins with transcription, i.e., the process of copying the genetic information on the RNA by using DNA as the template. Transcription is done by a protein complex known as RNA polymerase.

Only one of the two strands of the DNA double helix is read by the RNA polymerase. This strand is known as the *template strand*. Due to the complementary base pairing, the RNA sequence is same as the sequence of the *coding strand* of DNA, except for thymine replaced by uracil. RNA polymerase binds to the DNA at the start of a gene, opens the DNA double helix and manufactures the RNA polymer chain. It should be noted that not all genes yield proteins as the final product; for some of the genes, the final product may be an RNA molecule such as rRNA (ribosomal RNA) or tRNA (transfer RNA). Consequently, in eukaryotic cells, there are three different kinds of RNA polymerases. RNA polymerase I is mostly responsible for transcription of rRNA, RNA polymerase III is responsible for tRNA, rRNA and snRNA (small nuclear RNA) and RNA polymerase II mostly transcribes mRNA (or messenger RNA) and snRNA [19]. While all kinds of RNA are

involved at some stage of protein biosynthesis, it is the mRNA is the molecule that carries the specific instructions for the biosynthesis of one kind of protein from the DNA to the ribosomes (hence the term 'messenger'). Rest of the RNAs are common for all kinds of proteins.

The RNA polymerase can recognize the beginning of a gene through specific DNA sequences known as the promoter region (usually present upstream of the gene of interest) and forms the initiation complex around it. The RNA polymerase II initiation complex is a multimeric* biomolecular complex containing the subunits from the RNA polymerase and *the transcription factors*. After the formation of the initiation complex RNA synthesis starts in 5' → 3' direction. As the RNA polymerase moves along the template strand RNA gets elongated. Shortly after the start of RNA synthesis, the 5' end of the RNA is capped with cap0 structure (7meGTP). In mammalian cells, ribose sugar methylation of the first two bases of the elongating RNA gives rise to cap1 and cap2 structures as well. The addition of cap0, cap1 and cap2 structures is known as the capping of RNA and protects the newly formed RNA against nuclease degradation. Towards the end of the gene, the tail recognition signal (AAUAAA) and GU rich tract present on the RNA are identified by an endonuclease, which cleaves the RNA at CA cleavage signal and an associated Poly(A) polymerase adds 100-200 adenine residues on the 3' end of the newly formed RNA. This process is known as the tailing of RNA. The tailing marks the completion of RNA synthesis. However, the resulting RNA is not yet ready for translation and is known as primary transcript. The primary transcript undergoes splicing to form mRNA. Splicing is the process where segments of non-coding regions known as introns are removed from the primary transcript. After capping, tailing and splicing, the primary transcript finally becomes mRNA and is exported out of the nucleus.

1.2.2. Translation

Translation of mRNA occurs in the cytoplasm of the cells. In this process, the cellular decoding machinery, comprising of ribosomes and tRNA, reads the mRNA sequence and generates the corresponding polypeptide chain by repeatedly forming peptide bonds between amino acids. The '*message*' for recruiting any given amino acid is present in the mRNA sequence as a set of three nucleotides bases, known as a codon. Due to the four types of nucleotide bases present in the mRNA, a total of 64 (or 4^3) different 'triplet-codons' are possible†.

In the polypeptide biosynthesis machinery, tRNAs are the adapter molecules that translate the mRNA sequence to polypeptide sequence. These L-shaped molecules have four short base-paired double-helical stems (acceptor-stem, D-stem, anticodon-stem & T-stem), three loops (D-loop, anticodon-loop & T-loop) and a variable region [20]. The acceptor-stem and the anticodon-loop are on the opposite ends of the tRNA. The anticodon-loop of the tRNA recognizes the triplet-codons on the mRNA through nucleotide base-pairing. At the 3'end of the acceptor-stem, the cognate amino acid of a given tRNA is covalently attached by its aminoacyl-tRNA synthetase (or aaRS) via an ATP dependent biochemical reaction known as aminoacylation or the charging of tRNA, which results in the formation of aminoacylated tRNA ($tRNA^{AA}$) or charged tRNA. For protein biosynthesis to proceed correctly, it is imperative that aaRS are highly specific for both tRNA and amino acids.

The set of rules that define the relationship between the triplet-codons and the equivalent amino acids are referred to as the genetic code for the organism. For any given organism, one triplet-codon codes for a unique amino acid. Consequently, each tRNA carries a unique amino acid. However, one amino acid can be coded by multiple codons. Additionally, due to the imperfect base-pairing between the anticodon loop and the mRNA, wobbling is allowed for the third base of the codon-triplet, which allows one tRNA to read multiple codons. Out of the 64 available codons, the codon AUG (triplet-codon for methionine) and, with much less frequency, codon GUG (triplet-codon for valine) serve as the start codon [8]. For most of the known organisms, codons UAA, UAG and UGA do not code for any amino acid and serve as stop codons. They are respectively named as *ochre*, *amber* and *opal* codons [21]. The Universal Genetic Code is the version of genetic code used by almost all the organisms, with the exception of some protozoans, mycoplasmas and in the mitochondrial genomes of animals and fungi [8].

* Multimeric complex – A protein complex where more than one polypeptide chains are present.
† Codons are read in the direction 5' → 3' on the mRNA.

Translation step can be further subdivided into three stages – initiation, elongation and termination. In eukaryotic systems, during the initiation stage, the cap-binding complex (consisting of eIF4E*, eIF4G, eIF4A, eIF4B and poly(A)-binding protein PAPB) attaches to the cap of the mRNA and PAPB binds to the poly(A) tail of the mRNA to form a ring of mRNA. This structure binds the 43S preinitiation complex (containing the 40S subunit of ribosome, eIF1, eIF1A, eIF3, eIF2 and the initiator tRNA Met-tRNA$_i^{Met}$ [22]) and scans the mRNA from the 5' end for Kozak consensus sequence (GCCRCCAUGG) to align the initiator tRNA with the start codon (AUG). Once this alignment is complete, the 60S subunit of ribosome joins the complex with the help of eIF5, the cap-binding proteins and the initiation factors depart and the assembly of the 80S ribosome is complete, paving the way for polypeptide synthesis [8].

The 80S ribosome assembled around the mRNA has three sites for the tRNA: the A (acceptor) site, the P (peptide) site and the E (exit) site. Soon after the ribosome assembly, the initiator tRNA, which is covalently bound to methionine, occupies the P-site. The tRNA carrying the next amino acid enters the A-site of the ribosome, placing the P-site and A-site amino acids in proximity. The peptidyl transferase activity of the 28S rRNA creates a peptide bond between these two amino acids, while simultaneously cutting the bond between P-site amino acid and P-site tRNA. As a result, the tRNA at the A-site is covalently attached to a dipeptide. Next, the ribosome gets translocated by one codon, hence placing the now uncharged tRNA at the E-site and the tRNA carrying the dipeptide to the P-site. After the uncharged tRNA leaves the E-site, a new tRNA with the next amino acid enters the A-site. In this way the growing polypeptide chain is elongated by repeated transfer from the tRNA at the P-site to the amino acid at the A-site, till one of the three stop codons is encountered.

In eukaryotic systems, there are no tRNAs corresponding to the three stop codons TAA, TAG and TGA. As a result, when this codon is positioned in the A-site of the ribosome, the release factors (eRF1 and eRF3) recognize the stop codons and trigger the hydrolysis of the ester bond of the polypeptidyl-tRNA located at the P-site, hence releasing the polypeptide chain and completing the polypeptide synthesis [23].

1.2.3. Polypeptide folding and post-translational modifications

For functioning properly, the polypeptide chain must acquire the correct three-dimensional conformation. This is achieved by polypeptide folding. Traditionally it was believed that polypeptide folding is a self-assembly process resulting solely from the interaction of amino acid side chains. While this is true in case of secondary structure formation and for some small proteins, most of the polypeptides need assistance of other proteins known as chaperons for proper folding. While chaperons do not determine the folded confirmation of the proteins (which is still determined by the protein sequence), they bind to and stabilize the partially folded and unfolded conformations of the proteins [24]. In this way they prevent incorrect folding and aggregation of proteins within the cells. In addition to chaperons, two enzymes - protein disulfide isomerase (PDI), which forms disulfide bonds between two cysteine residues, and peptidyl prolyl isomerase, which catalyzes the isomerization between *cis* and *trans* configurations of prolyl peptide bonds, also play a role in proper folding of the polypeptides.

Post-translation modifications (PTMs) are reversible or irreversible covalent processing events that change the properties of proteins by proteolytic cleavage† and/or addition of modifying functional groups which extend the range of available amino acids. Proteolytic modification of proteins can play an important role in activation of some pathways (e.g., blood coagulation cascade), enzymes (e.g., digestive zymogens) and hormones (e.g., insulin) [24-26]. There are over 300 different types of PTMs where modifying functional groups are added to the amino acids of proteins [27]. Through these modifications, the cell diversifies the behavior and characteristics of proteins such as enzymatic activity, assembly, lifespan, intramolecular interactions, intracellular interactions, trafficking, activation, solubility, folding and localization. Phosphorylation, acetylation and ubiquitination account for more than 90% of such PTMs.

* eIF = Eukaryotic Initiation Factor

† Proteolytic cleavage - the cleavage of the polypeptide chain

1.3. PROBES FOR PROTEIN CHARACTERIZATION

Studying the behavior of a protein molecule is known as protein characterization. Protein characterization can either be *in-situ*, *i.e.*, in its natural environment, or *ex-situ*, *i.e.*, in a controlled environment after isolation. Several biophysical and biochemical methods are available for protein characterization [28-31]. Based on the underlying mechanisms for quantification, protein characterization methods can be grouped as 'label-based' or 'label-free' methods. Methods such as size exclusion chromatography, mass spectrometry, AFM and mass photometry are label-free, *i.e.*, they depend entirely on measurement of the intrinsic properties of proteins. These methods usually require a high degree of sample purity to reduce noise originating from non-target proteins. For some label-free methods (such as LC-MS and SPR) the target-selection step is inherently present, but most label-free methods can be used only after protein purification. Label-based protein characterization methods, such as EPR, radioimmunoassay and those involving fluorescence, are generally dependent on the attachment of an external probe on to the target protein. The probe produces unique measurable signal(s) whose magnitude depends on the applied physical stimulus and/or the physicochemical microenvironment of the probe. In this way, the external probe functions either as a beacon or as a microenvironment-reporter for the protein. As the signal from the probe is unique and/or well-distinguishable, label-based methods can be used for studying target proteins *ex-situ* as well as *in-situ*. However, this advantage for *in-situ* protein characterization is realized only if the external probe can selectively label the target protein.

Probes used for protein characterization usually contain three components 1) a target-selective ligand-group, 2) a reactive-group (a.k.a. the warhead) to covalently bind the probe with the target, and 3) a reporter-group with a 'unique' physical or chemical property which allows the identification of the reporter-group against the common background [32]. These three components of the probe are connected via covalent linkers, that may be of different lengths for different probes and that may also serve as protectors of the reporter group in some cases. For some probes, a strong affinity between the target and the ligand-group obviates the need for the reactive-group. Conversely, in some labelling methods, the reactive-group may have only a few reaction conjugates and hence it may simultaneously function as the target-selective ligand-group as well.

There is a large difference between the timescales of different biological events, as well as the range of the three-dimensional space in which these biological events may occur. For example, conformational changes in protein molecules, which are essential for protein functions, typically occur over microsecond timescales and can occupy a few nanometers of physical space. In comparison, sometimes for performing biological functions, proteins have to migrate for a few micrometers inside the cells, which can take anywhere between a few seconds to a few minutes. While studying both these events is essential for understanding the protein function, they represent two very different scales, and may not be studied by the same method or at the same time. For this reason, while addressing the research problem, the required spatiotemporal resolution, the scale and the duration of the protein-characterization are important considerations while choosing the method. Consequently, the probe (as well as the instrumentation) should be compatible with the experiment requirements. Numerous synthetic and biosynthetic probes are available for protein characterization, and together they can provide biophysical and biochemical information about the properties of the target-protein with varying degree of spatiotemporal range and resolutions.

The temporal resolution of a measurement determines the ability to separate two consecutive events and depends on the sampling rate, *i.e.*, the frequency with which the state and/or the location of the target can be accurately determined. For probe-based methods, the temporal resolution of the information depends on the excitation[*] frequency and the relaxation[†] time of the probe, as well as on how accurately the individual excitation-relaxation events can be measured by the instrument. The other characteristic of a measurement's time-scale is the temporal range of the measurement, which determines the duration for which the information can be collected and interpreted accurately. It depends on the physicochemical stability of the probe and the sample over time.

The spatial resolution of a measurement determines the ability to separate two adjacent objects and depends on the ability to unambiguously locate the objects with respect to each other. The spatial

[*] Excitation is defined as the application of an external stimulus on the system, to produce a response.
[†] Relaxation is defined as return of the perturbed system to equilibrium.

resolution primarily depends on the characterization method*, the instrumentation† and the strength of the signal from the targeted object as compared to the background. For protein characterization, the targeted object is the protein. However, since the measurements in label-based methods are made for the reporter-group (rather than for the target-protein), the spatial resolution of the gathered information is limited by the label's size and its degrees of freedom with respect to the target-protein. For example, a longer and/or more flexible protein-reporter linker or a higher degree of heterogeneity in protein-probe conjugation introduces uncertainty about the position of the reporter-group with respect to the protein (*i.e.*, increase in the degrees of freedom) thereby reduces the spatial resolution of the gathered information. Similarly, a larger size of the probe increases uncertainty about the position of the probe itself (and as a result for the protein as well), thus it also reduces the spatial resolution. On a separate note, if the size of the probe is significant as compared to the target, the unwanted non-specific interactions between the probe and the target may introduce artefacts in the measurements or the probe may not remain sensitive to the target's microenvironment, thereby violating the purpose of the probe. Conversely, a site-specifically introduced probe of a small size and with a small linker may even be able to resolve the relative arrangements/movements of the different domains of a protein [33]. During protein characterization, the spatial resolution of the measurement determines if the results are specific to the targeted amino acid, the targeted secondary structure, the targeted domain of the protein, the targeted protein molecule, the tertiary-complex/assembly of the proteins or the macromolecular-assembly of the cell that contains the labelled protein. The spatial resolution puts a lower limit on the scale of the measurement, while the spatial range puts the upper limit. The spatial range of the measurement determines the scale at which the information can be collected and accurately interpreted. The spatial range of a probe depends on the physicochemical interactions between the probe and its environment, and the change in the magnitude of such an interaction with distance.

1.3.1. Fluorescence Methods

Among all the biophysical methods that benefit from probes, fluorescence-based protein characterization methods are, by a huge margin, the biggest beneficiaries. Fluorescence is a form of luminescence caused by absorption of radiation at one wavelength and near-immediate emission of radiation at a different wavelength. Molecules that exhibit fluorescence are known as fluorophores. In very simple terms, after absorbing the energy of one or more photons, an electron of the fluorophore molecule loses some of this energy in vibrational relaxation and/or internal conversion, and then loses the remaining energy either radiatively (*i.e.*, by emitting a photon and exhibiting fluorescence) or non-radiatively (*i.e.*, by transferring the energy to an electron of a compatible neighboring molecule) [34-36]. Furthermore, in case of non-radiative energy transfer by the donor-fluorophore, the acceptor-molecule electron too would lose the energy either radiatively (*i.e.*, by emitting a photon and exhibiting FRET) or non-radiatively (*i.e.*, by losing the energy and exhibiting fluorescence-quenching). The emitted photon (from fluorescence as well as from FRET) always has a lower energy than the total energy initially absorbed by the fluorophore, which creates the difference between the 'excitation-wavelength' of the radiation and the 'emission-wavelength' of the radiation. All fluorescence-based methods invariably rely on fluorophores, which have characteristic photophysical properties (*i.e.*, excitation spectrum, emission spectrum, fluorescence decay lifetime, absorption coefficient, quantum yield and photochemical stability). As natural fluorophores usually have much lower brightness (*i.e.*, the product of absorption coefficient and quantum yield) than synthetic or biosynthetic ones, fluorescence-based methods usually have a very high signal to noise ratio. Also, since most fluorophores can be visually detected, there are two ways to acquire protein-characterization data from fluorescent probes – either the photophysical properties of fluorophores can be measured to interpret (or indirectly observe) the state of their environment, or the fluorophores can be used as beacons for the direct observation of the target. Both these modes, separately as well as collectively, can be used *in-vitro* or *in-vivo*.

The photophysical properties of fluorophores are generally very sensitive to their immediate surroundings, which makes them excellent probes for studying the target-protein's microenvironment such as pH, crowding, polarity, ion concentration etc. [37-40]. Fluorophores can also interact with neighboring (10 nm

* The underlying mechanism of a method sets a theoretical limit on the possible spatial resolution.
† Even if theoretically possible, spatial resolution may be technologically restricted.

or less distance) photoactive molecules to give rise to FRET*, PET† and BRET‡ phenomena [41-43]. Since the magnitude of these phenomena depends on the distance between the interacting photoactive molecules, they can be used as molecular rulers for nano-scale distance measurement to study the interactions and dynamics of protein molecules.

Fluorescence generally occurs on a timescale of 10^{-9} to 10^{-7} seconds (excitation timescale is 10^{-15} seconds; typical half-life is 10^{-9} to 10^{-8} seconds) [44, 45]. In comparison, the relevant events of protein dynamics (*e.g.*, free diffusion, oligomerization, domain movements, conformational changes, ligand-binding, catalysis and polypeptide folding-unfolding) have a timescale of 10^{-9} to 10^{6} seconds [46, 47]. Since modern-day fluorescence instruments are able to quantify fluorescence emissions for the most part of this timescale, or in some cases, even able to resolve the individual photon-emissions, fluorescence-based methods are suitable for obtaining high-temporal resolution information for a variety of biological processes involving proteins. For example, they are used to characterize the linear and rotational dynamics of their targets, which can be used to monitor the hydrodynamic properties of the target protein, protein-protein interactions, protein-ligand interactions or protein oligomerization etc. By monitoring the fluorescence lifetime (*i.e.*, average decay of individual photons), two dyes with similar emission can be differentiated as well as the state of their microenvironment can be interpreted. This can be used for fluorescence lifetime microscopy to study the microenvironments of the different regions of a living cell.

By appropriately illuminating the sample with the excitation-wavelength light, fluorophores can also be used as beacons for locating or tracking particles in the three-dimensional space. This property is used in fluorescence microscopy for studying the behavior of proteins *ex-vivo* . However, since fluorescence-based methods (generally) depend on UV-Vis region of the electromagnetic spectrum, conventionally, the spatial resolution of fluorescence-imaging based methods is diffraction-limited (usually 200-300 nm laterally and 500-800 nm axially). However, by enforcing spatially-selective excitation or multiphoton excitation of the fluorophore or by being able to collect signal from single event allowing to reconstruct gaussian profiles of emitters based on the point spread function of a microscope, the resolution can be pushed to as low as 2.4 nm laterally and 1.9 nm axially [36, 48, 49]. This is particularly beneficial for high-resolution fluorescence imaging [50].

The high spatial-resolution and high temporal-resolution of fluorescence-based methods can also be combined together for detecting individual fluorophores, thereby performing fluorescence-based single molecule studies. Single-molecule methods provide statistical information about the individual states present within the measured ensemble, which can be crucial for understanding the structural heterogeneity of the sample.

Fluorescence-based methods have revolutionized protein characterization, both *ex-situ* and *in-situ*, and have made significant contributions towards our understandings of functional mechanisms of proteins. Obviously, fluorophores are at the heart of all fluorescence-based methods. Therefore a considerable attention has also been given for developing thousands of unique fluorescent probes, both synthetic and biosynthetic in nature, as well as for introducing these probes on to the target biomolecules. Such probes differ in their photophysical properties and hence provide many options for fluorescence-based studies. Due to their unique biophysical properties, these fluorescent probes can be distinguished not only from other biomolecules, but also from each other. In this way they can serve as unique biophysical markers for protein identification and characterization.

The fluorescence of naturally fluorescent amino acids, *i.e.*, tryptophan, tyrosine and phenylalanine, are used for label-free *ex-situ* characterization of proteins. In particular, tryptophan fluorescence is often used to extract information about a protein's local environment, polypeptide folding or unfolding, conformational dynamics and interaction with ligands [51-55]. However, these 'natural' amino acids have limited applications for fluorescence-based studies *in-vivo*, due to their near-ubiquitous presence in most of the proteins, their low quantum yields and the fact that they exhibit fluorescent (excitation and emission) in the biologically damaging UV region of the spectrum [56]. Due to this reason, almost all the fluorescence-based protein characterization methods require labelling with a fluorescent probe. By the virtue of differences in their reporter-groups there are four different kinds of fluorescent probes – synthetic dye

* Förster or fluorescence resonance energy transfer.
† Photoinduced electron transfer.
‡ Bioluminescence resonance energy transfer.

based (*e.g.*, Cy3 (ThermoFisher), Cy5 (ThermoFisher), FITC [57], Alexa Flour (ThermoFisher)), nanoparticle based (*e.g.*, quantum dots, gold nanoparticles etc.), fluorescent protein based (*e.g.*, eGFP*, BFP†, mCherry etc.) and natural fluorophore based (*e.g.*, flavin, chlorophyll etc.) [58]. Except for natural fluorophores, the fluorescence properties of fluorescent probes can be tuned by modifying the physical or molecular structure. Furthermore, the fluorophore-core of these probes can be functionalized with various ligand-groups or reactive-groups for labelling biomolecules. This provides us with a large number of fluorescent probes suited for specific applications. However, each kind of fluorescent probe has its own advantages and disadvantages. Synthetic-dye based probes are basically organic molecules and have very small size 0.5-1 nm, because of which they can be used for site-specific labelling and protein characterization at a high spatiotemporal resolution. However, they are not very photostable (*i.e.*, they photobleach and lose fluorescence over time) and hence not suitable for prolonged data collection. Nanoparticle based fluorophores are very bright (*i.e.*, they have a very good absorption coefficient and quantum yield) and quite photostable, but they are very big (6-60 nm) as well hence not suited for site-specific labeling of proteins. Both organic fluorophores and nanoparticle-based fluorophores must be introduced in the cells for *in-situ* studies, which is a step that can be quite challenging for some of these fluorophores due to their physicochemical properties. Fluorescent proteins, however, can be genetically introduced in the cells, which makes them excellent reporter for monitoring protein production, protein localization and protein-protein interaction. However, apart from having relatively lower photostability, they too are quite large biomolecules and hence they are not suited for site specific studies. Naturally occurring fluorophores such as flavin, chlorophyll can be used for *in-vitro* studies, but their applications are limited to a small number of protein molecules [58].

1.3.2. Electron Paramagnetic Resonance Spectroscopy

Electron paramagnetic resonance (EPR) spectroscopy is a method to characterize systems that contain unpaired electrons or paramagnetic centers. When a paramagnetic center is present on the target protein, EPR spectroscopy can provide information about the rotational dynamics, solvent accessibility and microenvironment polarity of the protein or the domain [59]. When two paramagnetic centers are present (on the same protein molecule or between two different protein molecules), double-electron electron-resonance (DEER) can also quantify the mean distance and the distance distribution between them, thereby providing information on conformational dynamics [60]. EPR spectroscopy has been used for characterizing protein secondary structure, protein-protein interactions, protein-ligand interactions, protein unfolding kinetics and protein dynamics [61]. As unpaired electrons are energetically unfavored, most protein molecules are 'EPR-silent', which results in almost no background for EPR methods [59]. This makes EPR methods suitable for protein characterization *in-vivo* [62]. Paramagnetic centers are site-specifically or non-covalently introduced in proteins using spin labels. The reporter group of these spin labels is generally based on nitroxides, trityl radicals or transition metal cations [63].

1.3.3. Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy is used to observe local magnetic fields around atomic nuclei, which is a function of the electron distribution around an atom or the local molecular environment. It can provide atomic resolution information about the structure, conformational dynamics and interactions of proteins. NMR is primarily used for the *ex-situ* protein characterization, but can also be used for the *in-situ* characterization of proteins [64-66]. NMR spectroscopy requires presence of NMR-active atoms. ^{1}H is the most prominent NMR-active atom in proteins and produces sharp chemical shift peaks. ^{14}N is also an NMR-active atom, but produces wide and poorly resolved chemical shift peaks. The remaining elements of proteins, *i.e.*, ^{12}C, ^{16}O and ^{32}S, are NMR-silent atoms due to the absence of nuclear magnetic dipole moment. The ^{1}H NMR spectrum of proteins contains its structural information, but cannot be interpreted due to over-abundance of signals. Correlation of NMR information from other NMR-active atoms, such as naturally occurring isotopes ^{13}C and ^{15}N, benefits in the interpretation of the structural information [67]. For this reason and due to low natural abundance of these natural isotopes, protein molecules are enriched with stable NMR active atoms such as

* Enhanced Green Fluorescent Protein.
† Blue Fluorescent Protein.

^{13}C, ^{15}N, ^{17}O or ^{19}F. These atoms act as 'NMR-probe'* to detect the local environment of the protein molecules. NMR probes can be introduced in proteins by uniform labelling, residue-specific labelling or site-specific labelling. Uniform labelling and residue-specific labelling can be used for structural characterization of the entire protein, while site-specific or domain-specific labelling can help to understand the local organization and dynamics within a protein molecule [67].

1.3.4. Chemical Crosslinking

In general, crosslinking is the process of covalently joining two (or more) molecules. While the covalent attachment of probes on to the protein molecules also, by definition, qualifies as crosslinking, for protein characterization this term is mostly used for covalent attachment of the target protein with other proteins or biomolecules. Chemical crosslinking is primarily used, in conjugation with other biophysical methods, to understand the intramolecular and/or intermolecular transient interactions of a protein molecule, by stabilizing the interactions and mapping the spatial proximity of the atoms. Amine-reactive, sulfhydryl-reactive, carbonyl-reactive or guanidinyl-reactive crosslinker compounds can be used for stabilizing or mapping protein-protein interactions [68]. Formaldehyde and trifunctional crosslinking agents can be used for crosslinking proteins *in-vivo*, thereby facilitating the proteome-wide characterization of protein-protein interactions [69-71]. Photoactivated crosslinkers, which are based on benzophenone, aryl-azide or diazirine, are inert by default but can be 'switched-on' by UV light [68, 69]. Potentially, they can allow the characterization of time-dependent protein-protein interactions. When chemical crosslinkers are site-specifically attached to protein molecules, they can serve as probes for the local interactions.

1.4. ATTACHMENT OF PROBES ON TO THE PROTEIN MOLECULES

Selectivity is the preferential outcome of one chemical reaction or interaction over other possible chemical reactions or interactions. In practice this is achieved when, in any given environmental condition, one molecule chooses one specific functional group or molecular complex over others. As a myriad of biomolecules and functional groups are present living systems, selectivity is at the heart of labelling biomolecules with biophysical probes. In context of large biomolecules such as proteins, labelling selectivity could be defined in context of either selecting a specific protein out of many, or selecting a specific site on a given protein molecule. While both protein-specificity and site-specificity can be used to monitor the behavior of proteins, the site-specific selectivity becomes quite important when high spatial resolution is a consideration. For labelling a protein, a unique functional group on the protein molecule is targeted by a conjugate molecule present on the biophysical probe to attach the probe on the protein molecule. A preferable protein labelling method requires a highly specific labelling reaction, a high rate constant for the forward reaction (especially when reaction is performed in a hostile environment), high reactivity at moderate conditions, high affinity between the interacting partners, stability of the labelling reagent in aqueous environment, stable end product, very low or no background signal, easy and complete removal of leftover reagents, a low number of processing steps, either no or very low disruption in protein functions, versatility in the choice of labels, and, for labelling *in-vitro*, either no or very low cytotoxicity and high membrane permeability of the labelling reagent.

Traditionally, the chemically reactive groups on the protein molecule, such as carboxylic acids, amines, amides, thiols and alcohols, have been used as targets of synthetic probe labelling [72, 73]. Due to the low abundance of cysteine residues in proteins (0.5-2.3% of all proteogenic amino acids) and due to high nucleophilicity of the thiol-containing side chain, this amino acid has been a preferred choice for site-specific labelling of the protein molecules [74, 75]. Electrophiles such as maleimides, iodoacetamides, alkyl halides, pyridyl disulfides and sulfonylhydroxylamine can be used for cysteine-specific chemical reactions [72, 76]. Lysine residues and the N-terminus amino group can be used for labelling via primary-amine specific chemical reactions using activated esters, sulfonyl chlorides, isothiocyanates or aldehydes [72]. As lysine residues are quite abundant in proteins (*e.g.*, 6.6% of amino acids in proteins of human muscles [77]), this labelling approach is generally very non-specific for amino acid positions. Cysteine and lysine residues

* The term NMR-probe is used figuratively here. The term NMR-probe actually refers to the instrument in which the samples are present during the measurement.

are the most used amino acid residues for chemical modification of proteins. In addition, in presence of various catalysts, tyrosine can be modified using diazonium salts, iodine, nitrous acid, phenol, aldehyde & anilines, or π-allyls functional groups, tryptophan can be modified with diazo compounds, glutamate & aspartate can be modified with carbodiimides, and histidine can be modified using polycarbonates [72, 73]. Recently, serine selective bioconjugation has also been reported [78]. Apart from these amino acid residues, due to its lower pK_a value than the ε-amino functional groups, in some cases, the N-terminal amino group can also serve as an important target for biophysical probes via transamination reactions, native chemical ligation (NCL), expressed protein ligation (EPL) and split-intein mediated protein-*trans* splicing (PTS) [79-81].

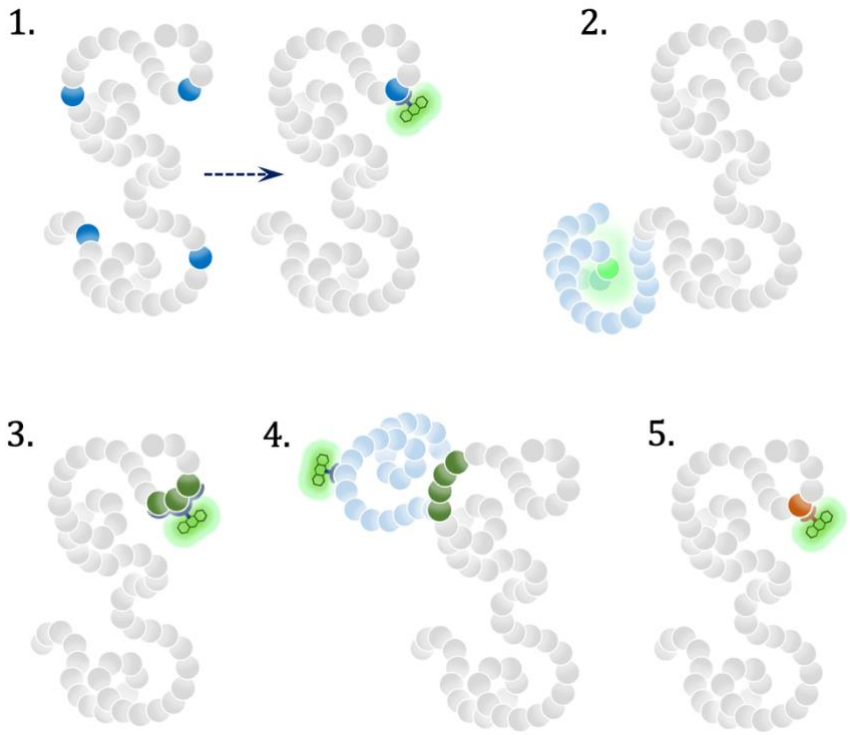


Figure 1: Common Strategies for Placing Probes on Proteins.
Several strategies are available for labelling proteins with probes. By using the reactive side chains of amino acids (blue residues), probes can be covalently attached on to the protein molecules, but for site-specific incorporation, the additional reactive residues must be removed by mutagenesis (1). Covalently attached reporter proteins can be genetically introduced on to the target-protein to function as probes (2). Special motifs (green residues) on the target-protein may be recognized by some dyes for site-specific covalent attachment (3). Special motifs or sequences (green residues) may also be recognized by antibodies for non-covalent attachment to the target-protein (4). Unnatural amino acids containing the probe or bioorthogonal reactive handles as their side chain can be genetically introduced into the polypeptide chain for site-specific probe labelling (5).

Traceless labelling* is another method of introducing synthetic probes on protein molecules *in-vivo* [82-84]. In this method, the synthetic probe is covalently attached to the given protein's ligand via a phenylsulfonate (tosylate) ester linkage. As the ligand binds to the protein, due to the proximity-induced reaction between the phenylsulfonate group and a nucleophilic amino acid on the surface of the protein, the probe gets covalently attached to the protein of interest. This method requires no genetic modification in proteins and can be used with endogenous proteins. However, it relies on a thorough knowledge of the interaction between a protein molecule and its substrate. Furthermore, since any nucleophilic group in the activity pocket's vicinity can be targeted by the tosylate group, this method provides a limited control over the site-specific attachment of the synthetic probes.

Targeting a specific type of amino acid in an attempt to label a single residue for conjugation reaction has its advantages and disadvantages. For small protein molecules this approach allows site specific

* Not to be confused with traceless ligation reaction.

biophysical probe labelling, generally without compromising the protein function. Using point mutagenesis, a single reactive amino acid side chain can be introduced into, or few unwanted amino acids be removed from the protein molecules. However, for large protein molecules or for molecular complexes, this approach of selectively introducing or removing reactive amino acids can be quite tedious due to the sheer number of such residues. For most proteins it might even be counterproductive due to the destabilizing effects of large number of mutations in proteins. To prevent mutagenesis within the protein of interest, (poly)peptide tags can be genetically incorporated in proteins.

In the past two decades, fluorescent and luminescent proteins (*e.g.*, GFP, mCherry, Renilla luciferase) have contributed significantly towards understanding the dynamics, localization and interactions of proteins *in-vivo* and *in-vitro* [41, 85-95]. Furthermore, new fluorescent proteins with unique photophysical properties are added every year to a long list of already existing ones [96]. These reporter proteins are covalently attached to the protein of interest via a short linker peptide and hence serve as a covalently attached biophysical probe. Similarly, self-labelling enzymes (*e.g.*, CLIP-tag, SNAP-tag, HaloTag) are also covalently attached to the protein of interest and can be used with fluorogenic compounds or fluorophores for monitoring protein behavior [97-99]. However, the relatively large size (3-5 nm or 20-25 kDa) of these covalently attached proteins can potentially interfere with protein-protein interaction and the long linker-length ensures that these protein molecules cannot be used for site-specific targeting of the protein of interest [100]. Additionally, for minimizing the disruption in the protein function, these protein molecules are generally added at either the N-terminal or the C-terminal, further restricting their site-specific applications. For localization of protein molecules *in-vivo* and *in-vitro*, primary antibodies coupled with reporter-tagged secondary antibodies can be used [101-105]. Since antibodies recognize the three-dimensional structure of the target protein for attachment, they are highly selective, and, in some cases can be used to distinguish apo and holo forms of proteins. However, each of these antibody adds ~10 nm to the size of the target protein and hence suffers the same drawbacks of aforementioned covalently attached reporter proteins [106]. Furthermore, antibodies are quite difficult to generate, so their widespread use is restricted to proteins with known biochemical properties [107].

Short peptide tags (0.6-6 kDa) with a unique amino acid sequence can also be genetically encoded in proteins and can be targeted by the biophysical probes via enzyme mediated labelling, peptide-peptide interaction or molecular complex dependent labelling [108]. The enzyme mediated labelling involves covalent attachment of biophysical probes to the protein by the ligases (*e.g.*, biotin ligase, lipoic acid ligase), transferases (*e.g.*, phosphopantetheinyl transferase, transglutaminase) or transpeptidases (*e.g.*, Sortase A), in response to their recognized peptide sequence. Biophysical probe labelling via peptide-peptide interaction is possible because of the mutual affinity of two peptide sequences. Coiled-coil motif-based interactions (*e.g.*, E3/K3 motif, E4/K4 motif or leucine zipper for non-covalent labelling) and its variants (for covalent labelling) have been heavily used for labelling the protein molecules for experiments *in-vitro* and *in-vivo* [108]. Other peptide-peptide interactions such as α-Bungarotoxin & α-Bungarotoxin Binding Sequence and Streptavidin & Streptavidin Binding Peptide have also been used in limited capacity for studying protein-protein interactions *in-vivo* [108, 109]. Previously mentioned labelling method protein-*trans* splicing also depends on peptide-peptide interaction, where the peptide tag excises itself, leaving only the covalently attached probe. Molecular complex dependent labelling is dependent on the recognition of metal ions and small molecules by the peptide tags. Examples of such peptide tags include tetracysteine motif (for biarsenical probes), tetraserine motif (for bisboronic acid probes), hexa-His or deca-His tags (for Ni-NTA based probes), oligo-Asp tag (for DpaTyr based probes), Lanthanide-binding tag (for Tb^{3+} probes), fluorette-tag (for Texas red, Rhodamine red, Oregon green 514, fluorescein fluorophores) and dC10α tag (for maleinimide-coumarin compounds) [100, 108]. Peptide tags offer a viable alternative of protein-based tags, not just for studying protein molecules, but also for applications such as purification, co-purification and immobilization [110, 111].

The shortest possible tag for introducing labels on proteins is just one residue in size, *i.e.*, an amino acid having a unique reporter-group as its side chain. Since such amino acids do not exist naturally, they are termed as synthetic or unnatural amino acids. The unnatural amino acids, when introduced in proteins, serve as a residue-type specific or residue-position probes. The reporter-group of the unnatural amino acids can be selected based on the protein characterization method and the target-protein's tolerance for chemical and structural changes. Furthermore, instead of the reporter-group, the side chain of the unnatural amino acid may also contain a unique chemical-reaction handle to introduce a variety of probes at a later timepoint. Unnatural amino acids can be introduced in proteins either proteome-wide or site-specifically. For proteome-wide incorporation of unnatural amino acids, those synthetic homologues of

proteinogenic amino acids which can be tolerated by the cells, are added to the protein expression media and the protein expression machinery substitutes some of the proteinogenic amino acids with their synthetic homologues [112-114]. This, however, is a random process and while it can provide some information on the properties of the protein(s), due to the high degree of heterogeneity in samples, the processing of the information could be either not possible or significantly difficult. Site-specific incorporation of unnatural amino acids can be achieved either via a synthetic or a biosynthetic approach. In the synthetic approach, unnatural amino acids can be added to the polypeptide chain during chemical synthesis [115]. This approach can be very beneficial when the required unnatural amino acid is not compatible with the protein biosynthesis machinery. However, there are many disadvantages of chemical synthesis, including high cost, low yield and an upper limit on the length of the polypeptide chain. For site-specific biosynthetic incorporation of unnatural amino acids, tRNA molecules charged with the unnatural amino acid (or, in short, $tRNA^{UAA}$) are introduced in the protein expression system and during the biosynthesis, in response to a specific codon on the mRNA, the ribosome introduces the unnatural amino in the polypeptide chain. This approach is known as genetic incorporation of unnatural amino acids. Furthermore, the $tRNA^{UAA}$ can either be produced synthetically (*i.e.*, by chemically joining the tRNA and the UAA) or enzymatically and introduced in the expression systems (usually cell-free expression systems or large biological cells such as *Xenopus* oocytes) [116-118], or the biological cells can be programmed to produce the required tRNA and enzymatically charge it with the required UAA. In case of synthetic production of $tRNA^{UAA}$, there is some flexibility in the choice of the UAA, but the yield of the protein is limited by the amount of $tRNA^{UAA}$ added to the system. In case of biosynthetic production of $tRNA^{UAA}$, the protein expression can be scaled-up by adding more UAA to the expression system. However, to ensure that this method works flawlessly, the expression system must be compatible with the unnatural amino acid and the protein-biosynthesis components specific to the required unnatural amino acid must be evolved, which is often a time-consuming process.

1.4.1. Bioorthogonal Chemistry in Synthetic Probe Labelling

In presence of many reactive groups with similar properties, bioorthogonal chemistry plays an important role for ensuring selectivity while labelling the biomolecules with synthetic probes. Bioorthogonal chemical reactions are orthogonal with respect to the natural biological reactions. Hence, a bioorthogonal functional group incorporated in a biomolecule serves as a unique reactive handle for the synthetic probe molecules bearing complementary bioorthogonal functionality [73]. Factors contributing to mutual orthogonality and bioorthogonality of chemical reactions include preferential affinity of reactants, differential rate of reaction in given environmental conditions, the equilibrium constant of the reactions and absence of competing functional groups. With some exceptions, most bioorthogonal reactions are either polar reactions (*i.e.*, reaction between electrophiles and nucleophiles) or cycloadditions (*i.e.*, reaction of π-bonded molecules to form σ-bonded cyclic products) [119].

The carbonyl oxygen of electrophile ketones and aldehyde groups reacts with α-effect enhanced amine nucleophiles such as aminooxy and hydrazide compounds to form physiologically-stable Schiff bases such as the oxime and hydrazone reaction products respectively [119, 120]. To exploit this reaction for labelling the biomolecules of interest (*e.g.*, proteins, glycans or glycoproteins), first, the carbonyl group is introduced in the biomolecule and then the reaction with amine nucleophile covalently attaches the synthetic probe to the biomolecule. While it is difficult to use this reaction within the cells or organisms (due to the presence of aldehyde-rich sugar molecules or ketone-containing hormones), this method has been used for biophysical-probe labelling of proteins *in-vitro* and on the cell-surface [120-124].

Unlike the ketone and aldehyde groups, which are bioorthogonal due to differential reactivity between target functional groups and other molecules, potential cross-reactivity with non-target biomolecules can be avoided by employing a truly non-natural functional group for synthetic probe labelling. Organic azides* are completely absent in biological systems [125]. Furthermore, azide is a very small functional group, hence introduces minimal perturbance to the target biomolecule. By Staudinger ligation, the azide group can readily react with triaryl phosphines (which is also a small molecule and absent from the living systems), without adversely affecting most of the other molecules or biological reactions [126, 127]. This

* Azide functional group ($-\overset{-}{N}-\overset{+}{N}\equiv N$) is highly reactive due to its 1,3 dipolar nature.

reaction has been used for targeting the biomolecules in mammalian cells [128, 129]. Staudinger ligation reaction mechanism has allowed the development of fluorogenic synthetic probes as well [130, 131].

In the presence of copper catalyst, the azide group can undergo rapid [3+2] cycloaddition reaction with terminal alkynes to form stable triazoles . This reaction, termed as the 'click-chemistry', is useful for biophysical studies *in-vitro* [132]. However, due to copper-induced toxicity, this reaction has limited applications *in-vivo*. The alkyne group of cyclooctyne-based molecules can react with azides via strain-promoted cycloadditions (SPAAC) or copper-free click chemistry, without the need of the copper [133]. This overcomes the copper-toxicity and enables the use of click chemistry in living systems [134, 135].

Apart from azides, other 1,3 dipolar functional groups, such as nitrones, nitrile oxides and diazo groups, can also react with cyclooctyne-derived strained alkynes. The rapid rate of such reactions allows labelling of biomolecules in living organisms [136]. Highly strained alkenes such as cyclopropane can also undergo dipolar cycloadditions with nitrile imines. However, owing to their high reactivity, most of these dipolar molecules are unstable in aqueous environment and hence are generated *in-situ* from their stable precursors. To generate the reactive species, the precursor molecules are photoactivated with UV light. Strained molecules such as trans-cyclooctene (TCO), norbornene or cyclopropene undergo inverse electron demand Diels-Alder (IED-DA) reactions with electron deficient tetrazines. Being the fastest bioorthogonal reaction, the TCO-tetrazine ligation is suitable for live animal imaging as well [137, 138].

1.5. GENETIC CODE EXPANSION – INSERTING UNNATURAL AMINO ACIDS IN PROTEINS

The polypeptide biosynthesis machinery of the cell introduces 20 proteinogenic amino acids in response to 61 triplet-codon signals present on the mRNA, while three codons are reserved for signaling the termination of polypeptide biosynthesis. The polypeptide elongation in the ribosome is at the heart of polypeptide biosynthesis. In short, the individual triplet-codons on the mRNA are recognized by the anticodon loop of the tRNA and the ribosome, along with the elongation factors, is responsible for recruiting the aminoacylated tRNA (*i.e.*, tRNA charged with the cognate amino acid or $tRNA^{AA}$), and for lowering the activation energy of peptide bond formation between the newly recruited amino acid and the growing polypeptide chain, thereby elongating the polypeptide chain. Limited amount of promiscuity is present during the elongation step of polypeptide biosynthesis, which allows the elongation factors and ribosome to detect several different kinds of aminoacylated tRNA (or $tRNA^{AA}$) [139]. It is because of this promiscuity that, when encountering a tRNA charged with an unnatural amino acid (or $tRNA^{UAA}$), the ribosome is able to introduce unnatural amino acids in the growing polypeptide chain. If the unnatural amino acids are biosynthetically incorporated in response to unique codons present on the mRNA, the process is termed as genetic code expansion. There are three key elements to this process, namely, 1) the presence of a $tRNA^{UAA}$, 2) the ability of this $tRNA^{UAA}$ to successfully recognize and base-pair with a unique codon on the mRNA, and 3) the ability of the ribosome to successfully accommodate this $tRNA^{UAA}$ for genetic code expansion. Consequently, these are the focal areas of research for genetic code expansion.

Genetic code expansion has been used for more than 30 years now [140]. The initial approach involved constructing Amber-suppressor tRNAs (based on the yeast $^{Phe}tRNA$) with the CUA anticodon loop and chemically aminoacylating this tRNA with unnatural amino acids, such as *p*-Nitrophenylalanine or *p*-fluorophenylalanine, for introducing these unnatural amino acids in response to the Amber stop codon (UAG codon on mRNA). However, at that time, this approach could only be used for *in-vitro* synthesis because of the lack of a method for biosynthetic production of $tRNA^{UAA}$ in living cells [116].

1.5.1. Genetic Code Expansion in Living Cells

In living cells, the aminoacylated tRNA can be generated biosynthetically by the aminoacyl tRNA synthetase (aaRS) enzyme. For each of the 20 amino acids, their isoacceptor tRNA (iso-tRNA) are charged by the cognate aaRS. To ensure that the correct amino acid is presented by the $tRNA^{AA}$ for polypeptide bond formation in the ribosome, the natural aminoacylation of tRNA has evolved as a highly specific bioprocess. This means that the aaRS recognizes only a specific kind of tRNA and amino acid for aminoacylation. The three-dimensional structure of the tRNA and the amino acid plays a role in this recognition. Since the anticodon loop of the tRNA is responsible for decoding the genetic message on the mRNA, this anticodon loop may also be involved during the selection of the tRNA by the aaRS. Nonetheless, the presence of a biosynthetic aminoacylation machinery allows for the possibility of introducing a similar but orthogonal mechanism into a living cell for unnatural amino acid incorporation *in-vivo*. However, for unambiguous site-specific unnatural amino acid incorporation, there are four important considerations for this approach – 1) the orthogonal aaRS must not recognize the native amino acids or tRNA molecules, 2) the orthogonal tRNA must not be aminoacylated by the native amino acids, 3) the orthogonal tRNA should be able to decode one and only one message on the mRNA, and 4) the codon assigned for the unnatural amino acid incorporation must recruit only the orthogonal $tRNA^{UAA}$ in the ribosome. By addressing these considerations several orthogonal translation systems (or OTS, *i.e.*, orthogonal aaRS/tRNA pairs) have been developed and, *in-vivo* genetic code expansion (GCE) has been made possible for prokaryotic as well eukaryotic cells.

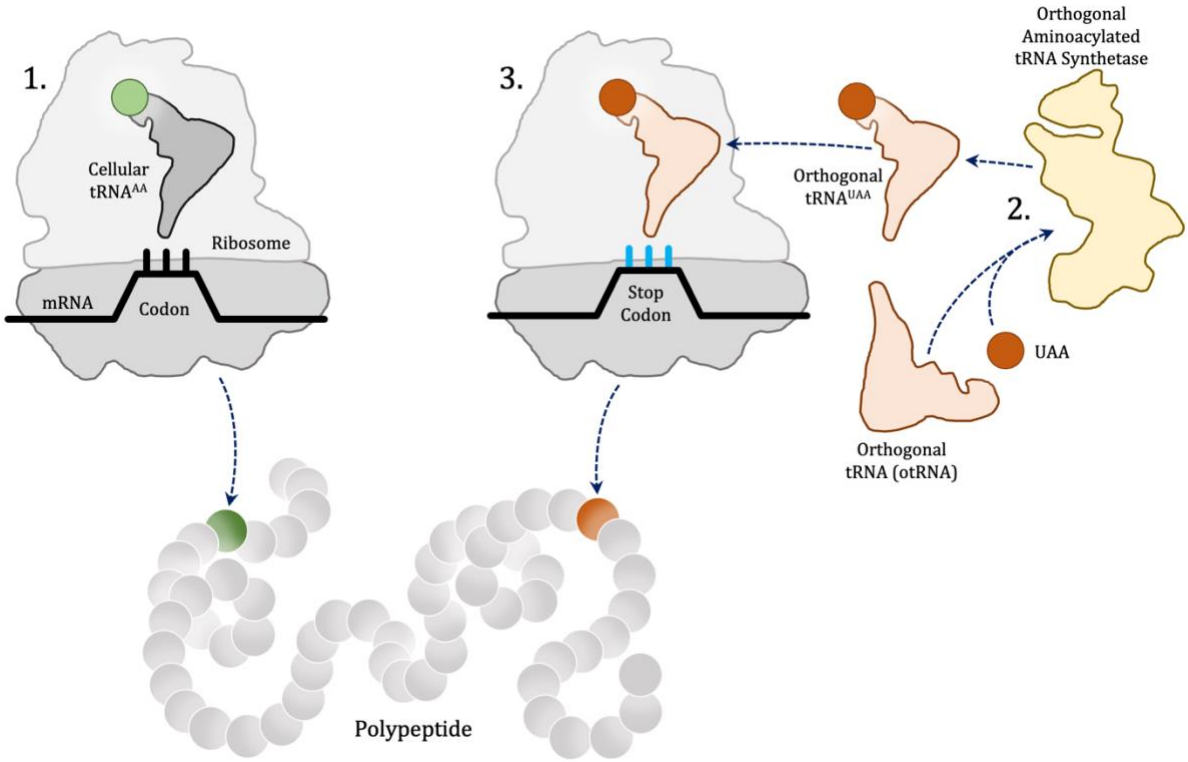


Figure 2: Genetic Code Expansion by Stop-Codon Suppression in Living Cells.
The translation machinery of the cell is 'tricked' for unnatural amino acid incorporation via genetic code expansion. For polypeptide biosynthesis, the ribosome facilitates decoding of the triplet-codons (present on the mRNA) by the aminoacylated tRNA ($tRNA^{AA}$) (1). Consequently, the amino acids are introduced in the growing polypeptide chain. For genetic code expansion, an orthogonal tRNA (otRNA) and an orthogonal tRNA synthetase (oRS) are introduced in the cell (2). The otRNA decodes one of the three stop-codons. In presence of the unnatural amino acid (UAA), the oRS can aminoacylate the otRNA with UAA, thereby creating the orthogonal aminoacylated tRNA ($otRNA^{UAA}$). As this tRNA is recognized by the ribosome, in response to the cognate stop-codon, the unnatural amino acid is introduced in the polypeptide chain (3).

1.5.2. Evolution of Orthogonal aaRS/tRNA Pairs for Genetic Code Expansion

Over the past two decades, various strategies have been adopted for genetic code expansion in living cells. The most common and validated approach is introduction of an orthogonal aaRS/tRNA pair. Such a pair

can either be evolved via reassignment of existing aaRS/tRNA pair or can be selected from an evolutionary divergent organism [141]. For the reassignment of aaRS/tRNA pair, the aaRS/tRNA pair is selected from the host organism, and is reassigned as an orthogonal aaRS/tRNA pair [142]. When selected from an evolutionary divergent organism, the aaRS/tRNA pair might already demonstrate orthogonality with respect to the host system, thereby simplifying the process of evolution [139]. For example, for bacterial cells, the orthogonal aaRS/tRNA pair can be imported either from archaea or eukaryotes [143]. Similarly, for eukaryotic cells the orthogonal aaRS/tRNA pair can be adopted either from archaea or bacteria [1].

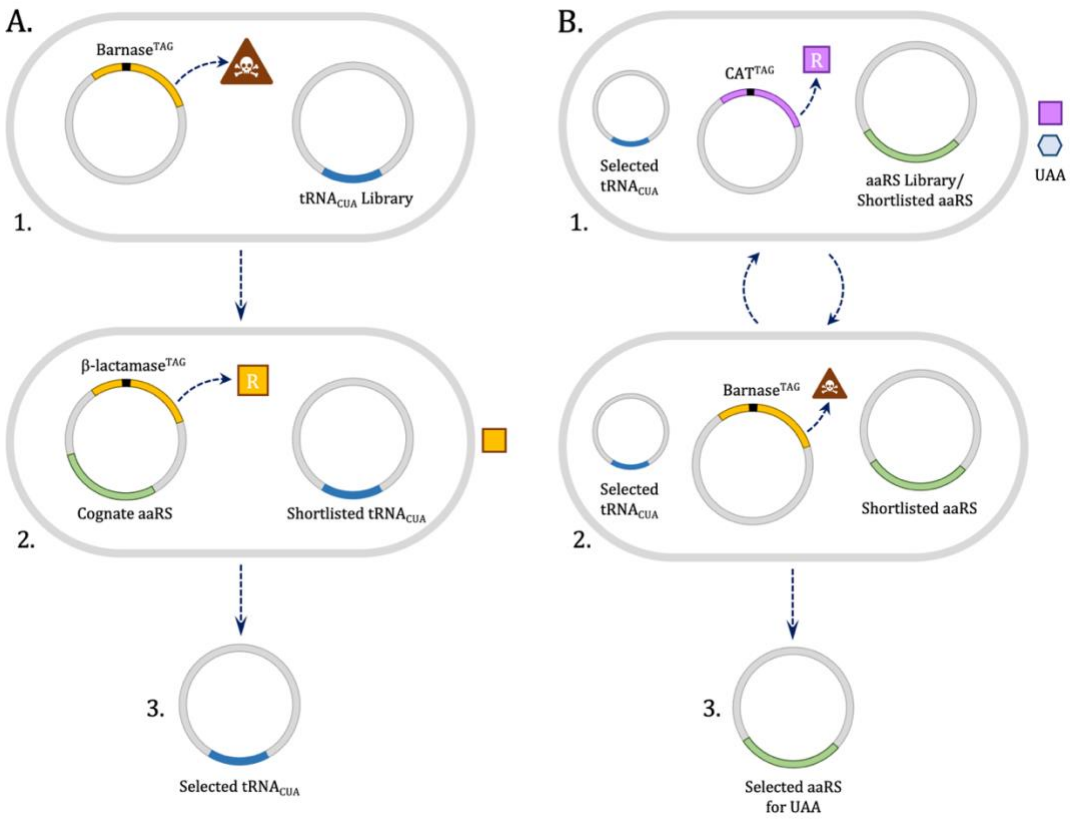


Figure 3: Synthetic Evolution of Orthogonal tRNA and Orthogonal aaRS in Bacterial Cells.
A. Synthetic evolution of the orthogonal tRNA (TAG-suppressing otRNA in this example). After introducing random mutations in selected region of the tRNA, the TAG-suppressing tRNA library is introduced in the bacteria. Simultaneously, a toxic gene (*e.g.*, *Barnase*) containing TAG stop codon is also introduced in the cells (1). When the cells are grown without the unnatural amino acid, if the tRNA is not orthogonal, toxic gene is expressed, killing the cells. After this negative selection step, the shortlisted tRNAs are introduced in cells expressing the cognate aminoacyl tRNA synthetase (aaRS). An antibiotic resistance gene (*e.g.*, for β-lactamase) with Tag stop-codon is also introduced in the cells (2). When the cells are grown in presence of the antibiotic, only those cells survive which contain the evolved tRNA that can be charged by the aaRS. These tRNAs are selected for evolving the orthogonal aaRS (3).
B. Synthetic evolution of the orthogonal aaRS. After introducing random mutations, the aaRS is introduced in cells expressing the selected tRNA. A second antibiotic resistance gene (*e.g.*, for Chloramphenicol acetyltransferase or CAT) containing the TAG stop-codon is also present in the cells. In presence of the antibiotic and the unnatural amino acids, only those cells survive that have a functional aaRS present (1). After this positive selection step, the shortlisted aaRS is introduced into the cells expressing the selected tRNA. A plasmid with a toxic gene (*e.g.*, *Barnase*) containing a TAG stop codon is also introduced in the cells (2). When cells are grown in absence of unnatural amino acid, only those cells survive where the aaRS is not able to recognize the cellular amino acids for aminoacylation of the selected tRNA. The positive and negative steps are repeated 2 or 3 times to select for the aaRS specific to the unnatural amino acid (3).

For the evolution of an orthogonal aaRS/tRNA pair, the first step is the selection of a blank-codon (or codon$_{BL}$) that does not code for any amino acid [139]. Most commonly, the naturally occurring stop codons, *i.e.*, Amber (TAG), Opal (TGA) and Ochre (TAA) codons, are designated as the blank codons. Afterwards, a suppressor-tRNA is developed that can decode the selected blank codon. While $_{Pyl}tRNA_{CUA}$ from methanogens such as *Methanosarcina barkeri* (*Mb*), *Methanosarcina mazei* (*Mm*) or *Desulfitobacterium hafniense* (*Dh*) naturally provide such suppressor-tRNA for the TAG codon suppression, for most suppressor-tRNA imported from a different domain of life, the anticodon loop must be modified [139].

If the selected suppressor-tRNA and its cognate aaRS are not already orthogonal to the host organism, by using directed evolution approach, orthogonal aaRS/tRNA pairs are evolved for the target unnatural amino acid [144]. First, a mutant library of the tRNA is created and by directed evolution, comprising of negative and positive selection steps, the orthogonality between the introduced the tRNA and the host organism is ensured. For selecting the orthogonal tRNA from the tRNA mutant library, a toxic gene (such as barnase gene for bacterial cells) containing the $codon_{BL}$ in its sequence is introduced in the cells along with the potential suppressor-tRNA. Through this negative selection step, those tRNA mutants are selected that demonstrate no $codon_{BL}$-suppression in presence of the native cellular machinery. The shortlisted tRNA mutants are then introduced into those cells that express the wild type cognate aaRS as well as an antibiotic resistance gene (such as β-lactamase gene for bacterial cells) containing $codon_{BL}$. The presence of the antibiotic kills all the cells containing non-functional tRNA. This positive selection step selects those tRNA mutants that can be aminoacylated by the imported aaRS and can suppress $codon_{BL}$.

After shortlisting the tRNA mutants, in the next selection step, the cognate aaRS is evolved for orthogonality to the host cell and specificity for the target unnatural amino acid. This involves structure-based mutagenesis and a similar directed-evolution strategy comprising of positive and negative selection steps. A library for the wild type aaRS is created to introduce random mutations in the amino-acid binding site. The aaRS mutants are then introduced in cells expressing the shortlisted suppressor-tRNA and a second antibiotic resistance gene (such as chloramphenicol acetyltransferase gene) containing $codon_{BL}$. When grown in presence of the second antibiotic and the target unnatural amino acid, due to positive selection, only those aaRS mutants are selected that can aminoacylate tRNA for $codon_{BL}$-suppression, either by using the unnatural amino acid or the cellular amino acids. The shortlisted aaRS are introduced in cells expressing the selected suppressor-tRNA and a toxic gene (such as barnase gene) containing $codon_{BL}$. When grown without the target unnatural amino acid, those aaRS mutants are eliminated that can recognize the cellular amino acids for aminoacylation. After two or three repetitions of this positive and negative selection step, an orthogonal aaRS for the target unnatural amino acid (or ^{UAA}RS) is obtained. While further evolution of the ^{UAA}RS can be similarly performed in the absence of the selected suppressor-tRNA, it is assumed that ^{UAA}RS does not significantly aminoacylate native tRNAs, otherwise it would lead to missense incorporation throughout proteome and cause cell death [139].

Due to multiple rounds of positive and negative selection steps, this approach is quite slow and, by design, it is targeted for generating one orthogonal aaRS/tRNA pair specific to the unnatural amino acid. Hence each unnatural amino acid may need its dedicated aaRS/tRNA pair. However, if two different unnatural amino acids have a similar overall shape, it is possible to evolve individual aaRS/tRNA pairs for each of these amino acids from a common aaRS/tRNA pair [145]. Furthermore, due to some promiscuity in the amino acid binding pocket of the evolved aaRS, it is sometimes possible to create polyspecific aaRS (for example, those derived from derived from $_{Pyl}RS$) which are able to recognize multiple unnatural amino acids [146]. For this reason, directed synthetic evolution of relatively small number of naturally occurring aaRS/tRNA pairs has provided orthogonal aaRS/tRNA pairs for many unnatural amino acids.

In theory, this approach of positive and negative selection, or its variants, can be used for all living organisms [147, 148]. However, this has been successfully used only for bacterial and yeast cells, since the limited transfection efficiency, slow growth rate and environmental sensitivity of mammalian cells and higher organisms makes this strategy impractical [1, 139, 144]. Nonetheless, if the evolved orthogonal aaRS/tRNA pair is orthogonal for the higher-organization organisms, they can be used for incorporating unnatural amino acids in those organisms as well [146, 149, 150]. Furthermore, if the interaction between ^{UAA}RS and $^{UAA}tRNA$ permits it, the anticodon loop of the $^{UAA}tRNA$ can be modified to recognize other blank-codons (other stop codons, reassigned codons, quadruplet codons or non-natural codons) as well for unnatural amino acid incorporation, thereby allowing the possibility of simultaneously incorporating multiple unnatural amino acids in proteins using mutually orthogonal aaRS/tRNA pairs [151-154].

1.5.3. Genetic Code Expansion in Bacterial Cells

In-vivo genetic code expansion for unnatural amino acid incorporation was first demonstrated in *E. coli*, when the $_{MjTyr}RS$ / $_{MjTyr}tRNA_{CUA}$ pair evolved from *Methanocaldococcus jannaschii* (*Mj*) was used to incorporate *O*-methyl-L-tyrosine unnatural amino acid in dihydrofolate reductase (DHFR) enzyme via Amber-codon suppression [155]. Since then, almost all developments related to genetic code expansion have occurred first in *E. coli* cells before they are adopted for eukaryotic cells or higher organisms. Using evolved orthogonal aaRS/tRNA pairs, over 150 unnatural amino acids have been incorporated in proteins via genetic code expansion in *E. coli* [156]. The OTS for genetic code expansion in *E. coli* have been derived from Leucyl aaRS/tRNA pair, Tyrosyl aaRS/tRNA pair, Pyrrolysyl aaRS/tRNA pair, Prolyl aaRS/tRNA pair, Glutamyl aaRS/tRNA pair and Tryptophanyl aaRS/tRNA pair to incorporate the homologues of these amino acids in proteins [145, 155, 157-160]. While *E. coli* has been the favorite bacterial host for unnatural amino acids incorporation in proteins, other bacteria (such as *Salmonella*) have also been used for genetic code expansion [161, 162].

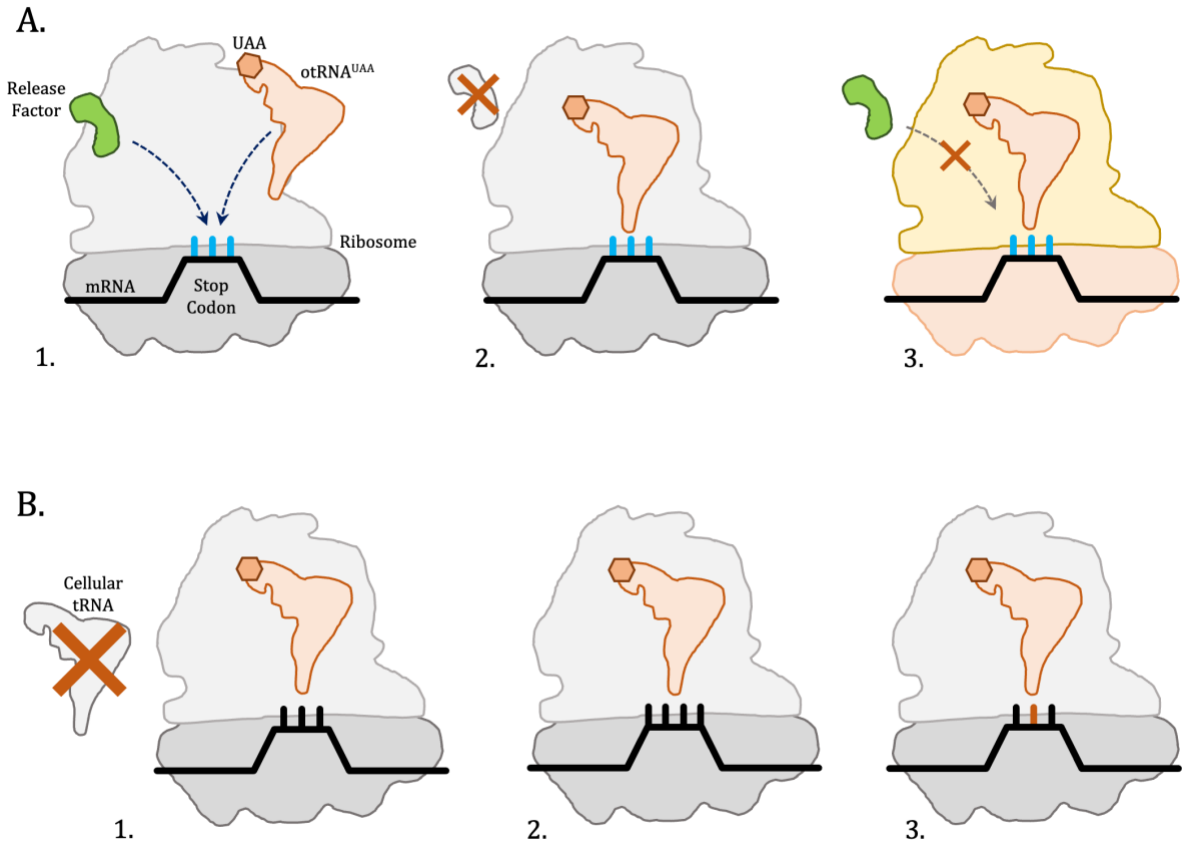


Figure 4: Developments in Genetic Code Expansion in Bacteria.
A. Strategies for improving stop-codon suppression. The orthogonal aminoacylated tRNA (otRNA^{UAA}) competes with release factors for the stop-codons (1). By knocking out the competing release factor, stop-codons suppression by the otRNA^{UAA} can be improved (2). Orthogonal ribosomes that do not recognize the competing release factors can also be used for improving the suppression efficiency (3).
B. Alternatives to stop-codons as blank-codons. By knocking out a cellular tRNA, the triplet-codon corresponding to that tRNA may be made available for the orthogonal tRNA (1). Quadruplet-codon decoding by the orthogonal tRNA (*i.e.*, frameshift-suppression) can be used for unnatural amino acid incorporation (2). Non-natural base pairs can increase the number of triplet-codons available by using more than four base-pairs of nucleotides (3).

When genetic code expansion (G.C.E.) occurs through stop-codon suppression, the G.C.E. machinery is competing against a more efficient cellular machinery for incorporation of unnatural amino acids in proteins. Several strategies have been used in bacterial expression systems to give a 'fighting chance' to the orthogonal machinery. The most straightforward one is reducing the competition with the release factors. Bacterial systems have two release factors, RF1 for TAA & TAG stop codons and RF2 for TAA & TGA stop codons [163]. As most genetic code expansion systems rely on Amber-suppression (TAG stop-codon) for

unnatural amino acid incorporation, knocking out the RF1 release factor or reducing the interaction between the RF1 and the ribosome could remove the ambiguous message for the Amber-codon (*i.e.*, unnatural amino acid incorporation or termination of polypeptide biosynthesis). By making the RF2 more efficient for TAA-codon mediated polypeptide termination, RF1-knockout *E. coli* strains could be created, which allow Amber-suppression mediated unnatural amino acid incorporation in proteins with a higher efficiency [163]. Synthetically evolved orthogonal ribosomes (Ribo-X) that only translate the 'orthogonal message' (due to decreased interactions with RF1) have also been developed for bacterial systems [164, 165].

Incorporation of multiple unnatural amino acids is limited by the number of available blank-codons. Since there are three stop codons, at-most two unnatural amnio acids can be incorporated using only the stop-codons (one stop-codon is needed to signal the termination of proteins). In search of new blank-codons, it has been possible to explore a few non-conventional approaches for genetic code expansion in *E. coli* cells. Such approaches include sense-codon compression, quadruplet-codon assignment, orthogonal ribosome evolution and use of non-natural base-pairs as codons [141]. Only two amino acids (MET and TRP) are coded by unique triplet-codons, rest are coded by more than one codons. While the redundancy in codons could have evolved as a mechanism to control the rate and amount of protein expression (thereby making it important for the survival of organisms), attempts have been made to 'free-up' some degenerate codons to be used as blank-codons during genetic code expansion [152, 166, 167]. By using quadruplet-codons, the available number of genetic codes can be increased to 256 (or 4^4) (as compared to 64 unique codes for triplet-codons). By modifying the otRNA anticodon loop and/or by synthetically evolving ribosomes (Ribo-Q), quadruplet codons have also been used as blank-codons for genetic code expansion in *E. coli* cells [153, 158, 168-170]. By using unnatural nucleotide bases (UNB), it has been possible to use more than four letters in the 'synthetic' genetic codons [171, 172]. Organisms harboring the unnatural base pairs have been created and have even been used for unnatural amino acid incorporation in proteins [154, 173].

1.5.4. Genetic Code Expansion in Yeast Cells

For eukaryotic protein production, yeast expression systems have several advantages such as fast growth, ease of genetic manipulation, high biomass, scalable fermentation, pathogen-free production and post-translational modifications [174]. Consequently, yeast-based expression systems have been used heavily for academic research as well as industrial applications. Genetic code expansion in yeast cells, particularly in *Saccharomyces cerevisiae* (*S. cerevisiae*), has been used for unnatural amino acid incorporation in proteins, as well as for evolving orthogonal aaRS/tRNA pairs for mammalian expression systems [1].

Bacterial aaRS/tRNA pairs that are orthogonal in yeast can be used for evolving orthogonal aaRS/tRNA pairs for yeast expression systems. For yeast cells as well, a directed evolution strategy with positive and negative selection steps can be used to evolve orthogonal aaRS/tRNA pairs specific to target amino acids. Usually this is done in a uracil-auxotrophic strain *S. cerevisiae* cells [1]. The blank codon ($codon_{BL}$) is added in the sequence of GAL4 transcriptional activator, which activates *ura3* gene thereby initiating uracil synthesis. For positive selection, the yeast cells are grown in the absence of uracil and presence of the unnatural amino acid, to select for those mutants of aaRS that can aminoacylate the suppressor tRNA for $codon_{BL}$. For negative selection, the yeast cells are grown without the unnatural amino acid but in presence of 5-fluoroorotic acid, which is converted to a toxic product by URA3 protein. Hence, those cells are killed where aminoacylation of the suppressor tRNA is independent of the unnatural amino acids.

Using this strategy, $_{EcTyr}RS$ / $_{EcTyr}tRNA$ derived aaRS/tRNA pairs could be used in *S. cerevisiae* cells to incorporate acetyl-L-phenylalanine, p-benzoyl-L-phenylalanine, p-azido-L-phenylalanine, o-methyl-L-tyrosine and p-iodo-L-tyrosine in proteins [1]. Similarly, $_{EcLeu}RS$ / $_{EcLeu}tRNA$ derived aaRS/tRNA pairs could also be used in *S. cerevisiae* to incorporate α-aminocaprylic acid, O-methyl tyrosine, o-nitrobenzyl cysteine and dansylalanine in proteins [175, 176]. Since the $_{Pyl}RS$ / $_{Pyl}tRNA$ pairs from *Methanosarcina barkeri* (*Mb*) and *Methanosarcina mazei* (*Mm*) are orthogonal for both bacterial and eukaryotic systems, it is possible to evolve an orthogonal aaRS/tRNA pair in *E. coli* and use it in eukaryotic cells, including yeast cells. For example, $_{MbPyl}RS$ / $_{MbPyl}tRNA$ pair could be used in *S. cerevisiae* for incorporation of *N*ε-acetyl-L-lysine, trifluoroacetyl-L-lysine, *N*ε-[(1-(6-nitrobenzo[d][1,3]dioxol-5-yl)ethoxy)carbonyl]-L-lysine, *N*ε-[(2-propynyloxy)carbonyl]-L-lysine and Nε-[(2-(3-methyl-3H-diazirin-3-yl)ethoxy)carbonyl]-L-lysine in proteins [177].

Due to fundamental differences in the transcription/translation machinery of bacterial and eukaryotic cells, OTS derived from prokaryotic systems pose some compatibility issues in eukaryotic systems [163, 178]. For example, the otRNA derived from bacterial expression systems may not efficiently express in eukaryotic expression systems due to lack of promoter elements. As many of these issues and their resolution are common between yeast and mammalian cells, they will be discussed later.

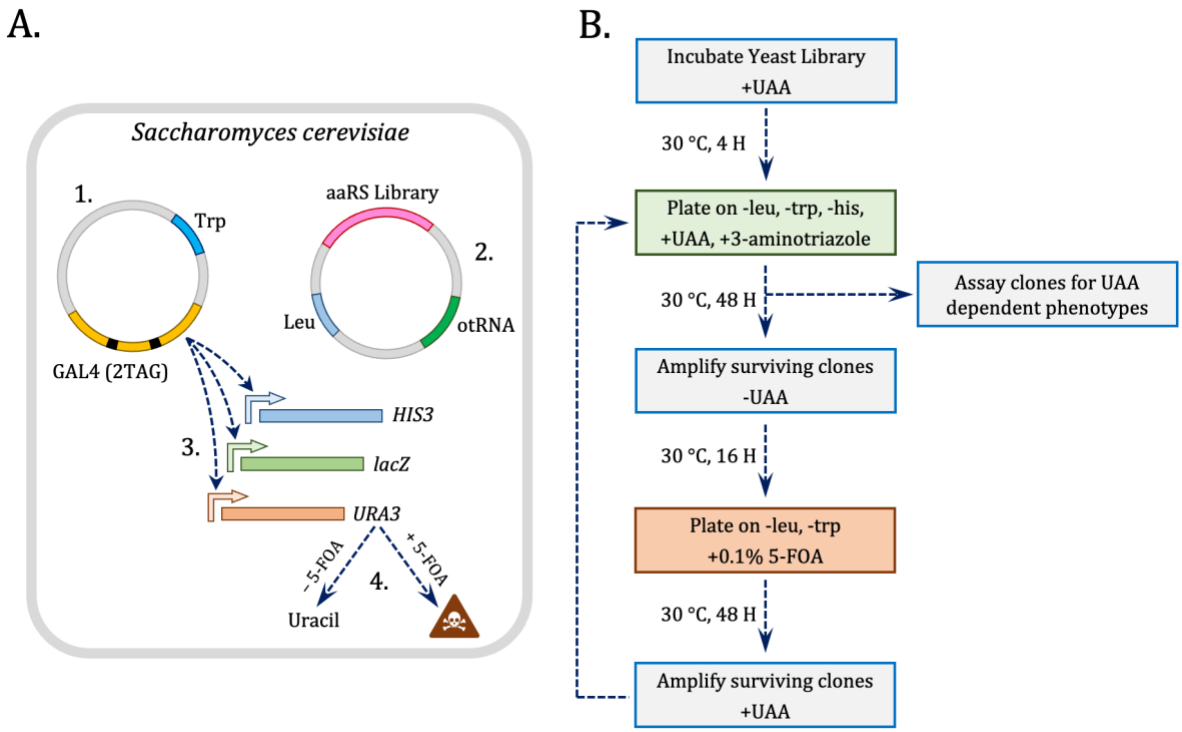


Figure 5: Synthetic Evolution of Orthogonal aaRS in Yeast.
A. Schematics and principle of positive and negative selection for orthogonal aaRS evolution in uracil-auxotrophic yeast cells (as reported in [1]). GAL4 gene, containing two stop codons, is introduced in yeast cells (1). Simultaneously, otRNA genes and the aaRS library is also introduced (2). After successful stop-codon suppression by the oRS and otRNA, the expression of GAL4 promotes the expression of *HIS3*, *lacZ* and *URA3* genes (3). In absence of 5-fluoroorotic acid (5-FOA), uracil is produced by URA3 protein, but in its presence a toxic product is generated that kills the cells (4).
B. Rounds of positive and negative selection to synthetically evolve an orthogonal aaRS to recognize an amino acid. The positive selection steps are in green and the negative selection steps are in red.

1.5.5. Unnatural Amino Acid Incorporation in Mammalian Cells

Mammalian expression systems for protein biosynthesis are more complex as compared to bacterial or yeast expression systems. However, they present the significant advantage of protein characterization in physiologically relevant environment. Therefore, many orthogonal aaRS/tRNA pairs have been evolved for use in mammalian systems. However, since high-throughput evolution of aaRS/tRNA pairs in mammalian cells is not practical, the orthogonal aaRS/tRNA pairs for mammalian cells have to be evolved in *E. coli* or *S. cerevisiae* (as previously discussed) and then shuttled into mammalian cells [142, 146]. To date, most, if not all, orthogonal aaRS/tRNA pairs for mammalian cells have been evolved using one of the four aaRS, namely, ${}_{EcTyr}RS$, ${}_{EcLeu}RS$, ${}_{Pyl}RS$ or ${}_{EcTrp}RS$ [142, 145, 179, 180]. Consequently, they are able to incorporate the homologues of tyrosine, leucine, pyrrolysine and tryptophan respectively in mammalian cells. So far, more than 100 unnatural amino acids have been incorporated in proteins expressed in mammalian cells [181].

The first successful attempt to incorporate tyrosine homologues in mammalian cells was involved the use of ${}_{EcTyr}RS$ / ${}_{BsTyr}tRNA$ pair for the incorporation of 3-iodo-L-tyrosine via amber codon suppression [179]. In this study, the ${}_{Tyr}tRNA$ from *Bacillus stearothermophilus* (*Bs*) was used because it had the natural promoter for expression in mammalian cells. However, it was shown that ${}_{EcTyr}RS$ / ${}_{EcTyr}tRNA$ pair is

orthogonal in yeast cells therefore it could be evolved in the yeast cells for the incorporation of unnatural amino acids in mammalian cells as well [1, 146, 182]. Furthermore, it was observed that the evolved $^{OMeY}_{EcTyr}RS$ (*i.e.*, the $_{EcTyr}RS$ evolved for o-methyl-L-tyrosine, and containing mutations Y37V, D182S, F183M and D256R) was polyspecific in nature, which could be used to incorporate many tyrosine homologues in mammalian cells. For introduction of tyrosine homologues in mammalian cells, a baculoviral-delivery compatible expression vector (based on pAcBac) was created [146]. In this expression vector, CMV-IE promoter was used for the expression of evolved $^{OMeY}_{EcTyr}RS$. However, due to the fear of possible homologous recombination during baculoviral amplification, rather than using multiple copies of $_{EcTyr}tRNA$ alone, two copies each of $_{EcTyr}tRNA$ as well as $_{BsTyr}tRNA$ genes (both kinds of tRNA are recognized by $_{EcTyr}RS$), promoted by U6 and H1 promoters were used to introduce the Amber suppressor-tRNAs in the mammalian cells. Using this setup, 9 different unnatural amino acids, such as p-azidophenylalanine (AzPhe) and p-propargyloxyphenylalanine (PrPhe), were introduced in proteins expressed in mammalian cells. This baculoviral delivery system was shown to be useful for many different kinds of mammalian cells (CHO, 3T3, HeLa, mouse embryonic fibroblast, rat cardiac fibroblast, cultured neurons and mouse embryonic stem cells).

The crystal structure of $_{Leu}RS$ from *Thermus thermophilus* showed that the leucine binding site the enzyme is a large cavity with side chains and without backbone elements [175]. Due to this, it was anticipated that the homologous $_{EcLeu}RS$ could tolerate unnatural amino acids, hence it could serve as an orthogonal aaRS for yeast cells. Furthermore, $_{EcLeu}tRNA_{CUA}$ was shown to be orthogonal in *S. cerevisiae* [175]. Hence by using previously described directed evolution methods, the $_{EcLeu}RS$ / $_{EcLeu}tRNA$ pair was evolved to introduce fluorescent unnatural amino acids, such as 2-amino- 3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (dansylalanine or DanAla) and 3-(6-acetylnaphthalen-2-ylamino)-2-aminopropanoic acid (ANAP) in *S. cerevisiae* cells in response to Amber stop codon [176, 183, 184]. The $^{DanAla}_{EcLeu}RS$ pair evolved for the incorporation of fluorescent unnatural amino acid dansylalanine in yeast contained six mutations (M40A, L41N, T252A, Y499I, Y527G and H537T) [176]. This oRS was used with the $_{EcLeu5}tRNA_{CUA}$ to incorporate dansylalanine in proteins expressed in yeast cells. Later, this pair was adopted to genetically incorporate dansylalanine in proteins expressed in mammalian cells as well [185]. The unnatural amino acid ANAP could serve as an environment sensing fluorophore. With respect to wild type $_{EcLeu}RS$, the $^{ANAP}_{EcLeu}RS$ (the one evolved for ANAP) contains nine mutations (L38F, M40G, L41P, Y499G, Y527A, H537T, L538C, F541S and A560V) [184]. Due to so many mutations for the functionally relevant aaRS, the evolution of this aaRS was quite challenging. For introducing the aaRS/tRNA pair in mammalian cells, the $^{ANAP}_{EcLeu}RS$ was expressed constitutively using a CMV promoter [186]. The $_{EcLeu}tRNA_{CUA}$ gene was modified to contain H1 promoter upstream and a poly(A) tail downstream to the tRNA gene. Eight tandem repeats of this expression cluster of H1-tRNA-poly(A) were introduced to increase the expression on tRNA [186]. Using 10 – 500 μM ANAP in the culture media of the CHO, HEK293 and HeLa cells, ANAP was demonstrated to be a good fluorophore for *in-vivo* imaging using scanning confocal microscopy and two-photon excitation fluorescence microscopy. Later, by using ANAP and GFP, a genetically encoded FRET pair was also developed for mammalian cells [187].

$_{Pyl}RS$ based aaRS/tRNA pairs for mammalian cells are usually derived from *Methanosarcina mazei* (*Mm*) and *Methanosarcina barkeri* (*Mb*) [165]. Since $_{Pyl}RS$ / $_{Pyl}tRNA$ pair is naturally orthogonal in both bacteria and eukaryotic cells, the derivatives of these aaRS/tRNA pairs can be evolved in bacterial cells and then easily migrated to eukaryotic cells. Due to this convenience, to date, $_{Pyl}RS$ based aaRS/tRNA pairs are the most widely used for genetic code expansion [141]. For introducing $_{Pyl}RS$ based lysine-homologue incorporation system in mammalian cells, the wild type $_{MmPyl}RS$ belonging to *M mazei* could be used as an orthogonal aaRS in mammalian cells to incorporate Nε-tert-butyloxycarbonyl-L-lysine (BOC-Lysine) in proteins [149]. The wild type $_{MmPyl}RS$ gene was expressed using CMV promoter using the pCDNA3.1/Zeo(+) plasmid. The $_{MmPyl}tRNA$ could be expressed by using human $_{Val}tRNA$ sequence, U6 promoter or T7 promoter at the 5' end of the $_{MmPyl}tRNA$ sequence. In each case, a terminator sequence was also present. Of these three conditions, U6 promoter gave the best results. By using nine tandem copies of the U6 and $_{MmPyl}tRNA$, the amber suppression was increased by a factor of 2. Furthermore, the wild type $_{MmPyl}RS$ could be evolved in *E. coli* cells to introduce Z-Lysine (which is larger than pyrrolysine) in proteins expressed in mammalian cells [149]. Polyspecific wild type $_{MbPyl}RS$ from *Methanosarcina barkeri* (*Mb*) could also be used for lysine homologue incorporation in mammalian cells, using the baculoviral gene delivery system [146]. For this expression vector, the wild-type $_{MbPyl}RS$ gene was expressed in mammalian cells using the CMV-IE promoter, while two different kinds of Amber suppressor-tRNA (both

compatible to the wild-type $_{MbPyl}RS$) from *M. mazei* and *Desulfitobacterium hafniense* (*Dh*) were expressed using H1 and U6 promoters. Using such a setup, 5 different lysine-homologues could be incorporated in proteins expressed in mammalian cells, including click-chemistry compatible unnatural amino acids azidolysine (N_3Lys) and propargyloxylysine (PrK) [146]. By modifying the anticodon loop of the tRNA, unnatural amino acids could be incorporated by TAA and TGA stop-codon suppression as well [188]. This paved the way for modifying the anticodon loop of the $_{MbPyl}tRNA_{CUA}$ to UCCU, and evolving a tRNA for quadruplet-codon (AGGA) suppression in mammalian cells [189]. This tRNA was used with $_{MmPyl}^{Y384F}RS$ ($_{MmPyl}RS$ having Y384F mutation) for incorporating BOC-Lysine unnatural amino acid in proteins using HEK293T cells [189].

Owing to their facile nature of evolution of orthogonal aaRS/tRNA pairs, bacterial cells are preferred over yeast cells as host systems for evolution of such pairs. However, aaRS/tRNA pair to be evolved must be orthogonal to the host system. Contrary to this, tryptophanyl-aaRS/tRNA based pair (*i.e.*, $_{Trp}RS$ /$_{Trp}tRNA$) for unnatural amino acid incorporation in mammalian cells has been derived from $_{EcTrp}RS$ /$_{EcTrp}tRNA$ as well as evolved in *E. coli* cells [142]. To allow the evolution of $_{EcTrp}RS$ /$_{EcTrp}tRNA$ pair in *E. coli* cells, first, the *trpS* and *trpT* genes (respectively encoding the endogenous $_{Trp}RS$ and $_{Trp}tRNA$) of *E. coli* (EcNR1 strain) were removed and simultaneously $_{ScTrp}RS$ /$_{ScTrp}tRNA$ pair from *S. cerevisiae*, which was shown to work in *E. coli* cells, was introduced [160]. By doing this, ATMW1 strain of *E. coli* was created, in which the endogenous $_{Trp}RS$ /$_{Trp}tRNA$ pair was replaced by the one from *S. cerevisiae*. For the evolution of $_{EcTrp}RS$ /$_{EcTrp}tRNA$ pair, the TGA-suppressor tRNA (*i.e.*, $_{EcTrp}tRNA_{UCA}$) was preferred over the more conventional TAG-suppressor tRNA because $_{EcTrp}tRNA_{CUA}$ could be aminoacylated by native $_{EcGln}RS$, thereby rendering the TAG-suppressor tRNA as non-orthogonal. After the evolution of orthogonal tRNA, $_{EcTrp}RS$ was evolved by positive and negative selection steps to identify a tryptophan homologue 5-hydroxytryptophan (5HTP). It was found that h9 (S8A, V144S and V146A mutations) and h14 (S8A, V144G and V146C mutations) variants of evolved $_{EcTrp}RS$ were polyspecific for unnatural amino acids. For introduction in mammalian cells, CMV promoter was used for the expression of the evolved $_{EcTrp}RS$ and after modifying $_{EcTrp}tRNA$ for TAG-suppression (*i.e.*, by using CUA anticodon), U6 promoter was used for the expression of the evolved $_{EcTrp}tRNA$. After introduction of this aaRS/tRNA pair in mammalian cells using the previously reported pAcBac1 plasmid, it was possible to introduce bioconjugation compatible unnatural amino acids 5-azidotryptophan (5-AzW) and 5-propargyloxytryptophan (5-PrW) into the test protein eGFP^{39TAG}.

1.5.6. Improving Genetic Code Expansion in Mammalian Systems

To ensure orthogonality, the aaRS/tRNA pairs used for genetic code expansion in mammalian systems are borrowed from prokaryotic or archaeal life forms. However, due to the fundamental differences between the transcription/translation machinery of prokaryotic and eukaryotic cells, some compatibility challenges may arise when prokaryotic genes are transferred in eukaryotic systems.

One such challenge is expression of functional orthogonal tRNA in eukaryotic cells [163]. While the orthogonal aaRS gene can be expressed by codon optimization and using a mammalian promoter, for the transcription of tRNA in eukaryotic cells promoter elements A-box and B-box are required. As these elements are absent in prokaryotic cells, the orthogonal tRNA evolved in bacterial cells must be modified for expression in eukaryotic cells. Initially, this was addressed by using prokaryotic tRNA that naturally contain the eukaryotic promoter sequence [179]. Later, by using type-3 Pol III promoters (such as H1, U6, snRNA, 7SK or MRP/7-2 for mammalian cells and RPR1 or SNR52 for yeast cells) and 3' flanking sequence (from human $_{fMet}tRNA$), a more general strategy could be developed for expression of prokaryotic tRNA in eukaryotic cells [179, 185].

Genetic code expansion in mammalian cells has been limited by the evolution of the orthogonal aaRS. For practical reasons, orthogonal aaRS/tRNA pairs are not evolved in mammalian cells, but rather rely on bacterial and yeast expression systems. However, for evolution of an orthogonal aaRS/tRNA pair, the starting aaRS/tRNA pair should also be orthogonal to the host systems. For this reason, aaRS/tRNA pairs derived from bacteria are evolved in yeast cells and aaRS/tRNA pairs derived from archaea are evolved in bacterial cells (or yeast cells). However, since directed evolution in bacteria is faster than that in yeast, most

mammalian G.C.E. systems are derived from aaRS/tRNA pairs that can be evolved in bacteria (particularly the ${}_{Pyl}RS\ /{}_{Pyl}tRNA$ pair). To resolve this, endogenous aaRS/tRNA pairs of bacterial cells can be replaced with the respective yeast aaRS/tRNA pairs to increase the number of aaRS/tRNA pairs that are orthogonal to both bacterial and eukaryotic expression systems [142].

Just as in bacterial cells, the G.C.E. machinery in mammalian cells also has to compete with the natural polypeptide elongation/termination machinery. By reducing this competition using engineered release factors, unnatural amino acid incorporation can be improved [190]. A low intracellular concentration of unnatural amino acid may also be restrictive for efficient unnatural amino acid incorporation. For addressing this, 'protected unnatural amino acids' have been used that can be de-protected by the cellular enzymes [163]. Blocking or knocking out the nonsense-mediated mRNA decay allows the gene for the target-protein (usually containing a stop-codon) to remain longer in the cells, thereby improves the production of UAA labelled target-protein [163, 165].

1.5.7. Genetic Code Expansion in Higher Complexity and Multicellular Living Organisms

Genetic code expansion in mammalian cells paves the way for genetic code expansion in higher eukaryotic organisms as well. However, for multicellular organisms, OTS gene delivery, continuous maintenance of these genes in the cells, nonsense-mediated decay (NMD) of the target-protein gene and sequestering of unnatural amino acids could be a challenge [191-193]. By using lentivirus to deliver the genes for orthogonal aaRS/tRNA (derived from ${}_{EcTyr}RS\ /{}_{EcTyr}tRNA$ pair) into neural stem cells (HCN-A94), the OTS genes could be integrated into the host genome, thereby allowing long-term genetic incorporation of unnatural amino acids into proteins [191].

Caenorhabditis elegans (*C. elegans*) was the first multicellular organism where genetic code expansion was successfully demonstrated [192]. Contrary to the lentiviral gene delivery approach, the OTS was delivered in the organism via injection and was maintained in the organism by using hygromycin B as the selection antibiotic. Additionally, *smg-2(e2008)* worms, which are deficient in the NMD mechanism, were used to prevent the loss of the target-protein gene. By using this approach, ${}_{MmPyl}RS\ /{}_{MmPyl}tRNA_{CUA}$ pair was used to introduce propargyloxylysine and BOC-lysine unnatural amino acids into the proteins expressed in different tissues of *C. elegans*. Soon, orthogonal aaRS/tRNA pairs, derived from ${}_{EcLeu}RS\ /{}_{EcLeu}tRNA$ and ${}_{EcTyr}RS\ /{}_{EcTyr}tRNA$ pairs, were also introduced (separately) in the organism for Amber-suppression mediated incorporation unnatural amino acids into proteins [193]. Contrary to the antibiotic selection approach, the OTS genes could be integrated into the chromosome via γ-irradiation, thereby ensuring successful transmission of these genes [193]. It was also demonstrated that dipeptides of unnatural amino acids could be used for improving the intracellular concentration of unnatural amino acids in multicellular organisms, thereby facilitating better UAA incorporation.

The ${}_{MmPyl}RS\ /{}_{MmPyl}tRNA_{CUA}$ pair has also been introduced into *Drosophila melanogaster*, where tissue specific incorporation of unnatural amino acids could be demonstrated [194]. An orthogonal aaRS/tRNA pair derived from the ${}_{MbPyl}RS\ /{}_{MbPyl}tRNA_{CUA}$ pair and evolved for Nε-acryllysine (AcrK) could incorporate unnatural amino acids for fluorogenic photoclick reactions in proteins expressed in bacterial cells, mammalian cells and plants (*Arabidopsis thaliana*) [195]. For *Mus musculus* (*M. musculus*), genetic code expansion was first performed with ${}^{Cmn}_{EcLeu}RS\ /{}_{Ecleu}tRNA_{CUA}$ pair (Cmn = 4,5-dimethoxy-2-nitrobenzyl-cysteine) via transient introduction of the OTS genes in mouse neocortex and diencephalon [196]. This was also the first example of genetic code expansion in a mammalian organism. Later, by integrating ${}^{AzF}_{EcTyr}RS\ /{}_{BsTyr}tRNA_{CUA}$ pair (evolved for *p*-azido-L-phenylalanine (AzF) incorporation) into the host genome, transgenic lines for *M. musculus* and *Danio rerio* (*D. rerio*) could be created for genetic code expansion *in-vivo* [197]. Recently, ${}_{Pyl}RS\ /{}_{Pyl}tRNA_{CUA}$ based orthogonal aaRS/tRNA pairs have been used for genetic code expansion in *M. musculus* [198-200].

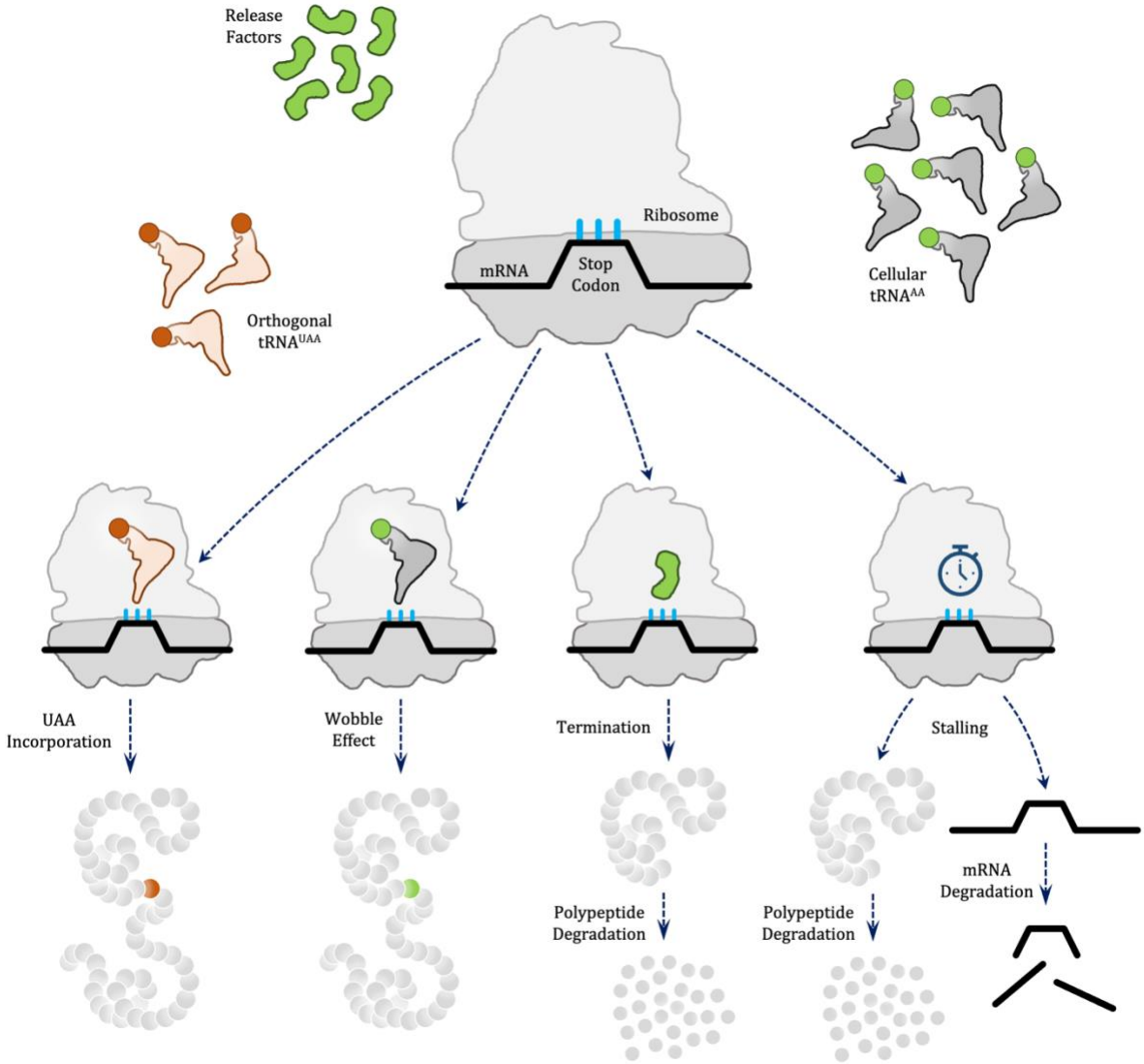


Figure 6: Possible Outcomes After Ribosome Encounters a Stop Codon During Polypeptide Synthesis.
The cell contains several components that compete for the stop codon. During genetic code expansion, as the ribosome encounters the stop codon, the orthogonal aminoacylated tRNA may introduce the unnatural amino acid in the polypeptide chain, or due to the wobble effect, a cellular amino acid may be introduced. Alternatively, the release factors may cause the termination of the polypeptide chain, eventually leading to polypeptide degradation. If the ribosome stalls at the stop-codon, the ribosome may be rescued by polypeptide termination and mRNA degradation. In this way the position of the stop-codon on the mRNA can determine the eventual fate of the polypeptide.

1.5.8. Protein Characterization Using Genetic Code Expansion

Genetic code expansion (G.C.E.) has been used for biophysical, biochemical and physiological characterization of proteins [201]. Unnatural amino acids introduced via G.C.E. have been used to site-specifically introduce post-translational modifications on proteins such as phosphorylation [202-208], ubiquitination [209-211], Nε-acetylation [212, 213] and sulfation [214]. They have also been used for *in-situ* photoactivation of proteins [215, 216], regulation of enzyme activity [217-219], controlling the folding or localization [145, 220, 221], or even for optical control of DNA recombination and gene expression *in-vivo* [222, 223]. Site-specifically introduced photocrosslinking agents can allow the spatiotemporal mapping of target-protein's interacting partners [224, 225]. Fluorescent amino acids or fluorescent probes introduced via unnatural amino acids have been used for high resolution fluorescence imaging, microenvironment detection or for studying protein dynamics [186, 187, 226, 227]. Unnatural amino acids can also be used to introduce NMR probes [65, 228] or EPR probes [229-231] on protein molecules. To suit

the requirements of the research question, new aaRS/tRNA pairs can be evolved for genetic incorporation of the targeted unnatural amino acid [205, 207].

1.6. AIMS OF THE THESIS

The starting point of this thesis is to site-specifically introduce two different fluorophores in proteins for fluorescence-based single-molecule biophysical studies. To ensure that our proteins of interest would undergo proper maturation, we have deemed mammalian protein expression systems to be the most suitable for our needs. At the onset of this project, by using the genetic code expansion in mammalian expression systems, I aim to site-specifically introduce bioorthogonal handles in proteins via unnatural amino acids (UAA) incorporation.

Aim 1: Assimilation of Existing Genetic Code Expansion Systems

I aim to develop a strategy for assimilating existing genetic code expansion systems that can be used for incorporating unnatural amino acids into proteins using mammalian expression systems.

Aim 2: Screening Assay for Quantification of UAA Incorporation

I aim to develop a screening assay for quantifying genetic incorporation of unnatural amino acids into proteins.

Aim 3: Site-specific Introduction of Two Fluorophores in Proteins

I aim to incorporate two different unnatural amino acids into proteins to facilitate chemical labelling of proteins by fluorophores.

Aim 4: Introduction of FRET Pair in Multienzymes

I aim to introduce two fluorophores in multienzymes for single molecule FRET studies.

1.7. DECLARATION OF OWN PROJECT CONTRIBUTION

I have carried out all parts of this work by myself. To be specific, I have created Gateway™ cloning compatible plasmid vectors, namely MX01, UCAP, UCZP, GIDC, GIDK, GIDS and GD54. I have created mammalian expression vectors (21 'G.C.E. Plasmids', 55 'G.E.M.S. Plasmids' and 3 Control-Plasmids) for performing call-based screening experiments. I have optimized large scale DNA purification for producing transfection grade DNA. I have purified the created mammalian expression plasmids for transient transfection of mammalian cells. I have optimized the mammalian transient transfection for unnatural amino acid incorporation in proteins. I have created scripts and pipelines for workflow-automation of cell-based assays using a laboratory automation setup. I have performed screening experiments for characterizing the efficiency of 21 acquired genetic code expansion systems and for understanding the site-specific incorporation of unnatural amino acids in enhanced green fluorescence protein (eGFP). I have created image-analysis pipelines for the CellProfiler software for analysis of the fluorescence images using high performance computing. I have created data-analysis and graphical-representation scripts for analyzing the output of image analysis using high performance computing.

2. MATERIALS AND METHODS

2.1. LIST OF CHEMICALS AND CONSUMABLES

2.1.1. Living Cell Cultures

HEK293 cells	ATCC Cat# CRL-1573 | additional glycerol stocks prepared inhouse.
NEB® 10-beta *E. coli* cells	NEB Cat# C2020K | electrocompetent cells prepared inhouse.
One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* cells	ThermoFisher Cat# A10460 | electrocompetent cells prepared inhouse.
Expi293F™ cells	ThermoFisher Cat# A14527 | additional glycerol stocks prepared inhouse.
One Shot™ PIR1 *E. coli* cells	ThermoFisher Cat# C101010 | electrocompetent cells prepared inhouse.

2.1.2. Cell-Culture Media

DMEM	Sigma Cat# D6429-500ML
Expi293™ Expression Medium	ThermoFisher Cat# A1435101
Fetal Bovine Serum	Gibco Cat# 10270106
Dulbecco's PBS	Sigma Cat# D8537-500ML
Phosphate Buffer Saline	Sigma Cat# D8537-500ML
SOB Medium	Carl Roth Cat# AE27.1
Terrific Broth	LLG-Labware Cat# 6771102

2.1.3. Chemicals

Acetic Acid Glacial (CH_3COOH)	CAS# 64-19-7	Merck Cat# 1.00063.1011
Acridine Orange ($C_{17}H_{19}N_3$)	CAS# 10127-02-3	Sigma-Aldrich Cat# 115931
Agarose	CAS# 9012-36-6	Sigma Cat# A9539-100G
Bestatin ($C_{16}H_{24}N_2O_4$)	CAS# 58-970-76-6	Carl Roth Cat# 2937.2
Betaine ($C_5H_{11}NO_2$)	CAS# 107-43-7	Sigma Cat# B0300-5VL
Calcium Chloride ($CaCl_2$)	CAS# 10043-52-4	AppliChem Cat # 131232.1211
Carbenicillin Disodium Salt ($C_{17}H_{16}N_2O_6S•2Na$)	CAS# 4800-94-6	GoldBio Cat# C-103-5
Chloramphenicol ($C_{11}H_{12}Cl_2N_2O_5$)	CAS# 56-75-7	Carl Roth Cat# 3886.3
DAPI ($C_{16}H_{15}N_5$)	CAS# 28718-90-3	ThermoFisher Cat# D1306
Dimethyl Sulfoxide ($(CH_3)_2SO$)	CAS# 67-68-5	Sigma Cat# D8418-250ML
EDTA ($C_{10}H_{16}N_2O_8$)	CAS# 139-33-3	Sigma Cat# E7889
Ethanol (C_2H_5OH)	CAS# 74-17-5	Merck Cat# 1009831000
Gel Loading Dye	---	NEB Cat# B7024S
Gentamycin Sulfate ($C_{19}H_{40}N_4O_{10}S$)	CAS# 1405-41-0	Sigma Cat# A4854.0025
Glycerol ($C_3H_8O_3$)	CAS# 56-81-5	AppliChem Cat# 131339-1211
HEPES ($C_8H_{18}N_2O_4S$)	CAS# 7365-45-9	AppliChem Cat# A1069.1000
Imidazole ($C_3H_4N_2$)	CAS# 288-32-4	Carl Roth Cat# 3899.4
Isopropanol (C_3H_8O)	CAS# 67-63-0	Merck Cat# 109634

Kanamycin Sulfate ($C_{18}H_{38}N_4O_{15}S$)	CAS# 70560-51-9	AppliChem Cat# A1493.0025
Manganese (II) chloride ($MnCl_2$)	CAS# 7773-01-5	AppliChem Cat# A2087.0500
Nonidet P-40	CAS# 9016-45-9	Sigma-Aldrich Cas# 74385-1L
PEI Max®	CAS# 49553-93-7	Polysciences Cat# 24765-1
Pepstatin A ($C_{34}H_{63}N_5O_9$)	CAS# 26305-03-3	AppliChem Cat# A2205.0025
Phenanthroline ($C_{12}H_8N_2$)	CAS# 66-71-7	AppliChem Cat# A3826.0010
Phosphoramidon ($C_{23}H_{34}N_3O_{10}P$)	CAS# 36357-77-4	AppliChem Cat# A2214.0010
Phosphoric Acid (H_3PO_4)	CAS# 7664-38-2	Macherey-Nagel Cat# 740317.1000
PIPES ($C_8H_{18}N_2O_6S_2$)	CAS# 5625-37-6	AppliChem Cat# A1079.0500
Poly-L-Lysine	CAS# 25988-63-0	Sigma Cat# P4707-50ML
Potassium Acetate (CH_3COOK)	CAS# 127-08-2	AppliChem Cat# 10027073
Potassium Chloride (KCl)	CAS# 7447-40-7	AppliChem Cat# 10021170
Red-Safe Solution	---	Chembio Cat# 21141
Sodium Chloride (NaCl)	CAS# 7440-23-5	AppliChem Cat# 131659.1214
Sodium Dodecyl Sulphate ($C_{12}H_{25}NaSO_4$)	CAS# 151-21-3	AppliChem Cat# A2263.0500
Sodium Hydroxide (NaOH)	CAS# 1310-73-2	AppliChem Cat# 131687.1211
Sorbitol ($C_6H_{14}O_6$)	CAS# 50-70-4	AppliChem Cat# A4992.1000
Spectinomycin Sulfate ($C_{14}H_{26}N_2O_{11}S$)	CAS# 64058-48-6	Sigma Cat# PHR1441-1G
TCEP ($C_9H_{15}O_6P$)	CAS# 51805-45-9	GoldBio Cat# TCEP100
Thymol Blue ($C_{27}H_{30}O_5S$)	CAS# 76-61-9	Sigma Cat# 861367
Tris ($C_4H_{11}NO_3$)	CAS# 77-86-1	Sigma-Aldrich Cat# T1503
Tris-Cl ($C_4H_{13}Cl_2NO_3$)	CAS# 1185-53-1	Sigma Cat# PHG00002
Triton-X 100	CAS# 9002-93-1	Merck Cat# L734943
Zeocin™	---	ThermoFisher Cat# R25001

2.1.4. Unnatural Amino Acids

L-5-HTP (or 5-HTP)	CAS# 4350-09-8	Sigma-Aldrich Cat# H9772
L-ANAP (or ANAP)	CAS# 2287346-69-2	AsisChem Cat# ASIS-0146
L-AzPhe (or AzPhe)	CAS# 33173-53-4	BaseClick Cat# BCAA-001
L-Dansylalanine	CAS# 53332-27-7	Sigma-Aldrich Cat# D0125-5G
L-N_3Lys (or N_3Lys)	CAS# 1167421-25-1	SiChem Cat# SC-8027
L-PrK (or PrK)	CAS# 1428330-91-9	SiChem Cat# SC-8002
L-PrPhe (or PrPhe)	CAS# 1080496-42-9	ChemImpex Cat# 29744

2.1.5. Single-use Consumables

Razor Blades	Apollo Cat# 3.7658.01
Bacterial Electroporation Cuvettes	Bio-Rad Cat# 1652083
10 mL Disposable Syringe	Braun Cat# 4616103V
NC-Slide A8™	ChemoMetec Cat# 942-0003
50 mL Centrifuge Tube	Corning Cat# 352070
150 mm Treated Cell-Culture Plate	Corning Cat# 430599

5 mL Centrifuge Tube	Eppendorf Cat# 0030122313
0.5 mL Microcentrifuge Tube	Eppendorf Cat# 0030123301
1.5 mL Microcentrifuge Tube	Eppendorf Cat# 0030123328
2.0 mL Microcentrifuge Tube	Eppendorf Cat# 0030123344
12.5 µL Filtered Tips, Sterile	Integra Cat# 3415
125 µL Filtered Tips, Sterile	Integra Cat# 3425
300 µL Filtered Tips, Sterile	Integra Cat# 3435
1250 µL Filtered Tips, Sterile	Integra Cat# 3445
12.5 µL Low-Retention, Long Filtered Tips, Sterile	Integra Cat# 6505
125 µL Low-Retention Filtered Tips, Sterile	Integra Cat# 6565
NucleoBond® Xtra-Midi Columns	Macherey-Nagel Cat# 740410.10S
NucleoBond® Xtra-Maxi Columns	Macherey-Nagel Cat# 740414.100S
0.2 µm Bottle Filter	Millipore Cat# 051266B
0.025 µm Membrane	Millipore Cat# VSWP04700
96-Well PCR Plates	Sarstedt Cat# 72.1978.202
PCR Strip	Sarstedt Cat# 72.991.002
PCR Tubes	Sarstedt Cat# 72.991.002 | Individual tubes separated by a blade.
0.2 µm Syringe Filter	Sarstedt Cat# 83.1826.001
384-Well Cell-Culture Plate	ThermoFisher Cat# 142761
96-Well Cell-Culture Plate	ThermoFisher Cat# 165305
1.25 mL Cryogenic Tube	ThermoFisher Cat# 374080
384-Well PCR Plates	ThermoFisher Cat# TF0384
125 mL Polycarbonate Flask	TriForest Cat# FPC0125S
250 mL Polycarbonate Flask	TriForest Cat# FPC0250S
500 mL Polycarbonate Flask	TriForest Cat# FPC0500S
1000 mL Polycarbonate Flask	TriForest Cat# FPC1000S

2.1.6. Instruments

12.5 µL 8-Channel Programmable Pipette	Integra (Voyager)
12.5 µL 8-Channel Programmable Pipette	Integra (Voyager)
125 µL 12-Channel Programmable Pipette	Integra (Voyager)
125 µL 8-Channel Programmable Pipette	Integra (Voyager)
1250 µL 8-Channel Programmable Pipette	Integra (Voyager)
300 µL 8-Channel Programmable Pipette	Integra (Voyager)
BioSpa Automated Incubator	BioTek
Cytation5 Plate Reader and Imager	BioTek
Humidified CO_2 Incubator	Binder
MultiFloFX Liquid Handler	BioTek
NucleoCounter® NC-3000™	ChemoMetec

2.1.7. Enzymes and Enzyme Mix

DNase A	AppliChem Cat# A3778.0100
Soybean Trypsin Inhibitor	Gibco Cat# 17075029
HF Buffer	NEB Cat# BS0518S
NEBuilder® HiFi DNA Assembly Master Mix	NEB Cat# E2621L
NdeI	NEB Cat# R0111L
XbaI	NEB Cat# R0145M
AvrII	NEB Cat# R0174L
DpnI	NEB Cat# R0176L
PvuI-HF	NEB Cat# R3150L
Phusion-Polymerase Mix	Prepared inhouse by Dr. Roman P. Jakob
dNTPs	ROTH Cat# L785.1
RNase A	Sigma Cat# R6148
Trypsin-EDTA	Sigma Cat# T3924-100ML
Gateway™ BP Clonase™ II Enzyme Mix	ThermoFisher Cat# 11789100
Gateway™ LR Clonase™ II Enzyme Mix	ThermoFisher Cat# 11791100
Proteinase K	ThermoFisher Cat# EO0492

2.1.8. Molecular Biology Workflow Kits

GenElute™ Plasmid Miniprep Kit	Sigma Cat# PLN70-1KT
Wizard® SV Gel and PCR Clean-Up System	Promega Cat# A9281

2.1.9. Oligonucleotides and Polynucleotides

Primers	Biomers and Sigma-Aldrich
Synthetic Genes	GenScript and TwistBiosciences

2.2. COMMON BUFFERS: COMPOSITION, PREPARATION AND STORAGE

2.2.1. Transfection-Grade DNA-Purification Buffers

Resuspension Buffer S1	50 mM Tris-HCl, 10 mM EDTA, 100 µg/mL RNase A in Milli-Q water.
	pH 8.0, adjusted with HCl. Filtered sterile* and stored in cold room (~4°C).
Lysis Buffer S2	200 mM NaOH, 1% SDS, 1 mg/L Thymol Blue (optional) in Milli-Q water.
	Stored at room temperature in dark.
Neutralization Buffer S3	2.8 M Potassium Acetate in Milli-Q water.
	pH 5.1, adjusted with Acetic Acid. Filtered sterile and stored in cold room (~4°C).
Equilibration Buffer N2	100 mM Tris, 15% ethanol, 900 mM KCl, 0.15% Triton X-100 in Milli-Q water.
	pH 6.3, adjusted with H_3PO_4. Filtered sterile and stored in cold room (~4°C).
Wash Buffer N3	100 mM Tris, 15% ethanol, 1.15 M KCl in Milli-Q water.
	pH 6.3, adjusted with H_3PO_4. Filtered sterile and stored in cold room (~4°C).
Elution Buffer N5	100 mM Tris, 15% isopropanol, 1 M KCl in Milli-Q water.
	pH 8.5, adjusted with H_3PO_4. Filtered sterile and stored in cold room (~4°C).

2.2.2. Cell-Culture Media and Buffers

HEK Cell Lysis Buffer	20 mM Tris, 250 mM NaCl, 1 mM TCEP, 10 mM imidazole, 2 µg/ml DNase A, 0.01% NP-40, 1 µM Phosphoramidon, 1 µM Pepstatin A, 10 µM Bestatin, 10 µM Phenanthroline in Milli-Q water.
	pH 7.4, adjusted with HCl. Filtered sterile and stored at -20°C in 10 mL aliquots.
SOC Medium	20 mM Glucose in SOB medium (prepared according to manufacturer's guidelines).
	Filtered sterile and stored at 4°C.
Amino Acid Solvent	0.2 M NaOH, 15% v/v DMSO in Milli-Q water. Always prepared fresh.

2.2.3. Analysis Buffers

50x TAE Buffer	242 g Tris, 57 mL Acetic Acid Glacial, 18.6 g EDTA in 1000 mL Milli-Q water.
	Stored at room temperature.
TE Buffer	10 mM Tris-Cl, 1 mM EDTA in Milli-Q water, pH 8.0. Filtered sterile and stored at 4°C.

2.2.4. Unnatural Amino Acid Stocks

100x 5-HTP	10 mM 5-HTP in PBS. Filtered sterile and stored at -20°C.
100x ANAP	10 mM ANAP in 100% DMSO. Stored at -20°C.
100x AzPhe	100 mM AzPhe in Amino Acid Solvent. Filtered sterile and stored at -20°C.
100x DanAla	100 mM DanAla in 100% DMSO. Stored at -20°C.
100x N_3Lys	100 mM N_3Lys in Amino Acid Solvent. Filtered sterile and stored at -20°C.
100x PrK	100 mM PrK in Amino Acid Solvent. Filtered sterile and stored at -20°C.
100x PrPhe	100 mM PrPhe in Amino Acid Solvent. Filtered sterile and stored at -20°C.

* Sterilization through filtration was performed by filtering the solution through a 0.2 µm filter.

2.3. COMMON PROTOCOLS

2.3.1. Analytical Polymerase Chain Reactions

For most PCR amplifications, analytical-PCR reactions were performed before preparative-PCR reactions to identify the optimal annealing temperature and extension time for the amplification of DNA fragments. For analytical-PCR amplifications, 80 µL PCR reaction mix was prepared with 1x Phusion HF Buffer, 1 mM dNTPs, 100 µM each of forward and reverse primers, 100 ng template DNA and 1% v/v Phusion-Polymerase Mix in autoclaved Milli-Q water. The 80 µL mix was then divided into a PCR strip of 8 tubes, such that each tube contained 10 µL reaction volume. The initial denaturation for the PCR reaction was performed at 98°C for 3 minutes. Afterwards, either 25 or 30 PCR cycles were performed, where each cycle involved denaturation at 98°C for 30 seconds, primer-annealing using a temperature gradient of 55-72°C and extension at a temperature of 72°C for 2-4 minutes, depending on the length of the expected PCR amplicon. After the PCR cycles were complete, final extension was performed at 72°C for 10 minutes, before storing the amplified DNA at 4°C. After the analytical-PCR reaction, the amplicons were analyzed on 1% agarose gels using electrophoresis and the most suitable PCR conditions were identified.

If the aforementioned PCR mix failed to give any PCR products, then the same procedure was repeated with 5%, 10% or 20% v/v betaine in the PCR-reaction mix.

2.3.2. Preparative Polymerase Chain Reactions

Preparative-PCR reactions were performed to amplify linear DNA for Gibson Assembly or Gateway™ cloning. The optimal conditions for PCR-mix and PCR reaction cycle, as obtained from the analytical-PCR reaction, were used for preparative-PCR reactions. Preparative-PCR reactions were performed in 200 µL total reaction volume, which was divided into 8x aliquots of 25 µL in a PCR strip. The PCR reaction mix contained 1x Phusion HF Buffer, 1 mM dNTPs, 100 µM each of forward and reverse primers, 100 ng template DNA and 1% v/v Phusion-Polymerase Mix in autoclaved Milli-Q water. Betaine was present if suggested by the analytical-PCR experiments. After the PCR reactions, the template DNA was digested using DpnI by following the recommended protocol from the manufacturer. Linear DNA was purified on 1% agarose gel, via electrophoresis.

2.3.3. Mutagenesis Polymerase Chain Reactions

Mutagenesis-PCR reactions were performed in 10 µL total reaction volume in a single PCR tube of a PCR strip. The PCR reaction mix contained 1x Phusion HF Buffer, 1 mM dNTPs, 100 µM each of forward and reverse primers, 100 ng template DNA and 1% v/v Phusion-Polymerase Mix in autoclaved Milli-Q water. After the PCR reaction, the template DNA was digested using DpnI by following the recommended protocol from the manufacturer.

2.3.4. Agarose-gel Purification of DNA

DNA mixed with 1x gel-loading dye was carefully loaded in the furrows of 1% agarose gel (prepared in 1x TAE buffer) containing 1x red-safe solution. The electrophoretic separation was achieved by applying a voltage of 95-105 V across the gel using the Bio-Rad PowerPac Basic Power Supply. After the separation, the appropriate bands of the linear DNA fragments were excised from the gel using a fresh razor blade under the UV light and were transferred to appropriately sized microcentrifuge tube (2 or 5 mL). DNA was separated purified from the gels by using the purification kit and the standard protocol of the Wizard® SV Gel and PCR Clean-Up System.

2.3.5. Restriction Digestion Reactions

Restriction digestion reactions were performed according to the protocols recommended by the manufacturers of the enzymes.

2.3.6. Gibson Assembly Reactions

Gibson Assembly reactions were performed according to the protocols recommended by NEB.

2.3.7. Gateway™ Cloning Reactions

Gateway™ cloning mediated recombination reactions were performed according to the protocols recommended by the manufacturer of the enzymes. Gateway™ BP Clonase™ II Enzyme Mix would be henceforth referred to as BP-Clonase. Gateway™ LR Clonase™ II Enzyme Mix would be henceforth referred to as LR-Clonase.

2.3.8. Preparation Electrocompetent Bacterial Cells

For preparation of electrocompetent cells, 5 mL sterile SOC medium was inoculated with 25 µL of freshly thawed bacterial cells (protocol is common for all kinds of cells). After overnight growth at 37°C, 300 µL of this culture was used to inoculate 15 mL sterile SOC medium. Cells were allowed to grow overnight at 37°C. 10 mL of this overnight culture was used to inoculate 1000 mL sterile SOC medium and cells were allowed to grow at 37°C at 200 rpm. At OD_{600} 0.6 - 0.7, cells were harvested at 4500 *xg*. After discarding the supernatant, cells were washed sequentially with 1000 mL chilled Milli-Q, 500 mL chilled solution of glycerol (15% v/v in Milli-Q water) and 100 mL chilled electro-transformation solution (15% v/v glycerol and 2% v/v sorbitol in Milli-Q water). After the final washing step, cells were homogeneously resuspended in 10 mL chilled electro-transformation solution and were distributed in sterile PCR-tubes as 100 µL aliquots. The aliquots were stored at -80°C after flash freezing in liquid nitrogen.

2.3.9. Transformation of Electrocompetent Bacterial Cells

For electro-transformation of competent cells, the frozen aliquot of cells was thawed on ice (0°C) for 5-10 minutes. Depending on the size and source of DNA, 10-100 pg of desalted plasmid DNA* was added to one aliquot cells. After thoroughly mixing the cells and the DNA, the mixture was transferred to sterile electroporation cuvette and was incubated on ice for 60 seconds. Electro-transformation of cells was done by giving two pulses of current using the standard '*E. Coli*' setting of Bio-Rad MicroPulser Electroporator. After electroporation, 500 µL SOC medium was added to cells, cells were recovered for 60 - 120 min in a shaker incubator (37°C). After post-transformation recovery, cells were plated on LB-Agar plates containing appropriate selection antibiotics.

2.3.10. Preparation Chemically Competent Bacterial Cells

For preparation of electrocompetent cells, 5 mL sterile SOC medium was inoculated with 25 µL of freshly thawed bacterial cells (protocol is common for all kinds of cells). After overnight growth at 37°C, 300 µL of this culture was used to inoculate 15 mL sterile SOC medium. Cells were allowed to grow overnight at 37°C. 10 mL of this overnight culture was used to inoculate 1000 mL sterile SOC medium and cells were allowed to grow at 37°C at 200 rpm At OD_{600} 0.6 - 0.7, cells were harvested at 4500 *xg*. After discarding the supernatant, cells were resuspended in 200 mL freshly prepared and chilled transformation buffer (10 mM PIPES, 55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl in Milli-Q water). After resuspended cells were incubated on ice (0°C) for 10 minutes, cells were harvested again and resuspended in 10 mL chilled transformation buffer. While mixing gently, cell culture grade DMSO was added to the cells to a final concentration of 7% v/v. Cells were then distributed in sterile 2 mL microcentrifuge tubes as 100 µL aliquots. The aliquots were stored at -80°C after flash freezing in liquid nitrogen.

* DNA can be desalted either by dilution in Milli-Q or dialysis using a 0.025 µm membrane.

2.3.11. Transformation of Chemically Competent Bacterial Cells

For chemical-transformation of competent cells, the frozen aliquot of cells was thawed on ice (0°C) for 5-10 min. Depending on the size and source of DNA, 10 pg - 10 ng of plasmid DNA was added to one aliquot cells. After thoroughly mixing the cells and the DNA, the microcentrifuge tube containing the mixture was incubated on ice (0°C) for 30 minutes. After incubation, heat-shock treatment was given to cells by incubating the microcentrifuge tube at 42°C for 30-40 seconds. Next, the microcentrifuge tube was again incubated on ice (0°C) for 5-10 minutes. After transformation, 900 µL SOC medium was added to cells, cells were recovered for 60 - 120 min in a shaker incubator (37°C). After post-transformation recovery, cells were plated on LB-Agar plates containing appropriate selection antibiotics.

2.3.12. Purification of DNA for Molecular Biology and Long-Term Storage

A single colony of cells was picked from the LB-Agar plates and were transferred to 5 mL SOC medium. The culture was incubated at 37°C with vigorous shaking (600-800 rpm) for 8-10 hours. For preparing glycerol stocks (if needed), 500 µL of this culture was removed into a cryogenic tube containing 500 µL of sterile 50% v/v glycerol and the mixture was stored at -80°C after flash freezing in liquid nitrogen. Remaining cells were harvested at 3000 *xg* and GenElute™ Plasmid Miniprep Kit was used for plasmid purification as per the recommended protocol. Plasmid DNA was eluted from the columns in 30-100 µL filtered sterile TE buffer (depending on the DNA concentration requirements of downstream operations). For use within a week, DNA was kept at 4°C. For long-term preservation, DNA was stored at -20°C.

2.3.13. Confirmation of Molecular Cloning Operations

All acquired and created plasmid DNA molecules were sequenced by using appropriate sequencing primers. Whenever new plasmids and backbones were assembled, the entire construct was sequenced. When a linear DNA fragment was inserted into a known plasmid backbone, only the region of interest was sequenced. Sequencing service was provided by *Microsynth AG*.

2.3.14. Measurement of Mammalian Cell-Density and Cell-Viability

95 µL of cells were removed from the cell-suspension into a PCR tube containing 5 µL of AO-DAPI solution (100 µg/ml DAPI and 30 µg/ml AO in Milli-Q water). After allowing the cells to stain for 60 seconds, 10-15 µL of stained cells were added to one well of the NC-Slide A8™ and the slide was loaded into the NucleoCounter® NC-3000™ cell counter for measuring cell-density, cell-viability & cell-aggregation.

2.3.15. Maintenance of Adherent Culture of Mammalian Cells

The frozen vial of HEK293 cells obtained from the vendor was stored in liquid nitrogen till it was needed. For starting the maintenance culture, the frozen cells were thawed in water bath (37°C) for about 2 minutes, outside of the vials was decontaminated with 70% ethanol and under sterile conditions, all cells were transferred to 30 mL cell culture medium C-DMEM (DMEM supplemented with 10% v/v FBS) present in 150 mm Treated Cell Culture (TCC) plate. Cells were allowed to grow at 37°C in a humidified CO_2 incubator (80-90% humidity and 5% CO_2) for about 2 days. For maintaining the culture, one-fourth of the cells were passaged to the next generation at about 90% confluence (visual observation under a light microscope). This was achieved by gently washing the cells with 10 mL sterile PBS, dislodging the cells with 2 mL of prewarmed (37°C) Trypsin-EDTA, deactivating the Trypsin using 10 mL C-DMEM, resuspending the cells by pipetting and adding 3 mL of resuspended cells to 27 mL cell culture medium C-DMEM present in 150 mm TCC plates.

2.3.16. Maintenance of Suspension Culture of Mammalian Cells

The frozen vial of Expi293F™ cells obtained from the vendor was stored in liquid nitrogen till it was needed. For starting the maintenance culture, the frozen cells were thawed in water bath (37°C) for about 2 minutes, outside of the vials was decontaminated with 70% ethanol and under sterile conditions, all cells were transferred to 30 mL Expi293™ Expression Medium present in 125 mL polycarbonate (PC) flask. Cells were allowed to grow at 37°C with agitation (120 rpm) in a humidified CO_2 incubator (80-90% humidity and 5% CO_2). Every 24-48 hours, cell-density and cell-viability was measured and recorded to keep an overview of cell-health. For culture maintenance, cells were passaged at cell-density of ~3 x 10^6 cells/ml in fresh Expi293™ Expression Medium such that the final concentration of cells was 0.5 x 10^6 cells/ml.

2.4. GATEWAY™ DESTINATION VECTORS AND PLASMID BACKBONES

To facilitate molecular cloning, a backbone plasmid vector *Plasmid GABB* was created. For screening experiments and large-scale protein production in mammalian cells, two Gateway™ Cloning [232] destination vectors, namely *Plasmid MX01* and *Plasmid UCAP* were created. To facilitate gene insertion into the *Plasmid UCAP*, a Gateway™ Cloning donor vector *Plasmid GD54* was created. For screening experiments, a mammalian expression vector backbone *Plasmid UTX0* was created. These five plasmid vectors were created using Gibson Assembly [233, 234].

Plasmids mentioned in this section are created *de-novo*, so they were characterized by restriction digestion and sequencing of the entire circular plasmid.

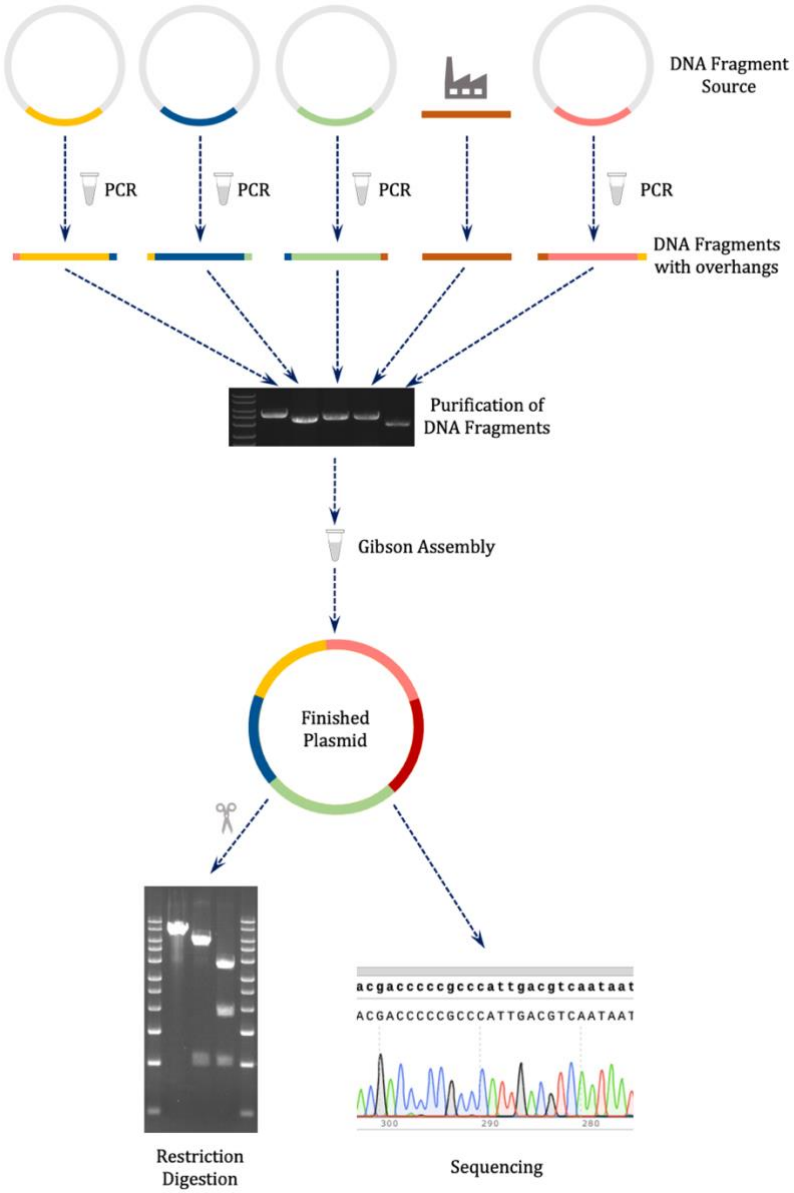


Figure 7: Strategy for Assembling the Plasmid Vectors.
Empty plasmids are assembled via Gibson Assembly. DNA fragments for the assembly are either amplified by PCR or are synthesized. Fragments are then purified on agarose gel, before Gibson Assembly. The identity of finished plasmids is confirmed by restriction digestion profiling and DNA sequencing.

2.4.1. Backbone Plasmid GABB

Four linear DNA fragments namely *GABB.Frag1*, *GABB.Frag2*, *GABB.Frag3* and *GABB.Frag4* were generated following the preparative-PCR reaction protocol. *GABB.Frag1* was amplified using *Plasmid SYN2.1* template and primers *GFA-1.5-0.0.1-1F* & *GFA-1.5-0.0.1-1R*. *GABB.Frag2* was amplified using *Plasmid SYN1.4* template and primers *GFA-1.5-0.0.1-2F* & *GFA-1.5-0.0.1-2R*. *GABB.Frag3* was amplified using *Plasmid pOMeY* template and primers *GFA-1.5-0.0.1-3F* & *GFA-1.5-0.0.1-3R*. *GABB.Frag4* was amplified using *Plasmid SYN1.1* template and primers *GFA-1.5-0.0.1-4F* & *GFA-1.5-0.0.1-4R*. After agarose-gel purification, the four linear DNA fragments were assembled into a plasmid via Gibson Assembly, by using the enzymes mix and the standard protocol of the NEBuilder® HiFi DNA Assembly Master Mix. The assembled *Plasmid GABB* was selected and amplified in NEB® 10-beta *E. coli* electrocompetent cells (*NEB Cat# C3020K*) using 100 μg/mL carbenicillin as the selection antibiotic.

2.4.2. Gateway™ Donor Vector GD54

Gateway™ donor vector *Plasmid GD54* was created to facilitate Gateway™ cloning mediated insertion of tRNA genes in *Plasmid UCAP*. Linear DNA fragments namely *GD21.GV.Frag1*, *GD21.GV.Frag2*, *GD21.GV.Frag3* and *GD21.GV.Frag4* were generated following the preparative-PCR reaction protocol. *GD21.GV.Frag1* was amplified using *Plasmid GABB* template and primers *GFA-1.6-1.1.1-F* & *GFA-1.6-1.1.1-R*. *GD21.GV.Frag2* was amplified using *Plasmid pDONR221* template and primers *GFA-1.6-1.1.2-F* & *GFA-1.6-1.1.2-R*. *GD21.GV.Frag3* was amplified using *Plasmid GABB* template and primers *GFA-1.6-1.1.3-F* & *GFA-1.6-1.1.3-R*. *GD21.GV.Frag4* was amplified using *Plasmid pDONR221* template and primers *GFA-1.6-1.1.4-F* & *GFA-1.6-1.1.4-R*. After agarose-gel purification, *GD21.GV.Frag1* & *GD21.GV.Frag2* were assembled into a *Plasmid GD21.GV.1* and *GD21.GV.Frag3* & *GD21.GV.Frag4* were assembled into a *Plasmid GD21.GV.2* via Gibson Assembly. Both *Plasmid GD21.GV.1* and *Plasmid GD21.GV.2* were selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 100 μg/mL carbenicillin disodium salt as the selection antibiotic. After DNA sequencing of both plasmids, *Plasmid GD54.GV.1* was generated by mutagenesis-PCR using *Plasmid GD21.GV.1* template and primers *PMI-1.6-1.1.1-F* & *PMI-1.6-1.1.1-R* and *Plasmid GD54.GV.2* was generated by mutagenesis PCR using *Plasmid GD21.GV.2* template and primers *PMI-1.6-1.1.2-F* & *PMI-1.6-1.1.2-R*. After DNA sequencing of *Plasmids GD54.GV.1* and *GD54.GV.2*, linear DNA fragment *GD54.Frag1* was PCR amplified (preparative-PCR) using *Plasmid GD54.GV.1* template and primers *GFA-1.6-1.1.5-F* & *GFA-1.6-1.1.5-R* and linear DNA fragment *GD54.Frag2* was PCR amplified (preparative-PCR) using *Plasmid GD54.GV.2* template and primers *GFA-1.6-1.1.6-F* & *GFA-1.6-1.1.6-R*. After agarose-gel purification, *GD54.Frag1* & *GD54.Frag2* were assembled into a *Plasmid GD54* via Gibson Assembly. *Plasmid GD54* was selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 50 μg/mL kanamycin sulfate and 25 μg/mL chloramphenicol as the selection antibiotics.

2.4.3. Gateway™ Destination Vector Plasmid MX01 for Mammalian Expression

Six linear DNA fragments namely *MX01.Frag1*, *MX01.Frag2*, *MX01.Frag3*, *MX01.Frag4*, *MX01.Frag5* and *MX01.Frag6* were generated following the preparative-PCR reaction protocol. *MX01.Frag1* was amplified using *Plasmid pOMeY* template and primers *GFA-1.5-6.0-1F* & *GVI-1.5-6.0-A-2R*. *MX01.Frag2* was amplified using *Plasmid pAB2G-N-HMF* template and primers *GVI-1.5-6.0-A-3F* & *GFA-1.5-6.1-1R*. *MX01.Frag3* was amplified using *Plasmid SYN3.1* template and primers *GFA-1.5-6.0-2F* & *GVI-1.5-6.0-B-2R*. *MX01.Frag4* was amplified using *Plasmid pAB2G-N-HMF* template and primers *GVI-1.5-6.0-B-3F* & *GFA-1.5-1.0-2R*. *MX01.Frag5* was amplified using *Plasmid SYN2.2* template and primers *GFA-1.5-1.0-3F* & *GVI-1.5-6.0-C-2R*. *MX01.Frag6* was amplified using *Plasmid pOMeY* template and primers *GVI-1.5-6.0-C-3F* & *GFA-1.5-6.0-3R*. After agarose-gel purification, the six linear DNA fragments were assembled into the *Plasmid MX01* via Gibson Assembly. The assembled *Plasmid MX01* was selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 20 μg/mL gentamycin as the selection antibiotic.

2.4.4. Gateway™ Destination Vector Plasmid UCAP for Mammalian Expression

We thank Dr. Dannel McCollum for providing pcDNA4-TO-Puromycin-mVenus-MAP (a.k.a. pPuro).

Linear DNA fragments namely *UCAA.Frag1*, *UCAA.Frag2*, *UCAA.Frag3*, *UCAA.Frag4* and *UCAA.Frag5* were generated following the preparative-PCR reaction protocol. *UCAA.Frag1* was amplified using *Plasmid SYN3.1* template and primers *GFA-1.5-7.0-1F* & *GVI-1.5-7.0-A-2R*. *UCAA.Frag2* was amplified using *Plasmid COM296_pDEST_pcD* template and primers *GVI-1.5-7.0-A-3F* & *GVI-1.5-7.0-A-3R*. *UCAA.Frag3* was amplified using *Plasmid GABB* template and primers *GVI-1.5-7.0-A-1F* & *GFA-1.5-7.0-1R*. *UCAA.Frag4* was amplified using *Plasmid Si-E038* template and primers *GFA-1.5-7.0-2F* & *GVA-1.5-7.0-B-2R*. *UCAA.Frag5* was amplified using *Plasmid pPuro* template and primers *GVI-1.5-7.0-B-3F* & *GFA-1.5-7.0-2R*. After agarose-gel purification, these five DNA fragments were assembled into a *Plasmid UCAA* via Gibson Assembly. The assembled *Plasmid UCAA* was selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 50 μg/mL kanamycin sulfate and 25 μg/mL chloramphenicol as the selection antibiotics.

However, after sequencing, it was discovered that *ccdB* gene of *Plasmid UCAA* was terminated prematurely, rendering it ineffective. To correct for this, *Plasmid UCAA* was linearized by simultaneous digestion with restriction enzymes NdeI & AvrII and the 3780 bp linear DNA fragment resulting from digestion was purified as *UCAA.Lin1*. Linear DNA fragments namely *UCAP.TRGW.Frag1*, *UCAP.TRGW.Frag2*, *UCAP.TRGW.Frag3* and *UCAP.TRGW.Frag4* were generated following the preparative-PCR reaction protocol. *UCAP.TRGW.Frag1* was amplified using linearized DNA *UCAA.Lin1* template and primers *GFA-1.6-6.6.1-F* & *GFA-1.6-6.6.1-R*. *UCAP.TRGW.Frag2* was amplified using *Plasmid pDONR221* template and primers *GFA-1.6-6.6.2-F* & *GFA-1.6-6.6.2-R*. *UCAP.TRGW.Frag3* was amplified using *Plasmid pDONR221* template and primers *GFA-1.6-6.6.3-F* & *GFA-1.6-6.6.3-R*. *UCAP.TRGW.Frag4* was amplified using linearized DNA *UCAA.Lin1* template and primers *GFA-1.6-6.6.4-F* & *GFA-1.6-6.6.4-R*. After agarose-gel purification, these four fragments were assembled into a *Plasmid UCAP.TRGW* via Gibson Assembly. *Plasmid UCAP.TRGW* was selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 50 μg/mL kanamycin as the selection antibiotic.

Additionally, *Plasmid UCAA* was separately linearized by digestion with restriction enzyme PvuI-HF as well and the 9162 bp linear DNA fragment resulting from digestion was purified as *UCAA.Lin2*. Linear DNA fragments namely *UCAP.Frag1*, *UCAP.Frag2*, *UCAP.Frag3* and *UCAP.Frag4* were generated following the preparative-PCR reaction protocol. *UCAP.Frag1* was amplified using linearized DNA *UCAA.Lin2* template and primers *GFA-1.6-6.7.1-F* & *GFA-1.6-6.7.1-R*. *UCAP.Frag2* was amplified using *Plasmid GABB* template and primers *GFA-1.6-6.7.2-F* & *GFA-1.6-6.7.2-R*. *UCAP.Frag3* was amplified using linearized DNA *UCAA.Lin2* template and primers *GFA-1.6-6.7.3-F* & *GFA-1.6-6.7.3-R*. *UCAP.Frag4* was amplified using *Plasmid UCAP.TRGW* template and primers *GFA-1.6-6.7.4-F* & *GFA-1.6-6.7.4-R*. After agarose-gel purification, these four fragments were assembled into a *Plasmid UCAP* via Gibson Assembly. *Plasmid UCAP* was selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 50 μg/mL kanamycin, 25 μg/mL chloramphenicol and 100 μg/mL carbenicillin as the selection antibiotics.

2.4.5. Gateway™ Destination Vector Plasmid UCZP for Mammalian Expression

Linear DNA fragments namely *UCZP.Frag1*, *UCZP.Frag2*, *UCZP.Frag3* and *UCZP.Frag4* were generated following the preparative-PCR reaction protocol. *UCZP.Frag1* was amplified using *Plasmid UCAP* template and primers *GFA-Z.A-1.1.1-F* & *GFA-Z.A-GWORI-R*. *UCZP.Frag2* was amplified using *Plasmid UCAP* template and primers *GFA-Z.A-GWORI-F* & *GFA-Z.A-GWCAM-R*. *UCZP.Frag3* was amplified using *Plasmid UCAP* template and primers *GFA-Z.A-GWCAM-F* & *GFA-Z.A-1.1.1-R*. *UCZP.Frag4* was amplified using *Plasmid pDONR221-Zeo* template and primers *GFA-Z.A-1.2.1-F* & *GFA-Z.A-1.2.1-R*. After agarose-gel purification, these five DNA fragments were assembled into a *Plasmid UCZP* via Gibson Assembly. The assembled *Plasmid UCAA* was selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 50 μg/mL Zeocin™, 50 μg/mL kanamycin and 25 μg/mL chloramphenicol as the selection antibiotics.

2.4.6. Mammalian Expression Vector Plasmid UTX0

mRaspberry gene present in synthetic gene *FP-mRaspberry* was inserted into mammalian expression vector *Plasmid MX01* via Gateway™ Cloning. Specifically, a linear DNA fragment *mRSP.attB12* was PCR-amplified (preparative-PCR) using synthetic gene *FP-mRaspberry* template and primers *GWA-1.6-11.1.1-F* & *GWA-1.6-11.1.1-R*. After agarose-gel purification, *mRaspberry* gene present in *mRSP.attB12* was inserted in Gateway™

Donor *Plasmid pDONR221*, using Gateway™ BP Clonase™ II Enzyme Mix assisted BP-recombination reaction, to generate the entry clone *GD21.9*. Using Gateway™ LR Clonase™ II Enzyme Mix assisted LR-recombination reaction between *GD21.9* and *Plasmid MX01*, *mRaspberry* gene was inserted in *Plasmid MX01* to generate the mammalian expression vector *Plasmid MX01.1*. After sequencing, *Plasmid MX01.1* was linearized by simultaneous digestion with restriction enzymes PvuI-HF & XbaI and 6268 bp linear DNA fragment resulting from the digestion was purified as linear DNA *UTX0.Frag1*. Another linear DNA fragment *UTX0.Frag2* was PCR-amplified (preparative-PCR) using *Plasmid eGFP-WPRE* template and primers *GFA-1.12-1.1.1-F* & *GFA-1.12-1.1.1-R* and was purified on agarose-gel. *UTX0.Frag1* and *UTX0.Frag2* were assembled into *Plasmid UTX0* via Gibson Assembly. *Plasmid UAT0* was selected and amplified in NEB® 10-beta *E. coli* electrocompetent cells using 20 μg/mL gentamycin as the selection antibiotic.

2.4.7. Gateway™ Destination Vectors GIDC, GIDK and GIDS for Insect Cell Expression

For baculovirus mediated protein production in insect cells, three Gateway™ cloning and MultiBac™ cloning [235] compatible destination vectors, namely *Plasmids GIDC*, *GIDK* & *GIDS* were also created via Gibson Assembly.

Three linear DNA fragments namely *GIDS.INT1.Frag1*, *GIDS.INT1.Frag2*, and *GIDS.INT1.Frag3* were generated following the preparative-PCR reaction protocol. *GIDS.INT1.Frag1* was amplified using *Plasmid pDONR221* template and primers *GFA-Z.D-1.1.1-F* & *GFA-Z.D-1.1.1-R*. *GIDS.INT1.Frag2* was amplified using *Plasmid GABB* template and primers *GFA-Z.D-1.1.2-F* & *GFA-Z.D-1.1.2-R*. *GIDS.INT1.Frag3* was amplified using *Plasmid pDONR221* template and primers *GFA-Z.D-1.1.3-F* & *GFA-Z.D-1.1.3-R*. After agarose-gel purification, the three DNA fragments were assembled into as a linear DNA Fragment *GIDS.INT1* via Gibson Assembly. Fragment *GIDS.INT1* was purified (without agarose-gel purification) using the Wizard® SV Gel and PCR Clean-Up System. Using purified *GIDS.INT1* as template and primers *Mex13-F* & *Mex13-R*, linear DNA fragment *GIDS.INT2.Frag1* was PCR-amplified (preparative-PCR). Additionally, linear DNA fragment *GIDS.INT2.Frag2* was PCR-amplified (preparative-PCR) using *Plasmid pIDS* template and primers *GFA-Z.D-4.1.1-F* & *GFA-Z.D-4.1.1-R*. After agarose-gel purification, these two fragments were assembled into *Plasmid GIDS.INT2* via Gibson Assembly. *Plasmid GIDS.INT2* was selected and amplified in One Shot™ ccdB Survival™ 2 $T1^R$ *E. coli* electrocompetent cells using 200 μg/mL carbenicillin and 50 μg/mL spectinomycin the selection antibiotics.

For changing the attP1-attP2 Gateway™ sites of *Plasmid GIDS.INT2* to attR1-attR2 Gateway™ sites, two linear DNA fragments namely *GIDS.INT3.Frag1* and *GIDS.INT3.Frag2* were generated following the preparative-PCR reaction protocol. *GIDS.INT3.Frag1* was amplified using *Plasmid GIDS.INT2* template and primers *GFA-Z.D-A.2.1-F* & *GFA-Z.D-A.2.1-R*. *GIDS.INT3.Frag2* was amplified using *Plasmid GIDS.INT2* template and primers *GFA-Z.D-K.2.2-F* & *GFA-Z.D-A.2.2-R*. After agarose-gel purification, these two fragments were assembled into *Plasmid GIDS.INT3* via Gibson Assembly. *Plasmid GIDS.INT3* was selected and amplified in One Shot™ ccdB Survival™ 2 $T1^R$ *E. coli* electrocompetent cells using 200 μg/mL carbenicillin and 50 μg/mL spectinomycin as the selection antibiotics.

Eventually, four linear DNA fragments namely *GIDX.BB*, *GIDC.Insert*, *GIDK.Insert* and *GIDS.Insert* were generated following the preparative-PCR reaction protocol. *GIDX.BB* was amplified using *Plasmid GIDS.INT3* template and primers *GFA-Z.D-A.3.1-F* & *GFA-Z.D-A.3.1-R*. *GIDC.Insert* was amplified using *Plasmid pIDC* template and primers *GFA-Z.D-1.4.1-F* & *GFA-Z.D-1.4.1-R*. *GIDK.Insert* was amplified using *Plasmid pIDK* template and primers *GFA-Z.D-2.4.1-F* & *GFA-Z.D-2.4.1-R*. *GIDS.Insert* was amplified using *Plasmid pIDS* template and primers *GFA-Z.D-2.4.1-F* & *GFA-Z.D-2.4.1-R*. These four fragments were purified on 1% agarose-gel.

Linear DNA fragment *GIDC.Insert* was inserted into the plasmid backbone *GIDX.BB* via Gibson Assembly, to create *Plasmid GIDC*. *Plasmid GIDC* was selected and amplified in One Shot™ ccdB Survival™ 2 $T1^R$ *E. coli* electrocompetent cells using 200 μg/mL carbenicillin and 25 μg/mL chloramphenicol as the selection antibiotics.

Linear DNA fragment *GIDK.Insert* was inserted into the plasmid backbone *GIDX.BB* via Gibson Assembly, to create *Plasmid GIDK*. *Plasmid GIDK* was selected and amplified in One Shot™ ccdB Survival™ 2 $T1^R$ *E. coli* electrocompetent cells using 200 μg/mL carbenicillin and 50 μg/mL kanamycin sulfate as the selection antibiotics.

Linear DNA fragment *GIDS.Insert* was inserted into the plasmid backbone *GIDX.BB* via Gibson Assembly, to create *Plasmid GIDS*. *Plasmid GIDS* was selected and amplified in One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* electrocompetent cells using 200 μg/mL carbenicillin and 50 μg/mL spectinomycin as the selection antibiotics.

2.5. MAMMALIAN EXPRESSION VECTORS FOR DELIVERING G.C.E. SYSTEMS

For genetic code expansion in mammalian cells, the orthogonal aaRS/tRNA pairs (or their sequences for synthetic genes) were obtained from external sources. $_{ANAP}RS$ and the corresponding otRNA has been evolved by the research group of Dr. Peter G. Schultz [186]. $_{DanAla}RS$ and the corresponding otRNA has been evolved by the research group of Dr. Peter G. Schultz [176]. $_{EcTrp(h9)}RS$, $_{EcTrp(h14)}RS$ and the corresponding otRNA has been evolved by the research group of Dr. Abhishek Chatterjee [142]. $_{OMeY}RS$ and the corresponding otRNA has been evolved by the research group of Dr. Peter G. Schultz [146]. $_{MbPyl}RS$, $_{MmPyl}RS$ and the corresponding otRNA has been evolved by the research group of Dr. Peter G. Schultz [146].

We thank Dr. Peter G. Schultz for providing us with plasmids pANAP, pAcBac2.tR4-OMeYRS/GFP* (a.k.a. pOMeY) and pAcBac1.tR4-MbPyl (a.k.a. pMbPyl).

Plasmids described in this section are derived from the plasmid backbones described in the previous section. For these plasmids, only those regions were sequenced which were subjugated to PCR amplification.

2.5.1. Entry Clones for Orthogonal tRNA Synthetase Gene

Table 1 details the starting materials and intermediates involved in creation of orthogonal tRNA Synthetase entry clones. The gene for the orthogonal tRNA Synthetase present in the template DNA was PCR-amplified (preparative-PCR) as attB12 insert, using the forward and reverse primers. attB12 insert was purified on agarose-gel and the gene of interest was inserted in the Gateway™ Donor by BP-Clonase mediated recombination cloning to make the oRS entry clone. All oRS entry clones were selected and amplified in NEB® 10-beta *E. coli* electrocompetent cells using 50 μg/mL kanamycin as the selection antibiotic.

Table 1: List of Orthogonal tRNA Synthetase Entry Clones

Template DNA Name	Forward Primer	Reverse Primer	attB12 Insert	Gateway™ Donor	oRS Entry Clone
pANAP	GWA-1.7-1.1.1-F	GWA-1.7-1.1.1-R	attB12.1	pDONR221	GD21.1
DanAla-RS	---	---	---	---	GD21.2 *
SYN4.6.a	GWA-1.7-4.1.1-F	GWA-1.7-4.1.1-R	attB12.3	pDONR221	GD21.3
SYN4.6.b	GWA-1.7-4.1.1-F	GWA-1.7-4.1.1-R	attB12.4	pDONR221	GD21.4
pMbPyl	GWA-1.7-3.1.2-F	GWA-1.7-3.1.2-R	attB12.5	pDONR221	GD21.5
mmPylRS-Y384F	----	----	----	pDONR221	GD21.6 †
pOMeY	GWA-1.7-2.1.2-F	GWA-1.7-2.1.2-R	attB12.7	pDONR221	GD21.7

* *Plasmid DanAla-RS* was already an oRS entry clone.
† Linear DNA *mmPylRS-Y384F* was already flanked with attB1-attB2 sites.

2.5.2. Entry Clones for Orthogonal tRNA Expression Cassette

Table 2 details the starting materials and intermediates involved in creation of orthogonal tRNA expression cassette entry clones. attB54 plasmid contains the expression cassette for the orthogonal tRNA, already flanked by attB5-attB4 recombination sites. This expression cassette was inserted in the Gateway™ Donor by BP-Clonase mediated recombination cloning to make the otRNA entry clone. All otRNA entry clones were selected and amplified in NEB® 10-beta *E. coli* electrocompetent cells using 50 μg/mL kanamycin as the selection antibiotic.

Table 2: List of Orthogonal tRNA Expression Cassette Entry Clones

attB54 Plasmid	tRNA	Gateway™ Donor	otRNA Entry Clone
SYN4.1.1	$_{EcLeu}tRNA_{UUA}$	GD54	GD54.1
SYN4.1.2	$_{EcLeu}tRNA_{CUA}$	GD54	GD54.2
SYN4.1.3	$_{EcLeu}tRNA_{UCA}$	GD54	GD54.3
SYN4.6.1	$_{EcTrp}tRNA_{UUA}$	GD54	GD54.4
SYN4.6.2	$_{EcTrp}tRNA_{CUA}$	GD54	GD54.5
SYN4.6.3	$_{EcTrp}tRNA_{UCA}$	GD54	GD54.6
SYN4.3.1	$_{MbPyl}tRNA_{UUA}$	GD54	GD54.7
SYN4.3.2	$_{MbPyl}tRNA_{CUA}$	GD54	GD54.8
SYN4.3.3	$_{MbPyl}tRNA_{UCA}$	GD54	GD54.9
SYN4.2.1	$_{OMeY}tRNA_{UUA}$	GD54	GD54.11
SYN4.2.2	$_{OMeY}tRNA_{CUA}$	GD54	GD54.12
SYN4.2.3	$_{OMeY}tRNA_{UCA}$	GD54	GD54.13

2.5.3. G.C.E. Plasmids - Vectors for Delivering the G.C.E.-Machinery Genes

For expressing a given G.C.E. system in mammalian cells, a corresponding 'G.C.E. Plasmid' (as listed in Table 3) was created by inserting the corresponding orthogonal tRNA Synthetase (oRS) gene and orthogonal tRNA (otRNA) expression cassette (containing 6x copies of the otRNA gene) into the Gateway™ cassettes of the *Plasmid UCAP*. This was achieved by first creating the Gateway™ entry clones for the genes of interest and then recombining the genes of interest into the Gateway™ destination vector to make expression vectors. For the oRS genes, the oRS entry clones (as listed in Table 3) were created using Gateway™ donor *Plasmid pDONR221* and for the otRNA expression cassettes, the otRNA entry clones (as listed in Table 3) were created using Gateway™ donor *Plasmid GD54*, as detailed in the supplementary information section 'Development of Expression Vectors for Screening Experiments'. Using these two kinds of entry clones and the destination vector *Plasmid UCAP*, 21 different kinds of 'G.C.E. Plasmids' (as listed in Table 3) were created via LR-Clonase mediated Gateway™ recombination cloning. For two simultaneous orthogonal Gateway™ recombination reactions, 100 ng of destination vector (*Plasmid UCAP*) and 150 ng of each of the oRS and otRNA entry clones (see Table 3) were mixed with 4 μL Gateway™ LR Clonase™ II Enzyme Mix in total 20 μL reaction volume buffered with TE buffer. After incubating the reaction mix for 3 hours at 25°C the reaction was stopped by adding 4 μg of Proteinase-K. The G.C.E. Plasmids were selected and amplified in One Shot™ PIR1 *E. coli* electrocompetent cells using 200 μg/mL carbenicillin as the selection antibiotic.

Table 3: Synthesis Intermediates of 'G.C.E. Plasmids' for Delivering G.C.E. Systems in Mammalian Cells

G.C.E. System	Destination Vector	oRS Entry Clone	otRNA Entry Clone	G.C.E. Plasmid
$_{ANAP}RS\ /_{EcLeu}tRNA$ [186]	UCAP	GD21.1	GD54.1	UCAP.1
			GD54.2	UCAP.2
			GD54.3	UCAP.3
$_{DanAla}RS\ /_{EcLeu}tRNA$ [185]		GD21.2	GD54.1	UCAP.4
			GD54.2	UCAP.5
			GD54.3	UCAP.6
$_{EcTrp(H9)}RS\ /_{EcTrp}tRNA$ [142]		GD21.3	GD54.4	UCAP.7
			GD54.5	UCAP.8
			GD54.6	UCAP.9
$_{EcTrp(H14)}RS\ /_{EcTrp}tRNA$ [142]		GD21.4	GD54.4	UCAP.10
			GD54.5	UCAP.11
			GD54.6	UCAP.12
$_{OMeY}RS\ /_{OMeY}tRNA$ [146]		GD21.7	GD54.11	UCAP.13
			GD54.12	UCAP.14
			GD54.13	UCAP.15
$_{MbPyl}RS\ /_{MbPyl}tRNA$ [146]		GD21.5	GD54.7	UCAP.16
			GD54.8	UCAP.17
			GD54.9	UCAP.18
$_{MmPyl}RS\ /_{MmPyl}tRNA$ [189]		GD21.6	GD54.7	UCAP.20
			GD54.8	UCAP.21
			GD54.9	UCAP.22

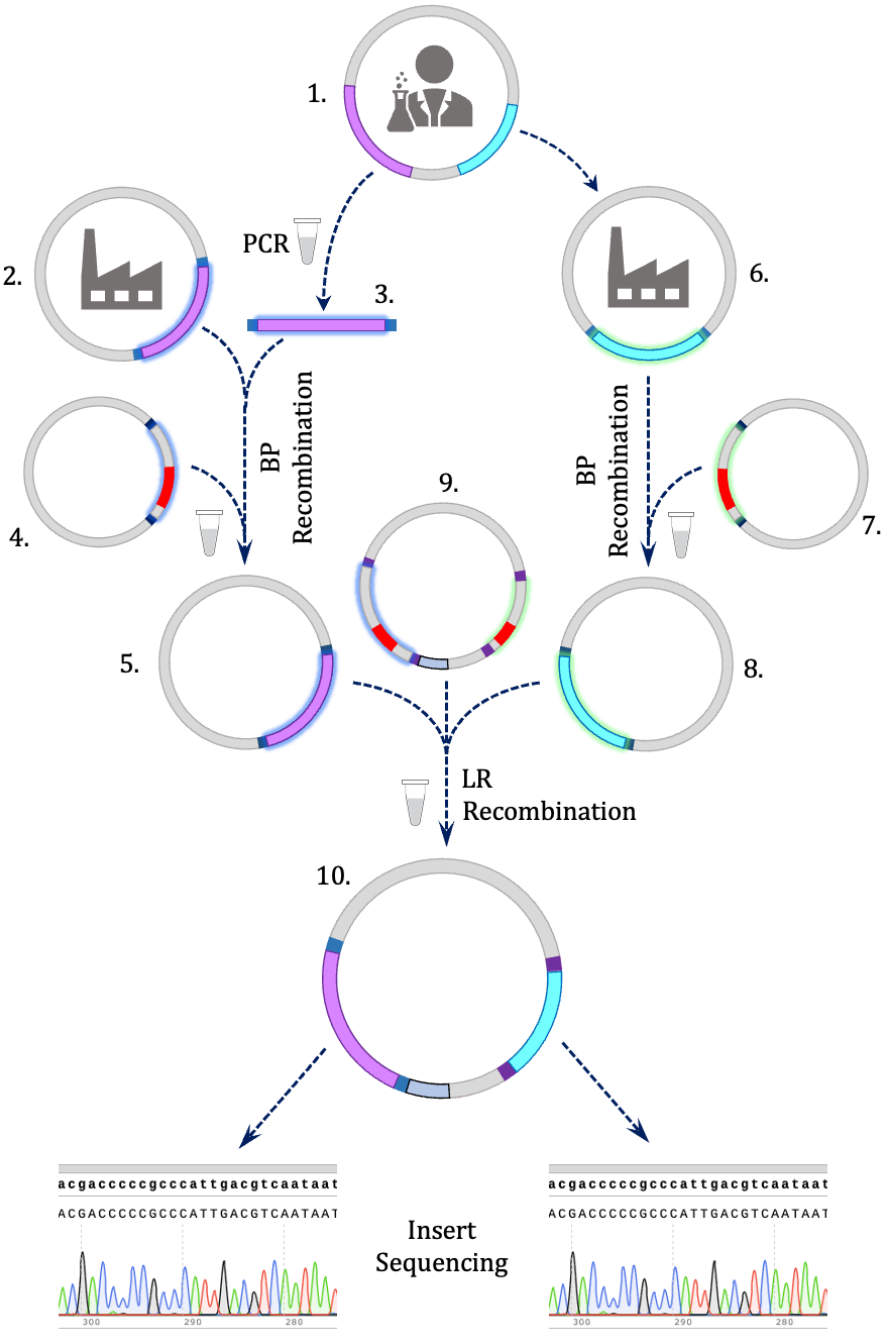


Figure 8: Strategy for Creating G.C.E. Plasmids.
G.C.E. Plasmids are created by introducing the orthogonal aaRS (oRS) gene and the orthogonal tRNA (otRNA) expression cassette in *Plasmid UCAP* (or *Plasmid UCZP*). oRS and otRNA are evolved by dedicated research groups (1) and we obtain either the genes or the sequences from them. The oRS gene is either synthesized with attB1-attB2 inserts (2) or attB1-attB2 flanking regions are added by PCR (3). BP recombination reaction with Gateway™ donor vector pDONR221 (4) gives oRS gene entry clone (5). otRNA genes are synthesized with attB5-attB4 inserts (6). After BP-recombination reaction with Gateway™ donor vector GD54 (7), entry clone for otRNA gene (8) is obtained. Both entry clones (5 & 8) are used for a one-pot LR-recombination reaction to insert their respective genes in the destination vector *Plasmid UCAP* (9), to give the mammalian expression vector (10). The identity of the inserted genes is confirmed by sequencing.

2.6. MAMMALIAN EXPRESSION VECTORS FOR G.C.E. EFFICIENCY QUANTIFICATION

For quantifying and comparing the UAA incorporation efficiency of the different G.C.E. systems, 55 different expression plasmids for the G.C.E. Efficiency Measurement and Screening (G.E.M.S.) were created by introducing the TAA, TAG and/or TGA stop codons at different amino acid positions in the eGFP gene of mammalian expression *Plasmid UTX0*.

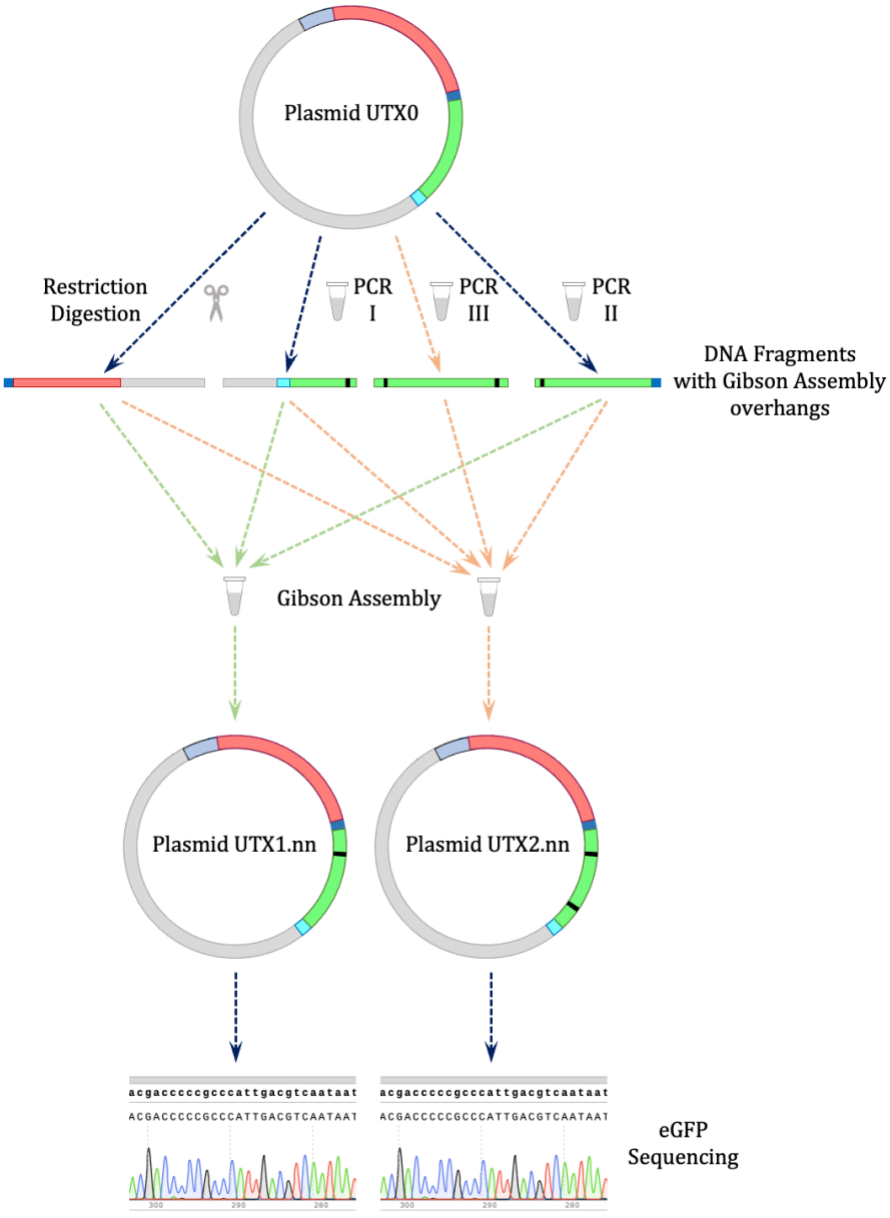


Figure 9: Strategy for Creating G.E.M.S. Plasmids.
Plasmid UTX0 expresses mRaspberry and eGFP proteins as separate polypeptides. For creating G.E.M.S. Plasmids, *Plasmid UTX0* is linearized via restriction digestion. Additionally, by using PCR, either two (for one stop-codon introduction) or three (for two stop codon introduction) linear fragments are created that contain targeted mutations in the overhangs. These fragments are then assembled with the backbone via Gibson Assembly and the mutations are confirmed via sequencing the eGFP gene of the plasmid.

2.6.1. G.E.M.S. Plasmids for Single-Site UAA Incorporation

For quantifying the efficiency of UAA incorporation at a single site in eGFP, one stop codon was introduced into the expression plasmid at a time. For any given amino acid position, three different expression

plasmids containing three different stop codons were created. By introducing each of the three stop codons at 15 different amino acid positions on eGFP gene, 45 different G.E.M.S. Plasmids (namely *Plasmids UTX1.01* to *UTX1.45*) were created by Gibson Assembly assisted mutagenesis. The details of the introduced mutations are listed in Table 4.

For creating the G.E.M.S. Plasmids for Single-Site UAA Incorporation Screening, the mammalian expression *Plasmid UTX0* was linearized by simultaneous digestion with restriction enzymes PvuI-HF & XbaI and 6268 bp linear DNA fragment backbone resulting from the digestion was purified as linear DNA *UTX0.BB*. For each of the expression vector, two linear DNA fragments (*GA Insert Fragment 1* and *GA Insert Fragment 2*) were created by preparative-PCR using *Plasmid UTX0* as template with respective primer pairs, as mentioned in Table 4. The template DNA was removed by DpnI digestion after PCR and the PCR-amplified GA Insert Fragments were purified (without agarose-gel purification) using the Wizard® SV Gel and PCR Clean-Up System. Eventually, the purified *GA Insert Fragments 1 and 2* were inserted in the linear DNA *UTX0.BB* via Gibson Assembly. The resulting expression vectors were selected and amplified in NEB® 10-beta *E. coli* chemically competent cells using 20 µg/mL gentamycin as the selection antibiotic.

Table 4: List of Expression Vectors for UAA Incorporation at One Site in eGFP

eGFP Mutation	GA Insert Fragment 1		GA Insert Fragment 2		Gibson Assembly Backbone	G.E.M.S. Plasmid
	Primer Pair 1	Fragment Name	Primer Pair 2	Fragment Name		
T9TAA	GFA-1.12-1.1.1-F & PMI-1.13-1.1.1-R	UTX1.01.A	PMI-1.13-1.1.1-F & GFA-1.12-1.1.1-R	UTX1.01.B	UTX0.BB	UTX1.01
T9TAG	GFA-1.12-1.1.1-F & PMI-1.13-1.1.2-R	UTX1.02.A	PMI-1.13-1.1.2-F & GFA-1.12-1.1.1-R	UTX1.02.B	UTX0.BB	UTX1.02
T9TGA	GFA-1.12-1.1.1-F & PMI-1.13-1.1.3-R	UTX1.03.A	PMI-1.13-1.1.3-F & GFA-1.12-1.1.1-R	UTX1.03.B	UTX0.BB	UTX1.03
Y39TAA	GFA-1.12-1.1.1-F & PMI-1.13-2.1.1-R	UTX1.04.A	PMI-1.13-2.1.1-F & GFA-1.12-1.1.1-R	UTX1.04.B	UTX0.BB	UTX1.04
Y39TAG	GFA-1.12-1.1.1-F & PMI-1.13-2.1.2-R	UTX1.05.A	PMI-1.13-2.1.2-F & GFA-1.12-1.1.1-R	UTX1.05.B	UTX0.BB	UTX1.05
Y39TGA	GFA-1.12-1.1.1-F & PMI-1.13-2.1.3-R	UTX1.06.A	PMI-1.13-2.1.3-F & GFA-1.12-1.1.1-R	UTX1.06.B	UTX0.BB	UTX1.06
K52TAA	GFA-1.12-1.1.1-F & PMI-1.13-3.1.1-R	UTX1.07.A	PMI-1.13-3.1.1-F & GFA-1.12-1.1.1-R	UTX1.07.B	UTX0.BB	UTX1.07
K52TAG	GFA-1.12-1.1.1-F & PMI-1.13-3.1.2-R	UTX1.08.A	PMI-1.13-3.1.2-F & GFA-1.12-1.1.1-R	UTX1.08.B	UTX0.BB	UTX1.08
K52TGA	GFA-1.12-1.1.1-F & PMI-1.13-3.1.3-R	UTX1.09.A	PMI-1.13-3.1.3-F & GFA-1.12-1.1.1-R	UTX1.09.B	UTX0.BB	UTX1.09
F99TAA	GFA-1.12-1.1.1-F & PMI-1.13-4.1.1-R	UTX1.10.A	PMI-1.13-4.1.1-F & GFA-1.12-1.1.1-R	UTX1.10.B	UTX0.BB	UTX1.10
F99TAG	GFA-1.12-1.1.1-F & PMI-1.13-4.1.2-R	UTX1.11.A	PMI-1.13-4.1.2-F & GFA-1.12-1.1.1-R	UTX1.11.B	UTX0.BB	UTX1.11
F99TGA	GFA-1.12-1.1.1-F & PMI-1.13-4.1.3-R	UTX1.12.A	PMI-1.13-4.1.3-F & GFA-1.12-1.1.1-R	UTX1.12.B	UTX0.BB	UTX1.12
D117TAA	GFA-1.12-1.1.1-F & PMI-1.13-5.1.1-R	UTX1.13.A	PMI-1.13-5.1.1-F & GFA-1.12-1.1.1-R	UTX1.13.B	UTX0.BB	UTX1.13
D117TAG	GFA-1.12-1.1.1-F & PMI-1.13-5.1.2-R	UTX1.14.A	PMI-1.13-5.1.2-F & GFA-1.12-1.1.1-R	UTX1.14.B	UTX0.BB	UTX1.14
D117TGA	GFA-1.12-1.1.1-F & PMI-1.13-5.1.3-R	UTX1.15.A	PMI-1.13-5.1.3-F & GFA-1.12-1.1.1-R	UTX1.15.B	UTX0.BB	UTX1.15
E132TAA	GFA-1.12-1.1.1-F & PMI-1.13-6.1.1-R	UTX1.16.A	PMI-1.13-6.1.1-F & GFA-1.12-1.1.1-R	UTX1.16.B	UTX0.BB	UTX1.16
E132TAG	GFA-1.12-1.1.1-F & PMI-1.13-6.1.2-R	UTX1.17.A	PMI-1.13-6.1.2-F & GFA-1.12-1.1.1-R	UTX1.17.B	UTX0.BB	UTX1.17

E132TGA	GFA-1.12-1.1.1-F & PMI-1.13-6.1.3-R	UTX1.18.A	PMI-1.13-6.1.3-F & GFA-1.12-1.1.1-R	UTX1.18.B	UTX0.BB	UTX1.18
Y143TAA	GFA-1.12-1.1.1-F & PMI-1.13-7.1.1-R	UTX1.19.A	PMI-1.13-7.1.1-F & GFA-1.12-1.1.1-R	UTX1.19.B	UTX0.BB	UTX1.19
Y143TAG	GFA-1.12-1.1.1-F & PMI-1.13-7.1.2-R	UTX1.20.A	PMI-1.13-7.1.2-F & GFA-1.12-1.1.1-R	UTX1.20.B	UTX0.BB	UTX1.20
Y143TGA	GFA-1.12-1.1.1-F & PMI-1.13-7.1.3-R	UTX1.21.A	PMI-1.13-7.1.3-F & GFA-1.12-1.1.1-R	UTX1.21.B	UTX0.BB	UTX1.21
Y151TAA	GFA-1.12-1.1.1-F & PMI-1.13-8.1.1-R	UTX1.22.A	PMI-1.13-8.1.1-F & GFA-1.12-1.1.1-R	UTX1.22.B	UTX0.BB	UTX1.22
Y151TAG	GFA-1.12-1.1.1-F & PMI-1.13-8.1.2-R	UTX1.23.A	PMI-1.13-8.1.2-F & GFA-1.12-1.1.1-R	UTX1.23.B	UTX0.BB	UTX1.23
Y151TGA	GFA-1.12-1.1.1-F & PMI-1.13-8.1.3-R	UTX1.24.A	PMI-1.13-8.1.3-F & GFA-1.12-1.1.1-R	UTX1.24.B	UTX0.BB	UTX1.24
Q157TAA	GFA-1.12-1.1.1-F & PMI-1.13-9.1.1-R	UTX1.25.A	PMI-1.13-9.1.1-F & GFA-1.12-1.1.1-R	UTX1.25.B	UTX0.BB	UTX1.25
Q157TAG	GFA-1.12-1.1.1-F & PMI-1.13-9.1.2-R	UTX1.26.A	PMI-1.13-9.1.2-F & GFA-1.12-1.1.1-R	UTX1.26.B	UTX0.BB	UTX1.26
Q157TGA	GFA-1.12-1.1.1-F & PMI-1.13-9.1.3-R	UTX1.27.A	PMI-1.13-9.1.3-F & GFA-1.12-1.1.1-R	UTX1.27.B	UTX0.BB	UTX1.27
V176TAA	GFA-1.12-1.1.1-F & PMI-1.13-10.1.1-R	UTX1.28.A	PMI-1.13-10.1.1-F & GFA-1.12-1.1.1-R	UTX1.28.B	UTX0.BB	UTX1.28
V176TAG	GFA-1.12-1.1.1-F & PMI-1.13-10.1.2-R	UTX1.29.A	PMI-1.13-10.1.2-F & GFA-1.12-1.1.1-R	UTX1.29.B	UTX0.BB	UTX1.29
V176TGA	GFA-1.12-1.1.1-F & PMI-1.13-10.1.3-R	UTX1.30.A	PMI-1.13-10.1.3-F & GFA-1.12-1.1.1-R	UTX1.30.B	UTX0.BB	UTX1.30
Y182TAA	GFA-1.12-1.1.1-F & PMI-1.13-11.1.1-R	UTX1.31.A	PMI-1.13-11.1.1-F & GFA-1.12-1.1.1-R	UTX1.31.B	UTX0.BB	UTX1.31
Y182TAG	GFA-1.12-1.1.1-F & PMI-1.13-11.1.2-R	UTX1.32.A	PMI-1.13-11.1.2-F & GFA-1.12-1.1.1-R	UTX1.32.B	UTX0.BB	UTX1.32
Y182TGA	GFA-1.12-1.1.1-F & PMI-1.13-11.1.3-R	UTX1.33.A	PMI-1.13-11.1.3-F & GFA-1.12-1.1.1-R	UTX1.33.B	UTX0.BB	UTX1.33
D190TAA	GFA-1.12-1.1.1-F & PMI-1.13-12.1.1-R	UTX1.34.A	PMI-1.13-12.1.1-F & GFA-1.12-1.1.1-R	UTX1.34.B	UTX0.BB	UTX1.34
D190TAG	GFA-1.12-1.1.1-F & PMI-1.13-12.1.2-R	UTX1.35.A	PMI-1.13-12.1.2-F & GFA-1.12-1.1.1-R	UTX1.35.B	UTX0.BB	UTX1.35
D190TGA	GFA-1.12-1.1.1-F & PMI-1.13-12.1.3-R	UTX1.36.A	PMI-1.13-12.1.3-F & GFA-1.12-1.1.1-R	UTX1.36.B	UTX0.BB	UTX1.36
Q204TAA	GFA-1.12-1.1.1-F & PMI-1.13-13.1.1-R	UTX1.37.A	PMI-1.13-13.1.1-F & GFA-1.12-1.1.1-R	UTX1.37.B	UTX0.BB	UTX1.37
Q204TAG	GFA-1.12-1.1.1-F & PMI-1.13-13.1.2-R	UTX1.38.A	PMI-1.13-13.1.2-F & GFA-1.12-1.1.1-R	UTX1.38.B	UTX0.BB	UTX1.38
Q204TGA	GFA-1.12-1.1.1-F & PMI-1.13-13.1.3-R	UTX1.39.A	PMI-1.13-13.1.3-F & GFA-1.12-1.1.1-R	UTX1.39.B	UTX0.BB	UTX1.39
N212TAA	GFA-1.12-1.1.1-F & PMI-1.13-14.1.1-R	UTX1.40.A	PMI-1.13-14.1.1-F & GFA-1.12-1.1.1-R	UTX1.40.B	UTX0.BB	UTX1.40
N212TAG	GFA-1.12-1.1.1-F & PMI-1.13-14.1.2-R	UTX1.41.A	PMI-1.13-14.1.2-F & GFA-1.12-1.1.1-R	UTX1.41.B	UTX0.BB	UTX1.41
N212TGA	GFA-1.12-1.1.1-F & PMI-1.13-14.1.3-R	UTX1.42.A	PMI-1.13-14.1.3-F & GFA-1.12-1.1.1-R	UTX1.42.B	UTX0.BB	UTX1.42
K214TAA	GFA-1.12-1.1.1-F & PMI-1.13-15.1.1-R	UTX1.43.A	PMI-1.13-15.1.1-F & GFA-1.12-1.1.1-R	UTX1.43.B	UTX0.BB	UTX1.43
K214TAG	GFA-1.12-1.1.1-F & PMI-1.13-15.1.2-R	UTX1.44.A	PMI-1.13-15.1.2-F & GFA-1.12-1.1.1-R	UTX1.44.B	UTX0.BB	UTX1.44

K214TGA	GFA-1.12-1.1.1-F & PMI-1.13-15.1.3-R	UTX1.45.A	PMI-1.13-15.1.3-F & GFA-1.12-1.1.1-R	UTX1.45.B	UTX0.BB	UTX1.45

2.6.2. G.E.M.S. Plasmids for Two-Site UAA Incorporation

For quantifying the efficiency of two-site UAA incorporation, TAG and TGA stop codons were simultaneously introduced in the eGFP gene at two different amino acid positions. The TAG stop codon was always upstream of the TGA stop codon. In this way 10 different G.E.M.S. Plasmids (namely *Plasmids UTX2.23* to *UTX2.26*, *UTX2.34* to *UTX2.36, UTX2.45, UTX2.46* and *UTX2.56*) were created by Gibson Assembly assisted mutagenesis. The details of the introduced mutations are listed in Table 5.

For creating the G.E.M.S. Plasmids for Single-Site UAA Incorporation Screening, the mammalian expression *Plasmid UTX0* was linearized by simultaneous digestion with restriction enzymes PvuI-HF & XbaI and 6268 bp linear DNA fragment backbone resulting from the digestion was purified as linear DNA *UTX0.BB*. For each of the expression vector, two linear DNA fragments (*GA Insert Fragment 1* and *GA Insert Fragment 2*) were created by preparative-PCR using the respective template and primer pairs, as mentioned in Table 5. The template DNA was removed by DpnI digestion after PCR and the PCR-amplified GA Insert Fragments were purified (without agarose-gel purification) using the Wizard® SV Gel and PCR Clean-Up System. Eventually, the purified *GA Insert Fragments 1 and 2* were inserted in the linear DNA *UTX0.BB* via Gibson Assembly. The resulting expression vectors were selected and amplified in NEB® 10-beta *E. coli* chemically competent cells using 20 μg/mL gentamycin as the selection antibiotic.

Table 5: List of Expression Vectors for UAA Incorporation at Two Sites in eGFP

eGFP Mutation	**GA Insert Fragment 1**			**GA Insert Fragment 2**			**Gibson Assembly Backbone**	**G.E.M.S. Plasmid**
	Template	**Primer Pair 1**	**Fragment Name**	**Template**	**Primer Pair 2**	**Fragment Name**		
Y39TAG-K52TGA	UTX1.05	GFA-1.12-1.1.1-F & PMI-1.13-3.1.3-R	UTX2.23.A	UTX1.05	PMI-1.13-3.1.3-F & GFA-1.12-1.1.1-R	UTX2.23.B	UTX0.BB	UTX2.23
Y39TAG-F99TGA	UTX1.05	GFA-1.12-1.1.1-F & PMI-1.13-4.1.3-R	UTX2.24.A	UTX1.05	PMI-1.13-4.1.3-F & GFA-1.12-1.1.1-R	UTX2.24.B	UTX0.BB	UTX2.24
Y39TAG-V176TGA	UTX1.05	GFA-1.12-1.1.1-F & PMI-1.13-10.1.3-R	UTX2.25.A	UTX1.05	PMI-1.13-10.1.3-F & GFA-1.12-1.1.1-R	UTX2.25.B	UTX0.BB	UTX2.25
Y39TAG-Q204TGA	UTX1.05	GFA-1.12-1.1.1-F & PMI-1.13-13.1.3-R	UTX2.26.A	UTX1.05	PMI-1.13-13.1.3-F & GFA-1.12-1.1.1-R	UTX2.26.B	UTX0.BB	UTX2.26
K52TAG-F99TGA	UTX1.08	GFA-1.12-1.1.1-F & PMI-1.13-4.1.3-R	UTX2.34.A	UTX1.08	PMI-1.13-4.1.3-F & GFA-1.12-1.1.1-R	UTX2.34.B	UTX0.BB	UTX2.34
K52TAG-V176TGA	UTX1.08	GFA-1.12-1.1.1-F & PMI-1.13-10.1.3-R	UTX2.35.A	UTX1.08	PMI-1.13-10.1.3-F & GFA-1.12-1.1.1-R	UTX2.35.B	UTX0.BB	UTX2.35
K52TAG-Q204TGA	UTX1.08	GFA-1.12-1.1.1-F & PMI-1.13-13.1.3-R	UTX2.36.A	UTX1.08	PMI-1.13-13.1.3-F & GFA-1.12-1.1.1-R	UTX2.36.B	UTX0.BB	UTX2.36
F99TAG-V176TGA	UTX1.11	GFA-1.12-1.1.1-F & PMI-1.13-10.1.3-R	UTX2.45.A	UTX1.11	PMI-1.13-10.1.3-F & GFA-1.12-1.1.1-R	UTX2.45.B	UTX0.BB	UTX2.45

F99TAG-Q204TGA	UTX1.11	GFA-1.12-1.1.1-F & PMI-1.13-13.1.3-R	UTX2.46.A	UTX1.11	PMI-1.13-13.1.3-F & GFA-1.12-1.1.1-R	UTX2.46.B	UTX0.BB	UTX2.46
V176TAG-Q204TGA	UTX1.29	GFA-1.12-1.1.1-F & PMI-1.13-13.1.3-R	UTX2.56.A	UTX1.29	PMI-1.13-13.1.3-F & GFA-1.12-1.1.1-R	UTX2.56.B	UTX0.BB	UTX2.56

2.7. CONTROL PLASMIDS FOR SCREENING EXPERIMENTS

For the control experiments, a WT-eGFP expressing mammalian vector *Plasmid MX01.3* and a WT-mRaspberry expressing mammalian vector *Plasmid MX01.4* were also created via Gateway™ cloning.

For expressing eGFP fluorescent protein in mammalian cells, *eGFP* gene present in synthetic gene *Plasmid SYN1.4* was inserted into mammalian expression vector *Plasmid MX01* via Gateway™ Cloning. Specifically, a linear DNA fragment *eGFP.attB12* was PCR-amplified (preparative-PCR) using synthetic gene *Plasmid SYN1.4* template and primers *GWA-Z.B-3.1.1-F* & *GWA-Z.B-3.1.1-R*. After agarose-gel purification, *eGFP* gene present in *mRSP.attB12* was inserted in *Plasmid pDONR221*, using BP-Clonase assisted BP-recombination reaction, to generate the entry clone *GD21.11*. Using LR-Clonase assisted LR-recombination reaction between *GD21.11* and *Plasmid MX01*, *eGFP* gene was inserted in *Plasmid MX01* to generate the mammalian expression vector *Plasmid MX01.3*. *Plasmid MX01.3* was selected and amplified in NEB® 10-beta *E. coli* electrocompetent cells using 20 μg/mL gentamycin as the selection antibiotic.

For expressing mRaspberry fluorescent protein in mammalian cells, *mRaspberry* gene present in synthetic gene *FP-mRaspberry* was inserted into mammalian expression vector *Plasmid MX01* via Gateway™ Cloning. Specifically, a linear DNA fragment *mRSP.attB12* was PCR-amplified (preparative-PCR) using synthetic gene *FP-mRaspberry* template and primers *GWA-Z.B-2.1.1-F* & *GWA-Z.B-2.1.1-R*. After agarose-gel purification, *mRaspberry* gene present in *mRSP.attB12* was inserted in *Plasmid pDONR221*, using BP-Clonase assisted BP-recombination reaction, to generate the entry clone *GD21.12*. Using LR-Clonase assisted LR-recombination reaction between *GD21.12* and *Plasmid MX01*, *mRaspberry* gene was inserted in *Plasmid MX01* to generate the mammalian expression vector *Plasmid MX01.4*. *Plasmid MX01.4* was selected and amplified in NEB® 10-beta *E. coli* electrocompetent cells using 20 μg/mL gentamycin as the selection antibiotic.

2.8. PLASMID PURIFICATION FOR TRANSIENT TRANSFECTION

Transfection-grade DNA was purified using NucleoBond® (*Macherey-Nagel*) anion exchange resin present either in Xtra-Maxi Columns or in Xtra-Midi Columns*. The DNA purification buffers (Buffers S1, S2, S3, N2, N3 and N5) were prepared inhouse, according to the buffer composition detailed in the supplementary information section 'Common Buffers' Composition'. Sterile Terrific Broth medium was prepared according to the manufacturer recommended protocol.

2.8.1. GigaPrep Plasmid Purification of G.C.E. Plasmids and Screening-Control Plasmids

Plasmids UCAP.01 to *UCAP.18*, *UCAP.20* to *UCAP.22*, *UTX0*, *MX01.3* and *MX01.4*, which were meant to be used repeatedly across multiple experiments, were purified using an optimized GigaPrep purification protocol (yield: 5-10 mg plasmid DNA per purification). For purifying low copy-number plasmids (*Plasmids UCAP.01* to *UCAP.18* and *UCAP.20* to *UCAP.22*), 3000 mL of sterile Terrific Broth medium, divided into two 5000 mL baffled base flasks and supplemented with 200 μg/mL carbenicillin disodium salt, was inoculated with 250 μL (each flask) of thawed glycerol stocks of cells. For purifying high copy-number plasmids (*Plasmids UAT0*, *MX01.3* and *MX01.4*), 1500 mL of sterile Terrific Broth medium, present in 5000 mL baffled base flask and

* Xtra-Maxi and Xtra-Midi are gravity flow columns and should not be used with vacuum manifold.

supplemented with 20 μg/mL gentamycin was inoculated with 200 μL thawed glycerol stocks of cells. After inoculation of the Terrific Broth medium, cells were allowed to grow at 37°C with agitation (85 rpm shaking for 50 mm shaker throw). After 16-18 hours of growth, the cells were harvested at 3000 *xg*.

Cells harvested from 1500 mL growth medium were resuspended in 50 mL Resuspension Buffer S1 by gentle shaking and were lysed using 50 mL Lysis Buffer S2 for 4-5 minutes. During cells lysis, homogeneous distribution of the lysis buffer was ensured by the homogeneous distribution of the blue color of the pH indicator thymol blue*. Cell lysis was stopped by adding 50 mL Neutralization Buffer S3. After neutralization the color of the cell lysate changed from blue to orange (due to thymol blue pH indicator). In this way, from 1500 mL cells, 150 mL of neutralized cell lysate was obtained. For binding the DNA to the NucleoBond® anion exchange resin, the neutralized cell lysate was passed through the Xtra-Maxi Column, while the cell debris was retained by the filter (provided with the Xtra-Maxi Column). For purifying high copy-number plasmids, 75 mL of neutralized cell lysate was passed through one Xtra-Maxi Column. For purifying low copy-number plasmids, 150 mL of neutralized cell lysate was passed through one Xtra-Maxi Column. In this way, the entire neutralized cell lysate was through one Xtra-Maxi Columns. After all the lysate was passed through the binding columns, each column was washed with 25 mL Equilibration Buffer N2. Afterwards, the debris filter was removed, and each column was washed with 35 mL Wash Buffer N3. After all the Wash Buffer passed through the column, 5 mL of prewarmed (60°C) Elution Buffer N5 was added to each of the columns and the flowthrough was collected. The columns were incubated in the Elution Buffer for 30 minutes, after which additional 5 mL of prewarmed Elution Buffer N5 was added to each of the columns for the elution of the DNA. After elution, the contents of both the columns were combined in a 50-mL centrifuge tube and DNA concentration of the combined elute was measured.

For the concentration of DNA, 14 mL of 100% isopropanol was added to the 20 mL of the DNA elution. After gently mixing by inversion, the mixture was placed on ice for 30 minutes for precipitating the DNA. The precipitated DNA was pelleted by centrifugation at 20,000 *xg* for 60 minutes at 2°C. The supernatant was discarded, and the DNA pellet was washed twice with 10 mL freshly prepared chilled ethanol solution (80% v/v in Milli-Q) while pelleting the DNA precipitate by centrifugation (20,000 *xg* for 60 minutes at 2°C) between each wash. After the second ethanol-wash, the centrifuge tube was moved into a sterile laminar flow hood without removing the supernatant†. Once in the sterile environment, all the supernatant was carefully aspirated without disturbing the DNA pellet and the DNA pellet was allowed to dry for 2-10 hours‡. 3 mL of sterile TE buffer was added to the dried pellet of DNA. To facilitate the solubilization of DNA, the centrifuge tube was left undisturbed at 37°C for 12-16 hours. After all DNA was redissolved, its concentration was measured (while maintaining sterility) and by using TE buffer, the DNA was diluted to a final concentration of 1 μg/μL (or 1 mg/mL). For long-term storage, the DNA was distributed as 500 μL aliquots in 2 mL microcentrifuge tubes and was stored at -80°C after flash freezing in liquid nitrogen. For use within the month, the DNA was stored at 4°C.

2.8.2. MidiPrep Plasmid Purification of G.E.M.S. Plasmids

Plasmids UTX1.01 to *UTX1.45*, *UTX2.23* to *UTX2.26*, *UTX2.34* to *UTX2.36*, *UTX2.45*, *UTX2.46* and *UTX2.56*, which were used for screening experiments, were purified using an optimized MidiPrep purification protocol (yield: 250-500 μg plasmid DNA per purification). 100 mL of sterile Terrific Broth medium, supplemented with 20 μg/mL gentamycin was inoculated with 50 μL thawed glycerol stocks of cells. After inoculation, cells were allowed to grow at 37°C with agitation (200 rpm shaking). After 14-16 hours of growth, the cells were harvested at 3000 *xg*. Cells were then resuspended in 10 mL Resuspension Buffer S1 by gentle shaking and were lysed for 3-4 minutes using 10 mL Lysis Buffer S2. Cell lysis was stopped by adding 10 mL Neutralization Buffer S3. The neutralized cell lysate was passed through the Xtra-Midi Column, while the cell debris was retained by the filter. After all the lysate was passed through, the binding column was washed with 5 mL Equilibration Buffer N2. Afterwards, the debris filter was removed, and the

* Thymol blue is a toxic chemical and must be handled with proper protection of the skin and eyes. In this protocol it is used only as a homogeneity indicator and has no effect on the purification steps. One can choose not to add it in the Lysis Buffer S2.

† After the second ethanol-wash, the DNA is considered sterilized (because of the ethanol solution) and must be handled as a sterile material.

‡ After 2-10 hours, the white pellet of the precipitated DNA turns transparent, marking the end of the drying process. DNA pellet should be monitored every hour to prevent over-drying of DNA.

column was washed with 10 mL Wash Buffer N3. After all wash buffer was passed through the column, 5 mL of prewarmed (60°C) Elution Buffer N5 was used for eluting the DNA in a 50-mL centrifuge tube and the DNA concentration in the elution was measured. For the concentration of DNA, 3.5 mL of 100% isopropanol was added to the 5 mL of the DNA elution. After gently mixing by inversion, the mixture was placed on ice for 30 min for precipitating the DNA. The precipitated DNA was pelleted by centrifugation at 20,000 *xg* for 30 minutes at 2°C. The supernatant was discarded, and the DNA pellet was transferred to a 2 mL microcentrifuge tube. The DNA pellet was washed twice with 1 mL freshly prepared chilled ethanol solution (80% v/v in Milli-Q) while pelleting the DNA precipitate by centrifugation (20,000 *xg* for 30 minutes at 2°C) between each wash. After the second ethanol-wash, the microcentrifuge tube was moved into a sterile laminar flow hood without removing the supernatant. Once in the sterile environment, all the supernatant was carefully aspirated without disturbing the DNA pellet and the DNA pellet was allowed to dry for 30-120 minutes, while checking the status of drying every 30 minutes. 200 µL of sterile TE buffer was added to the dried pellet of DNA and the microcentrifuge tube was left undisturbed at 37°C for 6-8 hours. After all DNA was redissolved, its concentration was measured (while maintaining sterility) and by using TE buffer, the DNA was diluted to a final concentration of 1 µg/µL (or 1 mg/mL). Since MidiPrep purified DNA was usually consumed within the month, it was stored at 4°C.

2.9. WORKFLOW AUTOMATION OF CELL-BASED ASSAY

Since we wanted to screen over 500 conditions (each condition was present in sextuplicate, hence over 3000 individual measurements were planned), we decided to automate the workflow of our screening experiment to ensure reproducibility of operations, minimize human error due to fatigue and for round-the-clock data acquisition. For this a MultiFloFX-BioSpa-Cytation5 Setup (*assembled and calibrated by Agilent-BioTek*) was programmed for sample handling and data acquisition using 98 well plates and 384 well plates.

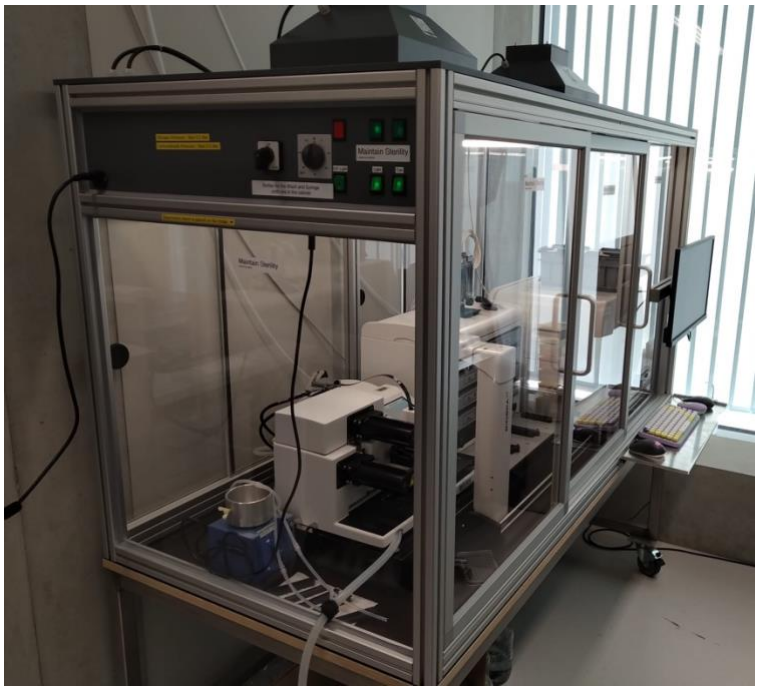


Figure 10: Workflow Automation Setup in Custom-Made Sterile Hood.

2.9.1. Instrument Setup and Sterilization

The MultiFloFX-BioSpa-Cytation5 Setup was placed in a custom-made sterile cabinet (*assembled and calibrated by Biozentrum Workshop*). For cell-based assays, the setup was sterilized before every

Figure 11: Workflow Automation of Cell-Based Assay.
Processes: Cell seeding (1), Transient transfection (2), Incubation for protein expression (3), Green- & Red-channel fluorescence imaging (4), Image analysis & quality control (5), Data extraction (6), Cell lysis & endpoint fluorescence measurement (7), Statistical analysis (8, 9) and Graphical representation (10).
Labels: Adherent HEK cells (A), MultiFloFX & computer (B, C), Multi-well plates containing the seeded cells (D), Programmable multichannel pipettes (E), Multi-well plates containing the transfected cells (F), Cytation5 plate imager & computer (G, H), Fluorescence images (I), Image-analysis on the HPC cluster (J) by using CellProfiler (K), Image-analysis quality control (L), Data from image-analysis (M), Cells at the end of the experiment (N), MultiFloFX-Cytation5 Setup & computer (O, P, Q), Data from the end-point fluorescence measurement (R), Statistical-Analysis on the HPC cluster (S) using R scripts (T) and Final graphs (U).

experiment. Inner and outer surfaces of the BioSpa automated incubator and Cytation5 multimode reader were sterilized by wiping with 70% ethanol solution. The tubing and pumps of the MultiFloFX liquid handler were sterilized by sequentially washing them with 70% isopropanol, 0.2 N NaOH (with soaking for 10 minutes), 0.2 N HCl, 70% isopropanol (with soaking for 5 minutes) and autoclaved Milli-Q water. Finally, the cabinet and the instrument surfaces were sterilized by irradiating with UV light for 2 hours.

2.9.2. Poly-L-Lysine Treatment

For Poly-L-Lysine treatment of 384 well plates, *Scripts ACA.1.1*, *ACA.1.2*, *ACA.1.3* and *ACA.1.4* were created for the MultiFloFX liquid handler. For user-free automated execution of these scripts, *Pipeline ACA.1* was created for the BioSpa automated incubator. In short, while executing this pipeline, the MultiFloFX dispenses 70 µL of 0.001% Poly-L-Lysine in all the wells of up to eight 384 well plates, the plates are incubated for 2 hours at 37°C and then MultiFloFX washes all the wells of the plates with 0.0002% Poly-L-Lysine before drying the plates and making them ready for cell seeding.

2.9.3. Cell Seeding

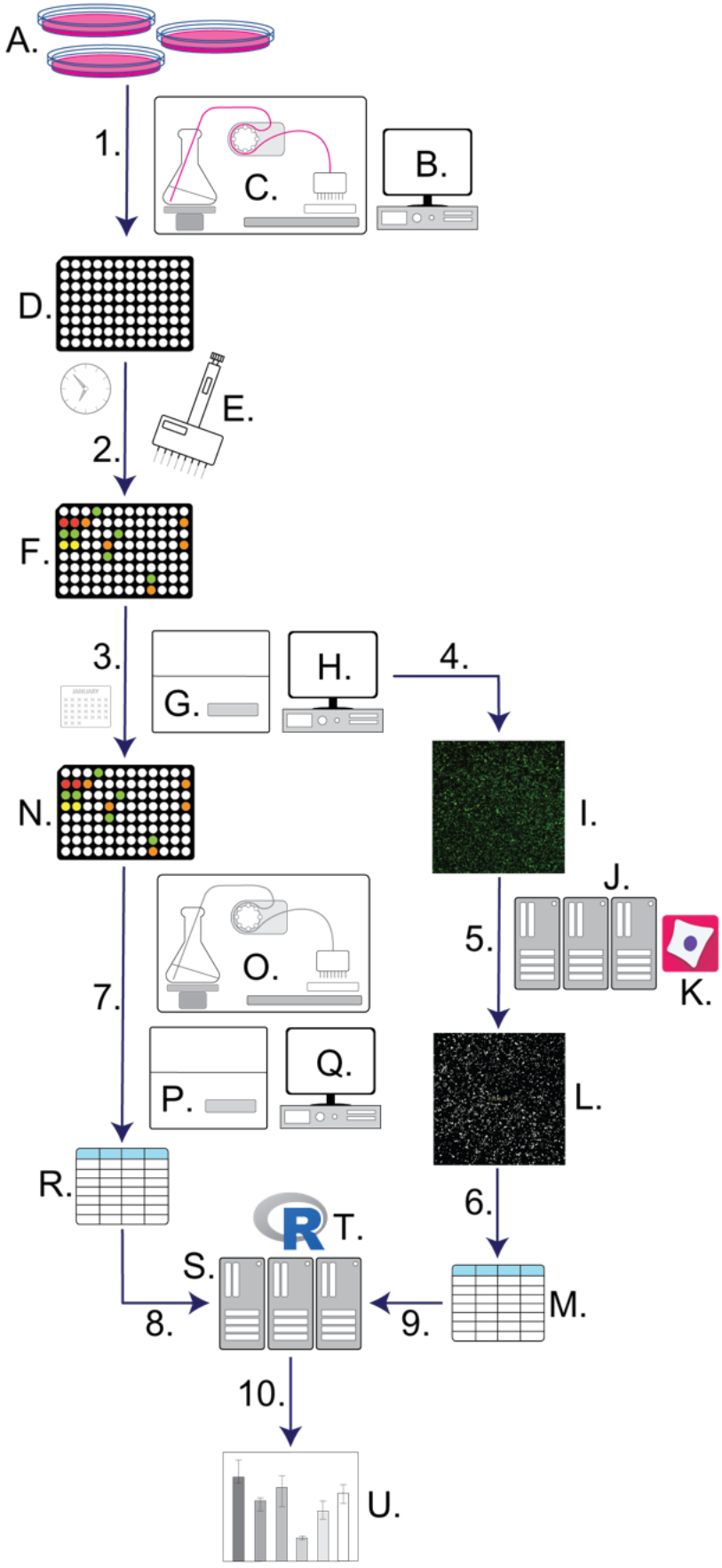
For seeding the suspended mammalian cells in 384 well plates, *Scripts ACA.2.1* and *ACA.2.2* were created for the MultiFloFX liquid handler. For user-free automated execution of these scripts, *Pipeline ACA.2* was created for the BioSpa automated incubator. In short, while executing this pipeline, the MultiFloFX dispenses 50 µL of suspended mammalian cells in all the wells of up to eight 384 well plates and the plates are incubated in the BioSpa automated incubator (37°C, 5% CO_2 and 90% humidity) until next use.

2.9.4. Media Exchange for Transient Transfection

Scripts ACA.3.0, *ACA.3.1*, *ACA.3.2* and *ACA.3.3* were created for the MultiFloFX liquid handler. For user-free automated media exchange in up to eight 384 well plates, *Pipeline ACA.3* was created for the BioSpa automated incubator. However, addition of DNA (for transient transfection) and unnatural amino acids (for protein expression) still required human intervention. In short, while executing this pipeline, the MultiFloFX first replaces the cell growth media with 25 µL transfection media (hence allowing the user to add DNA-PEI complex for transient transfection), and then, after 5 hours, the MultiFloFX replaces the transfection media with 45 µL cell growth media. After the second media exchange, the user adds unnatural amino acids to the wells.

2.9.5. Image Acquisition, Cell Lysis and Endpoint Fluorescence Data

For acquiring well-images and endpoint fluorescence of 384 well plates, *Protocols ACA.1*, *ACA.2* and *ACA.3* were created for the Cytation5 multimode reader. Additionally, *Scripts ACA.4.1* and *ACA.4.2* were created for the MultiFloFX liquid handler. For user-free automated execution of these scripts, *Pipeline ACA.4* was created for the BioSpa automated incubator. In short, while executing this pipeline, the Cytation5 acquires brightfield images of the wells immediately after transfection. Then, 36- and 72-hours post-transfection, the Cytation5 acquires the green-channel and red-channel fluorescence images of the wells. After 72 hours, the MultiFloFX lyses the cells by removing all the media from the wells before forcefully adding 60 µL lysis buffer in the wells. Finally, the Cytation5 acquires the endpoint fluorescence data for the green-channel and the red-channel.


A.
1.
B.
C.
D.
E.
2.
F.
3.
G.
H.
4.
N.
I.
J.
5.
K.
7.
O.
P.
Q.
L.
R.
6.
T.
S.
M.
8.
9.
10.
U.

2.10. SCREENING EXPERIMENTS IN MULTI-WELL PLATES

2.10.1. Poly-L-Lysine Treatment of Multi-Well Plates

To prevent the loss of cells during the washing or media exchange steps of the screening experiments, 96- and 384-well plates were treated with Poly-L-Lysine. All the operations of Poly-L-Lysine treatment were performed under sterile conditions in a laminar flow cabinet. For 96-well cell-culture plates, 100 µL of sterile 0.001% Poly-L-Lysine solution was dispensed in each of the wells using a programmable multichannel pipette (*Integra Voyager*). After incubating the (lid-covered) plates for 30 minutes at 4°C, all Poly-L-Lysine solution was removed from the wells and the wells were gently washed once with 350 µL sterile PBS solution. For Poly-L-Lysine treatment of 384-well cell-culture plates, empty 384-well plates were loaded in the BioSpa automated incubator, presterilized Syringe-A tubing of the MultiFloFX liquid handler was placed in sterile 0.001% Poly-L-Lysine solution, presterilized Syringe-B tubing of the MultiFloFX liquid handler was placed in sterile 0.0002% Poly-L-Lysine solution and *BioSpa Pipeline ACA.1* was executed.

2.10.2. Cell Seeding

HEK293 cells were used for screening experiments. By repeated passaging (as described in the supplementary information subsection 'Maintenance of Adherent Culture of Mammalian Cells') cells were grown in eight of the 150 mm Treated Cell Culture (TCC) plates. At nearly 70% confluence, the cells in each of the eight plates were washed once with 10 mL sterile PBS and 3 mL of prewarmed (37°C) Trypsin-EDTA was added to each of the TCC plates for dislodging the cells. After 3-4 minutes of incubation at room temperature, Trypsin was deactivated using 7 mL SBTI solution (1 mg/ml SBTI in DMEM) and cells were suspended by pipetting 10-15 times. Cell-suspension was then transferred to a sterile flask. After measuring the cell-density, cells were diluted to a final concentration of 0.25 x 10^6 cells/ml in C-DMEM (DMEM supplemented with 10% FBS). In this way, 350 mL cell suspension was prepared for cell-seeding in Poly-L-Lysine treated multi-well plates. For preventing cell aggregation and to keep them suspended, the cell-suspension was continuously agitated, either by hand or mechanically using a custom made low-rpm shaker. For manual dispensing of cells in 96-well cell-culture plates, 200 µL of cells were dispensed in each of the wells using a programmable 8-channel pipette. For automated dispensing of cells in 384-well cell-culture plates, the Poly-L-Lysine treated 384-well plates were loaded (if not already present) in the BioSpa automated incubator, presterilized Syringe-A tubing of the MultiFloFX liquid handler was placed in the cell suspension and *BioSpa Pipeline ACA.2* was executed. For both 96- and 384-well plates, cells were allowed to properly attach to the plates by incubation at 37°C in a humidified CO_2 incubator (5% CO_2 and ~90% humidity) for 6-8 hours before they were processed further for transient transfection.

2.10.3. Transient Transfection and Recovery

(Unless specified otherwise, for two-plasmid co-transfection, the individual plasmids were present in 1:1 ratio, and the individual stocks of DNA were premixed in this ratio to ensure sample homogeneity during pipetting. For three-plasmid co-transfection using a G.E.M.S. Plasmid and two G.C.E. Plasmids (see Section 3.3 for explanation), these plasmids were premixed in the ratio of 2:1:1 respectively to ensure sample homogeneity during pipetting).

For transient transfection of HEK293 cells, sterile reagents were used, and all operation were performed under sterile conditions in a laminar flow cabinet. For the transfection of cells seeded in 96-well cell-culture plates, first the DNA dilution was prepared in 96-well PCR-plates according to the screening conditions (as detailed in the section 'Screening Conditions'[*]). This was done by mixing 20 µL DMEM and 5 µg of transfection-grade DNA (stock concentration 1 µg/µL). To this DNA dilution, 25 µL of filtered sterile PEI solution (0.2 mg/ml PEI Max® in DMEM) was added and mixed three times by pipetting. For DNA-PEI

[*] A printed layout of the multi-well plates was used as a 'cheat-sheet' to keep track of the different screening conditions.

complex formation, the PCR-plates were incubated for 15 minutes at room temperature. Meanwhile, the 96-well cell-culture plates containing the seeded cells were moved to the laminar flow cabinet and culture-medium (C-DMEM) present in the wells was replaced with 180 µL transfection-medium (DMEM). After the completion of the DNA-PEI complex formation, by using a programmable multichannel pipette, 20 µL of DNA-PEI mix was gently added to the (respective) wells and gently mixed once (without disturbing the cells). In this way 2 µg DNA was added to each of the wells.

To facilitate the transient transfection, cells were left undisturbed for 5 hours at 37°C in a humidified CO_2 incubator (5% CO_2 and ~90% humidity). During this time, the protein expression medium was prepared in another 96-well PCR-plate by dividing 220 µL of UAA supplemented C-DMEM according to the screening conditions. After 5 hours of incubation, the transfection-media from the wells was gently replaced with 200 µL of the protein expression media. Cell-culture plates were then replaced in the humidified CO_2 incubator (37°C, 5% CO_2 and ~90% humidity) for recovery and protein expression.

For the transfection of cells seeded in 384-well cell-culture plates, first the DNA dilution was manually prepared in 384-well PCR-plates according to the screening conditions (as detailed in the section 'Screening Conditions'). This was done by mixing 30 µL DMEM and 0.7 µg of DNA (stock concentration 350 ng/µL, prepared by diluting the transfection grade DNA). To this DNA dilution, by using a programmable multichannel pipette, 3 µL of filtered sterile PEI solution (0.7 mg/ml PEI Max® in DMEM) was added and mixed five times by pipetting. For DNA-PEI complex formation, the PCR-plates were incubated for 15 minutes at room temperature. Meanwhile, *BioSpa Pipeline ACA.3* was executed on the MultiFloFX-BioSpa-Cytation5 Setup to aspirate the culture-medium (C-DMEM) present in the wells and replace it with 25 µL transfection-medium (DMEM). After the automated media-exchange and the DNA-PEI complex formation were completed, the cell-culture plate was moved into the laminar flow cabinet and by using a programmable multichannel pipette, 25 µL of DNA-PEI mix was gently added to the (respective) wells without disturbing the cells. In this way 0.5 µg DNA was added to each of the wells.

To facilitate the transient transfection, the cell-culture plate was replaced into the BioSpa automated incubator (37°C, 5% CO_2 and ~90% humidity) and was left undisturbed for 5 hours (the 5-hour transfection incubation is built in the *BioSpa Pipeline ACA.3*). During this time, for each of the needed unnatural amino acids, 4x-expression-media was separately prepared in 50 mL centrifuge tubes by diluting the 100x UAA-stocks solution to 4x UAA-solutions in C-DMEM. Further, for each cell-culture plate, according to the screening conditions, the 4x-expression-medium was divided in 96-well PCR-plate as 120 µL aliquots*. After 5 hours of incubation, because of the ongoing *BioSpa Pipeline ACA.3* the MultiFloFX-BioSpa-Cytation5 Setup aspirated the transfection-medium (DNA-PEI-DMEM) present in the wells and replaced it with 45 µL growth-medium (C-DMEM). After the automated media-exchange was completed, the cell-culture plate was moved into the laminar flow cabinet and by using a programmable multichannel pipette, 15 µL of the 4x-expression-media was gently added to the respective wells without disturbing the cells. After the addition of UAA, cell-culture plates were then replaced in the humidified CO_2 incubator (37°C, 5% CO_2 and ~90% humidity) for recovery and protein expression.

2.10.4. Image Acquisition

Soon after the completion of the transient transfection of cells, brightfield images of all the wells were collected on the Cytation5 multimode plate reader, either manually (only for 96-well cell-culture plates) or by using the *BioSpa Pipeline ACA.4*. The initial brightfield images were collected using the 4x phase-contrast objective, 70% LED brightness, 0 gain, default autofocus protocol (*i.e.*, autofocus is set by finding the sharpest image) and with an exposure time of 10 ms.

For 96-well cell-culture plates, green-channel and red-channel fluorescence images were manually collected 24-, 48- and 72-hours post-transfection. The green-channel images were collected using the BioTek filter cube 1225101 (*Excitation Filter: 469/35 nm* | *Emission Filter: 525/39 nm* | *Dichroic Mirror: 497 nm long-pass*), 4x phase-contrast objective, 50% LED brightness, 0 gain, default autofocus protocol and with an exposure time of 20 ms. Immediately after, the red-channel images were collected using the BioTek filter cube 1225117 (*Excitation Filter: 531/40 nm* | *Emission Filter: 685/40 nm* | *Dichroic Mirror: 660 nm*

* One of these 120 µL aliquots was meant for six of the wells of the 384-well cell culture plate.

long-pass), 4x phase-contrast objective, 70% LED brightness, 0 gain, default autofocus protocol and with an exposure time of 350 ms.

For 384-well cell-culture plates, because of the ongoing *BioSpa Pipeline ACA.4*, green-channel and red-channel fluorescence images were automatically collected 36- and 72-hours post transfection. The green-channel images were collected using the BioTek filter cube 1225101 (*Excitation Filter: 469/35 nm* | *Emission Filter: 525/39 nm* | *Dichroic Mirror: 497 nm long-pass*), 4x phase-contrast objective, 100% LED brightness, 0 gain, default autofocus protocol and with an exposure time of 15 ms. Immediately after, the red-channel images were collected using the BioTek filter cube 1225117 (*Excitation Filter: 531/40 nm* | *Emission Filter: 685/40 nm* | *Dichroic Mirror: 660 nm long-pass*), 4x phase-contrast objective, 100% LED brightness, 0 gain, default autofocus protocol and with an exposure time of 500 ms.

Collected images are stored as *.tif files in corresponding experiment folders in Maier lab central data folder provided by the sciCore facility*.

2.10.5. Cell Lysis and Endpoint Fluorescence Measurement

After collection of images the cells were lysed using the HEK Cell Lysis Buffer to release the proteins in the medium and for collecting the endpoint fluorescence. For 96-well plates, the culture medium present in the wells was gently aspirated using a programable multichannel pipette and 100 µL HEK Lysis Buffer was added forcefully. For 384-well plates, because of the ongoing *BioSpa Pipeline ACA.4*, the culture medium present in the wells was gently aspirated by the MultiFloFX liquid handler and 60 µL HEK Lysis Buffer, loaded in the secondary peristaltic pump tubing of the MultiFloFX, was forcefully added to the wells. For both kinds of plates, cells were completely lysed by shaking the plates for 3 minutes in the Cytation5 multimode reader. Afterwards, eGFP fluorescence (*Excitation: 465/15 nm* | *Emission: 510/15 nm*) and mRaspberry fluorescence (*Excitation: 570/30 nm* | *Emission: 640/30 nm*) were collected with extended gain of the detector by using the lower optical setup of Cytation5.

2.10.6. Image Analysis

Images collected using the MultiFloFX-BioSpa-Cytation5 Setup were processed on the SciCore High Performance Computing Cluster (University of Basel) using pipelines developed for CellProfiler image analysis software (ver. 4.2.1) [236-239]. For 96 well-plates, the raw images of dimension 1992 x 1992 pixels were used for analysis. For 384 well plates, since the well area was smaller than the image area, all images were cropped by 512 pixels from all the four edges (using the CellProfiler pipeline *384_Well_Image_Cropping.cppipe*) and cropped raw images of dimension 968 x 968 pixels were used for analysis. For each plate, the red and green channel images collected from the six untransfected wells were separately averaged (using the CellProfiler pipeline *Blank_Image_Average.cppipe*) to get the 'blank-image' for the red- and green-channel respectively. Afterwards, all images were analyzed (using the CellProfiler pipeline *Background_Subtraction_and_Analysis.cppipe*).

In short, for *Background_Subtraction_and_Analysis.cppipe*, the blank-images (obtained by averaging the images of the untransfected wells) of the red and green channel respectively were subtracted from red and green channel raw images (for both cropped and non-cropped), to get the 'blank-corrected-images'. Using the red-channel blank-corrected-images, individual objects (single cells) 7-40 pixels in diameter† were identified by using a global manual threshold of 4%‡. From these identified objects, cells containing saturated pixels in the red and/or green channels were identified and removed by using a global manual threshold of 98% on the raw images. Eventually, the red-channel intensity and green-channel intensity was measured for each 'non-saturated object' and the results were exported as text files (csv format).

These scripts are available via Zenodo electronic repository (https://doi.org/10.5281/zenodo.8351065).

---

* Cytation5/Shubham_PhD_Thesis/

† This range was obtained after optimization, to discard cell-debris and very large clumps of cells.

‡ Threshold of 4% means that the pixels below 4% of the saturation value were discarded for object identification.

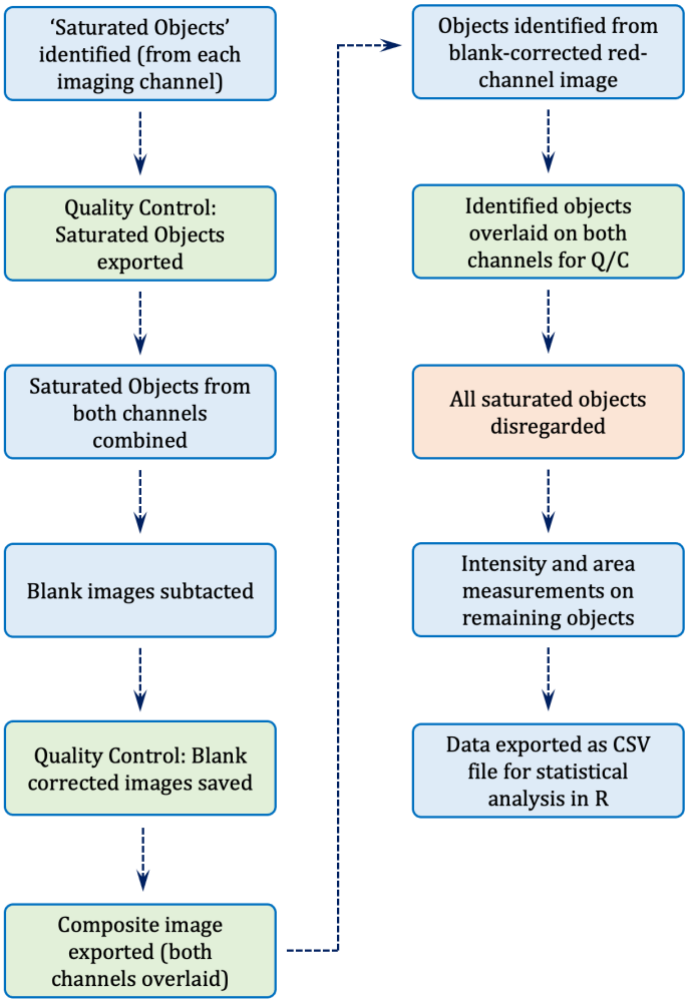


Figure 12: Image Processing Steps Involved in CellProfiler Application pipeline *Background_Subtraction_and_Analysis.cppipe*.

2.10.7. Statistical Analysis

Endpoint fluorescence measurements and CellProfiler image analysis output were statistically analyzed using R (ver. 4.1.2) on the SciCore High Performance Computing Cluster (University of Basel). R scripts for the analysis were written in R-Studio (ver. 1.4.1717-3).

For statistical analysis of endpoint fluorescence measurements, *Endpoint_Fluorescence_Analysis.R* script was used, which classifies the endpoint fluorescence data according to the conditions used in the experiment, calculates the Expression Ratio (total eGFP fluorescence to total mRaspberry fluorescence ratio) for each well and normalizes the expression ratio to the *Plasmid UTX0* condition (see Section 2.11.1).

For statistical analysis of data exported by CellProfiler pipelines, *CellProfiler_Output_Analysis.R* script was used. In short, the data exported from CellProfiler pipeline *Background_Subtraction_and_Analysis.cppipe* is sorted into respective conditions, some data is filtered out (*e.g.*, if number of identified objects in the individual wells are not statistically significant or if a cutoff is needed for total mRaspberry or eGFP fluorescence to consider individual objects for further processing), Expression Ratio is calculated for the individual objects and the whole wells, the whole-well expression ratio is normalized to the *Plasmid UTX0* condition (see Section 2.11.1) to get Normalized Expression Ratio (N.E.R.), average N.E.R. is calculated for each condition form the sextuplicate and the graphs are plotted for quality control.

For plotting the graphs displayed in this thesis, *Final_Graph.R* script was used.

These scripts are available via Zenodo electronic repository (https://doi.org/10.5281/zenodo.8351065).

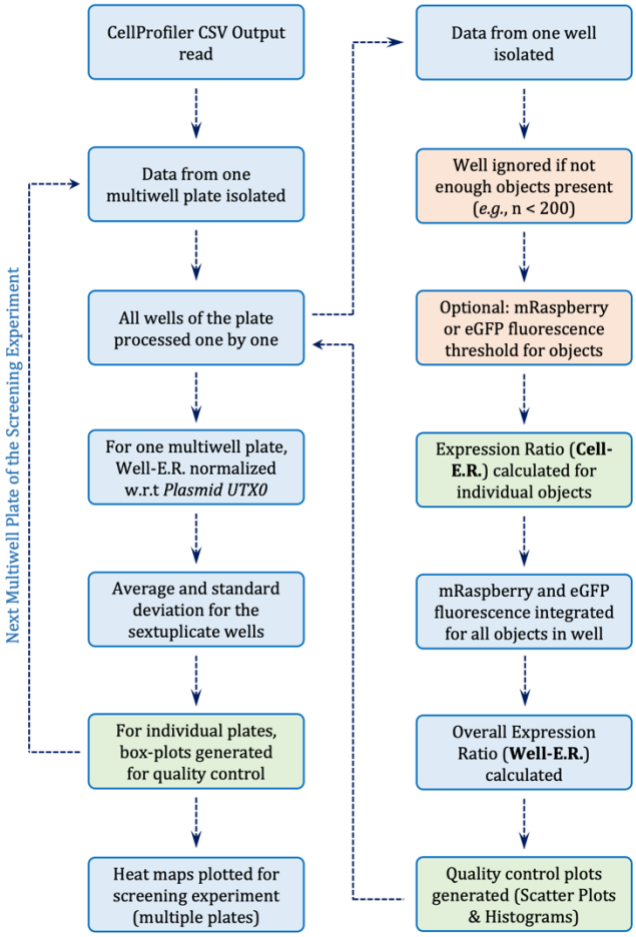


Figure 13: Data Analysis Steps Involved in the R Script *CellProfiler_Output_Analysis.R*.

2.11. SCREENING CONDITIONS FOR COMPARING UAA INCORPORATION

2.11.1. General Format of Screening Experiments

Each of the screening condition was present as a sextuplicate, *i.e.*, each of the tested condition was replicated in six of the wells of the multi-well cell-culture plates. Irrespective of the number and the nature of screening conditions in any given plate, four control-conditions (also as a sextuplicate) were always present in each of the multi-well plate. These four conditions included six wells each of the non-transfected cells, the cells transfected with *Plasmid MX01.3* (expressing only WT-eGFP), the cells transfected with *Plasmid MX01.4* (expressing only WT-mRaspberry) and the cells transfected with *Plasmid UTX0* (expressing WT-eGFP and WT-mRaspberry from same promoter but as different polypeptides).

Apart from these controls, in one of the 96-well cell culture plates, a maximum of 12 conditions could be screened and in one of the 384-well cell culture plates, a maximum of 60 conditions could be screened. If for any screening experiment, a higher number of conditions were needed, more than one multi-well cell culture plate was used. For automated cell-based assays, a maximum of eight multi-well plates could be processed in one session and if more than eight plates were needed for the screening experiment, the experiment was split over multiple sessions.

2.11.2. Two-Plasmid Co-transfection of Mammalian Cells

For optimizing the simultaneous delivery of two plasmids into the mammalian cells, three different co-transfection conditions were screened in the HEK cells seeded in 96-well cell culture plates. For the first condition, cells were co-transfected with 0.7 μL of *Plasmid MX01.3* & 1.4 μL of *Plasmid MX01.4*, for the second one, with 1.0 μL of *Plasmid MX01.3* & 1.0 μL of *Plasmid MX01.4* and for the third one, with 1.4 μL of *Plasmid MX01.3* & 0.7 μL of *Plasmid MX01.4*.

After the transient transfection, green- and red-channel images were collected as mentioned in Section 2.10.4. The images were analyzed on a local computer using the CellProfiler image analysis pipelines *Blank_Image_Average.cppipe* and *Background_Subtraction_and_Analysis.cppipe*, as mentioned in Section 2.10.6.

The CellProfiler output was analyzed in R (on the local computer) by using *CellProfiler_Output_Analysis.R* script as mentioned in Section 2.10.7. The graphs were plotted in R by using *Final_Graph.R* script as mentioned in Section 2.10.7.

2.11.3. Effects of Plasmid Ratio for UAA Incorporation in Proteins

To understand the effects of the ratio of the G.C.E. Plasmid and the G.E.M.S. Plasmid in the UAA incorporation and P.O.I. production, 12x conditions, as detailed in Table 6, were screened in the HEK cells seeded in 96-well cell culture plates.

After the transient transfection, green- and red-channel images were collected as mentioned in Section 2.10.4. The images were analyzed on a local computer using the CellProfiler image analysis pipelines *Blank_Image_Average.cppipe* and *Background_Subtraction_and_Analysis.cppipe*, as mentioned in Section 2.10.6.

The CellProfiler output was analyzed in R (on the local computer) by using *CellProfiler_Output_Analysis.R* script as mentioned in Section 2.10.7. The graphs were plotted in R by using *Final_Graph.R* script as mentioned in Section 2.10.7.

Table 6: 10x Conditions for Screening the Optimal Plasmid Ratio for Co-transfection

G.C.E. Plasmid	G.E.M.S. Plasmid	Total DNA	UAA
UCAP.03 (0.5 μg)	UTX1.06 (0.5 μg)	1.0 μg	ANAP (100 μM)
UCAP.03 (0.5 μg)	UTX1.06 (2.5 μg)	3.0 μg	ANAP (100 μM)
UCAP.03 (1.0 μg)	UTX1.06 (2.0 μg)	3.0 μg	ANAP (100 μM)
UCAP.03 (1.5 μg)	UTX1.06 (1.5 μg)	3.0 μg	ANAP (100 μM)
UCAP.03 (2.0 μg)	UTX1.06 (1.0 μg)	3.0 μg	ANAP (100 μM)
UCAP.03 (2.5 μg)	UTX1.06 (0.5 μg)	3.0 μg	ANAP (100 μM)
UCAP.21 (0.5 μg)	UTX1.05 (0.5 μg)	1.0 μg	N_3Lys (100 μM)
UCAP.21 (0.5 μg)	UTX1.05 (2.5 μg)	3.0 μg	N_3Lys (100 μM)
UCAP.21 (1.0 μg)	UTX1.05 (2.0 μg)	3.0 μg	N_3Lys (100 μM)
UCAP.21 (1.5 μg)	UTX1.05 (1.5 μg)	3.0 μg	N_3Lys (100 μM)
UCAP.21 (2.0 μg)	UTX1.05 (1.0 μg)	3.0 μg	N_3Lys (100 μM)
UCAP.21 (2.5 μg)	UTX1.05 (0.5 μg)	3.0 μg	N_3Lys (100 μM)

2.11.4. Effects of UAA Concentration

To understand the effects of UAA concentration in the UAA incorporation and P.O.I. production, 24x conditions, as detailed in Table 7, were screened in the HEK cells seeded in 96-well cell culture plates.

After the transient transfection, green- and red-channel images were collected as mentioned in Section 2.10.4. The images were analyzed on a local computer using the CellProfiler image analysis pipelines *Blank_Image_Average.cppipe* and *Background_Subtraction_and_Analysis.cppipe*, as mentioned in Section 2.10.6.

The CellProfiler output was analyzed in R (on the local computer) by using *CellProfiler_Output_Analysis.R* script as mentioned in Section 2.10.7. The graphs were plotted in R by using *Final_Graph.R* script as mentioned in Section 2.10.7.

Table 7: 24x Conditions for Optimization of UAA Concentration for UAA Incorporation

G.C.E. Plasmid	G.E.M.S. Plasmid	UAA (Final Concentration)
UCAP.13	UTX1.04	AzPhe (100 μM)
		AzPhe (5 mM)
		PrPhe (100 μM)
		PrPhe (5 mM)
UCAP.14	UTX1.05	AzPhe (100 μM)
		AzPhe (5 mM)
		PrPhe (100 μM)
		PrPhe (5 mM)
UCAP.15	UTX1.06	AzPhe (100 μM)
		AzPhe (5 mM)

		PrPhe (100 μM)
		PrPhe (5 mM)
UCAP.20	UTX1.04	N_3Lys (100 μM)
		N_3Lys (2 mM)
		PrK (100 μM)
		PrK (2 mM)
UCAP.21	UTX1.05	N_3Lys (100 μM)
		N_3Lys (2 mM)
		PrK (100 μM)
		PrK (2 mM)
UCAP.22	UTX1.06	N_3Lys (100 μM)
		N_3Lys (2 mM)
		PrK (100 μM)
		PrK (2 mM)

2.11.5. UAA Incorporation at Single-Site in eGFP

To determine the UAA incorporation efficiency of the different G.C.E. systems and to evaluate the functional tolerance of these G.C.E. systems by the P.O.I. (eGFP), 450x conditions were screened in the HEK cells seeded in 384-well cell culture plates. For these 450 conditions, UAA incorporation at 15 unique positions on eGFP (as described in Table 8) by 30x different G.C.E. Systems (as described in Table 9) were studied. In addition, 315x negative-control conditions, containing oRS and otRNA but no UAA, were also screened to quantify the non-specific amino acid incorporation at the 15x amino acid positions (Table 8) of eGFP by the 21x oRS-otRNA pairs listed in Table 3. In this way, a total of 765x conditions were screened in two sessions by employing the workflow automation for cell-based assays.

After the transient transfection of cells, green- and red-channel images were collected as mentioned in Section 2.10.4. The images were cropped and analyzed on the high-performance computing cluster using the CellProfiler image analysis pipelines *384_Well_Image_Cropping.cppipe, Blank_Image_Average.cppipe* and *Background_Subtraction_and_Analysis.cppipe*, as mentioned in Section 2.10.6.

The CellProfiler output was analyzed in R (on the high-performance computing cluster) by using *CellProfiler_Output_Analysis.R* script as mentioned in Section 2.10.7. The graphs were plotted in R by using *Final_Graph.R* script as mentioned in Section 2.10.7.

Table 8: 15x Amino Acid Positions on eGFP for One-Site UAA Incorporation Screening

Original Amino Acid & Position	**G.E.M.S. Plasmids** (TAA, TAG, TGA)
THR9	UTX1.01 – UTX1.03
TYR39	UTX1.04 – UTX1.06
LYS52	UTX1.07 – UTX1.09
PHE99	UTX1.10 – UTX1.12
ASP117	UTX1.13 – UTX1.15
GLU132	UTX1.16 – UTX1.18
TYR143	UTX1.19 – UTX1.21

TYR151	UTX1.22 – UTX1.24
GLN157	UTX1.25 – UTX1.27
VAL176	UTX1.28 – UTX1.30
TYR182	UTX1.31 – UTX1.33
ASP190	UTX1.34 – UTX1.36
GLN204	UTX1.37 – UTX1.39
ASN212	UTX1.40 – UTX1.42
LYS214	UTX1.43 – UTX1.45

Table 9: 30x Triplets of G.C.E. System for One-Site UAA Incorporation Screening

oRS	otRNA	UAA
${}^{ANAP}_{EcLeu}RS$	${}_{EcLeu}tRNA_{UUA}$	ANAP (100 μM)
	${}_{EcLeu}tRNA_{CUA}$	
	${}_{EcLeu}tRNA_{UCA}$	
${}^{DanAla}_{EcLeu}RS$	${}_{EcLeu}tRNA_{UUA}$	DanAla (1 mM)
	${}_{EcLeu}tRNA_{CUA}$	
	${}_{EcLeu}tRNA_{UCA}$	
${}_{EcTrp(h9)}RS$	${}_{EcTrp}tRNA_{UUA}$	5-HTP (100 μM)
	${}_{EcTrp}tRNA_{CUA}$	
	${}_{EcTrp}tRNA_{UCA}$	
${}_{EcTrp(h14)}RS$	${}_{EcTrp}tRNA_{UUA}$	
	${}_{EcTrp}tRNA_{CUA}$	
	${}_{EcTrp}tRNA_{UCA}$	
${}_{OMeY}RS$	${}_{OMeY}tRNA_{UUA}$	AzPhe (1 mM)
	${}_{OMeY}tRNA_{CUA}$	
	${}_{OMeY}tRNA_{UCA}$	
	${}_{OMeY}tRNA_{UUA}$	PrPhe (1 mM)
	${}_{OMeY}tRNA_{CUA}$	
	${}_{OMeY}tRNA_{UCA}$	
${}_{MbPyl}RS$	${}_{MbPyl}tRNA_{UUA}$	N_3Lys (1 mM)
	${}_{MbPyl}tRNA_{CUA}$	
	${}_{MbPyl}tRNA_{UCA}$	
	${}_{MbPyl}tRNA_{UUA}$	PrK (1 mM)
	${}_{MbPyl}tRNA_{CUA}$	
	${}_{MbPyl}tRNA_{UCA}$	
${}_{MmPyl}RS$	${}_{MbPyl}tRNA_{UUA}$	N_3Lys (1 mM)
	${}_{MbPyl}tRNA_{CUA}$	
	${}_{MbPyl}tRNA_{UCA}$	
	${}_{MbPyl}tRNA_{UUA}$	PrK (1 mM)
	${}_{MbPyl}tRNA_{CUA}$	
	${}_{MbPyl}tRNA_{UCA}$	

2.11.6. Crosstalk Between Different G.C.E. Systems

For quantifying the UAA incorporation crosstalk between the different G.C.E. systems, four G.E.M.S. Plasmids, namely *Plasmids UTX1.05*, *UTX1.06*, *UTX1.38* and *UTX1.39*, were tested with different G.C.E. Plasmid-UAA pairs. 60x conditions were screened in the HEK cells seeded in 384-well cell culture plates by employing the workflow automation for cell-based assays. Out of these 60 conditions, 20x conditions had mismatch between eGFP stop codon & otRNA anticodon (as detailed in Table 10) and 40x conditions had mismatch between the UAA and the G.C.E. machinery delivered in the cells (as detailed in Table 11).

After the transient transfection of cells, green- and red-channel images were collected as mentioned in Section 2.10.4. The images were cropped and analyzed on the high-performance computing cluster using the CellProfiler image analysis pipelines *384_Well_Image_Cropping.cppipe, Blank_Image_Average.cppipe* and *Background_Subtraction_and_Analysis.cppipe*, as mentioned in Section 2.10.6.

The CellProfiler output was analyzed in R (on the high-performance computing cluster) by using *CellProfiler_Output_Analysis.R* script as mentioned in Section 2.10.7. The graphs were plotted in R by using *Final_Graph.R* script as mentioned in Section 2.10.7.

Table 10: 20x Conditions with Mismatch Between P.O.I. Stop Codon and otRNA Anticodon

G.E.M.S. Plasmid	G.C.E. Machinery	UAA
UTX1.05 (TYR39TAG)	$^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{UCA}$	ANAP (100 µM)
	$_{MmPyl}RS$ / $_{MbPyl}tRNA_{UCA}$	N_3Lys (1 mM)
		PrK (1 mM)
	$_{OMeY}RS$ / $_{OMeY}tRNA_{UCA}$	AzPhe (1 mM)
		PrPhe (1 mM)
UTX1.06 (TYR39TGA)	$^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{CUA}$	ANAP (100 µM)
	$_{MmPyl}RS$ / $_{MbPyl}tRNA_{CUA}$	N_3Lys (1 mM)
		PrK (1 mM)
	$_{OMeY}RS$ / $_{OMeY}tRNA_{CUA}$	AzPhe (1 mM)
		PrPhe (1 mM)
UTX1.38 (GLN204TAG)	$^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{UCA}$	ANAP (100 µM)
	$_{MmPyl}RS$ / $_{MbPyl}tRNA_{UCA}$	N_3Lys (1 mM)
		PrK (1 mM)
	$_{OMeY}RS$ / $_{OMeY}tRNA_{UCA}$	AzPhe (1 mM)
		PrPhe (1 mM)
UTX1.39 (GLN204TGA)	$^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{CUA}$	ANAP (100 µM)
	$_{MmPyl}RS$ / $_{MbPyl}tRNA_{CUA}$	N_3Lys (1 mM)
		PrK (1 mM)
	$_{OMeY}RS$ / $_{OMeY}tRNA_{CUA}$	AzPhe (1 mM)
		PrPhe (1 mM)

Table 11: 40x Conditions with Mismatch Between G.C.E. Machinery and UAA

G.E.M.S. Plasmid	G.C.E. Machinery	UAA
UTX1.05 (TYR39TAG)	${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{CUA}$	AzPhe (1 mM)
		N_3Lys (1 mM)
		PrK (1 mM)
		PrPhe (1 mM)
	${}_{MmPyl}RS$ / ${}_{MbPyl}tRNA_{CUA}$	ANAP (100 μM)
		AzPhe (1 mM)
		PrPhe (1 mM)
	${}_{OMeY}RS$ / ${}_{OMeY}tRNA_{CUA}$	ANAP (100 μM)
		N_3Lys (1 mM)
		PrK (1 mM)
UTX1.06 (TYR39TGA)	${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{UCA}$	AzPhe (1 mM)
		N_3Lys (1 mM)
		PrK (1 mM)
		PrPhe (1 mM)
	${}_{MmPyl}RS$ / ${}_{MbPyl}tRNA_{UCA}$	ANAP (100 μM)
		AzPhe (1 mM)
		PrPhe (1 mM)
	${}_{OMeY}RS$ / ${}_{OMeY}tRNA_{UCA}$	ANAP (100 μM)
		N_3Lys (1 mM)
		PrK (1 mM)
UTX1.38 (GLN204TAG)	${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{CUA}$	AzPhe (1 mM)
		N_3Lys (1 mM)
		PrK (1 mM)
		PrPhe (1 mM)
	${}_{MmPyl}RS$ / ${}_{MbPyl}tRNA_{CUA}$	ANAP (100 μM)
		AzPhe (1 mM)
		PrPhe (1 mM)
	${}_{OMeY}RS$ / ${}_{OMeY}tRNA_{CUA}$	ANAP (100 μM)
		N_3Lys (1 mM)
		PrK (1 mM)
UTX1.39 (GLN204TGA)	${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{UCA}$	AzPhe (1 mM)
		N_3Lys (1 mM)
		PrK (1 mM)
		PrPhe (1 mM)
	${}_{MmPyl}RS$ / ${}_{MbPyl}tRNA_{UCA}$	ANAP (100 μM)
		AzPhe (1 mM)
		PrPhe (1 mM)
	${}_{OMeY}RS$ / ${}_{OMeY}tRNA_{UCA}$	ANAP (100 μM)
		N_3Lys (1 mM)
		PrK (1 mM)

2.11.7. UAA Incorporation at Two-Sites in eGFP

To determine the efficiency of UAA incorporation at two different sites in the P.O.I. (eGFP) by using two mutually orthogonal G.C.E. systems, 120x conditions were screened in the HEK cells seeded in 384-well cell culture plates by employing the workflow automation for cell-based assays. For each of the 10x G.E.M.S. Plasmid (as listed in Table 12) two-site UAA incorporation, using the 12x pairs of orthogonal G.C.E. systems (as listed in Table 13), was studied.

After the transient transfection of cells, green- and red-channel images were collected as mentioned in Section 2.10.4. The images were cropped and analyzed on the high-performance computing cluster using the CellProfiler image analysis pipelines *384_Well_Image_Cropping.cppipe, Blank_Image_Average.cppipe* and *Background_Subtraction_and_Analysis.cppipe*, as mentioned in Section 2.10.6.

The CellProfiler output was analyzed in R (on the high-performance computing cluster) by using *CellProfiler_Output_Analysis.R* script as mentioned in Section 2.10.7. The graphs were plotted in R by using *Final_Graph.R* script as mentioned in Section 2.10.7.

Table 12: 10x Amino Acid Positions-Pairs on eGFP for Two-Site UAA Incorporation Screening

Original Amino Acid Positions	**G.E.M.S. Plasmid**
TYR39 & LYS52	UTX2.23
TYR39 & PHE99	UTX2.24
TYR39 & VAL176	UTX2.25
TYR39 & GLN204	UTX2.26
LYS52 & PHE99	UTX2.34
LYS52 & VAL176	UTX2.35
LYS52 & GLN204	UTX2.36
PHE99 & VAL176	UTX2.45
PHE99 & GLN204	UTX2.46
VAL176 & GLN204	UTX2.56

Table 13: 12x Pairs of Orthogonal G.C.E. Systems for Two-Site UAA Incorporation Screening

Pair of Orthogonal G.C.E. Machinery	**UAA Pair**
$^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{CUA}$ & $_{OMeY}RS$ / $_{OMeY}tRNA_{UCA}$	ANAP (100 μM) & AzPhe (1 mM)
	ANAP (100 μM) & PrPhe (1 mM)
$^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{CUA}$ & $_{MmPyl}RS$ / $_{MbPyl}tRNA_{UCA}$	ANAP (100 μM) & N_3Lys (1 mM)
	ANAP (100 μM) & PrK (1 mM)
$_{MmPyl}RS$ / $_{MbPyl}tRNA_{CUA}$ & $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{UCA}$	N_3Lys (1 mM) & ANAP (100 μM)
	PrK (1 mM) & ANAP (100 μM)
$_{MmPyl}RS$ / $_{MbPyl}tRNA_{CUA}$ & $_{OMeY}RS$ / $_{OMeY}tRNA_{UCA}$	N_3Lys (1 mM) & PrPhe (1 mM)
	PrK (1 mM) & AzPhe (1 mM)
$_{OMeY}RS$ / $_{OMeY}tRNA_{CUA}$ & $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{UCA}$	AzPhe (1 mM) & ANAP (100 μM)
	PrPhe (1 mM) & ANAP (100 μM)
$_{OMeY}RS$ / $_{OMeY}tRNA_{CUA}$ & $_{MmPyl}RS$ / $_{MbPyl}tRNA_{UCA}$	AzPhe (1 mM) & PrK (1 mM)
	PrPhe (1 mM) & N_3Lys (1 mM)

3. RESULTS

For genetic code expansion, aaRS/tRNA pairs, which must be orthogonal to the host, are synthetically evolved for specific unnatural amino acids. Consequently, a significant amount of time has been devoted for evolving such pairs. While this process is at the heart of genetic code expansion, it is also a tedious and time-consuming process. For the end user who is not directly involved with this process, a successful integration of these genetic code expansion systems is important to make productive use of them.

In this work, we have tried to address three relevant aspects for a successful integration of existing mammalian genetic code expansion (G.C.E.) systems, namely, choosing the best conditions for unnatural amino acid incorporation in proteins, choosing the best G.C.E. system for incorporation of required functional group and choosing the best position on the target-protein for unnatural amino acid incorporation. In this study, we have provided a platform to assimilate the existing G.C.E. systems evolved for mammalian expression systems and to compare them for site-specific incorporation at various amino acid positions. In addition, we have also demonstrated a strategy for simultaneous incorporation of two different unnatural amino acids into proteins expressed in mammalian cells. As automation can significantly improve the throughput of the screening processes, we have also demonstrated the compatibility of our screening approach with automation and high-performance computing.

3.1. PLASMIDS AND PROTEIN EXPRESSION VECTORS

The orthogonal aaRS/tRNA pairs used by us have been developed by different research groups, and as a result they were in expression vectors with different plasmid backbones. To reduce the heterogeneity in the transfection conditions, we decided to have a common plasmid backbone. Therefore, we decided to develop our own mammalian expression vectors for delivering the genes for the protein of interest (P.O.I.) and the Orthogonal Translation System (OTS, or orthogonal aaRS/tRNA pair) into the cells. To this end, by combining Gateway™ and MultiBac™ technologies of molecular cloning, we have developed a set of two plasmids that can deliver the required genes into mammalian cells via transient or baculoviral transfection. All plasmids created for this thesis can be obtained from Prof. Dr. Timm Maier.

3.1.1. Gibson Assembly Helper Plasmid GABB

Plasmid GABB was created primarily for molecular cloning and optimization of cell-based assays (Figure 14 A). Upon digestion with KpnI and AvrII restriction enzymes, *Plasmid GABB* yields a 1920 bp linear DNA fragment containing ampicillin resistance gene *AmpR* and pBR322 origin of replication (*ori*). This linear DNA contains only the necessary elements for plasmid amplification and selection in bacterial cells and could be used as a backbone for Gibson Assembly of large P.O.I. genes. *Plasmid GABB* also has an eGFP expression cassette (*i.e.*, CMV Promoter, eGFP gene and β-globin poly(A) terminator) which constitutively expresses eGFP protein in mammalian cells. By utilizing this property, the plasmid was used for optimization of transient-transfection conditions for protein expression in mammalian cells.

3.1.2. Gateway™ Donor Vector GD54

Plasmid GD54 is a Gateway™ donor vector (Figure 14 B). This plasmid is commercially available as part of the "MultiSite Gateway™ Pro Plus" kit by Invitrogen [240]. However, since this kit is quite expensive and since we could not obtain this plasmid from other sources, we decided to create it ourselves. The architecture of this plasmid is similar to that of the Gateway™ donor *Plasmid pDONR221*, but with an exception that *Plasmid GD54* contains attP4 & attP5 recombination sites. The region flanked by these two sites is known as the Gateway™ cassette. By default, the Gateway™ cassette of *Plasmid GD54* contains a suicide gene *ccdB* and a chloramphenicol resistance gene *CmR*. Because of the suicide gene, *Plasmid GD54* must be amplified in ccdB-resistant strains of *E. coli* (*e.g.*, One Shot™ ccdB Survival™ 2 T1^{R} *E. coli* cells).

For Gateway™ recombination cloning using the *Plasmid GD54*, first, via PCR, attB4 and attB5 recombination sites are added on either ends of the target DNA. This PCR-amplified DNA is known as 'attB54 insert'. Alternatively, synthetic genes, which already have the gene of interest flanked by the attB4 and attB5 recombination sites, can be procured to be use with this plasmid. For example, all our otRNA expression cassettes were procured this way.

During the BP-Clonase™ mediated Gateway™ recombination cloning, the attB54 insert is transferred into the donor vector's (*i.e.*, *Plasmid GD54*) Gateway™ cassette, while simultaneously knocking out its 'default' *CmR-ccdB* bacterial expression cassette, thereby creating the 'GD54 entry clone'. Upon successful insertion of the expression cassette between the Gateway™ recombination sites, the ccdB protein is no longer expressed, hence, the positive constructs are amplified in usual cloning strains of *E. coli* (such as DH5α, BL21 or NEB10β) using kanamycin antibiotic selection.

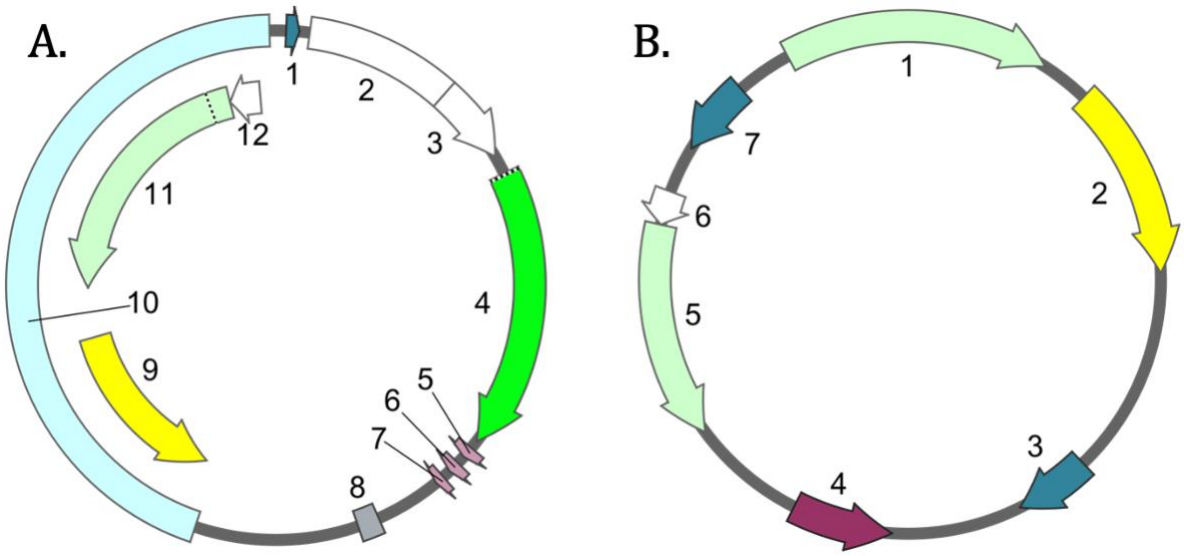


Figure 14: Architecture of Plasmids GABB and GD54.
A: Gibson Assembly helper vector *Plasmid GABB*. Plasmid features: LoxP recombination site (1), CMV-IE enhancer (2), CMV-IE promoter (3), eGFP gene (4), 10x-His purification tag (5), Myc purification tag (6), FLAG purification tag (7), β-globin poly(A) terminator (8), pBR322 origin of replication (9), AmpR-ori cassette (10), Ampicillin resistance gene AmpR (11) and AmpR promoter (12).
B: Gateway™ donor vector *Plasmid GD54*. Plasmid features: Kanamycin resistance gene KanR (1), pBR322 bacterial origin of replication (2), attP5 recombination site for Gateway™ cloning (3), ccdB suicide gene (4), Chloramphenicol resistance gene CmR (5), cat promoter (6) and attP4 recombination site for Gateway™ cloning (7).

3.1.3. Plasmid MX01 for Protein Expression in Mammalian Cells

Plasmid MX01 is a destination vector for Gateway™ cloning, as well as an acceptor vector for MultiBac™ cloning (Figure 16 A). The expression vector derived from this plasmid can be used to express the protein of interest in mammalian cells. The region of this plasmid flanked by the recombination sites attR1 and attR2 (also known as the Gateway™ cassette) can be used to insert the target-protein gene easily and rapidly via Gateway™ recombination cloning. Commercially available donor vectors such as *Plasmid pDONR221* or *Plasmid pDONR221Zeo*, which contain the attP1-attP2 Gateway™ recombination sites, are compatible with *Plasmid MX01* to insert the target-protein into its Gateway™ cassette via BP-Clonase and LR-Clonase mediated Gateway™ recombination.

By default, the Gateway™ cassette of *Plasmid MX01* contains a chloramphenicol resistance gene *CmR* and a suicide gene *ccdB*. Because of the suicide gene, *Plasmid MX01* must be amplified in ccdB-resistant strains of *E. coli*, using chloramphenicol and gentamycin antibiotic selection. During the LR-Clonase™ mediated Gateway™ recombination cloning, the protein of interest (P.O.I.) gene is transferred from an entry clone (derived from *Plasmid pDONR221*) into the destination vector's (*i.e.*, *Plasmid MX01*) Gateway™ cassette, while simultaneously knocking out the 'default' *CmR-ccdB* bacterial expression cassette from *Plasmid MX01*. In this way, the 'expression vector' is created for the P.O.I. Upon successful insertion of the P.O.I. gene in the Gateway™ cassette, the ccdB protein is no longer expressed, hence the positive constructs are amplified in

regular cloning strains of *E. coli* using gentamycin antibiotic selection. For successful selection of the correct expression vectors, these strains should not be ccdB resistant.

The CMV-IE promoter lies upstream to the Gateway™ cassette towards the attR1 site. This promoter allows *Plasmid MX01* derived expression vectors, to constitutively express the P.O.I. gene in mammalian cells. Immediately downstream to the Gateway™ cassette (towards the attR2 site) are the P2A and T2A sites. The P2A-T2A site is followed by a gene for eGFP. For the expression vectors derived from *Plasmid MX01*, if a terminal stop codon is not present in the inserted P.O.I. gene, and if P2A-T2A sites are in-frame with respect to the P.O.I. gene, then for each molecule of P.O.I. produced in the mammalian cells, one molecule of eGFP is also produced. Moreover, while translating the P2A and the T2A sites, the mammalian ribosome creates a nick in the polypeptide chain, hence the P.O.I. and the eGFP are expressed as two separate polypeptide chains. In this way, the eGFP gene serves as a reporter for eGFP-fluorescence based online quantification of target-protein expression. However, due to the attB2 site and the P2A site, which are present downstream to the P.O.I. gene (Figure 16 A, B), the P.O.I. contains 26 extra C-terminal amino acids (PSFLVQSG from the attB2 site and ATNFSLLKQAGDVEENPG from the P2A site).

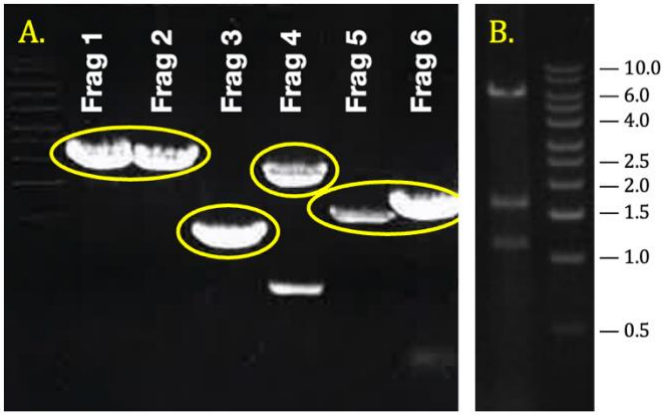


Figure 15: Development and Confirmation of Plasmid MX01.
A. The six fragments used for assembling *Plasmid MX01* were purified on agarose gel. The selected fragments are marked.
B. Before sequencing, the identity of *Plasmid MX01* was confirmed by digestion with SalI restriction enzyme. The three linear fragments (left lane, from top to bottom) are respectively 6041 bp, 1619 bp and 1095 bp in size. The right lane is 1 kbp DNA-ladder from Carl Roth.

For incorporating unnatural amino acids in proteins, the orthogonal translation system (OTS, *i.e.*, the orthogonal aaRS/tRNA pair) uses stop-codon suppression, which has to compete with translation-termination machinery of the cell. For this reason, often several expression conditions must be tested on a small scale, before the optimal conditions for the large-scale production of P.O.I.^{UAA} (*i.e.*, the protein of interest having the unnatural amino acid in its polypeptide chain) can be identified. Having an easily quantifiable marker such as eGFP-fluorescence is a useful screening tool for optimizing the expression conditions and expression duration and can be helpful in bypassing the tedious and time-consuming protein purification process for optimization of protein expression. Of course, the eGFP is marker only for P.O.I.^{UAA} expression, not for the proper folding and maturation of P.O.I.^{UAA}. For determining the quantity of properly folded and matured proteins, other protein characterization methods specific to the target-protein must be used.

Outside of the Gateway™ cassette, *Plasmid MX01* also contains pBR322 bacterial origin of replication, a LoxP recombination site, Tn7L and Tn7R recombination sites, a gentamycin resistance gene and a VSV-G glycoprotein expression cassette (*i.e.*, polH promoter, VSV-G gene, and SV40 poly (A) terminator) for insect cells.

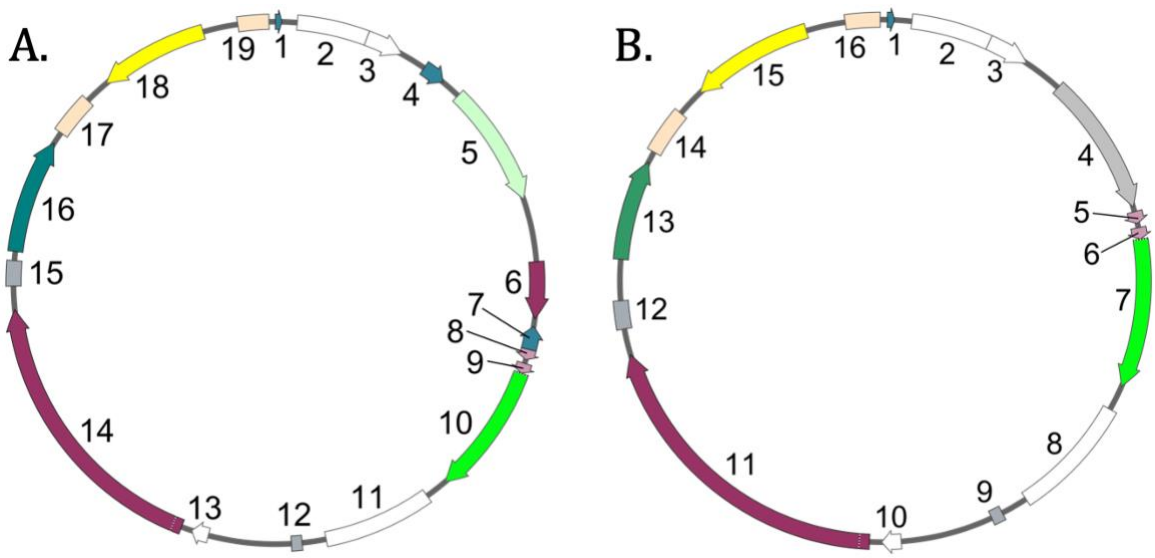


Figure 16: Architecture of Plasmid MX01 and Expression Vector Derived from it.
A: Gateway™ destination vector *Plasmid MX01* for mammalian expression. Plasmid features: LoxP recombination site (1), CMV-IE enhancer (2), CMV-IE promoter (3), attR1 recombination site of the P.O.I. Gateway™ cassette (4) Chloramphenicol resistance gene *CmR* (5), ccdB suicide gene *ccdB* (6), attR2 recombination site of the P.O.I. Gateway™ cassette (7), P2A site (8), T2A site (9), eGFP gene (10), WPRE enhancer (11), β-globin poly(A) terminator (12), Polyhedrin promoter (13), VSV-G glycoprotein gene (14), SV40 poly(A) terminator (15), Gentamycin resistance gene *GentR* (16) Tn7R transposon site (17), pBR322 origin of replication (18) and Tn7L transposon site (19).
B: Mammalian expression vector for the target-protein derived from *Plasmid MX01*. Plasmid features: LoxP recombination site (1), CMV-IE enhancer (2), CMV-IE promoter (3), Protein of interest gene (4), P2A site (5), T2A site (6), Gene for eGFP reporter protein (7), WPRE enhancer (8), β-globin poly(A) terminator (9), Polyhedrin promoter (10), VSV-G glycoprotein gene (11), SV40 poly(A) terminator (12), Gentamycin resistance gene *GentR* (13), Tn7R transposon site (14), pBR322 origin of replication (15) and Tn7L transposon site (16). attB1 and attB2 sites are not displayed.

The presence of pBR322 bacterial origin of replication (*ori*) on MX01-derived expression vectors (Figure 16 B) makes them high copy-number for bacterial cells. This can be helpful in large-scale plasmid purification for transient transfection. Due to the presence of a LoxP recombination site, the expression vector plasmids derived from *Plasmid MX01* can combine with other LoxP containing plasmids via the Cre-recombinase-mediated recombination reaction. Despite being created primarily for unnatural amino acid incorporation in proteins, *Plasmid MX01* is a standalone plasmid, and can be used for other applications requiring constitutive expression of proteins in mammalian cells [241]. For baculovirus mediated large scale protein production, the expression vectors derived from *Plasmid MX01* can be loaded on to the baculovirus genome via the Tn7L and Tn7R recombination sites.

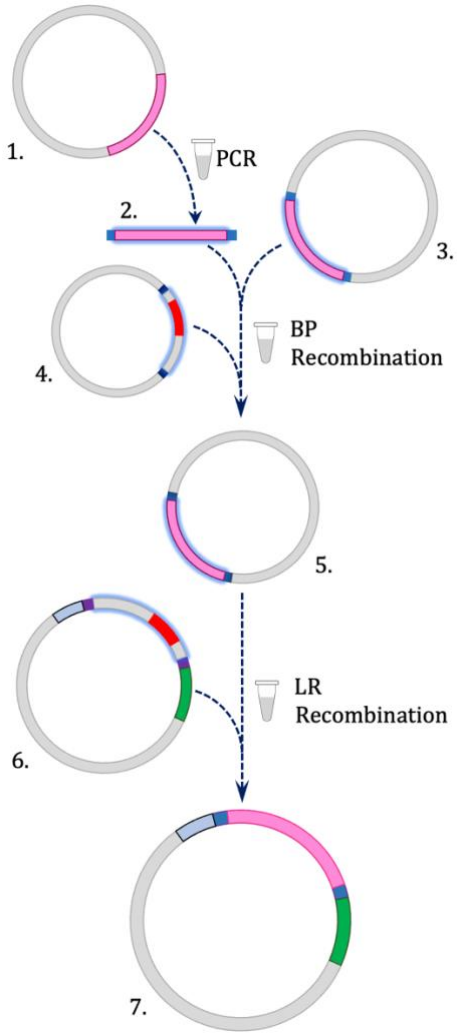


Figure 17: Creation of Mammalian Expression Vector from Plasmid MX01.
The gene for the protein of interest (P.O.I. gene) can be amplified via PCR by using a template DNA (1) to yield linear DNA with attB flanking regions (2). Alternatively, synthetic DNA with the gene already flanked with attB regions can be procured (3). BP-Clonase™ mediated Gateway™ recombination between the attB-insert and the Gateway™ donor vector (4) generates the entry clone (5). The exchanged regions are highlighted. The red region represents the suicide gene *ccdB*. LR-Clonase™ mediated Gateway™ recombination between the entry clone and the destination vector *Plasmid MX01* (6) generates the mammalian expression vector for the P.O.I. gene (7). If the same gene needs to be expressed in different expression systems, the same entry clone can be used with different destination vectors.

3.1.4. Plasmid UCAP for Delivery of Orthogonal Translation System in Mammalian Cells

Plasmid UCAP is a destination vector for Gateway™ cloning, as well as a donor vector for MultiBac™ cloning (Figure 19 A). This plasmid vector is used for delivering the oRS and otRNA genes of the OTS, into the mammalian protein expression systems. It contains two mutually orthogonal Gateway™ cassettes. The oRS Gateway™ cassette is flanked by attR1 & attR2 recombination sites and is used for inserting the oRS gene into the plasmid. The otRNA Gateway™ cassette is flanked by attR4 & attR5 recombination sites and is used for inserting the otRNA expression cassette. Outside of the two orthogonal Gateway™ cassettes, *Plasmid UCAP* also contains an ampicillin resistance gene *AmpR* with its bacterial promoter, the FRT & LoxP recombination sites, the R6K-γ bacterial origin of replication and a puromycin resistance expression cassette (SV40 promoter, puromycin resistance gene *PuroR* and SV40 poly(A) terminator) for mammalian cells.

The oRS Gateway™ cassette of the *Plasmid UCAP* contains a chloramphenicol resistance gene *CmR* and a suicide gene *ccdB*. CMV-IE mammalian promoter is present upstream of the oRS Gateway™ cassette (*i.e.*, on the attR1 side) and bGH poly(A) terminator signal is present downstream of the oRS Gateway™ cassette. For inserting the oRS gene in the plasmid, LR-Clonase mediated Gateway™ recombination is performed between this destination vector (*Plasmid UCAP*) and entry clones derived from *Plasmid pDONR221* or *Plasmid pDONR221Zeo*. While the oRS Gateway™ cassette needs its upstream as well as downstream elements (*i.e.*, the promoter and the terminator sequence) for protein expression, the otRNA Gateway™ cassette is used to deliver the entire expression cassette (*i.e.*, the promoter, the gene of interest and the terminator) for the otRNA expression. The otRNA Gateway™ cassette of the *Plasmid UCAP* contains a kanamycin resistance gene *KanR*, a pBR322 origin of replication (*ori*) and a suicide gene *ccdB*. These three elements together form the *KanR-ori-ccdB* cassette. For inserting the otRNA expression cassette in the plasmid, LR-Clonase mediated Gateway™ recombination is performed between *Plasmid UCAP* and entry clones derived from *Plasmid GD54*. Since the Gateway™ cassettes of *Plasmid UCAP* are mutually orthogonal, both oRS and otRNA genes can be inserted simultaneously (one-pot reaction) into the plasmid to form the mammalian expression vector.

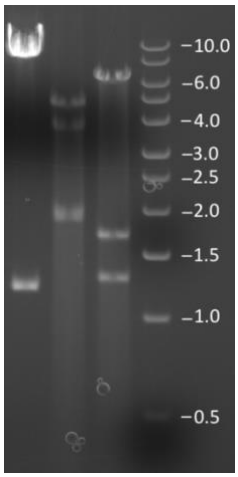


Figure 18: Restriction Digestion Confirmation for Plasmid UCAP.
Before sequencing, the identity of *Plasmid UCAP* was confirmed by restriction digestion. The first lane from the left contains *Plasmid UCAP* digested with BspEI restriction enzyme, which generates linear DNA fragments 8298 bp, 1210 bp and 64 bp in size (64 bp fragment is not visible). The second lane from the left contains *Plasmid UCAP* digested with KpnI restriction enzyme, which generates linear DNA fragments 4361 bp, 3419 bp and 1792 bp in size. The third lane from the left contains *Plasmid UCAP* digested with PvuII restriction enzyme, which generates linear DNA fragments 6602 bp, 1688 bp and 1282 bp in size. The right-most lane is 1 kbp DNA-ladder from Carl Roth.

Because of the two suicide genes, *Plasmid UCAP* must be amplified in ccdB-resistant strains of *E. coli*, using chloramphenicol, kanamycin, and ampicillin antibiotic selection. However, when the oRS gene and otRNA expression cassette are successfully inserted into the Gateway™ cassettes (Figure 19 C, D) both *ccdB* genes are knocked out and the resultant expression vector no longer requires ccdB-resistant *E. coli* strains for amplification. However, since *KanR-ori-ccdB* cassette is also knocked out of the expression vector plasmid, it no longer has the pBR322 bacterial origin of replication and relies solely on R6K-γ bacterial origin of replication. Only Pir^+ strains of *E. coli* can recognize R6K-γ bacterial origin of replication hence *Plasmid UCAP* derived expression vectors need Pir^+ bacterial strains (such as One Shot™ PIR1 *E. coli* cells) for amplification, while using ampicillin or carbenicillin (preferred) as selection antibiotic.

Due to the presence of a LoxP recombination site, the expression vector plasmids derived from *Plasmid UCAP* can also recombine with other LoxP containing plasmids. The presence of R6K-γ *ori* in these plasmids makes them low copy number plasmids. For this reason, large culture volumes and Pir^+ strain of *E. coli* is needed for large scale production of these plasmids. However, this origin of replication was chosen intentionally to make the plasmids compatible with MultiBac™ cloning (described later).

The FRT recombination site on the expression vectors derived from *Plasmid UCAP* can be used to integrate these plasmids into the genomic DNA of Flp-In™ cells lines (Invitrogen, ThermoFisher), thereby creating constitutive-expression cell lines for a given OTS genes. While creating the new cell lines, the puromycin resistance expression cassette present on these plasmids helps in selection of those cells where the plasmid of interest is successfully integrated into the genome.

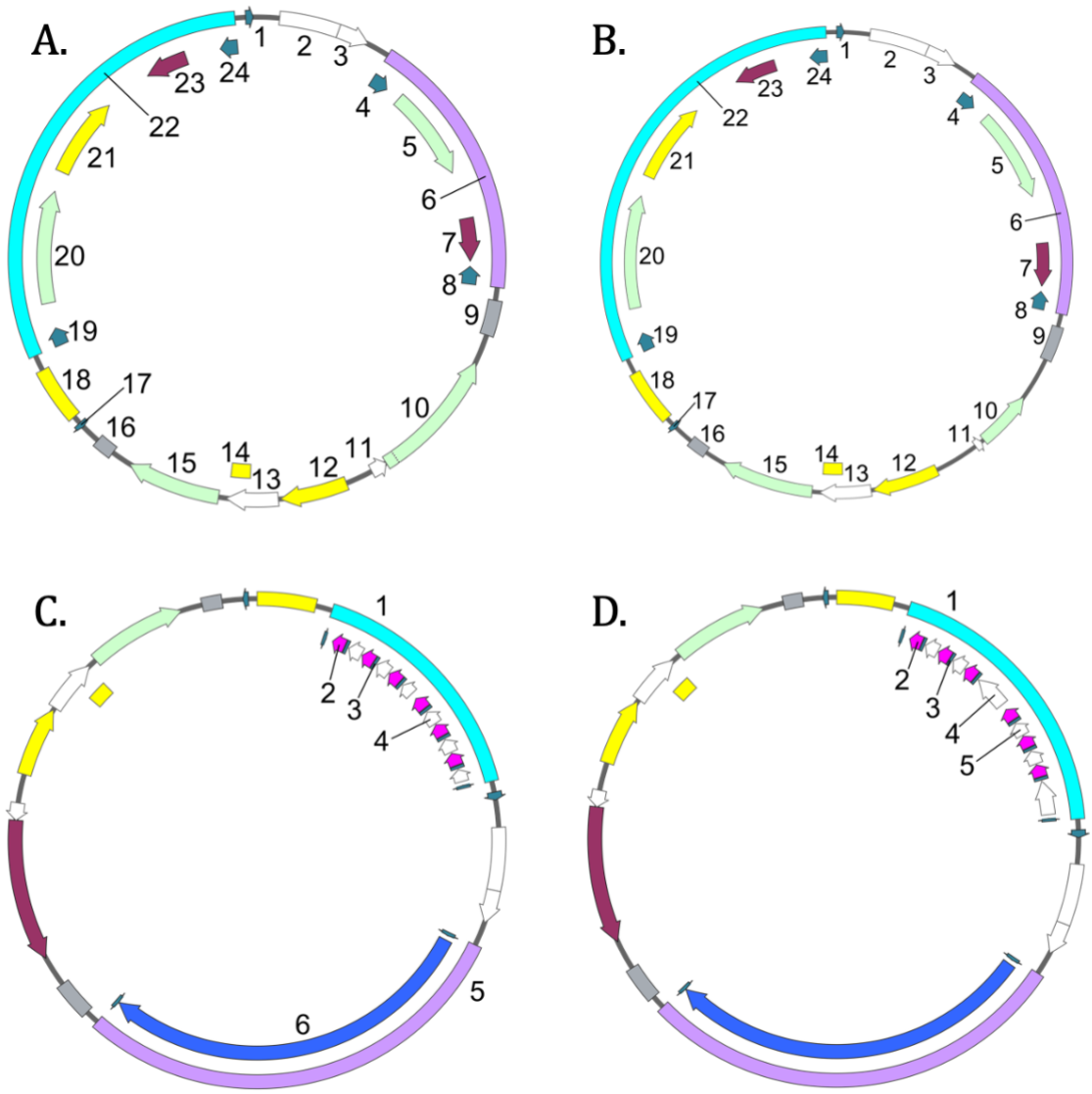


Figure 19: Plasmids UCAP & UCZP and Mammalian Expression Vectors Derived from Plasmid UCAP.
A. Gateway™ Destination vector *Plasmid UCAP* for mammalian expression. Plasmid features: FRT recombination site (1), CMV-IE enhancer (2), CMV-IE promoter (3), attR1 recombination site of the oRS Gateway™ cassette (4), Chloramphenicol resistance gene *CmR* (5), Gateway™ cassette for the orthogonal tRNA Synthetase gene (6), *ccdB* suicide gene (7), attR2 recombination site of the oRS Gateway™ cassette (8), bGH poly(A) terminator (9), Ampicillin resistance gene *AmpR* (10), AmpR promoter (11), f1 origin of replication (12), SV40 promoter (13), SV40 origin of replication (14), Puromycin resistance gene *PuroR* (15), SV40 poly(A) terminator (16), LoxP recombination site (17), R6K-γ origin of replication (18), attR4 recombination site of the otRNA Gateway™ cassette (19), Kanamycin resistance gene *KanR* (20), pBR322 origin of replication (21), Gateway™ cassette for the orthogonal tRNA expression cassette (22), *ccdB* suicide gene (23) and attR5 recombination site of the otRNA Gateway™ cassette (24).
B. Gateway™ destination Vector UCZP for mammalian expression. Plasmid features: FRT recombination site (1), CMV-IE enhancer (2), CMV-IE promoter (3), attR1 recombination site of the oRS Gateway™ cassette (4), Chloramphenicol resistance gene *CmR* (5), Gateway™ cassette for the orthogonal tRNA Synthetase gene (6), *ccdB* suicide gene (7), attR2 recombination site of the oRS Gateway™ cassette (8), bGH poly(A) terminator (9), Zeocin™ resistance gene *BleoR* (10), EM7 promoter (11), f1 origin of replication (12), SV40 promoter (13), SV40 origin of replication (14), Puromycin resistance gene *PuroR* (15), SV40 poly(A) terminator (16), LoxP recombination site (17), R6K-γ origin of replication (18), attR4 recombination site of the otRNA Gateway™ cassette (19), Kanamycin resistance gene *KanR* (20), pBR322 origin of replication (21), Gateway™ cassette for the orthogonal tRNA expression cassette (22), *ccdB* suicide gene (23) and attR5 recombination site of the otRNA Gateway™ cassette (24).
C. Mammalian expression vector for delivering orthogonal aaRS and orthogonal tRNA genes. Through the otRNA Gateway™ cassette (1), six copies of the otRNA gene (2) are inserted into the *Plasmid UCAP*. For allowing tetracycline-induced expression of otRNA, three Tetracycline-Operator sites (3) are present immediately before of each copy of the otRNA gene. For EcLeu- and EcTrp-otRNA, each copy of the otRNA gene is expressed by H1 promoter (4). Through the oRS Gateway™ cassette (5) of the plasmid, the oRS gene (6) is inserted in the plasmid.
D. Alternative architecture of the otRNA expression cassette. Through the otRNA Gateway™ cassette (1), six copies of the otRNA gene (2) are inserted into the *Plasmid UCAP*. However, for OMeY- and MbPyl-otRNA expression cassettes, two copies of the otRNA are expressed by U6 promoter (2) and the remaining four copies are expressed by H1 promoter (3).

3.1.5. Plasmid UCZP for Delivery of Orthogonal Translation System in Mammalian Cells

Plasmid UCZP is a destination vector for Gateway™ cloning, as well as a donor vector for MultiBac™ cloning (Figure 19 B). It is a variant of *Plasmid UCAP*, with Zeocin™ resistance gene and its bacterial promoter, instead of the bacterial expression cassette for ampicillin resistance. Accordingly, *Plasmid UCZP* must be amplified in ccdB-resistant strains of *E. coli*, using chloramphenicol, kanamycin & Zeocin™ antibiotic selection and expression vectors derived from *Plasmid UCZP* need Zeocin™ as the selection antibiotic (while still using Pir^+ bacterial strains for amplification). In addition, for inserting the oRS gene in the plasmid, LR-Clonase mediated Gateway™ recombination is performed between this destination vector (*Plasmid UCZP*) and entry clones derived from *Plasmid pDONR221* (due to the Zeocin™ resistance marker *Plasmid pDONR221Zeo* is not compatible for Gateway™ cloning with *Plasmid UCZP*).

The remaining backbone and properties of the two plasmids are same. If the selection on ampicillin or carbenicillin is not efficient *Plasmid UCZP* can be used instead of *Plasmid UCAP*, for OTS gene delivery into mammalian cells (Zeocin™ is a more potent antibiotic, but it is toxic to humans as well). Additionally, by using the Cre-LoxP recombination, the expression vectors derived from *Plasmid UCZP* can be used alongside the expression vectors derived from *Plasmid UCAP*, to deliver mutually orthogonal OTS genes (*i.e.*, two mutually orthogonal oRS/otRNA pairs) in mammalian cells using the baculoviral gene delivery system.

3.1.6. Architecture of otRNA Expression Cassette

For each of our G.C.E. Plasmid, while only one copy of the oRS gene is present in the oRS Gateway™ cassette, the otRNA expression cassette used in our screening experiments contains six copies of the otRNA gene. All these six copies have their own promoter and terminator sequences, that is why otRNA Gateway™ of *Plasmid UCAP* is not flanked by promoter or terminator sequences. The six copies of the otRNA gene are present in two different architectures. For *Plasmids UCAP.01 to UCAP.12*, each of these six copies is promoted by the H1 RNA promoter (Figure 19 C). For *Plasmids UCAP.13 to UCAP.18* and *Plasmids UCAP.19 to UCAP.22*, the otRNA expression cassette contained two copies of a sub-cassette with three otRNA genes (Figure 19 D). One of the otRNA gene of the sub-cassette is promoted by U6 RNA polymerase III promoter and the remaining two were promoted by the H1 RNA promoter.

When UAA is incorporated using G.C.E., one of the stop-codons is 'repurposed' as a signal for UAA incorporation. Although, the efficiency of UAA incorporation is low, there may be some effects on the cellular functions when termination of essential proteins fails due to UAA incorporation. To delay this, each of the six copies of the otRNA gene also contains three copies of tetracycline-operator sites (TO). If the Tetracycline Repressor Element (TRE) is constitutively expressed in the cells (*e.g.*, in HEK-TREx cells from Sigma), the TRE protein binds to the TO site on the plasmid and prevents the expression of otRNA. However, when the cell culture media is supplemented with the tetracycline antibiotic, TRE competitively binds with the tetracycline, frees the TO site and otRNA can be expressed, thereby initiating UAA incorporation in proteins.

For each kind of oRS, otRNA for Ochre (TAA), Amber (TAG) and Opal (TGA) suppression were screened. Because of the six copies of the otRNA in each of the otRNA expression cassette, it was not possible to perform sequence-based molecular biology operations (such as restriction cloning or PCR) within the cassette. Therefore, otRNA cassettes flanked by attB4 & attB5 recombination sites were always ordered as synthetic genes.

3.1.7. Mammalian Expression Vector UTX0

Plasmid UTX0 (Figure 20 A) has an architecture similar to *Plasmid MX01*, with two key differences: a) the plasmid contains the gene for mRaspberry fluorescent protein in its Gateway™ cassette, and b) the eGFP gene, present downstream to the Gateway™ cassette contains an HMF (10x-His-Myc-FLAG) peptide sequence at its C-terminal to aid the purification of the eGFP, whenever necessary. Figure 22 shows the arrangement of the genes expressed by *Plasmid UTX0*. When *Plasmid UTX0* is transiently transfected in mammalian cells, WT-mRaspberry and WT-eGFP are constitutively expressed in 1:1 molar ratio. Due to the presence of the P2A-T2A sites, these two proteins are expressed as separate polypeptides.

3.1.8. GIDx Plasmids for Baculoviral Production of Proteins in Insect Cells

Plasmids GIDC, GIDK and GIDS (Figure 20 B, C and D respectively) are destination vectors for Gateway™ cloning, as well as a donor vectors for MultiBac™ cloning. They were created for multiprotein expression in insect cells. All three of them have a Gateway™ cassette flanked by attR1 and attR2 recombination sites. By default, the Gateway™ cassette of these three plasmids contains an ampicillin resistance gene *AmpR*, a pBR322 origin of replication (*ori*) and a suicide gene *ccdB*. Together, these three elements form the *AmpR-ori-ccdB* cassette. Because of the suicide gene, these plasmids can only be amplified in ccdB-resistant strains of *E. coli*. *Plasmid GIDC* needs chloramphenicol & carbenicillin antibiotics for selection, *Plasmid GIDK* needs kanamycin & carbenicillin and *Plasmid GIDS* needs spectinomycin & carbenicillin for selection.

Plasmid GIDC has polH promoter upstream and SV40 poly(A) terminator downstream, with respect to its Gateway™ cassette. *Plasmids GIDK and GIDS* have p10 promoter upstream and HSV TK poly(A) terminator downstream, with respect to their Gateway™ cassette. For inserting the P.O.I. gene in *Plasmids GIDC and GIDS*, LR-Clonase™ mediated Gateway™ recombination can be performed between these destination vectors and entry clones derived from *Plasmid pDONR221*. However, since *Plasmid pDONR221* and *Plasmid GIDK* both have kanamycin resistance, they are incompatible with each other for Gateway™ recombination cloning. However, *Plasmid pDONOR221Zeo* (ThermoFisher), which is a variant of *Plasmid pDONR221* with Zeocin™ resistance gene, can be used to create entry clones that are compatible with all three of the GIDx plasmids (used as substitute of GIDC, GIDK and GIDS plasmids).

When the gene of interest is successfully inserted in the Gateway™ cassette of GIDx plasmids the *ccdB* gene is knocked. But simultaneously, as *Amp-ori-ccdB* cassette is lost, the expression vectors derived from GIDx plasmids do not have the pBR322 bacterial origin of replication and must rely solely on R6K-γ bacterial origin of replication for amplification in bacterial cells. As mentioned before, since only Pir^+ strains of *E. coli* can recognize R6K-γ bacterial origin of replication, *Plasmid GIDx* derived expression vectors also need Pir^+ bacterial strains for amplification. *Plasmid GIDC* derived expression vectors are selected on chloramphenicol antibiotic, *Plasmid GIDK* derived expression vectors on kanamycin antibiotic and *Plasmid GIDS* derived expression vectors are selected on spectinomycin antibiotic.

The expression vectors derived from plasmids GIDx contain a LoxP recombination site as well. Because of the LoxP site and R6K-γ bacterial origin of replication, plasmids GIDx serve as MultiBac™ donor vectors. They can undergo Cre-LoxP recombination with MultiBac™ acceptor vectors (such as *Plasmid pAB2G-N-HMF*), to form multi-protein expression constructs.

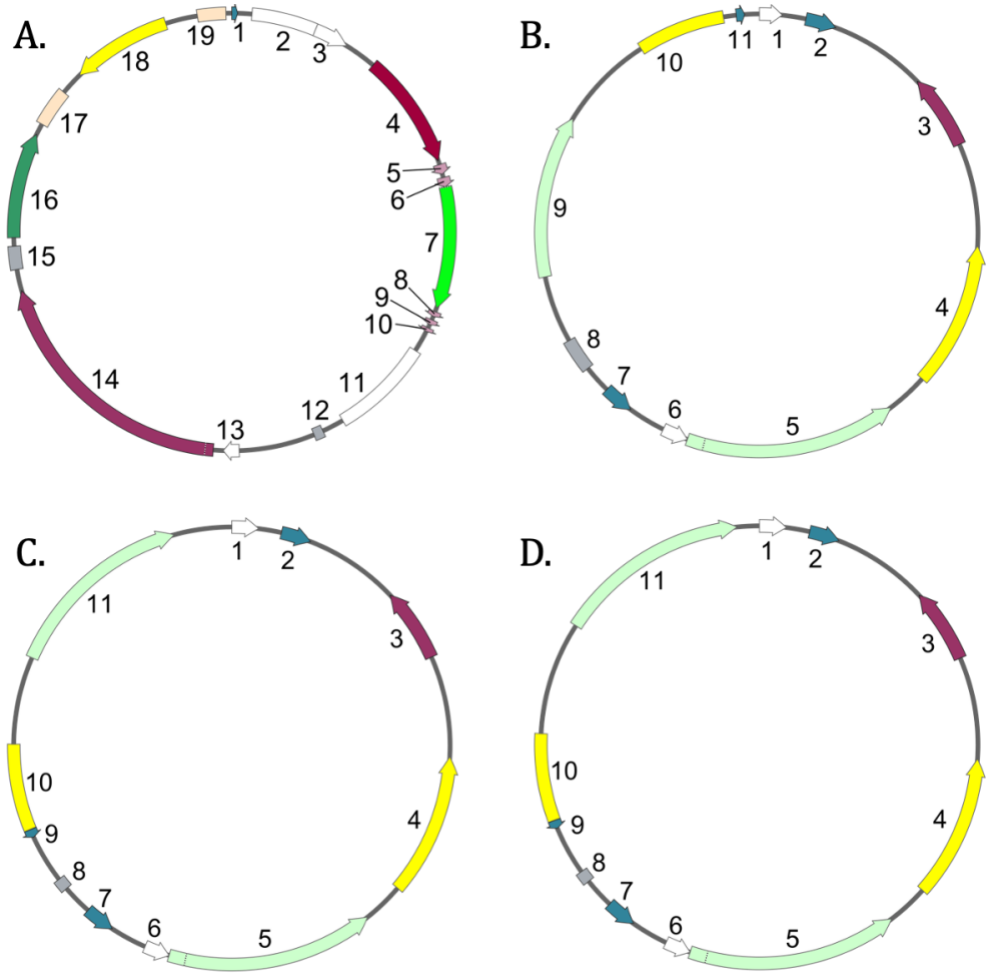


Figure 20: Architecture of Plasmids UTX0, GIDC, GIDK and GIDS.

A. Mammalian expression vector UTX0. Plasmid features: LoxP recombination site (1), CMV-IE enhancer (2), CMV-IE promoter (3), mRaspberry protein gene (4), P2A site (5), T2A site (6), eGFP gene (7), 10x-His purification tag (8), Myc purification tag (9), FLAG purification tag (10), WPRE enhancer (11), β-globin poly(A) terminator (12), Polyhedrin promoter (13), VSV-G glycoprotein gene (14), SV40 poly(A) terminator (15), Gentamycin resistance gene *GentR* (16), Tn7R transposon site (17), pBR322 origin of replication (18) and Tn7L transposon site (19). attB1 and attB2 sites are not displayed.

B. Gateway™ destination vector *Plasmid GIDC* for insect cell expression. Plasmid features: Polyhedrin insect expression promoter (1), attR1 recombination site of the Gateway™ cassette (2), *ccdB* suicide gene (3), pBR322 origin of replication (4), Ampicillin resistance gene *AmpR* (5), AmpR promoter (6), attR2 recombination site of the Gateway™ cassette (7), SV40 poly(A) terminator (8), Chloramphenicol resistance gene *CmR* (9), R6K-γ bacterial origin of replication (10) and LoxP recombination site (11).

C. Gateway™ destination vector *Plasmid GIDK* for insect cell expression. Plasmid features: Polyhedrin insect expression promoter (1), attR1 recombination site of the Gateway™ cassette (2), *ccdB* suicide gene (3), pBR322 origin of replication (4), Ampicillin resistance gene *AmpR* (5), AmpR promoter (6), attR2 recombination site of the Gateway™ cassette (7), SV40 poly(A) terminator (8), LoxP recombination site (9), R6K-γ bacterial origin of replication (10) and Kanamycin resistance gene *KanR* (11).

D. Gateway™ destination vector *Plasmid GIDS* for insect cell expression. Plasmid features: Polyhedrin insect expression promoter (1), attR1 recombination site of the Gateway™ cassette (2), *ccdB* suicide gene (3), pBR322 origin of replication (4), Ampicillin resistance gene *AmpR* (5), AmpR promoter (6), attR2 recombination site of the Gateway™ cassette (7), SV40 poly(A) terminator (8), LoxP recombination site (9), R6K-γ bacterial origin of replication (10) and Spectinomycin resistance gene *SpecR* (11).

3.2. MULTIBAC™ COMPATIBILITY OF PLASMIDS MX01, UCAP AND UCZP

The MultiBac™ baculovirus based BacMam™ technology can be used for baculovirus mediated gene delivery into mammalian cells [242-244]. The Tn7 recombination site present on MultiBac™ acceptor plasmids can be used to integrate these plasmids into the baculovirus genome.

Since *Plasmid MX01* contains the Tn7 recombination site, the expression vectors derived from *Plasmid MX01* can be used for baculovirus-mediated large-scale protein production using the BacMam™ technology. While amplifying the baculovirus in the insect cells, the expression of VSV-G glycoprotein (due to the polH promoter) and its incorporation in the assembled baculovirus increases the mammalian transduction efficiency of the assembled baculoviruses [245]. Since there is no mammalian origin of replication sites on this plasmid or in the baculovirus genome, the baculoviral transduction is transient in nature, *i.e.*, the plasmid DNA is lost after several generations. Hence, the baculoviral transduction for large scale protein production by using the expression vectors derived from *Plasmid MX01* can be safely performed in BSL1 biosafety environment (provided that a toxic protein is not loaded in the Gateway™ cassette of the plasmid).

For co-expression of multiple proteins MultiBac™ technology also involves joining MultiBac™ acceptor and donor plasmids via Cre-LoxP recombination. Due to the presence of the LoxP site and the pBR322 *ori*, the expression vectors derived from *Plasmid MX01* (*i.e.*, P.O.I. Plasmids) work as MultiBac™ acceptor plasmids. Since *Plasmids UCAP and UCZP* contain the LoxP site & the R6K-γ *ori*, and because pBR322 *ori* is knocked out of the plasmid after Gateway™ recombination, the expression vectors derived from *Plasmids UCAP or UCZP* (*i.e.*, G.C.E. Plasmids) work as MultiBac™ donor plasmids, and can be combined with *Plasmid MX01*. Therefore P.O.I. Plasmids and G.C.E. Plasmids can be combined together via Cre-LoxP recombination.

When a MultiBac™ donor plasmid (*e.g.*, G.C.E. Plasmids) fuses with a MultiBac™ acceptor plasmid (*e.g.*, P.O.I. Plasmids), the fused-plasmid contains the resistance gene from the MultiBac™ donor, the resistance gene from the MultiBac acceptor and the pBR322 *ori* from the MultiBac acceptor. Since only the successfully fused plasmid contains all the three components needed for the propagation of the plasmid, only this plasmid is amplified in usual cloning strains of *E. coli* when both the selection-antibiotics are present in the growth medium (Figure 21). This fusion of the two expression vectors can be very helpful for baculovirus-mediated protein co-expression, where the MultiBac™ acceptor plasmid gives a 'piggyback-ride' to the MultiBac™ donor plasmid for insertion into the MultiBac™ baculovirus genome. Using the same principle, a MultiBac™ acceptor plasmid can be combined with more than one MultiBac™ donor plasmids if the MultiBac™ donor plasmids use different antibiotics for selection (*e.g.*, *Plasmid UCAP* and *Plasmid UCZP*).

To our knowledge, *Plasmid UCAP* is the first plasmid vector that can work both as a Gateway™ destination vector and a MultiBac™ donor vector. Previously, for insect cell expression, we tried to combine these two cloning technologies using a PCR-based approach, but the ease of handling could not be achieved [unpublished data]. In general, the Gateway™ destination vectors require a ccdB resistant strain of *E. coli* for amplification while the MultiBac™ donor vectors require a Pir^+ strain. This means that these two types of plasmid vectors are dependent on two different strains of *E. coli* for their amplification. For *Plasmid UCAP*, this incompatibility was circumvented by inserting the pBR322 *ori* into the otRNA Gateway™ cassette. In this way, because of the pBR322 *ori*, *Plasmid UCAP*, which is a Gateway™ destination vector, could be amplified in a ccdB resistant strain of *E. coli* and since the pBR322 is knocked out during the formation of expression vector, the expression vector derived from *Plasmid UCAP* behaves as a MultiBac™ donor vector due to the R6K-γ *ori* and can be amplified only in Pir^+ cells. The newly developed *Plasmids UCZP, GIDC, GIDK and GIDS* use the same strategy to work as Gateway™ destination vectors and MultiBac™ donor vectors for insect expression systems.

Baculovirus preparation is a time consuming and space intensive process. Having one baculovirus instead of two (or more) for protein co-expression can save a lot of virus-preparation time. Additionally, fusion of the plasmids can guarantee the co-delivery of all relevant genes, thereby ensuring homogeneity of protein production. Some studies demonstrate that the number of copies of the suppressor-tRNA can play a crucial role in improving the yield of the UAA labelled target-protein [146]. The number of copies can be easily controlled using the otRNA Gateway™ cassette of the *Plasmid UCAP* or *Plasmid UCZP*.

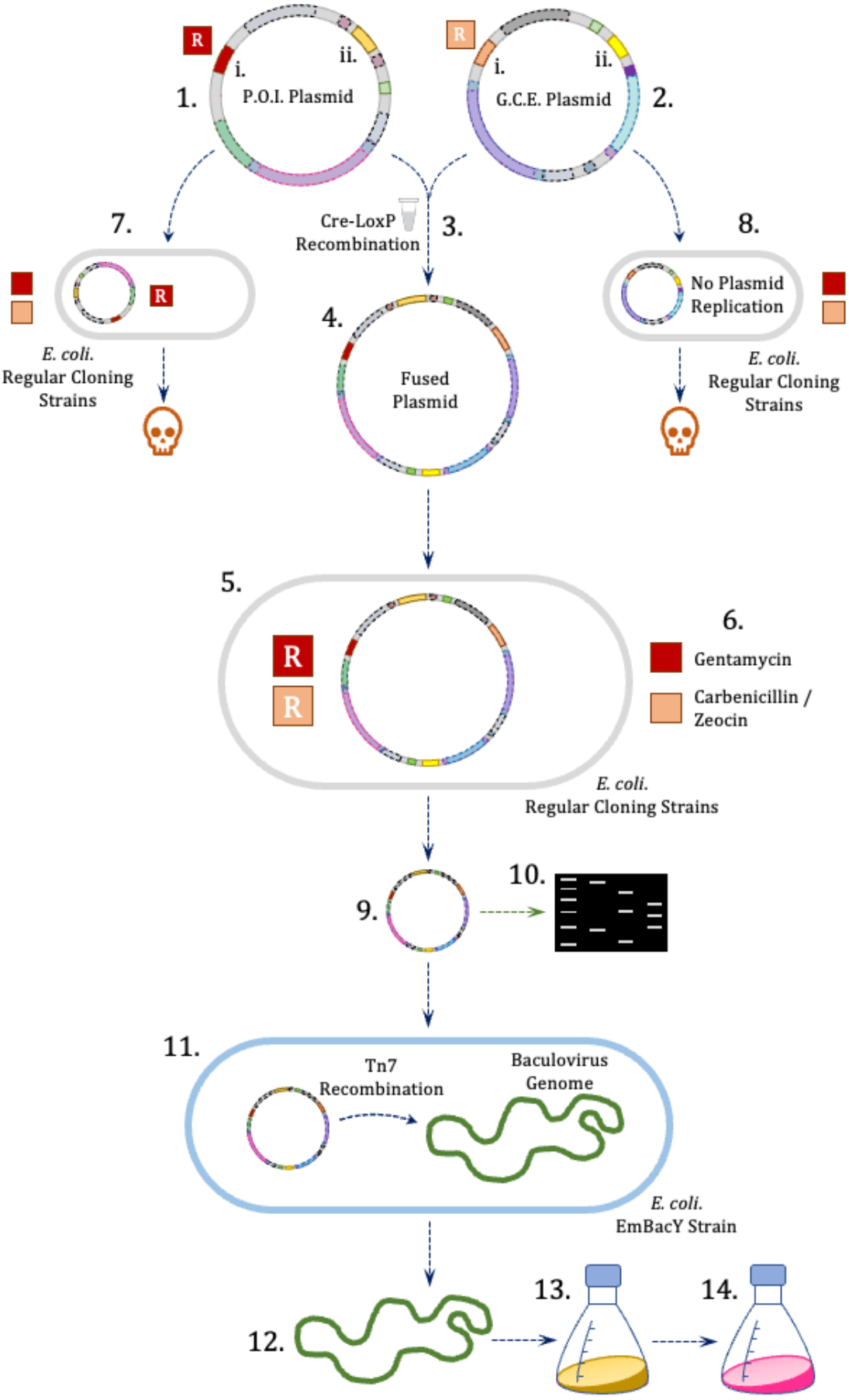

1.
R
i.
ii.
P.O.I. Plasmid
2.
R
i.
ii.
G.C.E. Plasmid
Cre-LoxP
Recombination
3.
4.
Fused
Plasmid
5.
R
R
6.
Gentamycin
Carbenicillin /
Zeocin
E. coli.
Regular Cloning Strains
7.
R
E. coli.
Regular Cloning
Strains
8.
No Plasmid
Replication
E. coli.
Regular Cloning
Strains
9.
10.
11.
Tn7
Recombination
Baculovirus
Genome
E. coli.
EmBacY Strain
12.
13.
14.

Figure 21: Production of Baculovirus for Genetic Code Expansion in Mammalian Cells.
The P.O.I. Plasmid (used for delivering the target-protein gene) (1) contains gentamycin resistance gene (1.i) and pBR322 origin of replication (1.ii). The G.C.E. Plasmid (used for delivering the genes for oRS/otRNA pair) (2) contains ampicillin or Zeocin™ resistance gene (2.i) and R6kγ origin of replication (2.ii). Due to the R6kγ origin of replication, the G.C.E. Plasmid cannot replicate in normal cloning strains of *E. coli*. Due to the presence of LoxP recombination sites, P.O.I. Plasmid and G.C.E. Plasmid can be joined via Cre-recombinase mediated recombination reaction (3, 4). For the selection of the fused-plasmid, the DNA is transferred into the regular cloning strains of *E. coli* (5) and is grown in presence of the two resistances (6). All the cells having only P.O.I. Plasmid do not survive due to the lack of ampicillin or Zeocin™ resistance gene (7). All the cells having only G.C.E. Plasmid do not survive because of the lack of gentamycin resistance gene (8). All cells that have separate copies of these two plasmids do not survive because G.C.E. Plasmid does not amplify in the regular cloning strains of *E. coli*. Only those cells that have the fused- plasmid can survive due to the amplification of the plasmid by pBR322 origin of replication and, as a result, expression of both antibiotic resistances (5). The fused-plasmid is purified (9) and recombination is confirmed via restriction digestion (10). The fused-plasmid is then delivered into EmBacY cells (*E. coli*) that contains the baculovirus genome (11). Due to the Tn7 recombination sites present on the P.O.I. Plasmid, the fused-plasmid gets integrated into the baculovirus genome. Those cells where the fused-plasmid has integrated into the baculovirus genome produce white colonies when grown on X-Gal (5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside). These colonies are amplified and the DNA is purified (12). The purified DNA is used for producing baculovirus in the insect cells (13). The baculovirus can then be used for the transduction of mammalian cells (14).

3.3. SCREENING ASSAY FOR G.C.E. EFFICIENCY QUANTIFICATION

For incorporating unnatural amino acids in the target-protein via genetic code expansion (G.C.E.), four essential components must be delivered into the cells, namely, the orthogonal aaRS (oRS), the orthogonal tRNA (otRNA), the unnatural amino acid (UAA) and the blank-codon ($codon_{BL}$) on the mRNA of the target-protein. For mammalian systems, which can use either stop-codon or quadruplet-codon as the $codon_{BL}$, the suppression of the $codon_{BL}$ by the otRNA may not work with 100% efficiency because the G.C.E. machinery competes with the cellular machinery involved in polypeptide biosynthesis and termination. In case of inefficient $codon_{BL}$-suppression, either the polypeptide is prematurely terminated or due to the frameshift mutation, a non-intended polypeptide is synthesized.

The efficiency of $codon_{BL}$-suppression may be a result of a complex interplay between several factors such as the cellular concentration of unnatural amino acids, the rate of aminoacylation of otRNA by the oRS, the concentration of the $otRNA^{UAA}$ in the cells, competition between the $otRNA^{UAA}$ & the release factors and the position of the $codon_{BL}$ on the mRNA [146]. Consequently, the UAA incorporation in proteins can be improved by optimizing these factors. Furthermore, for introducing the required functional group on the target-protein, several OTS may be available. To efficiently explore such a large sample space for selecting the most efficient UAA incorporation system and/or condition, a screening strategy is necessary.

For quantification of UAA incorporation efficiency, among other methods, fluorescent proteins have been used [146]. By introducing one or more blank-codons within the gene of a fluorescent protein, $codon_{BL}$ suppression by the G.C.E. System can be monitored. For a fluorescent protein, the correct folding of the protein domains brings the fluorophore-forming amino acid side-chains in close proximity and allows fluorophore formation. Since all strands of its β-barrel are needed for the correct folding of the fluorescent protein, prematurely terminated polypeptide (*e.g.*, when $codon_{BL}$ suppression does not work) shows no fluorescence. Furthermore, the fluorescence exhibited by the protein's fluorophore is proportional to the amount of correctly folded protein produced. Therefore, by monitoring its fluorescence, the amount of correctly folded fluorescent protein can be quantified without the need for protein purification.

Two fluorescent proteins in tandem have also been used for quantifying the UAA incorporation efficiency by G.C.E. systems [192, 213]. Since fluorescent proteins have characteristic fluorescence, when two different fluorescent proteins are co-expressed, it is possible to quantify their relative amounts by monitoring their fluorescence. If two fluorescent protein genes are present in tandem and if $codon_{BL}$ is introduced between these genes, during polypeptide biosynthesis, the first protein would be expressed continuously, but the second protein would be expressed only when the orthogonal tRNA is able to introduce an amino acid in response to the $codon_{BL}$. In such a scenario, the efficiency of $codon_{BL}$-suppression can be determined by comparing the expression of the first and the second protein.

3.3.1. Our Approach for Quantification of Unnatural Amino Acid Incorporation Efficiency

The total amount of fluorescent protein(s) expressed in the cells can be quantified by using cell-based fluorescence-quantification methods such as F.A.C.S. or fluorescence-imaging. For comparing the UAA incorporation efficiency of different genetic code expansion (G.C.E.) systems, we developed a fluorescence-imaging based screening assay. For this, we have created a mammalian expression vector *Plasmid UTX0* where mRaspberry and eGFP genes are present in tandem, *i.e.*, they are expressed by the same promoter one after another (Figure 22). Since eGFP is the brighter protein of these two (which means it can be detected at lower concentrations), by introducing stop-codon(s) as $codon_{BL}$ within the eGFP gene of *Plasmid UTX0* we have used eGFP as the reporter of the UAA incorporation (*i.e.*, expression of eGFP is dependent on successful $codon_{BL}$-suppression). As mRaspberry is expressed continuously, we have used mRaspberry protein as the reporter of transfection and normal protein production. Since the emission spectrum of eGFP overlaps with the excitation spectrum of mRaspberry, we express them as two separate polypeptides to discourage FRET pair formation and possible loss of eGFP fluorescence [246]. This is done by introducing P2A and T2A sites between the two proteins (Figure 22), both of which signal the ribosome to create a nick in the polypeptide chain during polypeptide biosynthesis. In this way the two fluorescent proteins are expressed from the same promoter, but as two different polypeptides.

Since mRaspberry and eGFP are part of the same expression cassette, ideally, one molecule of eGFP protein should be produced for each molecule of mRaspberry protein. However, since the eGFP gene also contains the blank-codon to signal the UAA incorporation, the expression of eGFP is dependent on the $codon_{BL}$-suppression efficiency of the used G.C.E. system. By comparing the total eGFP expression and total mRaspberry expression the $codon_{BL}$-suppression efficiency of the G.C.E. system can be quantified. Since fluorescence from a fluorescent protein can be used to quantify its amount, the expression ratio of eGFP and mRaspberry proteins can be determined by comparing their fluorescence. For our screening experiments, we have used the ratio of eGFP fluorescence and mRaspberry fluorescence as their Expression Ratio. Since the Expression Ratio depends on the efficiency of $codon_{BL}$-suppression, it can be used to compare UAA incorporation efficiency of different G.C.E. systems or in different expression conditions.

Figure 22: Expression cassette of Plasmid UTX0.
The genes for mRaspberry and eGFP fluorescent proteins are expressed by the same promoter (CMV-IE), but due to P2A and T2A sites, they are expressed as two separate polypeptide chains. Downstream to the eGFP, 10xHis-Myc-FLAG peptide sequence is present to aid the purification of eGFP, if needed.

3.3.2. Normalized Expression Ratio (N.E.R.)

The measured value of total fluorescence depends on the configuration of instrument. For example, for fluorescence imaging, a higher value of exposure time or a brighter intensity of excitation light would lead to a higher value of measured fluorescence. Since these parameters can differ between two different experiments, a normalization method is required to compare the results between different experiments.

For a given G.C.E. system, if the efficiency of $codon_{BL}$-suppression is 100%, for each molecule of mRaspberry, one molecule of $eGFP^{UAA}$ would be produced. Provided that UAA incorporation does not adversely affect fluorophore formation, in practice, this is same as having WT-eGFP gene after the mRaspberry gene (since in this case as well one eGFP protein molecule is produced for one mRaspberry protein molecule). On the other hand, if the efficiency of $codon_{BL}$-suppression is 0%, the eGFP would not be produced at all. In practice, this would be same as not having the eGFP gene after the mRaspberry gene. For any given measurement condition, these two conditions provide us with the two extreme possibilities (*i.e.*, the minimum and the maximum possible values). For anything in between, the molar ratio of eGFP and mRaspberry (*i.e.*, the Expression Ratio) can provide an idea of how good or bad is the UAA incorporation efficiency of the given G.C.E. system. Therefore, for a screening experiment Expression Ratio for *Plasmid UTX0* (which expresses mRaspberry and eGFP proteins via the same promoter but as two different polypeptides) can be used as the standard to normalize the Expression Ratio for any given G.C.E. system.

When the Expression Ratio for a G.C.E. system is normalized to the Expression Ratio of *Plasmid UTX0*, the result is known as the Normalized Expression Ratio or N.E.R.

3.3.3. Control-Conditions for the Screening Assay

In our screening assays, we have used two plasmids, namely, *Plasmid UTX0* and *Plasmid MX01.4* as controls. When introduced in mammalian cells, *Plasmid UTX0* produces mRaspberry and eGFP proteins from the same promoter but as two different polypeptide chains, while *Plasmid MX01.4* produces only mRaspberry protein. mRaspberry fluorescence from *Plasmid MX01.4* can also be used for quantifying the crosstalk between the green and the red fluorescence-imaging channels. We have also used *Plasmid MX01.3*, which expresses only $eGFP^{WT}$ in the mammalian cells, as another control for quantifying crosstalk between the red and the green fluorescence-imaging channels.

3.3.4. G.E.M.S. Plasmids and G.C.E. Plasmids Used for Screening Experiments

For quantifying the UAA incorporation efficiency of G.C.E. Systems and for studying the effects of amino acid positions on the UAA incorporation efficiency, either one or two stop-codons are introduced at different amino acid positions within the eGFP gene of *Plasmid UTX0*. For the sake of convenience, all screening plasmid(s) obtained by mutagenesis in the eGFP gene of *Plasmid UTX0* would be collectively referred to as G.E.M.S. Plasmid(s) (G.E.M.S. = G.C.E. Efficiency Measurement and Screening). For delivering the orthogonal aaRS (oRS) and orthogonal tRNA (otRNA) genes in mammalian cells, expression vectors derived from *Plasmid UCAP* are used. For the sake of convenience, they would be collectively referred to as G.C.E. Plasmids.

3.3.5. Implementation of the Screening Assay

For the screening assay, the relevant G.E.M.S. Plasmid and G.C.E. Plasmid(s)* are introduced in mammalian cells via transient transfection (Figure 23). Five hours after transfection, the expression medium is supplemented with unnatural amino acids (except for negative-control conditions). Due to the mRaspberry gene present on the G.E.M.S. Plasmid, mRaspberry is constitutively expressed in the cells. Due to the G.C.E. Plasmid, the oRS and otRNA are also produced in the cells. If the unnatural amino acid (UAA) is not added in the expression medium, the stop-codon present on the eGFP gene would not be suppressed by the otRNA, therefor no eGFP would be produced. Upon addition of the UAA, the oRS aminoacylates the otRNA with the UAA to generate the orthogonal aminoacylated tRNA ($otRNA^{UAA}$), which can suppress the stop-codon and $eGFP^{UAA}$ is produced. If otRNA can be aminoacylated by the cellular amino acids (either by oRS or by cellular machinery), eGFP would be expressed without addition of UAA in the protein expression medium (*i.e.*, in negative-control conditions). This determines the non-specific incorporation by the G.C.E. system.

The screening assay is performed in multi-well plates. Each screened condition is present in sextuplicate (six individual wells on a multi-well plate). For quantifying the UAA incorporation, fluorescence images are acquired for each well using red and green fluorescence channels (as detailed in Section 2.10.4). Acquired images are analyzed by CellProfiler pipelines (Figure 12 and Figure 25) to subtract the blank form each channel, identify the individual objects (single cells) from blank-corrected images, disregard the saturated objects (overexposed pixels) and measure the total eGFP and mRaspberry fluorescence for each object (Details in Section 2.10.6). The output of CellProfiler pipelines is statistically analyzed in R (Figure 13 and Figure 26) to classify the individual wells according to their expression conditions, calculate the Expression Ratio (E.R.) of individual wells and to calculate the Normalized Expression Ratio (N.E.R.) for individual wells (Details in Section 2.10.7). Endpoint-fluorescence analysis (measurement of the total mRaspberry and eGFP fluorescence after cell-lysis) is used alongside the fluorescence-image analysis for quantifying UAA incorporation.

As a part of this screening assay, the pipelines for fluorescence-image analysis and scripts for statistical analysis have also been developed (as detailed in Sections 2.10.6 and 2.10.7) and will be made available on a public research repository.

* G.C.E. Plasmids express the orthogonal aaRS and the orthogonal tRNA in mammalian cells.

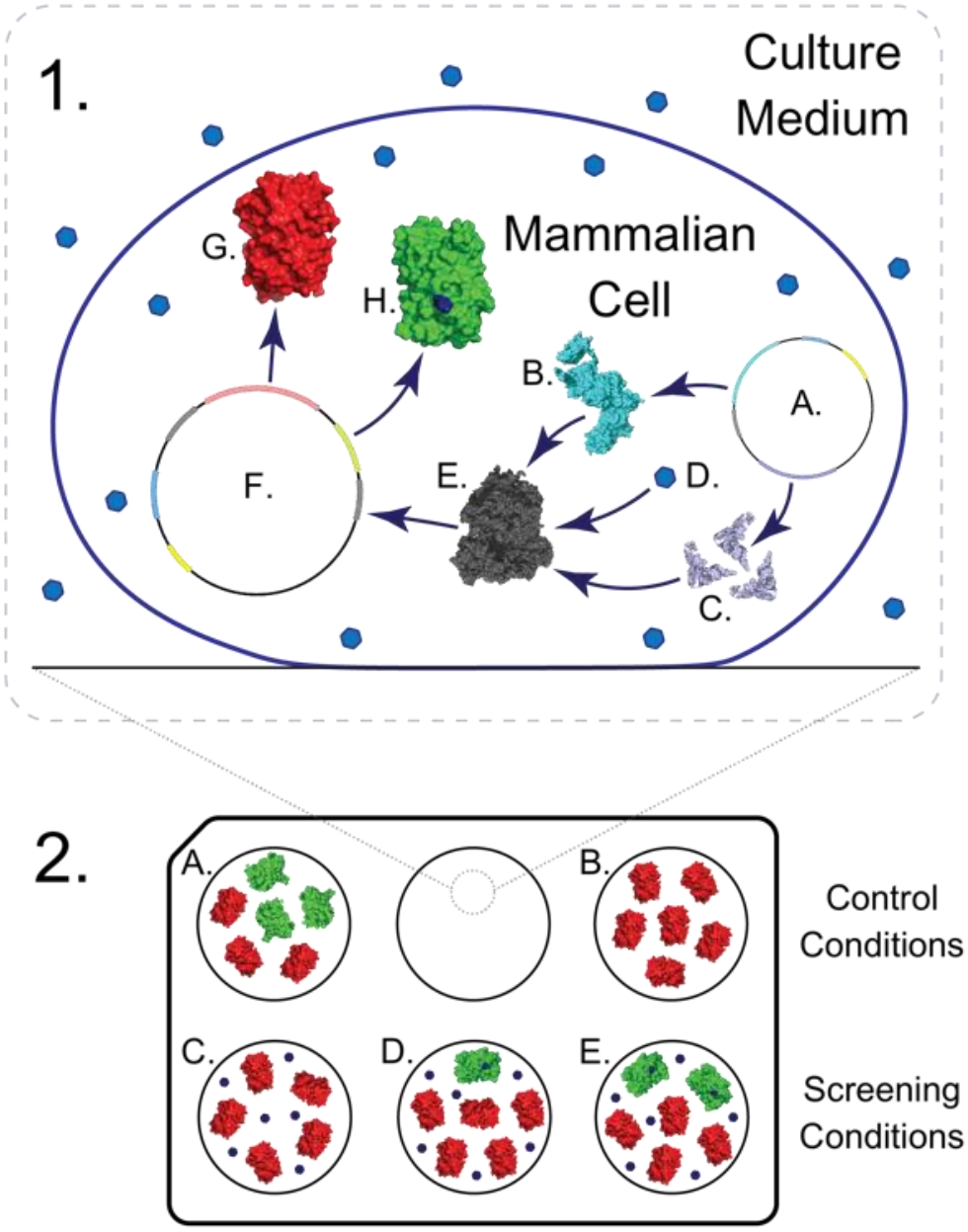


Figure 23: Schematics for Screening of UAA Incorporation Efficiency

1. Schematics of the screening assay: The G.C.E. Plasmid and the G.E.M.S. Plasmid are transiently transfected into the mammalian cells. Plasmid UCAP derived G.C.E. machinery expression vector, a.k.a. G.C.E. Plasmid (A), expresses orthogonal RS (oRS) (B, representation from PDB ID - 2AMC) and orthogonal tRNA (otRNA) (C, representation from PDB ID - 1EHZ). When the unnatural amino acid (UAA) (D) is added to the medium, oRS charges otRNA to produce aminoacylated otRNA (not shown). The protein-expression machinery of the cell (E, representation from PDB ID - 6Y0G) can recognize the aminoacylated otRNA. Meanwhile, G.E.M.S. plasmid (F) constitutively expresses WT-mRaspberry protein (G, representation from PDB ID - 2H5Q). When the aminoacylated otRNA can identify and suppress the stop codon present on the eGFP gene, genetic code expansion is successful and eGFP protein containing the UAA (H, representation from PDB ID - 2Y0G) is produced as well.
2. Different transfection conditions present in well: For cell-based screening experiments, the conditions present in a multi-well plate may be broadly classified either as control conditions or screening conditions. The output of the control conditions is used to standardize the result of screening conditions. For screening UAA incorporation efficiency of G.C.E. systems, Plasmids UTX0 and MX01.4 provide the control conditions. For Plasmid UTX0, eGFP and mRaspberry are expressed in 1:1 molar ratio (A). For Plasmid MX01.4 only mRaspberry is expressed (B). For screening conditions, after co-transfection with G.C.E. Plasmid and G.E.M.S. plasmid, eGFP and mRaspberry proteins can be expressed in varying ratio (C, D and E), indicative of the UAA incorporation efficiency of the G.C.E. system. If eGFP is expressed without UAA, it suggests a leak in the orthogonality of the G.C.E. system.

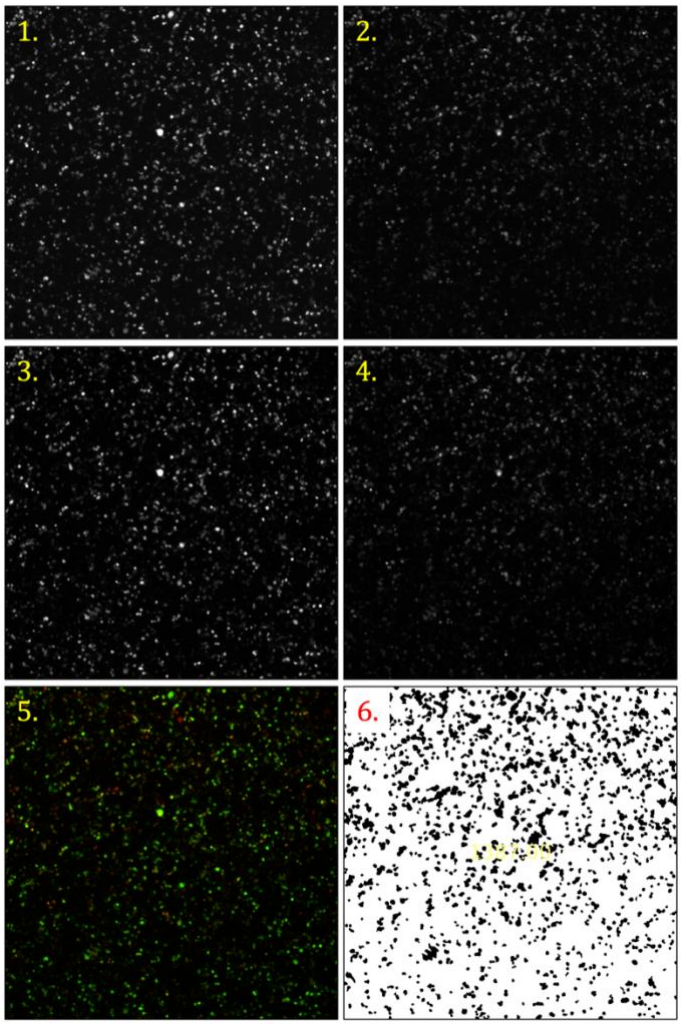


Figure 24: Stages of Image Analysis.
Images obtained from the Cytation5 multimode reader. Are processed using CellProfiler pipelines. Different stages of processing are represented. Unprocessed image from the green-fluorescence channel (1). Unprocessed image from the red-fluorescence channel (2). Blank-corrected green-channel image (3). Blank corrected red-channel image (4). Overlay of blank-corrected green-channel image and blank-corrected red-channel image, represented in their respective colors (5). Object identification (dark spots) by CellProfiler pipelines (6).

Endpoint fluorescence analysis presents the 'overall picture' for the screening conditions after cell-lysis (end of experiment). In comparison, the fluorescence image-based analysis is a non-invasive method that can provide important information about the intermediate stages, especially, protein expression over time, to optimize the duration of protein production. Since fluorescence image-based analysis method identifies individual cells for analysis, this 'single-particle' approach presents a more comprehensive picture of the protein expression condition, and can be used for optimization of transfection conditions or for characterizing the homogeneity of protein expression [213]. The biggest advantage of image-based analysis is that it works well even when a low proportion of cells are transfected.

3.3.6. Workflow Automation of the Screening Assay

The screening-assay used in our screening experiments is compatible with process automation and high-throughput approach. Figure 11 summarizes our approach for automating the processes involved in this assay. For this study, all the steps for this assay, except the transfection step, from could be automated by creating appropriate pipelines for the automation setup. The transfection step could not be automated due to the lack of appropriate sterile automation-compatible instruments for transient transfection of

mammalian cells in multi-well plates. More recent lab-automation setups acquired by our research group will address this issue in future.

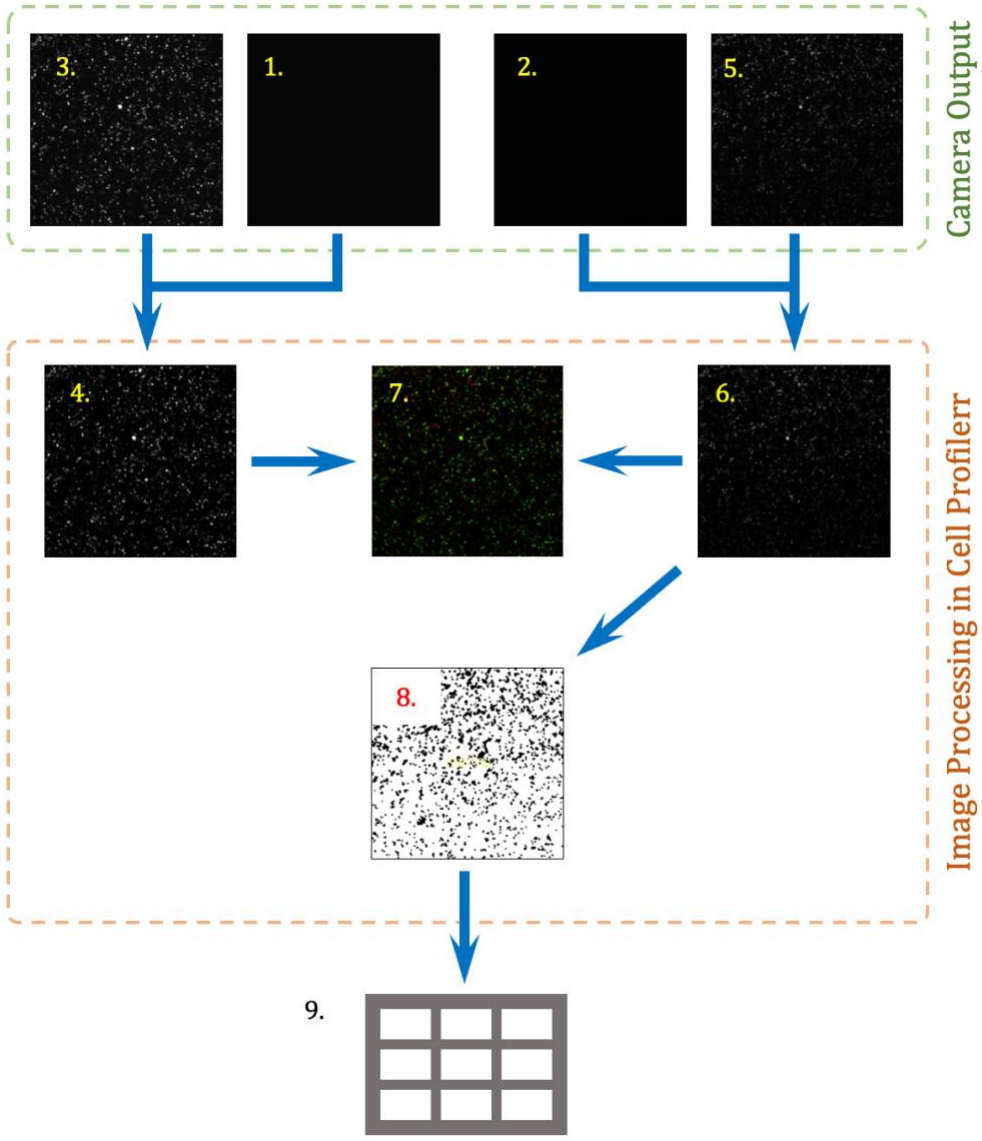


Figure 25: Image Processing Using CellProfiler.
For each multi-well plate the averaged-blank-image is generated by averaging all six untransfected wells. For each multi-well plate separate averaged-blank-image is generated for the green channel (1) and the red channel (2). For each well, the 'green-channel averaged-blank-image' (1) is subtracted from the 'acquired-image' of the green channel (3) to give 'blank-corrected green-channel-image' (4). Similarly, the 'red-channel averaged-blank-image' (2) is subtracted from the 'acquired-image' of the red channel (5) to give 'blank-corrected red-channel image' (6). From the blank-corrected images of both channels, a composite image (7) is generated, where blank-corrected green and red channel images are overlaid. From the blank corrected red channel image, individual objects (single cells) are identified (8). For each object, its area, its position and its blank corrected eGFP and mRaspberry fluorescence exported as a CSV file (9).

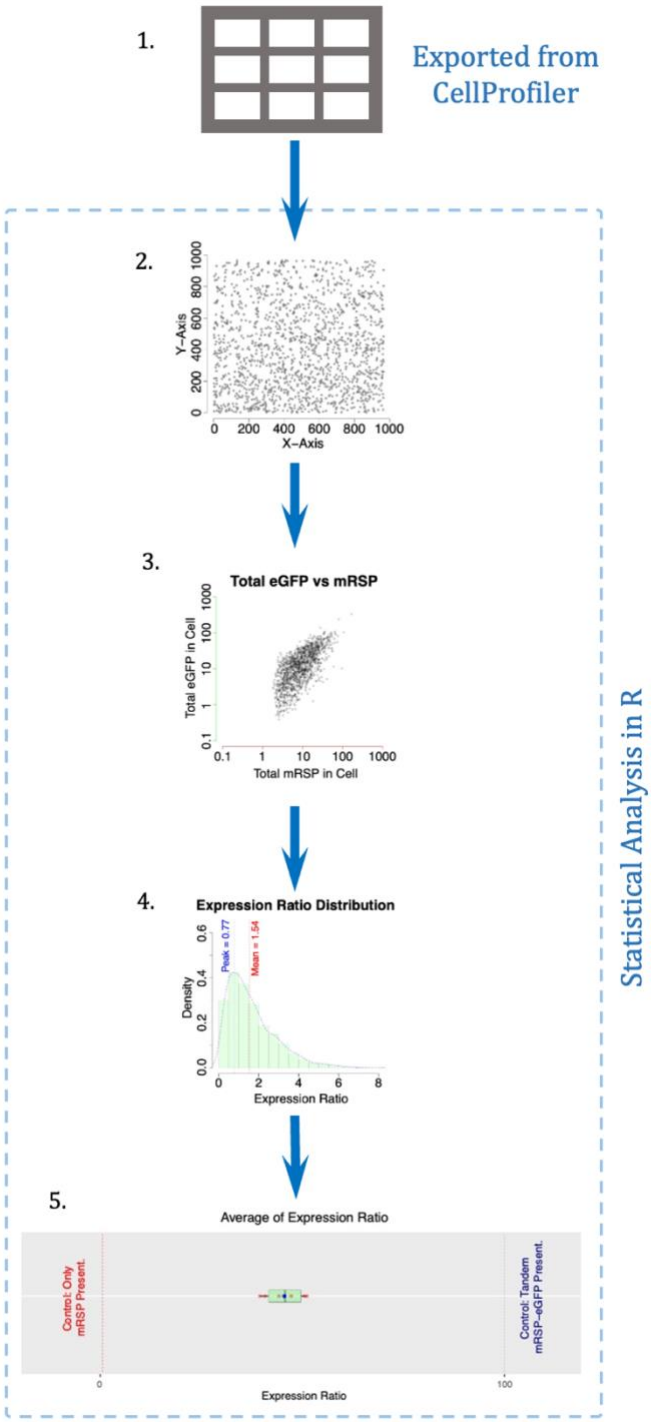


Figure 26: Schematics of Statistical Analysis in R.
From the CSV file (1) exported by the CellProfiler pipeline *Background_Subtraction_and_Analysis.cppipe*, for each well, we obtain a list of identified objects (single cells) with information such as location of the object (2) and total eGFP and mRaspberry fluorescence exhibited by the object (3). For each object, the Expression Ratio can be individually calculated and the distribution of the Expression Ratio can be studied (4). For the whole well, the total eGFP fluorescence and total mRaspberry fluorescence are separately calculated and overall Expression Ratio of the well is calculated (step not displayed in image). The Expression Ratio of the well is normalized against the Expression Ratio of *Plasmid UTX0* condition and the data is plotted for the sextuplicate of the condition (5). The mean value (blue circle in (5)) is used for box-plots.

3.4. SCREENING EXPERIMENTS FOR COMPARING UAA INCORPORATION

Using the Orthogonal Translation Systems (or OTSs, *i.e.*, the oRS and otRNA pair), we wanted to introduce two fluorophores in proteins for smFRET studies. By using genetic code expansion (G.C.E.), fluorophores can be introduced in proteins via two ways - either a fluorescent UAA can be directly incorporated in the target-proteins, or a fluorophore can be attached to a unique reactive handle present on the incorporated unnatural amino acid (UAA). We decided to explore both these options.

At the time of commencement of this study, there were limited reports of two-UAA incorporation in proteins expressed in mammalian cells. Moreover, since this field is still evolving, we anticipated that more OTSs would soon join the expanding universe of the orthogonal aaRS/tRNA pairs to incorporate an even more versatile set of UAAs, and our needs may not be limited to incorporation of fluorophores in proteins. Therefore, we decided against using already reported G.C.E. systems for two-UAA incorporation and decided to develop a strategy for multiple UAA incorporation in proteins, to have more freedom in the choice of unnatural amino acids.

We identified and procured seven different orthogonal tRNA synthetases (oRSs) evolved for mammalian systems that were known to incorporate either fluorescent UAAs or UAAs with bioorthogonal reactive handles. Of these seven, the $^{ANAP}_{EcLeu}RS$ can incorporate L-3-(6-acetylnaphthalen-2-ylamino)-2-aminopropionic acid (ANAP), the $^{DanAla}_{EcLeu}RS$ can incorporate dansylalanine (DanAla), the $_{EcTrp(h9)}RS$ and $_{EcTrp(h14)}RS$ can incorporate azidotryptophan (AzW) & propargyloxytryptophan (PrW), the $_{OMeY}RS$ can incorporate azidophenylalanine (AzPhe) & propargyloxyphenylalanine (PrPhe), and the $_{MbPyl}RS$ and $_{MmPyl}RS$ can incorporate azidolysine (N_3Lys) & propargyloxylysine (PrK) [142, 146, 176, 185, 186]. Out of these eight UAAs mentioned here, ANAP and DanAla are fluorescent amino acids and the rest contain bioorthogonal click-reaction handles, which can be used, among other applications, to introduce fluorophores after protein biosynthesis.

For introducing two unnatural amino acids in proteins, the two OTS must be orthogonal to each other, *i.e.*, they should not recognize each other's aaRS, tRNA or UAA for aminoacylation reaction. We also need two different blank-codons to signal the incorporation of the two UAAs. For mammalian systems, the three stop codons and quadruplet codons can be used as the blank-codons. Of these two options, stop-codon suppression is well-established for mammalian expression systems, so we decided to use two different stop codons as blank-codons for UAA incorporation in proteins. To provide unique blank-codons for the two OTS that could be used together, we decided to characterize the UAA incorporation efficiency of each of the seven aforementioned oRS for each of the three stop codons. Accordingly, the respective $tRNA_{UUA}$, $tRNA_{CUA}$ and $tRNA_{UCA}$ variants were procured via synthetic genes for these oRS. By combining the seven different oRS and the three kinds of otRNA for each oRS we could have 21 orthogonal aaRS/tRNA pairs for UAA incorporation using mammalian expression system. By separately introducing these 21 pairs of genes in *Plasmid UCAP*, 21 mammalian expression vectors were created. For the sake of convenience, these are collectively called as G.C.E. Plasmids.

Before incorporating two UAAs in our target-proteins, we wanted to characterize the efficiency these OTS to select for the best G.C.E. System suited for our needs. As the aforementioned eight amino acids were of interest to us, we decided to use them in our screening assay. However, since we could not procure AzW and PrW, we decided to use 5-Hydroxytryptophan (5-HTP) instead for quantifying the UAA incorporation efficiency of $_{EcTrp(h9)}RS$ and $_{EcTrp(h14)}RS$ based G.C.E. systems. In this way, by using the screening assay described in Section 3.3, we decided to characterize the UAA incorporation efficiency of 30 sets of OTS-UAA (*i.e.*, a triplet of oRS-otRNA-UAA).

Since stop-codon suppression by the otRNA can be dependent on the context of the stop-codon on the mRNA, we also decided to develop a strategy to screen for the effects of the position of the stop-codon on UAA incorporation. For doing so, we decided to introduce UAA at several amino acid positions in eGFP. Since our cell-based assay relies on the fluorescence of mRaspberry and eGFP, we wanted to make sure that UAA incorporation does not adversely affect the fluorescence of these proteins. Since mRaspberry is expressed as a separate polypeptide, we reasoned that it was unlikely that UAA incorporation in eGFP would have any effect on its fluorescence properties. To make sure that UAA incorporation in eGFP would not interfere with the fluorophore formation, we chose 14 amino acid positions where the side-chains were solvent exposed and/or were pointing away from the β-barrel of eGFP. As the control, we also selected one

position (TYR143) that was buried. In this way we selected 15 different amino acid positions on the eGFP for UAA incorporation (Figure 27). Since we wanted to test the suppression of all three stop codons, for each of the 15 amino acid positions, TAA, TAG and TGA stop-codons were introduced in *Plasmid UTX0* by mutagenesis. These plasmids are collectively referred to as G.E.M.S. Plasmids. Eventually, by using different combinations of G.C.E. Plasmids, G.E.M.S. Plasmids and the UAA, for these 15 amino acid positions in the eGFP, and by using 30 sets of OTS-UAA, we screened 450 conditions for incorporation of unnatural amino acids in eGFP.

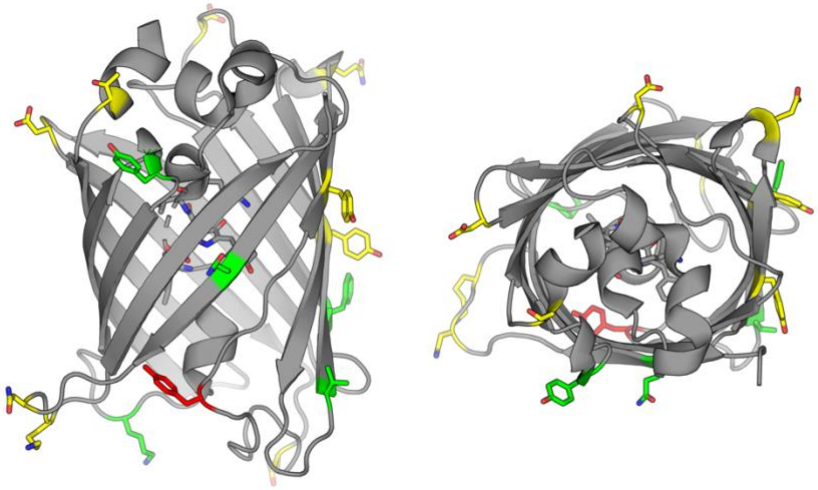


Figure 27: Amino Acid Positions Selected for Screening Experiments for UAA Incorporation in eGFP.
Red, yellow and green colored side chains represent those amino acid positions that were screened for UAA incorporation in eGFP. The yellow and green colored side chains were selected because they point out of the β-barrel of eGFP. The red colored side chain represents a buried amino acid TYR143. After one-UAA incorporation screening, green colored side chains were shortlisted for simultaneously introducing two UAAs in eGFP.

In such a screening setup, between two different conditions, the difference in N.E.R. values could arise because of a combination of three factors: 1) difference in stop-codon suppression efficiency of the G.C.E. systems, 2) difference in stop-codon suppression at given amino acid positions, and 3) (despite our best efforts to avoid this) the disruption in the fluorophore formation due to destabilization of the protein structure. Nonetheless, a high N.E.R. value (after removing the background from the non-specific amino acid incorporation) signifies two things: the UAA could be incorporated and the protein could fold properly to form the fluorophore. So based on single-UAA incorporation screening, it should be possible to shortlist those conditions and amino acid positions that have a higher possibility for successfully incorporating two unnatural amino acids simultaneously. So before screening for simultaneous incorporation of two unnatural amino acids, we screened for incorporation of one unnatural amino acid at different positions in eGFP and by using different G.C.E. systems.

3.4.1. Two-Plasmid Co-transfection

By the design of our screening assay, to quantify the UAA incorporation efficiency, we need to co-transfect the cells with two different plasmids: the G.C.E. Plasmid and the G.E.M.S. Plasmid. Our first concern was how efficiently the two plasmids can co-transfect the cells. To understand this, screening experiments were performed where HEK cells were transiently transfected with different ratios of eGFP and mRaspberry expressing plasmids, *i.e.*, *Plasmid MX01.3* and *Plasmid MX01.4* respectively (Figure 28). For *Plasmid MX01.3*, no mRaspberry signal was observed in the cells (Figure 28, A). Similarly, for *Plasmid MX01.4*, no eGFP signal was observed in the cells (Figure 28, B). This indicated that crosstalk between the red and the green fluorescence channels was negligible.

After transfecting the cells with *Plasmid UTX0*, which expresses eGFP and mRaspberry from the same promoter but as different polypeptide chains, both eGFP and mRaspberry signals were observed in the cells (Figure 28, C). In this case, the average expression ratio for the cells is 3.54. However, it is important to reiterate here that if the exposure time for image-acquisition is changed for either of the two fluorescence channels, the value of average expression will change as well. Hence it is necessary to keep this control-

condition (*i.e.*, transfection with *Plasmid UTX0*) in every experiment to normalize the screening results and calculate the Normalized Expression Ratio or N.E.R. for the samples. When cells were co-transfected with three different ratio of *Plasmid MX01.3* and *Plasmid MX01.4*, we found no indication that cells expressing only eGFP or only mRaspberry were present (Figure 28, D-F).

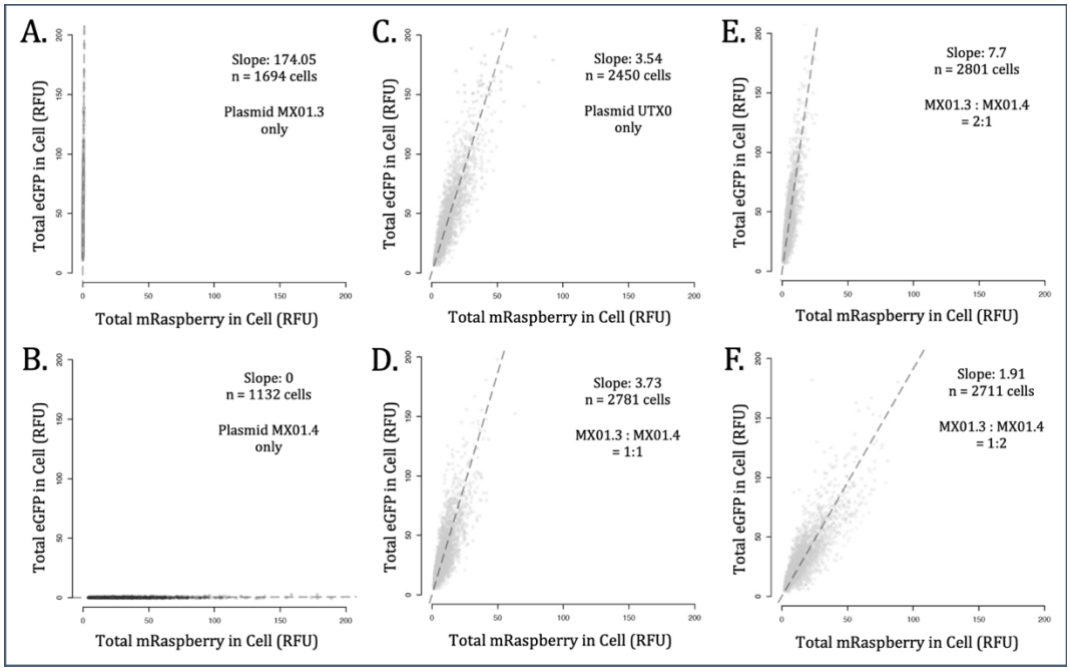


Figure 28: Co-transfection of HEK Cells with Two Different Plasmids
Total eGFP fluorescence and total mRaspberry fluorescence, as measured within the individual cells by image analysis, is plotted on Y-axis and X-axis respectively. Each point represents one identified cell. The dashed line represents the trendline of the plotted points and the slope is the slope of the trendline. n denotes the total number of cells that were identified in the image. Plot (A) is condition where cells were transfected with eGFP protein expressing *Plasmid MX01.3*. Cells in plot (B) were transfected with mRaspberry protein expressing *Plasmid MX01.4*. For plot (C), cells were transfected with *Plasmid UTX0*, which expresses mRaspberry and eGFP proteins using the same promoter. Cells in plot (D, E and F) were co-transfected with *Plasmids MX01.3 and MX01.4*. For plot (D), these two plasmids were present in 1:1 ratio, for plot (E) MX01.3 : MX01.4 ratio was 2:1 and for plot (F) this ratio was 1:2.

These results demonstrate that for the tested conditions, irrespective of the ratio of two co-transfecting plasmids, both plasmids were always delivered in the cells. Since the G.E.M.S. Plasmids and G.C.E. Plasmids have similar size, they too could be used for successful co-transfection of HEK cells.

3.4.2. Different Ratio of G.C.E. and G.E.M.S. Plasmids for Transient Transfection

After ensuring that co-transfection works well, we wanted to understand the impact of different ratios of the G.C.E. Plasmid and the G.E.M.S. Plasmid. Since unnatural amino acid incorporation at position TYR39 of eGFP has already been widely reported in the literature, to find the optimal ratio of these two plasmids, the G.E.M.S. plasmid containing eGFP^{TYR39TAG} mutation was used with the ${}_{MmPyl}RS$ / ${}_{MmPyl}tRNA_{CUA}$ pair & N_3Lys and the G.E.M.S. plasmid containing eGFP^{TYR39TGA} mutation was used with the ${}^{ANAP}_{EcLeu}RS$ / ${}^{ANAP}_{EcLeu}tRNA_{UCA}$ pair & ANAP. Figure 29 and Figure 30 summarize the results of this experiment.

When 1 µg or 3 µg total DNA was used for transfection (highlighted labels on Y-axis), we observed that at 1:1 ratio of these two plasmids, the total amount of DNA used for transfection did not have a drastic effect on N.E.R. as well as on total eGFP^{UAA} production for both the tested G.C.E. Systems. Contrary to expectations, the average value of total eGFP-fluorescence was higher when lower amount of DNA (1 µg vs. 3 µg) was used for transfection. This result demonstrates that 1 µg DNA was sufficient for protein expression in one well of a 96-well plate. Higher amount of DNA was either not taken up by the cells or was taken up, but the transcription-translation machinery of the cells was already saturated. The slightly lower amount of total eGFP-fluorescence for a higher amount of DNA could be a result of possible cytotoxicity by the DNA-PEI complex.

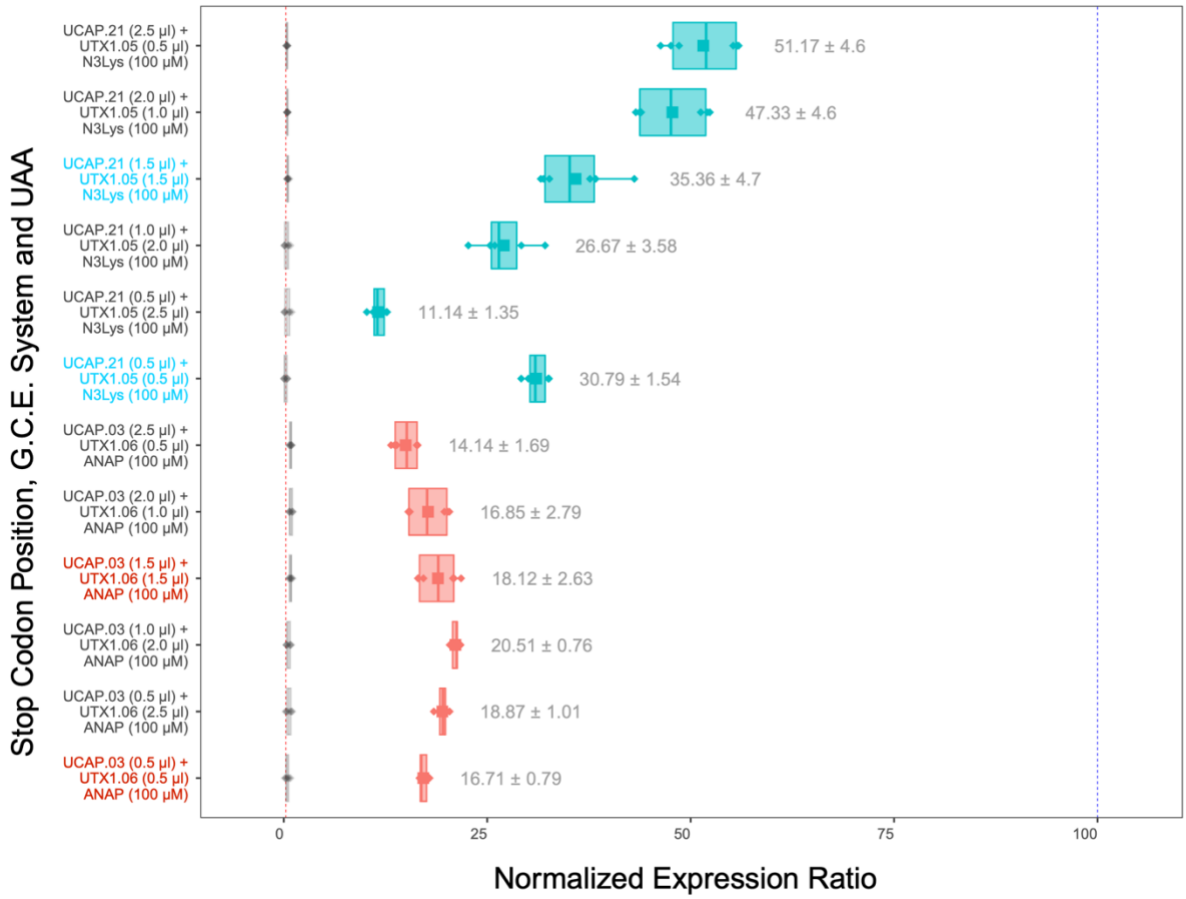


Figure 29: Effect of the Ratio of G.C.E. Plasmid and G.E.M.S. Plasmid on UAA Incorporation.
UCAP.21 delivers ${}_{MmPyl}RS$ / ${}_{MmPyl}tRNA_{CUA}$ pair and UCAP.03 delivers ${}^{ANAP}_{EcLeu}RS$ / ${}^{ANAP}_{EcLeu}tRNA_{UCA}$ pair in mammalian cells. UTX1.05 contains eGFP^{TYR39TAG} and UTX1.06 contains eGFP^{TYR39TGA}. Y-Axis details the screening condition (G.C.E. Plasmid & its amount, G.E.M.S. Plasmid & its amount, and UAA & its concentration). X-Axis represents the linear scale for the N.E.R. Diamonds represent the individual data points (i.e., the N.E.R. value) for the six replicates. Squares represent the mean N.E.R. value. Box plots represent the second and the third quartile values for the given data while whiskers represent the first and the fourth quartile. Numerical values represent the value of the average N.E.R. for the given condition. Red dashed line represents the N.E.R. for the control-condition from transfection with *Plasmid MX01.4* and blue line represents the N.E.R. for the control condition from transfection with *Plasmid UTX0*. Box plots are colored according to the oRS present in the G.C.E. system - red is for ANAP-RS and green is for MmPyl-RS. Grey box plots are for those conditions where the same G.C.E. and G.E.M.S. plasmids were present, but UAA was not added. On Y-Axis, those labels are colored which have 1:1 ratio of G.C.E. Plasmid and G.E.M.S. Plasmid in transfection mix.

For the ${}^{ANAP}_{EcLeu}RS$ / ${}^{ANAP}_{EcLeu}tRNA_{UCA}$ pair in presence of 100 µM ANAP, the N.E.R. values (Figure 29) were about the same for different G.C.E. Plasmid to G.E.M.S. Plasmid ratio while total eGFP-fluorescence (Figure 30) generally increased with the proportion of G.E.M.S. Plasmid. This indicated that for most of the tested conditions, the amount of orthogonal aaRS/tRNA pair was not the limiting factor for the TGA-suppression and ANAP incorporation. However, eGFP^{ANAP} expression was limited by the amount of G.E.M.S. Plasmid present in the cells. Increasing the relative amount of G.E.M.S. Plasmid in the cells increased total eGFP-fluorescence, till the point when the expression of the orthogonal aaRS/tRNA pair became the limiting factor. In absence of ANAP, almost no eGFP-fluorescence was observed (Figure 29 and Figure 30), which demonstrates low non-specific incorporation by ${}^{ANAP}_{EcLeu}RS$ / ${}^{ANAP}_{EcLeu}tRNA_{UCA}$ pair.

A different trend was observed for the ${}_{MmPyl}RS$ / ${}_{MmPyl}tRNA_{CUA}$ pair in presence of 100 µM N_3Lys. Both N.E.R. and total eGFP-fluorescence generally increased with the proportion of G.C.E. Plasmid. This indicated that, for most of the tested conditions, the presence of insufficient amount of the orthogonal aaRS/tRNA pair in the cell was the limiting factor for UAA incorporation in eGFP. By increasing the proportion of G.C.E. Plasmid, more orthogonal aaRS/tRNA could be produced and the G.C.E. machinery could compete better for the TAG-codon suppression. For 5:1 ratio of G.C.E. Plasmid to G.E.M.S. Plasmid, the N.E.R. still improved slightly as compared to 2:1 ratio, but total eGFP^{N3Lys} production reduced because presence of insufficient

amount of G.E.M.S. Plasmid became the overall limiting factor. For $_{MmPyl}RS$ / $_{MmPyl}tRNA_{CUA}$ pair as well the non-specific incorporation was low.

For screening experiments, it is practical to use same ratio of the G.C.E. Plasmid and G.E.M.S. Plasmid throughout the screened conditions. Therefore, we decided that using 1:1 ratio of G.C.E. Plasmid and G.E.M.S. Plasmid was a good compromise for ensuring high rate of stop-codon suppression while keeping up the eGFP^{UAA} production. Furthermore, since for 1:1 ratio of G.C.E. Plasmid and G.E.M.S. plasmid, N.E.R. and total eGFP-fluorescence was comparable when either 1 µg or 3 µg of total DNA was present during transfection, we decided that lower amount of total DNA was better for avoiding possible transfection-induced cytotoxicity. However, for better pipetting accuracy, we decided to use 2 µg of total DNA, rather than 1 µg, in our screening experiments, when 96-well plates were used.

By screening for different ratios of G.C.E. and G.E.M.S. Plasmids, we find that the ratio of the two plasmids may affect the UAA incorporation as well as overall eGFP^{UAA} production. By monitoring both N.E.R. and total eGFP fluorescence, we can determine the limiting factors and optimize the transfection conditions accordingly.

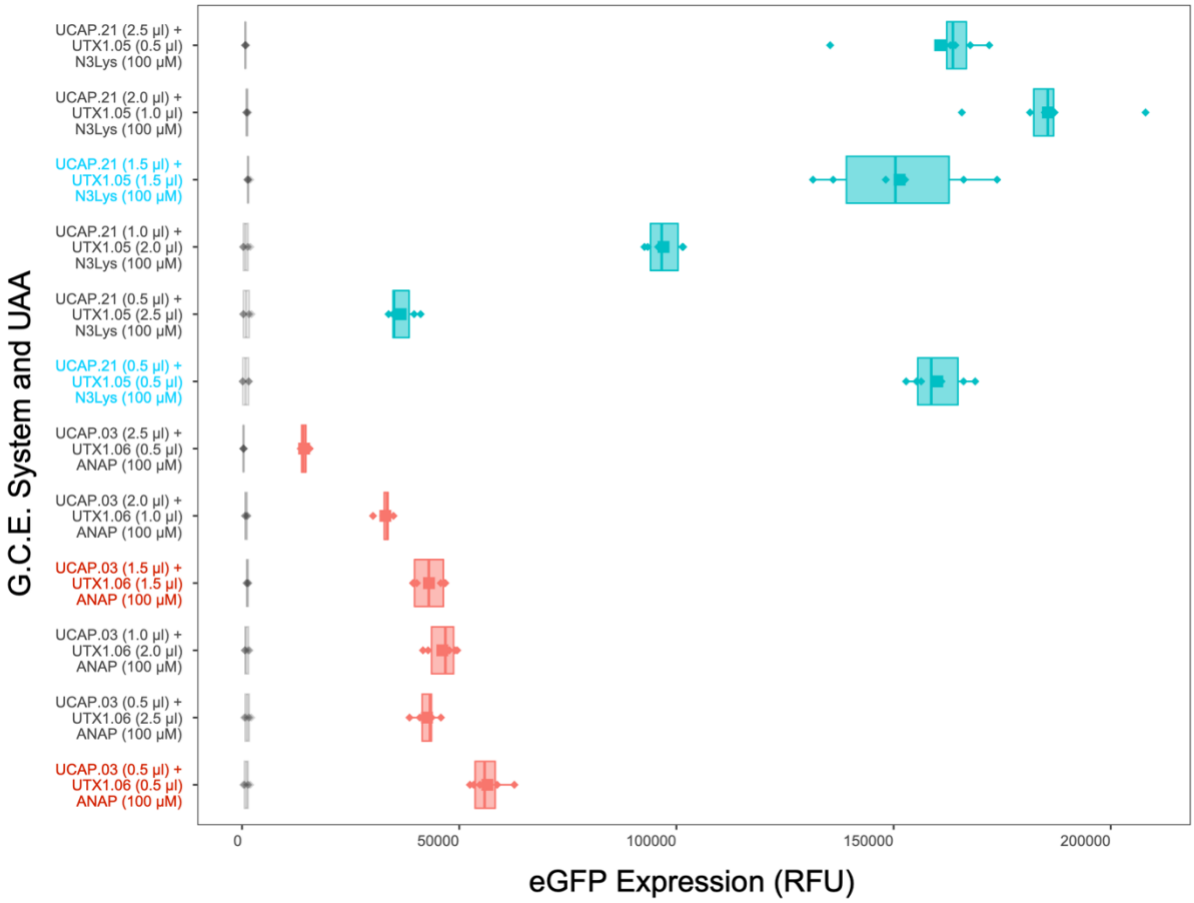


Figure 30: Optimal Ratio of G.C.E. Plasmid and G.E.M.S. Plasmid - Total eGFP Production.
UCAP.21 delivers $_{MmPyl}RS$ / $_{MmPyl}tRNA_{CUA}$ pair and UCAP.03 delivers $^{ANAP}_{EcLeu}RS$ / $^{ANAP}_{EcLeu}tRNA_{UCA}$ pair in mammalian cells. UTX1.05 contains eGFP^{TYR39TAG} and UTX1.06 contains eGFP^{TYR39TGA}. Y-Axis details the screening conditions. X-Axis represents the linear scale for total eGFP-fluorescence. Diamonds represent the individual data points (i.e., the total eGFP-fluorescence) for the six replicates. Squares represent the mean value. Box plots represent the second and the third quartile values. Whiskers represent the first and the fourth quartile values. Box plots are colored according to the oRS present in the G.C.E. system - red is for ANAP-RS and green is for MmPyl-RS. Grey box plots are for those conditions where the same G.C.E. and G.E.M.S. plasmids were present, but UAA was not added. On Y-Axis, those labels are colored which have 1:1 ratio of G.C.E. Plasmid and G.E.M.S. Plasmid in transfection mix.

3.4.3. Effects of Unnatural Amino Acid Concentration

$_{MbPyl}RS$ and $_{OMeY}RS$ based G.C.E. systems are polyspecific, *i.e.*, they can recognize more than one UAAs for aminoacylating the respective otRNAs. To understand the effects of UAA concentration for stop-codon suppression and UAA incorporation, all three otRNA variants for these two oRS were tested with two different UAAs and with two different UAA concentrations present in the expression media. The result from this screening is summarized in Figure 31 and Figure 32.

For the $_{MmPyl}RS$ / $_{MmPyl}tRNA$ pair, either 100 μM or 2 mM of Azidolysine (N_3Lys) or Propargyloxylysine (PrK) unnatural amino acids were added to the culture medium during protein expression. Among the three stop-codons, for this aaRS/tRNA pair, TAG-suppression gave the best results, both in terms of N.E.R. values and total eGFP^{UAA} production. A 20-fold increase in PrK concentration drastically increased the N.E.R. value as well as total eGFP^{UAA} production in case of TAG-suppression and TGA-suppression. For the TAA-suppression, these values increased slightly but the overall values remained low. In comparison, a 20-fold increase in N_3Lys concentration only slightly increased both N.E.R. and eGFP^{UAA} production. In absence of unnatural amino acids, we observed very low non-specific incorporation from the $_{MmPyl}RS$ / $_{MmPyl}tRNA$ pair.

For the $_{OMeY}RS$ / $_{OMeY}tRNA$ pair either 100 μM or 5 mM of Propargyloxyphenylalanine (PrPhe) or Azidophenylalanine (AzPhe) unnatural amino acids were added to the culture medium during protein expression. Among the three stop-codons, for this aaRS/tRNA pair as well, TAG-suppression gave the best results, both in terms of N.E.R. values and total eGFP^{UAA} production. A higher concentration of unnatural amino acid always resulted in about 2-3 times higher N.E.R. value for the AzPhe incorporation. However, the increase in the N.E.R. was not always accompanied by increase in the total eGFP-fluorescence. For example, in case of AzPhe incorporation using TAG stop-codon, between 100 μM and 5 mM AzPhe, the average N.E.R. value increased from 31.63 ± 1.04 to 53.29 ± 2.71 (Figure 31) but the total eGFP-fluorescence value reduced (Figure 32). This demonstrates that high UAA concentration can adversely impact protein production. In case of PrPhe incorporation, an increase in the N.E.R. value was always accompanied by an increase in the total eGFP fluorescence. We also observed about 2.5% non-specific incorporation for the $_{OMeY}RS$ / $_{OMeY}tRNA_{CUA}$ pair.

Based on these results, we decided to use 1 mM unnatural amino acids for screening experiments. However, since L-ANAP is very expensive, in the interest of cost, we decided to keep using 100 μM of this amino acid. Furthermore, we hoped to eventually acquire AzW and PrW unnatural amino acids and were using 5-HTP only for characterizing the $_{EcTrp(h9)}RS$ and $_{EcTrp(h14)}RS$ based G.C.E. systems, without the intention for using it in further studies, so in the interest of material conservation, for 5-HTP as well we used 100 μM concentration in our screening experiments.

By screening two different UAA concentrations in the protein expression medium we find that UAA concentration could impact UAA incorporation efficiency and overall protein production. This underscores the importance of a screening strategy to optimize protein production.

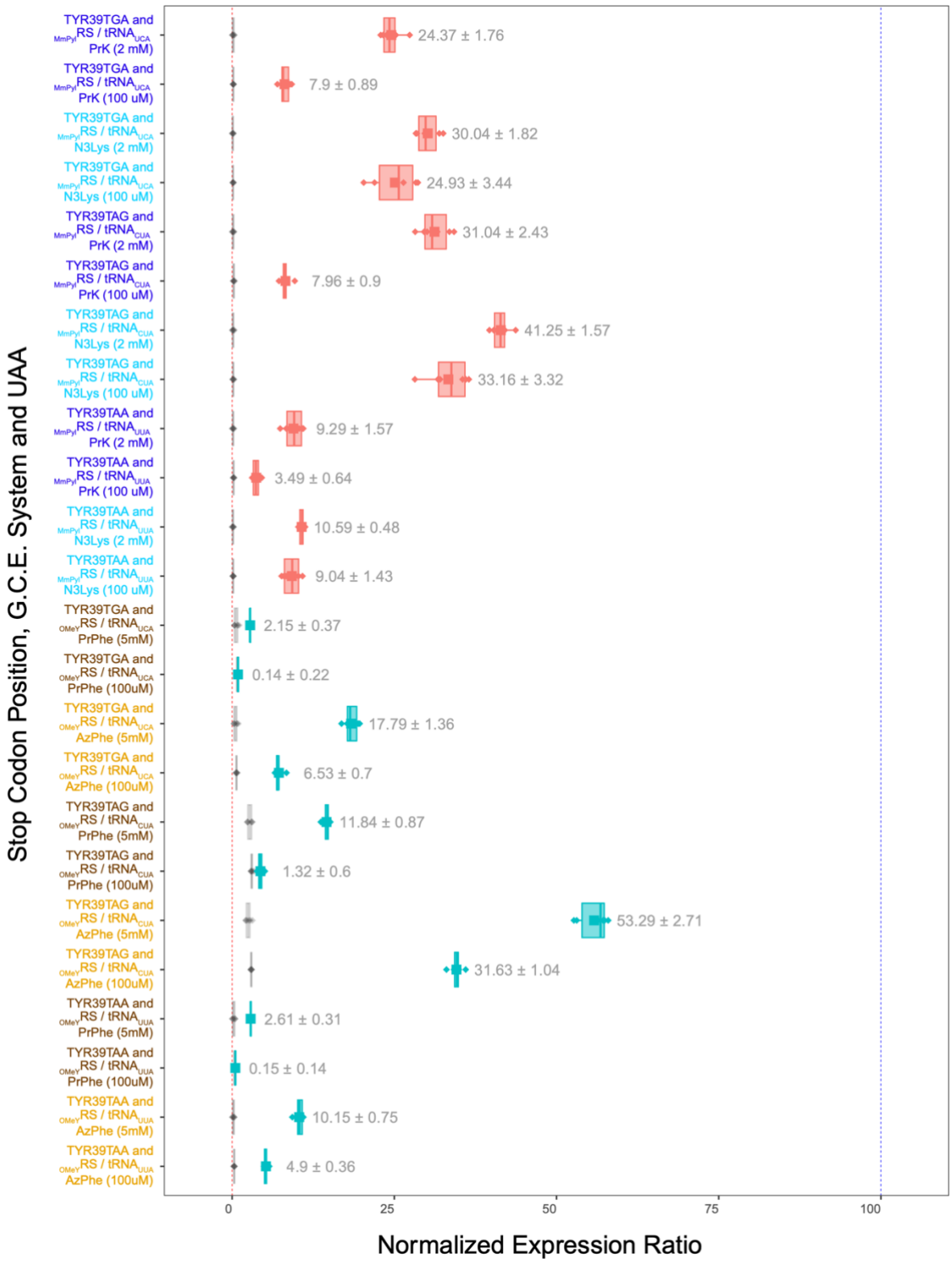


Figure 31: Effects of Unnatural Amino Acid Concentration.
Y-Axis details the screening condition (eGFP Mutation, G.C.E. system and UAA & its concentration). X-Axis represents the linear scale for the N.E.R. Diamonds represent the individual data points (i.e., the N.E.R. value) for the six replicates. Squares represent the mean N.E.R. value. Box plots represent the second and the third quartile values for the given data while whiskers represent the first and the fourth quartile. Numerical values represent the value of the average N.E.R. for the given condition. Red dashed line represents the N.E.R. for the control-condition from transfection with *Plasmid MX01.4* and blue line represents the N.E.R. for the control condition from transfection with *Plasmid UTX0*. Box plots are colored according to the oRS present in the G.C.E. system - red is for MmPyl-RS and green is for OMeY-RS. Grey box plots are for those conditions where the same G.C.E. and G.E.M.S. plasmids were present, but UAA was not added. Y-Axis labels are colored according to the incorporated UAA: orange is for AzPhe, brown is for PrPhe, sky-blue is for N_3Lys and dark blue is for PrK.

Figure 32: Effects of Unnatural Amino Acid Concentration - Total eGFP Production.
Y-Axis details the screening conditions. X-Axis represents the linear scale for total eGFP-fluorescence. Diamonds represent the individual data points (i.e., the total eGFP-fluorescence) for the six replicates. Squares represent the mean value. Box plots represent the second and the third quartile values. Whiskers represent the first and the fourth quartile values. Box plots are colored according to the oRS present in the G.C.E. system - red is for MmPyl-RS and green is for OMeY-RS. Grey box plots are for those conditions where the same G.C.E. and G.E.M.S. plasmids were present, but UAA was not added. Y-Axis labels are colored according to the incorporated UAA: orange is for AzPhe, brown is for PrPhe, sky-blue is for N_3Lys and dark blue is for PrK.

3.4.4. Incorporation of One Unnatural Amino Acid in eGFP

As mentioned before in Section 3.4, by using 30 sets of OTS-UAA (*i.e.*, oRS/otRNA/UAA triplet), we decided to incorporate single unnatural amino acids in eGFP at 15 different amino acid positions. In this way 450 unique conditions were screened. Furthermore, for the 21 OTSs (*i.e.*, oRS/otRNA pairs) that were used in this screening experiment, we also wanted to characterize their non-specific amino acid incorporation. Therefore, we also performed a screening experiment for the 21 OTSs in absence of any unnatural amino acids. Figure 33 summarizes the average N.E.R. for the screening experiment when the UAA was added to the screening conditions and Figure 34 summarizes the average N.E.R. when UAA was absent.

By performing the screening experiments in the absence of UAAs, the non-specific incorporation by the different G.C.E. systems could be quantified (Figure 34). In the absence of unnatural amino acid, stop-codon suppression happens because of mischarging of the otRNA by a cellular amino acid. This mischarging could either be due to the oRS when it recognizes the cellular amino acid as its substrate, or due to the cellular aaRS when it recognizes the otRNA as its substrate. By subtracting the non-specific UAA incorporation form the screening experiments, we could derive the 'actual incorporation' of UAAs by these tested G.C.E. Systems at the different amino acid positions (Figure 35).

In Figure 35, for any given UAA, while moving along the X-axis and by comparing the N.E.R. values, we can select for the best amino acid positions for the UAA incorporation. We observed that the relative N.E.R. values for the different eGFP positions were not always the same for the different UAAs. For example, for ANAP incorporation, LYS52, GLN204, TYR151 & TYR39 were the best amino acid positions, for AzPhe incorporation, VAL176, GLN204, PHE99 & ASN212 were the best, for incorporating N_3Lys, TYR39, LYS52, THR9 & GLN204 were the best, and for 5-HTP incorporation, TYR39, LYS214, VAL176 and GLN204 were the best amino acid positions (Figure 35). For amino acid position TYR151, we observed relatively good incorporation for ANAP and AzPhe, but almost no incorporation for N_3Lys or PrK. For the buried amino acid position TYR143, no eGFP-fluorescence was observed for any of the screened conditions.

For any given amino acid position, while moving along the Y-axis (Figure 35) and comparing the N.E.R. values, we can select for the best G.C.E. system for incorporation of the required unnatural amino acid. For example, ANAP is incorporated by ${}^{ANAP}_{EcLeu}RS$ via TAA-, TAG- and TGA-codon suppression by using ${}_{EcLeu}tRNA_{UUA}$, ${}_{EcLeu}tRNA_{CUA}$ and ${}_{EcLeu}tRNA_{UCA}$ respectively. For each of the amino acid positions, ANAP incorporation by ${}_{EcLeu}tRNA_{CUA}$ is the best, followed by ${}_{EcLeu}tRNA_{UCA}$. Similarly, for ${}_{OMeY}RS$, AzPhe and PrPhe incorporation by ${}_{OMeY}tRNA_{CUA}$ is the best, followed by ${}_{OMeY}tRNA_{UCA}$. For ${}_{EcTrp(h9)}RS$ and ${}_{EcTrp(h14)}RS$, 5-HTP incorporation by ${}_{EcTrp}tRNA_{UCA}$ is the best, followed by ${}_{EcTrp}tRNA_{CUA}$. However, for ${}_{MmPyl}RS$ mediated UAA incorporation, the efficiency of both ${}_{MmPyl}tRNA_{CUA}$ and ${}_{MmPyl}tRNA_{UCA}$ is about the same (Figure 35).

For N_3Lys, which is incorporated by ${}_{MbPyl}RS$ and ${}_{MmPyl}RS$, the UAA incorporation is significantly higher for ${}_{MmPyl}RS$ for all three stop codons (as compared to ${}_{MbPyl}RS$). Between these two G.C.E. systems, the only difference is their oRS. This result demonstrates that the difference in UAA incorporation for these two systems is due to the rate of aminoacylation of otRNA by the oRS.

The observations from the endpoint fluorescence measurements (Figure 36) corroborate the observations from the image analysis (except some differences in the individual N.E.R. values). Total eGFP expression also follows the same overall trend (Figure 37).

Using such a screening experiment, it is possible to select for best conditions for UAA incorporation. For example, out of 450 screened conditions, for 22 conditions, the N.E.R. was more than the chosen cutoff of 30%. They are summarized in Figure 38.

The screening experiment for single-UAA incorporation demonstrates that UAA incorporation efficiency depends on the orthogonal aaRS (*e.g.*, ${}_{MbPyl}RS$ vs ${}_{MmPyl}RS$) and the stop-codon used for suppression (*e.g.*, TAG vs TAA). It also demonstrates that for introducing the same functional group at the required amino acid position (*e.g.*, azido group), screening different G.C.E. systems can provide a more efficient solution. A high value of non-specific incorporation for some G.C.E. systems demonstrates the importance of negative controls for assimilating G.C.E. systems. For same G.C.E. systems, different N.E.R. values for different amino

acid positions highlights the importance of screening for multiple amino acid positions for most efficient UAA incorporation.

3.4.5. Lessons for Simultaneous Incorporation of Two Different UAAs in eGFP

From the single-UAA incorporation screening experiments, poorly performing G.C.E. systems and eGFP-positions could be identified and excluded from further studies and two-site UAA incorporation experiments. For example, due to the low overall efficiency of ${}_{MbPyl}RS$ and ${}^{DanAla}_{EcLeu}RS$ based G.C.E. systems, they were not considered for further experiments. Additionally, due to the low overall efficiency of $tRNA_{UUA}$ based systems, TAA stop-codon suppression was not considered for two-site UAA incorporation. ${}_{EcTrp(h9)}RS$ and ${}_{EcTrp(h14)}RS$ based G.C.E. systems were also excluded from further studies because of relatively high non-specific incorporation and because click-chemistry compatible UAAs (*i.e.*, AzW and PrW) for this system could not be obtained.

From the single UAA incorporation screening experiments, we observed that no amino acid position was 'the obvious best' for UAA incorporation. However, between the different G.C.E. Systems, the same amino acid positions were often the ones with one of the highest N.E.R. values. We concluded that these were the amino acid positions that could better tolerate the incorporation of a variety of UAAs. While this is not always correct, as exemplified by amino acid position TYR151, this generalization helped to select for the amino acid positions for double-UAA incorporation experiments.

For incorporating a pair of unnatural amino acids in eGFP, the amino acids positions TYR39, LYS52, PHE99, VAL176 and GLN204 were selected for the two-site UAA incorporation experiments, because they showed relatively high UAA incorporation for multiple G.C.E. Systems. For the shortlisted G.C.E. systems, the N.E.R. values for TAG stop-codon suppression were higher than that of TGA stop codon suppression. For this reason, it made sense to introduce the TGA stop codon downstream to the TAG stop codon, so that the first amino acid position for UAA incorporation is not the rate-limiting one.

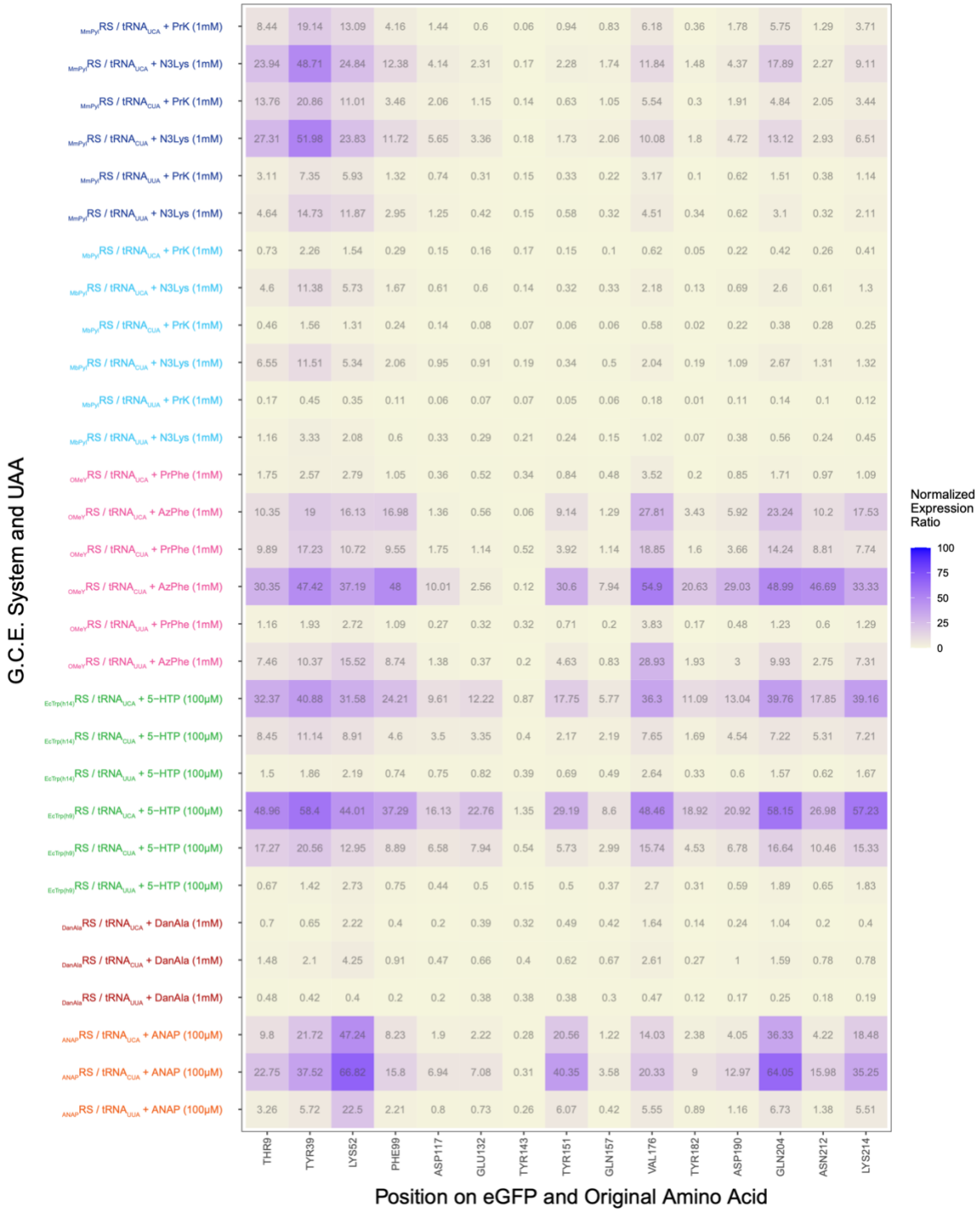


Figure 33: One-Site UAA Incorporation - N.E.R. Measured by Image Analysis.
Y-Axis details the G.C.E. systems (oRS, otRNA-anticodon and UAA with its concentration). X-Axis represents amino acid positions where the stop-codon is present for UAA incorporation. For any given eGFP position, the stop-codon along the Y-Axis is complementary to the otRNA anticodon. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the oRS present in the G.C.E. system - orange is for ANAP-RS, brown is for DanAla-RS, green is for EcTrp-RS (both h9 and h14 variants), pink is for OMeY-RS, sky-blue is for MbPyl-RS and dark blue is for MmPyl-RS.

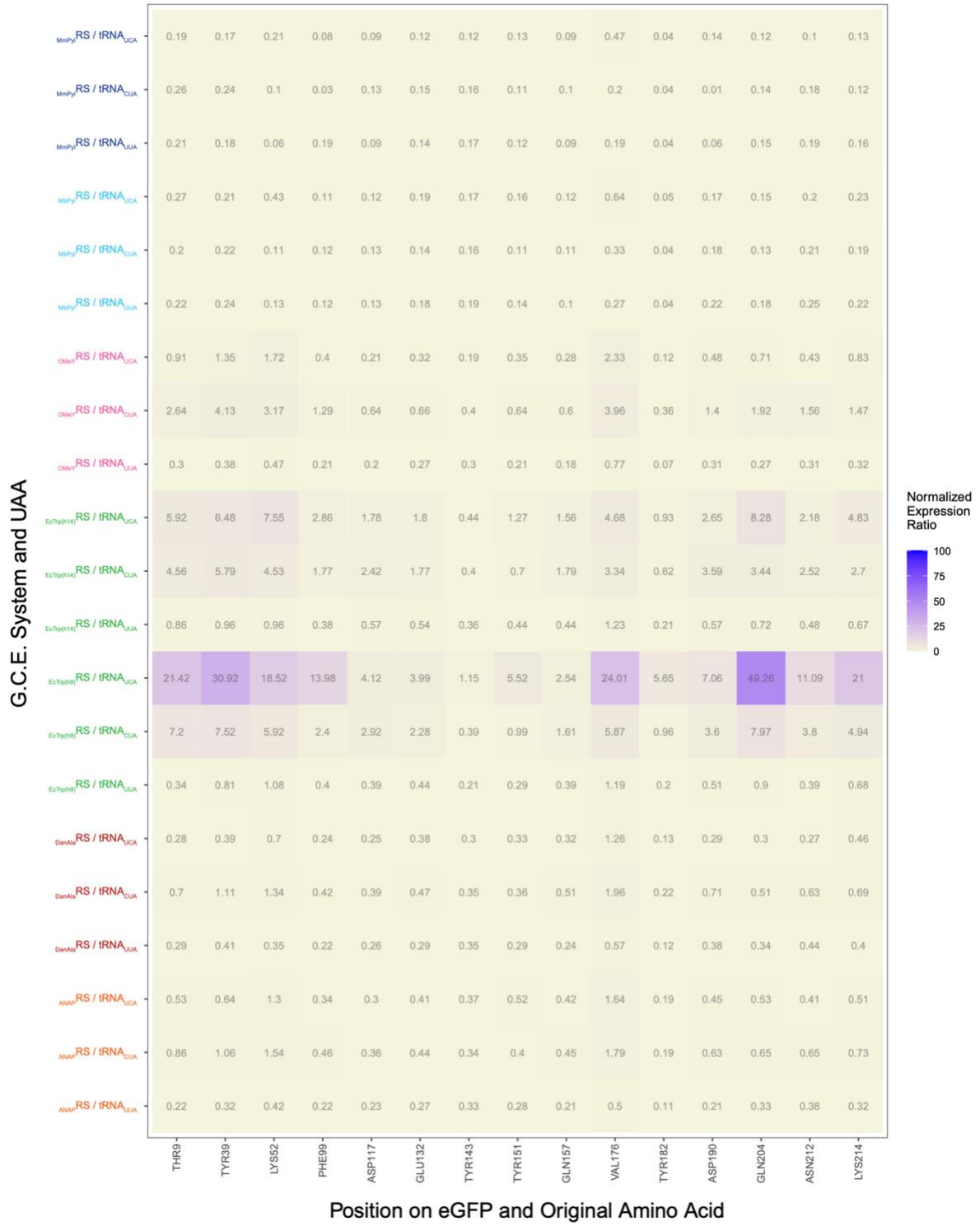


Figure 34: Non-Specific UAA Incorporation by G.C.E. Machinery.
Y-Axis details the G.C.E. Machinery (oRS and otRNA-anticodon). X-Axis represents amino acid positions where the stop-codon is present. For any given eGFP position, the stop-codon along the Y-Axis is complementary to the otRNA anticodon. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the oRS present in the G.C.E. system - orange is for ANAP-RS, brown is for DanAla-RS, green is for EcTrp-RS (both h9 and h14 variants), pink is for OMeY-RS, sky-blue is for MbPyl-RS and dark blue is for MmPyl-RS.

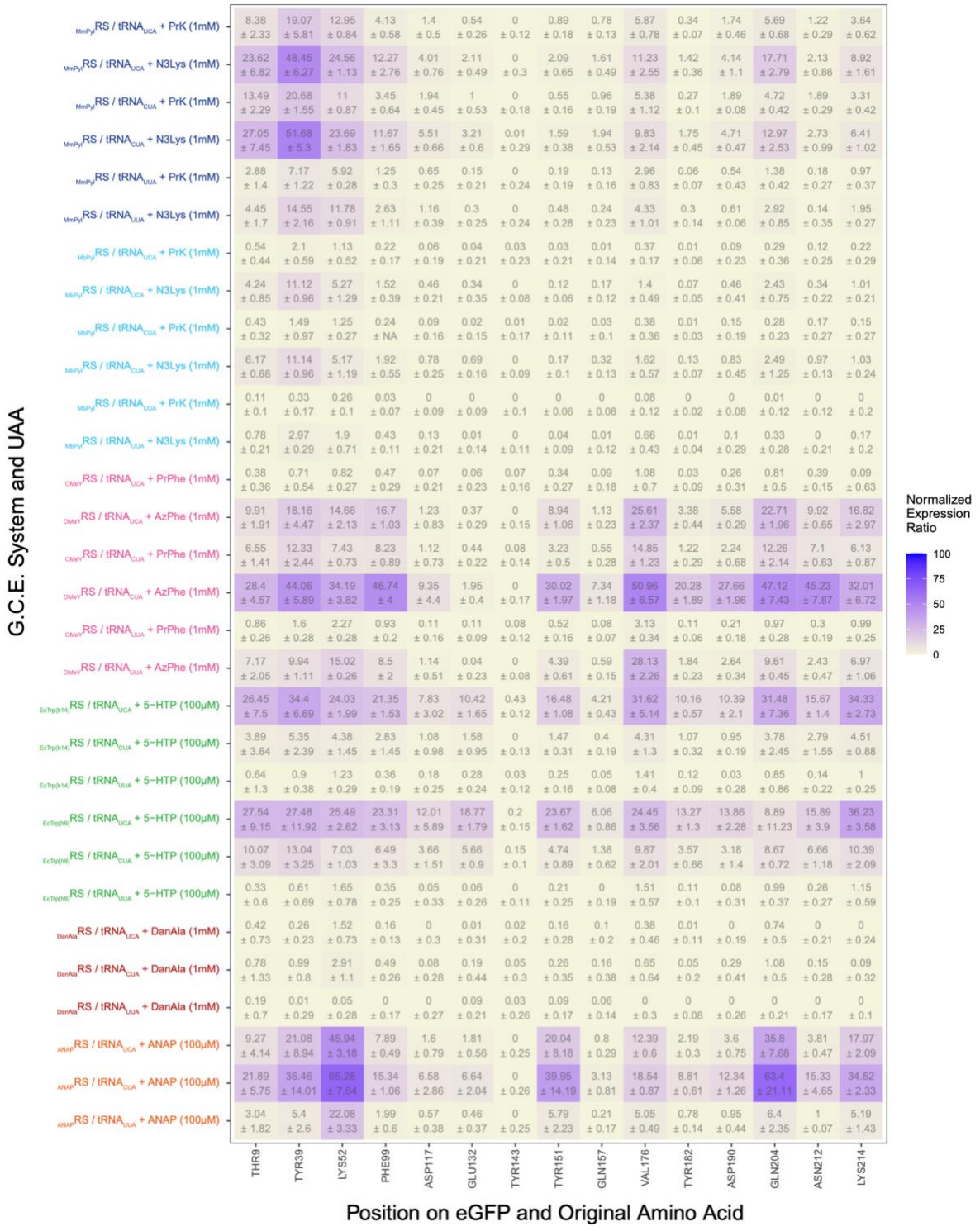


Figure 35: One-Site UAA Incorporation - N.E.R. Values Corrected for Non-Specific Amino Acid Incorporation.
Y-Axis details the G.C.E. systems and X-Axis represents amino acid positions where the stop-codon is present for UAA incorporation. For any given eGFP position, the stop-codon along the Y-Axis is complementary to the otRNA anticodon. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the oRS present in the G.C.E. system - orange is for ANAP-RS, brown is for DanAla-RS, green is for EcTrp-RS (both h9 and h14 variants), pink is for OMeY-RS, sky-blue is for MbPyl-RS and dark blue is for MmPyl-RS.

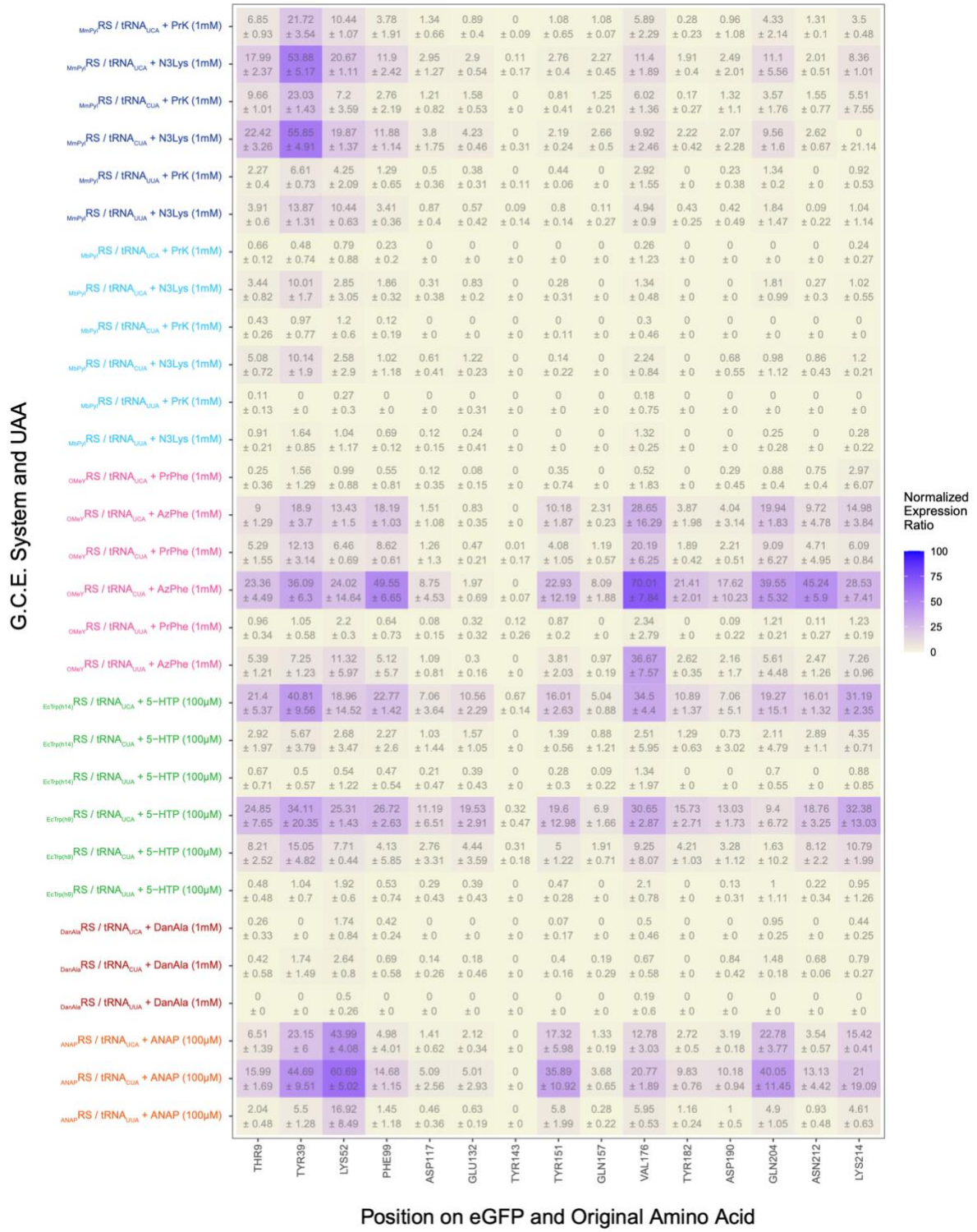


Figure 36: One-Site UAA Incorporation - N.E.R. Measured by Endpoint Fluorescence Analysis.
Y-Axis details the G.C.E. systems and X-Axis represents amino acid positions where the stop-codon is present for UAA incorporation. For any given eGFP position, the stop-codon along the Y-Axis is complementary to the otRNA anticodon. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the oRS present in the G.C.E. system - orange is for ANAP-RS, brown is for DanAla-RS, green is for EcTrp-RS (both h9 and h14 variants), pink is for OMeY-RS, sky-blue is for MbPyl-RS and dark blue is for MmPyl-RS.

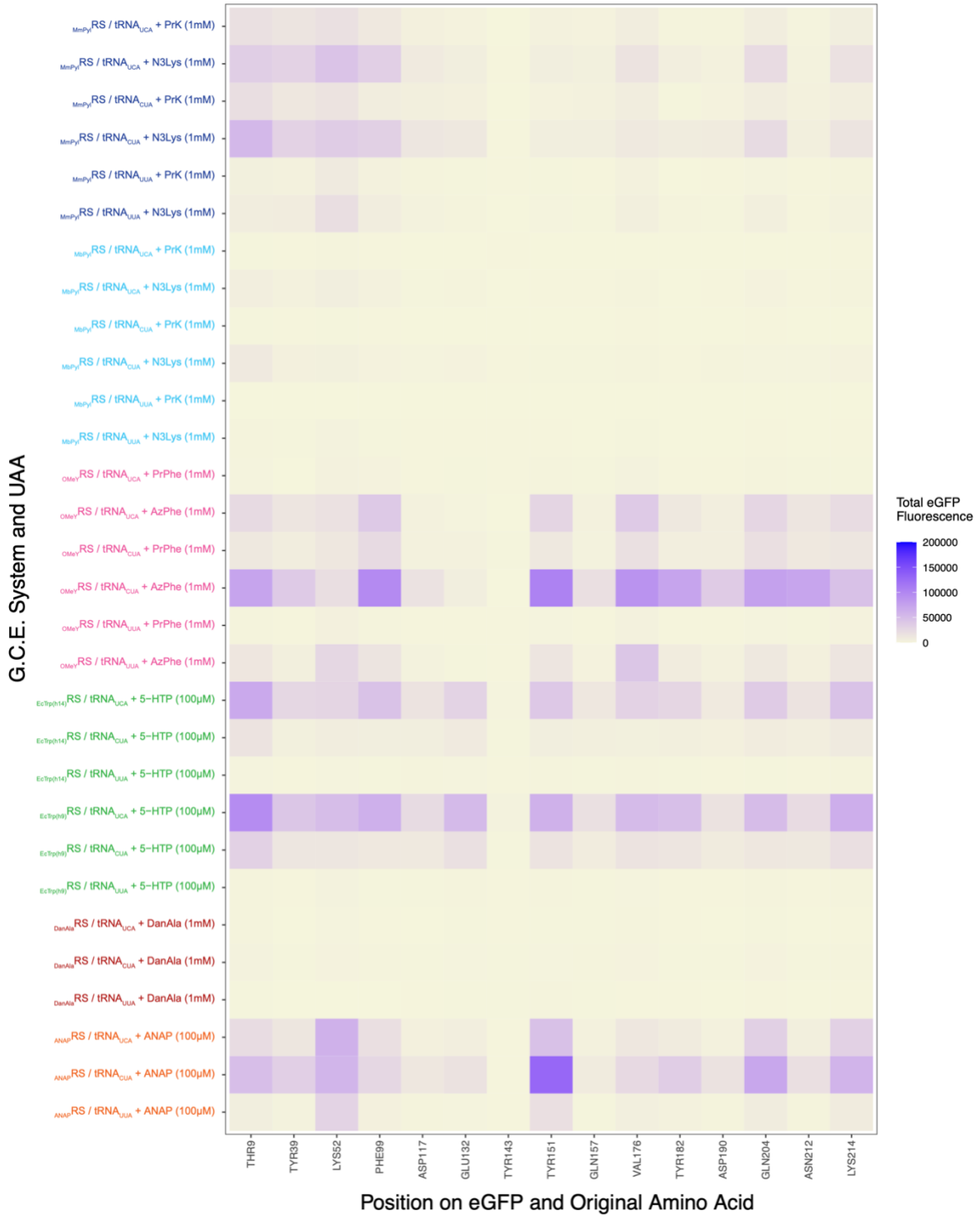


Figure 37: One-Site UAA Incorporation - Total eGFP Fluorescence (calculated from Image Analysis).
Y-Axis details the G.C.E. systems and X-Axis represents amino acid positions where the stop-codon is present for UAA incorporation. For any given eGFP position, the stop-codon along the Y-Axis is complementary to the otRNA anticodon. Colors of the boxes represent the mean values of total eGFP fluorescence for the six replicates of the condition, according to the adjoining scale. Y-Axis labels are colored according to the oRS present in the G.C.E. system - orange is for ANAP-RS, brown is for DanAla-RS, green is for EcTrp-RS (both h9 and h14 variants), pink is for OMeY-RS, sky-blue is for MbPyl-RS and dark blue is for MmPyl-RS.

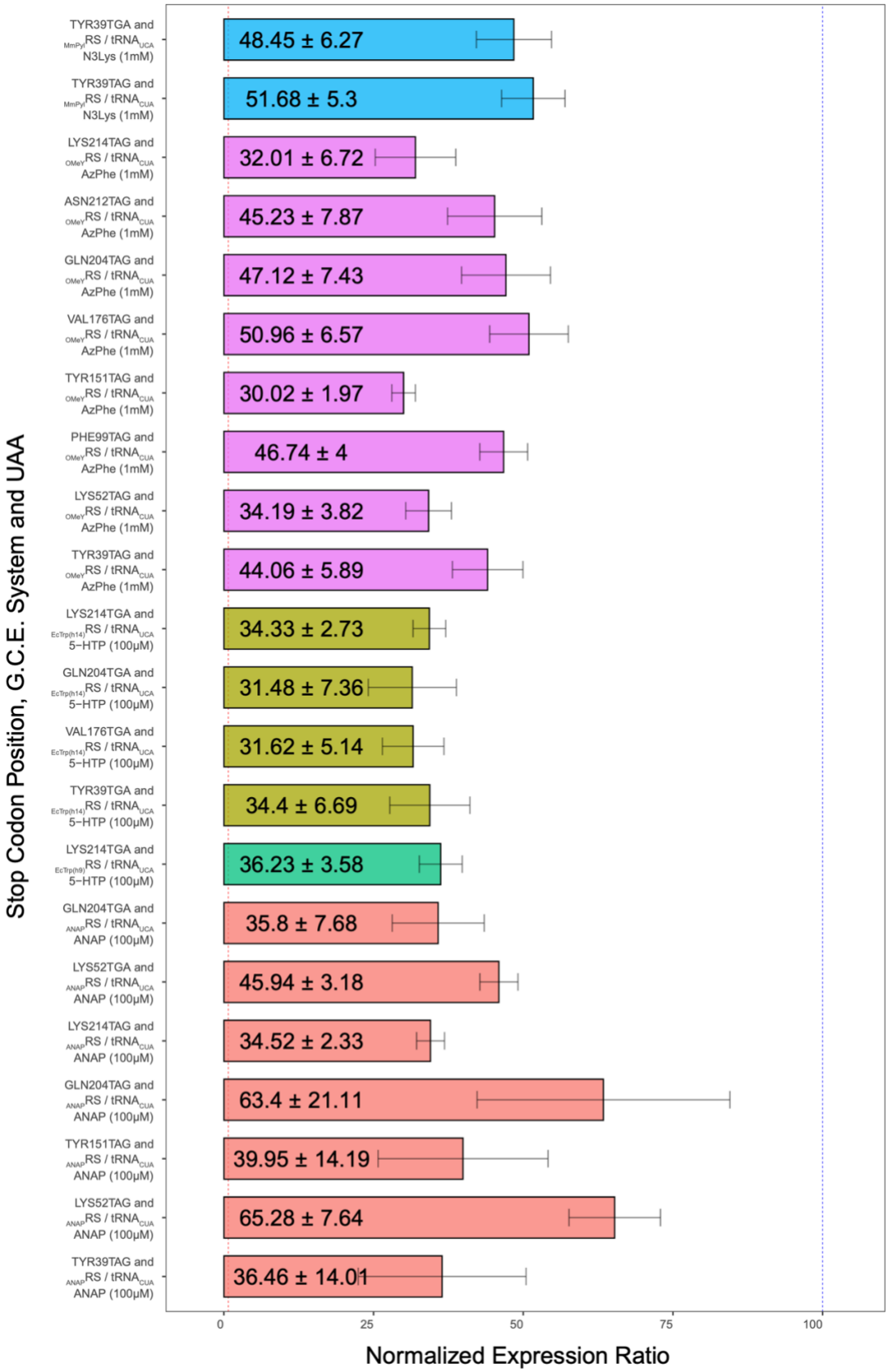


Figure 38: One-Site UAA Incorporation - Screening Conditions with N.E.R. Cutoff 30%.
Y-Axis details the screening condition (eGFP Mutation, G.C.E. system and UAA & its concentration). X-Axis represents the linear scale for the N.E.R. Length of the bar represents the mean value and the whiskers represent the standard error. Numerical values represent the value of the average N.E.R. for the given condition. Red dashed line represents the N.E.R. for the control-condition from transfection with *Plasmid MX01.4* and blue line represents the N.E.R. for the control condition from transfection with *Plasmid UTX0*. Bar plots are grouped and colored according to the oRS present in the G.C.E. system - red is for ANAP-RS, green is for EcTrp(h9)-RS, olive is for EcTrp(h14)-RS, pink is for OMeY-RS and blue is for MmPyl-RS.

3.4.6. Crosstalk Between G.C.E. Systems

For UAA incorporation at two different positions, two mutually orthogonal G.C.E. systems must be used. This means that the two G.C.E. systems must meet three essential criteria: 1) the orthogonal aaRS/tRNA pair of either G.C.E. systems must not recognize the unnatural amino acid for the other G.C.E. system (*i.e.*, no UAA-crosstalk), 2) oRS for one G.C.E. system must not charge the otRNA for the other system (*i.e.*, no otRNA-crosstalk), and 3) otRNA for one G.C.E. system should recognize only its own stop-codon signal for the UAA incorporation (*i.e.*, no stop-codon-crosstalk).

After screening for one UAA incorporation at different amino acid positions in eGFP, we decided study the mutual orthogonality of the G.C.E. Systems. For this, stop-codon-crosstalk (Figure 39), and UAA-crosstalk (Figure 40) between the G.C.E. systems was studied. The otRNA-crosstalk could not be studied using our plasmids because the oRS and otRNA are present on the same plasmid vector.

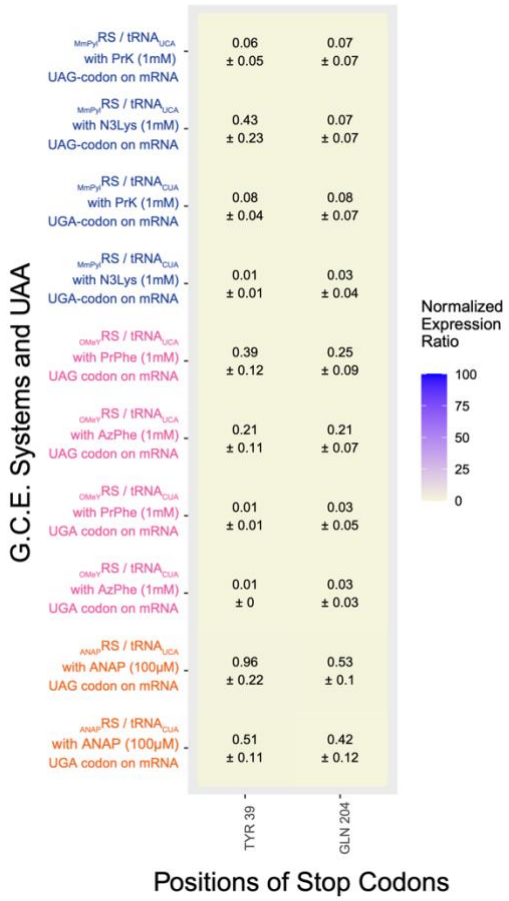


Figure 39: Stop-Codon Crosstalk for G.C.E. Systems.
Y-Axis details the G.C.E. systems. X-Axis represents amino acid positions where the stop-codon is present for UAA incorporation. For the two mentioned positions, the eGFP contains TAG stop-codon when tRNA-UCA is present and TGA stop codon when tRNA-CUA is present. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the oRS present in the G.C.E. system - orange is for ANAP-RS, pink is for OMeY-RS and dark blue is for MmPyl-RS.

3.4.6.1. Stop-Codon Crosstalk

For two amino acid positions on eGFP (TYR39 and GLN204), stop-codon crosstalk was studied for the shortlisted ${}^{ANAP}_{EcLeu}RS$, ${}^{OMeY}_{OMeY}RS$ and ${}_{MmPyl}RS$ based G.C.E. systems. For this, TAG stop-codon suppression was tested with cognate $tRNA_{UCA}$ and TGA stop-codon suppression was tested with cognate $tRNA_{CUA}$. Figure 39 summarizes the crosstalk between the G.C.E. systems due to misidentification of stop-codon by the

otRNA. In general, all N.E.R. values were less than 1% and were comparable to the average background N.E.R. (obtained from transfection with *Plasmid MX01.4* [*]) for this experiment, which was 0.28%.

For $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{UCA}$ pair, we observed a maximum N.E.R. value of 0.96 ± 0.22 for eGFP^{39TAG}. In contrast, for eGFP^{39TGA}, the N.E.R. value for this oRS/otRNA pair was 21.08 ± 8.94 (Figure 35). In other words, when both TAG and TGA stop-codons are present on an mRNA, the G.C.E. system $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{UCA}$ is about 30 times more likely to incorporate unnatural amino acid via the TGA-suppression, as compared to TAG-suppression. Hence, this result demonstrates that UAA incorporation due to mismatch between the stop-codon and otRNA of G.C.E. system is possible but is unlikely. For other oRS/otRNA pairs, the N.E.R. was either very close to the average background N.E.R., or was less than it, so it was considered as insignificant.

3.4.6.2. Unnatural Amino Acid Crosstalk

For two amino acid positions on eGFP (TYR39 and GLN204), misidentification and misincorporation of unnatural amino acids by $^{ANAP}_{EcLeu}RS$, $^{OMeY}_{OMeY}RS$ and $_{MmPyl}RS$ based G.C.E. systems was studied using TAG and TGA stop-codon suppression. The results of this experiment are summarized in Figure 40 A. From the individual values in Figure 40 A, the nonspecific incorporation value obtained from Figure 34 were subtracted to generate 'background corrected UAA-crosstalk' Figure 40 B. For Figure 40 B, the negative values were set to 0. The average background N.E.R. for this experiment was 0.28% (for Figure 40 A).

For $^{ANAP}_{EcLeu}RS$ based G.C.E. systems, incorporation of AzPhe, PrPhe, N_3Lys and PrK was tested. By using the $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{CUA}$ pair for eGFP^{TYR39TAG}, nonspecific-incorporation-corrected N.E.R. value of 0.3%, 5.02%, 2.1% and 1.69% were obtained for these unnatural amino acids respectively. For the same oRS/otRNA pair and eGFP^{GLN204TAG}, these values were 0%, 0.14%, 0% and 0% respectively. By using the $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA_{UCA}$ pair for eGFP^{TYR39TGA}, nonspecific-incorporation-corrected N.E.R. value of 0%, 1.21%, 0.37% and 0% were obtained for AzPhe, PrPhe, N_3Lys and PrK respectively. For the same oRS/otRNA pair and eGFP^{GLN204TGA}, these values were 0% for all amino acids.

These results demonstrate that there is at-least some crosstalk between $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA$ pair and N_3Lys & PrK amino acids. Amino acid position 39 on eGFP showed highest nonspecific-incorporation-corrected N.E.R. value (5.02%) for this oRS/otRNA pair. However, ANAP specific incorporation for this oRS/otRNA pair is 36.46% and 21.08% for TAG- and TGA-suppression respectively (Figure 41 and Figure 35). This shows that when both ANAP and N_3Lys unnatural amino acids are present, ANAP incorporation is at least 7 times more likely to occur this oRS/otRNA pair, as compared to N_3Lys incorporation.

For $_{OMeY}RS$ based G.C.E. systems, incorporation of ANAP, N_3Lys and PrK was tested. By using the $_{OMeY}RS$ / $_{OMeY}tRNA_{CUA}$ pair for eGFP^{TYR39TAG}, nonspecific-incorporation-corrected N.E.R. value of 1.04%, 0.18% and 0.93% were obtained for these unnatural amino acids respectively. For the same oRS/otRNA pair and eGFP^{GLN204TAG}, these values were 0.7%, 0% and 0.11% respectively. By using the $_{OMeY}RS$ / $_{OMeY}tRNA_{UCA}$ pair for eGFP^{TYR39TGA}, nonspecific-incorporation-corrected N.E.R. value of 2.03%, 0% and 0% were obtained for ANAP, N_3Lys and PrK unnatural amino acids respectively. For the same oRS/otRNA pair and eGFP^{GLN204TAG}, these values were 0.16%, 0% and 0% respectively.

These results demonstrate that there is at-least some crosstalk between $_{OMeY}RS$ / $_{OMeY}tRNA$ pair and ANAP amino acid. Amino acid position 39 on eGFP showed highest nonspecific-incorporation-corrected N.E.R. value (2.03%) for this oRS/otRNA pair. However, AzPhe specific incorporation for this oRS/otRNA pair is 44.06% and 18.16% for TAG- and TGA-suppression respectively (Figure 41 and Figure 35). This shows that when both AzPhe and ANAP unnatural amino acids are present, AzPhe incorporation is at least 9 times more likely to occur for this oRS/otRNA pair, as compared to ANAP incorporation. PrPhe specific incorporation for this oRS/otRNA pair was 0.71% for TGA-suppression (Figure 41 and Figure 35). This shows that ANAP can compete with PrPhe for incorporation in proteins.

[*] While ideally, the N.E.R. value of the 'no-eGFP' conditions should be zero, due to the noise in the imaging system, this value is very close to zero, but seldom equal to zero.

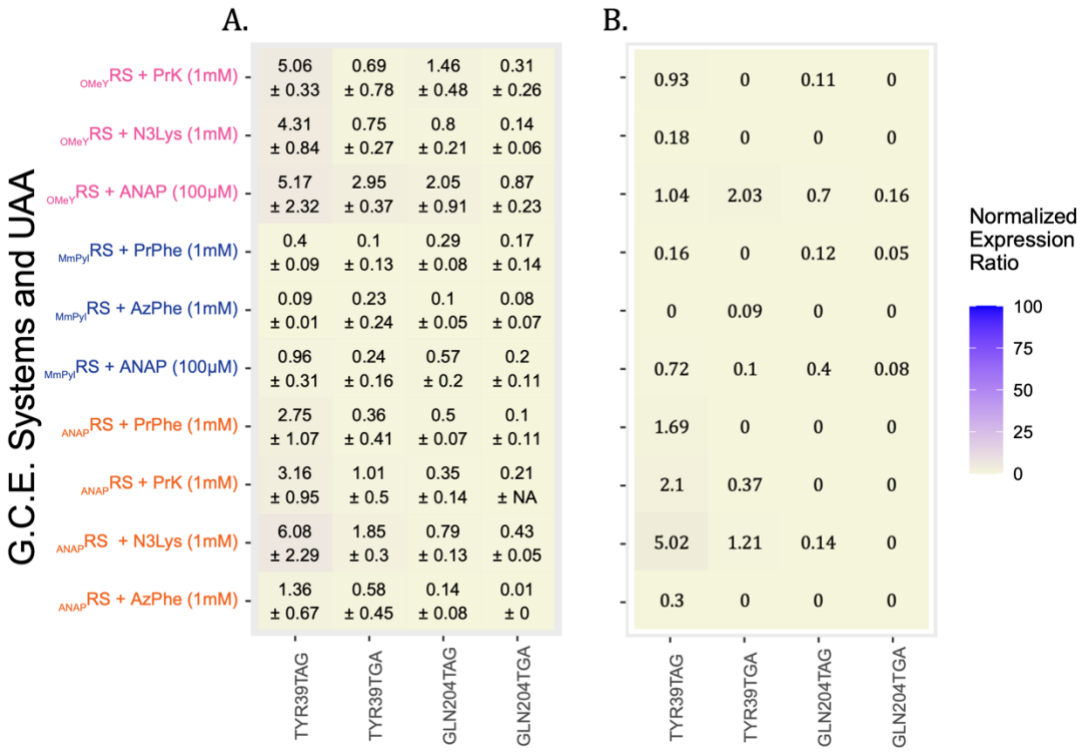


Figure 40: UAA Crosstalk Between G.C.E. Systems.
A. UAA crosstalk between G.C.E. systems.
B. UAA crosstalk after subtracting non-specific amino acid incorporation.
This plot was generated in MS-Excel. Negative values were set to zero. Y-Axis details the G.C.E. systems. X-Axis represents amino acid positions and the stop-codon present for UAA incorporation. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the oRS present in the G.C.E. system - orange is for ANAP-RS, pink is for OMeY-RS and dark blue is for MmPyl-RS.

For $_{MmPyl}RS$ based G.C.E. systems, incorporation of ANAP, AzPhe and PrPhe was tested. By using the $_{MmPyl}RS$ / $_{MmPyl}tRNA_{CUA}$ pair for eGFP^{TYR39TAG}, nonspecific-incorporation-corrected N.E.R. value of 0.72%, 0% and 0.16% were obtained for these unnatural amino acids respectively. For eGFP^{GLN204TAG}, these values were 0.4%, 0% and 0.12% respectively. By using the $_{MmPyl}RS$ / $_{MmPyl}tRNA_{UCA}$ pair for eGFP^{TYR39TGA}, nonspecific-incorporation-corrected N.E.R. value of 0.1%, 0.09% and 0% were obtained for ANAP, AzPhe and PrPhe respectively. For eGFP^{GLN204TAG}, these values were 0.08%, 0% and 0.05% respectively. Since these values are negligible as compared to PrK and N_3Lys incorporation (Figure 41 and Figure 35), we conclude that $_{MmPyl}RS$ / $_{MmPyl}tRNA$ pair does not suffer from crosstalk with the tested amino acids.

To summarize the results of crosstalk screening experiments, from the tRNA crosstalk screening, we learn that TAG and TGA stop codons are correctly identified by their respective otRNAs. UAA crosstalk screening experiments demonstrate that some crosstalk exists between $^{ANAP}_{EcLeu}RS$ / $_{EcLeu}tRNA$ pair and N_3Lys and PrK unnatural amino acids, and between $_{OMeY}RS$ / $_{OMeY}tRNA$ pair and ANAP amino acid. However, from the cognate-amino-acid-specific incorporation data (Figure 41), we conclude that in presence of both cognate amino acids and misidentified amino acids, it is more likely that the cognate amino acid would be incorporated into proteins.

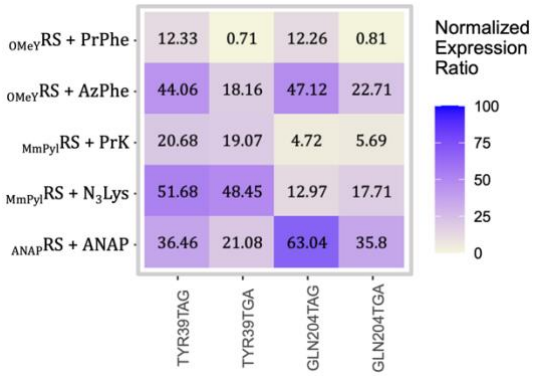


Figure 41: Background-Corrected UAA Incorporation for Selected Conditions.
This graph was created in MS-Excel to zoom on the selected conditions. Y-Axis details the oRS and UAA. X-Axis represents amino acid positions and the stop-codon present for UAA incorporation. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition.

3.4.7. Incorporation of Two Different Unnatural Amino Acids in eGFP

For incorporating two different UAA in eGFP, five amino acid positions were shortlisted (TYR39, LYS52, PHE99, VAL176 and GLN204, as explained in Section 3.4.5). Since TAG and TGA stop codon suppression demonstrated the highest N.E.R. values for one UAA incorporation screening experiments, these two stop-codons were chosen for simultaneously incorporating two UAAs in proteins. By introducing TAG and TGA stop codons in different combinations of these five amino acid positions, 10 different variants of *Plasmid UTX0* were created by mutagenesis, each containing two stop-codons in the eGFP gene (these plasmids are listed in the Methods Section 2.11.7 in Table 12). These plasmids too are collectively referred to as G.E.M.S. Plasmids.

Six OTS with relatively high N.E.R. values and a potential to incorporate either fluorescent UAA or bioorthogonal reaction handles in proteins, were also shortlisted. These were based on $^{ANAP}_{EcLeu}RS$, $_{OMeY}RS$ & $_{MmPyl}RS$, with their TAG- & TGA-suppressor otRNAs. By using these orthogonal aaRS/tRNA pairs, we decided to screen for the incorporation of six pairs of UAAs in eGFP. These UAA-pairs were ANAP-AzPhe, ANAP-PrPhe, ANAP-N_3Lys, ANAP-PrK, AzPhe-PrK and N_3Lys-PrPhe. ANAP was selected as a candidate for screening because it is a fluorescent unnatural amino acid. AzPhe, PrPhe, N_3Lys and PrK were selected as they have bioorthogonal reaction handles for click-chemistry and could facilitate introduction of fluorophores at a later stage. For chemically introducing two different fluorophores on proteins, two distinct functional groups are required. Therefore, to ensure chemical orthogonality between the incorporated UAAs, AzPhe was paired with PrK and PrPhe was paired with N_3Lys. For each pair of unnatural amino acids, the individual UAA could either be incorporated using TAG-suppressing OTS or by using TGA-suppressing OTS. In this way, for incorporation of these six pairs of unnatural amino acids, 12 combinations of mutually orthogonal OTS pairs are needed (these are listed in the Methods Section 2.11.7 in Table 13).

Therefore, for each of the 10 variants of *Plasmid UTX0*, we decided to incorporate a pair of unnatural amino acids in the eGFP by using the 12 combinations of mutually orthogonal OTS pairs. In this way, we decided to screen 120 conditions for two-UAA incorporation in eGFP.

Before going further with this screening experiment, we were curious to know if we could predict the UAA incorporation efficiency of the OTS pairs. For this, we assumed that for both the stop-codons, the UAA incorporation is independent of each other. Since stop-codon suppression depends on the context of the stop codon on the mRNA, this assumption is not entirely correct. However, we argued that if the two stop codons are far away from each other, this could be a safe assumption. According to this assumption, each individual stop-codon suppression has a certain probability of succeeding and producing the full-length protein. Therefore, to get the probability of full-length protein production after two stop-codon suppressions, the individual probabilities can be multiplied. From Figure 33, we know the probability of stop-codon suppression for each amino acid position on eGFP and for each of the G.C.E. system. Therefore, to calculate the probability of full-length eGFP production after stop-codon suppression at two different

amino acid positions, we multiplied the probabilities of individual 'stop-codon suppression' at these positions. By doing so, we created a prediction table, represented in Figure 42.

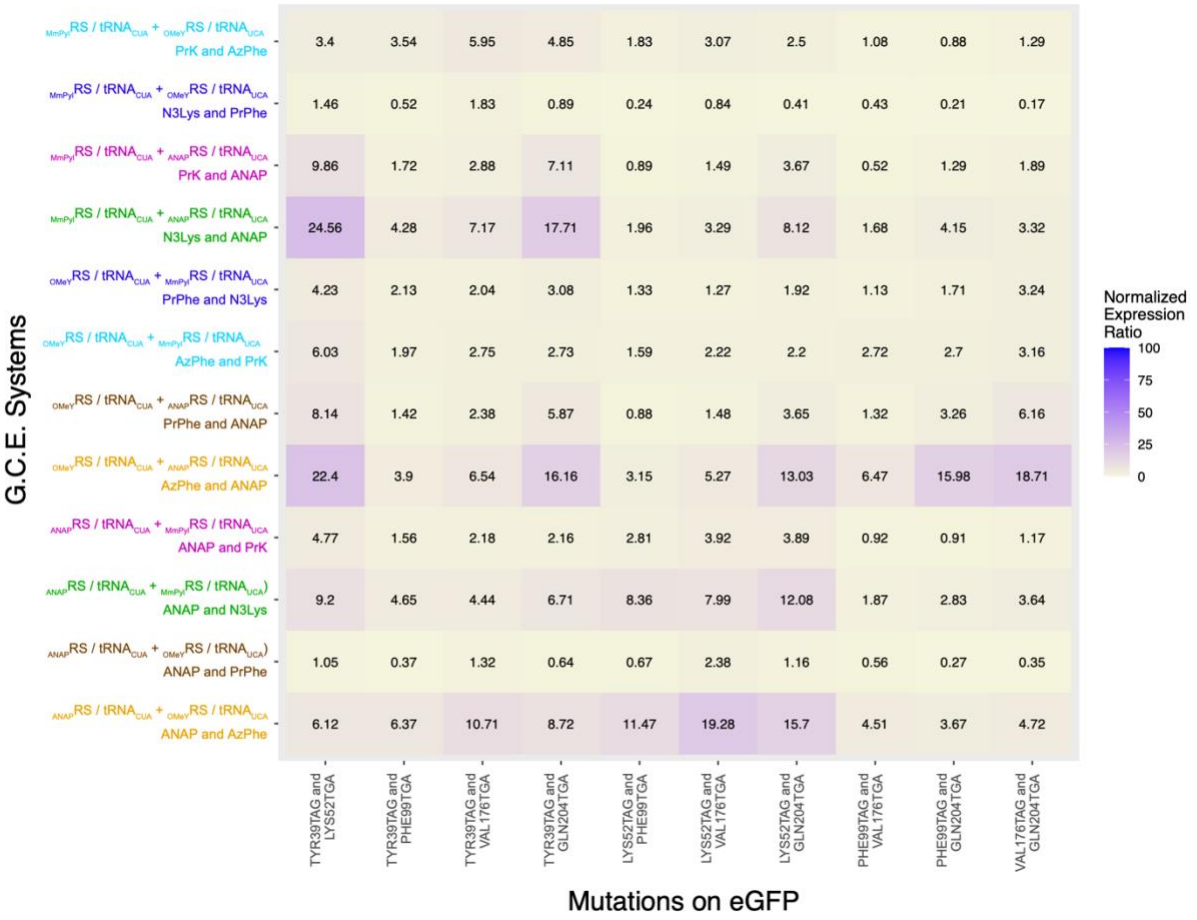


Figure 42: Prediction for Two-Site UAA Incorporation.
Y-Axis details the two G.C.E. systems with corresponding unnatural amino acids. X-Axis represents amino acid positions and the stop codons used. Colors of the boxes represent the predicted N.E.R. value according to the adjoining scale. Numerical values represent the predicted N.E.R. for the given condition. Y-Axis labels are colored according to the UAA pair present in the medium: orange is for ANAP-AzPhe, Brown is for ANAP-PrPhe, green is for ANAP-N_3Lys, pink is for ANAP-PrK, sky-blue is for AzPhe-PrK and dark blue is for PrPhe-N_3Lys. For each pair of UAA, the first UAA is introduced via TAG stop codon suppression and the second one is introduced via TGA stop codon suppression.

Afterwards, we went ahead with the screening experiment for two-UAA incorporation. For this screening experiment, we have slightly modified our screening assay protocol. In this screening experiment, we need to co-transfect the cells with three plasmids, *i.e.*, one G.E.M.S. Plasmid (with TAG & TGA stop codons in the eGFP gene) and two G.C.E. Plasmids (having mutually orthogonal OTS). These three plasmids are present in the ratio of 2:1:1 respectively. The concentration of the two unnatural amino acids in the protein expression media was not changed.

Figure 43 summarizes the results of this screening experiment. The N.E.R. values for two-site UAA incorporation were significantly lower than those for the one site screening experiment. This was expected because for incorporation of two UAAs, the UAA incorporation machinery has to compete twice with the native machinery.

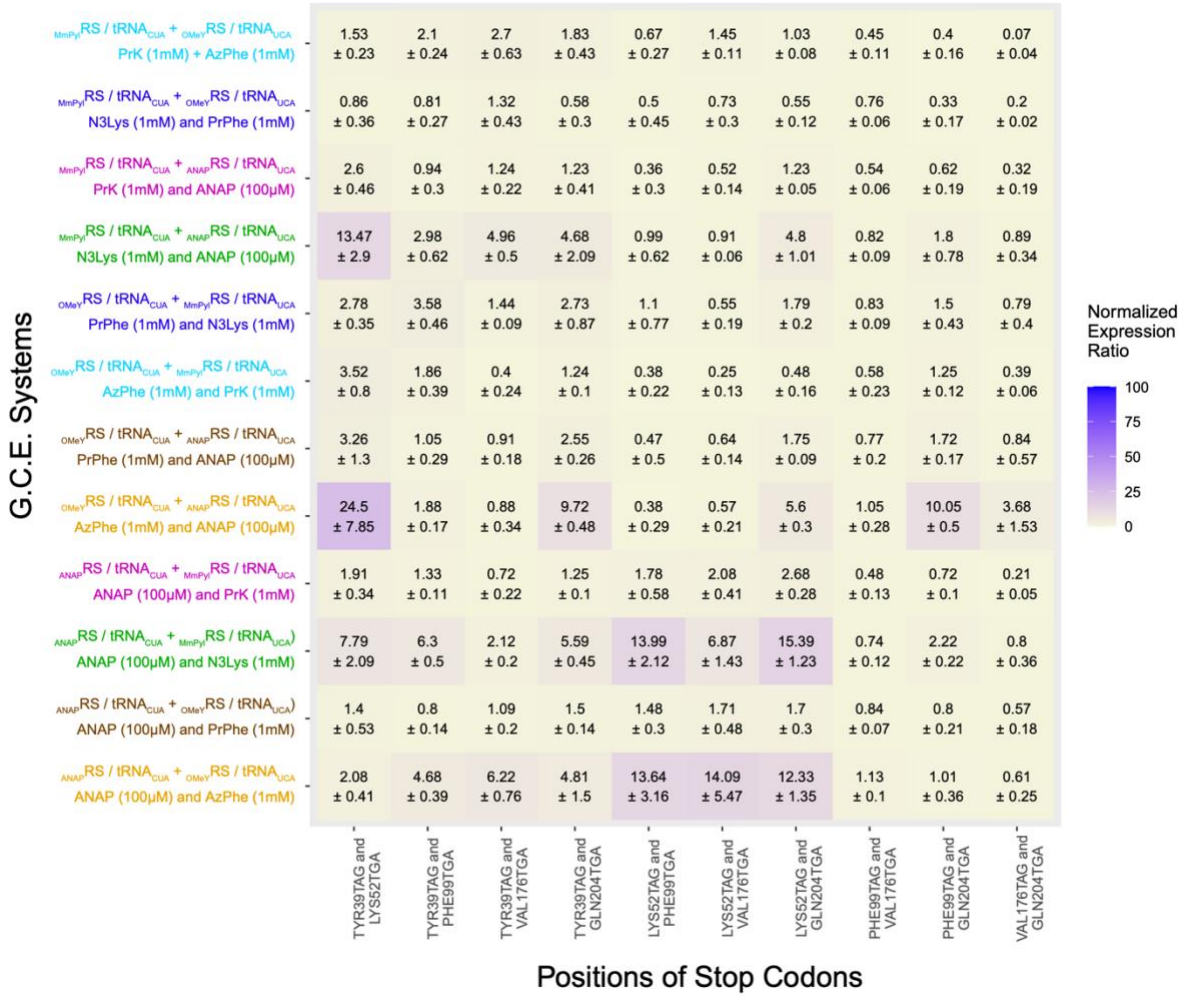


Figure 43: Two-Site UAA Incorporation - N.E.R. Measured by Image Analysis.
Y-Axis details the two G.C.E. systems (both oRS, both otRNA with their anticodons and both UAAs with their concentrations). X-Axis represents amino acid positions and the stop codons used. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the UAA pair present in the medium: orange is for ANAP-AzPhe, Brown is for ANAP-PrPhe, green is for ANAP-N_3Lys, pink is for ANAP-PrK, sky-blue is for AzPhe-PrK and dark blue is for PrPhe-N_3Lys. For each pair of UAA, the first UAA is introduced via TAG stop codon suppression and the second one is introduced via TGA stop codon suppression.

Three of the five most efficient conditions and seven of the ten most efficient conditions were predicted by our model (Figure 42 and Figure 43). However, the actual values for most of these conditions were different from our predictions, possibly because UAA incorporation at two positions is more complex than our simple assumptions. Some of the possible reasons for deviation from the prediction could be the context of the two stop-codons on the mRNA, the interaction between G.C.E. systems or the disruption of protein folding due to the two incorporated UAAs. Despite the differences in the predicted and measured values, this strategy may be helpful in shortlisting the candidates for a more dedicated and detailed screening.

For nine of the screened conditions, the N.E.R. values were either close to, or greater than the chosen cutoff of 10% (Figure 46). Each one of these conditions had ANAP as one of the UAAs. The other UAA was either AzPhe (six conditions) or N_3Lys (three conditions). For five conditions, ANAP was incorporated by TAG stop-codon suppression.

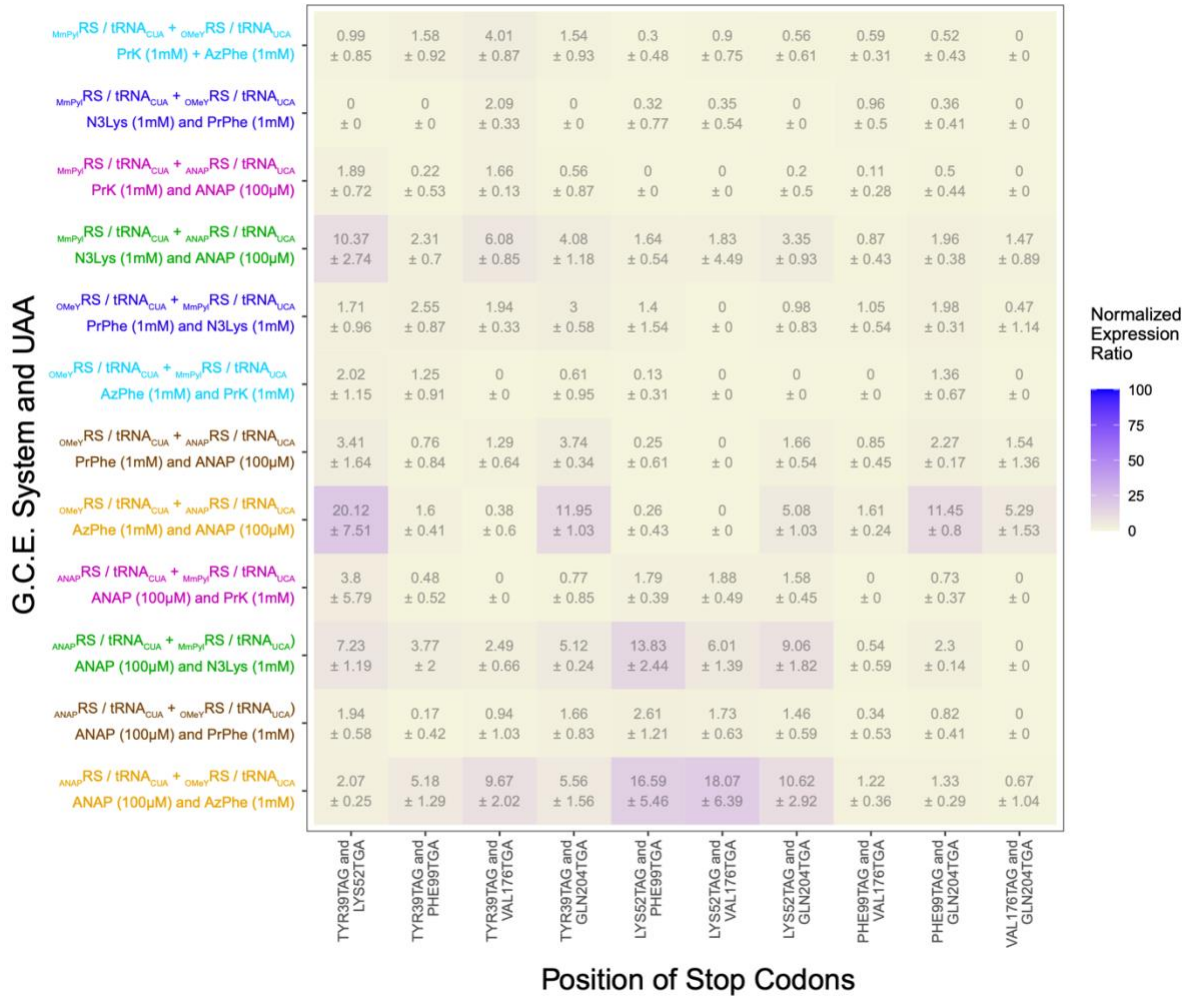


Figure 44: Two-Site UAA Incorporation - N.E.R. Measured by Endpoint Fluorescence Analysis.
Y-Axis details the two G.C.E. systems and X-Axis represents amino acid positions and the stop codons used. Colors of the boxes represent the mean N.E.R. value for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the average N.E.R. for the given condition. Y-Axis labels are colored according to the UAA pair present in the medium: orange is for ANAP-AzPhe, Brown is for ANAP-PrPhe, green is for ANAP-N_3Lys, pink is for ANAP-PrK, sky-blue is for AzPhe-PrK and dark blue is for PrPhe-N_3Lys. For each pair of UAA, the first UAA is introduced via TAG stop codon suppression and the second one is introduced via TGA stop codon suppression.

When two mutually orthogonal OTS are present in the screening conditions, non-intended amino acid incorporation may happen because of mis-suppression of a stop-codon by the incorrect otRNA (*i.e.*, stop-codon crosstalk), aminoacylation of either of the otRNA by a cellular amino acid (*i.e.*, due to non-specific amino acid incorporation by the G.C.E. system) or by aminoacylation of the otRNA by incorrect UAA (*i.e.*, UAA crosstalk or otRNA crosstalk). In Section 3.4.6.1 we have established that stop-codon crosstalk between TAG- and TGA-suppressing otRNA is negligible, therefore when two mutually orthogonal OTS are present, the stop-codons are suppressed by the correct otRNA. In Section 3.4.6.2, the N.E.R. values measured for the UAA crosstalk (Figure 40 A) have two components - non-specific incorporation of cellular amino acids by the G.C.E. system and incorporation of the incorrect UAA by the G.C.E. system. Therefore, these values can be used for quantifying the 'total non-intended amino acid incorporation' (due to the cellular amino acid or due to the incorrect UAA) by a G.C.E. system in presence of another G.C.E. system. As mentioned in Section 3.4.6, we did not screen for otRNA crosstalk as both oRS and otRNA are present on the same plasmid.

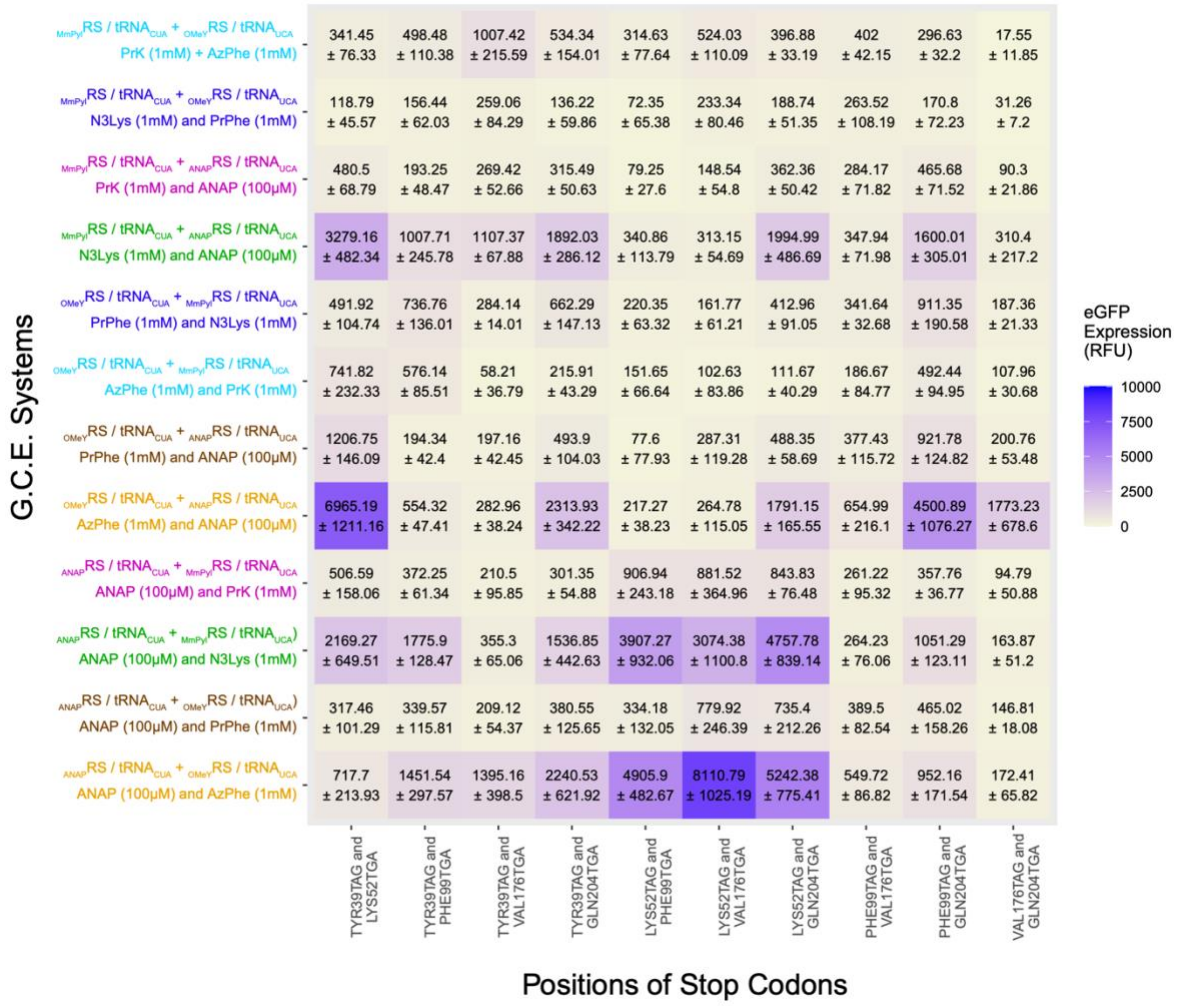


Figure 45: Two-Site UAA Incorporation - Total eGFP Fluorescence (Calculated by Image Analysis).
Y-Axis details the two G.C.E. systems and X-Axis represents amino acid positions and the stop codons used. Colors of the boxes represent the mean value of total eGFP-fluorescence for the six replicates of the condition, according to the adjoining scale. Numerical values represent the value of the eGFP-fluorescence for the given condition. Y-Axis labels are colored according to the UAA pair present in the medium: orange is for ANAP-AzPhe, Brown is for ANAP-PrPhe, green is for ANAP-N_3Lys, pink is for ANAP-PrK, sky-blue is for AzPhe-PrK and dark blue is for PrPhe-N_3Lys. For each pair of UAA, the first UAA is introduced via TAG stop codon suppression and the second one is introduced via TGA stop codon suppression.

3.4.7.1. ANAP Incorporation by TAG Stop-Codon Suppression:

For ANAP incorporation at amino acid position LYS52, when AzPhe was incorporated at PHE99, VAL176 & GLN204 positions, the N.E.R. values were 13.64%, 14.09% and 13.33% respectively (Figure 43). The maximum N.E.R. value for total non-intended incorporation between ${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{CUA}$ pair and AzPhe was 1.36% while for ${}^{OMeY}_{OMeY}RS$ / ${}_{OMeY}tRNA_{UCA}$ pair and ANAP it was 2.95% (Figure 40 A). Since, for two-UAA incorporation, the measured N.E.R. values for this condition are significantly higher than the N.E.R. values for total non-intended amnio acid incorporation, it demonstrates that when ANAP was incorporated at LYS39TAG position, AzPhe was incorporated at PHE99TGA, VAL176TGA & GLN204TGA positions.

For ANAP incorporation at amino acid position LYS52, when N_3Lys was incorporated at PHE99 & GLN204 positions, the N.E.R. values were 13.99% and 15.39% respectively (Figure 43). The maximum N.E.R. value for total non-intended incorporation between the ${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{CUA}$ pair and N_3Lys was 6.08% while for ${}_{MmPyl}RS$ / ${}_{MmPyl}tRNA_{UCA}$ pair and ANAP it was 0.2% (Figure 40 A). Since, for this condition as well, the measured N.E.R. values for two-site incorporation are higher than the N.E.R. values total non-intended incorporation, these results also demonstrate that when ANAP was incorporated at LYS39TAG position, N_3Lys was incorporated at PHE99TGA & GLN204TGA positions.

3.4.7.2. AzPhe Incorporation by TAG Stop-Codon Suppression:

For two-site UAA incorporation, when AzPhe was introduced in eGFP at amino acid position TYR39TAG, the N.E.R. value for ANAP incorporation via TGA stop-codon suppression was 24.5% and 9.72% respectively for LYS52 and GLN204 positions (Figure 43). When AzPhe is introduced at position PHE99TAG, the N.E.R. value for ANAP incorporation at GLN204TGA position is 10.05%. The maximum N.E.R. value for the crosstalk between ${}^{OMeY}_{OMeY}RS$ / ${}_{OMeY}tRNA_{CUA}$ pair and ANAP was 5.17% while for ${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{UCA}$ pair and AzPhe it was 0.58% (Figure 40 A). Since, for two-UAA incorporation, the measured N.E.R. values for this condition are higher than the corresponding UAA-crosstalk values, these results demonstrate that for these three discussed conditions, AzPhe and ANAP unnatural amino acids were incorporated via TAG and TGA stop codon suppression respectively.

3.4.7.3. N_3Lys Incorporation by TAG Stop-Codon Suppression:

For two-site UAA incorporation, when N_3Lys was introduced in eGFP at position TYR39TAG, N.E.R. value was 13.47% when ANAP was incorporated at position LYS52 via TGA stop codon suppression. The maximum N.E.R. value for the crosstalk between ${}_{MmPyl}RS$ / ${}_{MmPyl}tRNA_{CUA}$ pair and ANAP was 0.96% while for ${}^{ANAP}_{EcLeu}RS$ / ${}_{EcLeu}tRNA_{UCA}$ pair and N_3Lys it was 6.08% (Figure 40). Since, for two-UAA incorporation, the measured N.E.R. values for this condition are higher than the corresponding UAA-crosstalk values, these results demonstrate that when N_3Lys was incorporated at the TYR39 position via TAG stop-codon suppression, ANAP was incorporated at the LYS52 position via TGA stop codon suppression.

Observations from the endpoint fluorescence measurements (Figure 44) corroborated the observations from the image analysis (except a few differences in the individual N.E.R. values). Total eGFP expression also follows the same overall trend (Figure 45).

For each of the nine conditions, ANAP was always incorporated, either at amino acid position LYS52 or at GLN204. For both these positions, the other amino acids incorporated along with ANAP are AzPhe in 6 conditions and N_3Lys in 3 conditions. Both these amino acids can be used for introducing fluorophores using click chemistry. Since from sections 3.4.4 and 3.4.6 we also know that there is at-least some non-specific incorporation as well as UAA misidentification by the oRS in each condition, the true incorporation-efficiency determination would require peptide ratio determination using mass spectrometry.

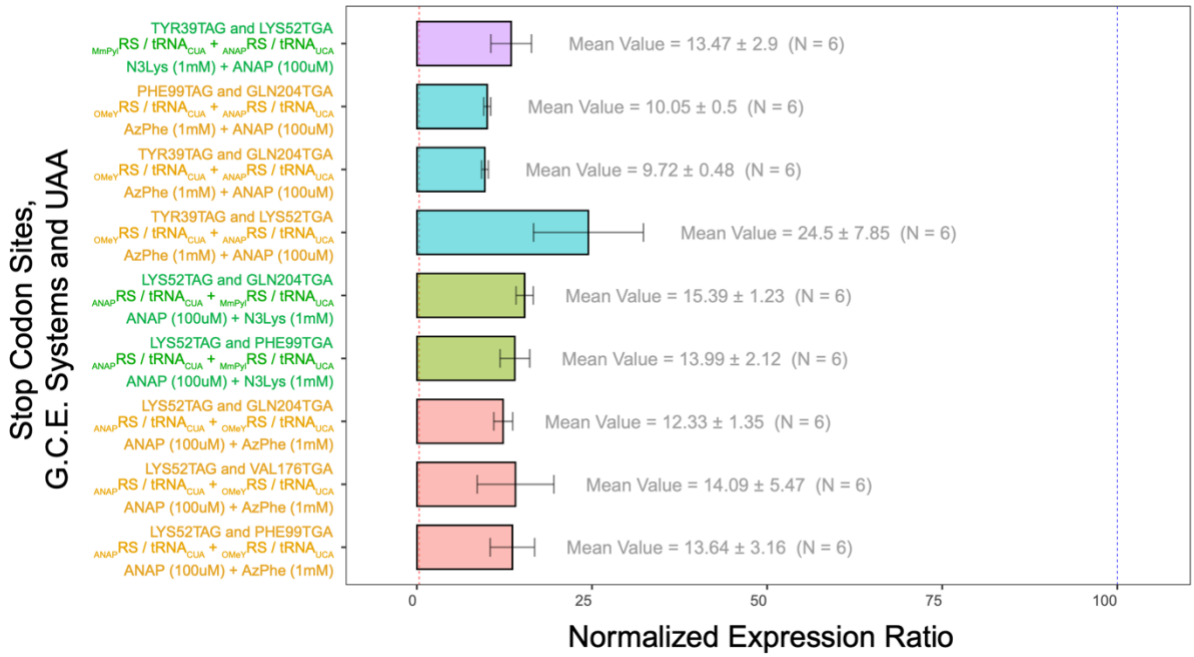


Figure 46: Two-Site UAA Incorporation - Screening Conditions with N.E.R. Cutoff ~10%.
Y-Axis details the screening condition (the two eGFP Mutations, the two G.C.E. systems and the two UAAs & their concentrations). X-Axis represents the linear scale for the N.E.R. Length of the bar represents the mean value and the whiskers represent the standard error. Red dashed line represents the N.E.R. for the control-condition from transfection with *Plasmid MX01.4* and blue line represents the N.E.R. for the control condition from transfection with *Plasmid UTX0*. Bar plots are grouped and colored according to the two oRS present in the screening conditions. Y-Axis labels are colored according to the UAA pair present in the medium: orange is for ANAP-AzPhe, and green is for ANAP-N_3Lys. For each pair of UAA, the first UAA is introduced via TAG stop codon suppression and the second one is introduced via TGA stop codon suppression.

From the two-UAA incorporation screening experiments, by analyzing the two-UAA incorporation pattern and non-intended UAA incorporation data, we conclude that the targeted UAAs are incorporated in the eGFP for the nine conditions where the N.E.R. values were around or above our cutoff of 10%. Overall, two-UAA incorporation has the same trend as was predicted by our model, but the individual values for the incorporation are different. Seven of the ten most efficient conditions could be predicted by our model.

4. DISCUSSION AND OUTLOOK

4.1. SUMMARY OF RESULTS

Three Gateway™ cloning and MultiBac™ cloning compatible novel destination vectors, namely *Plasmids GIDC, GIDK and GIDS* were created for expressing multiprotein complexes in insect cells. Three Gateway™ cloning compatible destination vectors, namely *Plasmid MX01*, *Plasmid UCAP* and *Plasmid UCZP* were created for mammalian expression systems. One Gateway™ cloning compatible donor vector, namely *Plasmid GD54* was created to facilitate Gateway™ cloning in *Plasmid UCAP*. Using these plasmids, twenty-one 'G.C.E. Plasmids' were created for delivering the genes for the Orthogonal Translation Systems or OTS (*i.e.*, orthogonal aminoacylated tRNA synthetase and orthogonal tRNA pairs) into mammalian cells. Fifty-five 'G.E.M.S. Plasmids' were also created to screen for unnatural amino acid incorporation in eGFP. A cell-based screening assay was developed to compare the efficiency of different G.C.E. systems. Using the G.E.M.S. Plasmids and G.C.E. Plasmids, screening experiments for single unnatural amino acid incorporation were performed to compare the efficiency of thirty G.C.E. systems for UAA incorporation at fifteen different amino acid positions in eGFP. Based on the results of this screening, a strategy was developed for simultaneous incorporation of two different unnatural amino acids in eGFP. Best performing G.C.E. systems and amino acid positions on eGFP were shortlisted to screen for simultaneous incorporation of a fluorophore and a bioorthogonal handle into eGFP protein. The last aim, i.e., introduction of FRET pair in multienzymes, could not be completed due to lack of time.

4.2. PLASMID VECTORS FOR ASSIMILATING MAMMALIAN OTS

We have developed two plasmid vectors, namely *Plasmid UCAP* and *Plasmid UCZP*, to assimilate existing mammalian OTS for unnatural amino acid incorporation of proteins expressed in mammalian cells. In addition, we have also developed a plasmid vector *Plasmid MX01* for delivering the gene for the target-protein (P.O.I.) in the mammalian cells. This two-plasmid system allows the target-protein gene to be expressed with different G.C.E. systems and *vice-versa*. As the overall architecture of our screening plasmids (*i.e.*, the G.E.M.S.* Plasmids) is derived from *Plasmid MX01* (with a few changes in the reporter gene), we were able to 'mix-and-match' these plasmids to screen nearly 1000 conditions (including negative controls) for UAA-incorporation in mammalian cells by using a set of 21 G.C.E. Plasmids and 55 G.E.M.S. Plasmids (45 plasmids for single unnatural amino acid incorporation and 10 plasmids for simultaneous incorporation of two different unnatural amino acids).

For *Plasmid UCAP* and *Plasmid UCZP*, we have introduced SV40 origin of replication to allow continued multiplication of the plasmid in mammalian cells expressing the T-antigen (*e.g.*, HEK293T cells). To have the possibility of integrating the oRS and otRNA genes into the genome of cells, we have also included a FRT site for facilitating Flippase-mediated recombination into the genomic DNA of HEK-FlpIn cells. These features would allow us to explore multiple options for optimization of protein production in mammalian cells.

We decided to use Gateway™ technology for introducing the target-protein and OTS genes. Gateway™ cloning is a recombination-based cloning. Due to the directional nature of the Gateway™ cassette and the presence of the suicide gene in this cassette, the Gateway™ cloning is very efficient in generating expression vectors. Gateway™ recombination is also suitable for cloning large genes into the expression vectors, which allows the possibility to use *Plasmid MX01* for expressing multienzymes and large protein complexes in mammalian cells.

Gateway™ cloning relies only on PCR for introduction of genes into the expression cassette of the destination vector (*i.e.*, the precursor of the expression vector, *e.g.*, *Plasmid MX01*) [232, 240]. If the gene can be procured from a commercial source as the attB flanked DNA, even the need for PCR can be obviated. We obtained some of the orthogonal aaRS and all of the orthogonal tRNA from commercial sources. By obtaining $_{MmPyl}RS$ and all the otRNA genes as attB flanked DNA, we were able to create some of the G.C.E.

* G.E.M.S. = G.C.E. Efficiency Measurement and Screening.

Plasmids by using only Gateway™ recombination cloning. Thus, we have demonstrated that synthetic genes can be viable route for assimilating the existing genetic code expansion systems. The presence of mutually orthogonal Gateway™ cassettes in *Plasmid UCAP* allowed us to simultaneously introduce the oRS and otRNA genes in the mammalian expression vector by using a one-pot recombination reaction.

Regarding the plasmid architecture for genetic code expansion, two approaches are followed. Either all the genes are delivered using single plasmid, or the genes are split over two plasmids (one-plasmid gene delivery vs two-plasmid gene delivery) [247, 248]. For the two-plasmid approach, either the target-protein and G.C.E. genes (for the oRS and otRNA) are delivered separately, or multiple copies of the orthogonal tRNA are split between the two plasmids, while delivering the target-protein and oRS genes via separate plasmids. The expression levels of the genes are dependent on the overall architecture of the plasmids, which includes elements present on the expression vector, such as the promoter, the enhancers, replication motifs, relative arrangement of genes etc. It has been reported that genetic code expansion from single-plasmid may result in higher UAA incorporation efficiency [247]. Recently, the effects of the plasmid architecture were studied in NIH3T3 and HEK293T cells for several $_{Pyl}RS$ / $_{Pyl}tRNA_{CUA}$ pairs (TAG-suppression) [249]. This study claimed that two-plasmid gene delivery, with two copies of the otRNA in each plasmid, is a better approach for UAA incorporation, than delivering the P.O.I. gene and the OTS genes separately [249]. These conflicting results demonstrate that the impact of the plasmid architecture on the UAA-incorporation efficiency may be dependent on the G.C.E. system. The presence of SV40 *ori* on the plasmids was demonstrated to be better for UAA incorporation when the cells could express T-antigen (*i.e.*, HEK293T cells) [249].

In a recent study, three-plasmid approach was used to study the cross-reactivity between the different oRS and otRNA of supposedly mutually orthogonal OTS [250]. While this approach is faster for the intended purpose, three-plasmid co-transfection may not be ideal for the homogeneity of the transfection. As we have demonstrated in this study, the ratio of the two plasmids (G.C.E. Plasmid and P.O.I. Plasmid) can be an important factor for unnatural amino acid incorporation in proteins, a more homogeneously transfected sample could be better suited for the characterization of the OTS and for UAA incorporation in the target-protein [213].

Since two independent plasmids (one plasmid for delivering OTS gene and the other for delivering the target-protein gene) are better suited for screening applications, we believe that our two-plasmid approach is better suited for assimilating OTS for mammalian cells and evaluating their efficiency.

4.3. BACMAM™ COMPATIBILITY FOR LARGE SCALE PROTEIN PRODUCTION

Our plasmid vectors are MultiBac™ and BacMam™ cloning compatible. BacMam™ technology allows baculoviral gene delivery into mammalian cells. *Plasmid MX01* has been designed such that the expression vectors (P.O.I. Plasmids) derived from it are compatible with transient transfection as well as baculovirus-mediated viral transfection (by using the BacMam technology) [243, 244]. To ensure that same G.C.E. plasmids can be used with multiple P.O.I. plasmids for baculoviral gene delivery, we designed *Plasmid UCAP* and *Plasmid UCZP* as a MultiBac donor vector, which means that they can be combined *in-vitro* with the P.O.I. Plasmids by using the Cre-LoxP recombination (Figure 21). The large foreign gene capacity (>100 kbp) of the baculovirus is equipped to deal with the resultant increase in the size of the plasmids [251].

The low efficiency of UAA incorporation is a limiting factor for producing of the full-length UAA-containing protein, as a result, a higher amount of culture must be used for the production of adequate amounts of protein for subsequent operations on the purified protein. For large volumes of culture, cationic lipid or cationic polymer mediated gene delivery may not be the most efficient option. As an alternative, baculovirus mediated transfection provides a cheaper and more scalable option for gene delivery [252]. While virus production takes around three weeks, once produced, the virus cultures are viable for around three months so they can be used for multiple transductions. As the virus cannot multiply in mammalian cells or integrate in the host genome, the nature of the gene delivery is transient. Even before the start of this work, baculoviral gene delivery for mammalian cells has been used for gene delivery in immortalized cell lines, stem cells, primary cells and animals [146, 253]. These considerations encouraged us to develop a gene delivery system compatible with the BacMam™ system to have an option of large-scale protein production, if needed.

Recently, two $_{MmPyl}RS$ / $_{MmPyl}tRNA$ pairs based on two different polyspecific $_{MmPyl}RS$ were cloned into the bacmid genome to create the MultiBacTAG protein expression system [254]. Using this expression system, unnatural amino acids (allowed pyrrolysine homologues) can be introduced in proteins expressed in large cultures of insect cells (*Spodoptera frugiperda* cells, *SF9* or *SF21*). Since the two $_{MmPyl}RS$ / $_{MmPyl}tRNA$ pairs are polyspecific, they can be used for multiple unnatural amino acids. In comparison, our approach of joining two independent plasmids via Cre-LoxP recombination gives the flexibility for the selection of the required orthogonal aaRS/tRNA pair.

This can allow baculoviral delivery of mutually orthogonal OTS genes to mammalian cells for incorporation of multiple unnatural amino acids in proteins.

4.4. AUTOMATION COMPATIBLE SCREENING ASSAY

We have developed a cell-based screening assay for quantifying the unnatural amino acid incorporation by genetic code expansion (G.C.E.) in mammalian cells. By using this assay, the most efficient G.C.E. systems and expression conditions can be shortlisted for further optimizations. The assay involves fluorescence-imaging-based monitoring of the relative fluorescence of two proteins (mRaspberry and eGFP) that are expressed by the same promoter but as different polypeptides. Of these, the second protein (eGFP) is expressed only when the genetic code expansion is successful. As a control of the image-based assay, we also use the endpoint fluorescence (after cell lysis) of the two fluorescent proteins to quantify their relative abundance.

For a protein where limited structural information is available, unnatural amino acid incorporation by genetic code expansion can provide valuable information such as its dynamics, interacting partners and effects of post-translational modifications. Since unnatural amino acid incorporation can be dependent on many factors such as the G.C.E. system, unnatural amino acid concentration or position of the stop-codon, a screening strategy, where the effects of such factors can be assessed in a short time, is necessary to shortlist the most viable conditions. Fluorescence imaging and F.A.C.S. allow a direct readout of the total amount of fluorescent proteins expressed in the living cells. The most significant advantage of such living-cell based monitoring approaches is that the time-consuming protein purification step can be avoided for total protein quantification. Of these two, F.A.C.S. requires cells in suspension culture and to be used for adherent cultures, it requires an extra step of cell-detachment and suspension. Since manipulation of adherent culture is easier as compared to suspension culture, we decided to proceed with fluorescence-imaging-based analysis.

This approach was very useful when we decided to automate our workflow for large-scale screening, as our cell-culture operations could be easily integrated with the automation setup. By using this assay, in about two weeks, we screened over 450 conditions and their respective negative controls for single-UAA incorporation to select for best suited conditions for simultaneous two-UAA incorporation. The automation setup required some time (2-3 months) for preparing the automated workflow pipeline and the trial runs. However, since this setup is now ready, newer screening assays would require much less time.

High-throughput screening approaches have been reported for *in-vitro* expression and for bacterial cells using $_{MmPyl}RS$ based G.C.E. Systems [255, 256]. The *E. coli* extract based cell-free *in-vitro* screening takes about 48 hours and uses split-GFP fluorescence and lysozyme activity as reporters [255]. Using the *in-vitro* expression, 24 amino acid positions on lysozyme could be screened for optimal UAA incorporation in nearly 48 hours [255]. For screening UAA incorporation in bacterial cells, the cell mass and eGFP fluorescence were used as reporters for online monitoring [256]. Fluorescence-imaging based quantification and comparison of genetic code expansion efficiency was also reported recently, although with manual imaging using an inverted fluorescence microscope [249]. Together, these studies confirm that the context of the stop-codon on the mRNA is important for determining UAA-incorporation efficiency [255, 257] and UAA-incorporation efficiency depends on the G.C.E. system and the concentration of the unnatural amino acids [249]. During our screening experiments we have found similar results, *i.e.*, unnatural amino acid incorporation may be dependent on the G.C.E. system (*e.g.*, $_{MmPyl}RS$ was more efficient than $_{MbPyl}RS$ when same otRNA and the unnatural amino acids were present), the unnatural amino acid incorporation (*e.g.*,

100 μM PrK vs 2 mM PrK $_{MmPyl}RS$ based G.C.E. Systems) and on the amino acid position (*e.g.*, different positions on eGFP).

For the end user of synthetically evolved G.C.E. Systems, these observations underscore the importance of having a robust screening strategy for choosing the best G.C.E. system or expression conditions for introducing special functional groups in proteins. Such a strategy should not be limited by the G.C.E. system or the amino acid position on the target-protein, but should be able to accommodate both of them simultaneously, along with other factors such as amino acid concentration, presence of additives etc. Since this creates a large sample space for screening, workflow automation may be very helpful of taking the load off the user and for the reproducibility of operations and results. An automation-compatible screening assay, such as ours, would be very useful for such screening experiments.

4.5. EGFP REPORTER FOR THE SCREENING ASSAY

Green Fluorescent Protein and its variants have been the most widely used reporters for quantifying unnatural amino acid incorporation efficiency, either alone, or with mCherry [146, 185, 207, 213, 248]. We decided to use eGFP protein as the reporter for unnatural amino acid incorporation because it is very bright and can be detected even at a low concentration. We reasoned that this would allow efficiency-quantification for low-efficiency G.C.E. systems as well. To minimize the effects of mRaspberry, on eGFP, we expressed it as a separate polypeptide using P2A and T2A sites. For selecting amino acid positions on eGFP for UAA incorporation, we relied on its structure and selected those amino acid positions where side chains pointed away from the β-barrel (except one position that was buried) (Figure 27). Many of these positions have been reported as permissive positions (*e.g.*, TYR39, TYR151 etc.). These precautions were taken to ensure that unnatural amino acid incorporation does not interfere with the formation of the eGFP fluorophore, and hence eGFP fluorescence could be an accurate reporter of unnatural amino acid incorporation.

However, despite our best efforts, it may be possible that unnatural amino acid incorporation may affect fluorophore formation. While using eGFP as the test protein, our aim was to select for the best G.C.E. systems and amino acid positions for unnatural amino acid incorporation, to develop a strategy for eventually introducing two different unnatural amino acids simultaneously. Since by measuring eGFP fluorescence, we could simultaneously confirm unnatural amino acid incorporation as well as the activity of protein (*i.e.*, fluorescence), this approach served its intended purpose for us.

For screening a target-protein, we recommend putting the target-protein gene between mRaspberry and eGFP genes. This can be done by putting the target-protein gene between P2A and T2A sites after XbaI-mediated linearization of *Plasmid UTX0*. Since all three proteins would be now expressed as separate polypeptides and since eGFP fluorescence would be completely decoupled from UAA-incorporation effects, by using the same assay principles the UAA-incorporation efficiency can be screened for the target protein.

4.6. INCORPORATION OF TWO UNNATURAL AMINO ACIDS IN PROTEINS

By using eGFP as the test protein, we have demonstrated a strategy for incorporation of two unnatural amino acids into proteins expressed in mammalian cells. Using this strategy, by screening 30 different G.C.E. systems for single-UAA incorporation at 15 different amino acid positions in eGFP, we were able to shortlist those conditions where two UAA incorporation could work for introducing either a fluorophore and a click-reaction handle or mutually orthogonal click-reaction handles into eGFP. Consequently, after another round of screening, this time for two-UAA incorporation, we could shortlist mutually orthogonal oRS/otRNA pairs and amino acid positions that show >10% UAA incorporation efficiency in eGFP. These conditions can be optimized further for improving the UAA incorporation and overall protein production. In this study, we have also demonstrated that screening just one amino acid position for comparing the efficiency of different G.C.E. systems or screening just one G.C.E. system for determining the optimal amino acid position of UAA incorporation may not be the best strategy for two-UAA incorporation in proteins.

For mammalian cells, the incorporation of two different unnatural amino acids using mutually orthogonal G.C.E. systems was first reported for the permissive sites of eGFP (amino acid positions 39 and 151) and for heavy & light chains of Herceptin-auristatin antibody [258]. In this study, $_{MbPyl}RS$ / $_{MmPyl}tRNA_{UUA}$ and $_{EcTyr}RS$ / $_{EcTyr}tRNA_{CUA}$ pairs were used for introducing mutually orthogonal click-reaction handles in these proteins. For selection of a stop-codon orthogonal to the TAG stop-codon, TAA-suppressing $_{MmPyl}tRNA_{UUA}$ was selected over TGA-suppressing otRNA as it demonstrated a higher UAA incorporation. In a more recent study, by using $_{MbPyl}RS$ / $_{MmPyl}tRNA_{CUA}$ and $_{EcTyr}RS$ / $_{EcTyr}tRNA_{UCA}$ pairs, two unnatural amino acids were incorporated at TYR39TAG and TYR151TGA positions of eGFP [250]. In this study, for the different unnatural amino acid pairs, the yield for the double mutant was between 0.6% and 7.5% with respect to the WT-eGFP. The study also demonstrated 4-6 times lower efficiency for TGA-suppression (for amino acid position TYR39), as compared to the TAG-suppression for G.C.E. systems derived from $_{EcTyr}RS$, $_{MbPyl}RS$ and $_{EcLeu}RS$. In comparison, for our study, we found TAG-suppression to be 1.5-2 times more efficient than TGA-suppression for $_{ANAP}RS$ (based on $_{EcLeu}RS$), 2-3 times for the $_{OMeY}RS$ (based on $_{EcTyr}RS$) and almost the same for $_{MmPyl}RS$.

For efficient incorporation of multiple different unnatural amino acids in proteins via genetic code expansion, mutually orthogonal oRS/otRNA pairs are needed. In addition, two different blank-codons on the mRNA are also required to function as unique signals for UAA incorporation. Finally, since the efficiency of UAA incorporation is reduced with every new unnatural amino acid on the polypeptide chain, the individual genetic code expansion systems need to have a high UAA-incorporation efficiency, if multiple UAA incorporation can be productively used.

In a recent study, mutual orthogonality of the different G.C.E. systems was systematically studied [250]. It was demonstrated that TAA-suppressing tRNAs could suppress TAG stop codons as well, possibly due to wobble effect. It was also observed that $_{EcTyr}RS$ could aminoacylate $_{EcLeu}tRNA_{CUA}$. For simultaneous incorporation of two-UAA in proteins, such an approach is essential for establishing true mutual orthogonality of the G.C.E. systems. Recently, a $_{Pyl}RS$ / $_{Pyl}tRNA$ pair was discovered, that was orthogonal to the $_{MmPyl}RS$ / $_{MmPyl}tRNA$ pair [259]. By evolving these two to be specific to two different unnatural amino acids, it was possible to site specifically incorporate Nε-benzyloxycarbonyl-L-lysine (CbzK) and Nε-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine (CypK) in proteins expressed in *E. coli* cells. This is an important development since it demonstrates that mutually orthogonal G.C.E. systems can be evolved from evolutionary divergent aaRS/tRNA pair even though they may have the same amino acid substrate. Using the same principle, it was further demonstrated, in *E. coli* cells, that given enough number of blank-codons, and high enough individual UAA incorporation efficiencies, multiple unnatural amino acids can be incorporated in proteins via genetic code expansion [260]. In this study, the TAG stop codon and two quadruplet codons were used with mutually orthogonal $_{Pyl}RS$ / $_{Pyl}tRNA$ pairs to incorporate three distinct unnatural amino acids in proteins expressed in *E. coli*. Since the archaea-derived $_{Pyl}RS$ / $_{Pyl}tRNA$ pairs are orthogonal in eukaryotes as well, such newly-evolved pairs can be easily transferred to mammalian systems (although with stop-codon suppression, as quadruplet-codon suppression is not yet sufficient for mammalian systems). Indeed, by using two such pairs, and by improving their UAA incorporation efficiency via synthetic evolution, it was possible to incorporate two different unnatural amino acids (in response to TAG and TAA stop codons) on a synthetic minimal notch receptor SynNotch expressed in mammalian cells [261]. Furthermore, since the position of the incorporated unnatural amino acids was on the cell surface, it was possible to label the cell surface receptors with fluorescent dyes.

These examples demonstrate that considerable attention is being given (and rightly so) on the development of the mutually orthogonal oRS/otRNA pairs for multiple UAA incorporation in proteins. However, a systematic approach for integrating these newly developed G.C.E. systems is still missing.

While developing our strategy for two-UAA incorporation in proteins, we have used eGFP as the proof of concept. The same strategy can be used for incorporating two different unnatural amino acids into a target protein. For multiple UAA incorporation in proteins, due to the loss of efficiency upon subsequent addition of unnatural amino acids, choosing the most efficient conditions is very important. The efficiency of UAA-incorporation depends on the G.C.E. system, the expression conditions and on the position of the stop-codon [255, 257]. By using our screening approach, these can be optimized and compared to select for the best mutually orthogonal pair. In this study, we used a test protein, whose activity (*i.e.*, fluorescence) could be monitored online. We understand that the same may not be the true for the target-protein. However, by

slightly altering our screening platform (*e.g.*, by introducing the target-protein gene between mRaspberry and WT-eGFP genes) and screening for successful stop-codon suppression in the target-protein and full-length protein production, the most efficient conditions can be shortlisted for further assay and optimizations. For those proteins where the structural information is not prevalent, this approach can provide a significant advantage for incorporation of multiple UAA incorporation.

4.7. OUTLOOK

Life has evolved on earth for over 3.5 billion years. During this time, biological systems were 'programmed' to incorporate 22 amino acids in proteins. However, in past 20 years, more than 150 unnatural amino acids have been added into proteins via genetic code expansion (G.C.E.). There is still the question of low efficiency of these systems, but nonetheless G.C.E. is a feat of mankind, and possibly one of the first steps towards an expanded synthetic-biology universe. Genetic code expansion has already been used in many areas of biological research to characterize proteins and study their impact on living organisms. Our understanding and reprogramming of the protein biosynthesis machinery have contributed significantly to these developments. Fluorescent proteins once revolutionized our understanding of biological systems. Genetic code expansion has the same potential. However, G.C.E. in living cells is significantly more complex than fluorescent protein expression in living cells. There are many individual parameters that must be optimized by the end user, which can be a daunting task and may discourage the use of such systems. This thesis has presented a systematic approach for such optimizations.

Unnatural amino acid incorporation allows targeted modifications of proteins, which can have many biotechnological, therapeutic and industrial applications. Alongside the development of new approaches for the G.C.E., standardization and commercialization of G.C.E. is an important step towards improving its wide-spread acceptance. Owing to their vast resources, biopharmaceutical industries can play a significant role in this. They might also have some stake in commercialization of the G.C.E.

Therapeutic proteins are used to replace abnormal or deficient proteins in body to reduce or neutralize the impact of diseases. However, renal elimination significantly reduces the half-life of many therapeutic proteins in the blood circulation, thereby reducing their efficacy and requiring frequent administration. Covalent modification of therapeutic-proteins improves their affinity, stability and pharmacokinetic properties, and can also significantly improve their stability against proteolytic degradation, thereby reducing renal elimination [262]. Such covalent modification strategies may include PEGylation, lipidation, glycosylation, polypeptide-fusion (FC-domain-fusion or albumin-fusion). These covalent modification strategies usually involve the use of reactive amino acid side-chains present on the therapeutic proteins. However, this introduces sample heterogeneity, which brings challenges for therapeutic-characterization and could be relevant for the regulatory approval. By using unnatural amino acids, unique reactive-handles can be introduced into therapeutic-proteins for site-specific modification. This would allow better characterization of the therapeutic and pharmacokinetic properties of the protein post its modification.

Similarly, UAA-mediated site-specific modification of antibodies for creating the antibody-drug conjugates can improve the sample homogeneity and facilitate better characterization. Such conjugates are used for targeted drug delivery. By introducing radioactive nuclei reporters on the antibody, these conjugates also be used for theranostic applications. Bispecific antibodies that are used to treat multifaceted diseases can be stabilized by UAA-assisted site-specific conjugation in addition to the regular electrostatic interaction between the antibody fragments. As in all areas of manufacturing, for biological drugs as well, counterfeits present a serious problem. By using genetic code expansion, fluorescent markers with characteristic photophysical properties can be introduced in protein-based therapeutics for rapid on-site validation of the therapeutic proteins.

Apart from creating proteins with customized functionalities, organisms with customized functionalities can also be created using genetic code expansion. These can be useful for research as well as for manufacturing processes. By developing pathogens auxotrophic for unnatural amino acids, another layer of biosafety can be added. This can be done by linking the expression of survival genes to genetic code expansion in the cells. Such a development may eventually facilitate less stringent safety protocols for research and development on pathogenic strains, which would improve the quantity of much needed research done on such organisms.

By developing transgenic organisms with G.C.E. integrated in the genomes, the production of target-protein (that contains the blank-codon for the G.C.E.) can be turned on or off depending upon the addition of the unnatural amino acids. This can be particularly useful for expressing those genes that require specific biological conditions for production (*e.g.*, polyketides or antibiotic precursors, which require late exponential phase of growth). Such a switch can be placed anywhere in the biosynthesis machinery to control the behavior of the organism.

Genetic code expansion can also be important for the manufacturing industry, particularly related to development of polymers. Polypeptides are polymers of amino acids. Their sequence and environment determine their three-dimensional structure, and as a result, these factors can affect its function. Synthetic polypeptides are created either by solid phase peptide synthesis (SPPS) or by N-carboxyanhydride (NCA) ring-opening polymerization [263]. The former method suffers from low yield and in the latter one the polypeptide cannot be sequenced. Biological or enzymatic synthesis of polypeptides, on the other hand, is both cheap and precise. Once the issue of low efficiency is addressed, genetic code expansion can provide a solution for cost effective, scalable and homogeneous synthesis of polypeptides. Since for G.C.E., the polypeptide sequence is controlled by the DNA sequence, just by modifying the DNA sequence, many different kinds of polypeptides can be created by using almost similar protocols. By modifying the length of polypeptides and the side-chains of unnatural amino acids, custom functionalities can be introduced into the polypeptides. These biopolymers can be used to create nanoparticles for gene-, protein- or drug-delivery, hence they can be used in therapeutics. For therapeutic proteins, aggregation at a high concentration is a major challenge, and can cause immunogenicity. Functionalized biopolymers can be used for improving the solubility of proteins without the need for covalent modification. Unnatural amino acid biopolymers can also be used for functionalization of surfaces for biotechnological and biomedical applications. By modulating their sensitivity to their environment, UAA-polymers can also be used as biosensors. By using UAA with crosslinking side chains, biodegradable biopolymers can also be produced from such polypeptides. As we move towards an eco-friendlier approach, such biodegradable polymers would have numerous medical and industrial applications.

Some challenges still exist for commercial viability of unnatural amino acid incorporation in proteins, the most important being scalability and case-by case optimization of genetic code expansion systems. As the G.C.E. systems are evolved to introduce a specific kind of unnatural amino acid in proteins, their characterization and optimization parameters differ from one G.C.E. system to the other. For a wider acceptance of the genetic code expansion technology, some lessons can also be learnt from the success of fluorescent proteins. The most important reason for their impact is the ease with which they can be integrated into an 'existing workflow' of protein expression. Fluorescent protein genes can be easily obtained as synthetic genes and fused to the gene of interest. Non-profit databases such as *fpbase.org* can allow the end users to quickly access their requirements, compare the available fluorescent proteins and get the protein sequence, all from the same source. This is much easier than, for example, going through ten different sources and still having doubts about the correct sequence of the required genes because 'some critical information could have been overlooked'. Having a centralized database of evolved orthogonal translation system (OTS) would also prevent duplication of efforts to evolve and characterize OTS for a specific purpose. In addition, having an unnatural amino acid bank would also improve the outreach of this technology. A typical research group has individuals with almost similar expertise. Genetic code expansion, on the other hand, may require individuals with several different expertise, for example, molecular biologists, cell-biologists, chemists etc. For some of these areas, the overlap between them is very limited. While the genes for the OTS can be procured from commercial sources, many unnatural amino acids are synthesized inhouse by the research groups, and they may not be commercially available yet. Having an unnatural amino acid bank would help those research groups where the expertise or the resources for chemical synthesis of unnatural amino acids is not present.

While synthetic evolution of new G.C.E. systems (orthogonal aaRS/tRNA pairs as well as available blank-codons) is paramount for pushing this technology further, efforts should also be made to make this technology accessible for a wider base of researchers. The most efficient way to do this would be for the academia and the industry to work together for the standardization and commercialization of this technology.

VII. ACKNOWLEDGEMENTS

I am filled with immense gratitude towards everyone who helped me and supported me to achieve this milestone. Looking back, I realize that I could not have done it all by myself. My foremost thanks go to Prof. Dr. Timm Maier, without whom this endeavor would not have been possible. Not only has he been an amazing mentor for my academic life, but he has also been a terrific supervisor as well. He always gave me freedom when I was struggling in life and always nudged me in the right direction when I was faltering. My interaction with him has been the biggest factor of my personal and professional development during my Ph.D. years. At the same time, I also want to that my Ph.D. committee members, Prof. Dr. Roderick Lim and Prof. Dr. Benjamin Schuler, for their regular support and constructive feedback. I am also thankful to Prof. Dr. Markus Affolter for agreeing to participate in my Ph.D. thesis examination.

I would also like to thank the Swiss Nanoscience Institute (S.N.I.), not only for the S.N.I. Ph.D. Fellowship which sponsored a major portion of my Ph.D., but also for numerous Ph.D. retreats which allowed me to share my knowledge and experience with fellow scientists from other disciplines of science. I believe that dissemination of scientific knowledge and inspiring the general public for adopting scientific thinking is an important responsibility of a scientist. Hence, I want to thank S.N.I. for also allowing me the opportunity to participate in their outreach program and share my knowledge with young citizens of Basel area.

As I was starting this project, I was given the responsibility for setting up the mammalian cell culture facility. This was a challenging task and the mechanical and electrical workshops of the Biozentrum ensured that I had my customized instruments ready on time. Almost all of my data was processed on SciCore computing nodes. Handling the information at this scale would not have been possible without their support. While I was deciding on the best route for my project, the Imaging Core Facility and the F.A.C.S. Core Facility were very supporting and provided me with the technical knowhow of how to proceed further. Therefore, I also want to thank these facilities of the Biozentrum. I also want to thank Dr. Peter Schultz and Dr. Dannel McCollum, who were very kind to provide some of the plasmids used in this study.

My Ph.D. was a journey with many ups and downs. Despite being so far away from my home (6710 ± 1 km, if someone is interested in the context), I have never felt alone. I cannot express in words how thankful I am, that Bhanupriya, my wife, has been here with me during all of this. There were times when I was too weak to continue and she motivated me to bring out the best in me. There were times when she put her life on hold to make sure that I could move forward in mine. There were good times and there were bad times and she has been there in all of them to give me company. I could not have done this without her love and support. On a side note, for this very reason, I am also thankful to my in-laws and my parents for 'forcing me' to marry her even when I did not believe that I was ready for this life altering decision. I guess they knew something I did not.

I am glad that I started my Ph.D. in a lab full of tremendous scientists and was able to develop a friendship with many of them. Being in their company has been very motivating. I am also thankful that I could share many cheerful moments of my life with them. It was good to have Francesca as a good friend, although sometimes she was a bit bossy one. I still have a so-called PTSD from that one board game that we played a few years ago. In the free time, it was always good to participate in non-scientific conversations with Yves, Janine and Malik. It was good to share food and recipes with Yana. I could bond with Leonie over our mutual love for slightly darkish humor. I have always seen Roman as an overall positive influence in the lab and having him in the same office was very motivating. Florian and Hugo have always been there with their (ahem… in the interest of safety, let's call it witty) humor. Florian has also been quite helpful due to his muscles and positive attitude. It was nice participate in good scientific discussions with Niko who helped shaped some of my research ideas. During the thesis writing, sometimes when I was feeling low, it was nice to talk to Karo and Louise-Marie. These conversations have always uplifted my spirits. Stefan was very helpful in brainstorming sessions for thesis writing. A few friends from back home, particularly Raj, Shveta, Anurag, Sid and Dhwani deserve special mention for making sure that I survived the COVID times.

I would also like to mention my parents and my brother here, although they have not been directly involved in my academic journey. As a matter of fact, I think that they do not know even half of it. I believe they think that I am working on developing some new vaccine or solving world hunger so something like that. I do not blame them. I have kept my journey a little bit ambiguous. Nonetheless, they have been very supportive of

my journey and my dreams. When I migrated to Europe to learn and grow, I also left a big emotional hole in their lives, and I know that they have struggled with those emotions. So last, but certainly not the least, I would like to acknowledge their struggle and their sacrifice, and their positive thoughts for me.

VIII. REFERENCES

1. Chin, J.W., et al., *An Expanded Eukaryotic Genetic Code.* Science, 2003. **301**(5635): p. 964-967.

2. Gómez-Márquez, J., *What is life?* Mol Biol Rep, 2021. **48**(8): p. 6223-6230.

3. Rees, D.C., T.N. Williams, and M.T. Gladwin, *Sickle-cell disease.* Lancet, 2010. **376**(9757): p. 2018-31.

4. Rehman, I., M. Farooq, and S. Botelho. *Biochemistry, Secondary Protein Structure*. StatPearls 2021 December 15, 2021; Available from: https://www.ncbi.nlm.nih.gov/books/NBK470235/.

5. Griffith, E.C. and V. Vaida, *In situ observation of peptide bond formation at the water-air interface.* Proceedings of the National Academy of Sciences, 2012. **109**(39): p. 15697-15701.

6. Ryle, A.P., et al., *The disulphide bonds of insulin.* Biochem J, 1955. **60**(4): p. 541-56.

7. Goodsell, D. *Titin*. Molecule of the Month 2015; Available from: https://pdb101.rcsb.org/motm/185.

8. Clark, D.P., N.J. Pazdernik, and M.R. McGehee, *Chapter 13 - Protein Synthesis*, in *Molecular Biology (Third Edition)*, D.P. Clark, N.J. Pazdernik, and M.R. McGehee, Editors. 2019, Academic Cell. p. 397-444.

9. Zwanzig, R., A. Szabo, and B. Bagchi, *Levinthal's paradox.* Proc Natl Acad Sci U S A, 1992. **89**(1): p. 20-2.

10. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**(6968): p. 884-890.

11. Bonduelle, C., *Secondary structures of synthetic polypeptide polymers.* Polymer Chemistry, 2018. **9**(13): p. 1517-1529.

12. Dill, K.A. and J.L. MacCallum, *The protein-folding problem, 50 years on.* Science, 2012. **338**(6110): p. 1042-6.

13. Reeb, J. and B. Rost, *Secondary Structure Prediction*, in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, et al., Editors. 2019, Academic Press: Oxford. p. 488-496.

14. Mallamace, F., et al., *Energy landscape in protein folding and unfolding.* Proceedings of the National Academy of Sciences, 2016. **113**(12): p. 3159-3163.

15. Ponomarenko, E.A., et al., *The Size of the Human Proteome: The Width and Depth.* Int J Anal Chem, 2016. **2016**: p. 7436849.

16. Godbey, W.T., *Chapter 3 - Proteins*, in *Biotechnology and its Applications (Second Edition)*, W.T. Godbey, Editor. 2022, Academic Press. p. 47-72.

17. Alberts, B., et al. *Protein Function*. Molecular Biology of the Cell (4th Ed) 2002; Available from: https://www.ncbi.nlm.nih.gov/books/NBK26911/.

18. Lu, R.-M., et al., *Development of therapeutic antibodies for the treatment of diseases.* Journal of Biomedical Science, 2020. **27**(1): p. 1.

19. Clark, D.P., N.J. Pazdernik, and M.R. McGehee, *Chapter 11 - Transcription of Genes*, in *Molecular Biology (Third Edition)*, D.P. Clark, N.J. Pazdernik, and M.R. McGehee, Editors. 2019, Academic Cell. p. 332-361.

20. Ishitani, R., et al., *Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme.* Cell, 2003. **113**(3): p. 383-94.

21. Ivanova, N.N., et al., *Stop codon reassignments in the wild.* Science, 2014. **344**(6186): p. 909-913.

22. Hashem, Y., et al., *Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29.* Cell, 2013. **153**(5): p. 1108-19.

23. Jackson, R.J., C.U. Hellen, and T.V. Pestova, *Termination and post-termination events in eukaryotic translation.* Adv Protein Chem Struct Biol, 2012. **86**: p. 45-93.

24. Cooper, G.M. *Protein Folding and Processing*. The Cell: A Molecular Approach. 2nd Edition. 2000; Available from: https://www.ncbi.nlm.nih.gov/books/NBK9843/.

25. Thrower, E.C., et al., *Zymogen activation in a reconstituted pancreatic acinar cell system.* Am J Physiol Gastrointest Liver Physiol, 2006. **290**(5): p. G894-902.

26. Smith, S.A., R.J. Travers, and J.H. Morrissey, *How it all starts: Initiation of the clotting cascade.* Crit Rev Biochem Mol Biol, 2015. **50**(4): p. 326-36.

27. Ramazi, S. and J. Zahiri, *Post-translational modifications in proteins: resources, tools and prediction methods.* Database, 2021. **2021**: p. baab012.

28. Alberts, B., et al., *Analyzing Protein Structure and Function*, in *Molecular Biology of the Cell*. 2002, Garland Science: New York.

29. Büyükköroğlu, G., et al., *Chapter 15 - Techniques for Protein Analysis*, in *Omics Technologies and Bio-Engineering*, D. Barh and V. Azevedo, Editors. 2018, Academic Press. p. 317-351.

30. Houde, D.J. and S.A. Berkowitz, *Chapter 3 - Biopharmaceutical industry's biophysical toolbox*, in *Biophysical Characterization of Proteins in Developing Biopharmaceuticals (Second Edition)*, D.J. Houde and S.A. Berkowitz, Editors. 2020, Elsevier. p. 55-85.

31. Wilson, C.G.M. and M.R. Arkin, *Chapter 5 - Screening and biophysics in small molecule discovery*, in *Small Molecule Drug Discovery*, A. Trabocchi and E. Lenci, Editors. 2020, Elsevier. p. 127-161.

32. van der Zouwen, A.J. and M.D. Witte, *Modular Approaches to Synthesize Activity- and Affinity-Based Chemical Probes.* Frontiers in Chemistry, 2021. **9**.

33. de Boer, M., et al., *Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers.* Elife, 2019. **8**.

34. *Introduction to Fluorescence*, in *Principles of Fluorescence Spectroscopy*, J.R. Lakowicz, Editor. 2006, Springer US: Boston, MA. p. 1-26.

35. *Fluorophores*, in *Principles of Fluorescence Spectroscopy*, J.R. Lakowicz, Editor. 2006, Springer US: Boston, MA. p. 63-95.

36. Larson, A.M., *Multiphoton microscopy.* Nature Photonics, 2011. **5**(1): p. 1-1.

37. Ueno, T. and T. Nagano, *Fluorescent probes for sensing and imaging.* Nature Methods, 2011. **8**(8): p. 642-645.

38. Swanson, W.B., et al., *A library of Rhodamine6G-based pH-sensitive fluorescent probes with versatile in vivo and in vitro applications.* RSC Chemical Biology, 2022. **3**(6): p. 748-764.

39. Yamamoto, J., et al., *Quantitative evaluation of macromolecular crowding environment based on translational and rotational diffusion using polarization dependent fluorescence correlation spectroscopy.* Scientific Reports, 2021. **11**(1): p. 10594.

40. Xiao, H., P. Li, and B. Tang, *Recent progresses in fluorescent probes for detection of polarity.* Coordination Chemistry Reviews, 2021. **427**: p. 213582.

41. Algar, W.R., et al., *FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls.* Nature Methods, 2019. **16**(9): p. 815-829.

42. Kobayashi, H., et al., *Bioluminescence resonance energy transfer–based imaging of protein–protein interactions in living cells.* Nature Protocols, 2019. **14**(4): p. 1084-1107.

43. Pantazis, A., et al., *Harnessing photoinduced electron transfer to optically determine protein sub-nanoscale atomic distances.* Nature Communications, 2018. **9**(1): p. 4738.

44. Jaffe, H.H. and A.L. Miller, *The fates of electronic excitation energy.* Journal of Chemical Education, 1966. **43**(9): p. 469.

45. Usai, C. and A. Diaspro, *Fluorescence: General Aspects*, in *Encyclopedia of Biophysics*, G.C.K. Roberts, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 826-832.

46. Schuler, B. and H. Hofmann, *Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales.* Current Opinion in Structural Biology, 2013. **23**(1): p. 36-47.

47. Kawale, A.A. and B.M. Burmann, *Characterization of backbone dynamics using solution NMR spectroscopy to discern the functional plasticity of structurally analogous proteins.* STAR Protoc, 2021. **2**(4): p. 100919.

48. Schermelleh, L., et al., *Super-resolution microscopy demystified.* Nature Cell Biology, 2019. **21**(1): p. 72-84.

49. Schmidt, R., et al., *MINFLUX nanometer-scale 3D imaging and microsecond-range tracking on a common fluorescence microscope.* Nature Communications, 2021. **12**(1): p. 1478.

50. Carter, M. and J. Shieh, *Chapter 5 - Microscopy*, in *Guide to Research Techniques in Neuroscience (Second Edition)*, M. Carter and J. Shieh, Editors. 2015, Academic Press: San Diego. p. 117-144.

51. Hellmann, N. and D. Schneider, *Hands On: Using Tryptophan Fluorescence Spectroscopy to Study Protein Structure*, in *Protein Supersecondary Structures: Methods and Protocols*, A.E. Kister, Editor. 2019, Springer New York: New York, NY. p. 379-401.

52. Cheng, Z., et al., *Fluorescent amino acids as versatile building blocks for chemical biology.* Nature Reviews Chemistry, 2020. **4**(6): p. 275-290.

53. Ghisaidoobe, A.B. and S.J. Chung, *Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques.* Int J Mol Sci, 2014. **15**(12): p. 22518-38.

54. Moon, C.P. and K.G. Fleming, *Chapter six - Using Tryptophan Fluorescence to Measure the Stability of Membrane Proteins Folded in Liposomes*, in *Methods in Enzymology*, M.L. Johnson, J.M. Holt, and G.K. Ackers, Editors. 2011, Academic Press. p. 189-211.

55. Sindrewicz, P., et al., *Intrinsic tryptophan fluorescence spectroscopy reliably determines galectin-ligand interactions.* Scientific Reports, 2019. **9**(1): p. 11851.

56. Yang, H., et al. *Intrinsic Fluorescence Spectra of Tryptophan, Tyrosine and Phenyloalanine*. 2015. Atlantis Press.

57. COONS, A.H., et al., *The Demonstration of Pneumococcal Antigen in Tissues by the Use of Fluorescent Antibody.* The Journal of Immunology, 1942. **45**(3): p. 159-170.

58. van den Wildenberg, S.M.J.L., B. Prevo, and E.J.G. Peterman, *A Brief Introduction to Single-Molecule Fluorescence Methods*, in *Single Molecule Analysis: Methods and Protocols*, E.J.G. Peterman, Editor. 2018, Springer New York: New York, NY. p. 93-113.

59. Roser, P., et al., *Site-directed spin labeling of proteins for distance measurements in vitro and in cells.* Organic & Biomolecular Chemistry, 2016. **14**(24): p. 5468-5476.

60. Jeschke, G., *DEER Distance Measurements on Proteins.* Annual Review of Physical Chemistry, 2012. **63**(1): p. 419-446.

61. Klug, C.S. and J.B. Feix, *Methods and Applications of Site-Directed Spin Labeling EPR Spectroscopy*, in *Methods in Cell Biology*. 2008, Academic Press. p. 617-658.

62. Martorana, A., et al., *Probing Protein Conformation in Cells by EPR Distance Measurements using Gd3+ Spin Labeling.* Journal of the American Chemical Society, 2014. **136**(38): p. 13458-13465.

63. Torricella, F., et al., *Nitroxide spin labels and EPR spectroscopy: A powerful association for protein dynamics studies.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2021. **1869**(7): p. 140653.

64. Hu, Y., et al., *NMR-Based Methods for Protein Analysis.* Analytical Chemistry, 2021. **93**(4): p. 1866-1879.

65. Deiters, A., B.H. Geierstanger, and P.G. Schultz, *Site-Specific in vivo Labeling of Proteins for NMR Studies.* ChemBioChem, 2005. **6**(1): p. 55-58.

66. Reardon, P.N. and L.D. Spicer, *Multidimensional NMR Spectroscopy for Protein Characterization and Assignment inside Cells.* Journal of the American Chemical Society, 2005. **127**(31): p. 10848-10849.

67. Vogl, D.P., A.C. Conibear, and C.F.W. Becker, *Segmental and site-specific isotope labelling strategies for structural analysis of posttranslationally modified proteins.* RSC Chemical Biology, 2021. **2**(5): p. 1441-1461.

68. Chu, F., D.T. Thornton, and H.T. Nguyen, *Chemical cross-linking in the structural analysis of protein assemblies.* Methods, 2018. **144**: p. 53-63.

69. Piersimoni, L., et al., *Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions—A Method for All Seasons.* Chemical Reviews, 2022. **122**(8): p. 7500-7531.

70. Sutherland, B.W., J. Toews, and J. Kast, *Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions.* J Mass Spectrom, 2008. **43**(6): p. 699-715.

71. Rey, M., et al., *Advanced In Vivo Cross-Linking Mass Spectrometry Platform to Characterize Proteome-Wide Protein Interactions.* Analytical Chemistry, 2021. **93**(9): p. 4166-4174.

72. Spicer, C.D. and B.G. Davis, *Selective chemical protein modification.* Nature Communications, 2014. **5**(1): p. 4740.

73. Sletten, E.M. and C.R. Bertozzi, *Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality.* Angewandte Chemie International Edition, 2009. **48**(38): p. 6974-6998.

74. Chen, F.-J., et al., *Fast and Cysteine-Specific Modification of Peptides, Proteins and Bacteriophage Using Chlorooximes.* Chemistry – A European Journal, 2022. **28**(20): p. e202200058.

75. Wiedemann, C., et al., *Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR.* Frontiers in Chemistry, 2020. **8**.

76. Zhang, Y., et al., *Cysteine-specific protein multi-functionalization and disulfide bridging using 3-bromo-5-methylene pyrrolones.* Nature Communications, 2020. **11**(1): p. 1015.

77. Gorissen, S.H.M., et al., *Protein content and amino acid composition of commercially available plant-based protein isolates.* Amino acids, 2018. **50**(12): p. 1685-1695.

78. Vantourout, J.C., et al., *Serine-Selective Bioconjugation.* Journal of the American Chemical Society, 2020. **142**(41): p. 17236-17242.

79. Dixon, H.B.F., *N-terminal modification of proteins—a review.* Journal of Protein Chemistry, 1984. **3**(1): p. 99-108.

80. Dawson, P.E., et al., *Synthesis of proteins by native chemical ligation.* Science, 1994. **266**(5186): p. 776-9.

81. Muralidharan, V. and T.W. Muir, *Protein ligation: an enabling technology for the biophysical analysis of proteins.* Nature Methods, 2006. **3**(6): p. 429-438.

82. Hayashi, T. and I. Hamachi, *Traceless Affinity Labeling of Endogenous Proteins for Functional Analysis in Living Cells.* Accounts of Chemical Research, 2012. **45**(9): p. 1460-1469.

83. Tsukiji, S., et al., *Ligand-directed tosyl chemistry for protein labeling in vivo.* Nat Chem Biol, 2009. **5**(5): p. 341-3.

84. Tsukiji, S. and I. Hamachi, *Ligand-directed tosyl chemistry for selective native protein labeling in vitro, in cells, and in vivo.* Methods Mol Biol, 2015. **1266**: p. 243-63.

85. Lippincott-Schwartz, J. and G.H. Patterson, *Development and use of fluorescent protein markers in living cells.* Science, 2003. **300**(5616): p. 87-91.

86. Rodriguez, E.A., et al., *The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins.* Trends Biochem Sci, 2017. **42**(2): p. 111-129.

87. Saito, K., et al., *Luminescent proteins for high-speed single-cell and whole-body imaging.* Nature Communications, 2012. **3**(1): p. 1262.

88. Saito, K. and T. Nagai, *Recent progress in luminescent proteins development.* Current Opinion in Chemical Biology, 2015. **27**: p. 46-51.

89. Suzuki, K., et al., *Five colour variants of bright luminescent protein for real-time multicolour bioimaging.* Nature Communications, 2016. **7**(1): p. 13718.

90. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.

91. Brown, N.E., J.B. Blumer, and J.R. Hepler, *Bioluminescence resonance energy transfer to detect protein-protein interactions in live cells.* Methods Mol Biol, 2015. **1278**: p. 457-65.

92. Dale, N.C., et al., *NanoBRET: The Bright Future of Proximity-Based Assays.* Frontiers in Bioengineering and Biotechnology, 2019. **7**.

93. Nissim-Rafinia, M. and E. Meshorer, *Photobleaching assays (FRAP & FLIP) to measure chromatin protein dynamics in living embryonic stem cells.* Journal of visualized experiments : JoVE, 2011(52): p. 2696.

94. Yoon, S. and P. Penzes, *A fluorescence recovery after photobleaching protocol to measure surface diffusion of DAGLα in primary cultured cortical mouse neurons.* STAR Protocols, 2022. **3**(1): p. 101118.

95. Kremers, G.J., et al., *Fluorescent proteins at a glance.* J Cell Sci, 2011. **124**(Pt 2): p. 157-60.

96. Lambert, T.J., *FPbase: a community-editable fluorescent protein database.* Nature Methods, 2019. **16**(4): p. 277-278.

97. Hoelzel, C.A. and X. Zhang, *Visualizing and Manipulating Biological Processes by Using HaloTag and SNAP-Tag Technologies.* Chembiochem, 2020. **21**(14): p. 1935-1946.

98. Gautier, A., et al., *An Engineered Protein Tag for Multiprotein Labeling in Living Cells.* Chemistry & Biology, 2008. **15**(2): p. 128-136.

99. Hori, Y. and K. Kikuchi, *Protein labeling with fluorogenic probes for no-wash live-cell imaging of proteins.* Current Opinion in Chemical Biology, 2013. **17**(4): p. 644-650.

100. Andresen, M., R. Schmitz-Salue, and S. Jakobs, *Short Tetracysteine Tags to β-Tubulin Demonstrate the Significance of Small Labels for Live Cell Imaging.* Molecular Biology of the Cell, 2004. **15**(12): p. 5616-5622.

101. Robitaille, A.M., et al., *Quantitative phosphoproteomics reveal mTORC1 activates de novo pyrimidine synthesis.* Science, 2013. **339**(6125): p. 1320-3.

102. Yu, Y., et al., *Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling.* Science, 2011. **332**(6035): p. 1322-6.

103. Jia, J.J., et al., *mTORC1 promotes TOP mRNA translation through site-specific phosphorylation of LARP1.* Nucleic Acids Res, 2021. **49**(6): p. 3461-3489.

104. Wang, J., et al., *Mass spectrometry-based proteomic analysis reveals the interacting partners of lipin1.* IUBMB Life, 2018. **70**(8): p. 753-762.

105. Martina, J.A. and R. Puertollano, *Rag GTPases mediate amino acid-dependent recruitment of TFEB and MITF to lysosomes.* J Cell Biol, 2013. **200**(4): p. 475-91.

106. Strack, R., *Protein labeling in cells.* Nature Methods, 2016. **13**(1): p. 33-33.

107. McKay, Craig S. and M.G. Finn, *Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation.* Chemistry & Biology, 2014. **21**(9): p. 1075-1101.

108. Lotze, J., et al., *Peptide-tags for site-specific protein labelling in vitro and in vivo.* Molecular BioSystems, 2016. **12**(6): p. 1731-1745.

109. McCann, C.M., et al., *Peptide tags for labeling membrane proteins in live cells with multiple fluorophores.* Biotechniques, 2005. **38**(6): p. 945-52.

110. Kimple, M.E., A.L. Brill, and R.L. Pasker, *Overview of affinity tags for protein purification.* Current protocols in protein science, 2013. **73**: p. 9.9.1-9.9.23.

111. Pina, A.S., et al., *Affinity Tags in Protein Purification and Peptide Enrichment: An Overview*, in *Protein Downstream Processing: Design, Development, and Application of High and Low-Resolution Methods*, N.E. Labrou, Editor. 2021, Springer US: New York, NY. p. 107-132.

112. Chen, X., et al., *Quantitative proteomics using SILAC: Principles, applications, and developments.* PROTEOMICS, 2015. **15**(18): p. 3175-3192.

113. Suchanek, M., A. Radzikowska, and C. Thiele, *Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells.* Nature Methods, 2005. **2**(4): p. 261-268.

114. Yang, T., et al., *Photo-lysine captures proteins that bind lysine post-translational modifications.* Nature Chemical Biology, 2016. **12**(2): p. 70-72.

115. Hou, W., X. Zhang, and C.-F. Liu, *Progress in Chemical Synthesis of Peptides and Proteins.* Transactions of Tianjin University, 2017. **23**(5): p. 401-419.

116. Cornish, V.W., D. Mendel, and P.G. Schultz, *Probing Protein Structure and Function with an Expanded Genetic Code.* Angewandte Chemie International Edition in English, 1995. **34**(6): p. 621-633.

117. Albayrak, C. and J.R. Swartz, *Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation.* Nucleic Acids Res, 2013. **41**(11): p. 5949-63.

118. Dougherty, D.A. and E.B. Van Arnam, *In vivo incorporation of non-canonical amino acids by using the chemical aminoacylation strategy: a broadly applicable mechanistic tool.* Chembiochem, 2014. **15**(12): p. 1710-20.

119. Patterson, D.M., L.A. Nazarova, and J.A. Prescher, *Finding the Right (Bioorthogonal) Chemistry.* ACS Chemical Biology, 2014. **9**(3): p. 592-605.

120. Cornish, V.W., K.M. Hahn, and P.G. Schultz, *Site-Specific Protein Modification Using a Ketone Handle.* Journal of the American Chemical Society, 1996. **118**(34): p. 8150-8151.

121. Zhang, Z., et al., *A New Strategy for the Site-Specific Modification of Proteins in Vivo.* Biochemistry, 2003. **42**(22): p. 6735-6746.

122. Zeng, Y., et al., *High-efficiency labeling of sialylated glycoproteins on living cells.* Nat Methods, 2009. **6**(3): p. 207-9.

123. Carrico, I.S., B.L. Carlson, and C.R. Bertozzi, *Introducing genetically encoded aldehydes into proteins.* Nat Chem Biol, 2007. **3**(6): p. 321-2.

124. Chen, I., et al., *Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase.* Nat Methods, 2005. **2**(2): p. 99-104.

125. Griffin, R.J., *3 The Medicinal Chemistry of the Azido Group*, in *Progress in Medicinal Chemistry*, G.P. Ellis and D.K. Luscombe, Editors. 1994, Elsevier. p. 121-232.

126. Saxon, E. and C.R. Bertozzi, *Cell Surface Engineering by a Modified Staudinger Reaction.* Science, 2000. **287**(5460): p. 2007-2010.

127. van Berkel, S.S., M.B. van Eldijk, and J.C.M. van Hest, *Staudinger ligation as a method for bioconjugation.* Angewandte Chemie (International ed. in English), 2011. **50**(38): p. 8806-8827.

128. Prescher, J.A., D.H. Dube, and C.R. Bertozzi, *Chemical remodelling of cell surfaces in living animals.* Nature, 2004. **430**(7002): p. 873-7.

129. Hang, H.C., et al., *Chemical Probes for the Rapid Detection of Fatty-Acylated Proteins in Mammalian Cells.* Journal of the American Chemical Society, 2007. **129**(10): p. 2744-2745.

130. Lemieux, G.A., C.L. de Graffenried, and C.R. Bertozzi, *A Fluorogenic Dye Activated by the Staudinger Ligation.* Journal of the American Chemical Society, 2003. **125**(16): p. 4708-4709.

131. Hangauer, M.J. and C.R. Bertozzi, *A FRET-based fluorogenic phosphine for live-cell imaging with the Staudinger ligation.* Angew Chem Int Ed Engl, 2008. **47**(13): p. 2394-7.

132. Dieterich, D.C., et al., *Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging.* Nat Protoc, 2007. **2**(3): p. 532-40.

133. Jewett, J.C. and C.R. Bertozzi, *Cu-free click cycloaddition reactions in chemical biology.* Chemical Society Reviews, 2010. **39**(4): p. 1272-1279.

134. Baskin, J.M., et al., *Copper-free click chemistry for dynamic in vivo imaging.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 16793-7.

135. Chang, P.V., et al., *Copper-free click chemistry in living animals.* Proc Natl Acad Sci U S A, 2010. **107**(5): p. 1821-6.

136. McKay, C.S., et al., *Strain-promoted cycloadditions of cyclic nitrones with cyclooctynes for labeling human cancer cells.* Chemical Communications, 2011. **47**(36): p. 10040-10042.

137. Rossin, R., et al., *Highly Reactive trans-Cyclooctene Tags with Improved Stability for Diels–Alder Chemistry in Living Systems.* Bioconjugate Chemistry, 2013. **24**(7): p. 1210-1217.

138. Devaraj, N.K., et al., *Reactive polymer enables efficient in vivo bioorthogonal chemistry.* Proceedings of the National Academy of Sciences, 2012. **109**(13): p. 4762-4767.

139. Liu, C.C. and P.G. Schultz, *Adding New Chemistries to the Genetic Code.* Annual Review of Biochemistry, 2010. **79**(1): p. 413-444.

140. Noren, C.J., et al., *A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins.* Science, 1989. **244**(4901): p. 182-188.

141. de la Torre, D. and J.W. Chin, *Reprogramming the genetic code.* Nature Reviews Genetics, 2021. **22**(3): p. 169-184.

142. Italia, J.S., et al., *An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes.* Nat Chem Biol, 2017. **13**(4): p. 446-450.

143. Wang, L., et al., *Expanding the Genetic Code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

144. Wang, L. and P.G. Schultz, *A general approach for the generation of orthogonal tRNAs.* Chemistry & Biology, 2001. **8**(9): p. 883-890.

145. Chen, P.R., et al., *A facile system for encoding unnatural amino acids in mammalian cells.* Angew Chem Int Ed Engl, 2009. **48**(22): p. 4052-5.

146. Chatterjee, A., et al., *Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells.* Proc Natl Acad Sci U S A, 2013. **110**(29): p. 11803-8.

147. Santoro, S.W., et al., *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity.* Nature Biotechnology, 2002. **20**(10): p. 1044-1048.

148. Melançon, C.E. and P.G. Schultz, *One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases.* Bioorganic & Medicinal Chemistry Letters, 2009. **19**(14): p. 3845-3847.

149. Mukai, T., et al., *Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases.* Biochemical and Biophysical Research Communications, 2008. **371**(4): p. 818-822.

150. Chen, P.R., et al., *A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells.* Angewandte Chemie International Edition, 2009. **48**(22): p. 4052-4055.

151. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon.* Proceedings of the National Academy of Sciences, 2004. **101**(20): p. 7566-7571.

152. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon.* ACS Chem Biol, 2015. **10**(7): p. 1648-53.

153. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(20): p. 7566-7571.

154. Zhang, Y., et al., *A semi-synthetic organism that stores and retrieves increased genetic information.* Nature, 2017. **551**(7682): p. 644-647.

155. Wang, L., et al., *A New Functional Suppressor tRNA/Aminoacyl–tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

156. Gan, R., et al., *Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins.* Biotechnol Bioeng, 2017. **114**(5): p. 1074-1086.

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

158. Chatterjee, A., H. Xiao, and P.G. Schultz, *Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli.* Proceedings of the National Academy of Sciences, 2012. **109**(37): p. 14841-14846.

159. Santoro, S.W., et al., *An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli.* Nucleic Acids Research, 2003. **31**(23): p. 6700-6709.

160. Chatterjee, A., et al., *A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in E. coli.* Angew Chem Int Ed Engl, 2013. **52**(19): p. 5106-9.

161. Gan, Q., et al., *Expanding the genetic code of Salmonella with non-canonical amino acids.* Scientific Reports, 2016. **6**(1): p. 39920.

162. Singh, M.K., et al., *Genetic code expansion enables visualization of Salmonella type three secretion system components and secreted effectors.* eLife, 2021. **10**: p. e67789.

163. Wang, L., *Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts.* Acc Chem Res, 2017. **50**(11): p. 2767-2775.

164. Wang, K., et al., *Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion.* Nature Biotechnology, 2007. **25**(7): p. 770-777.

165. Chin, J.W., *Expanding and Reprogramming the Genetic Code of Cells and Animals.* Annual Review of Biochemistry, 2014. **83**(1): p. 379-408.

166. Plotkin, J.B. and G. Kudla, *Synonymous but not the same: the causes and consequences of codon bias.* Nature Reviews Genetics, 2011. **12**(1): p. 32-42.

167. Mukai, T., et al., *Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli.* Nucleic Acids Res, 2015. **43**(16): p. 8111-22.

168. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome.* Nature, 2010. **464**(7287): p. 441-444.

169. Wang, K., W.H. Schmied, and J.W. Chin, *Reprogramming the genetic code: from triplet to quadruplet codes.* Angew Chem Int Ed Engl, 2012. **51**(10): p. 2288-97.

170. Wang, K., et al., *Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET.* Nature Chemistry, 2014. **6**(5): p. 393-403.

171. Yang, Z., et al., *Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System.* Journal of the American Chemical Society, 2011. **133**(38): p. 15105-15112.

172. Hoshika, S., et al., *Hachimoji DNA and RNA: A genetic system with eight building blocks.* Science, 2019. **363**(6429): p. 884-887.

173. Malyshev, D.A., et al., *A semi-synthetic organism with an expanded genetic alphabet.* Nature, 2014. **509**(7500): p. 385-388.

174. Baghban, R., et al., *Yeast Expression Systems: Overview and Recent Advances.* Mol Biotechnol, 2019. **61**(5): p. 365-384.

175. Wu, N., et al., *A Genetically Encoded Photocaged Amino Acid.* Journal of the American Chemical Society, 2004. **126**(44): p. 14306-14307.

176. Summerer, D., et al., *A genetically encoded fluorescent amino acid.* Proceedings of the National Academy of Sciences, 2006. **103**(26): p. 9785-9789.

177. Hancock, S.M., et al., *Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair.* J Am Chem Soc, 2010. **132**(42): p. 14819-24.

178. Wang, Q. and L. Wang, *Genetic incorporation of unnatural amino acids into proteins in yeast.* Methods Mol Biol, 2012. **794**: p. 199-213.

179. Sakamoto, K., et al., *Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells.* Nucleic Acids Res, 2002. **30**(21): p. 4692-9.

180. Ai, H.W., *Biochemical analysis with the expanded genetic lexicon.* Anal Bioanal Chem, 2012. **403**(8): p. 2089-102.

181. Nodling, A.R., et al., *Using genetically incorporated unnatural amino acids to control protein functions in mammalian cells.* Single Cell Biology, 2019. **63**(2): p. 237-266.

182. Edwards, H. and P. Schimmel, *A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase.* Mol Cell Biol, 1990. **10**(4): p. 1633-41.

183. Wang, J., J. Xie, and P.G. Schultz, *A genetically encoded fluorescent amino acid.* J Am Chem Soc, 2006. **128**(27): p. 8738-9.

184. Lee, H.S., et al., *Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in Saccharomyces cerevisiae.* Journal of the American Chemical Society, 2009. **131**(36): p. 12921-12923.

185. Wang, W., et al., *Genetically encoding unnatural amino acids for cellular and neuronal studies.* Nature Neuroscience, 2007. **10**(8): p. 1063-1072.

186. Chatterjee, A., et al., *A genetically encoded fluorescent probe in mammalian cells.* J Am Chem Soc, 2013. **135**(34): p. 12540-3.

187. Mitchell, A.L., et al., *A Unique Genetically Encoded FRET Pair in Mammalian Cells.* Chembiochem, 2017. **18**(6): p. 511-514.

188. Wan, W., et al., *A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli.* Angew Chem Int Ed Engl, 2010. **49**(18): p. 3211-4.

189. Niu, W., P.G. Schultz, and J. Guo, *An Expanded Genetic Code in Mammalian Cells with a Functional Quadruplet Codon.* ACS Chemical Biology, 2013. **8**(7): p. 1640-1645.

190. Schmied, W.H., et al., *Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1.* Journal of the American Chemical Society, 2014. **136**(44): p. 15577-15583.

191. Shen, B., et al., *Genetically Encoding Unnatural Amino Acids in Neural Stem Cells and Optically Reporting Voltage-Sensitive Domain Changes in Differentiated Neurons.* Stem Cells, 2011. **29**(8): p. 1231-1240.

192. Greiss, S. and J.W. Chin, *Expanding the Genetic Code of an Animal.* Journal of the American Chemical Society, 2011. **133**(36): p. 14196-14199.

193. Parrish, A.R., et al., *Expanding the Genetic Code of Caenorhabditis elegans Using Bacterial Aminoacyl-tRNA Synthetase/tRNA Pairs.* ACS Chemical Biology, 2012. **7**(7): p. 1292-1302.

194. Bianco, A., et al., *Expanding the genetic code of Drosophila melanogaster.* Nature Chemical Biology, 2012. **8**(9): p. 748-750.

195. Li, F., et al., *Expanding the Genetic Code for Photoclick Chemistry in E. coli, Mammalian Cells, and A. thaliana.* Angewandte Chemie International Edition, 2013. **52**(37): p. 9700-9704.

196. Kang, J.-Y., et al., *In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids.* Neuron, 2013. **80**(2): p. 358-370.

197. Chen, Y., et al., *Heritable expansion of the genetic code in mouse and zebrafish.* Cell Research, 2017. **27**(2): p. 294-297.

198. Ernst, R.J., et al., *Genetic code expansion in the mouse brain.* Nature Chemical Biology, 2016. **12**(10): p. 776-778.

199. Han, S., et al., *Expanding the genetic code of Mus musculus.* Nature Communications, 2017. **8**(1): p. 14568.

200. Maywood, E.S., et al., *Translational switching of Cry1 protein expression confers reversible control of circadian behavior in arrhythmic Cry-deficient mice.* Proceedings of the National Academy of Sciences, 2018. **115**(52): p. E12388-E12397.

201. Davis, L. and J.W. Chin, *Designer proteins: applications of genetic code expansion in cell biology.* Nat Rev Mol Cell Biol, 2012. **13**(3): p. 168-82.

202. Luo, X., et al., *Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria.* Nature Chemical Biology, 2017. **13**(8): p. 845-849.

203. Hoppmann, C., et al., *Site-specific incorporation of phosphotyrosine using an expanded genetic code.* Nature Chemical Biology, 2017. **13**(8): p. 842-844.

204. He, X., et al., *Genetic Encoding of A Nonhydrolyzable Phosphotyrosine Analog in Mammalian Cells.* bioRxiv, 2021: p. 2021.11.30.470659.

205. Park, H.-S., et al., *Expanding the Genetic Code of Escherichia coli with Phosphoserine.* Science, 2011. **333**(6046): p. 1151-1154.

206. Rogerson, D.T., et al., *Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog.* Nature Chemical Biology, 2015. **11**(7): p. 496-503.

207. Beránek, V., et al., *Genetically Encoded Protein Phosphorylation in Mammalian Cells.* Cell Chemical Biology, 2018. **25**(9): p. 1067-1074.e5.

208. Zhang, M.S., et al., *Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing.* Nature Methods, 2017. **14**(7): p. 729-736.

209. Virdee, S., et al., *Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase.* Nature Chemical Biology, 2010. **6**(10): p. 750-757.

210. Virdee, S., et al., *Traceless and Site-Specific Ubiquitination of Recombinant Proteins.* Journal of the American Chemical Society, 2011. **133**(28): p. 10708-10711.

211. Fottner, M., et al., *Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase.* Nature Chemical Biology, 2019. **15**(3): p. 276-284.

212. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding Nε-acetyllysine in recombinant proteins.* Nature Chemical Biology, 2008. **4**(4): p. 232-234.

213. Elsässer, S.J., et al., *Genetic code expansion in stable cell lines enables encoded chromatin modification.* Nat Methods, 2016. **13**(2): p. 158-64.

214. Italia, J.S., et al., *Genetically encoded protein sulfation in mammalian cells.* Nature Chemical Biology, 2020. **16**(4): p. 379-382.

215. Courtney, T. and A. Deiters, *Recent advances in the optical control of protein function through genetic code expansion.* Current Opinion in Chemical Biology, 2018. **46**: p. 99-107.

216. Baker, A.S. and A. Deiters, *Optical Control of Protein Function through Unnatural Amino Acid Mutagenesis and Other Optogenetic Approaches.* ACS Chemical Biology, 2014. **9**(7): p. 1398-1407.

217. Edwards, W.F., D.D. Young, and A. Deiters, *Light-Activated Cre Recombinase as a Tool for the Spatial and Temporal Control of Gene Function in Mammalian Cells.* ACS Chemical Biology, 2009. **4**(6): p. 441-445.

218. Nguyen, D.P., et al., *Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells.* Journal of the American Chemical Society, 2014. **136**(6): p. 2240-2243.

219. Klippenstein, V., et al., *Optocontrol of glutamate receptor activity by single side-chain photoisomerization.* eLife, 2017. **6**.

220. Gautier, A., et al., *Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells.* Journal of the American Chemical Society, 2010. **132**(12): p. 4086-4088.

221. Luo, J., et al., *Genetically Encoded Optochemical Probes for Simultaneous Fluorescence Reporting and Light Activation of Protein Function with Two-Photon Excitation.* Journal of the American Chemical Society, 2014. **136**(44): p. 15551-15558.

222. Brown, W., et al., *Cell-Lineage Tracing in Zebrafish Embryos with an Expanded Genetic Code.* ChemBioChem, 2018. **19**(12): p. 1244-1249.

223. Davis, L., et al., *Precise optical control of gene expression in C elegans using improved genetic code expansion and Cre recombinase.* eLife, 2021. **10**: p. e67075.

224. Nguyen, T.-A., M. Cigler, and K. Lang, *Expanding the Genetic Code to Study Protein–Protein Interactions.* Angewandte Chemie International Edition, 2018. **57**(44): p. 14350-14361.

225. Wu, X., et al., *Site-Specific Photo-Crosslinking Proteomics Reveal Regulation of IFITM3 Trafficking and Turnover by VCP/p97 ATPase.* Cell Chemical Biology, 2020. **27**(5): p. 571-585.e6.

226. Uttamapinant, C., et al., *Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins.* J Am Chem Soc, 2015. **137**(14): p. 4602-5.

227. Sachdeva, A., et al., *Concerted, rapid, quantitative, and site-specific dual labeling of proteins.* J Am Chem Soc, 2014. **136**(22): p. 7785-8.

228. Jones, D.H., et al., *Site-specific labeling of proteins with NMR-active unnatural amino acids.* Journal of Biomolecular NMR, 2009. **46**(1): p. 89.

229. Fleissner, M.R., et al., *Site-directed spin labeling of a genetically encoded unnatural amino acid.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21637-42.

230. Schmidt, M.J., et al., *Genetically Encoded Spin Labels for In Vitro and In-Cell EPR Studies of Native Proteins.* Methods Enzymol, 2015. **563**: p. 483-502.

231. Schmidt, M.J., et al., *EPR Distance Measurements in Native Proteins with Genetically Encoded Spin Labels.* ACS Chem Biol, 2015. **10**(12): p. 2764-71.

232. Esposito, D., L.A. Garvey, and C.S. Chakiath, *Gateway Cloning for Protein Expression*, in *High Throughput Protein Expression and Purification: Methods and Protocols*, S.A. Doyle, Editor. 2009, Humana Press: Totowa, NJ. p. 31-54.

233. Gibson, D.G., et al., *Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome.* Science, 2010. **329**(5987): p. 52-56.

234. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.

235. Bieniossek, C., T.J. Richmond, and I. Berger, *MultiBac: multigene baculovirus-based eukaryotic protein complex production.* Curr Protoc Protein Sci, 2008. **Chapter 5**: p. Unit 5.20.

236. Carpenter, A.E., et al., *CellProfiler: image analysis software for identifying and quantifying cell phenotypes.* Genome Biol, 2006. **7**(10): p. R100.

237. Kamentsky, L., et al., *Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software.* Bioinformatics, 2011. **27**(8): p. 1179-80.

238. McQuin, C., et al., *CellProfiler 3.0: Next-generation image processing for biology.* PLOS Biology, 2018. **16**(7): p. e2005970.

239. Stirling, D.R., A.E. Carpenter, and B.A. Cimini, *CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis.* Bioinformatics, 2021. **37**(21): p. 3992-3994.

240. Petersen, L.K. and R.S. Stowers, *A Gateway MultiSite Recombination Cloning Toolkit.* PLOS ONE, 2011. **6**(9): p. e24531.

241. Scaiola, A., et al., *The 3.2-Å resolution structure of human mTORC2.* Sci Adv, 2020. **6**(45).

242. Fitzgerald, D.J., et al., *Protein complex expression by using multigene baculoviral vectors.* Nat Methods, 2006. **3**(12): p. 1021-32.

243. Boyce, F.M. and N.L. Bucher, *Baculovirus-mediated gene transfer into mammalian cells.* Proceedings of the National Academy of Sciences, 1996. **93**(6): p. 2348-2352.

244. Hofmann, C., et al., *Efficient gene transfer into human hepatocytes by baculovirus vectors.* Proceedings of the National Academy of Sciences, 1995. **92**(22): p. 10099-10103.

245. Barsoum, J., et al., *Efficient transduction of mammalian cells by a recombinant baculovirus having the vesicular stomatitis virus G glycoprotein.* Hum Gene Ther, 1997. **8**(17): p. 2011-8.

246. Albertazzi, L., et al., *Quantitative FRET analysis with the EGFP-mCherry fluorescent protein pair.* Photochem Photobiol, 2009. **85**(1): p. 287-97.

247. Cohen, S. and E. Arbely, *Single-Plasmid-Based System for Efficient Noncanonical Amino Acid Mutagenesis in Cultured Mammalian Cells.* Chembiochem, 2016. **17**(11): p. 1008-11.

248. Schmied, W.H., et al., *Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1.* J Am Chem Soc, 2014. **136**(44): p. 15577-83.

249. Zhou, W., et al., *Genetic code expansion in mammalian cells: A plasmid system comparison.* Bioorganic & Medicinal Chemistry, 2020. **28**(24): p. 115772.

250. Zheng, Y., et al., *Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells.* Chem Sci, 2017. **8**(10): p. 7211-7217.

251. Mansouri, M. and P. Berger, *Baculovirus for gene delivery to mammalian cells: Past, present and future.* Plasmid, 2018. **98**: p. 1-7.

252. Dukkipati, A., et al., *BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies.* Protein Expr Purif, 2008. **62**(2): p. 160-70.

253. Heikura, T., et al., *Baculovirus-mediated vascular endothelial growth factor-DΔNΔC gene transfer induces angiogenesis in rabbit skeletal muscle.* The Journal of Gene Medicine, 2012. **14**(1): p. 35-43.

254. Koehler, C. and E.A. Lemke, *MultiBacTAG-Genetic Code Expansion Using the Baculovirus Expression System in Sf21 Cells.* Noncanonical Amino Acids: Methods and Protocols, 2018. **1728**: p. 297-311.

255. Schinn, S.-M., et al., *Rapid in vitro screening for the location-dependent effects of unnatural amino acids on protein expression and activity.* Biotechnology and Bioengineering, 2017. **114**(10): p. 2412-2417.

256. Wandrey, G., et al., *Probing unnatural amino acid integration into enhanced green fluorescent protein by genetic code expansion with a high-throughput screening platform.* Journal of Biological Engineering, 2016. **10**(1): p. 11.

257. Chemla, Y., et al., *Context effects of genetic code expansion by stop codon suppression.* Current Opinion in Chemical Biology, 2018. **46**: p. 146-155.

258. Xiao, H., et al., *Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells.* Angewandte Chemie International Edition, 2013. **52**(52): p. 14080-14083.

259. Willis, J.C.W. and J.W. Chin, *Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs.* Nature Chemistry, 2018. **10**(8): p. 831-837.

260. Dunkelmann, D.L., et al., *Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids.* Nature Chemistry, 2020. **12**(6): p. 535-544.

261. Meineke, B., et al., *Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells.* Cell Rep, 2020. **31**(12): p. 107811.

262. Sánchez-Trasviña, C., et al., *Purification of Modified Therapeutic Proteins Available on the Market: An Analysis of Chromatography-Based Strategies.* Frontiers in Bioengineering and Biotechnology, 2021. **9**.

263. Lee, D., et al., *Stimuli-Responsive Polypeptides for Biomedical Applications.* Polymers, 2018. **10**(8): p. 830.

264. Ma, H., et al., *A highly efficient multifunctional tandem affinity purification approach applicable to diverse organisms.* Mol Cell Proteomics, 2012. **11**(8): p. 501-11.

IX. APPENDIX

APPENDIX 1. DNA SEQUENCES OF PLASMIDS AND PRIMERS

A. List of Acquired Plasmids and Synthetic Gene Constructs

PLASMID NAME & SOURCE	PLASMID SEQUENCE
SYN1.1 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag* *acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac* *catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa* *gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt* *cccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctcgcgaatgcatctagatgcaacaaacttctcactactcaaa* *caagcaggtgacgtggaggagaatcccgggccttctagaggatccggctccggcgagggcaggggaagtctactaacatgcggggacgt* *ggaggaaaatcccggcccagtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacg* *gccacaagttcagcgtgtccggcgagggcgagggcgatgccacctaaggcaagctgaccctgaagttcatctgcaccaccggcaagctgc* *ccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttca* *agtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcg* *agggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaac* *tacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacgg* *cagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccag* *tccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacg* *agctgtacaagggttcaactagtggttctggtcaccaccaccaccaccaccaccaccaccaccacggttcaactagtggttctggtgagcagaag* *ctgatctccgaggaggacctgggttcaactagtggttctggtgactacaaggacgacgacgacaagctttgatagtaatagtaatgatagt* *aatgaatcggatcccgggcccgtcgactgcagaggcctgcatgcaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttat* *ccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgtt* *gcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattg* *ggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggtt* *atccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgct* *ggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaaga* *taccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagc* *gtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcc* *cgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacagga* *ttagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgc* *gctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgca* *agcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacg* *ttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatat* *gagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccc* *cgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatt* *tatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgg* *gaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatgg* *cttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgtt* *gtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgt* *gactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgc* *cacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcga* *tgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaa* *aaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgag* *cggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaacc* *attattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc*
SYN1.4 Circular Plasmid Synthetic Gene (*Genscript*)	*gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgt* *gtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagg* *gttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggt* *cattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattga* *cgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttgg* *cagtacatcaagtgtatcatatgccaagtacgcccccattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgacct* *tatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggata* *gcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgt* *cgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaac* *ccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacttaagcttggtaccgagctcgg* *atccgcaacaaacttctcactactcaaacaagcaggtgacgtggaggagaatcccgggccttctagaggatccggctccggcgagggcag* *gggaagtctactaacatgcggggacgtggaggaaaatcccggcccagtgagcaagggcgaggagctgttcaccggggtggtgcccatcc* *tggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctg* *aagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctacccc* *gaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaact* *acaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaac* *atcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttc*

	aagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctg cccgacaaccactacctgagcacccagtccgccctgagcaaagacccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccg ccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtggttctggtcaccacccaccaccaccaccaccaccacggt tcaactagtggttctggtgagcagaagctgatctccgaggaggacctgggttcaactagtggttctggtgactacaaggacgacgacgaca agctttagtaatgatagtaatgatagtaatgaggatccactagtccagtgtggtggaattctgcagatatccagcacagtggcggccgctcg agtctagagggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttga ccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtgg ggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaa agaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgacc gctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggg gcatcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctg atagacggtttttcgccctttgacgttggagtccacgttcttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtcta ttcttttgatttataagggatttttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaa tgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtg tggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgccca tcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcag cacgtgatgaaaaagcctgaactcaccgcgacgtctgtcgagaagtttctgatcgaaaagttcgacagcgtctccgacctgatgcagctctc ggagggcgaagaatctcgtgctttcagcttcgatgtaggagggcgtggatatgtcctgcgggtaaatagctgcgccgatggtttctacaaag atcgttatgtttatcggcactttgcatcggccgcgctcccgattccggaagtgcttgacattggggaattcagcgagagcctgacctattgcat ctcccgccgtgcacagggtgtcacgttgcaagacctgcctgaaaccgaactgcccgctgttctgcagccggtcgcggaggccatggatgcga tcgctgcggccgatcttagccagacgagcgggttcggcccattcggaccgcaaggaatcggtcaatacactacatggcgtgatttcatatgc gcgattgctgatccccatgtgtatcactggcaaactgtgatggacgacaccgtcagtgcgtccgtcgcgcaggctctcgatgagctgatgctt tgggccgaggactgccccgaagtccggcacctcgtgcacgcggatttcggctccaacaatgtcctgacggacaatggccgcataacagcgg tcattgactggagcgaggcgatgttcggggattcccaatacgaggtcgccaacatcttcttctggaggccgtggttggcttgtatggagcagc agacgcgctacttcgagcggaggcatccggagcttgcaggatcgccgcggctccgggcgtatatgctccgcattggtcttgaccaactctatc agagcttggttgacggcaatttcgatgatgcagcttgggcgcagggtcgatgcgacgcaatcgtccgatccggagccgggactgtcgggcg tacacaaatcgcccgcagaagcgcggccgtctggaccgatggctgtgtagaagtactcgccgatagtggaaaccgacgccccagcactcgt ccgagggcaaaggaatagcacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgc cggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagc aatagcatcacaaatttcacaaataaagcattttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtatac cgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagc cggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaa acctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactc gctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcagga aagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgac gagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctc gtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtagg tatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcg tcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgct acagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaa aagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaa ggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatca aaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaat gcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggag ggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaa gggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagtta atagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatca aggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgtta tcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattc tgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcat cattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatc ttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaa atgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaa aataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtc
SYN2.1 Circular Plasmid Synthetic Gene (*Genscript*)	*aaagtagccgaagatgacggtttgtcacatggagttggcaggatgtttgattaaaaacataacaggaagaaaaatgccccgctgtgggcg gacaaaatagttgggaactgggaggggtggaaatggagttttaaggattatttagggaagagtgacaaaatagatgggaactgggtgt agcgtcgtaagctaatacgaaaattaaaaatgacaaaatagtttggaactagatttcacttatctggttcggatctccttaggctcaagcagt gatcagatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgt gatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcagggggagg tgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctctagtacttctcgacaagcttgtcgagactgc aggctctagattcgaaagcggccgcgactagtgagctcgtcgacgtaggcctttgaattccgcgcgcttcggaccgggatccggtaccataa cttcgtatagcatacattatacgaagttatctgtgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaatta cggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcc cattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgccc acttggcagtacatcaagtgtatcatatgccaagtacgcccccattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtaca tgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgt ggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttcca aaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaacta gagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggta ccgcgggcccgggatccaccggtatctattagaaaggatccgcgcccgatggtgggacggtatgaataatccggaatatttataggttttttt attacaaaactgttacgaaaacagtaaaatacttatttatttgcgagatggttatcattttaattatctccatgatctattaatattccggagtat*

	acggacctttaattcaacccaacacaatatattatagttaaataagaattattatcaaatcatttgtatattaattaaaatactatactgtaaa ttacattttatttacaatcactcgacgaagacttgatcacccgggatctcgagccatggtgctagcagctgatgcatagcatgcggtaccggg agatgggggaggctaactgaaacacggaaggagacaataccggaaggaacccgcgctatgacggcaataaaaagacagaataaaac gcacgggtgttgggtcgtttgttcataaacgcggggttcggtcccagggctggcactctgtcgatacccccaccgagaccccattgggaccaat acgcccgcgtttcttccttttccccaccccaaccccccaagttcgggtgaaggcccagggctcgcagccaacgtcggggcggcaagccctgcca tagccactacgggtacgtaggccaaccactagaactatagctagagtcctgggcgaacaaacgatgctcgccttccagaaaaccgaggat gcgaaccacttcatccggggtcagcaccaccggcaagcgccgcgacggccgaggtctaccgatctcctgaagcaagggcagatccgtgca cagcaccttgccgtagaagaacagcaaggccgccaatgcctgacgatgcgtggagaccgaaaccttgcgctcgttcgccagccaggacag aaatgcctcgacttcgctgctgcccaaggttgccgggtgacgcacaccgtggaaacggatgaaggcacgaacccagttgacataagcctgt tcggttcgtaaactgtaatgcaagtagcgtatgcgctcacgcaactggtccagaaccttgaccgaacgcagcggtggtaacggcgcagtgg cggttttcatggcttgttatgactgttttttgtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgt tatggagcagcaacgatgttacgcagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaggtggctcaagtatgggc atcattcgcacatgtaggctcggccctgaccaagtcaaatccatgcgggctgctcttgatcttttcggtcgtgagttcggagacgtagccacct actcccaacatcagccggactccgattacctcgggaacttgctccgtagtaagacattcatcgcgcttgctgccttcgaccaagaagcggttg ttggcgctctcgcggcttacgttctgcccaggtttgagcagccgcgtagtgagatctatatctatgatctcgcagtctccggcgagcaccggag gcagggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcgcttggtgcttatgtgatctacgtgcaagcagattacggtg acgatcccgcagtggctctctatacaaagttgggcatacgggaagaagtgatgcactttgatatcgacccaagtaccgccacctaacaattc gttcaagccgagatcggcttcccggccgcggagttgttcggtaaattgtcacaacgccgcgaatatagtctttaccatgcccttggccacgcc cctctttaatacgacgggcaatttgcacttcagaaaatgaagagtttgctttagccataacaaaagtccagtatgcttttcacagcataactg gactgatttcagtttacaactattctgtctagtttaagactttattgtcatagtttagatctattttgttcagtttaagactttattgtccgcccacac ccgcttacgcagggcatccatttattactcaaccgtaaccgattttgccaggttacgcggctggtctgcggtgtgaaataccgcacagatgcg taaggagaaaataccgcatcaggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctc actcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaa accgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcg aaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctg tccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgt gtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactg gcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacacta gaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgct ggtagcggtggttttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgac gctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaag ttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttc gttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgag acccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctcc atccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtgg tgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggt tagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcat gccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcg tcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatctt accgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaa caggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcat ttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagt gccacctgaaattgtaaacgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaaccaataggccgaaatcggca aaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtggactc caacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgt aaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaag aaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgcta cagggcgcgtcccattcgccattcaggctgcaaataagcgttgatattcagtcaattacaaacattaataacgaagagatgacagaaaaat tttcattctgtgacagagaa
SYN2.2 Circular Plasmid Synthetic Gene (*Genscript*)	*gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgt gtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagg gttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggt cattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattga cgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttgg cagtacatcaagtgtatcatatgccaagtacgcccccattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgacct tatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggata gcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgt cgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaac ccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacttaagcttggtaccgagctcgg atccggttcaactagtggttctggtcaccaccaccaccaccaccaccaccaccacggttcaactagtggttctggtgagcagaagctgatctc cgaggaggacctgggttcaactagtggttctggtgactacaaggacgacgacgacaagcttgcaacaaacttctcactactcaaacaagca ggtgacgtggaggagaatcccgggccttctagaggatccggctccggcgagggcaggggaagtctactaacatgcggggacgtggagga aaatcccggcccagtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccaca agttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgcc ctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgc catgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcga caccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaaca gccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtg cagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccct gagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgta caagtagtaatgatagtaatgatagtaatgaggatccactagtccagtgtggtggaattctgcagatatccagcacagtggcggccgctcg agtctagagggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttga*

	ccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtgg ggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaa agaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgacc gctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggg gcatcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctg atagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtcta ttcttttgatttataagggatttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaa tgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtg tggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgccca tcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccatttcggatctgatcag cacgtgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatggccaagttgaccagtgc cgttccggtgctcaccgcgcgcgacgtcgccggagcggtcgagttctggaccgaccggctcgggttctcccgggacttcgtggagggacgact tcgccggtgtggtccgggacgacgtgaccctgttcatcagcgcggtccaggaccaggtggtgccggacaacaccctggcctgggtgtgggt gcgcggcctggacgagctgtacgccgagtggtcggaggtcgtgtccacgaacttccgggacgcctccgggccggccatgaccgagatcgg cgagcagccgtgggggcgggagttcgccctgcgcgacccggccggcaactgcgtgcacttcgtggccgaggagcaggactgacacgtgct acgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggga tctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaa gcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgta atcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctgggg tgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatc ggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcga gcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagc aaaaggccaggaaccgtaaaaaggccgcgttgctggcgttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaag tcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgct taccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctc caagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacga cttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaac tacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaa acaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttc tacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaa ttaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagc gatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaa tgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaa ctttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgct acaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtg caaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataa ttctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagtt gctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaa ctctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctg ggtgagcaaaaacaggaaggcaaaatgccgcaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaat attattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacat ttccccgaaaagtgccacctgacgtc
SYN3.1 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattctttctaatagatacggtggatcccgggcccgcggtaccgtcgactgcagaatt cgaagcttgagctcgagatctgagtccggtagcgctagccctgcaggtgataagccagtaagcagtgggttctctagttagccagagagct ctgcttatatagacctcccaccgtacacgcctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccgttgattttg gtgccaaaacaaactcccattgacgtcaatggggtggagacttggaaatccccgtgagtcaaaccgctatccacgcccattgatgtactgcc aaaaccgcatcaccatggtaatagcgatgactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatg ccaggcgggccatttaccgtcattgacgtcaataggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaa atagtccacccattgacgtcaatggaaagtccctattggcgttactatgggaacatacgtcattattgacgtcaatgggcgggggtcgttggg cggtcagccaggcgggccatttaccgtaagttatgtaacgcggaactccatatatgggctatgaactaatgaccccgtaattgattactatta ataactagtcaataatcaatgtcaacgcgtatatctggcccgtacatctattaatattccggagtaggtcgcgaatcgatactagtaaaaccc atgtgcctggcagataacttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctg ttgatagtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggcta acgtactaagctctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaac ttaaatagcctctaaggttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaag ccagggatgtaacgcactgagaagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttc acgctgccgcaagcactcagggcgcaagggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaac tttgtatagaaaagttgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaacagactacataatact gtaaaacacaacatatccagtcactatgaatcaactacttagatggtattcggccgcattaggcacccccaggctttacactttatgcttccgg ctcgtataatgtgtggattttgagttaggatccgtcgagatttcaggagctaaggaagctaaaatggagaaaaaaatcactggatatacc accgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggata ttacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccgga attccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttc atcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttc cctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaactt cttcgccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgat ggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaatctagaggatccggcttactaa*

	aagccagataacagtatgcgtatttgcgcgctgatttttgcggtataagaatatatactgatatgtatacccgaagtatgtcaaaaagaggt atgctatgaagcagcgtattacagtgacagttgacagcgacagctatcagttgctcaaggcatatatgatgtcaatatctccggttcggtaa gcacaaccatgcagaatgaagcccgtcgtctgcgtgccgaacgctggaaagcggaaaatcaggaagggatggctgaggtcgcccggttt attgaaatgaacggctcttttgccgacgagaacaggggctggtgaaatgcagtttaaggtttacacctataaaagagagagccgttatcgt ctgtttgtggatgtacagagtgatattattgacacgcccgggcgacggatggtgatccccctggccagtgcacgtctgctgtcagataaagtc ccccgtgaactttacccggtggtgcatatcggggatgaaagctggcgcatgatgaccaccgatatggccagtgtgccggtctccgttatcgg ggaagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaacctgatgttctggggaatataaatgtcaggctcccctt atacacagccagtctgcaggtcgatctaagtagttgattcatagtgactggatatgttgtgttttacagtattatgtagtctgttttttatgcaaa atctaatttaatatattgatatttatatcattttacgtttctcgttcaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaa actgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgctttt gaagctagaactagtggatccgcgccaaaaccgcggtctagagtcgacctgcagatctagatgcattcgcgaggtaatcggatcccgggcc cgtcgactgcagaggcctgcatgcaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaac atacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttcca gtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgct cactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggata acgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgc ccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctgga agctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcac gctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccgg taactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgta ggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttacc ttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcag aaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatga gattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgaca gttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacg atacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagcc agccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagtt cgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttccc aacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccg cagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaacca agtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaa gtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccc aactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcg acacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgt atttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaac ctataaaaataggcgtatcacgaggccctttcgtc
SYN4.1.1 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctcgcgaatgcatctagatgggggacaactttgtatacaaaa gttgtgcaatatggacaacttcttcgcccccctcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataa gttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttaaa tccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccat aaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgccc ggatggtggaatcggtagacacaagggatttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatcta atatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagag acttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttaaatccctcggcgttcgcgctgt gcgggttcaagtcccgctccgggtatttttggatctaatatttgcatggttggcaccatgggcggatattatatcgctatgtgttctgggaaat caccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagaca ccgcccggatggtggaatcggtagacacaagggatttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttgg atctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtga tagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttaaatccctcggcgttcgc gctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccc tatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatc ggtagacacaagggatttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgcag ccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtcccctatcggatcccgggcccgtcgactgcagaggcctgcatgcaa gcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgta aagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgc attaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcg gctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaa aaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatc gacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccg accctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtag gtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgg taagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtg gtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttg atccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcct ttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctag atccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcac ctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggcccc agtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagt*

	ggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgtt gccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccc catgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagc actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcg gggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcac cagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactct tccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggtt ccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggcc ctttcgtc
SYN4.1.2 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctcgcgaatgcatctagatgggggacaactttgtatacaaaa gttgtgcaatatggacaacttcttcgccccctcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataa gttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaa tccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccat aaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgccc ggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatcta atatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagag acttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgt gcgggttcaagtcccgctccgggtatttttggatctaatatttgcatggttggcaccatgggcggatattatatcgctatgtgttctgggaaat caccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagaca ccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttg gatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtg atagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcg cgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatcc ctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatc ggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgcag ccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtcccctatcggatcccgggcccgtcgactgcagaggcctgcatgcaa gcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgta aagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgc attaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcg gctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaa aaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatc gacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccg accctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtag gtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgg taagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtg gtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttg atccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctt tgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctag atccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcac ctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggcccc agtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagt ggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgt tgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccc ccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagc actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcg gggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcac cagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactct tccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggtt ccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggcc ctttcgtc*
SYN4.1.3 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctcgcgaatgcatctagatgggggacaactttgtatacaaaa gttgtgcaatatggacaacttcttcgccccctcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataa gttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttcaaa tccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccat aaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgccc ggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatcta atatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagag acttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcgcgctgt gcgggttcaagtcccgctccgggtatttttggatctaatatttgcatggttggcaccatgggcggatattatatcgctatgtgttctgggaaat caccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagaca*

	ccgcccggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttg gatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtg atagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcg cgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatcc ctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatc ggtagacacaagggatttcaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgcag ccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtcccctatcggatcccgggcccgtcgactgcagaggcctgcatgcaa gcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgta aagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgc attaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcg gctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaa aaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatc gacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccg accctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtag gtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgg taagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtg gtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttg atccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcct ttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctag atccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcac ctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggcccc agtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagt ggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgt tgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccc ccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagc actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcg gggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcac cagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactct tcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggtt ccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggcc ctttcgtc
SYN4.2.1 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctcgcgaatgcatctagatggggacaactttgtatacaaaag ttgtgcaatatggacaacttcttcgcccccctcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgtt agagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgc agttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacg aatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggt ggggttcccgagcggccaaagggagcagactttaaatctgccgtcacagacttcgaaggttcgaatccttcccccaccatttttaatatttgc atgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttata agttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggactttaaatccgctccctttgggttcggcggttcg aatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttat aagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggacttt aaatccgctccctttgggttcggcggttcgaatccgtcccccctccatttttgttggcaccatgggcggatattatatcgggcaggaagagggc ctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagatattag tacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaact tgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgataga gacttataagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagactttaaatctgccgtca cagacttcgaaggttcgaatccttcccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatc agtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggct aaacgcggcggactttaaatccgctccctttgggttcggcggttcgaatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctggg aaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgataga gacaccggaggggtagcgaagtggctaaacgcggcggactttaaatccgctccctttgggttcggcggttcgaatccgtcccccctccattttt cagccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtccccatcggatcccgggcccgtcgactgcagaggcctgcatgc aagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtg taaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagct gcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgtt cggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagc aaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaa tcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttcc gaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgta ggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccg gtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagt ggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctctt gatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcc tttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctag atccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcac ctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggcccc agtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagt*

	ggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgtt gccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccc catgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagc actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcg gggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcac cagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactct tccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggtt ccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggcc ctttcgtc
SYN4.2.2 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctcgcgaatgcatctagatggggacaactttgtatacaaaag ttgtgcaatatggacaacttcttcgcccccctcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgtt agagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgc agttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacg aatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggactctaaatccgctcccttgggttcggcggttcg aatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttat aagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggactct aaatccgctcccttgggttcggcggttcgaatccgtcccccctccatttttgttggcaccatgggcggatattatatcgggcaggaagagggc ctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagatattag tacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaact tgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgataga gacttataagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagactctaaatctgccgtca cagacttcgaaggttcgaatccttccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatc agtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggct aaacgcggcggactctaaatccgctcccttgggttcggcggttcgaatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctgg gaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatag agacaccggaggggtagcgaagtggctaaacgcggcggactctaaatccgctccctttgggttcggcggttcgaatccgtcccccctccatttt ttcagccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtccccatcggatcccgggcccgtcgactgcagaggcctgcat gcaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaag tgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccag ctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcg ttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgag caaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaa atcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgtt ccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtg taggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacc cggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttga agtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagc tcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaag atcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcac ctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgag gcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctgg ccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcag aagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaac gttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatg atccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatg gcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtat gcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgtt cttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttact ttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcat actcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatag ggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacg aggccctttcgtc*
SYN4.2.3 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctcgcgaatgcatctagatggggacaactttgtatacaaaag ttgtgcaatatggacaacttcttcgcccccctcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgta aacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactat catatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagt tccctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcag acttcaaatctgccgtcacagacttcgaaggttcgaatccttccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccat*

	aaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgga ggggtagcgaagtggctaaacgcggcggacttcaaatccgctccctttgggttcggcggttcgaatccgtcccctccatttttaatatttgca tgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataa gttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggacttcaaatccgctccctttgggttcggcggttcga atccgtcccctccatttttgttggcaccatgggcggatattatatcgggcaggaagagggcctatttcccatgattccttcatatttgcatata cgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataa tttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatat cttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagac accgaaacaccggtggggttcccgagcggccaaagggagcagacttcaaatctgccgtcacagacttcgaaggttcgaatccttcccccac catttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcag tgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggacttcaaatccgctcccttt gggttcggcggttcgaatccgtcccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccatataaacgtgaaatccctatca gtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggcta aacgcggcggacttcaaatccgctcccttgggttcggcggttcgaatccgtccccctccatttttcagccaatccctgggtgagtttcaccacc caacttttctatacaaagttgtccccaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaac atacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttcca gtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgct cactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggata acgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgc ccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctgga agctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcac gctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccgg taactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgta ggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttacc ttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcag aaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatga gattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgaca gttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacg atacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagcc agccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagtt cgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttccc aacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccg cagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaacca agtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaa gtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccc aactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcg acacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgt atttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaac ctataaaaatagcgtatcacgaggccctttcgtc
SYN4.3.1 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcggggacaacttttgtatacaaaagttgtgcaatatggacaacttcttcgcccccctc gggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgta aacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactat catatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttataagttcc ctatcagtgatagagacttataagttccctatcagtgatagagacaccggaaacctgatcatgtagatcgaacggactttaaat ccgttcagccgggttagattcccgggtttccgtttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatc agtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggtggatcgaatagatc acacggactttaaatccgtgcaggcgggtgaaactcccgcacccccgttttttaatatttgcatgtcgctatgtgttctgggaaatcaccata aacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgggg ggtggatcgaatagatcacacggactttaaatccgtgcaggcgggtgaaactcccgcacccccgttttttgttggcaccatggcggatatt atatcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttga ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatgg actatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttata agttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaaacctgatcatgtagatcgaacggact ttaaatccgttcagccgggttagattcccgggtttccgtttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatc cctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggtggatcgaa tagatcacacggactttaaatccgtgcaggcgggtgaaactcccgcacccccgttttttaatatttgcatgtcgctatgtgttctgggaaatc accataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacac cgggggtggatcgaatagatcacacggactttaaatccgtgcaggcgggtgaaactcccgcacccccgttttttcagccaatccctgggt gagtttcaccacccaacttttctatacaaagttgtccccaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcac aattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctca ctgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctc ttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccaca gaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttt tccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccagg cgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgct ttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgct gcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcaga gcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgct*

	gaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagc agattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagg gattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaa acttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtg tagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagcta gagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattc agctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcaga agtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggt gagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatag cagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacc cactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagg gaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggata catatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattat catgacattaacctataaaaataggcgtatcacgaggccctttcgtc
SYN4.3.2 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcggggacaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccctc gggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgta aacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactat catatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttataagttcc ctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggaaacctgatcatgtagatcgaacggactctaaa tccgttcagccgggttagattcccggggtttccgtttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctat cagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggggggtggatcgaatagat cacacggactctaaatccgtgcaggcgggtgaaactcccgcacccccgtttttaatatttgcatgtcgctatgtgttctgggaaatcaccat aaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggg gggtggatcgaatagatcacacggactctaaatccgtgcaggcgggtgaaactcccgcacccccgtttttgttggcaccatgggcggata ttatatcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttg actgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatg gactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttat aagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggaaacctgatcatgtagatcgaacgga ctctaaatccgttcagccgggttagattcccggggtttccgtttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaa atccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggggggtggatc gaatagatcacacggactctaaatccgtgcaggcgggtgaaactcccgcacccccgtttttaatatttgcatgtcgctatgtgttctgggaa atcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagaga caccggggggtggatcgaatagatcacacggactctaaatccgtgcaggcgggtgaaactcccgcacccccgtttttcagccaatccctg ggtgagtttcaccacccaacttttctatacaaagttgtccccaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgct cacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgc tcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcg ctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatcca cagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgt ttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatacca ggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggc gctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgacc gctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagc agagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctct gctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagca gcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaa gggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagt aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcg tgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatca gcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagc tagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcat tcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcag aagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactg gtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacat agcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaa cccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa gggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcgga tacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattatt atcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc*
SYN4.3.3 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcggggacaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccctc gggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgta aacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactat*

	catatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttataagttcc ctatcagtgatagagacttataagttccctatcagtgatagagacaccggaaacctgatcatgtagatcgaacggacttcaaa tccgttcagccgggttagattcccggggtttccgtttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctat cagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatcgaatagat cacacggacttcaaatccgtgcaggcgggtgaaactcccgcacccccgtttttttaatatttgcatgtcgctatgtgttctgggaaatcaccat aaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggg gggtggatcgaatagatcacacggacttcaaatccgtgcaggcgggtgaaactcccgcacccccgtttttttgttggcaccatgggcggata ttatatcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttg actgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatg gactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttat aagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggaaacctgatcatgtagatcgaacgga cttcaaatccgttcagccgggttagattcccggggtttccgtttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaa atccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatc gaatagatcacacggacttcaaatccgtgcaggcgggtgaaactcccgcacccccgtttttttaatatttgcatgtcgctatgtgttctgggaa atcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagaga caccgggggggtggatcgaatagatcacacggacttcaaatccgtgcaggcgggtgaaactcccgcacccccgtttttttcagccaatccctg ggtgagtttcaccacccaacttttctatacaaagttgtccccaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgct cacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgc tcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcg ctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatcca cagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaggccgcgttgctggcgt ttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaa ggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggc gctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgacc gctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagc agagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctct gctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagca gcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaa gggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagt aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcg tgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatca gcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagc tagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcat tcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcag aagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactg gtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacat agcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaa cccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa gggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcgga tacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattatt atcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc
SYN4.6.1 Circular Plasmid Synthetic Gene (*Genscript*)	tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcggggacaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccctc gctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttc cctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctccgcccct gccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttc cctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgg gtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatc cctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaat tggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatggttggcaccat gggcggatattatatcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgat agagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttgggagtt cgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgat agagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcac cggtctttaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatca ccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacacc aggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatattt gcatgcagccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtccccaagcttggcgtaatcatggtcatagctgtttcctg tgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactc acattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagagg cggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaagg cggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaa aggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgac aggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttct cccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaa ccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcc actggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaaca gtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggt ggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtgg

	aacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatca atctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccat agttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgct caccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtct attaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgct cgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctcctt cggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccg taagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacg ggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgt tgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaag gcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagg gttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctg acgtctaagaaaccattattatcatgacattaacctataaaaatagGcgtatcacgaggccctttcgtc
SYN4.6.2 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcggggacaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccctc gctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttcccctatcagtgatagagacttataagttc cctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctccgcccct tgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagtt ccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccg ggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaa tccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttca attggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatggttggcacc atgggcggatattatatcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtg atagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttggga gttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagt gatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagag caccggtctctaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaa tcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagac accaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaat atttgcatgcagccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtccccaagcttggcgtaatcatggtcatagctgttt cctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagcta actcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggag aggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaa aggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgta aaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccc gacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcct ttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcac gaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagca gccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaga acagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtag c ggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagt ggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaat caatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcc atagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacg ctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagt ctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacg ctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctcc ttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatcc gtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatac gggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctg ttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaa ggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcag ggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacct gacgtctaagaaaccattattatcatgacattaacctataaaaatagGcgtatcacgaggccctttcgtc*
SYN4.6.3 Circular Plasmid Synthetic Gene (*Genscript*)	*tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcag acaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcac catatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggtttt cccagtcacgacgttgtaaaacgacggccagtgaattcggggacaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccctc gctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttcccctatcagtgatagagacttataagttc cctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctccgcccct tgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagtt ccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccg ggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaa tccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttca attggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatggttggcacc atgggcggatattatatcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtg*

	atagagacttataagttcccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttggga gttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagt gatagagacttataagttcccctatcagtgatagagacttataagttcccctatcagtgatagagacaccaggggcgtagttcaattggtagag caccggtcttcaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaa tcaccataaacgtgaaatccctatcagtgatagagacttataagttcccctatcagtgatagagacttataagttcccctatcagtgatagagac accaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaat atttgcatgcagccaatccctgggtgagtttcaccacccaacttttctatacaaagttgtccccaagcttggcgtaatcatggtcatagctgttt cctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagcta actcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggag aggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaa aggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgta aaaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccc gacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcct ttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcac gaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagca gccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaga acagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagc ggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagt ggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaat caatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcc atagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacg ctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagt ctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacg ctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctcc ttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatcc gtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatac gggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctg ttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaa ggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcag ggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacct gacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc
SYN4.6.a Circular Plasmid Synthetic Gene (*Genscript*)	*gacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgt gtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagg gttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggt cattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattga cgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttgg cagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgacct tatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggata gcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgt cgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaac ccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacttaagcttggtaccgagctcgg atccatgactaagcccatcgttttgctggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaa catgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaagcgacg ctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttag gctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaac gctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgagtcctgctggtgaagaccagaaacagcac ctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgtttattccgaaatctggcgcg cgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctggaagatccgaaat cggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggcgtttcca acctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaagg tgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgat gaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaatg atagggatccactagtccagtgtggtggaattctgcagatatccagcacagtggcggccgctcgagtctagaggaagcggagctactaact tcagcctgctgaagcaggctggagacgtggaggagaaccctggacctatgagcaagggcgaggagctgttcaccggggtggtgcccatcc tggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctg aagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctacccc gaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaact acaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaac atcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttc aagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctg cccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccg ccgccgggatcactcacggcatggacgagctgtacaagtaagggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccag ccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcat tgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctgg ggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagc gcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacg ttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgatta gggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttcca aactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattt aacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaa agcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagt cagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttat*

	ttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctccc gggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctcc ggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcagg ggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacg ggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtc atctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccac caagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctc gcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatat catggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtg atattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgc cttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccg ccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgc ccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattcta gttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcc tgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaa ctcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggaga ggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaa ggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaa aaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccg acaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctt tctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacg aaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcag ccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaa cagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcg gtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtg gaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatc aatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcca tagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgct caccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtct attaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgct cgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctcctt cggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccg taagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacg ggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgt tgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaag gcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagg gttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctg acgtc
SYN4.6.b Circular Plasmid Synthetic Gene (*Genscript*)	*atgcattagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccg cctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaa tgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaa tggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgat gcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttt tggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtct atataagcagagctggtttagtgaaccgtcagatccgctagcgattacgccaagctcgaaattaaccctcactaaagggaacaaaagctgg agctccaccgcggtggcggccgccaccatggattacaaggatgacgacgataaggactataaggacgatgatgacaaggactacaaag atgatgacgataaagcccgggcgggatccatgactaagcccatcgttttgctggcgcacagccctcaggtgaattgaccattggtaactac atgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccagga tgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtc ccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaat ctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgggtcc ttgtggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccgg agccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgtt atcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatg tgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaag gcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaac gatgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagc gattggttttgtggcgaagccgtaatgataggggatccccccgggctgcaggaattcgatatcggaagcggagctactaacttcagcctgctga agcaggctggagacgtggaggagaaccctggacctaagcttatcgataccgtcgacctcgaggggggggcccggtaccttaattaattaag gtaccaggtaagtgtacccaattcgccctatagtgagtcgtattacaattcactcgatcggctcgctgatcagcctcgactgtgccttctagttg ccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtccttttcctaataaaatgaggaaattgcatc gcattgtctgagtaggtgtcattctattctgggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatg ctggggaacgcgtaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccgaaat cggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtg gactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagttttttggggtcgaggt gccgtaaagcactaaatcggaaccctaaagggagccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagg gaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgc cgctacagggcgcgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctc atgagacaataaccctgataaatgcttcaataatattgaaaaaggaagaatcctgaggcggaaagaaccagctgtggaatgtgtgtcagt tagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccc caggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgccccta actccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctcggcctctgagctattccag*

	aagtagtgaggaggctttttggaggcctaggcttttgcaaagatcgatcaagagacaggatgaggatcgtttcgcatgattgaacaagat ggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgt gttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggc tatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtg ccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccg gctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctgg acgaagaacatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgagcatgcccgacggcgaggatctcgtcgtgacccatg gcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcagg acatagcgttggctacccgtgatattgctgaagaacttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattc gcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgcca tcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggg gatctcatgctggagttcttcgcccaccctagggggaggctaactgaaacacggaaggagacaatacccggaaggaacccgcgctatgacg gcaataaaaagacagaataaaacgcacggtgttgggtcgtttgttcataaacgcggggttcggtcccagggctggcactctgtcgatacccc caccgagaccccattggggccaatacgcccgcgtttcttccttttccccaccccaccccccaagttcgggtgaaggcccagggctcgcagcca acgtcggggcggcaggccctgccatagcctcaggttactcatatatacttttagattgatttaaaacttcatttttaatttaaaaggatctaggt gaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatc ttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctac caactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaaga actctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactca agacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccga actgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcg gaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcga tttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctca catgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcc
eGFP-WPRE Circular Plasmid Synthetic Gene (*Twist Biosciences*)	*ttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaa ggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaa tttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttctttt ccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgag acgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatat tttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggat aaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgcca tgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacc catataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttcccgttgaatatggctcataacacccttgtattactgttt atgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacacaacgtggctttccccgc cgctctagaactagtggatccaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcgcattatacgagac gtccaggttgggatacctgaaacaaaacccatcgtacggccaaggaagtctccaataactgtgatccaccacaagcgccagggttttccca gtcacgacgttgtaaaacgacggccagtcatgcataatccgcacgcatctggaataaggaagtgccattccgcctgacctgcaacaaacttc tcactactcaaacaagcaggtgacgtggaggagaatcccgggccttctagaggatccggctccggcgagggcaggggaagtctactaaca tgcggggacgtggaggaaaatcccggcccagtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacgg cgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctaagttcatctgcaccacc ggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagca cgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgag gtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagct ggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaacttcaagattaggcacaacat cgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctg agcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg gcatggacgagctgtacaagggttcaactagtgggtctggtcaccaccaccatcaccatcaccaccatcacggttcaactagtggctctggt gagcagaagctgatctccgaggaggacctgggttcaactagtggttctggtgactacaaggacgacgacaagcttttagtaatgatagt aatgatagtaatgagggcccttcgaacaaaaactcatctcagaagaggatctgaatatgcataccggtcatcatcaccatcaccattgaga attcaacgcgttaagtcgacaatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgt ggatacgctgctttaatgccttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgag gagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcag ctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttg ggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttc tgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcaga cgagtcggatctcccttgggccgcctccccgcggcctggccggccgccagcacagtggtcgatcgaccaatgccctggctcacaaataccac tgagatcttttcccctctgccaaaaattatggggaggctaggtggaggctcagtgatgataagtctgcgatggtggatgcatgtgtcatggtc atagctgtttcctgtgtgaaattgttatccgctcagagggcacaatcctattccgcgctatccgacaatctccaagacattaggtggagttcag ttcggcgtatggcatatgtcgctggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcg tttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccc aggcgtttcccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtgg cgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgac cgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagc agagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctct gctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagca gcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctctattcaacaaagccgccgtcccgtc aagtcagcgtaaatgggtaggggggcttcaaatcgtccgctctgccagtgttacaaccaattaacaaattctga*
FP-mRaspberry Linear DNA Synthetic Gene	*agtgaaccgtcagatccgctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatcc accggtggctgccgccaccatggtgagcaagggcgaggaggtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgaa cggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggc*

(*Twist Biosciences*)	*cccctgccctttcgcctgggacatcctgtcccctcagtgcatgtacggctccaagggctacgtgaagcaccccgccgacatccccgactacttga agctgtccttcccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgca ggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacgg ccccgtaatgcagaagaagaccatgggctggga ggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatgaagatgaggctgaagctgaaggacggcggccactacg acgccgaggtcaagaccacctacatggccaagaagcccgtgcagctgcccggcgcctacaagaccgacatcaagctggacatcacctccc acaacgaggactacaccatcgtggaacagtacgagcgcgccgagggccgccactccaccggcgcctaagcggccgcgactctagatcata atcagccataccacatttgtagaggtt*
mmPylRS-Y384F Linear DNA Synthetic Gene (*Twist Biosciences*)	*ggggacaagtttgtacaaaaaagcaggcttagccgccaccatggataaaaaaccactaaacactctgatatctgcaaccgggctctggat gtccaggaccggaacaattcataaaataaaacaccacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgtt gtaaacaactccaggagcagcaggactgcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatg aggatctcaataagttcctcacaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaag gcaatgccaaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaacctttctggatctaaatttttcacctgc gataccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactg gtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgccccgcacttacgaagagccagactgacaggcttg aagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctctctcgcagaaaaaaa gacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacagggttttc tggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagg gttgacaagaacttctgcctgagacccatgcttgctccaaacctttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaa atttttgaaataggcccatgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgccagatgggat cgggatgcacacgggaaaatcttgaaagcataattacggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatg gtctttggggatacccttgatgtaatgcacggagacctggaactttcctctgcagtagtcggacccatacccgcttgaccgggaatggggtatt gataaaccctggatagggggcaggtttcgggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccg agtcttactataacgggatttctaccaacctgtaatgatagaacccagctttcttgtacaaagtggtcccc*
DanAla-RS Plasmid DNA Synthetic Gene (*Twist Biosciences*)	*aggttgggatacctgaaacaaaacccatcgtacggccaaggaagtctccaataactgtgatccaccacaagcgccagggttttcccagtca cgacgttgtaaaacgacggccagtcatgcataatccgcacgcatctggaataaggaagtgccattccgcctgacctcaaataatgattttatt ttgactgatagtgacctgttcgttgcaacagattgatgagcaatgctttttataatgccaactttgtacaaaaaagcaggctgcaccatgca agagcagtaccgccctgaagagattgagtccaaagtgcagttacactgggatgagaagcgaacattcgaggtgaccgaagacgagtcca aggagaagtactactgcttaagcgccaatccttacccgagcggtagactgcatatgggacatgtgcgaaattacaccatcggcgacgtgat cgcccggtaccagagaatgctcggaaagaacgtgctccagccaatcggatgggacgcattcgggttacctgcagaagggggcggccgtga aaaacaatacagctccagctccatggacctatgataacatcgcgtatatgaaaaatcagctgaagatgcttgggtttggctacgactggtcc cgggaactggctacttgcacacccgaatactatagatgggagcagaagttttttacagagctctacaagaaagggctggtgtacaaaaag accagcgccgtaaactggtgcccaaatgaccagaccgtgttggcgaatgaacaggtgatcgatgggtgctgttggagatgtgacactaaa gtagaaaggaaggagatcccacagtggttcataaagatcactgcgtacgctgatgagctcctgaacgacctggacaagttagaccattgg cctgacacagtaaagactatgcaaaggaattggataggacgtagtgagggcgtggagattacatttaacgtaaatgattatgataacacc cttacagtgtacactactcgaccggacgcctttatgggctgcacttatcttgcagtggccgccggacatccgttagcccagaaagctgcagaa aataaccctgagctggcagcctttatcgacgagtgcagaaacaccaaggtcgctgaggccgagatggctacaatggaaaagaaaggagt agataccggctttaaggctgtccaccctctcacaggggaagaaattcctgtatgggctgccaacttcgtcctgatggagtatggcactggcgc agtgatggcagtacccggacatgaccagcgagattacgagttcgcctccaaatacgggctcaatatcaagccagtgattctggcggctgat ggttctgaaccagacctcagtcagcaagctctgactgaaaagggtgttctcttcaactctggggaatttaatggtcttgaccacgaagcggct ttcaacgcaatcgcagacaagttgaccgccatgggggttggggagagaaaggtgaactaccggctgagggactggggagttagccggca gcgctattggggcgccccaatccctatggtcaccttggaggatggaacagtgatgcccacaccagatgaccagctccctgtgattcttcccga agacgtggtgatggacggaattacttcccccattaaggccgacccagagtgggcgaaaacaacagttaatggtatgccagcgctgcggga aactgacaccttcgacactttcatggaaagctcttggatctacgcacgctacacgtgcccccaatacaaggagggaatgctggacagcgag gcggcaaactattggctccccgttgacattggaattggagggatcgaacacgccataatgacattgctgtattttagatttttccataaactga tgcgagatgcaggcatggtgaactctgacgagccagcaaaacagctgttatgccagggaatggtgctggccgacgctttctactacgttgg cgaaaacggcgagcgtaattgggtaagtcctgttgacgcaatagtagagcgcgacgagaagggaaggatagttaaggccaaagacgcg gctgggcacgaattggtgtacacgggcatgtcaaagatgagtaaatccaaaaacaatggtattgacccccaggtcatggtcgaacggtat ggggcagatactgtacggctctttatgattcgcctccccgcagatatgacccttgagtggcaggaatccggggtggaggggggcgaaca gattcttaaagcgcgtctggaagctggtgtacgagcacaccgctaaagctgatgtggccgcactcaacgtggatgccttgacagaaaacca gaaggccctgcggcgcgacgtgcacaagactatcgcaaaggttacagatgacatcgggaggcggcaaacgttcaatacagccatcgcgg caattatggagttgatgaataagctggcaaaggctccaacggatggagagcaggaccgagccctcatgcaggaggctctgctggccgtgg taagaatgttaaacccttcactcctcatatctgtttcacattgtggcaggaactgaagggggagggagacatcgataacgctccgtggcca gtggccgacgagaaggctatggtcgaagatagcaccctcgtggtggtccaagtaaacgggaaggtgagagctaaaataacggtgcccgt cgacgccactgaggagcaagtacgagagagggccggccaggagcacctggtggccaagtacctcgacggagtcacagtcagaaaggtc atctatgtgccgggcaaactgctgaacctcgttgtgggctaagacccagctttcttgtacaaagttggcattataagaaagcattgcttatca atttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgaggctaggtggaggctcagtgatgataagtctgcgatggtgg atgcatgtgtcatggtcatagctgtttcctgtgtgaaattgttatccgctcagagggcacaatcctattccgcgctatccgacaatctccaagac attaggtggagttcagttcggcgtatggcatatgtcgctggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaa ggccgcgttgctggcgttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgaca ggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctc ccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaac cccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcca ctggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacag tatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtg gttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctctattcaa caaagccgccgtcccgtcaagtcagcgtaaatgggtaggggggcttcaaatcgtccgctctgccagtgttacaaccaattaacaaattctgatt agaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaag gagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaat ttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttctttc cagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgaga cgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattt*

	tcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggata aaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccat gtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccc atataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttcccgttgaatatggctcataacacccccttgtattactgtttа tgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacacaacgtggctttccccgcc gctctagaactagtggatccaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcgcattatacgagacg tcc
pAcBac2.tR4-OMeYRS/GFP* a.k.a. pOMeY Plasmid DNA (*AddGene Cat# 50831)* *(This plasmid was a gift from Dr. Peter G. Schultz* [146]*)*	*ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaa tgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccg ctccttttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctt tacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttgg agtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccga tttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttt cggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttc aataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcaccc agaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcctt gagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaa gagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagt aagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaac cgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacccaaacgacgagcgtgac accacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactg gatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtg ggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggat gaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgat ttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactga gcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgcta ccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttc tagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcaca cagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgccacgcttcccgaagggagaaa ggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtc ctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcct ttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgag ctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttac gcatctgtgcggtatttcacaccgcagaccagccgcgtaacctggcaaaatcggttacggttgagtaataaatggatgccctgcgtaagcgg gtgtgggcggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactagacagaatagttgtaaactg aaatcagtccagttatgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcatttttctgaagtgcaaattgcccgtcgta ttaaagaggggcgtggccaagggcatggtaaagactatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgat ctcggcttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttcttcccgtatgcccaactttgtatagagagccact gcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtggc aatgccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcggctgctcaaacctgggcagaacgtaagcc gcgagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccgaggtaatcggagtcc ggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcatggatttgacttggtcagggccga gcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgccctgctgcgtaacatcgttgctgctgcgtaacatc gttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatagactgtacaaaaaaacagtcat aacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttctggaccagttgcgtgagcgcatacgctacttgcatt acagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccacggtgtgcgtcacccggcaaccttgggcagcagcg aagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtctccacgcatcgtcaggcattggcggccttgctgttcttctacgg caaggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggccgtcgcggcgcttgccggtggtgctgaccccggatgaag tggttcgcatcctcggttttctggaaggcgagcatcgtttgttcgcccaggactctagctatagttctagtggttggctacgtacccgtagtggc tatggcagggcttgcgcttaatgcgccgctacagggcgcgtggggataccccctagagccccagctgaaaaaatggaggggacggattc gaaccgccgaacccaaagggagcggatttagagtccgccgcgtttagccacttcgctacccctccggtgtctctatcactgatagggaacttа taagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaatggtgggggaaggа ttcgaaccttcgaagtctgtgacggcagatttagagtctgctcccctttggccgctcgggaaccccaccggtgtttcgtcctttccacaagatata taaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccatttttaaaacataatttttaaaactgcaaactacccaagaa attattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgca aatatgaaggaatcatgggaaataggccctcttcctgcccgagatctggttctttccgcctcagaagccatagagcccaccgcatccccagc atgcctgctattgtcttcccaatcctccccctttgctgtcctgccccacccccccccagaatagaatgacacctactcagacaatgcgatgcaa tttcctcatttttattaggaaaggacagtgggagtggcaccttccagggtcaaggaaggcacgggggaggggcaaacaacagatggctggc aactagaaggcacagtcgaggctgatcagcgggtttaaacgggccctctagactcgagttaaagtcgacttaacgcgttgaattcttaaac gggcccttccagcaaatcagacagtaattcttttttaccgcgacgcagtaaggtaaaacgaccaaacagacgatcttcttctttaaagaagt attcaggatcggactgtttttcaccgttaatggtgatggcattggaggcgatagtttacgtgcctgaccacgggaaggttgcagttcagaat cgaccagtgcctgcatcaggtctgcgcccttttccatctcaaccatcggtacgccgtcctgcgccagctgttcgaagtccgcttcactcagcgca ctcaaagaaccgctgaacaggcattcggtaatacgttttgccgcctgtaaaccttcttcaccgtgaaccagacgagtcacctgctccgccagt acatactgggcgcgcggtgctttaccgctgtttttatcttcttcttccagggcgttgatctcttcaatgctcataaaggtgaagaacttcaggaa gcggtaaacgtcggcacgcgcagtgttgatccagaactggtagaatttgtacgggctggttttcttcggatccaaccagactgcgccgccttc agttttaccaaatttggtgccatctgctttagtgatcagcggaacggtcaggccaaacacctgattctgatgcagacgacgggtcaggtcgat accagaagtgatgttaccccactggtcagaaccaccaatttgcagcaccacaccgtactgtttgttcaaacaggccatactataaccctgcag caggttgtaggaaaactcagtgaacgaaatcccctgatcttcacggttgagacgctgcttaaccgcttctttgttgatcatctggttaacggag aagtgtttgccaatatcgcgcaggaaggtcagcacattcatattgccgaaccagtcataattattggccgcgatagcagagttttctccacag tcgaaatcgaggaacggggcaacctgcttacggattttgtccacccactcctgaacagtttcttcggtgttcagcttacgctcggcagctttga agctcgggtcgccaatcagacccgtcgcgccgcctaccagcgcaaccggcttgtggcccgcctgctggaagcgtttcaggcataacaatgg*

aacaagatgccccaaatgcaagctgtcagcggtaggatcgaagccacacacgagtgcgatcgggccttgcgccagtcgctctgctaacgct
tcctcgtccgtcacctgggctaccagccccgctcttgcaattgtttaatcaagttactgcttgccatggtggcgctagccagcttgggtctccct
atagtgagtcgtattaatttcgataagccagtaagcagtgggttctctagttagccagagagctctgcttatatagacctcccaccgtacacg
cctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccgttgattttggtgccaaaacaaactcccattgacgtca
atggggtggagacttggaaatccccgtgagtcaaaccgctatccacgcccattgatgtactgccaaaaccgcatcaccatggtaatagcgat
gactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgt
caataggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaaatagtccacccattgacgtcaatggaaa
gtccctattggcgttactatgggaacatacgtcattattgacgtcaatgggcgggggtcgttgggcggtcagccaggcgggccatttaccgta
agttatgtaacgcggaactccatatatgggctatgaactaatgaccccgtaattgattactattaataactagtcaataatcaatgtcaacgc
gtatatctggcccgtacatcgcgaagcagcgcaaaacggatcctgcaggtatttgcggccgcggtccgtatactccggaatattaatagatc
atggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattcc
ggattattcataccgtcccaccatcgggcgcgaactcctaaaaaaccgccaccatgaagtgccttttgtacttagccttttttattcattgggggtg
aattgcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatt
taaattggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtca
tgcttccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaat
gcaaggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatg
ccgaagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaa
tgcagcaattacatatgccccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccat
ggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaact
ggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctct
ttgctgcagccagattcccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttg
agaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcc
taaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgctcc
aatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaa
ttggacccaatggagttctgaggaccagttcaggatataagtttcctttatacatgattggacatggtatgttggactccgatcttcatcttagc
tcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttttggtgatactgggctatc
caaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctctttttctttatcataggattaatcattggactatt
cttggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttg
gaaagtgataagtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcc
ccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaa
atttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatcactgcttga
gcctagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcct
ggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgg
gtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatgg
cccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtcgaggt
gagccccacgttctgcttcactctcccatctccccccctccccacccccaattttgtatttatttattttttaattattttgtgcagcgatgggggc
ggggggggggggggcgcgcgccaggcggggcggggcggggcgaggggcggggcggggcgaggcggagaggtgcggcggcagcca
atcagagcggcgcgctccgaaagtttccttttatggcgaggcggcggcggcggccctataaaaagcgaagcgcgcggcgggcgggga
gtcgctgcgttgccttcgccccgtgccccgctccgcgccgcctcgcgccgcccgccccggctctgactgaccgcgttactcccacaggtgagcg
ggcgggacggcccttctcctccgggctgtaattagcgcttggtttaatgacggctcgtttcttttctgtggctgcgtgaaagccttaaagggctc
cgggagggccctttgtgcgggggggagcggctcgggggggtgcgtgcgtgtgtgtgcgtgggagcgccgcgtgcggcccgcgctgccc
ggcggctgtgagcgctgcgggcgcggcgcggggctttgtgcgctccgcgtgtgcgcgaggggagcgcggccgggggcggtgccccgcggt
gcgggggggctgcgaggggaacaaaggctgcgtgcggggtgtgtgcgtgggggggtgagcaggggggtgtgggcgcggcggtcgggct
gtaaccccccccgcaccccctccccgagttgctgagcacggcccggcttcgggtgcggggctccgtgcggggcgtggcgcggggctcgcc
gtgccgggcgggggggtggcggcaggtgggggtgccgggcggggcggggccgcctcgggccgggggagggctcgggggagggggcgcggc
ggccccggagcgccggcggctgtcgaggcgcggcgagccgcagccattgccttttatggtaatcgtgcgagagggcgcagggacttccttt
gtcccaaatctggcggagccgaaatctgggaggcgccgccgcaccccctctagcgggcgcgggcgaagcggtgcggcgccggcaggaag
gaaatgggcggggagggccttcgtgcgtcgccgcgccgccgtcccctctccatctccagcctcggggctgccgcaggggacggctgccttt
cgggggggacggggcagggcggggttcggcttctggcgtgtgaccggcggctctagagcctctgctaaccatgttcatgccttcttcttttttcc
tacagctcctgggcaacgtgctggttattgtgctgtctcatcattttggcaaagaattggccaaggaggccaccatggtgagcaagggcgag
gagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggc
gatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctac
ggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgca
ccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagg
gcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagc
agaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacc
cccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatc
acatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggggcccttcgaacaaaaactcatctca
gaagaggatctgaatatgcataccggtcatcatcaccatcaccattgagaattcaacgcgttaagtcgacaatcaacctctggattacaaaa
tttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgccttgtatcatgctattgcttcccgt
atggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactg
tgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcgg
aactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctt
tccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgc
ggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctcccttttgggccgcctccccgcggcctggccg
gccgccagcacagtggtcgatcgaccaatgccctggctcacaaataccactgagatcttttcccctctgccaaaaattatggggacatcatga
agccccttgagcatctgacttctggctaataaaggaaatttattttcattgcaatagtgtgttggaatttttgtgtctctcactcggaaggacat
atgggagggcaaatcatttaaaacatcagaatgagtatttggtttagagtttggcaacatatgcccatatgctggctgccatgaacaaaggt
tggctataaagaggtcatcagtatatgaaacagcccccctgctgtccattccttattccatagaaaagccttgacttgaggttagattttttttat
attttgttttgtgttatttttttctttaacatccctaaaattttccttacatgttttactagccagatttttcctcctctcctgactactcccagtcatag
ctgtccctcttctcttatggagatccctcgacctgcctaggtcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgat
acaaggctgttagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttctt
tgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgt
ggaaaggacgaaacaccggtggggttcccgagcggccaaagggagcagactctaaatctgccgtcacagacttcgaaggttcgaatcctt

	ccccccaccatttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttcc ctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggactctaaatccgctccctttgggttcggcggttcgaatcc gtccccctccatttttgctaggagatccgaaccagataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttacgacg ctacacccagttcccatctattttgtcactcttccctaaataatccttaaaaactccatttccacccctcccagttcccaactattttgtccgcccac agcggggcatttttcttcctgttatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttcggctacttt
pDONR221 a.k.a. GD21 (*ThermoFisher Cat# 12536017*)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagc gagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgac aggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaagg ccatccgtcaggatggccttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgtt caaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcc tttcgttttatttgatgcctggcagttccctactctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtct taagctcgggccccaaataatgattttattttgactgatagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaacttt gtacaaaaaagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaacagactacataatactgt aaaacacaacatatccagtcactatgaatcaactacttagatggtattagtgacctgtagtcgaccgacagccttccaaatgttcttcgggtg atgctgccaacttagtcgaccgacagccttccaaatgttcttctcaaacggaatcgtcgtatccagcctactcgctattgtcctcaatgccgtat taaatcataaaaagaaataagaaaaagaggtgcgagcctcttttttgtgtgacaaaataaaaacatctacctattcatatacgctagtgtca tagtcctgaaaatcatctgcatcaagaacaatttcacaactcttatacttttctcttacaagtcgttcggcttcatctggattttcagcctctatac ttactaaacgtgataaagtttctgtaatttctactgtatcgacctgcagactggctgtgtataagggagcctgacatttatattccccagaacat caggttaatggcgtttttgatgtcattttcgcggtggctgagatcagccacttcttcccgataacggagaccggcacactggccatatcggtg gtcatcatgcgccagctttcatcccccgatatgcaccaccgggtaaagttcacgggagactttatctgacagcagacgtgcactggccagggg gatcaccatccgtcgcccgggcgtgtcaataatatcactctgtacatccacaaacagacgataacggctctctcttttataggtgtaaacctta aactgcatttcaccagcccctgttctcgtcagcaaaagagccgttcatttcaataaaccgggcgacctcagccatcccttcctgattttccgctttt ccagcgttcggcacgcagacgacgggcttcattctgcatggttgtgcttaccagaccggagatattgacatcatatatgccttgagcaactga tagctgtcgctgtcaactgtcactgtaatacgctgcttcatagcatacctcttttttgacatacttcgggtatacatatcagtatatattcttatacc gcaaaaatcagcgcgcaaatacgcatactgttatctggcttttagtaagccggatccacgcggcgtttacgccccccctgccactcatcgcag tactgttgtaattcattaagcattctgccgacatggaagccatcacaaacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcg ccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcaccca gggattggctgagacgaaaaacatattctcaataaacccttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatata tgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtg aacactatcccatatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaagg ccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactga ctgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcct gaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcatt ttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattc ggcgcaaagtgcgtcgggtgatgctgccaacttagtcgactacaggtcactaataccatctaagtagttgattcatagtgactggatatgttg tgttttacagtattatgtagtctgtttttttatgcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcagctttcttgtacaaa gttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgata tcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaa aatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggcc cgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgt atgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactg gctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaaca gcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttg taattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgac gagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttct cacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgcca tcctatggaactgcctcggtgagttttctccttcattacagaaacggcttttcaaaaatatggtattgataatcctgatatgaataaattgcag tttcatttgatgctcgatgagttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaa gctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctt ttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccg aaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcacc gcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagtta ccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacct acagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggaga gcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgc tcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgtt*
pAB2G-N-HMF (*Dr. Edward Stuttfeld, prepared inhouse*)	*ttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaa cgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaat ctaaagtatatatgagtaaacttggtctgacagttgacccaagtcaaccggttgtgggcggacaaaatagttgggaactgggaggggtgg aaatggagttttaaggattatttagggaagagtgacaaaatagatgggaactgggtgtagcgtcgtaagctaatacgaaaattaaaaat gacaaaatagtttggaactagatttcacttatctggttccaattagatgggtataccctagtatacggaccttttaattcaacccaacacaatat attatagttaaataagaattattatcaaatcatttgtatattaattaaaatactatactgtaaattacattttatttacaatcactcgacgaaga cttgatcacccgggatctcgagctagcgaccatggcacaccaccaccaccaccaccaccaccaccacggttcaactagtggttctggtgagc agaagctgatctccgaggaggacctgggttcaactagtggttctggtgactacaaggacgacgacgacaagcttacaagtttgtacaaaa aagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaacagactacataaatactgtaaaacac aacatatccagtcactatggcggccgcattaggcaccccaggctttacactttatgcttccggctcgtataatgtgtggattttgagttaggatc cgtcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaag aacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaa ataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagct ggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttc*

	cggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgt ctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcacgatgggcaaatatt atacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcggcagaatgcttaatgaat tacaacagtactgcgatgagtggcagggcggggcgtaaacgcgtggatccggcttactaaaagccagataacagtatgcgtatttgcgcg ctgatttttgcggtataagaatatatactgatatgtatacccgaagtatgtcaaaaagaggtatgctatgaagcagcgtattacagtgacag ttgacagcgacagctatcagttgctcaaggcatatatgatgtcaatatctccggtctggtaagcacaaccatgcagaatgaagcccgtcgtct gcgtgccgaacgctggaaagcggaaaatcaggaagggatggctgaggtcgcccggtttattgaaatgaacggctcttttgctgacgagaa cagggggctggtgaaatgcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttgtggatgtacagagtgatattattgac acgcccgggcgacggatggtgatccccctggccagtgcacgtctgctgtcagataaagtctcccgtgaactttacccggtggtgcatatcgg gggatgaaagctggcgcatgatgaccaccgatatggccagtgtgccggtctccgttatcggggaagaagtggctgatctcagccaccgcgaaa aatgacatcaaaaacgccattaacctgatgttctggggaatataaatgtcaggctcccttatacacagccagtctgcaggtcgaccatagtg actggatatgttgtgttttacagtattatgtagtctgtttttttatgcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcag ctttcttgtacaaagtggtaccgggagatgggggaggctaactgaaacacggaaggagacaatacggaaggaacccgcgctatgacgg caataaaaagacagaataaaacgcacgggtgttgggtcgtttgttcataaacgcggggttcggtcccagggctggcactctgtcgataccc caccgagaccccattgggaccaatacgcccgcgtttcttccttttccccaccccaacccccaagttcgggtgaaggcccagggctcgcagcca acgtcggggcggcaagccctgccatagccactacgggtacgtttaaactcgctaccttaggaccgttatagttacagataacttcgtataatg tatgctatacgaagttatggcgacttcgctgctgcccaaggttgccgggtgacgcacaccgtggaaacggatgaaggcacgaacccagttg acataagcctgttcggttcgtaaactgtaatgcaagtagcgtatgcgctcacgcaactggtccagaaccttgaccgaacgcagcggtggtaa cggcgcagtggcggttttcatggcttgttatgactgttttttgtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtc gatgtttgatgttatggagcagcaacgatgttacgcagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaggtggct caagtatgggcatcattcgcacatgtaggctcggccctgaccaagtcaaatccatgcgggctgctcttgatcttttcggtcgtgagttcggag acgtagccacctactcccaacatcagccggactccgattacctcgggaacttgctccgtagtaagacattcatcgcgcttgctgccttcgacca agaagcggttgttggcgctctcgcggcttacgttctgcccaagtttgagcagccgcgtagtgagatctatatctatgatctcgcagtctccggc gagcaccggaggcagggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcgcttggtgcttatgtgatctacgtgcaagc agattacggtgacgatcccgcagtggctctctatacaaagttgggcatacgggaagaagtgatgcactttgatatcgacccaagtaccgcc acctaacaattcgttcaagccgagatcggcttcccggccgcggagttgttcggtaaattgtcacaacgccgcgaaatatagtctttacatgccct tggccacgcccctctttaatacgacgggcaatttgcacttcagaaaatgaagagtttgctttagccataacaaaagtccagtatgcttttcac agcataactggactgatttcagtttacaactattctgtctagtttaagacttattgtcatagtttagatctatttgttcagtttaagactttattg tccgcccacacccgcttacgcagggcatctgcgcatccacaggaagagcgacccaagtcaatcaggggataacgcaggaaagaacatgt gagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcaca aaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcct gttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcg gtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtcca acccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttctt gaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggta gctcttgatccggcaaacaaaccaccgctggtagcggtgg
COM296_pDEST_pcD (*Dr. Edward Stuttfeld, prepared inhouse*)	*gacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgt gtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagg gttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggt cattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattga cgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttgg cagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgacct tatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggata gcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgt cgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaac ccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacgggccctctagactcgagacc atgggaggagagcagaagctgatctccgaggaggacctgaacgagcagaagctgatctccgaggaggacctgaacgctcgagcggccg ccactgtgctggatatcacaagtttgtacaaaaaagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcat aaaaaacagactacatataatactgtaaaacacaacatatccagtcactatggcggccgcattaggcaccccaggctttacactttatgcttcc ggctcgtataatgtgtggattttgagttaggatccgtcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatata ccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctgga tattacggcctttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccgg aattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgtttt catcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctattt ccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaact tcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtga tggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaaacgcgtggatccggcttacta aaagccagataacagtatgcgtatttgcgcgctgatttttgcggtataagaatatatactgatatgtatacccgaagtatgtcaaaaagagg tatgctatgaagcagcgtattacagtgacagttgacagcgacagctatcagttgctcaaggcatatatgatgtcaatatctccggtctggtaa gcacaaccatgcagaatgaagcccgtcgtctgcgtgccgaacgctggaaagcggaaaatcaggaagggatggctgaggtcgcccggttt attgaaatgaacggctcttttgctgacgagaacaggggctggtgaaatgcagtttaaggtttacacctataaaagagagagccgttatcgtc tgtttgtggatgtacagagtgatattattgacacgcccgggcgacggatggtgatcccccctggccagtgcacgtctgctgtcagataaagtct cccgtgaactttacccggtggtgcatatcggggatgaaagctggcgcatgatgaccaccgatatggccagtgtgccggtctccgttatcggg gaagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaacctgatgttctggggaatataaatgtcaggctcccttat acacagccagtctgcaggtcgaccatagtgactggatatgttgtgttttacagtattatgtagtctgtttttttatgcaaaatctaatttaatatat tgatatttatatcattttacgtttctcgttcagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagct cggtaccaagcttaagtttaaaccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgac cctggaaggtgccactcccactgtccttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggg gtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaa gaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgc tacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggg ctcccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgat*

	agacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctatt ctttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatg tgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtgg aaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcc cgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagc tattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaaga gacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgact gggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccg gtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcact gaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcat ggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactc ggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgc gcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcg actgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctga ccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctgggg ttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgtt ttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggtta caaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatc atgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacac aacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttt ccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctc gctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggg ataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctc cgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccct ggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagc tcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatc cggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtat gtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagtt accttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggttttttgtttgcaagcagcagattacgcgcag aaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatga gattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgaca gttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacg atacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagcc agccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagtt cgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttccc aacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccg cagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaacca agtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaa gtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccc aactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaagggaataagggcg acacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgt atttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtc
SI-E038 (*Dr. Stefan Imseng, prepared inhouse*)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagc gagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgac aggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaagg ccatccgtcaggatggccttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgtt caaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcc tttcgttttatttgatgcctggcagttccctactctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtct taagctcgggccccaaataatgattttattttgactgatagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaacttt gtacaaaaaagcaggcttaatgcgttcctccaagaacgtgatcaaggagttcatgcgtttcaaggtgcgtatggagggcaccgtgaacggt cacgagttcgagatcgagggtgagggtgagggtcgtccctacgagggtcacaacaccgtgaagctgaaggtgaccaagggtggtcccctg cccttcgcttgggacatcctgtccccccagttccagtacggttccaaggtgtacgtgaagcaccccgctgacatccccgactacaagaagctg tccttccccgagggtttcaagtgggagcgtgtgatgaacttcgaggacggtggtgtggtgaccgtgacccaggactcctccctgcaggacgg ttgcttcatctacaaggtgaagttcatcggtgtgaacttcccctccgacggtcccgtgatgcagaagaagactatgggttgggaggcttccac cgagcgtctgtacccccgtgacggtgtgctgaagggtgagatccacaaggctctgaagctgaaggacggtggtcactacctggtggagttc aagtccatctacatggctaagaagcccgtgcagctgcccggttactactacgtggactccaagctggacatcacctcccacaacgaggacta caccatcgtggagcagtacgagcgtaccgagggtcgtcaccacctgttcctgtaacacccagctttcttgtacaaagttggcattataagaaa gcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgatatcccctatagtgagtcgt attacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaac aataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaac atggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgc gccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgc ctcttccgaccatcaagcatttttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggttatta ga agaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacag cgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcc tgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttt tgacgagggaaattaatagttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcg gtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatg agtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatccc ttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgc tgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagc agagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgc*

	taatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcgg tcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagc tatgagaa agcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttcca gggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcc tatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt
pcDNA4-TO-Puromycin-mVenus-MAP a.k.a. pPuro Plasmid DNA (*AddGene Cat# 44118)* *(Plasmid was a gift from Dr. Dannel McCollum* [264]*)*	*acggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtg tgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggt taggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtca ttagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgt caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcag tacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttat gggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcg gtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggaaccaaaatcaacgggactttccaaaatgtcgt aacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctccctatcagtgatagagatc tccctatcagtgatagagatcgtcgacgagctcgtttagtgaaccgtcagatcgcctggagacgccatccacgctgttttgacctccatagaa gacaccgggaccgatccagcctccggactctagcgtttaaacttaagcttggtaccgagctcggatccactagtccagtgtggtggaattctg cagatatccagcacagtggcggccgcggcagcagcatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagc tggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagctgatct gcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgggctacggcctgcagtgcttcgcccgctaccccgaccacatga agcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccg cgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggc acaagctggagtacaactacaacagccacaacgtctatatcaccgccgacaagcagaagaacggcatcaaggccaacttcaagatccgc cacaacatcgaggacggcagcctcgagtccgctcatcaccatcaccatcaccatcacggcagcgccatggacgagaagaccaccggctgg cggggcggccacgtggtggagggcctggccggcgagctggagcagctgcgggccaggctggagcaccaccctcagggccagcgggagc cctccggcggcggtagcgctgactacaaagacgatgacgacaagggcagcgctagcagcggcggcggcgtgcagctcgccgaccactac cagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagctaccagtccaagctgagcaaagaccccaac gagaagcgcgatcacatggtcctgaaggagcgcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaatctagaggg cccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtg ccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcagg acagcaaggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctg gggctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccctttagg gttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttttt cgcccttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttat aagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagtta gggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcccca ggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaact ccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaag tagtgaggaggctttttggaggcctaggaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcct aggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagtacaagcccacggtgcgcctcgc cacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgacccggaccgcc acatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccg cggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttccc ggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcc cgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagac ctccgcgccccgcaacctcccctttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcat gacccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccggga cgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataa agcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgt ataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacg agccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgg gaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactg actcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgc aggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccc tgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaag ctc cctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgt aggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaact atcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcgg tgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcgg aaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggttttttgtttgcaagcagcagattacgcgcagaaaaaaa ggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatca aaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaat gcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggag ggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaa gggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagtta atagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatca aggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgtta tcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattc tgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcat cattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatc ttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaa*

	atgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaa aataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc
pANAP Plasmid DNA (*AddGene Cat# 49696)* *(Plasmid was a gift from Dr. Peter G. Schultz [186])*	*gacggatcgggagatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataa gttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttc aagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgataga gacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctg tgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatc agtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggc gttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtga aatccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaa atccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcacca taaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaag ggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctggg aaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggt agacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctat gtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggt ggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgc atgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacaccgc ccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatct cccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctga gtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctg cttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccat atatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatg ttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatc atatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttg gcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacgggg atttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgcccca ttgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttat cgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacttaagcttgccaccatggaagagcaataccgcccggaag agatagaatccaaagtacagcttcattgggatgagaagcgcacatttgaagtaaccgaagacgagagcaaagagaagtattactgctttt ctggccctccctatccttctggtcgactacacatgggccacgtacgtaactacaccatcggtgacgtgatcgcccgctaccagcgtatgctggg caaaaacgtcctgcagccgatcggctgggacgcgtttggtctgcctgcggaaggcgcggcggtgaaaaacaacaccgctccggcaccgtg gacgtacgacaacatcgcgtatatgaaaaaccagctcaaaatgctgggctttggttatgactggagccgcgagctggcaacctgtacgccg gaatactaccgttgggaacagaaattcttcaccgagctgtataaaaaaggcctggtatataagaagacttctgcggtcaactggtgtccga acgaccagaccgtactggcgaacgaacaagttatcgacggctgctgctggcgctgcgataccaaagttgaacgtaaagagatcccgcagt ggtttatcaaaatcactgcttacgctgacgagctgctcaacgatctggataaactggatcactggcctgacaccgttaaaaccatgcagcgt aactggatcggtcgttccgaaggcgtggagatcaccttcaacgttaacgactatgacaacacgctgaccgtttacactacccgcccggacac ctttatgggttgtacctacctggcggtagctgcgggtcatccgctggcgcagaaagcggcggaaaataatcctgaactggcggcctttattg acgaatgccgtaacaccaaagttgccgaagctgaaatggcgacgatggagaaaaaaggcgtcgatactggctttaaagcggttcacccat taacgggcgaagaaattcccgtttgggcagcaaacttcgtattgatggagtacggcacgggcgcagttatggcggtaccggggcacgacc agcgcgactacgagtttgcctctaaatacggcctgaacatcaaaccggttatcctggcagctgacggctctgagccagatctttctcagcaag ccctgactgaaaaaggcgtgctgttcaactctggcgagttcaacggtcttgaccatgaagcggccttcaacgccatcgccgataaactgact gcgatgggcgttggcgagcgtaaagtgaactaccgcctgcgcgactggggtgtttcccgtcagcgttactggggcgcgccgattccgatggt gactctagaagacggtaccgtaatgccgaccccggacgaccagctgccggtgatcctgccggaggatgtggtaatggacggcattaccag cccgattaaagcagatccggagtgggcgaaaactaccgttaacggtatgccagcactgcgtgaaaccgacactttcgacacctttatggag tcctcctgggttcttgcgcgctacacttgcccgcagtacaaagaaggtatgctggattccgaagcggctaactactggctgccggtggatatc gcgattggtggtattgaacacgccattatggagagtctctactgtcgcttcttccacaaactgatgcgtgatgcaggcatggtgaactctgac gaaccagttaaacagttgctgtgtcagggtatggtgctggcagatgccttctactatgttggcgaaaacggcgaacgtaactgggtttcccc ggttgatgctatcgttgaacgtgacgagaaaggccgtatcgtgaaagcgaaagatgcggcaggccatgaactggtttataccggcatgag caaaatgtccaagtcgaagaacaacggtatcgacccgcaggtgatggttgaacgttacggcgcggacaccgttcgtctgtttatgatgtttg cttctccggctgatatgactctcgaatggcaggaatccggtgtggaaggggctaaccgcttcctgaaacgtgtctggaaactggtttacgag cacacagcaaaaagtgatgttgcggcactgaacgttgatgcgctgactgaaaatcagaaagcgctgcgtcgcgatgtgcataaaacgatc gctaaagtgaccgatgatatcggccgtcgtcagaccttcaacaccgcaattgcggcgattatggagctgatgaacaaactggcgaaagcac caaccgatggcgagcaggatcgcgctctgatgcaggaagcactgctggccgttgtccgtatgcttaacccgttcaccccgcacatctgcttca cgctgtggcaggaactgaaaggcgaaggcgatatcgacaacgcgccgtggccggttgctgacgaaaaagcgatggtggaagactccacg ctggtcgtggtgcaggttaacggtaaagtccgtgccaaaatcaccgttccggtggacgcaacggaagaacaggttcgcgaacgtgctggcc aggaacatctggtagcaaaatatcttgatggcgttactgtacgtaaagtgatttacgtaccaggtaaactcctcaatctggtcgttggcgggc ccgtttaactcgagtctagagggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccg tgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattc tggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttc tgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcg cagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctc taaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggc catcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccct atctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaatta attctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagc aaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccta actccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcc tctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcg gatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggct attcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaa gaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtg*

	ctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccga gaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcg agcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcg ccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccg cttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttgg cggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctg agcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggtt gggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattg cagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactc atcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccg ctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgc gctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattggg cgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatc cacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggc gtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatac caggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtg gcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccga ccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattag cagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctc tgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggttttttgtttgcaagcagc agattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagg gattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaa acttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtg tagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagcta gagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattc agctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcaga agtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggt gagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatag cagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacc cactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagg gaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggata catatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtc
pAcBac1.tR4-MbPyl a.k.a. pMbPyl Plasmid DNA (*AddGene Cat#* 50832*)* *(Plasmid was a gift from Dr. Peter G. Schultz [146])*	*ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaa tgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccg ctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctt tacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttgg agtccacgttcttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgat ttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttc ggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttca ataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcaccca gaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttt gagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaag agcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagta agagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaacc gctttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaa acgacgagcgtgaca ccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactgg atggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgg gtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatg aacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatt taaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgag cgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctac cagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttct agtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagt ggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcaca cagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaa ggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtc ctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcct ttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgag ctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttac gcatctgtgcggtatttcacaccgcatagaccagccgcgtaacctggcaaaatcggttacggttgagtaataaatggatgccctgcgtaagc gggtgtgggcggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactagacagaatagttgtaaac tgaaatcagtccagttatgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcattttctgaagtgcaaattgcccgtcg tattaaagaggggcgtggccaagggcatggtaaagactatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccg atctcggcttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttcttcccgtatgcccaactttgtatagagagcca ctgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtg gcaatgccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcggctgctcaaacttgggcagaacgtaagc cgcgagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccgaggtaatcggagtcc ggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcatggatttgacttggtcagggccga gcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgccctgctgcgtaacatcgttgctgctgcgtaacatc gttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatagactgtacaaaaaaacagtcat aacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttctggaccagttgcgtgagcgcatacgctacttgcatt*

	acagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccacggtgtgcgtcacccggcaaccttgggcagcagcg aagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtctccacgcatcgtcaggcattggcggccttgctgttcttctacgg caaggtgctgtgcacggatctgcccttgcttcaggagatcggtagacctcggccgtcgcggcgcttgccggtggtgctgacccccggatgaag tggttcgcatcctcggttttctggaaggcgagcatcgtttgttcgcccaggactctagctatagttctagtggttggctacgtacccgtagtggc tatggcagggcttgcgcttaatgcgccgctacagggcgcgtggggataccccctagagccccagctgaaaaaacgggggggtgcgggagtt tcacccgcctgcacggatttagagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcact gatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacggaaaccccgggaatctaacccggctg aacggatttagagtccgttcgatctacatgatcaggtttccggtgtttcgtcctttccacaagatatataaagccaagaaatcgaaatactttc aagttacggtaagcatatgatagtccatttttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgt actaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaatag gccctcttcctgcccgagatctggttctttccgcctcagaagccatagagcccaccgcatccccagcatgcctgctattgtcttcccaatcctccc ccttgctgtcctgccccaccccaccccccagaatagaatgacacctactcagacaatgcgatgcaatttcctcattttattaggaaaggacagt gggagtggcaccttccagggtcaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaaggcacagtcgaggctga tcagcgggtttaaacgggccctctagactcgagttaaagtcgacgcggggaggcggcccaaagggagatccgactcgtctgagggcgaa ggcgaagacgcggaagaggccgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaagggacgtagcag aaggacgtcccgcgcagaatccaggtggcaacacaggcgagcagccaaggaaaggacgatgatttccccgacaacaccacggaattgtc agtgcccaacagccgagcccctgtccagcagcgggcaaggcaggcggcgatgagttccgccgtggcaatagggaggggggaaagcgaaa gtcccggaaaggagctgacaggtggtggcaatgccccaaccagtgggggttgcgtcagcaaacacagtgcacaccacgccacgttgcctg acaacgggccacaactcctcataaagagacagcaaccaggatttatacaaggaggagaaaatgaaagccatacgggaagcaatagcat gatacaaaggcattaaagcagcgtatccacatagcgtaaaaggagcaacatagttaagaataccagtcaatctttcacaaattttgtaatc cagaggttgattgtcgacttaacgcgttgaattcttacagattggttgaaatcccattatagtaagattcggaccttgatgccctcttaatgtttt taaagccgtgcataaccttgagcaagcgttcaagaccaaaacctgcacctatccatggtttgtcaatacccccattctctatcaagagaaactg gcccgacgactgccgaagaaagctccaggtccccgtgcattatatcaagagtatccccaaagaccatacaggaatctcctacgatttcgaa gtcgatttccagatagtccagaaactctttgatgagagcttcaagattttcccgagtacatcccgaacccatctgacagaagttcaccatagt aaattcttccaggtgctctttgccgtcagactctttccggtaacaaggtccgacttcgaaaatttttattgggcctggtaaaatcctatcgagttt tcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagagatttttatccacccggaagatctgttttgaaagttcagtatcatt attaatacccattctctccacgtattccgccggaataaggataggagactttatctccagaaaaccccggtctacgaaaaatttcgtaatatc acgttcgagtttaccgaggtagtcttctctatcattggtatagagccgctgaaaatcgtttttttcttcttgtcacaagttcaggctcaagttccctg aaaggctttgccatatttagagaaatttttatcctctggacttaagagagcctcaaccctatcaagctggcttcttgtaagtgaaggagcaggga gccgatgcgggaacagacgaatttggagttgattttgcaggcgaaggtacagatctggatgtgttcgttgatgcctttgcagaaacagaatt ttccagaggcttcggagcccttgaaactgatttcggcatagcttttttgaccttggagcagaaactaccctaactttcacactgtttttgctttcg gttgatcttgtgagaaaattattgatatcctcgtccgaaaccctacatcgtttgcaggtttttctgtacttatgatgtctgaatgctctggctgttc tacaactcctggaattattcacaacaagatggtctccacacgccatttcaatgtatattttacttcttgagacctcatggtgcttgattttgtgga gcgtgccagtcctggacatccagagcccggtcgcagatattaaaacatctaatggtttttttatccatggtggcgctagccagcttgggtctccc tatagtgagtcgtattaatttcgataagccagtaagcagtgggttctctagttagccagagagctctgcttatatagacctcccaccgtacacg cctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccgttgattttggtgccaaaacaaactcccattgacgtca atggggtggagacttggaaatccccgtgagtcaaaccgctatccacgcccattgatgtactgccaaaaccgcatcaccatggtaatagcgat gactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgt caataggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaaatactccacccattgacgtcaatggaaa gtccctattggcgttactatgggaacatacgtcattattgacgtcaatgggcgggggtcgttgggcggtcagccaggcgggccatttaccgta agttatgtaacgcggaactccatatatgggctatgaactaatgaccccgtaattgattactattaataactagtcaataatcaatgtcaacgc gtatatctggcccgtacatcgcgaagcagcgcaaaacggatcctgcaggtatttgcggccgcggtccgtatactccggaatattaatagatc atggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattcc ggattattcataccgtcccaccatcgggcgcgaactcctaaaaaaccgccaccatgaagtgccttttgtacttagccttttattcattggggtg aattgcaagttcaccatagttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatt taaattggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtca tgcttccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaat gcaaggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatg ccgaagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaa tgcagcaattacatatgccccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccat ggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaact ggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctct ttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttg agaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcc taaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgctcc aatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaa ttggacccaatggagttctgaggaccagttcaggatataagtttcctttatacatgattggacatggtatgttggactccgatcttcatcttagc tcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttggtgatactgggctatc caaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctcttttttctttatcataggttaatcattggactatt cttggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttg gaaagtgataaggccaggccggccaagcttgtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgcttt aaaaaacctcccacacctcccccgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaa taaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtc tggatctgatcactgcttgagcctaggtcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttag agagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcag ttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaa acaccggaaacctgatcatgtagatcgaacggactctaaatccgttcagccgggttagattcccggggtttccgtttttaatatttgcatgtc gctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggg tggatcgaatagatcacacggactctaaatccgtgcaggcgggtgaaactcccgcacccccgtttttttgctaggagatccgaaccagataa gtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttacgacgctacacccagttcccatctattttgtcactcttccctaaata atccttaaaaactccatttccacccctcccagttcccaactattttgtccgcccacagcggggcatttttcttcctgttatgtttttaatcaaacatc ctgccaactccatgtgacaaaccgtcatcttcggctacttt

pIDC	*tgtctggaatctgacgggggcattggtgacgaggtgacccagctcatccccaagactaagatccctgcccacacgcgctacgccctgcagt gtcgcttcagcccgactccacgctcctcgccacctgctcggctgatcagacgtgcaagatctggaggacgtccaacttctccctgatgacgg agctgagcatcaagagcggcaaccccggggagtcctcccgcggctggatgtggggctgcgccttctcgggggactcccagtacatcgtcac tgcttcctcggacaacctggcccggctctggtgtgtggagactggagagatcaagagagagtatggcggccaccagaaggctgttgtctgc ctggccttcaatgacagtgtgctgggctagtctagagcctgcagtctcgacaagcttgtcgagaagtactagaggatcataatcagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccgaacctgaaacataaaatgaatgcaattgttgttgttaacttgt ttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtcc aaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagcctagaagatccggctgctaacaaagcccgaaaggaagctgagtt ggctgctgccaccgctgagcaataactatcataaccctaggtgccatttcattacctctttctccgcacccgacatagatctgggccaacttttt ggcgaaaatgagacgttgatcggcacgtaagagggtccaactttcaccataatgaaataagatcactaccgggcgtatttttgagttatcg agattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacat tttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggcctttttaaagaccgtaaagaaaaataag cacaagttttatccggccttttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagctggtga tatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccggca gtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttcccctaaagggtttattgagaatatgtttttcgtctcag ccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacg caaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaa cagtactgcgatgagtggcagggcggggcgtaatttttttaaggcagttattggtgcccttaaacgcctggttgctacgcctgaataagtgat aataagcggatgaatggcagaaattcgaaagcaaattcgacccggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattg ctggtttaccggtttattgactaccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggccagtttgctcaggctctcccgtggaggta ataattgacgatatgatcatttattctgcctcccagctgacattcatccggggtcagcaccgtttctgcggactggctttctacgtgttccgcttcc tttagcagcccttgcgccctgagtgcttgcggcagcgtgaagctaatttctgtcagccgttaagtgttcctgtgtcactcaaaattgctttgagag gctctaagggcttctcagtgcgttacatccctggcttgttgtccacaaccgttaaaccttaaaagctttaaaagccttatatattcttttttttctta taaaacttaaaaccttagaggctatttaagttgctgatttatattaattttattgttcaaacatgagagcttagtacgtgaaacatgagagctt agtacgttagccatgagagcttagtacgttagccatgagggtttagttcgttaaacatgagagcttagtacgttaaacatgagagcttagta cgtgaaacatgagagcttagtacgtactatcaacaggttgaactgctgatcaacagatcctctacgcggccgcggtaccataacttcgtata gcatacattatacgaagttatctgccaggcacatgggttttactagtatcgattcgcgacctactccggaatattaatagatcatggagataa ttaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcat accgtcccaccatcgggcgcggatccaccatgaacacctccccaggcacggtgggcagtgacccggtcatcctggccactgcaggctacga ccacaccgtgcgcttctggcaggcccacagcggcatctgcacccggacggtgcagcaccaggactcccaggtgaatgccttggaggtcaca ccggaccgcagcatgattgctgctgcaggttaccagcacatccgcatgtatgatctcaactccaataaccctaaccccatcatcagctacgac ggcgtcaacaagaacatcgcgtctgtgggcttccacgaagacggccgctggatgtacacgggcggcgaggactgcacagccaggatctgg gacctcaggtcccggaacctgcagtgccagcggatcttccaggtgaacgcacccattaactgcgtgtgcctgcaccccaaccaggcagagct catcgtgggtgaccagagcggggctatccacatctgggacttgaaaacagaccacaacgagcagctgatccctgagcccgaggtctccatc acgtccgcccacatcgatcccgacgccagctacatggcagctgtcaatagcaccggaaactgcta*
pIDK	*cttgctggtggctagtctgtttcgaaatttttattggcggaaaggattatgaggtcgtataactgcactcccgtcagcagcccgcgtctgccgc ccacgtacatgcacgccatgtggcaagcctgggacctggctgttgacatctgtctgtctcagctgccgacgatcatcgaggaaggcactgcg tttcggcacagcccgttcttcgccgagcagctgaccgcattccaggtgtggctcaccatgggcgtggagaaccgaaacccaccgaagcagct gcccatcgtcctgcaggtgctgttaagccaagtgcaccggctgagagcattggacttgcttggaagatttttggacctgggtcccctgggcagt gagcctggccttgtctgtcggcatcttcccctacgtgctgaagctgctccagagctcggcccgagagctgcggccacttctcgttttcatctgggg ccaagatcctcgcagtggacagctcgtgccaggcggacctcgtgaaggacaacggccacaagtacttcctgtcggtcctggcggacccctа catgccagctgaacaccggaccatgacggctttcattctcgccgtgatcgtcaacagctatcacacggggcaggaagcctgccttcagggaa acctcattgccatctgcctggagcagctcaacgacccgcacccttgctgcgccagtgggtggccatctgcctcggcaggatctggcagaact tcgactcggcgaggtggtgcggcgtgagggacagcgctcatgagaagctctacagcctcctctccgaccccattcccgaggtccgctgcgca gcggtcttcgcccttggcacgttcgtgggcaactctgcagagaggacggaccactccaccaccatcgaccacaacgtggccatgatgctggc ccagctggtcagcgacgggagccccatggtccggaaggagctggtggtggctctgagtcatcttgtggttcagtatgaaagcaatttctgca ccgtggccctgcagttcatagaagaggaaaagaactacgccttgccttaggtaccgggagatgggggaggctaactgaaacacggaagga agacaatacggaaggaacccgcgctatgacggcaataaaaagacagaataaaacgcacgggtgttgggtcgtttgttcataaacgcgg ggttcggtcccagggctggcactctgtcgatacccccaccgagaccccattgggaccaatacgcccgcgtttcttccttttccccaccccaacccc caagttcgggtgaaggcccagggctcgcagccaacgtcggggcggcaagccctgccatagccactacgggtacgtttaaacccatgtgcct ggcagataacttcgtataatgtatgctatacgaagttatggtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaa gctctcatgtttaacgaactaaaccctcatggctaacgtactaagctctcatggctaacgtactaagctctcatgtttcacgtactaagctctca tgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaaggttttaagttttataagaaaaaaaagaatatataaggctttta aagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgagaagcccttagagcctctcaaagcaattttcagtgaca caggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaagggctgctaaaggaagcggaacacgtaga aagccagtccgcagaaacggtgctgaccccggatgaatgtcagctactgggctatctggacaagggaaaacgcaagcgcaaagagaaa gcaggtagcttgcagtgggcttacatggcgatagctagactgggcggttttatggacagcaagcgaaccggaattgccagctggggcgccc tctggtaaggttgggaagccctgcaaagtaaactggatggctttcttgccgccaaggatctgatggcgcaggggatcaagatctgatcaag agacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatga ctgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttgtcaagaccgacctgtc cggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtca ctgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatc atggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgta ctcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggc gcgcatgcccgacggcgaggatctcgtcgtgacacatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcat cgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggct gaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctgg ggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatc gttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccgggatctatgtcgggtgcggagaa agaggtaatgaaatggcacctaggtatcgatactagtatacggacctttaattcaacccaacacaatatattatagttaaataagaattatt atcaaatcatttgtatattaattaaaatactatactgtaaattacattttatttacaatcactcgacgaagacttgatcacccgggatctcgaga ccatggagtccgaaatgctgcaatcgcctcttctgggcctgggggaggaagatgaggctgatcttacagactggaacctacctttggctttta*

	tgaaaaagaggcactgtgagaaaattgaaggctccaaatccttagctcagagctggaggatgaaggatcggatgaagacagtcagtgtt gccttagttttgtgcctgaatgttggtgtggaccctcccgatgtggtgaagaccacgccctgtgcacgcttggaatgctggatcgatcctctgtc gatgggtcctcagaaagctctggaaaccatcggtgcaaatttacagaagcagtacgagaactggcagccaagggcccggtacaagcaga gccttgacccaactgtggatgaagtcaagaagctctgcacgtccttacgtcgcaacgccaaggaggagcgagtcctctttcactacaatggc cacggggtgcccggcccacagtcaacggggaggtctgggtcttcaacaagaactacacgcagtacatccctctgtccatatatgacctgca gacgtggatgggcagcccgtcgatcttcgtctacgactgctccaatgctggcttgatcgtcaagtccttcaagcagttcgcactacagcggga gcaggagctggaggtagctgcaatcaacccaaatcaccctcttgctcagatgcctttgcctccgtcgatgaaaaactgcatccagctggcag cctgcgaggccaccgagctgctgcccatgatccccgacctcccggctgacctattcacctcctgcctcaccacccccatcaagatcgccctgcg ctggttttgcatgcagaaatgtgtcagtctggtgcctggcgtcacactggatttgatagaaaagatccctggccgcctgaacgacaggagga cgcccctgggtgaactgaactggatcttcacagccatcacagacaccatcgcgtggaacgtgctcccccgggatctcttccaaaagctcttca gacagga
pIDS	*cgatactagtatacggacctttaattcaacccaacacaatatattatagttaaataagaattattatcaaatcatttgtatattaattaaaata ctatactgtaaattacattttatttacaatcactcgacgaagacttgatcacccgggatctcgagccatggccttcttggacaatccaactatca ttctagctcatattcgacagtcacatgtgaccagtgatgacacgggaatgtgtgagatggttctcattgatcatgatgttgacctagagaaga ttcatcctccttcaatgcctggagacagtgggtcagaaattcagggaagcaatggtgagactcagggctatgtatatgcccagtcagtcgat attacctcaagttgggactttggtattagaagacgctcaaacacagctcaaagattagaacgactccgaaaagagagacaaaaccagatc aaatgcaaaaatattcagtggaaagaaagaaattctaagcaatcagcccaggagttaaagtcactgtttgaaaaaaaatctctcaaaga gaagcctccaatttctgggaagcagtcgatattatctgtacgcctagaacagtgccctctgcagctgaataacccttttaacgagtattccaa atttgatggcaagggtcatgtaggtacaacagcaaccaagaagatcgatgtctacctccctctgcactcgagccaggacagactgctgcca atgaccgtggtgacaatggccagcgccagggtgcaggacctgatcgggctcatctgctggcagtatacaagcgaaggacgggagccgaa gctcaatgacaatgtcagtgcctactgcctgcatattgctgaggatgatggggaggtggacaccgatttccccccgctggattccaatgagcc cattcataagtttggcttcagtactttggccctggttgaaaagtactcatctcctggtctgacatccaaagagtcactctttgttcgaataaatg ctgctcatggattctcccttattcaggtggacaacacaaaggttaccatgaaggaaatcttactgaaggcagtgaagcgaagaaaaggatc ccagaaagtttcaggccctcagtaccgcctggagaagcagagcgagcccaatgtcgccgttgacctggacagcactttggagagccagag cgcatgggagttctgcctggtccgcgagaacagttcaagggcagacgggtttttgaggaggattcgcaaattgacatagccacagtacag gatatgcttagcagccaccattacaagtcattcaaagtcagcatgatccacagactgcgattcacaaccgacgtacagctaggtatctctgg agacaaagtagagatagaccctgttacgaatcagaaagccagcactaagttttggattaagcagaaacccatctcaatcgattccgacctg ctctgtgcctgtgaccttgctgaagagaaaagccccagtcacgcaatatttaaactcacgtatctaagcaatcacgactataaacacctctac tttgaatcggacgctgctaccgtcaatgaaattgtgctcaaggttaactacatcctggaatcgcgagctagcactgcccgggctgactacttt gctcaaaaacaaagaaaactgaacagacgtacgagcttcagcttccagaaggagaagaaatccgggcagcagtgatgctagcagctga tgcatagcatgcggtaccgggagatgggggaggctaactgaaacacggaaggagacaataccggaaggaacccgcgctatgacggca ataaaaagacagaataaaacgcacgggtgttgggtcgtttgttcataaacgcggggttcggtcccagggctggcactctgtcgataccccca ccgagacccattgggaccaatacgcccgcgtttcttccttttccccaccccaacccccaagttcgggtgaaggcccagggctcgcagccaac gtcggggcggcaagccctgccatagccactacgggtacgtttaaacccatgtgcctggcagataacttcgtataatgtatgctatacgaagtt atggtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaac gtactaagctctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaactt aaatagcctctaaggttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagc cagggatgtaacgcactgagaagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacatAattcagcttc acgctgccgcaagcactcagggcgcaagggctgctaaaggaagcggaacacgtagaaagccagtccgcagaaacggtgctgaccccgg atgaatgtcagctgggaggcagaataaatgatcatatcgtcaattattacctccacggggagagcctgagcaaactggcctcaggcatttg agaagcacacggtcacactgcttccggtagtcaataaaccggtaagtagcgtatgcgctcacgcaactggtccagaaccttgaccgaacgc agcggtggtaacggcgcagtggcggttttcatggcttgttatgactgttttttggggtacagtctatgcctcgggcatccaagcagcaagcg cgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaaacatcatga gggaagcggtgatcgccgaagtatcgactcaactatcagaggtagttggcgtcatcgagcgccatctcgaaccgacgttgctggccgtaca tttgtacggctccgcagtggatggcggcctgaagccacacagtgatattgatttgctggttacggtgaccgtaaggcttgatgaaacaacgc ggcgagctttgatcaacgacctttttggaaacttcggcttcccctggagagagcgagattctccgcgctgtagaagtcaccattgttgtgcacg acgacatcattccgtggcgttatccagctaagcgcgaactgcaatttggagaatggcagcgcaatgacattcttgcaggtatcttcgagcca gccacgatcgacattgatctggctatcttgctgacaaaagcaagagaacatagcgttgccttggtaggtccagcggcggaggaaactctttga tccggttcctgaacaggatctatttgaggcgctaaatgaaaccttaacgctatggaactcgccgcccgactgggctggcgatgagcgaaatg tagtgcttacgttgtcccgcatttggtacagcgcagtaaccggcaaaatcgcgccgaaggatgtcgctgccgactgggcaatggagcgcctg ccggcccagtatcagcccgtcatacttgaagctagacaggcttatcttggacaagaagaagatcgcttggcctcgcgcgcagatcagttgga agaatttgtccactacgtgaaaggcgagatcaccaaggtagtcggcaaataatgtctaacaattcgttcaagccgacggatctatgtcgggt gcggagaaagaggtaatgaaatggcacctaggtat*

B. List of Primers

PRIMER NAME	SEQUENCE
GFA-1.12-1.1.1-F	GGTGACGTGGAGGAGAATCCCGGGCCTTCTAGAGGATCCGGCTCCGGCGAGGGCAGGGG
GFA-1.12-1.1.1-F	GGTGACGTGGAGGAGAATCCCGGGCCTTCTAGAGGATCCGGCTCCGGCGAGGGCAGGGG
GFA-1.12-1.1.1-R	GGGCATTGGTCGATCGACCACTGTGCTGGCGGCCGGCCAGGCCGCGGGGAGGCGGCCC
GFA-1.12-1.1.1-R	GGGCATTGGTCGATCGACCACTGTGCTGGCGGCCGGCCAGGCCGCGGGGAGGCGGCCC
GFA-1.5-0.0.1-1F	CATTAACCTATAAAAATAGGGGTACCATAACTTCGTATAGCA
GFA-1.5-0.0.1-1R	TCCTCGCCCTTGCTCACCATCCTGCAGGTGATAAGCCAGT
GFA-1.5-0.0.1-2F	ACTGGCTTATCACCTGCAGGATGGTGAGCAAGGGCGAG
GFA-1.5-0.0.1-2R	GGCCAGGCCGCGGGGAGGCGTCATTACTATCATTACTATCATTACT
GFA-1.5-0.0.1-3F	GATAGTAATGATAGTAATGACGCCTCCCCGCGG
GFA-1.5-0.0.1-3R	TCTGTGGATAACCGTATTACCTGCCCGACCTAGGGCA
GFA-1.5-0.0.1-4F	CTATACGAAGTTATGGTACCCCTATTTTTATAGGTTAATGTCATGA
GFA-1.5-0.0.1-4R	ACCTGCCCTAGGTCGGGCAGGTAATACGGTTATCCACAGAATCAGG
GFA-1.5-1.0-2R	AGTAGTGAGAAGTTTGTTGCACCACTTTGTACAAGAAAGCTGA
GFA-1.5-1.0-3F	GCTTTCTTGTACAAAGTGGTGCAACAAACTTCTCACTACTCA
GFA-1.5-6.0-1F	ACCTGCCCTAGGTCGGGCAGCAACGCGTATATCTGGCCCG
GFA-1.5-6.0-2F	ATGGGTATACCCTAGTATACGGTACCATAACTTCGTATAGCA
GFA-1.5-6.0-3R	CGGGCCAGATATACGCGTTGCTGCCCGACCTAGGGCA
GFA-1.5-6.1-1R	CTATACGAAGTTATGGTACCGTATACTAGGGTATACCCATCT
GFA-1.5-7.0-1F	ATAGCTGTTTCCTGTGTGAACTGCAGGTCGACTCTAGACC
GFA-1.5-7.0-1R	AGGGGATATCAGCTGGATGGTGTCAGACCAAGTTTACTCA
GFA-1.5-7.0-2F	TGAGTAAACTTGGTCTGACACCATCCAGCTGATATCCCCT
GFA-1.5-7.0-2R	GGTCTAGAGTCGACCTGCAGTTCACACAGGAAACAGCTATGAC
GFA-1.6-1.1.1-F	GAAATCGTCGTGGTATTCACGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
GFA-1.6-1.1.1-R	AAGGACAATTACAAACAGGATTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCAC
GFA-1.6-1.1.2-F	ATCATGACATTAACCTATAATCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCG
GFA-1.6-1.1.2-R	TTTTGCTCACATGTTCTTTCGTGAATACCACGACGATTTCCGGCAGTTTCT
GFA-1.6-1.1.3-F	CTGCGCCGGTTGCATTCGATACAGAATCAGGGGATAACGCAGGAAAGAACAT
GFA-1.6-1.1.3-R	AACGTTTTCATCGCTCTGGATATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCAC
GFA-1.6-1.1.4-F	TATCATGACATTAACCTATATCCAGAGCGATGAAAACGTTTCAGTTTGCTCAT
GFA-1.6-1.1.4-R	GCGTTATCCCCTGATTCTGTATCGAATGCAACCGGCGCAGG
GFA-1.6-1.1.5-F	CTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGT
GFA-1.6-1.1.5-R	AACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCT
GFA-1.6-1.1.6-F	GAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGG
GFA-1.6-1.1.6-R	AAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAG
GFA-1.6-6.6.1-F	TTTAGCTTCCTTAGCTCCTGAAAATCTCGACGGAT
GFA-1.6-6.6.1-R	GTTCTTGCAGCTCGGTGACCCGCTCGATGTAATATTCCGGAGTAGGTCGCGAATCGATACTAGTAAAACCCATGTGC
GFA-1.6-6.6.2-F	TCGAGATTTTCAGGAGCTAAGGAAGCTAAAATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACA
GFA-1.6-6.6.2-R	GGACTATGACACTAGCGTATATGAATAGGTTCGGTCGTTCGGCTGCGGC
GFA-1.6-6.6.3-F	TGATACCGCTCGCCGCAGCCGAACGACCGAACCTATTCATATACGCTAGTGTCATAGTCCTGAAAATCATCTGC
GFA-1.6-6.6.3-R	CAGTCACTATGAATCAACTACTTAGATCGAAGAGGTATGCTATGAAGCAGCGTATTACAGTGAC
GFA-1.6-6.6.4-F	GTATCGATTCGCGACCTACTCCGGAATATTACATCGAGCGGGTCACCGAGCTGCAAGAACTCT
GFA-1.6-6.6.4-R	TCGATCTAAGTAGTTGATTCATAGTGACTGGATATGTTGTGTTTTACAG
GFA-1.6-6.7.1-F	GTGGATCCGCGCCAAAACCGCGGTCTAGAGCGACCTACTCCGGAATATTAATAGATGTACGGGC
GFA-1.6-6.7.1-R	CCTTTTTGATAATCTCATGACCAAAATCCCATCCCCTGATTCTGTGGATAACCGTATTACCGG
GFA-1.6-6.7.2-F	GTAATACGGTTATCCACAGAATCAGGGGATGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC
GFA-1.6-6.7.2-R	GAGCTTGCGCCGTCCCGTCAAGTCAGCGTATAGACGTCAGGTGGCACTTTTCGGGG
GFA-1.6-6.7.3-F	ATTTCCCCGAAAAGTGCCACCTGACGTCTATACGCTGACTTGACGGGACGGC
GFA-1.6-6.7.3-R	GGGTTTTACTAGTATCGATTCGCGACCTACCGACCTGCAGTTCACACAGGAAACAGCTATGA
GFA-1.6-6.7.4-F	ATAGCTGTTTCCTGTGTGAACTGCAGGTCGGTAGGTCGCGAATCGATACTAGTAAAACCCATGTGCCT
GFA-1.6-6.7.4-R	GTACATCTATTAATATTCCGGAGTAGGTCGCTCTAGACCGCGGTTTTGGCGC
GFA-Z.A-1.1.1-F	CCGTAGAAAAGATCAAAGGATCTTCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTATACGC
GFA-Z.A-1.1.1-R	ATTTTGAGACACGGGCCAGAGCTGCACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTC ATTTTT
GFA-Z.A-1.2.1-F	ACAGTGCAGCTCTGGCCCGTGTCTCAAAATGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAA G
GFA-Z.A-1.2.1-R	TGTGCGAAGATCCTTTGATCTTTTCTACGGGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAC GACGCG
GFA-Z.A-GWCAM-F	CCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGG
GFA-Z.A-GWCAM-R	CCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGG
GFA-Z.A-GWORI-F	GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGG
GFA-Z.A-GWORI-R	CCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
GFA-Z.D-1.1.1-F	CGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACC
GFA-Z.D-1.1.1-R	TGGGAAAACATCCATGCTAGCGTTAACGCGCGGATAAAATGCTTGATGGTCGGAAGAGGC
GFA-Z.D-1.1.2-F	GGCTTTTAGTAAGCCGGATCCACGCGGCGTACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCG
GFA-Z.D-1.1.2-R	TACCGGGAAGCCCTGGGCCAACTTTTGGCGGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
GFA-Z.D-1.1.3-F	CAGGAACCGTAAAAAGGCCGCGTTGCTGGCCGCGTTAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGA CGGCC
GFA-Z.D-1.1.3-R	ACGCCGCGTGGATCCGGCTTACTAAAAGCCAGATAACAGTATGCGTATTTGCGCGC

GFA-Z.D-1.4.1-F	GCTGTTTCCTGTGTGAAATTGTTATCCGCCTCCCACACCTCCCCCTGAACCTG
GFA-Z.D-1.4.1-R	ACTGGCCGTCGTTTTACAACGTCGTGACTGGGCCGCGCCCGATGGTGGGACGG
GFA-Z.D-2.4.1-F	GCTGTTTCCTGTGTGAAATTGTTATCCGCCGGGAGATGGGGGAGGCTAACTGAAACACG
GFA-Z.D-2.4.1-R	GGCCGTCGTTTTACAACGTCGTGACTGGGGGTCTCGAGATCCCGGGTGATCAAGTCTTCGTCG
GFA-Z.D-4.1.1-F	TACATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAGAGGGCCGGCAATAAAAAGACAGAATAAAACGCACG GGTGTTGGGTCGT
GFA-Z.D-4.1.1-R	CGTCGTGACTGGGAAAACCCTGGCGCGATTGTAAATAAAATGTAATTTACAGTATAGTATTTTAATTAATATACAAATG ATTTGATAATAATTCTTATTTAACTATAATATATTGTGTTGGG
GFA-Z.D-A.2.1-F	CGTAAAATGATATAAATATCAATATATTAAATTAGATTTTGCATAAAAAAACAGACTACATAATACTGTAAAACACAACA TATCCAGTCACTATGGTCGACCGACAGCCTTCCAAATGTTCTTCGGGTGATGCTGCC
GFA-Z.D-A.2.1-R	GCTGAACGAGAAACGTAAAATGATATAAATATCAATATATTAAATTAGATTTTGCATAAAAAAACAGACTACATAATACT GTAAAACACAACATATCCGTCGACTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCG
GFA-Z.D-A.2.2-R	TAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGGTCCCCTATAGTGAGTC GTATTACATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAGAGGGC
GFA-Z.D-A.3.1-F	CCCAGTCACGACGTTGTAAAACGACGGCCAGTGCGCCAGGGTTTTCCCAGTCACGACG
GFA-Z.D-A.3.1-R	CGGATAACAATTTCACACAGGAAACAGCTATGACCATGTAATACGACTCACTATAGGGG
GFA-Z.D-K.2.2-F	ATGCAAAATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTTTTGTACAAACTTGTCGTCG TGACTGGGAAAACCCTGGCGCGATTGTAAATAAAATGTAATTTACAG
GVA-1.5-7.0-B-2R	TTCCGCCTCAGAAGCCATAGTTTCTACGGGGTCTGACGCT
GVI-1.5-6.0-A-2R	CCACTGCGCCGTTACCACCGAACTAGGCTCAAGCAGTGATCA
GVI-1.5-6.0-A-3F	ATCACTGCTTGAGCCTAGTTCGGTGGTAACGGCGCA
GVI-1.5-6.0-B-2R	GCTTTTTTGTACAAACTTGTTTTCTAATAGATACCGGTGGATCCC
GVI-1.5-6.0-B-3F	CCACCGGTATCTATTAGAAAACAAGTTTGTACAAAAAAGCTGAACG
GVI-1.5-6.0-C-2R	AGTTTTTGTTCGAAGGGCCCTCATTACTATCATTACTATCATTACT
GVI-1.5-6.0-C-3F	GATAGTAATGATAGTAATGAGGGCCCTTCGAACAAAAACTC
GVI-1.5-7.0-A-1F	GGGCTCTAGGGGGTATCCCGGTAATACGGTTATCCACAGAATCAGG
GVI-1.5-7.0-A-2R	TTTTTGTACAAACTTGTGATTTTCTAATAGATACCGGTGGATCCC
GVI-1.5-7.0-A-3F	CCACCGGTATCTATTAGAAAATCACAAGTTTGTACAAAAAAGCTGA
GVI-1.5-7.0-A-3R	TCTGTGGATAACCGTATTACCGGGATACCCCCTAGAGCC
GVI-1.5-7.0-B-3F	AGCGTCAGACCCCGTAGAAACTATGGCTTCTGAGGCGGAA
GWA-1.6-11.1.1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACCACCGGTGGCTGCCGCCACCATGGTGAGCAAGGGCG
GWA-1.6-11.1.1-R	GGGGACCACTTTGTACAAGAAAGCTGGGGGCGCCGGTGGAGTGGCGGCCCTCGGCGCGCTCG
GWA-1.7-1.1.1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGGCTAGCGTTTAAACTTAAGCTTGCCACCATGGAAGAGCAATACCGCCC GG
GWA-1.7-1.1.1-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTCTATCATTAAACGGGCCCGCCAACGACCAGATTGAGGAGTTTACCTGG
GWA-1.7-2.1.2-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAACCCAAGCTGGCTAGCGCCACCATGGCAAGC
GWA-1.7-2.1.2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTATTAAACGGGCCCTTTCCAGCAAATCAGACAGTAATTCTTTTTAC CGCGACGC
GWA-1.7-3.1.2-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATAAAAAACCATTAGATGTTTTAATATCTGCGACCGGGCTCTGG ATGTCCAGG
GWA-1.7-3.1.2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTATTACAGATTGGTTGAAATCCCATTATAGTAAGATTCGGACCTTG ATGCCCTCTTAATGTTTTTAAAGCCGTGC
GWA-1.7-4.1.1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGGATCCATGACTAAGCCCATCGTTTTTGCTGGCGCACAGCCC
GWA-1.7-4.1.1-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTGGATCCCTATCATTACGGCTTCGCCACAAAACCAATCGC
GWA-Z.B-2.1.1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGCCACCATGGTGAGCAAGGGCGAGGAGGTCATCAAGGAGTTCATGCGC
GWA-Z.B-2.1.1-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTATTAGGCGCCGGTGGAGTGGCGGCCCTCGGCGCGC
GWA-Z.B-3.1.1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG
GWA-Z.B-3.1.1-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTATTCATTACTACTTGTACAGCTCGTCCATGCCGAGAGTGATCCCG GCGGCGG
Mex13-F	GCGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCC
Mex13-R	GCCCTCTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACC
PMI-1.13-1.1.1-F	GGGCGAGGAGCTGTTCTAAGGGGTGGTGCCCATCCTGG
PMI-1.13-1.1.1-R	CCAGGATGGGCACCACCCCTTAGAACAGCTCCTCGCCC
PMI-1.13-1.1.2-F	GGGCGAGGAGCTGTTCTAGGGGGTGGTGCCCATCCTGG
PMI-1.13-1.1.2-R	CCAGGATGGGCACCACCCCCTAGAACAGCTCCTCGCCC
PMI-1.13-1.1.3-F	GGGCGAGGAGCTGTTCTGAGGGGTGGTGCCCATCCTGG
PMI-1.13-1.1.3-R	CCAGGATGGGCACCACCCCTCAGAACAGCTCCTCGCCC
PMI-1.13-10.1.1-F	GGCACAACATCGAGGACGGCAGCTAACAGCTCGCCGACC
PMI-1.13-10.1.1-R	GGTCGGCGAGCTGTTAGCTGCCGTCCTCGATGTTGTGCC
PMI-1.13-10.1.2-F	GGCACAACATCGAGGACGGCAGCTAGCAGCTCGCCGACC
PMI-1.13-10.1.2-R	GGTCGGCGAGCTGCTAGCTGCCGTCCTCGATGTTGTGCC
PMI-1.13-10.1.3-F	GGCACAACATCGAGGACGGCAGCTGACAGCTCGCCGACC
PMI-1.13-10.1.3-R	GGTCGGCGAGCTGTCAGCTGCCGTCCTCGATGTTGTGCC
PMI-1.13-11.1.1-F	GCAGCTCGCCGACCACTAACAGCAGAACACCCCC
PMI-1.13-11.1.1-R	GGGGGTGTTCTGCTGTTAGTGGTCGGCGAGCTGC
PMI-1.13-11.1.2-F	GCAGCTCGCCGACCACTAGCAGCAGAACACCCCC
PMI-1.13-11.1.2-R	GGGGGTGTTCTGCTGCTAGTGGTCGGCGAGCTGC
PMI-1.13-11.1.3-F	GCAGCTCGCCGACCACTGACAGCAGAACACCCCC
PMI-1.13-11.1.3-R	GGGGGTGTTCTGCTGTCAGTGGTCGGCGAGCTGC
PMI-1.13-12.1.1-F	GCAGAACACCCCCATCGGCTAAGGCCCCGTGCTGCTGCCC
PMI-1.13-12.1.1-R	GGGCAGCAGCACGGGGCCTTAGCCGATGGGGGTGTTCTGC
PMI-1.13-12.1.2-F	GCAGAACACCCCCATCGGCTAGGGCCCCGTGCTGCTGCCC
PMI-1.13-12.1.2-R	GGGCAGCAGCACGGGGCCCTAGCCGATGGGGGTGTTCTGC
PMI-1.13-12.1.3-F	GCAGAACACCCCCATCGGCTGAGGCCCCGTGCTGCTGCCC

PMI-1.13-12.1.3-R	GGGCAGCAGCACGGGGCCTCAGCCGATGGGGGTGTTCTGC
PMI-1.13-13.1.1-F	CCACTACCTGAGCACCTAATCCGCCCTGAGCAAAGACCCC
PMI-1.13-13.1.1-R	GGGGTCTTTGCTCAGGGCGGATTAGGTGCTCAGGTAGTGG
PMI-1.13-13.1.2-F	CCACTACCTGAGCACCTAGTCCGCCCTGAGCAAAGACCCC
PMI-1.13-13.1.2-R	GGGGTCTTTGCTCAGGGCGGACTAGGTGCTCAGGTAGTGG
PMI-1.13-13.1.3-F	CCACTACCTGAGCACCTGATCCGCCCTGAGCAAAGACCCC
PMI-1.13-13.1.3-R	GGGGTCTTTGCTCAGGGCGGATCAGGTGCTCAGGTAGTGG
PMI-1.13-14.1.1-F	CCGCCCTGAGCAAAGACCCCTAAGAGAAGCGCGATCACATGGTCC
PMI-1.13-14.1.1-R	GGACCATGTGATCGCGCTTCTCTTAGGGGTCTTTGCTCAGGGCGG
PMI-1.13-14.1.2-F	CCGCCCTGAGCAAAGACCCCTAGGAGAAGCGCGATCACATGGTCC
PMI-1.13-14.1.2-R	GGACCATGTGATCGCGCTTCTCCTAGGGGTCTTTGCTCAGGGCGG
PMI-1.13-14.1.3-F	CCGCCCTGAGCAAAGACCCCTGAGAGAAGCGCGATCACATGGTCC
PMI-1.13-14.1.3-R	GGACCATGTGATCGCGCTTCTCTCAGGGGTCTTTGCTCAGGGCGG
PMI-1.13-15.1.1-F	CCTGAGCAAAGACCCCAACGAGTAACGCGATCACATGGTCCTGCTGG
PMI-1.13-15.1.1-R	CCAGCAGGACCATGTGATCGCGTTACTCGTTGGGGTCTTTGCTCAGG
PMI-1.13-15.1.2-F	CCTGAGCAAAGACCCCAACGAGTAGCGCGATCACATGGTCCTGCTGG
PMI-1.13-15.1.2-R	CCAGCAGGACCATGTGATCGCGCTACTCGTTGGGGTCTTTGCTCAGG
PMI-1.13-15.1.3-F	CCTGAGCAAAGACCCCAACGAGTGACGCGATCACATGGTCCTGCTGG
PMI-1.13-15.1.3-R	CCAGCAGGACCATGTGATCGCGTCACTCGTTGGGGTCTTTGCTCAGG
PMI-1.13-2.1.1-F	CCGGCGAGGGCGAGGGCGATGCCACCTAAGGCAAGCTGACCCTTAAGTTCATCTGCACC
PMI-1.13-2.1.1-R	GGTGCAGATGAACTTAAGGGTCAGCTTGCCTTAGGTGGCATCGCCCTCGCCCTCGCCGG
PMI-1.13-2.1.2-F	CCGGCGAGGGCGAGGGCGATGCCACCTAGGGCAAGCTGACCCTTAAGTTCATCTGCACC
PMI-1.13-2.1.2-R	GGTGCAGATGAACTTAAGGGTCAGCTTGCCCTAGGTGGCATCGCCCTCGCCCTCGCCGG
PMI-1.13-2.1.3-F	CCGGCGAGGGCGAGGGCGATGCCACCTGAGGCAAGCTGACCCTTAAGTTCATCTGCACC
PMI-1.13-2.1.3-R	GGTGCAGATGAACTTAAGGGTCAGCTTGCCTCAGGTGGCATCGCCCTCGCCCTCGCCGG
PMI-1.13-3.1.1-F	CCCTTAAGTTCATCTGCACCACCGGCTAACTGCCCGTGCCCTGGCCC
PMI-1.13-3.1.1-R	GGGCCAGGGCACGGGCAGTTAGCCGGTGGTGCAGATGAACTTAAGGG
PMI-1.13-3.1.2-F	CCCTTAAGTTCATCTGCACCACCGGCTAGCTGCCCGTGCCCTGGCCC
PMI-1.13-3.1.2-R	GGGCCAGGGCACGGGCAGCTAGCCGGTGGTGCAGATGAACTTAAGGG
PMI-1.13-3.1.3-F	CCCTTAAGTTCATCTGCACCACCGGCTGACTGCCCGTGCCCTGGCCC
PMI-1.13-3.1.3-R	GGGCCAGGGCACGGGCAGTCAGCCGGTGGTGCAGATGAACTTAAGGG
PMI-1.13-4.1.1-F	GGCTACGTCCAGGAGCGCACCATCTAATTCAAGGACGACGGC
PMI-1.13-4.1.1-R	GCCGTCGTCCTTGAATTAGATGGTGCGCTCCTGGACGTAGCC
PMI-1.13-4.1.2-F	GGCTACGTCCAGGAGCGCACCATCTAGTTCAAGGACGACGGC
PMI-1.13-4.1.2-R	GCCGTCGTCCTTGAACTAGATGGTGCGCTCCTGGACGTAGCC
PMI-1.13-4.1.3-F	GGCTACGTCCAGGAGCGCACCATCTGATTCAAGGACGACGGC
PMI-1.13-4.1.3-R	GCCGTCGTCCTTGAATCAGATGGTGCGCTCCTGGACGTAGCC
PMI-1.13-5.1.1-F	GCCGAGGTGAAGTTCGAGGGCTAAACCCTGGTGAACCGC
PMI-1.13-5.1.1-R	GCGGTTCACCAGGGTTTAGCCCTCGAACTTCACCTCGGC
PMI-1.13-5.1.2-F	GCCGAGGTGAAGTTCGAGGGCTAGACCCTGGTGAACCGC
PMI-1.13-5.1.2-R	GCGGTTCACCAGGGTCTAGCCCTCGAACTTCACCTCGGC
PMI-1.13-5.1.3-F	GCCGAGGTGAAGTTCGAGGGCTGAACCCTGGTGAACCGC
PMI-1.13-5.1.3-R	GCGGTTCACCAGGGTTCAGCCCTCGAACTTCACCTCGGC
PMI-1.13-6.1.1-F	GGGCATCGACTTCAAGTAAGACGGCAACATCCTGGGGC
PMI-1.13-6.1.1-R	GCCCCAGGATGTTGCCGTCTTACTTGAAGTCGATGCCC
PMI-1.13-6.1.2-F	GGGCATCGACTTCAAGTAGGACGGCAACATCCTGGGGC
PMI-1.13-6.1.2-R	GCCCCAGGATGTTGCCGTCCTACTTGAAGTCGATGCCC
PMI-1.13-6.1.3-F	GGGCATCGACTTCAAGTGAGACGGCAACATCCTGGGGC
PMI-1.13-6.1.3-R	GCCCCAGGATGTTGCCGTCTCACTTGAAGTCGATGCCC
PMI-1.13-7.1.1-F	GGGGCACAAGCTGGAGTAAAACTATAACAGCCACAACG
PMI-1.13-7.1.1-R	CGTTGTGGCTGTTATAGTTTTACTCCAGCTTGTGCCCC
PMI-1.13-7.1.2-F	GGGGCACAAGCTGGAGTAGAACTATAACAGCCACAACG
PMI-1.13-7.1.2-R	CGTTGTGGCTGTTATAGTTCTACTCCAGCTTGTGCCCC
PMI-1.13-7.1.3-F	GGGGCACAAGCTGGAGTGAAACTATAACAGCCACAACG
PMI-1.13-7.1.3-R	CGTTGTGGCTGTTATAGTTTCACTCCAGCTTGTGCCCC
PMI-1.13-8.1.1-F	CTATAACAGCCACAACGTCTAAATCATGGCTGACAAGCAGAAGAACGGC
PMI-1.13-8.1.1-R	GCCGTTCTTCTGCTTGTCAGCCATGATTTAGACGTTGTGGCTGTTATAG
PMI-1.13-8.1.2-F	CTATAACAGCCACAACGTCTAGATCATGGCTGACAAGCAGAAGAACGGC
PMI-1.13-8.1.2-R	GCCGTTCTTCTGCTTGTCAGCCATGATCTAGACGTTGTGGCTGTTATAG
PMI-1.13-8.1.3-F	CTATAACAGCCACAACGTCTGAATCATGGCTGACAAGCAGAAGAACGGC
PMI-1.13-8.1.3-R	GCCGTTCTTCTGCTTGTCAGCCATGATTCAGACGTTGTGGCTGTTATAG
PMI-1.13-9.1.1-F	CGTCTATATCATGGCTGACAAGTAAAAGAACGGCATCAAGGTGAACTTCAAGATTAGGC
PMI-1.13-9.1.1-R	GCCTAATCTTGAAGTTCACCTTGATGCCGTTCTTTTACTTGTCAGCCATGATATAGACG
PMI-1.13-9.1.2-F	CGTCTATATCATGGCTGACAAGTAGAAGAACGGCATCAAGGTGAACTTCAAGATTAGGC
PMI-1.13-9.1.2-R	GCCTAATCTTGAAGTTCACCTTGATGCCGTTCTTCTACTTGTCAGCCATGATATAGACG
PMI-1.13-9.1.3-F	CGTCTATATCATGGCTGACAAGTGAAAGAACGGCATCAAGGTGAACTTCAAGATTAGGC
PMI-1.13-9.1.3-R	GCCTAATCTTGAAGTTCACCTTGATGCCGTTCTTTCACTTGTCAGCCATGATATAGACG
PMI-1.6-1.1.1-F	GCCAACTTTGTATACAAAAGTTGAACGAGAAACG
PMI-1.6-1.1.1-R	CGTTTCTCGTTCAACTTTTGTATACAAAGTTGGC
PMI-1.6-1.1.2-F	CGTTTCTCGTTCAACTTTTCTATACAAAGTTGGC
PMI-1.6-1.1.2-R	GCCAACTTTGTATAGAAAAGTTGAACGAGAAACG

C. List of Plasmids Backbones Assembled Inhouse

PLASMID NAME	PLASMID SEQUENCE
GABB	*cattaacctataaaaatagggtaccataacttcgtatagcatacattatacgaagttatctgtgtacgggccagatatacgcgttgacattgattatt gactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctga ccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactattta cggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcc cagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcg tggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatg tcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccact gcttactggcttatcacctgcaggatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacg gccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgcc ctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcc cgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtga accgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatc atggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagc agaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcg atcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtggttctggtcaccaccacc accaccaccaccaccaccaccacggttcaactagtggttctggtgagcagaagctgatctccgaggaggacctgggttcaactagtggttctggtgactac aaggacgacgacgacaagctttagtaatgatagtaatgatagtaatgacgcctccccgcggcctggccggccgccagcacagtggtcgatcgaccа atgccctggctcacaaataccactgagatcttttcccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaag gaaatttattttcattgcaatagtgtgttggaatttttgtgtctctcactcggaaggacatatgggagggcaaatcatttaaaacatcagaatgagtat ttggtttagagtttggcaacatatgcccatatgctggctgccatgaacaaaggttggctataaagaggtcatcagtatatgaaacagcccctgctgtc cattccttattccatagaaaagccttgacttgaggttagatttttttatatttgttttgtgttatttttttctttaacatcccaaaatttccttacatgtttta ctagccagatttttcctcctctcctgactactcccagtcatagctgtccctcttctcttatggagatccctcgacctgccctaggtcgggcaggtaatacgg ttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcg ttttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgt ttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatag ctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggt aactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggt gctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaag agttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaa gaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacc tagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacct atctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgc aatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaacttt atccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatc gtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggtta gctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccg taagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggata ataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagt tcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaa aaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggata catatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatga*
GD54	*ctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggtt gatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtc actcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccag gatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggcttttcaaaaatatggtattgataatcctgatatgaataaatt gcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagc tcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgc gcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggct tcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgct aatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggct gaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgctt cccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggta tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgc ggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagc tgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgc gttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccagg aagagtttgtagaaacgcaaaaaggccatccgtcaggatggccttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctcc gggccgttgcttcacaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccag tcttccgactgagcctttcgttttatttgatgcctggcagttccctactctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgac ggccagtcttaagctcgggccccaaataatgattttattttgactgatagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaa ctttgtatacaaaagttgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaacagactacataatactgtaa aacacaacatatccagtcactatgaatcaactacttagatggtattagtgacctgtagtcgaccgacagccttccaaatgttcttcgggtgatgctgcc aacttagtcgaccgacagccttccaaatgttcttctcaaacggaatcgtcgtatccagcctactcgctattgtcctcaatgccgtattaaatcataaaaa gaaataagaaaaagaggtgcgagcctcttttttgtgtgacaaaataaaaacatctacctattcatatacgctagtgtcatagtcctgaaaatcatctg catcaagaacaatttcacaacttctatactttttctcttacaagtcgttcggcttcatctggattttcagcctctatacttactaaacgtgataaagtttctgt aatttctactgtatcgacctgcagactggctgtgtataagggagcctgacatttatattccccagaacatcaggttaatggcgtttttgatgtcattttcg cggtggctgagatcagccacttcttccccgataacggagaccggcacactggccatatcggtggtcatcatgcgccagctttcatccccgatatgcacc*

	accgggtaaagttcacgggagactttatctgacagcagacgtgcactggccaggggatcaccatccgtcgcccgggcgtgtcaataatatcactct gtacatccacaaacagacgataacggctctctcttttataggtgtaaaccttaaactgcatttcaccagccctgttctcgtcagcaaaagagccgttca tttcaataaaccgggcgacctcagccatcccttcctgattttccgctttccagcgttcggcacgcagacgacgggcttcattctgcatggttgtgcttacc agaccggagatattgacatcatatatgccttgagcaactgatagctgtcgctgtcaactgtcactgtaatacgctgcttcatagcatacctcttttttgaca tacttcgggtatacatatcagtatatattcttatacCgcaaaaatcagcgcgcaaatacgcatactgttatctggcttttagtaagccggatccacgcgg cgtttacgcccccccctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacaaacggcatgatgaacctgaatc gccagcggcatcagcaccttgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaa actggtgaaactcacccagggattggctgagacgaaaaacatattctcaataaacccctttagggaaataggccaggttttcaccgtaacacgccaca tcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaaca agggtgaacactatcccatatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaag gccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaa atgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttcctttagctcctgaaaatctcga taactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcc cagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgat gctgccaacttagtcgactacaggtcactaataccatctaagtagttgattcatagtgactggatatgttgtgttttacagtattatgtagtctgttttttat gcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcaactttctatacaaagttggcattataagaaagcattgcttatcaatttgt tgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcct ggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagta atacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcg cgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgcc aatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggt tactcaccactgcgatccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttc
MX01	*acctgccctaggtcgggcagcaacgcgtatatctggcccgtacatcgcgaagcagcgcaaaacggatcctgcaggtatttgcggccgcggtccgtat actccggaatattaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaa aacctataaatattccggattattcataccgtcccaccatcgggcgcgaactcctaaaaaaccgccaccatgaagtgccttttgtacttagccttttttatt cattggggtgaattgcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctc agatttaaattggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcat gcttccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgcaagg aaagcattgaacaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtga attgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgc cccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccatggacatcaccttcttctcagagga cggagagctatcatccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactg caagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctctttgctgcagccagattccctgaatgcccagaagggt caagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttgagaggatcttggattattccctctgccaagaaacctggagca aaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcctaaaaacccaggaaccggtcctgctttcaccataatcaatggtaccct aaaatactttgagaccagatacatcagagtcgatattgctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggga actgtgggatgactgggcaccatatgaagacgtggaaattggacccaatggagttctgaggaccagttcaggatataagtttcctttatacatgattg gacatggtatgttggactccgatcttcatcttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatga gagtttatttttggtgatactgggctatccaaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctcttttttctttatc atagggttaatcattggactattcttggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacata gagatgaaccgacttggaaagtgataagtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacc tcccacacctcccccgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcat cacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgag cctagttcggtggtaacggcgcagtggcggttttcatggcttgttatgactgttttttgtacagtctatgcctcgggcatccaagcagcaagcgcgttac gccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttagg tggctcaagtatgggcatcattcgcacatgtaggctcggccctgaccaagtcaaatccatgcgggctgctcttgatcttttcggtcgtgagttcggagac gtagccacctactcccaacatcagccggactccgattacctcgggaacttgctccgtagtaagacattcatcgcgcttgctgccttcgaccaagaagcg gttgttggcgctctcgcggcttacgttctgcccaagtttgagcagccgcgtagtgagatctatatctatgatctcgcagtctccggcgagcaccggaggc agggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcgcttggtgcttatgtgatctacgtgcaagcagattacggtgacgatccc gcagtggctctctatacaaagttgggcatacgggaagaagtgatgcactttgatatcgacccaagtaccgccacctaacaattcgttcaagccgaga tcggcttcccggccgcggagttgttcggtaaattgtcacaacgccgcgaatatagtctttacatgcccttggccacgcccctctttaatacgacgggcaa tttgcacttcagaaaatgaagagtttgctttagccataacaaaagtccagtatgcttttcacagcataactggactgatttcagtttacaactattctgt ctagtttaagactttattgtcatagtttagatctattttgttcagtttaagactttattgtccgcccacacccgcttacgcagggcatctgcgcatccacag gaagagcgacccaagtcaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgt tgctggcgttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagata ccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgct ttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgcct tatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgt aggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcgg aaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaagg atctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggat cttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttgacccaagtcaaccggttgt gggcggacaaaatagttgggaactgggaggggtggaaatggagtttttaaggattatttagggaagagtgacaaaatagatgggaactgggtgt agcgtcgtaagctaatacgaaaattaaaaatgacaaaatagtttggaactagatttcacttatctggttccaattagatgggtataccctagtatacg gtaccataacttcgtatagcatacattatacgaagttatctgccaggcacatgggttttactagtatcgattcgcgacctactccggaatattaatagat gtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcg ttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatag ggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgt caatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattacc atggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttg ttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatat aagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcga*

	attctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaacaagtttgtacaaaaaagctgaacgagaaacgtaaaatgata taaatatcaatatattaaattagattttgcataaaaaacagactacataatactgtaaaacacaacatatccagtcactatggcggccgcattaggca ccccaggctttacactttatgcttccggctcgtataatgtgtggattttgagttaggatccgtcgagattttcaggagctaaggaagctaaaatggaga aaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccaga ccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgct catccggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttt tcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttcccta aagggtttattgagaatatgttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgccccc gttttcacgatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcggc agaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaaacgcgtggatccggcttactaaaagccagataacagtatgc gtatttgcgcgctgatttttgcggtataagaatatatactgatatgtatacccgaagtatgtcaaaaagaggtatgctatgaagcagcgtattacagtg acagttgacagcgacagctatcagttgctcaaggcatatatgatgtcaatatctccggtctggtaagcacaaccatgcagaatgaagcccgtcgtctg cgtgccgaacgctggaaagcggaaaatcaggaagggatggctgaggtcgcccggtttattgaaatgaacggctcttttgctgacgagaacagggg ctggtgaaatgcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttgtggatgtacagagtgatattattgacacgcccgggcga cggatggtgatccccctggccagtgcacgtctgctgtcagataaagtctcccgtgaactttacccggtggtgcatatcggggatgaaagctggcgcat gatgaccaccgatatggccagtgtgccggtctccgttatcggggaagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaa cctgatgttctggggaatataaatgtcaggctcccttatacacagccagtctgcaggtcgaccatagtgactggatatgttgtgttttacagtattatgt agtctgttttttatgcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcagctttcttgtacaaagtggtgcaacaaacttctcact actcaaacaagcaggtgacgtggaggagaatcccgggccttctagaggatccggctccggcgagggcaggggaagtctactaacatgcggggac gtggaggaaaatcccggcccagtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggcca caagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctgg cccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaa ggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgc atcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggc cgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaac acccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcac atggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtagtaatgatagtaatgatagtaatgagggccctt cgaacaaaaactcatctcagaagaggatctgaatatgcataccggtcatcatcaccatcaccattgagaattcaacgcgttaagtcgacaatcaacc tctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctatt gcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgca ctgtgtttgctgacgcaacccccactggttggggcattgccacccacctgtcagctcctttccgggactttcgctttcccctccctattgccacggcggaac tcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctg ctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggcctcaatccagcggaccttccttcccgcggcctgctgccggct ctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctcccttgggccgcctccccgcggcctggccggccgccagcacagtggtcgat cgaccaatgccctggctcacaaataccactgagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggcta ataaaggaaatttatttcattgcaatagtgtgttggaatttttgtgtctctcactcggaaggacatatgggagggcaaatcatttaaaacatcagaa tgagtatttggtttagagtttggcaacatatgcccatatgctggctgccatgaacaaaggttggctataaagaggtcatcagtatatgaaacagcccc ctgctgtccattccttattccatagaaaagccttgacttgaggttagatttttttatatttgttttgtgttatttttttctttaacatcccctaaaatttccttac atgttttactagccagatttttcctcctctcctgactactcccagtcatagctgtccctcttctcttatggagatccctcg
UCAP	*gtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagt tattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgccca acgaccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaa ctgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccTattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtaca tgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatag cggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac aactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttact ggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggt atctattagaaaatcacaagtttgtacaaaaaagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaaac agactacatataatactgtaaaacacaacatatccagtcactatggcggccgcattaggcaccccaggctttacactttatgcttccggctcgtataatgt gtggattttgagttaggatccgtcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaa tggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaa agaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagc tggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccggca gtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatc cctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaa ggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtg gcagggcggggcgtaaacgcgtggatccggcttactaaaagccagataacagtatgcgtatttgcgcgctgatttttgcggtataagaatatatact gatatgtatacccgaagtatgtcaaaaagaggtatgctatgaagcagcgtattacagtgacagttgacagcgacagctatcagttgctcaaggcata tatgatgtcaatatctccggtctggtaagcacaaccatgcagaatgaagcccgtcgtctgcgtgccgaacgctggaaagcggaaaatcaggaaggg atggctgaggtcgcccggtttattgaaatgaacggctcttttgctgacgagaacaggggctggtgaaatgcagtttaaggtttacacctataaaaga gagagccgttatcgtctgtttgtggatgtacagagtgatattattgacacgcccgggcgacggatggtgatcccctggccagtgcacgtctgctgtca gataaagtctcccgtgaactttacccggtggtgcatatcggggatgaaagctggcgcatgatgaccaccgatatggccagtgtgccggtctccgttat cggggaagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaacctgatgttctggggaatataaatgtcaggctcccttata cacagccagtctgcaggtcgaccatagtgactggatatgttgtgttttacagtattatgtagtctgttttttatgcaaaatctaatttaatatattgatattt atatcattttacgtttctcgttcagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagctt aagtttaaaccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcc cactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggg aggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatccc ggtaatacggttatccacagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttt taaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcca tagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccg gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgc*

	cgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggctt cattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaa gtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtact caaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaa agtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactg atcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatg ttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaaca aataggggttccgcgcacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacg tgagttacgcgtcgttccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgc gccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttc ccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaa aaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactctt gttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattta acaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgc atctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagt cccgccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggc cgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggaggccgaggccgcctctgcctctgagctattccagaagtagt gaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagtacaagc ccacggtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgac ccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggc gccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttcccg gctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaggagcccgcgtggttcctggccaccgtcggcgtctcccgccgaccac cagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgc aacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgc ctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgc ggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaa gcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcat ggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtataatgtatgct atacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgatagtacgtactaagctctcatgtttcacgtactaagct ctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagctctcatggctaacgtactaagctctcatgtttcacg tactaagctctcatgtttgaacaatataaattaatatataaatcagcaacttaaatagcctctaaggttttaagttttataagaaaaaaagaatatataa ggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgagaagcccttagagcctctcaaagcaattttgagtga cacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaagggctgctccatgttggaatttaatcgcggcctcg acgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgc ataaaaaacagactacataatactgtaaaacacaacatatccagtcactatgaatcaactacttagatggtattcggccgcattaggcaccccccagg ctttacactttatgcttccggctcgtataatgtgtggattttgagttaggatccgtcgagattttcaggagctaaggaagctaaaatatcatcatgaaca ataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatg ctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttc tgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagca tttttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaa tattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaa tcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttt gccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgag tcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggcttttttcaaaaatatggtatt gataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgct gacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaagg atcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaa ctcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagc accgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccg gataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgag ctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagct tccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagccta tggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataacc gtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgaacctattcatatacgctagtgtcatagtcctgaaaatcatctgcatc aagaacaatttcacaactcttatacttttctcttacaagtcgttcggcttcatctggattttcagcctctatacttactaaacgtgataaagtttctgtaatttt ctactgtatcgacctgcagactggctgtgtataagggagcctgacatttatattccccagaacatcaggttaatggcgtttttgatgtcattttcgcggtg gctgagatcagccacttcttcccgataacggagaccggcacactggccatatcggtggtcatcatgcgccagctttcatcccgatatgcaccaccggg gtaaagttcacgggagactttatctgacagcagacgtgcactggccaggggatcaccatccgtcgcccgggcgtgtcaataatatcactctgtacat ccacaaacagacgataacggctctctcttttataggtgtaaaccttaaactgcatttcaccagccctgttctcgtcagcaaagagccgttcatttcaa taaaccgggcgacctcagccatcccttcctgattttccgctttccagcgttcggcacgcagacgacgggcttcattctgcatggttgtgcttaccagaccg gagatattgacatcatatatgccttgagcaactgatagctgtcgctgtcaactgtcactgtaatacgctgcttcatagcatacctcttcgatctaagtagt tgattcatagtgactggatatgttgtgttttacagtattatgtagtctgttttttatgcaaaatctaatttaatatattgatatttatatcattttacgtttctc gttcaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcct attccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaacta
UCZP	*ccgtagaaaagatcaaaggatcttcgcacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaa aatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctagggg gtatcccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctc ctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacc tcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaat agtggactcttgttccaaactggaacaacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaat gagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtat gcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtca*

gcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcag
aggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatat
ccattttcggatctgatcagcacgtgatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgcc
gccgcgttcgccgactaccccgccacgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgt
cgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcg
ccgagatcggcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagc
ccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcg
cgccggggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccga
aggaccgcgcacctggtgcatgacccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcg
gaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggt
tacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtc
tgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgcaggtcggtaggtcgcgaatcgatactagtaa
aacccatgtgcctggcagataacttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgt
tgatagtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtac
taagctctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctc
taaggttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgca
ctgagaagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcaggg
cgcaagggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgaacgagaaac
gtaaaatgatataaatatcaatatattaaattagattttgcataaaaaacagactacataatactgtaaaacacaacatatccagtcactatgaatca
actacttagatggtattcggccgcattaggcacccccaggctttacactttatgcttccggctcgtataatgtgtggattttgagttaggatccgtcgaga
ttttcaggagctaaggaagctaaaatatcatcatgaacaataaaactgtctgcttacataaacagtaacacaaggggtgttatgagccatattcaac
gggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaat
ctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagacta
aactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacag
cattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgt
ccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggct
ggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttg
acgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttt
tctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatca
gaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtc
gttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaccacc
gctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttcta
gtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgata
agtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttgg
agcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggt
aagcggcagggtcggaacaggagagcgcactagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt
gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttg
ctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgaacct
attcatatacgctagtgtcatagtcctgaaaatcatctgcatcaagaacaatttcacaactcttatacttttctcttacaagtcgttcggcttcatctggatt
ttcagcctctatacttactaaacgtgataaagtttctgtaatttctactgtatcgacctgcagactggctgtgtataagggagcctgacatttatattcccc
agaacatcaggttaatggcgtttttgatgtcattttcgcggtggctgagatcagccacttcttccccgataacggagaccggcacactggccatatcgg
tggtcatcatgcgccagctttcatccccgatatgcaccaccgggtaaagttcacgggagactttatctgacagcagacgtgcactggccaggggatc
accatccgtcgcccgggcgtgtcaataatatcactctgtacatccacaaacagacgataacggctctctcttttataggtgtaaaccttaaactgcattt
caccagcccctgttctcgtcagcaaaagagccgttcatttcaataaaccgggcgacctcagccatcccttcctgattttccgctttccagcgttcggcacg
cagacgacgggcttcattctgcatggttgtgcttaccagaccggagatattgacatcatatatgccttgagcaactgatagctgtcgctgtcaactgtca
ctgtaatacgctgcttcatagcatacctcttcgatctaagtagttgattcatagtgactggatatgttgtgttttacagtattatgtagtctgttttttatgca
aaatctaatttaatatattgatatttatatcattttacgtttctcgttcaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaact
gcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctag
aactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccccggtcta
gagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcat
tagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataa
tgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgt
atcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggca
gtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagt
ctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggc
ggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctac
cggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtaca
aaaaagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaacagactacataatactgtaaaacacaac
atatccagtcactatggcggccgcattaggcaccccaggctttacacttttatgcttccggctcgtataatgtgtggattttgagttaggatccgtcgagat
tttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatt
tcagtcagttgatcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggc
ctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgtt
acaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggc
gtgttacggtgaaaacctggcctatttcccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaa
cgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttca
tcatgccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaaacgcgtggatccg
gcttactaaaagccagataacagtatgcgtatttgcgcgctgatttttgcggtataagaatatatactgatatgtatacccgaagtatgtcaaaaaga
ggtatgctatgaagcagcgtattacagtgacagttgacagcgacagctatcagttgctcaaggcatatatgatgtcaatatctccggtctggtaagca
caaccatgcagaatgaagcccgtcgtctgcgtgccgaacgctggaaagcggaaaatcaggaagggatggctgaggtcgcccggtttattgaaatg
aacggctcttttgctgacgagaacaggggctggtgaaatgcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttgtggatgtac
agagtgatattattgacacgcccgggcgacggatggtgatccccctggccagtgcacgtctgctgtcagataaagtctcccgtgaactttacccggtg
gtgcatatcggggatgaaagctggcgcatgatgaccaccgatatggccagtgtgccggtctccgttatcggggaagaagtggctgatctcagccacc
gcgaaaatgacatcaaaaacgccattaacctgatgttctggggaatataaatgtcaggctcccttatacacagccagtctgcaggtcgaccatagtg
actggatatgttgtgttttacagtattatgtagtctgtttttttatgcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcagctttctt

	gtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctgatcagcctcgactg tgccttctagttgccagccatctgttgtttgccccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaa ttgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatg ctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatcccggtaatacggttatccacagaatcaggg gatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaa acttggtctgacagtgcagctctggcccgtgtctcaaaatgcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatc atcatgatcagtcctgctcctcggccacgaagtgcacgcagttgccggccgggtcgcgcagggcgaactcccgcccccacggctgctcgccgatctcg gtcatggccggcccggaggcgtcccggaagttcgtggacacgacctccgaccactcggcgtacagctcgtccaggccgcgcacccacacccaggcca gggtgttgtccggcaccacctggtcctggaccgcgctgatgaacagggtcacgtcgtcccggaccacaccggcgaagtcgtcctccacgaagtcccg ggagaacccgagccggtcggtccagaactcgaccgctccggcgacgtcgcgcgcggtgagcaccggaacggcactggtcaacttggccatggttta gttcctcaccttgtcgtattatactatgccgatatactatgccgatgattaattgtcaacacgtgctgatcatgaccaaaatcccttaacgtgagttacgc gtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttc
UTX0	*ataacttcgtatagcatacattatacgaagttatctgccaggcacatgggttttactagtatcgattcgcgacctactccggaatattaatagatgtacg ggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttaca taacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggac tttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaat gacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggt gatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttgg caccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagc agagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattct gcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaacaagtttgtacaaaaaagcaggcttaccaccggtggctgccgccacc atggtgagcaagggcgaggaggtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgaacggccacgagttcgagatcgaggg cgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggacatcctgtcccct cagtgcatgtacggctccaagggctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttcccgagggcttcaagtgggagcgcgt gatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaa cttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcga gatgaagatgaggctgaagctgaaggacggcggccactacgacgccgaggtcaagaccacctacatggccaagaagcccgtgcagctgcccggc gcctacaagaccgacatcaagctggacatcacctcccacaacgaggactacaccatcgtggaacagtacgagcgcgccgagggccgccacctccacc ggcgcccccagctttcttgtacaaagtggtgcaacaaacttctcactactcaaacaagcaggtgacgtggaggagaatcccgggccttctagaggat ccggctccggcgagggcagggagtctactaacatgcggggacgtggaggaaaatcccggcccagtgagcaaggcgaggagctgttcaccgg ggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagct gacccttaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccct gaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaag acccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggca caagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaacttcaagattaggcacaaca tcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagca cccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacga gctgtacaagggttcaactagtgggtctggtcaccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctc cgaggaggacctgggttcaactagtggttctggtgactacaaggacgacgacgacaagctttagtaatgatagtaatgatagtaatgagggcccttc gaacaaaaactcatctcagaagaggatctgaatatgcataccggtcatcatcaccatcaccattgagaattcaacgcgttaagtcgacaatcaacct ctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattg cttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcact gtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttcccctccctattgccacggcggaactc atcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgct cgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttcctttcccgcggcctgctgccggctct gcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctcccttgggccgcctccccgcggcctggccggccgccagcacagtggtcgatc gaccaatgccctggctcacaaataccactgagatcttttcccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaa taaaggaaatttatttttcattgcaatagtgtgttggaatttttgtgtctctcactcggaaggacatatgggagggcaaatcatttaaaacatcagaat gagtatttggtttagagtttggcaacatatgcccatatgctggctgccatgaacaaaggttggctataaagaggtcatcagtatatgaaacagccccc tgctgtccattcctttattccatagaaaagccttgacttgaggttagatttttttatattttgttttgtgttatttttttctttaacatccctaaaatttttcctac atgttttactagccagatttttcctcctctcctgactactcccagtcatagctgtccctcttctcttatggagatccctcgacctgccctaggtcgggcagca acgcgtatatctggcccgtacatcgcgaagcagcgcaaaacggatcctgcaggtatttgcggccgcggtccgtatactccggaatattaatagatcat ggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaatagaaaaaacctataaatattccggattatt cataccgtcccaccatcgggcgcgaactcctaaaaaaccgccaccatgaagtgccttttgtacttagccttttattcattggggtgaattgcaagttca ccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgactt aataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgt gatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaac aaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtccaggtgactcctcaccat gtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaac ctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctggga aaggagggcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagact cccatcaggtgtctggttcgagatggctgataaggatctctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcag acctcagtggatgtaagtctaattcaggacgttgagaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaat ctctccagtggatctcagctatcttgctcctaaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatac atcagagtcgatattgctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccat atgaagacgtggaaattggacccaatggagttctgaggaccagttcaggatataagtttccttatacatgattggacatggtatgttggactccgat cttcatcttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagttttattttttggtgatactgg gctatccaaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctcttttttctttatcataggggttaatcattggactatt cttggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttggaaagt gataagtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccgaacctga aacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcat tttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagcctagttcggtggtaacggcgca*

	gtggcggttttcatggcttgttatgactgttttttttgtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaggtggctcaagtatgggcatcattcgcacatgtaggctcggccctgaccaagtcaaatccatgcgggctgctcttgatcttttcggtcgtgagttcggagacgtagccacctactcccaacatcagccggactccgattacctcggggaacttgctccgtagtaagacattcatcgcgcttgctgccttcgaccaagaagcggttgttggcgctctcgcggcttacgttctgcccaagtttgagcagccgcgtagtgagatctatatctatgatctcgcagtctccggcgagcaccggaggcagggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcgcttggtgcttatgtgatctacgtgcaagcagattacggtgacgatcccgcagtggctctctatacaaagttgggcatacgggaagaagtgatgcactttgatatcgacccaagtaccgccacctaacaattcgttcaagccgagatcggcttcccggccgcggagttgttcggtaaattgtcacaacgccgcgaatatagtctttacatgcccttggccacgcccctctttaatacgacgggcaatttgcacttcagaaaatgaagagtttgctttagccataacaaaagtccagtatgcttttcacagcataactggactgatttcagtttacaactattctgtctagtttaagactttattgtcatagtttagatctattttgttcagtttaagactttattgtccgcccacacccgcttacgcagggcatctgcgcatccacaggaagagcgacccaagtcaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttgacccaagtcaaccggttgtgggcggacaaaatagttgggaactgggaggggtggaaatggagttttaaggattatttagggaagagtgacaaaatagatgggaactgggtgtagcgtcgtaagctaatacgaaaattaaaaatgacaaaatagtttggaactagatttcacttatctggttccaattagatgggtataccctagtatacggtacc
GIDC	*atcatggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccaccatcgggcgcggcccagtcacgacgttgtaaaacgacggccagtgcgccagggttttcccagtcacgacgacaagtttgtacaaaaaagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaaacagactacataatactgtaaaacacaacatatccagtcactatggtcgaccgacagccttccaaatgttcttcgggtgatgctgccaacttagtcgaccgacagccttccaaatgttcttctcaaacggaatcgtcgtatccagcctactcgctattgtcctcaatgccgtattaaatcataaaaagaaataagaaaaagaggtgcgagcctcttttttgtgtgacaaaataaaaacatctacctattcatatacgctagtgtcatagtcctgaaaatcatctgcatcaagaacaatttcacaactcttatacttttctcttacaagtcgttcggcttcatctggattttcagcctctatacttactaaacgtgataaagtttctgtaatttctactgtatcgacctgcagactggctgtgtataagggagcctgacatttatattccccagaacatcaggttaatggcgtttttgatgtcattttcgcggtggctgagatcagccacttcttccccgataacggagaccggcacactggccatatcggtggtcatcatgcgccagctttcatccccgatatgcaccaccgggtaaagttcacgggagactttatctgacagcagacgtgcactggccaggggatcaccatccgtcgcccgggcgtgtcaataatatcactctgtacatccacaaacagacgataacggctctctcttttataggtgtaaaccttaaactgcatttcaccagcccctgttctcgtcagcaaaagagccgttcatttcaataaaccgggcgacctcagccatcccttcctgatttttccgctttccagcgttcggcacgcagacgacgggcttcattctgcatggttgtgcttaccagaccggagatattgacatcatatatgccttgagcaactgatagctgtcgctgtcaactgtcactgtaatacgctgcttcatagcatacctctttttgacatacttcgggtatacatatcagtatatattcttatacccgcaaaaatcagcgcgcaaatacgcatactgttatctggcttttagtaagccggatccacgcggcgtacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttgacccaagtcaaccggttgtgggcggacaaaatagttgggaactgggaggggtggaaatggagttttaaggattatttagggaagagtgacaaaatagatgggaactgggtgtagcgtcgtaagctaatacgaaaattaaaaatgacaaaatagtttggaactagatttcacttatctggttccaattagatgggtataccctagtatacggtacc*

	cagtgcgttacatccctggcttgttgtccacaaccgttaaaccttaaaagctttaaaagccttatatattcttttttttcttataaaacttaaaaccttagag gctatttaagttgctgatttatattaattttattgttcaaacatgagagcttagtacgtgaaacatgagagcttagtacgttagccatgagagcttagta cgttagccatgagggtttagttcgttaaacatgagagcttagtacgttaaacatgagagcttagtacgtgaaacatgagagcttagtacgtactatca acaggttgaactgctgatcaacagatcctctacgcggccgcggtaccataacttcgtatagcatacattatacgaagttatctgccaggcacatgggtt ttactagtatcgattcgcgacctactccggaatattaatag
GIDK	*gacctttaattcaacccaacacaatatattatagttaaataagaattattatcaaatcatttgtatattaattaaaatactatactgtaaattacattttat ttacaatcactcgacgaagacttgatcacccgggatctcgagacccccagtcacgacgttgtaaaacgacggccagtgcgccagggttttcccagtca cgacgacaagtttgtacaaaaaagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaaacagactacat aatactgtaaaacacaacatatccagtcactatggtcgaccgacagccttccaaatgttcttcgggtgatgctgccaacttagtcgaccgacagccttc caaatgttcttctcaaacggaatcgtcgtatccagcctactcgctattgtcctcaatgccgtattaaatcataaaaagaaataagaaaaagaggtgcg agcctcttttttgtgtgacaaaataaaaaaatctacctattcatatacgctagtgtcatagtcctgaaaatcatctgcatcaagaacaatttcacaactc ttatactttctcttacaagtcgttcggcttcatctggattttcagcctctatacttactaaacgtgataaagtttctgtaatttctactgtatcgacctgcag actggctgtgtataagggagcctgacatttatattccccagaacatcaggttaatggcgtttttgatgtcattttcgcggtggctgagatcagccacttct tccccgataacggagaccggcacactggccatatcggtggtcatcatgcgccagctttcatccccgatatgcaccaccgggtaaagttcacgggagac tttatctgacagcagacgtgcactggccaggggggatcaccatccgtcgcccgggcgtgtcaataatatcactctgtacatccacaaacagacgataac ggctctctctttttataggtgtaaaccttaaactgcatttcaccagcccctgttctcgtcagcaaaagagccgttcatttcaataaaccgggcgacctcagc catcccttcctgattttccgctttccagcgttcggcacgcagacgacgggcttcattctgcatggttgtgcttaccagaccggagatattgacatcatata tgccttgagcaactgatagctgtcgctgtcaactgtcactgtaatacgctgcttcatagcatacctctttttgacatacttcgggtatacatatcagtatat attcttatatccgcaaaaatcagcgcgcaaatacgcatactgttatctggcttttagtaagccggatccacgcggcgtacgcaggaaagaacatgtga gcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcga cgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgcc gcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaa gctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcc actggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactag aagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcg gtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacg aaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagta tatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccg tcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagca ataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaa gtagttcgccagttaatagtctgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttccc aacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgt tatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgag aatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaa acgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactt tcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttc ctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgca catttccccgaaaagtgccacctgacgtccgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgat cttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttagtcgacggatatgttgtgttttacagtattatgtagtctgttttttt atgcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcagctttcttgtacaaagtggtcccctatagtgagtcgtattacatggtc atagctgtttcctgtgtgaaattgttatccgccgggagatgggggaggctaactgaaacacggaaggagacaataccggaaggaacccgcgctatg acggcaataaaaagacagaataaaacgcacgggtgttgggtcgtttgttcataaacgcggggttcggtcccagggctggcactctgtcgataccccca ccgagacccattgggaccaatacgcccgcgtttcttccttttccccaccccaacccccaagttcgggtgaaggcccagggctcgcagccaacgtcggg gcggcaagccctgccatagccactacgggtacgtttaaacccatgtgcctggcagataacttcgtataatgtatgctatacgaagtgatggtacgtact aagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagctctcatggc taacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaaggttttaagttt tataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgagaagccctta gagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacatgggaattagcttcacgctgccgcaagcactcagggcgcaagggct gctaaaggaagcggaacacgtagaaagccagtccgcagaaacggtgctgaccccggatgaatgtcagctactgggctatctggacaagggaaaa cgcaagcgcaaagagaaagcaggtagcttgcagtgggcttacatggcgatagctagactgggcggttttatggacagcaagcgaaccggaattgc cagctggggcgccctctggtaaggttgggaagccctgcaaagtaaactggatggctttcttgccgccaaggatctgatggcgcaggggatcaagatc tgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctat gactgggcacaactgacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggt gccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgg gaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatg cggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgt cgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgt cgtgacacatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgcta tcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcg cagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgag atttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctgga gttcttcgcccaccccgggatctatgtcgggtgcggagaaagaggtaatgaaatggcacctaggtatcgatactagtatacg*
GIDS	*gacctttaattcaacccaacacaatatattatagttaaataagaattattatcaaatcatttgtatattaattaaaatactatactgtaaattacattttat ttacaatcactcgacgaagacttgatcacccgggatctcgagacccccagtcacgacgttgtaaaacgacggccagtgcgccagggttttcccagtca cgacgacaagtttgtacaaaaaagctgaacgagaaacgtaaaatgatataaatatcaatatattaaattagattttgcataaaaaaacagactacat aatactgtaaaacacaacatatccagtcactatggtcgaccgacagccttccaaatgttcttcgggtgatgctgccaacttagtcgaccgacagccttc caaatgttcttctcaaacggaatcgtcgtatccagcctactcgctattgtcctcaatgccgtattaaatcataaaaagaaataagaaaaagaggtgcg agcctcttttttgtgtgacaaaataaaaacatctacctattcatatacgctagtgtcatagtcctgaaaatcatctgcatcaagaacaatttcacaactct tatactttctcttacaagtcgttcggcttcatctggattttcagcctctatacttactaaacgtgataaagtttctgtaatttctactgtatcgacctgcaga ctggctgtgtataagggagcctgacatttatattccccagaacatcaggttaatggcgtttttgatgtcattttcgcggtggctgagatcagccacttctt*

	ccccgataacggagaccggcacactggccatatcggtggtcatcatgcgccagctttcatccccgatatgcaccaccgggtaaagttcacgggagact ttatctgacagcagacgtgcactggccaggggatcaccatccgtcgcccgggcgtgtcaataatatcactctgtacatccacaaacagacgataac ggctctctctttttataggtgtaaaccttaaactgcatttcaccagcccctgttctcgtcagcaaaagagccgttcatttcaatataaccgggcgacctcagc catcccttcctgattttccgctttccagcgttcggcacgcagacgacgggcttcattctgcatggttgtgcttaccagaccggagatattgacatcatata tgccttgagcaactgatagctgtcgctgtcaactgtcactgtaatacgctgcttcatagcatacctctttttgacatacttcgggtatacatatcagtatat attcttataccgcaaaaatcagcgcgcaaatacgcatactgttatctggcttttagtaagccggatccacgcggcgtacgcaggaaagaacatgtga gcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataagctccgccccccctgacgagcatcacaaaaatcga cgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgcc gcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaa gctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcca ctggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactaga agaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcgg tggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacga aaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtat atatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgt cgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaa taaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagt agttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaa cgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgtta tcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgaga atagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaa cgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttc accagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcct ttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcaca tttccccgaaaagtgccacctgacgtccgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatct tccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttagtcgacggatatgttgtgttttacagtattatgtagtctgttttttat gcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcagctttcttgtacaaagtggtcccctatagtgagtcgtattacatggtcat agctgtttcctgtgtgaaattgttatccgcgggagatgggggaggctaactgaaacacggaaggagacaatacccggaaggaacccgcgctatga cggcaataaaaagacagaataaaacgcacgggtgttgggtcgtttgttcataaacgcggggttcggtcccagggctggcactctgtcgatacccccac cgagaccccattgggaccaatacgcccgcgtttcttccttttccccacccccaaccccccaagttcgggtgaaggcccagggctcgcagccaacgtcggg gcggcaagccctgccatagccactacgggtacgtttaaacccatgtgcctggcagataacttcgtataatgtatgctatacgaagttatggtacgtact aagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagctctcatggc taacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaaggttttaagttt tataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgagaagcccttа gagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacatgggaattagcttcacgctgccgcaagcactcagggcgcaagggct gctaaaggaagcggaacacgtagaaagccagtccgcagaaacggtgctgaccccggatgaatgtcagctgggaggcagaataaatgatcatatc gtcaattattacctccacggggagagcctgagcaaactggcctcaggcatttgagaagcacacggtcacactgcttccggtagtcaataaaccggta agtagcgtatgcgctcacgcaactggtccagaaccttgaccgaacgcagcggtggtaacggcgcagtggcggttttcatggcttgttatgactgttttt ttggggtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcag ggcagtcgcccttaaaacaaagttaaacatcatgagggaagcggtgatcgccgaagtatcgactcaactatcagaggtagttggcgtcatcgagcgc catctcgaaccgacgttgctggccgtacatttgtacggctccgcagtggatggcggcctgaagccacacagtgatattgatttgctggttacggtgacc gtaaggcttgatgaaacaacgcggcgagctttgatcaacgacctttggaaacttcggcttcccctggagagagcgagattctccgcgctgtagaagt caccattgttgtgcacgacgacatcattccgtggcgttatccagctaagcgcgaactgcaatttggagaatggcagcgcaatgacattcttgcaggtat cttcgagccagccacgatcgacattgatctggctatcttgctgacaaaagcaagagaacatagcgttgccttggtaggtccagcggcggaggaactc tttgatccggttcctgaacaggatctatttgaggcgctaaatgaaaccttaacgctatggaactcgccgcccgactgggctggcgatgagcgaaatgt agtgcttacgttgtcccgcatttggtacagcgcagtaaccggcaaaatcgcgccgaaggatgtcgctgccgactgggcaatggagcgcctgccggcc cagtatcagcccgtcatacttgaagctagacaggcttatcttggacaagaagaagatcgcttggcctcgcgcgcagatcagttggaagaatttgtcca ctacgtgaaaggcgagatcaccaaggtagtcggcaaataatgtctaacaattcgttcaagccgacggatctatgtcgggtgcggagaaagaggta atgaaatggcacctaggtatcgatactagtatacg

D. List of Generated Entry Clones

PLASMID NAME & Detail	PLASMID SEQUENCE
GD21.1 ANAP-RS	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggcccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttttataatgccaactttgtacaaaaaagcaggcttaggctagcgtttaaacttaagct tgccaccatggaagagcaataccgcccggaagagatagaatccaaagtacagcttcattgggatgagaagcgcacatttgaagtaaccgaagacg agagcaaagagaagtattactgcttttctggccctccctatccttctggtcgactacacatgggccacgtacgtaactacaccatcggtgacgtgatcg cccgctaccagcgtatgctgggcaaaaacgtcctgcagccgatcggctgggacgcgtttggtctgcctgcggaaggcgcggcggtgaaaaacaaca ccgctccggcaccgtggacgtacgacaacatcgcgtatatgaaaaaccagctcaaaatgctgggctttggttatgactggagccgcgagctggcaac ctgtacgccggaatactaccgttgggaacagaaattcttcaccgagctgtataaaaaaggcctggtatataagaagacttctgcggtcaactggtgtc cgaacgaccagaccgtactggcgaacgaacaagttatcgacggctgctgctggcgctgcgataccaaagttgaacgtaaagagatcccgcagtggt ttatcaaaatcactgcttacgctgacgagctgctcaacgatctggataaactggatcactggcctgacaccgttaaaaccatgcagcgtaactggatc ggtcgttccgaaggcgtggagatcaccttcaacgttaacgactatgacaacacgctgaccgtttacactacccgcccggacacctttatgggttgtacc tacctggcggtagctgcgggtcatccgctggcgcagaaagcggcggaaaataatcctgaactggcggcctttattgacgaatgccgtaacaccaaa gttgccgaagctgaaatggcgacgatggagaaaaaaggcgtcgatactggctttaaagcggttcacccattaacgggcgaagaaattcccgtttgg gcagcaaacttcgtattgatggagtacggcacgggcgcagttatggcggtaccggggcacgaccagcgcgactacgagtttgcctctaaatacggcc tgaacatcaaaccggttatcctggcagctgacggctctgagccagatctttctcagcaagccctgactgaaaaaggcgtgctgttcaactctggcgag ttcaacggtcttgaccatgaagcggccttcaacgccatcgccgataaactgactgcgatgggcgttggcgagcgtaaagtgaactaccgcctgcgcg actggggtgtttcccgtcagcgttactggggcgcgccgattccgatggtgactctagaagacggtaccgtaatgccgaccccggacgaccagctgccg gtgatcctgccggaggatgtggtaatggacggcattaccagcccgatttaaagcagatccggagtgggcgaaaactaccgttaacggtatgccagca ctgcgtgaaaccgacactttcgacacctttatggagtcctcctgggttcttgcgcgctacacttgcccgcagtacaaagaaggtatgctggattccgaa gcggctaactactggctgccggtggatatcgcgattggtggtattgaacacgccattatggagagtctctactgtcgcttcttccacaaactgatgcgtg atgcaggcatggtgaactctgacgaaccagttaaacagttgctgtgtcagggtatggtgctggcagatgccttctactatgttggcgaaaacggcga acgtaactgggtttccccggttgatgctatcgttgaacgtgacgagaaaggccgtatcgtgaaagcgaaagatgcggcaggccatgaactggtttat accggcatgagcaaaatgtccaagtcgaagaacaacggtatcgacccgcaggtgatggttgaacgttacggcgcggacaccgttcgtctgtttatga tgtttgcttctccggctgatatgactctcgaatggcaggaatccggtgtggaaggggctaaccgcttcctgaaacgtgtctggaaactggtttacgagc acacagcaaaaagtgatgttgcggcactgaacgttgatgcgctgactgaaaatcagaaagcgctgcgtcgcgatgtgcataaaacgatcgctaaa gtgaccgatgatatcggccgtcgtcagaccttcaacaccgcaattgcggcgattatggagctgatgaacaaactggcgaaagcaccaaccgatggc gagcaggatcgcgctctgatgcaggaagcactgctggccgttgtccgtatgcttaacccgttcaccccgcacatctgcttcacgctgtggcaggaactg aaaggcgaaggcgatatcgacaacgcgccgtggccggttgctgacgaaaaagcgatggtggaagactccacgctggtcgtggtgcaggttaacgg taaagtccgtgccaaaatcaccgttccggtggacgcaacggaagaacaggttcgcgaacgtgctggccaggaacatctggtagcaaaatatcttga tggcgttactgtacgtaaagtgatttacgtaccaggtaaactcctcaatctggtcgttggcgggcccgtttaatgatagaacccagctttcttgtacaaa gttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgatatcccct atagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatc atgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaac atggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccag agttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgacc atcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattagaagaatatcctgattca ggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctc aggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgc ataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatg ttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaa atatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcaga gcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaa gatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaa gagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaag aactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagac gatagttaccggataaggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatac ctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgc acgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcagggg ggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt*
GD21.2 DanAla-RS	*atgcaagagcagtaccgccctgaagagattgagtccaaagtgcagttacactgggatgagaagcgaacattcgaggtgaccgaagacgagtcca aggagaagtactactgcttaagcgccaatccttacccgagcggtagactgcatatgggacatgtgcgaaattacaccatcggcgacgtgatcgcccg gtaccagagaatgctcggaaagaacgtgctccagccaatcggatgggacgcattcgggttacctgcagaaggggcggccgtgaaaaacaataca gctccagctccatggacctatgataacatcgcgtatatgaaaaatcagctgaagatgcttgggtttggctacgactggtcccgggaactggctacttgc acacccgaatactatagatgggagcagaagttttttacagagctctacaagaaagggctggtgtacaaaaagaccagcgccgtaaactggtgccca aatgaccagaccgtgttggcgaatgaacaggtgatcgatgggtgctgttggagatgtgacactaaagtagaaaggaaggagatcccacagtggtt cataaagatcactgcgtacgctgatgagctcctgaacgacctggacaagttagaccattggcctgacacagtaaagactatgcaaaggaattggat aggacgtagtgagggcgtggagattacatttaacgtaaatgattatgataacacccttacagtgtacactactcgaccggacgccttttatgggctgca cttatcttgcagtggccgccggacatccgttagcccagaaagctgcagaaaataaccctgagctggcagccttttatcgacgagtgcagaaacaccaa ggtcgctgaggccgagatggctacaatggaaaagaaaggagtagataccggctttaaggctgtccaccctctcacaggggaagaaattcctgtatg ggctgccaacttcgtcctgatggagtatggcactggcgcagtgatggcagtacccggacatgaccagcgagattacgagttcgcctccaaatacggg ctcaatatcaagccagtgattctggcggctgatggttctgaaccagacctcagtcagcaagctctgactgaaaagggtgttctcttcaactctggggaa tttaatggtcttgaccacgaagcggctttcaacgcaatcgcagacaagttgaccgccatgggggttggggagagaaaggtgaactaccggctgagg*

	gactggggagttagccggcagcgctattggggcgccccaatccctatggtcaccttggaggatggaacagtgatgcccacaccagatgaccagctcc ctgtgattcttcccgaagacgtggtgatggacggaattacttccccattaaggccgacccagagtgggcgaaaacaacagttaatggtatgccagc gctgcgggaaactgacaccttcgacactttcatggaaagctcttggatctacgcacgctacacgtgcccccaatacaaggagggaatgctggacagc gaggcggcaaactattggctccccgttgacattggaattggagggatcgaacacgccataatgacattgctgtatttagatttttccataaactgatg cgagatgcaggcatggtgaactctgacgagccagcaaaacagctgttatgccagggaatggtgctggccgacgctttctactacgttggcgaaaac ggcgagcgtaattgggtaagtcctgttgacgcaatagtagagcgcgacgagaagggaaggatagttaaggccaaagacgcggctgggcacgaat tggtgtacacgggcatgtcaaagatgagtaaatccaaaaacaatggtattgacccccaggtcatggtcgaacggtatggggcagatactgtacggc tctttatgatgttcgcctcccccgcagatatgacccttgagtggcaggaatccggggtggaggggcgaacagattcttaaagcgcgtctggaagctg gtgtacgagcacaccgctaaagctgatgtggccgcactcaacgtggatgccttgacagaaaaccagaaggccctgcggcgcgacgtgcacaagac tatcgcaaaggttacagatgacatcgggaggcggcaaacgttcaatacagccatcgcggcaattatggagttgatgaataagctggcaaaggctcc aacggatggagagcaggaccgagccctcatgcaggaggctctgctggccgtggtaagaatgttaaacccctcactcctcatatctgtttcacattgtg gcaggaactgaaggggggagggagacatcgataacgctccgtggccagtggccgacgagaaggctatggtcgaagatagcaccctcgtggtggtc caagtaaacgggaaggtgagagctaaaataacggtgcccgtcgacgccactgaggagcaagtacgagagagggccggccaggagcacctggtg gccaagtacctcgacggagtcacagtcagaaaggtcatctatgtgccgggcaaactgctgaacctcgttgtgggctaagacccagctttcttgtacaa agttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgaggctaggtggaggctc agtgatgataagtctgcgatggtggatgcatgtgtcatggtcatagctgtttcctgtgtgaaattgttatccgctcagagggcacaatcctattccgcgc tatccgacaatctccaagacattaggtggagttcagttcggcgtatggcatatgtcgctggaaagaacatgtgagcaaaaggccagcaaaaggcca ggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaa acccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgccttt ctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccc ccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaaca ggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgc tctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagca gattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcttattcaacaaagccgccgtcccgtcaagtcagc gtaaatgggtaggggcttcaaatcgtccgctctgccagtgttacaaccaattaacaaattctgattagaaaaactcatcgagcatcaaatgaaact gcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggca agatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttccccctcgtcaaaaataaggttatcaagtgagaaatcacca tgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactc gcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgc aaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtg gtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgt aacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgaca ttatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctgagcaagacgtttcccgttgaatatggctcataacac cccttgtattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacacaacgtggc tttccccgccgctctagaactagtggatccaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcgcattatacga gacgtccaggttgggatacctgaaacaaaacccatcgtacggccaaggaagtctccaataactgtgatccaccacaagcgccagggttttcccagtc acgacgttgtaaaacgacggccagtcatgcataatccgcacgcatctggaataaggaagtgccattccgcctgacctcaaataatgatttta ttttga ctgatagtgacctgttcgttgcaacagattgatgagcaatgctttttataatgccaactttgtacaaaaaagcaggctgcacc
GD21.3 EcTrp(h9)-RS	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgatttttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaactttgtacaaaaaagcaggcttaggatccatgactaagcccatc gtttttgctggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatt tactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtgg tatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaact gagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcc tgctgtatcaaactaatctgagtcctgctggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggc gagatctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgata atcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttc gctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcg aaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgat gaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggtttt gtggcgaagccgtaatgatagggatccaacccagctttcttgtacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggt cactatcagtcaaaataaaatcattatttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccg tgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgtt atgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggc aatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagat gagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcga tccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcga ttcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatg acgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcact tgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatgga actgcctcggtgagttttctccttcattacagaaacggcttttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcg atgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttа acgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgca aacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagata ccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggc tgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgc acacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaagg*

	cggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt
GD21.4 EcTrp(h14)-RS	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggccttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctactctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactgatagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtacaaaaaagcaggcttaggatccatgactaagcccatcgtttttgctggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgggtccttgtggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatccctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaatgatagggatccaaccccagctttcttgtacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggttattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggcttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt*
GD21.5 MbPyl-RS	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggccttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctactctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactgatagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtacaaaaaagcaggcttaatggataaaaaaccattagatgttttaatatctgcgaccgggctctggatgtccaggactggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatacattgaaatggcgtgtggagaccatcttgttgtgaataattccaggagttgtagaacagccagagcattcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcggacgaggatatcaataattttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagtttctgctccaaaggtcaaaaaagctatgccgaaatcagtttcaagggctccgaagcctctggaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgtaccttcgcctgcaaaatcaactccaaattcgtctgttcccgcatcggctcctgctccttcacttacaagaagccagcttgatagggttgaggctctcttaagtccagaggataaaatttctctaaatatggcaaagcctttcagggaacttgagcctgaacttgtgacaagaagaaaaaacgattttcagcggctctataccaatgatagagaagactacctcggtaaactcgaacgtgatattacgaaatttttcgtagaccggggttttctggagataaagtctcctatccttattccggcggaatacgtggagagaatgggtattaataatgatactgaacttcaaaacagatctttcgggtggataaaaatctctgcttgaggccaatgcttgcccccgactctttacaactatctgcgaaaactcgataggattttaccaggcccaataaaaattttcgaagtcggaccttgttaccggaaagagtctgacggcaaagagcacctggaagaatttactatggtgaacttctgtcagatgggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagtttctggactatctggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgatataatgcacggggacctggagctttcttcggcagtcgtcgggccagtttctcttgatagagaatggggtattgacaaaccatggataggtgcaggttttggtcttgaacgcttgctcaaggttatgcacggcttaaaaacattaagagggcatcaaggtccgaatcttactataatgggatttcaaccaatctgtaatagtgaaacccagctttcttgtacaaagttggcatataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggttattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggcttttcaaaaatatggtattga*

	taatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctga cttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatc ttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactc tttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcacc gcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggat aaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagct atgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttcc agggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatgg aaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt
GD21.6 MmPyl-RS	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctcccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaacttttgtacaaaaaagcaggcttagccgccaccatggataaaaa accactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacaccacgaagtctctcgaagcaaaatc tatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggactgcaagagcgctcaggcaccacaaatacaggaagac ctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcacaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgc ccctaccagaacgaaaaaggcaatgccaaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatc taaattttcacctgcgataccggtttccaccccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcat ccgcactggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactgacagg ctt gaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagccttcagggagcttgagtccgaattgctctctcgcagaaaaaaagacct gcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacaggggttttctggaaataaa atccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaagaacttctg cctgagacccatgcttgctccaaaccttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgcta cagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaaatcttgaaa gcataattacggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctttggggataccctgatgtaatgcacggag acctggaactttcctctgcagtagtcggacccatacgcttgaccgggaatggggtattgataaaccctggatagggcaggtttcgggctcgaacgc cttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtctttactataacgggatttctaccaacctgtaatgagaacccc agctttcttgtacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccat ccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaaga taaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccg cgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaa gcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaattt atgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattagaag aatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcg tatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtct ggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgagggggaaattaat aggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaa cggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttg taacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcag accccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggttt gtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttag gccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgg gttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctaca ccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgg aacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtg atgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt*
GD21.7 OMeY-RS	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctcccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggcccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaacttttgtacaaaaaagcaggcttaacccaagctggctagcgccac catggcaagcagtaacttgattaaacaattgcaagagcgggggctggtagcccaggtgacggacgaggaagcgttagcagagcgactggcgcaa ggcccgatcgcactcgtgtgtggcttcgatcctaccgctgacagcttgcatttggggcatcttgttccattgttatgcctgaaacgcttccagcaggcgg gccacaagccggttgcgctggtaggcggcgcgacgggtctgattggcgacccgagcttcaaagctgccgagcgtaagctgaacaccgaagaaact gttcaggagtgggtggacaaaatccgtaagcaggttgccccgttcctcgatttcgactgtggagaaaactctgctatcgcggccaataattatgactg gttcggcaatatgaatgtgctgaccttcctgcgcgatattggcaaacacttctccgttaaccagatgatcaacaaagaagcggttaagcagcgtctca accgtgaagatcaggggatttcgttcactgagttttcctacaacctgctgcagggttatagtatggcctgtttgaacaaacagtacggtgtggtgctgc aaattggtggttctgaccagtggggtaacatcacttctggtatcgacctgacccgtcgtctgcatcagaatcaggtgtttggcctgaccgttccgctgat cactaaagcagatggcaccaaatttggtaaaactgaaggcggcgcagtctggttggatccgaagaaaaccagcccgtacaaattctaccagttctg gatcaacactgcgcgtgccgacgtttaccgcttcctgaagttcttcaccttatgagcattgaagagatcaacgccctggaagaagaagataaaaac agcggtaaagcaccgcgcgcccagtatgtactggcggagcaggtgactcgtctggttcacggtgaagaaggtttacaggcggcaaaacgtattacc gaatgcctgttcagcggttctttgagtgcgctgagtgaagcggacttcgaacagctggcgcaggacggcgtaccgatggttgagatggaaaagggc gcagacctgatgcaggcactggtcgattctgaactgcaaccttcccgtggtcaggcacgtaaaactatcgcctccaatgccatcaccattaacggtga aaaacagtccgatcctgaatacttctttaaagaagaagatcgtctgtttggtcgttttaccttactgcgtcgcggtaaaaagaattactgtctgatttgct ggaaagggcccgtttaatagtgaaacccagctttcttgtacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcact*

	atcagtcaaaataaaatcattatttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtc tcaaaatctctgatgttacattgcacaagataaaaatatatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatga gccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatc aggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgag atggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccc cggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcc tgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacg agcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttga taaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactg cctcggtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatga gtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgt gagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaaca aaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaa atactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgct gccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcaca cagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcgg acaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttc gccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggcct tttgctggccttttgctcacatgtt
GD21.9 mRaspberry (without terminal stop codon)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtacaaaaaagcaggcttaccaccggtggctgccgccacc atggtgagcaagggcgaggaggtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgaacggccaccgagttcgagatcgaggg cgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggacatcctgtcccct cagtgcatgtacggctccaagggctacgtgaagcaccccgccgacatcccgactacttgaagctgtccttcccgagggcttcaagtgggagcgcgt gatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaa cttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcga gatgaagatgaggctgaagctgaaggacggcggccactacgacgccgaggtcaagaccacctacatggccaagaagcccgtgcagctgcccggc gcctacaagaccgacatcaagctggacatcacctcccacaacgaggactacaccatcgtggaacagtacgagcgcgccgagggccgccactccacc ggcgcccccagctttcttgtacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcat tatttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacatt gcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacg tcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgctt gtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctg acggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccagg tattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaaca gcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttg aacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggg gaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttca ttacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggtt aattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactg agcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagc ggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccg tagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtc ttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacg acctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggca gggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg attttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt t*
GD21.12 mRaspberry (with terminal stop codon)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtacaaaaaagcaggcttagccaccatggtgagcaaggg cgaggaggtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggc cgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggacatcctgtcccctcagtgcatgtacgg ctccaagggctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttcccgagggcttcaagtgggagcgcgtgatgaacttcgagg acggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacgg ccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatgaagatgagg ctgaagctgaaggacggcggccactacgacgccgaggtcaagaccacctacatggccaagaagcccgtgcagctgcccggcgcctacaagaccga catcaagctggacatcacctcccacaacgaggactacaccatcgtggaacagtacgagcgcgccgagggccgccactccaccggcgcctaatagtg aaacccagctttcttgtacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattat ttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgc acaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtc*

	gaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgt atgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctga cggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggt attagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacag cgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttga acaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgagggg aaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcat tacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggtt aattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactg agcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagc ggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccg tagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtc ttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacg acctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggca gggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg atttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt t
GD21.11 eGFP (with terminal Stop codons)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttttataatgccaactttgtacaaaaaagcaggcttagccaccatggtgagcaaggg cgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcg atgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgc agtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaag gacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggagg acggcaacatcctgggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaa cttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgccc gacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccggg atcactctcggcatggacgagctgtacaagtagtaatgaatagtgaaacccagctttcttgtacaaagttggcattataagaaagcattgcttatcaa tttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgatatcccctatagtgagtcgtattacatggtcatagctgtt tcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaac agtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggg ctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgtt gccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcat ggttactcaccactgcgatccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttc ctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggtt gatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtc actcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccag gatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaatt gcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagc tcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgc gcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggct tcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgct aatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggct gaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgctt cccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggta tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgc ggcctttttacggttcctggccttttgctggccttttgctcacatgtt*
GD54.1 EcLeu-tRNA (TAA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgccccct cgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacacccgcccggatggtggaatcggtagacacaagggattttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccggg tatttttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagt gatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattttaaatccctcggcgttcgcgct gtgcgggttcaagtcccgctccgggtatttttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtga tagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggg attttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtga tagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggg attttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtga tagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattttaaatccctcggcgttcgcgct gtgcgggttcaagtcccgctccgggtatttttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtga tagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggg attttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtga tagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattttaaatccctcggcgttcgcgct gtgcgggttcaagtcccgctccgggtatttttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagt*

	gatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaag ggatttttaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttttggatctaatatttgcatgcagccaatccctgggtgagtttc accacccaacttttctatacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcatta tttgccatccagctgcagggcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatct ctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattc aacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcga caatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcag actaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaa acagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgta attgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaa tggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttа tttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtg agttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttct aatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttac gcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaa ccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtt cttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtgg cgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccag cttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatcc ggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctga cttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttt gctcacatgtt
GD54.2 EcLeu-tRNA (TAG)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgatttttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatc agtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccggg tatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagt gatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgct gtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtga tagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggg attctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatggttggcaccatgggcggatattat atcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttcccct atcagtgatagagacacccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctcc gggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctat cagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggattctaaatccctcggcgttcg cgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatca gtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacaca agggattctaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgcagccaatccctgggtgagt ttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcat tatttgccatccagctgcagggcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaat ctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatat tcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcg acaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtca gactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaa aacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgt aattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgta atggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccctt atttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggt gagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttct aatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttac gcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaa ccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtt cttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtgg cgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccag cttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatcc ggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctga cttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggcctt tttgctcacatgtt*
GD54.3 EcLeu-tRNA (TGA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctcccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgatttttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgccccct*

	cgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatc agtgatagagacaccgcccggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgctcgg tatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagt gatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcgcgct gtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtga tagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaaggg atttcaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatggttggcaccatgggcggatattat atcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccct atcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctcc gggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctat cagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacacaagggatttcaaatccctcggcgttcg cgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatca gtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgcccggatggtggaatcggtagacaca agggatttcaaatccctcggcgttcgcgctgtgcgggttcaagtcccgctccgggtatttttggatctaatatttgcatgcagccaatccctgggtgagt ttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcat tatttgccatccagctgcagggcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaat ctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatat tcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcg acaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtca gactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaa aacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgt aattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgta atggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttt atttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggt gagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagttttct aatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttac gcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaa ccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtt cttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtgg cgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccag cttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatc cggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctg acttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggcct tttgctcacatgtt
GD54.4 EcTrp-tRNA (TAA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccacccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatc agtgatagagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggg atctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagag acttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctcc gcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttcc ctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttg ggagttcgagtctctccgcccctgccatttttggatctaatatttgcatggttggcaccatgggcggatattatatgcgctatgtgttctgggaaatcacc ataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcg tagttcaattggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaatatttgcatgtcgctatgt gttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgata gagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctccgcccctgccatttttggatctaat atttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagac ttata agttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctttaaaaccgggtgttgggagttcgagtctctccgcccctg ccatttttggatctaatatttgcatgcagccaatccctgggtgagtttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttatc aatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgcagggcggccgcgatatcccctatagtgagtcgtatt acatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaac tgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttta tatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaaca tggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccg tactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggttattagaagaatatcctgattcaggtgaaaatattgttga tgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaat gaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctc accggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatc gcagaccgataccaggatcttgccatcctatggaactgcctcggtgagtttctccttcattacagaaacggctttttcaaaaatatggtattgataatcc tgatatgaataaattgcagtttcatttgatgctcgatgagttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgac gggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttg agatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttc cgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcct acatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataag gcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatg agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccag*

	ggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaa aaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt
GD54.5 EcTrp-tRNA (TAG)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatc agtgatagagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctccgccccgccattttttgg atctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagag acttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctcc gcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttc cctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttg ggagttcgagtctctccgccctgccatttttggatctaatatttgcatggttggcaccatgggcggatattatatcgctatgtgttctgggaaatcacc ataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcg tagttcaattggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctccgccccgccattttttggatctaatatttgcatgtcgctatgt gttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgata gagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctccgccccgccattttttggatctaat atttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttata agttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtctctaaaaccgggtgttgggagttcgagtctctccgccccc gccatttttggatctaatatttgcatgcagccaatccctgggtgagtttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttat caatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgcagggcggccgcgatatccccctatagtgagtcgtat tacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaa ctgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgattt atatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaac atggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatcc gtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttg atgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaa tgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattct caccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaat cgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatc ctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttga cgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttctt gagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttt ccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcct acatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataag gcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatg agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccag ggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaa aaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt*
GD54.6 EcTrp-tRNA (TGA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatc agtgatagagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctccgccctgccatttttggg atctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagag acttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctcc gcccctgccatttttggatctaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttc cctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttg ggagttcgagtctctccgccccgccattttttggatctaatatttgcatggttggcaccatgggcggatattatatcgctatgtgttctgggaaatcacc ataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccaggggcg tagttcaattggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctccgccccgccattttttggatctaatatttgcatgtcgctatgt gttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgata gagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctccgccccgccattttttggatctaat atttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttata agttccctatcagtgatagagacaccaggggcgtagttcaattggtagagcaccggtcttcaaaaccgggtgttgggagttcgagtctctccgccccc gccatttttggatctaatatttgcatgcagccaatccctgggtgagtttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttat caatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgcagggcggccgcgatatccccctatagtgagtcgtat tacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaa ctgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgattt atatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaac atggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatcc gtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttg atgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaa*

	tgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattct caccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaat cgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatc ctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttga cgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttctt gagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttt ccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcct acatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataag gcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatg agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccag ggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaa aaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt
GD54.7 MbPyl-tRNA (TAA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttacc gtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttataagttccctatcagtgatagagac ttataagttccctatcagtgatagagacaccgaaacaccggaaacctgatcatgtagatcgaacggactttaaatccgttcagccgggttagattccc ggggtttccgttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatc agtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatcgaatagatcacacggactttaaatccgtgcaggcgggtgaa actcccgcacccccgttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttc cctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatcgaatagatcacacggactttaaatccgtgcaggcgg gtgaaactcccgcacccccgttttttgttggcaccatgggcggatattatatcgggcaggaagagggcctatttcccatgattccttcatatttgcatat acgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttt gggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaa ggacgtccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggaaa cctgatcatgtagatcgaacggactttaaatccgttcagccgggttagattcccggggtttccgttttttaatatttgcatgtcgctatgtgttctgggaaa tcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgg ggggtggatcgaatagatcacacggactttaaatccgtgcaggcgggtgaaactcccgcacccccgttttttaatatttgcatgtcgctatgtgttctg ggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagac accgggggggtggatcgaatagatcacacggactttaaatccgtgcaggcgggtgaaactcccgcacccccgttttttcagccaatccctgggtgagt ttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcat tatttgccatccagctgcagggcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaat ctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatat tcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcg acaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtca gactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaa aacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgt aattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgta atggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttt atttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggt gagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttct aatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttac gcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaa ccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtt cttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtgg cgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccag cttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatc cggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctg acttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggcct tttgctcacatgtt*
GD54.8 MbPyl-tRNA (TAG)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttacc gtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttataagttccctatcagtgatagagac ttataagttccctatcagtgatagagacaccgaaacaccggaaacctgatcatgtagatcgaacggactctaaatccgttcagccgggttagattccc ggggtttccgttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatc agtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatcgaatagatcacacggactctaaatccgtgcaggcgggtgaa actcccgcacccccgttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttc*

	cctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatcgaatagatcacacggactctaaatccgtgcaggcgg gtgaaactcccgcacccccgttttttgttggcaccatgggcggatattatatcgggcaggaagagggcctatttcccatgattccttcatatttgca tat acgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttt gggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaa ggacgtccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggaaa cctgatcatgtagatcgaacggactctaaatccgttcagccgggttagattcccggggtttccgtttttaatatttgcatgtcgctatgtgttctgggaa atcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccg gggggtggatcgaatagatcacacggactctaaatccgtgcaggcgggtgaaactcccgcacccccgttttttaatatttgcatgtcgctatgtgttct gggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagaga caccgggggggtggatcgaatagatcacacggactctaaatccgtgcaggcgggtgaaactcccgcacccccgttttttcagccaatccctgggtgag tttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatca ttatttgccatccagctgcagggcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaa tctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccata ttcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgc gacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtc agactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatcccggaa aaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttg taattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgt aatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataacc ttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcg gtgagtttttcttcattacagaaacggcttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagttttt ctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagtt acgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaa aaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatact gttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcc cagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacagg tatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacc tctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctg gccttttgctcacatgtt
GD54.9 MbPyl-tRNA (TGA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttacc gtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgtccctatcagtgatagagacttataagttccctatcagtgatagagac ttataagttccctatcagtgatagagacaccgaaacaccggaaacctgatcatgtagatcgaacggacttcaaatccgttcagccgggttagattccc gggggtttccgtttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatc agtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatcgaatagatcacacggacttcaaatccgtgcaggcgggtgaa actcccgcacccccgttttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttc cctatcagtgatagagacttataagttccctatcagtgatagagacaccgggggggtggatcgaatagatcacacggacttcaaatccgtgcaggcgg gtgaaactcccgcacccccgttttttgttggcaccatgggcggatattatatcgggcaggaagagggcctatttcccatgattccttcatatttgcatat acgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttt gggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaa ggacgtccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccgaaacaccggaaa cctgatcatgtagatcgaacggacttcaaatccgttcagccgggttagattcccggggtttccgtttttaatatttgcatgtcgctatgtgttctgggaa atcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccg gggggtggatcgaatagatcacacggacttcaaatccgtgcaggcgggtgaaactcccgcacccccgttttttaatatttgcatgtcgctatgtgttct gggaaatcaccataaacgtgaaatccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagaga caccgggggggtggatcgaatagatcacacggacttcaaatccgtgcaggcgggtgaaactcccgcacccccgttttttcagccaatccctgggtgag tttcaccacccaacttttctatacaaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatca ttatttgccatccagctgcagggcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaa tctctgatgttacattgcacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccata ttcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgc gacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtc agactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaa aaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttg taattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgt aatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataacc ttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcg gtgagtttttcttcattacagaaacggcttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagttttt ctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagtt acgcgtcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaa aaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatact gttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcc cagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacagg tatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacc*

	tctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctg gccttttgctcacatgtt
GD54.11 OMeY-tRNA (TAA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgatttttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttacc gtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttcccatcagtgatagaga gacttataagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagactttaaatctgccgtcacagact tcgaaggttcgaatccttcccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagaga cttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggacttta aatccgctcccttgggttcggcggttcgaatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatc cctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggct aaacgcggcggacttttaaatccgctcccttgggttcggcggttcgaatccgtcccccctccatttttgttggcaccatgggcggatatttatatcgggcag gaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagata ttagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttg aaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgatagagacttata agttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagactttaaatctgccgtcacagacttcgaaagg ttcgaatccttcccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttataa gttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggacttttaaatccgc tcccttgggttcggcggttcgaatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatca gtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcg gcggactttaaatccgctcccttgggttcggcggttcgaatccgtcccccctccatttttcagccaatccctgggtgagtttcaccacccaacttttctata caaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgcagg gcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaa gataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgaggg ccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatggg aagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaa tttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattaga agaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcg cgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagt ctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaa taggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaa acggcttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggtt gtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtca gaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtt tgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttag gccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgg gttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctaca ccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgg aacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgattttgtg atgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt*
GD54.12 OMeY-tRNA (TAG)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgatttttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgctttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttacc gtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgatagaga gacttataagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagactctaaatctgccgtcacagac ttcgaaggttcgaatccttcccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagag acttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggactct aaatccgctcccttgggttcggcggttcgaatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaat ccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggct aaacgcggcggactctaaatccgctcccttgggttcggcggttcgaatccgtcccccctccatttttgttggcaccatgggcggatatttatatcgggca ggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagat attagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaactt gaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgatagagacttat aagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagactctaaatctgccgtcacagacttcgaag gttcgaatccttcccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttata agttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggactctaaatccg ctcccttgggttcggcggttcgaatccgtcccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatca gtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcg*

	gcggactctaaatccgctccctttgggttcggcggttcgaatccgtccccctccatttttcagccaatccctgggtgagtttcaccacccaacttttctata caaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgcagg gcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaa gataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgagg ccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatggg aagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaa tttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattaga agaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcg cgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagt ctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaa taggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaa acggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagttttttctaatcagaattggttaattggtt gtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtca gaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtt tgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttag gccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgg gttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctaca ccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgg aacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtg atgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt
GD54.13 OMeY-tRNA (TGA)	*ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca gtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact ggaaagcgggcagtgagcgcaacgcaattaatacgcgtaccgctagccaggaagagtttgtagaaacgcaaaaaggccatccgtcaggatggcc ttctgcttagtttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcacaacgttcaaatccgctcccggcggatttgtc ctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtcttccgactgagcctttcgttttatttgatgcctggcagttccctac tctcgcgttaacgctagcatggatgttttcccagtcacgacgttgtaaaacgacggccagtcttaagctcgggccccaaataatgattttattttgactg atagtgacctgttcgttgcaacaaattgatgagcaatgcttttttataatgccaactttgtatacaaaagttgtgcaatatggacaacttcttcgcccccct cgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccg taacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgatagaga gacttataagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagacttcaaatctgccgtcacagac ttcgaaggttcgaatccttccccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagag acttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggacttc aaatccgctccctttgggttcggcggttcgaatccgtccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaat ccctatcagtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggct aaacgcggcggacttcaaatccgctccctttgggttcggcggttcgaatccgtccccctccatttttgttggcaccatgggcggatatttatatcgggca ggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattagaattaatttgactgtaaacacaaagat attagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaactt gaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaatccctatcagtgatagagacttataagttccctatcagtgatagagacttat aagttccctatcagtgatagagacaccgaaacaccggtggggttcccgagcggccaaagggagcagacttcaaatctgccgtcacagacttcgaag gttcgaatccttccccccaccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagacttata agttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcggcggacttcaaatccg ctcccctttgggttcggcggttcgaatccgtccccctccatttttaatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatca gtgatagagacttataagttccctatcagtgatagagacttataagttccctatcagtgatagagacaccggaggggtagcgaagtggctaaacgcg gcggacttcaaatccgctcccctttgggttcggcggttcgaatccgtccccctccatttttcagccaatccctgggtgagtttcaccacccaacttttctata caaagttggcattataagaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgccatccagctgcagg gcggccgcgatatcccctatagtgagtcgtattacatggtcatagctgtttcctggcagctctggcccgtgtctcaaaatctctgatgttacattgcacaa gataaaaatatatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcgagg ccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatggg aagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaa tttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattaga agaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcg cgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagt ctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaa taggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaa acggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggtt gtaacactggcagagcattacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtca gaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtt tgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttag gccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgg gttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctaca ccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgg aacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtg atgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgtt*

E. List of Generated Mammalian Expression Vectors

PLASMID NAME & Detail	PLASMID SEQUENCE
MX01.1 mRaspberry (without terminal stop codon)	*atgggtataccctagtatacggtaccataacttcgtatagcatacattatacgaagttatctgccaggcacatgggttttactagtatcgattcgcgacc tactccggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcata gcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatg ttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgcc aagtacgcccccttattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctac gtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccca ttgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgt gtacggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcag atctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaacaagtttgtacaaaaaagcaggc ttaccaccggtggctgccgccaccatggtgagcaagggcgaggaggtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgaac ggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgc ccttcgcctgggacatcctgtcccctcagtgcatgtacggctccaagggctacgtgaagcacccgccgacatccccgactacttgaagctgtccttccc cgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctcccctgcaggacggcgagttcatcta caaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccc cgaggacggcgccctgaagggcgagatgaagatgaggctgaagctgaaggacggcggccactacgacgccgaggtcaagaccacctacatggc caagaagcccgtgcagctgcccggcgcctacaagaccgacatcaagctggacatcacctcccacaacgaggactacaccatcgtggaacagtacg agcgcgccgagggccgccactccaccggcgcccccagctttcttgtacaaagtggtgcaacaaacttctcactactcaaacaagcaggtgacgtgga ggagaatcccgggccttctagaggatccggctccggcgagggcaggggaagtctactaacatgcggggacgtggaggaaaatcccggcccagtg agcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgaggg cgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacct acggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccat cttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttt caaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatc aaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgt gctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgacc gccgccgggatcactctcggcatggacgagctgtacaagtagtaatgatagtaatgatagtaatgagggcccttcgaacaaaaactcatctcagaa gaggatctgaatatgcataccggtcatcatcaccatcaccattgagaattcaacgcgttaagtcgacaatcaacctctggattacaaaatttgtgaaa gattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcatttttctc ctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccca ctggttggggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgct gctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggatt ctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgc cttcgccctcagacgagtcggatctcccttgggccgcctccccgcggcctggccggccgccagcacagtggtcgatcgaccaatgccctggctcacaa ataccactgagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatttatttttca ttg caatagtgtgttggaatttttgtgtctctcactcggaaggacatatgggagggcaaatcatttaaaacatcagaatgagtatttggtttagagtttggc aacatatgcccatatgctggctgccatgaacaaaggttggctataaagaggtcatcagtatatgaaacagccccctgctgtccattccttattccatag aaaagccttgacttgaggttagatttttttatatttgttttgtgttatttttttctttaacatcccctaaaatttccttacatgttttactagccagatttttcct cctctcctgactactcccagtcatagctgtcccctcttctcttatggagatccctcgacctgccctaggtcgggcagcaacgcgtatatctggcccgtacatc gcgaagcagcgcaaaacggatcctgcaggtatttgcggccgcggtccgtatactccggaatattaatagatcatggagataattaaaatgataacc atctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccaccatcgggcgc gaactcctaaaaaaccgccaccatgaagtgccttttgtacttagccttttttattcattggggtgaattgcaagttcaccatagtttttccacacaaccaaa aaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgacttaataggcacagccttacaagtca aaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgtgatttccgctggtatggaccgaag tatataacacattccatccgatccttcactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggct tccctcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacagga gaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaacctggcattctgactataaggtcaa agggctatgtgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaa gtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagat ggctgataaggatctctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaatt caggacgttgagaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatctt gctcctaaaaacccaggaaccggtcctgctttcaccataatcaatggtacccctaaaatactttgagaccagatacatcagagtcgatattgctgctcca atcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaattggacc caatggagttctgaggaccagttcaggatataagtttccttatacatgattggacatggtatgttggactccgatcttcatcttagctcaaaggctcag gtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttggtgatactgggctatccaaaaatccaatcgagct tgtagaaggttggttcagtagttggaaaagctctattgcctctttttctttatcataggttaatcattggactattcttggttctccgagttggtatccat ctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttggaaagtgataagtcgagaagtactagag gatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccgaacctgaaacataaaatgaatgcaattgtt gttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggt ttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagcctagttcggtggtaacggcgcagtggcggttttcatggcttgttatg actgttttttgtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgca gcagggcagtcgcccctaaaacaaagttaggtggctcaagtatgggcatcattcgcacatgtaggctcggccctgaccaagtcaaatccatgcgggct gctcttgatcttttcggtcgtgagttcggagacgtagccacctactcccaacatcagccggactccgattacctcgggaacttgctccgtagtaagacat tcatcgcgcttgctgccttcgaccaagaagcggttgttggcgctctcgcggcttacgttctgcccaagtttgagcagccgcgtagtgagatctatatctat gatctcgcagtctccggcgagcaccggaggcagggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcgcttggtgcttatgtgat ctacgtgcaagcagattacggtgacgatcccgcagtggctctctatacaaagttgggcatacgggaagaagtgatgcactttgatatcgacccaagt accgccacctaacaattcgttcaagccgagatcggcttcccggccgcggagttgttcggtaaattgtcacaacgccgcgaatatagtctttacatgccc ttggccacgcccctctttaatacgacgggcaatttgcacttcagaaaatgaagagtttgctttagccataacaaaagtccagtatgcttttcacagcat aactggactgatttcagtttacaactattctgtctagtttaagacttattgtcatagtttagatctattttgttcagtttaagactttattgtccgcccacac*

	ccgcttacgcagggcatctgcgcatccacaggaagagcgacccaagtcaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagca aaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagag gtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacct gtccgcctttctcccttcgggaagcgtggcgctttctatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgc acgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcca ctggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttg gtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttg caagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaa gggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaactt ggtctgacagttgacccaagtcaaccggttgtgggcggacaaaatagttgggaactgggaggggtggaaatggagtttttaaggattatttaggga agagtgacaaaatagatgggaactgggtgtagcgtcgtaagctaatacgaaaattaaaaatgacaaaatagtttggaactagatttcacttatctg gttccaattag
MX01.3 eGFP (with terminal stop codon)	*ataacttcgtatagcatacattatacgaagttatctgccaggcacatgggttttactagtatcgattcgcgacctactccggaatattaatagatgtacg ggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttaca taacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggac tttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaat gacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggt gatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttgg caccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagc agagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattct gcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaacaagtttgtacaaaaaagcaggcttagccaccatggtgagcaagggg cgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcg atgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgc agtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaag gacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggagg acggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaa cttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgccc gacaaccactacctgagcacccagtccgccctgagcaaagacccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccggg atcactctcggcatggacgagctgtacaagtagtaatgaatagtgaaacccagctttcttgtacaaagtggtgcaacaaacttctcactactcaaaca agcaggtgacgtggaggagaatcccgggccttctagaggatccggctccggcgagggcaggggaagtctactaacatgcggggacgtggaggaa aatcccggcccagtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagc gtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcg tgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtc caggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagct gaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagc agaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatc ggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctg ctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtagtaatgatagtaatgatagtaatgagggcccttcgaacaaa aactcatctcagaagaggatctgaatatgcataccggtcatcatcaccatcaccattgagaatttcaacgcgttaagtcgacaatcaacctctggatta caaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgccttgtatcatgctattgcttcccgta tggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgct gacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttcccctccctattgccacggcggaactcatcgccgc ctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtg ttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctc ttccgcgtcttcgccttcgccctcagacgagtcggatctcccctttgggccgcctccccgcggcctggccggccgccagcacagtggtcgatcgaccaatg ccctggctcacaaataccactgagatctttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaagga aatttattttcattgcaatagtgtgttggaatttttgtgtctctcactcggaaggacatatgggagggcaaatcatttaaaacatcagaatgagtatttg gtttagagtttggcaacatatgcccatatgctggctgccatgaacaaaggttggctataaagaggtcatcagtatatgaaacagccccctgctgtcca ttccttattccatagaaaagccttgacttgaggttagatttttttatatttgttttgtgttatttttttctttaacatcccctaaaatttccttacatgttttact agccagatttttcctcctctcctgactactcccagtcatagctgtccctcttctcttatggagatccctcgacctgccctaggtcgggcagcaacgcgtata tctggcccgtacatcgcgaagcagcgcaaaacggatcctgcaggtatttgcggccgcggtccgtatactccggaatattaatagatcatggagataa ttaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtc ccaccatcgggcgcgaactcctaaaaaaccgccaccatgaagtgccttttgtacttagccttttttattcatttttttctttaacatcccctaaaatttccttacatgttttact tccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgacttaataggca cagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgtgatttccgct ggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaacaaggaacttt ggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtccaggtgactcctcaccatgtgctggttg atgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaacctggcattct gactataaggtcaaagggctatgtgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggaggg cacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggt gtctggttcgagatggctgataaggatctctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtgg atgtaagtctaattcaggacgttgagaggatcttggattattcccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtgg atctcagctatcttgctcctaaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcg atattgctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacg tggaaattggacccaatggagttctgaggaccagttcaggatataagtttccttatacatgattggacatggtatgttggactccgatcttcatcttag ctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttggtgatactgggctatccaaa aatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctcttttttctttatcataggttaatcattggactattcttggttctcc gagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttggaaagtgataagtcg agaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccgaacctgaaacataaa atgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcact gcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagcctagttcggtggtaacggcgcagtggcggt tttcatggcttgttatgactgttttttttgtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgttatggagca*

	gcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaggtggctcaagtatgggcatcattcgcacatgtaggctcggccctgaccaagt caaatccatgcgggctgctcttgatcttttcggtcgtgagttcggagacgtagccacctactcccaacatcagccggactccgatttacctcgggaacttg ctccgtagtaagacattcatcgcgcttgctgccttcgaccaagaagcggttgttggcgctctcgcggcttacgttctgcccaagtttgagcagccgcgta gtgagatctatatctatgatctcgcagtctccggcgagcaccggaggcagggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcg cttggtgcttatgtgatctacgtgcaagcagattacggtgacgatcccgcagtggctctctatacaaagttgggcatacgggaagaagtgatgcacttt gatatcgacccaagtaccgccacctaacaattcgttcaagccgagatcggcttcccggccgcggagttgttcggtaaattgtcacaacgccgcgaata tagtctttacatgcccttggccacgcccctctttaatacgacgggcaatttgcacttcagaaaatgaagagtttgctttagccataacaaaagtccagta tgcttttcacagcataactggactgatttcagtttacaactattctgtctagtttaagactttattgtcatagtttagatctattttgttcagtttaagacttt attgtccgcccacacccgcttacgcagggcatctgcgcatccacaggaagagcgacccaagtcaatcaggggataacgcaggaaagaacatgtga gcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcga cgatcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgcc gcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaa gctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcc actggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactag aagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcg gtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacg aaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagta tatatgagtaaacttggtctgacagttgacccaagtcaaccggttgtgggcggacaaaatagttgggaactgggaggggtggaaatggagttttca aggattatttagggaagagtgacaaaatagatgggaactgggtgtagcgtcgtaagctaatacgaaaattaaaaatgacaaaatagtttggaact agatttcacttatctggttccaattagatgggtataccctagtatacggtacc
MX01.4 mRaspberry (with terminal stop codon)	*ataacttcgtatagcatacattatacgaagttatctgccaggcacatgggttttactagtatcgattcgcgacctactccggaatattaatagatgtacg ggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttaca taacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggac tttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccttattgacgtcaat gacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggt gatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggc caccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagc agagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattct gcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaacaagtttgtacaaaaaagcaggcttagccaccatggtgagcaaggg cgaggaggtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggc cgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgccctttcgcctgggacatcctgtcccctcagtgcatgtacgg ctccaagggctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgagg acggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacgg ccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatgaagatgaggg ctgaagctgaaggacggcggccactacgacgccgaggtcaagaccacctacatggccaagaagcccgtgcagctgcccggcgcctacaagaccga catcaagctggacatcacctcccacaacgaggactacaccatcgtggaacagtacgagcgcgccgagggccgccacctccaccggcgcctaatagtg aaacccagctttcttgtacaaagtggtgcaacaaacttctcactactcaaacaagcaggtgacgtggaggagaatcccgggccttctagaggatccg gctccggcgagggcaggggaagtctactaacatgcggggacgtggaggaaaatcccggcccagtgagcaagggcgaggagctgttcaccggggt ggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctga ccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccg accacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagac ccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcaca agctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatc gaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacc cagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgag ctgtacaagtagtaatgatagtaatgatagtaatgagggccccttcgaacaaaaactcatctcagaagaggatctgaatatgcataccggtcatcatc accatcaccattgagaattcaacgcgttaagtcgacaatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttt tacgctatgtggatacgctgctttaatgccttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttat gaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctc ctttccgggactttcgctttcccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactga caattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcg gccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctttgg gccgcctccccgcggcctggccggccgccagcacagtggtcgatcgaccaatgccctggctcacaaataccactgagatcttttcccctctgccaaaaa ttatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatttattttcattgcaatagtgtgttggaatttttgtgtctctcact cggaaggacatatgggagggcaaatcatttaaaacatcagaatgagtatttggtttagagtttggcaacatatgcccatatgctggctgccatgaac aaaggttggctataaagaggtcatcagtatatgaaacagccccctgctgtccattccttattccatagaaaagccttgacttgaggttagatttttttat atttgttttgtgttattttttttctttaacatcccctaaaatttttccttacatgttttactagccagatttttcctcctcctcctgactactcccagtcatagctgtccc tcttctcttatggagatccctcgacctgccctaggtcgggcagcaacgcgtatatctggcccgtacatcgcgaagcagcgcaaaacggatcctgcagg tatttgcggccgcggtccgtatactccggaatattaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttc gtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccaccatcgggcgcgaactcctaaaaaaccgccaccatgaagtg ccttttgtacttagccttttattcattgggggtgaattgcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaatta ccattattgcccgtcaagctcagatttaaattggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaag cagacggttggatgtgtcatgcttccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactcca tctgtagaacaatgcaaggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtg acggatgccgaagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaa aatgcagcaattacatatgccccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccatgga catcaccttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaactggaggcaag gcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctctttgctgcagccagatt ccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttgagaggatcttggattattccctc tgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcctaaaaacccaggaaccggtcctgctttc accataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgctccaatcctctcaagaatggtcggaatgatcagtgg aactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaattggacccaatggagttctgaggaccagttcaggatata*

	agtttcctttatacatgattggacatggtatgttggactccgatcttcatcttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgctt cgcaacttcctgatgatgagagtttatttttttggtgatactgggctatccaaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctc tattgcctcttttttctttatcatagggttaatcattggactattcttggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaaga cagatttatacagacatagagatgaaccgacttggaaagtgataagtcgagaagtactagaggatcataatcagccataccacatttgtagaggttt tacttgctttaaaaaacctcccacacctcccccgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttac aaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctg gatctgatcactgcttgagcctagttcggtggtaacggcgcagtggcggttttcatggcttgttatgactgttttttgtacagtctatgcctcgggcatcc aagcagcaagcgcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaggt ggctcaagtatgggcatcattcgcacatgtaggctcggccctgaccaagtcaaatccatgcgggctgctcttgatcttttcggtcgtgagttcggagac gtagccacctactcccaacatcagccggactccgattacctcgggaacttgctccgtagtaagacattcatcgcgcttgctgccttcgaccaagaagcg gttgttggcgctctcgcggcttacgttctgcccaagtttgagcagccgcgtagtgagatctatatctatgatctcgcagtctccggcgagcaccggaggc agggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcgcttggtgcttatgtgatctacgtgcaagcagattacggtgacgatccc gcagtggctctctatacaaagttgggcatacgggaagaagtgatgcactttgatatcgacccaagtaccgccacctaacaattcgttcaagccgaga tcggcttcccggccgcggagttgttcggtaaattgtcacaacgccgcgaatatagtctttacatgcccttggccacgcccctctttaatacgacgggcaa tttgcacttcagaaaatgaagagtttgctttagccataacaaaagtccagtatgcttttcacagcataactggactgatttcagtttacaactattctgt ctagtttaagactttattgtcatagtttagatctatttgttcagtttaagactttattgtccgcccacacccgcttacgcagggcatctgcgcatccacag gaagagcgacccaagtcaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgt tgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgatcaagtcagaggtggcgaaacccgacaggactataaagata ccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgct ttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgcct tatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgt aggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcgg aaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaagg atctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggat cttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttgacccaagtcaaccggttgt gggcggacaaaatagttgggaactgggaggggtggaaatggagttttcaaggattatttagggaagagtgacaaaatagatgggaactgggtgt agcgtcgtaagctaatacgaaaattaaaaatgacaaaatagtttggaactagatttcacttatctggttccaattagatgggtataccctagtatacg gtacc
UCAP.1 ANAP-TAA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttaaaatcccttgt gtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcg aacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagg gaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccgg agcgggacttgaacccgcacagcgcgaacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagg gaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataa tatccgcccatggtgccaaccatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttaaaa tcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatca ctgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgca cagcgcgaacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcac tgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaat acccggagcgggacttgaacccgcacagcgcgaacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcact gatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagc gaggggcgaagaagttgtccatattgcacaactttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat caggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccg cgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactcc ggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccca tatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttccc atagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagt acgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtat tagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattga cgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtac ggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatct cgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggct taggctagcgtttaaacttaagcttgccaccatggaagagcaataccgcccggaagagatagaatccaaagtacagcttcattgggatgagaagcg cacatttgaagtaaccgaagacgagagcaaagagaagtattactgcttttctggccctcccctatccttctggtcgactacacatgggccacgtacgtaa ctacaccatcggtgacgtgatcgcccgctaccagcgtatgctgggcaaaaacgtcctgcagccgatcggctgggacgcgtttggtctgcctgcggaag gcgcggcggtgaaaaacaacaccgctccggcaccgtggacgtacgacaacatcgcgtatatgaaaaaccagctcaaaatgctgggctttggttatg actggagccgcgagctggcaacctgtacgccggaatactaccgttgggaacagaaattcttcaccgagctgtataaaaaaggcctggtatataaga agacttctgcggtcaactggtgtccgaacgaccagaccgtactggcgaacgaacaagttatcgacggctgctgctggcgctgcgataccaaagttga acgtaaagagatcccgcagtggtttatcaaaatcactgcttacgctgacgagctgctcaacgatctggataaactggatcactggcctgacaccgtta aaaccatgcagcgtaactggatcggtcgttccgaaggcgtggagatcaccttcaacgttaacgactatgacaacacgctgaccgtttacactacccgc ccggacaccttttatgggttgtacctacctggcggtagctgcgggtcatccgctggcgcagaaagcggcggaaaatatcctgaactggcggcctttat tgacgaatgccgtaacaccaaagttgccgaagctgaaatggcgacgatggagaaaaaaggcgtcgatactggctttaaagcggttcacccattaac gggcgaagaaattcccgtttgggcagcaaacttcgtattgatggagtacggcacgggcgcagttatggcggtaccggggcacgaccagcgcgacta cgagtttgcctctaaatacggcctgaacatcaaaccggttatcctggcagctgacggctctgagccagatctttctcagcaagccctgactgaaaaag gcgtgctgttcaactctggcgagttcaacggtcttgaccatgaagcggccttcaacgccatcgccgataaactgactgcgatgggcgttggcgagcgt aaagtgaactaccgcctgcgcgactggggtgtttcccgtcagcgttactggggcgcgccgattccgatggtgactctagaagacggtaccgtaatgcc*

	gaccccggacgaccagctgccggtgatcctgccggaggatgtggtaatggacggcattaccagcccgattaaagcagatccggagtgggcgaaaa ctaccgttaacggtatgccagcactgcgtgaaaccgacactttcgacacctttatggagtcctcctgggttcttgcgcgctacacttgcccgcagtacaa agaaggtatgctggattccgaagcggctaactactggctgccggtggatatcgcgattggtggtattgaacacgccattatggagagtctctactgtc gcttcttccacaaactgatgcgtgatgcaggcatggtgaactctgacgaaccagttaaacagttgctgtgtcagggtatggtgctggcagatgccttct actatgttggcgaaaacggcgaacgtaactgggtttccccggttgatgctatcgttgaacgtgacgagaaaggccgtatcgtgaaagcgaaagatg cggcaggccatgaactggtttatacccggcatgagcaaaatgtccaagtcgaagaacaacggtatcgacccgcaggtgatggttgaacgttacggcg cggacaccgttcgtctgtttatgatgtttgcttctccggctgatatgactctcgaatggcaggaatccggtgtggaaggggctaaccgcttcctgaaacg tgtctggaaactggtttacgagcacacagcaaaaagtgatgttgcggcactgaacgttgatgcgctgactgaaaatcagaaagcgctgcgtcgcga tgtgcataaaacgatcgctaaagtgaccgatgatatcggccgtcgtcagaccttcaacaccgcaattgcggcgattatggagctgatgaacaaactg gcgaaagcaccaaccgatggcgagcaggatcgcgctctgatgcaggaagcactgctggccgttgtccgtatgcttaacccgttcaccccgcacatctg cttcacgctgtggcaggaactgaaaggcgaaggcgatatcgacaacgcgccgtggccggttgctgacgaaaaagcgatggtggaagactccacgc tggtcgtggtgcaggttaacggtaaagtccgtgccaaaatcaccgttccggtggacgcaacggaagaacaggttcgcgaacgtgctggccaggaac atctggtagcaaaatatcttgatggcgttactgtacgtaaagtgatttacgtaccaggtaaactcctcaatctggtcgttggcgggcccgtttaatgata gaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctg atcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaat aaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagac aatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatcccggtaatacggttatcc acagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaag tatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccc cgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcag caataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagt aagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttc ccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgga aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcccctagcgcccgctcctttcgctttcttcccttcctttctcgccac gttcgccagctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaac aacactcaaccctatctcgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccgcggagagcgtcgaagcgggggcggtgttcgccgagatcggccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggct ggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.2 ANAP-TAG	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttagaatcccttgt gtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcg aacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagg gaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccgg agcgggacttgaacccgcacagcgcgaacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagg gaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataa tatccgcccatggtgccaaccatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttagaa tcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatca ctgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgca cagcgcgaacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcac tgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaat acccggagcgggacttgaacccgcacagcgcgaacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcact gatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagc gaggggcgaagaagttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat caggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccg cgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactcc ggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccca*

	tatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttccc atagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagt acgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtat tagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattga cgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtac ggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatct cgaggctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggct taggctagcgtttaaacttaagcttgccaccatggaagagcaataccgcccggaagagatagaatccaaagtacagcttcattgggatgagaagcg cacatttgaagtaaccgaagacgagagcaaagagaagtattactgcttttctggccctccctatccttctggtcgactacacatgggccacgtacgtaa ctacaccatcggtgacgtgatcgcccgctaccagcgtatgctgggcaaaaacgtcctgcagccgatcggctgggacgcgtttggtctgcctgcggaag gcgcggcggtgaaaaacaacaccgctccggcaccgtggacgtacgacaacatcgcgtatatgaaaaaccagctcaaaatgctgggctttggttatg actggagccgcgagctggcaacctgtacgccggaatactaccgttgggaacagaaattcttcaccgagctgtataaaaaaggcctggtatataaga agacttctgcggtcaactggtgtccgaacgaccagaccgtactggcgaacgaacaagttatcgacggctgctgctggcgctgcgataccaaagttga acgtaaagagatcccgcagtggtttatcaaaatcactgcttacgctgacgagctgctcaacgatctggataaactggatcactggcctgacaccgtta aaaccatgcagcgtaactggatcggtcgttccgaaggcgtggagatcaccttcaacgttaacgactatgacaacacgctgaccgtttacactacccgc ccggacacctttatgggttgtacctacctggcggtagctgcgggtcatccgctggcgcagaaagcggcggaaaataatcctgaactggcggccttttat tgacgaatgccgtaacaccaaagttgccgaagctgaaatggcgacgatggagaaaaaaggcgtcgatactggctttaaagcggttcacccattaac gggcgaagaaattcccgtttgggcagcaaacttcgtattgatggagtacggcacgggcgcagttatggcggtaccggggcacgaccagcgcgacta cgagtttgcctctaaatacggcctgaacatcaaaccggttatcctggcagctgacggctctgagccagatctttctcagcaagccctgactgaaaaag gcgtgctgttcaactctggcgagttcaacggtcttgaccatgaagcggccttcaacgccatcgccgataaactgactgcgatgggcgttggcgagcgt aaagtgaactaccgcctgcgcgactggggtgtttcccgtcagcgttactggggcgcgccgattccgatggtgactctagaagacggtaccgtaatgcc gaccccggacgaccagctgccggtgatcctgccggaggatgtggtaatggacggcattaccagcccgattaaagcagatccggagtgggcgaaaa ctaccgttaacggtatgccagcactgcgtgaaaccgacactttcgacacctttatggagtcctcctgggttcttgcgcgctacacttgcccgcagtacaa agaaggtatgctggattccgaagcggctaactactggctgccggtggatatcgcgattggtggtattgaacacgccattatggagagtctctactgtc gcttcttccacaaactgatgcgtgatgcaggcatggtgaactctgacgaaccagttaaacagttgctgtgtcagggtatggtgctggcagatgccttct actatgttggcgaaaacggcgaacgtaactgggtttcccggttgatgctatcgttgaacgtgacgagaaaggccgtatcgtgaaagcgaaagatg cggcaggccatgaactggtttatacccggcatgagcaaaatgtccaagtcgaagaacaacggtatcgacccgcaggtgatggttgaacgttacggcg cggacaccgttcgtctgtttatgatgtttgcttctccggctgatatgactctcgaatggcaggaatccggtgtggaaggggctaaccgcttcctgaaacg tgtctggaaactggtttacgagcacacagcaaaaagtgatgttgcggcactgaacgttgatgcgctgactgaaaatcagaaagcgctgcgtcgcga tgtgcataaaacgatcgctaaagtgaccgatgatatcggccgtcgtcagaccttcaacaccgcaattgcggcgattatggagctgatgaacaaactg gcgaaagcaccaaccgatggcgagcaggatcgcgctctgatgcaggaagcactgctggccgttgtccgtatgcttaacccgttcaccccgcacatctg cttcacgctgtggcaggaactgaaaggcgaaggcgatatcgacaacgcgccgtggccggttgctgacgaaaaagcgatggtggaagactccacgc tggtcgtggtgcaggttaacggtaaagtccgtgccaaaatcaccgttccggtggacgcaacggaagaacaggttcgcgaacgtgctggccaggaac atctggtagcaaaatatcttgatggcgttactgtacgtaaagtgatttacgtaccaggtaaactcctcaatctggtcgttggcgggcccgtttaatgata gaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctg atcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaat aaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagac aatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatcccggtaatacggttatcc acagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaag tatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccc cgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcag caataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagt aagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttc ccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgga aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcg cacatttcccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac gttcgccagctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaac aacactcaaccctatctctgtctattcttttgatttataagggatttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccatttttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccgggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.3	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg*

ANAP-TGA	*ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttgaaatcccttgt gtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcg aacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagg gaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccgg agcgggacttgaacccgcacagcgcgaacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagg gaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataa tatccgcccatggtgccaaccatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttgaaa tcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatca ctgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgca cagcgcgaacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcac tgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaat acccggagcgggacttgaacccgcacagcgcgaacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcact gatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagc gaggggggcgaagaagttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat caggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccg cgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactcc ggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccca tatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttccc atagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagt acgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtat tagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattga cgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtac ggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatct cgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggct taggctagcgtttaaacttaagcttgccaccatggaagagcaataccgcccggaagagatagaatccaaagtacagcttcattgggatgagaagcg cacatttgaagtaaccgaagacgagagcaaagagaagtattactgcttttctggccctccctatccttctggtcgactacacatgggccacgtacgtaa ctacaccatcggtgacgtgatcgcccgctaccagcgtatgctgggcaaaaacgtcctgcagccgatcggctgggacgcgtttggtctgcctgcggaag gcgcggcggtgaaaaacaacaccgctccggcaccgtggacgtacgacaacatcgcgtatatgaaaaaccagctcaaaatgctgggctttggttatg actggagccgcgagctggcaacctgtacgccggaatactaccgttgggaacagaaattcttcaccgagctgtataaaaaaggcctggtatataaga agacttctgcggtcaactggtgtccgaacgaccagaccgtactggcgaacgaacaagttatcgacggctgctgctggcgctgcgataccaaagttga acgtaaagagatcccgcagtggtttatcaaaatcactgcttacgctgacgagctgctcaacgatctggataaactggatcactggcctgacaccgtta aaaccatgcagcgtaactggatcggtcgttccgaaggcgtggagatcaccttcaacgttaacgactatgacaacacgctgaccgtttacactacccgc ccggacacctttatgggttgtacctacctggcggtagctgcgggtcatccgctggcgcagaaagcggcggaaaataatcctgaactggcggcctttat tgacgaatgccgtaacaccaaagttgccgaagctgaaatggcgacgatggagaaaaaaggcgtcgatactggctttaaagcggttcacccattaac gggcgaagaaattcccgtttgggcagcaaacttcgtattgatggagtacggcacgggcgcagttatggcggtaccggggcacgaccagcgcgacta cgagtttgcctctaaatacggcctgaacatcaaaccggttatcctggcagctgacggctctgagccagatctttctcagcaagccctgactgaaaaag gcgtgctgttcaactctggcgagttcaacggtcttgaccatgaagcggccttcaacgccatcgccgataaactgactgcgatgggcgttggcgagcgt aaagtgaactaccgcctgcgcgactggggtgtttcccgtcagcgttactggggcgcgccgattccgatggtgactctagaagacggtaccgtaatgcc gaccccggacgaccagctgccggtgatcctgccggaggatgtggtaatggacggcattaccagcccgattaaagcagatccggagtgggcgaaaa ctaccgttaacggtatgccagcactgcgtgaaaccgacactttcgacacctttatggagtcctcctgggttcttgcgcgctacacttgcccgcagtacaa agaaggtatgctggattccgaagcggctaactactggctgccggtggatatcgcgattggtggtattgaacacgccattatggagagtctctactgtc gcttcttccacaaactgatgcgtgatgcaggcatggtgaactctgacgaaccagttaaacagttgctgtgtcagggtatggtgctggcagatgccttct actatgttggcgaaaacggcgaacgtaactgggtttcccggttgatgctatcgttgaacgtgacgagaaaggccgtatcgtgaaagcgaaagatg cggcaggccatgaactggtttataccggcatgagcaaaatgtccaagtcgaagaacaacggtatcgacccgcaggtgatggttgaacgttacggcg cggacaccgttcgtctgtttatgatgtttgcttctccggctgatatgactctcgaatggcaggaatccggtgtggaaggggctaaccgcttcctgaaacg tgtctggaaactggtttacgagcacacagcaaaaagtgatgttgcggcactgaacgttgatgcgctgactgaaaatcagaaagcgctgcgtcgcga tgtgcataaaacgatcgctaaagtgaccgatgatatcggccgtcgtcagaccttcaacaccgcaattgcggcgattatggagctgatgaacaaactg gcgaaagcaccaaccgatggcgagcaggatcgcgctctgatgcaggaagcactgctggccgttgtccgtatgcttaacccgttcaccccgcacatctg cttcacgctgtggcaggaactgaaaggcgaaggcgatatcgacaacgcgccgtggccggttgctgacgaaaaagcgatggtggaagactccacgc tggtcgtggtgcaggttaacggtaaagtccgtgccaaaatcaccgttccggtggacgcaacggaagaacaggttcgcgaacgtgctggccaggaac atctggtagcaaaatatcttgatggcgttactgtacgtaaagtgatttacgtaccaggtaaactcctcaatctggtcgttggcgggcccgtttaatgata gaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctg atcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaat aaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggtggggtggggcaggacagcaagggggaggattgggaagac aatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccggtaatacggttatcc acagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaag tatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccc cgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcag caataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagt aagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttc ccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggga aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac gttcgccagctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt*

	gatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaac aacactcaaccctatctctgttctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttccgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgcccgcaacctcccctctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.4 DanAla-TAA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttaaaatcccttgt gtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcg aacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagg gaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccgg agcgggacttgaacccgcacagcgcgaacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagg gaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataa tatccgcccatggtgccaaccatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttaaaa tcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatca ctgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgca cagcgcgaacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcac tgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaat acccggagcgggacttgaacccgcacagcgcgaacgccgagggatttaaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcact gatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagc gaggggcgaagaagttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat caggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccg cgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactcc ggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccca tatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttccc atagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagt acgcccccttattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtat tagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattga cgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtac ggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatct cgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggct gcaccatgcaagagcagtaccgccctgaagagattgagtccaaagtgcagttacactgggatgagaagcgaacattcgaggtgaccgaagacga gtccaaggagaagtactactgcttaagcgccaatccttacccgagcggtagactgcatatgggacatgtgcgaaattacaccatcggcgacgtgatc gcccggtaccagagaatgctcggaaagaacgtgctccagccaatcggatgggacgcattcgggttacctgcagaaggggcggccgtgaaaaaca atacagctccagctccatggacctatgataacatcgcgtatatgaaaaatcagctgaagatgcttgggtttggctacgactggtcccgggaactggct acttgcacacccgaatactatagatgggagcagaagttttttacagagctctacaagaaagggctggtgtacaaaaagaccagcgccgtaaactgg tgcccaaatgaccagaccgtgttggcgaatgaacaggtgatcgatgggtgctgttggagatgtgacactaaagtagaaaggaaggagatcccaca gtggttcataaagatcactgcgtacgctgatgagctcctgaacgacctggacaagttagaccattggcctgacacagtaaagactatgcaaaggaat tggataggacgtagtgagggcgtggagattacatttaacgtaaatgattatgataacacccttacagtgtacactactcgaccggacgcctttatggg ctgcacttatcttgcagtggccgccggacatccgttagcccagaaagctgcagaaaataaccctgagctggcagcctttatcgacgagtgcagaaac accaaggtcgctgaggccgagatggctacaatggaaaagaaaggagtagataccggctttaaggctgtccacacctctcacaggggaagaaattcc tgtatgggctgccaacttcgtcctgatggagtatggcactggcgcagtgatggcagtacccggacatgaccagcgagattacgagttcgcctccaaat acgggctcaatatcaagccagtgattctggcggctgatggttctgaaccagacctcagtcagcaagctctgactgaaaagggtgttctcttcaactctg gggaatttaatggtcttgaccacgaagcggctttcaacgcaatcgcagacaagttgaccgccatgggggttggggagagaaagggtgaactaccgg ctgagggactggggagttagccggcagcgctattggggcgccccaatccctatggtcaccttggaggatggaacagtgatgcccacaccagatgac cagctccctgtgattcttcccgaagacgtggtgatggacggaattacttcccccattaaggccgacccagagtgggcgaaaacaacagttaatggtat gccagcgctgcgggaaactgacaccttcgacactttcatggaaagctcttggatctacgcacgctacacgtgccccccaatacaaggagggaatgctg gacagcgaggcggcaaactattggctccccgttgacattggaattggagggatcgaacacgccataatgacattgctgtattttagatttttccataa actgatgcgagatgcaggcatggtgaactctgacgagccagcaaaacagctgttatgccagggaatggtgctggccgacgctttctactacgttggc gaaaacggcgagcgtaattgggtaagtcctgttgacgcaatagtagagcgcgacgagaagggaaggatagttaaggccaaagacgcggctggg cacgaattggtgtacacgggcatgtcaaagatgagtaaatccaaaaacaatggtattgaccccaggtcatggtcgaacggtatggggcagatact gtacggctctttatgatgttcgcctcccccgcagatatgacccttgagtggcaggaatccggggtggaggggcgaacagattcttaaagcgcgtctg gaagctggtgtacgagcacaccgctaaagctgatgtggccgcactcaacgtggatgccttgacagaaaaccagaaggccctgcggcgcgacgtgc acaagactatcgcaaaggttacagatgacatcgggaggcggcaaacgttcaatacagccatcgcggcaattatggagttgatgaataagctggca aaggctccaacggatggagagcaggaccgagccctcatgcaggaggctctgctggccgtggtaagaatgttaaaccccttcactcctcatatctgttt cacattgtggcaggaactgaaggggagggagacatcgataacgctccgtggccagtggccgacgagaaggctatggtcgaagatagcaccctc*

	gtggtggtccaagtaaacgggaaggtgagagctaaaataacggtgcccgtcgacgccactgaggagcaagtacgagagagggccggccaggag cacctggtggccaagtacctcgacggagtcacagtcagaaaggtcatctatgtgccgggcaaactgctgaacctcgttgtgggctaagacccagcttt cttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaag tttaaaccgctgatcagcctcgac tgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgagga aattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggc atgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccggtaatacggttatccacagaatcag gggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagt aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtag ataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaacc agccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcg ccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatc aaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcact catggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagt gtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttct tcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccag cgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaa tgttgaatactcatactcttcctttttca atattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccc cgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagc gtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgc ggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccag ctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcac gtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaac cctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattct gtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgt ggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcc cctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaa gtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagta caagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccg tcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacga cggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggtt cccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccga ccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgcc ccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccg gtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctcca gcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaat aaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaat catggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtataatgtat gctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.5 DanAla-TAG	gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttagaatcccttgt gtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcg aacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagg gaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccgg agcgggacttgaacccgcacagcgcgaacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagg gaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataa tatccgcccatggtgccaaccatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttagaa tcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatca ctgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgca cagcgcgaacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcac tgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaat acccggagcgggacttgaacccgcacagcgcgaacgccgagggatttagaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcact gatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagc gagggggcgaagaagttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat caggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccg cgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactcc ggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccca tatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttccc atagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagt acgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtat tagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattga cgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtac ggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatct cgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggct gcaccatgcaagagcagtaccgccctgaagagattgagtccaaagtgcagttacactgggatgagaagcgaacattcgaggtgaccgaagacga gtccaaggagaagtactactgcttaagcgccaatccttacccgagcggtagactgcatatgggacatgtgcgaaattacaccatcggcgacgtgatc gcccggtaccagagaatgctcggaaagaacgtgctccagccaatcggatgggacgcattcgggttacctgcagaaggggcggccgtgaaaaaca atacagctccagctccatggacctatgataacatcgcgtatatgaaaatcagctgaagatgcttgggtttggctacgactggtcccgggaactggct

	acttgcacacccgaatactatagatgggagcagaagttttttacagagctctacaagaaagggctggtgtacaaaaagaccagcgccgtaaactgg tgcccaaatgaccagaccgtgttggcgaatgaacaggtgatcgatgggtgctgttggagatgtgacactaaagtagaaaggaaggagatcccaca gtggttcataaagatcactgcgtacgctgatgagctcctgaacgacctggacaagttagaccattggcctgacacagtaaagactatgcaaaggaat tggataggacgtagtgagggcgtggagattacatttaacgtaaatgattatgataacacccttacagtgtacactactcgaccggacgccttttatggg ctgcacttatcttgcagtggccgccggacatccgttagcccagaaagctgcagaaaataaccctgagctggcagccttttatcgacgagtgcagaaac accaaggtcgctgaggccgagatggctacaatggaaaagaaaggagtagataccggctttaaggctgtccacccctctcacaggggaagaaattcc tgtatgggctgccaacttcgtcctgatggagtatggcactggcgcagtgatggcagtacccggacatgaccagcgagattacgagttcgcctccaaat acgggctcaatatcaagccagtgattctggcggctgatggttctgaaccagacctcagtcagcaagctctgactgaaaagggtgttctcttcaactctg gggaatttaatggtcttgaccacgaagcggctttcaacgcaatcgcagacaagttgaccgccatgggggttggggagagaaaggtgaactaccgg ctgagggactggggagttagccggcagcgctattggggcgccccaatccctatggtcaccttggaggatggaacagtgatgcccacaccagatgac cagctccctgtgattcttcccgaagacgtggtgatggacggaattacttccccattaaggccgacccagagtgggcgaaaacaacagttaatggtat gccagcgctgcgggaaactgacaccttcgacactttcatggaaagctcttggatctacgcacgctacacgtgcccccaatacaaggagggaatgctg gacagcgaggcggcaaactattggctccccgttgacattggaattggagggatcgaacacgccataatgacattgctgtattttagatttttccataa actgatgcgagatgcaggcatggtgaactctgacgagccagcaaaacagctgttatgccagggaatggtgctggccgacgctttctactacgttggc gaaaacggcgagcgtaattgggtaagtcctgttgacgcaatagtagagcgcgacgagaagggaaggatagttaaggccaaagacgcggctggg cacgaattggtgtacacgggcatgtcaaagatgagtaaatccaaaaacaatggtattgacccccaggtcatggtcgaacggtatggggcagatact gtacggctctttatgatgttcgcctccccgcagatatgacccttgagtggcaggaatccggggtggaggggcgaacagattcttaaagcgcgtctg gaagctggtgtacgagcacaccgctaaagctgatgtggccgcactcaacgtggatgccttgacagaaaaccagaaggccctgcggcgcgacgtgc acaagactatcgcaaaggttacagatgacatcgggaggcggcaaacgttcaatacagccatcgcggcaattatggagttgatgaataagctggca aaggctccaacggatggagagcaggaccgagccctcatgcaggaggctctgctggccgtggtaagaatgttaaacccttcactcctcatatctgttt cacattgtggcaggaactgaaggggggagggagacatcgataacgctccgtggccagtggccgacgagaaggctatggtcgaagatagcaccctc gtggtggtccaagtaaacgggaaggtgagagctaaaataacggtgcccgtcgacgccactgaggagcaagtacgagagagggccggccaggag cacctggtggccaagtacctcgacggagtcacagtcagaaaggtcatctatgtgccgggcaaactgctgaacctcgttgtgggctaagacccagctttt cttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctgatcagcctcgac tgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaataaaatgagga aattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggc atgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccggtaatacggttatccacagaatcag gggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagt aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtag ataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaacc agcccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcg ccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatc aaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcact catggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagt gtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttct tcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccag cgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttca atattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccc cgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagc gtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagcgc ggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttttcgctttcttcccttcctttctcgccacgttcgccag ctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcac gtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaac cctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattct gtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgt ggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcc cctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaa gtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccatttcggatctgatcagcacgtgatgaccgagta caagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccg tcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacga cggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggtt cccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccga ccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgcc ccgcaacctcccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccg gtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctcca gcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaat aaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaat catggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtatatgtat gctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.6 DanAla-TGA	gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttgaaatcccttgt gtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcg aacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagg gaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccgg agcgggacttgaacccgcacagcgcgaacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagg gaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataa tatccgcccatggtgccaaccatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttgaaa

tcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatca
ctgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatacccggagcgggacttgaacccgca
cagcgcgaacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcactgatagggaacttataagtctctatcac
tgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaat
acccggagcgggacttgaacccgcacagcgcgaacgccgagggatttgaaatcccttgtgtctaccgattccaccatccgggcggtgtctctatcact
gatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagc
gaggggcgaagaagttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat
caggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccg
cgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactcc
ggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccca
tatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttccc
atagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagt
acgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtat
tagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattga
cgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtac
ggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatct
cgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggct
gcaccatgcaagagcagtaccgccctgaagagattgagtccaaagtgcagttacactgggatgagaagcgaacattcgaggtgaccgaagacga
gtccaaggagaagtactactgcttaagcgccaatccttacccgagcggtagactgcatatgggacatgtgcgaaattacaccatcggcgacgtgatc
gcccggtaccagagaatgctcggaaagaacgtgctccagccaatcggatgggacgcattcgggttacctgcagaaggggcggccgtgaaaaaca
atacagctccagctccatggacctatgataacatcgcgtatatgaaaaatcagctgaagatgcttgggtttggctacgactggtcccgggaactggct
acttgcacacccgaatactatagatgggagcagaagttttttacagagctctacaagaaagggctggtgtacaaaaagaccagcgccgtaaactgg
tgcccaaatgaccagaccgtgttggcgaatgaacaggtgatcgatgggtgctgttggagatgtgacactaaagtagaaaggaaggagatcccaca
gtggttcataaagatcactgcgtacgctgatgagctcctgaacgacctggacaagttagaccattggcctgacacagtaaagactatgcaaaggaat
tggataggacgtagtgagggcgtggagattacatttaacgtaaatgattatgataacacccttacagtgtacactactcgaccggacgcctttatggg
ctgcacttatcttgcagtggccgccggacatccgttagcccagaaagctgcagaaaataaccctgagctggcagccttttatcgacgagtgcagaaac
accaaggtcgctgaggccgagatggctacaatggaaaagaaaggagtagataccggctttaaggctgtccaccctctcacaggggaagaaattcc
tgtatgggctgccaacttcgtcctgatggagtatggcactggcgcagtgatggcagtacccggacatgaccagcgagattacgagttcgcctccaaat
acgggctcaatatcaagccagtgattctggcggctgatggttctgaaccagacctcagtcagcaagctctgactgaaaagggtgttctcttcaactctg
gggaatttaatggtcttgaccacgaagcggctttcaacgcaatcgcagacaagttgaccgccatgggggttggggagagaaaggtgaactaccgg
ctgagggactgggagttagccggcagcgctattggggcgccccaatccctatggtcaccttggaggatggaacagtgatgcccacaccagatgac
cagctccctgtgattcttcccgaagacgtggtgatggacggaattacttcccccattaaggccgacccagagtgggcgaaaacaacagttaatggtat
gccagcgctgcgggaaactgacaccttcgacactttcatggaaagctcttggatctacgcacgctacacgtgccccccaatacaaggagggaatgctg
gacagcgaggcggcaaactattggctccccgttgacattggaattggagggatcgaacacgccataatgacattgctgtattttagatttttccataa
actgatgcgagatgcaggcatggtgaactctgacgagccagcaaaacagctgttatgccagggaatggtgctggccgacgctttctactacgttggc
gaaaacggcgagcgtaattgggtaagtcctgttgacgcaatagtagagcgcgacgagaagggaaggatagttaaggccaaagacgcggctgggg
cacgaattggtgtacacgggcatgtcaaagatgagtaaatccaaaaacaatggtattgacccccaggtcatggtcgaacggtatggggcagatact
gtacggctctttatgatgttcgcctccccgcagatatgacccttgagtggcaggaatccggggtggaggggcgaacagattcttaaagcgcgtctg
gaagctggtgtacgagcacaccgctaaagctgatgtggccgcactcaacgtggatgccttgacagaaaaccagaaggccctgcggcgcgacgtgc
acaagactatcgcaaaggttacagatgacatcgggaggcggcaaacgttcaatacagccatcgcggcaattatggagttgatgaataagctggca
aaggctccaacggatggagagcaggaccgagccctcatgcaggaggctctgctggccgtggtaagaatgttaaaccccttcactcctcatatctgttt
cacattgtggcaggaactgaaggggagggagacatcgataacgctccgtggccagtggccgacgagaaggctatggtcgaagatagcaccctc
gtggtggtccaagtaaacgggaaggtgagagctaaaataacggtgcccgtcgacgccactgaggagcaagtacgagagagggccggccaggag
cacctggtggccaagtacctcgacggagtcacagtcagaaaggtcatctatgtgccgggcaaactgctgaacctcgttgtgggctaagacccagcttt
cttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctgatcagcctcgac
tgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaataaaatgagga
aattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggc
atgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatcccggtaatacggttatccacagaatcag
gggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagt
aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtag
ataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaacc
agccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcg
ccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatc
aaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcact
catggttatggcagcactgcataatttctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagt
gtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttct
tcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccag
cgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttca
atattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccc
cgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagc
gtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagcgc
ggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccag
ctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcac
gtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaac
cctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattct
gtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgt
ggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcc
cctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaa
gtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagta
caagcccacggtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccg
tcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacga
cggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggtt
cccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccga
ccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgcc

	ccgcaacctcccctctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccg gtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctcca gcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttatataatggttacaaataaagcaatagcatcacaaatttcacaaat aaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaat catggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtataatgtat gctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.7 EcTrp(h9)-TAA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagccctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaattttaatcgcggcctgagcttttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttaaagaccggtgctctacca attgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcac gtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttta aagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctcta tcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcga actcccaacacccggttttaaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataatatccgcccatggtgccaaccatg caaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttaaagaccggtgctctaccaattgaactacgccccta gtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccc agaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttaaagaccggtgctcta ccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggattt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggt tttaaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgaggggcgaagaagttgtccatattgcacaacttttgtatacaaag ttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctcta gaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagat gtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgatta ttgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggct gaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatt tacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatg cccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatggg cgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaa tgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccca ctgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggat ccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaggatccatgactaagcccatcgtttttgctggcgcacagccctcaggt gaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatc accgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttt tgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagata aatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgagtcctgc tggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgttt attccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctgg aagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggc gtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaag gtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaa gatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaatgatagggatcc aacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctga tcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaata aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagaca atagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatccccggtaatacggttatcca cagaatcaggggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagt atatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcccc gtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagta agtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcc caacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaa aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac gttcgccagctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttgacgttggagtccacgttcttaatagtggactcttgttccaaactggaac aacactcaaccctatctcgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc*

	acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctcccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.8 EcTrp(h9)-TAG	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttagagaccggtgctctacca attgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcac gtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttta gagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctcta tcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcga actcccaacacccggttttagagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataatatccgcccatggtgccaaccatg caaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttagagaccggtgctctaccaattgaactacgcccctg gtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccc agaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttagagaccggtgctcta ccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggattt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggt tttagagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgaggggggcgaagaagttgtccatattgcacaacttttgtatacaaag ttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctcta gaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagat gtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgatta ttgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggct gaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatt tacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatg cccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatggg cgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaa tgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccca ctgcttactggcttatcaccgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggat ccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaggatccatgactaagcccatcgttttgctggcgcacagccctcaggt gaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatc accgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttt tgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagata aatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgagtcctgc tggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgttt attccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctgg aagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggc gtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaag gtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaa gatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaatgatagggatcc aacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctga tcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaata aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagaca atagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatcccggtaatacggttatcca cagaatcaggggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagt atatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcccc gtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagta agtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcc caacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaa aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac gttcgccagctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttgacgttggagtccacgttcttaatagtggactcttgttccaaactggaac aacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag*

	caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.9 EcTrp(h9)-TGA	gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttgaagaccggtgctctacca attgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcac gtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttg aagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctcta tcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcga actcccaacacccggttttgaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataatatccgcccatggtgccaaccatg caaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttgaagaccggtgctctaccaattgaactacgccctg gtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccc agaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttgaagaccggtgctcta ccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggt tttgaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgaggggcgaagaagttgtccatattgcacaacttttgtatacaaag ttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctcta gaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagat gtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgatta ttgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggct gaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatt tacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatg cccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatggg cgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaa tgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccca ctgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggat ccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaggatccatgactaagcccatcgtttttgctggcgcacagccctcaggt gaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatc accgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttt tgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagata aatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgttatcaaactaatctgagtcctgc tggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgttt attccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctgg aagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggc gtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaag gtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaa gatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaatgatagggatcc aacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaacccgctga tcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaata aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagaca atagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccggtaatacggttatcca cagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagt atatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcccc gtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagta agtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcc caacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaa aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccaactgatcttcagcatcttttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac

	gttcgccagctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttcttaatagtggactcttgttccaaactggaac aacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctcccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.10 EcTrp(h14)-TAA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggctttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttaaagaccggtgctctacca attgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcac gtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttaa aagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctcta tcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcga actcccaacacccggttttaaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataatatccgcccatggtgccaaccatg caaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttaaagaccggtgctctaccaattgaactacgccctg gtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccc agaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttaaagaccggtgctcta ccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggattt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggt tttaaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgaggggcgaagaagttgtccatattgcacaacttttgtatacaaag ttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctcta gaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagat gtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgatta ttgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggct gaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatt tacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatg cccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatggg cgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaa tgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccca ctgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggat ccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaggatccatgactaagcccatcgtttttgctggcgcacagccctcaggt gaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatc accgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttt tgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagata aatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgggtccttg tggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgttt attccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctgg aagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggc gtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaag gtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaa gatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaatgatagggatcc aacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctga tcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaata aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagaca atagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccggtaatacggttatcca cagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagt atatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcccc gtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagta agtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcc caacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgga aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc*

	ttccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaaactatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac gttcgccagctttccccgtcaagctctaaatcgggggctcccttttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaac aacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggct ggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.11 EcTrp(h14)-TAG	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttagagaccggtgctctacca attgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggatttcac gtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttta gagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctcta tcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcga actcccaacacccggttttagagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataatatccgcccatggtgccaaccatg caaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttagagaccggtgctctaccaattgaactacgcccctg gtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccc agaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttagagaccggtgctcta ccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggattt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggt tttagagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgaggggcgaagaagttgtccatattgcacaacttttgtatacaaag ttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctcta gaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagat gtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgatta ttgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggct gaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatt tacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccattgacgtcaatgacggtaaatggcccgcctggcattatg cccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatggg cgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaa tgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccа ctgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggat ccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaggatccatgactaagcccatcgtttttgctggcgcacagccctcaggt gaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatc accgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttt tgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagata aatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgggtccttg tggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgttt attccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaatаacgttatcggcctgctgg aagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggc gtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaag gtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaa gatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaatgatagggatcc aacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctga tcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaata aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagaca atagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccggtaatacggttatcca cagaatcaggggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagt atatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcccc gtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagta agtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcc caacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt*

	gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaa aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac gttcgccagctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaac aacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctcccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.12 EcTrp(h14)-TGA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgcatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttgaagaccggtgctctacca attgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcac gtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttg aagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctcta tcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcga actcccaacacccggttttaaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgatataatatccgcccatggtgccaaccatg caaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttgaagaccggtgctctaccaattgaactacgcccctg gtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccc agaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggttttgaagaccggtgctcta ccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggattt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattagatccaaaaaatggcaggggcggagagactcgaactcccaacacccggt tttgaagaccggtgctctaccaattgaactacgcccctggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgaggggcgaagaagttgtccatattgcacaacttttgtatacaaag ttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctcta gaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagat gtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgatta ttgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggct gaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatt tacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatg cccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatggg cgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaa tgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccca ctgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggat ccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaggatccatgactaagcccatcgtttttgctggcgcacagccctcaggt gaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatc accgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttt tgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagata aatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgggtccttg tggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgttt attccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctgg aagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggc gtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaag gtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaa gatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggtttgtggcgaagccgtaatgatagggatcc aacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgaagctcggtaccaagcttaagtttaaaccgcctga tcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaata aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagaca atagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatccccggtaatacggttatcca cagaatcaggggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagt atatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcccc*

	gtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagc aataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagta agtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcc caacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgga aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttta ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcg cacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgtt ccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcccgctcctttcgctttcttcccttcctttctcgccacc gttcgccagctttcccgtcaagctctaaatcgggggctcccttttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggt gatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaac aacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactcc gcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctga gctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgt gatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccgggagagcgtcgaagcgggggcggtgttcgccgagatcggccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.13 OMeY-TAA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttaaagtccgccgcgtttagccacttcgctacc cctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtga tttcccagaacacatagcgacatgcaaatattaaaaaatggagggggacggattcgaaccgccgaacccaaagggagcggatttaaagtccgccg cgtttagccacttcgctacccctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaatggtgggggaaggattcgaaccttcgaagtctgtgacgg cagatttaaagtctgctccctttggccgctcgggaaccccaccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgataggg aacttataagtctctatcactgatagggattcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgata gtccattttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaatta attctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggt gccaacaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttaaagtccgccgcgtttagccacttcgctacccctccgg tgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccca gaacacatagcgacatgcaaatattaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttaaagtccgccgcgttta gccacttcgctacccctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggattt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaatggtgggggaaggattcgaaccttcgaagtctgtgacggcagatt taaagtctgctccctttggccgctcgggaaccccaccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaactta taagtctctatcactgatagggattcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccat ttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattcta attatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggggcgaagaagttgtccatat tgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcc tattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgac ctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccag atatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactta cggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccatt gacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggta aatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcgg ttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaa atcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctct ctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcga cggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaacccaagctggctagcgccaccatgg caagcagtaacttgattaaacaattgcaagagcgggggctggtagcccaggtgacggacgaggaagcgttagcagagcgactggcgcaaggccc gatcgcactcgtgtgtggcttcgatcctaccgctgacagcttgcatttggggcatcttgttccattgttatgcctgaaacgcttccagcaggcgggccac aagccggttgcgctggtaggcggcgcgacgggtctgattggcgacccgagcttcaaagctgccgagcgtaagctgaacaccgaagaaactgttcag gagtgggtggacaaaatccgtaagcaggttgccccgttcctcgatttcgactgtggagaaaactctgctatcgcggccaataattatgactggttcgg caatatgaatgtgctgaccttcctgcgcgatattggcaaacacttctccgttaaccagatgatcaacaaagaagcggttaagcagcgtctcaaccgtg aagatcaggggatttcgttcactgagttttcctacaacctgctgcagggttatagtatggcctgtttgaacaaacagtacggtgtggtgctgcaaattg gtggttctgaccagtggggtaacatcacttctggtatcgacctgacccgtcgtctgcatcagaatcaggtgtttggcctgaccgttccgctgatcactaa agcagatggcaccaaatttggtaaaactgaaggcggcgcagtctggttggatccgaagaaaaccagcccgtacaaattctaccagttctggatcaa*

	cactgcgcgtgccgacgtttaccgcttcctgaagttcttcacctttatgagcattgaagagatcaacgccctggaagaagaagataaaaacagcggt aaagcaccgcgcgcccagtatgtactggcggagcaggtgactcgtctggttcacggtgaagaaggtttacaggcggcaaaacgtattaccgaatgc ctgttcagcggttctttgagtgcgctgagtgaagcggacttcgaacagctggcgcaggacggcgtaccgatggttgagatggaaaagggcgcagac ctgatgcaggcactggtcgattctgaactgcaaccttcccgtggtcaggcacgtaaaactatcgcctccaatgccatcaccattaacggtgaaaaaca gtccgatcctgaatacttctttaaagaagaagatcgtctgtttggtcgttttaccttactgcgtcgcggtaaaaagaattactgtctgatttgctggaaag ggcccgtttaatagtgaaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagct taagtttaaaccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcc cactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaaggggg aggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccc ggtaatacggttatccacagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttt taaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcca tagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccg gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgc cgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggctt cattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaa gtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtact caaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaa agtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactg atcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatg ttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaaca aataggggttccgcgcacatttccccgaaaagtgccacctgacgtctataccgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacg tgagttacgcgtcgttccactgagcgtcagaccccgtagaaaactatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgc gccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttcgctttcttc ccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaa aaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactctt gttccaaactggaacaacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttta acaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgc atctcaattagtcagcaaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagt cccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggc cgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcgga tctgatcagcacgtgatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgc cgactaccccgccacgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgac atcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcg gcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggt tcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccgggg tgcccgccttcctggagacctccgcgccccgcaacctcccctttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgc gcacctggtgcatgacccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgtttt ccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataa agcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccg tcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgt gcctggcagataacttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.14 OMeY-TAG	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttagagtccgccgcgtttagccacttcgctacc cctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtga tttcccagaacacatagcgacatgcaaatattaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttagagtccgccg cgtttagccacttcgctacccctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaatggtgggggaaggattcgaaccttcgaagtctgtgacgg cagatttagagtctgctccctttggccgctcgggaaccccaccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgataggg aacttataagtctctatcactgatagggattcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgata gtccattttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaatta attctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggt ggccaacaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttagagtccgccgcgtttagccacttcgctacccctccgg tgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccca gaacacatagcgacatgcaaatattaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttagagtccgccgcgtttta gccacttcgctacccctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaatggtgggggaaggattcgaaccttcgaagtctgtgacggcagatt tagagtctgctcccttggccgctcgggaaccccaccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaactta taagtctctatcactgatagggattcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccat ttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattcta attatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggggcgaagaagttgtccatat tgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcc tattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgac ctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccag atatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactta cggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccatt gacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccctattgacgtcaatgacggta aatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcgg ttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaa*

	atcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctct ctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcga cggtaccgcgggcccgggatccaccggtatctatttagaaaatcacaagtttgtacaaaaaagcaggctttaaccctcagctggctagcgccaccatgg caagcagtaacttgattaaacaattgcaagagcgggggctggtagcccaggtgacggacgaggaagcgttagcagagcgactggcgcaaggccc gatcgcactcgtgtgtggcttcgatcctaccgctgacagcttgcatttggggcatcttgttccattgttatgcctgaaacgcttccagcaggcgggccac aagccggttgcgctggtaggcggcgcgacgggtctgattggcgacccgagcttcaaagctgccgagcgtaagctgaacaccgaagaaactgttcag gagtgggtggacaaaatccgtaagcaggttgccccgttcctcgatttcgactgtggagaaaactctgctatcgcggccaataattatgactggttcgg caatatgaatgtgctgaccttcctgcgcgatattggcaaacacttctccgttaaccagatgatcaacaaagaagcggttaagcagcgtctcaaccgtg aagatcaggggatttcgttcactgagttttcctacaacctgctgcagggttatagtatggcctgtttgaacaaacagtacggtgtggtgctgcaaattg gtggttctgaccagtggggtaacatcacttctggtatcgacctgacccgtcgtctgcatcagaatcaggtgtttggcctgaccgttccgctgatcactaa agcagatggcaccaaatttggtaaaactgaaggcggcgcagtctggttggatccgaagaaaaccagcccgtacaaattctaccagttctggatcaa cactgcgcgtgccgacgtttaccgcttcctgaagttcttcacctttatgagcattgaagagatcaacgccctggaagaagaagataaaaacagcggt aaagcaccgcgcgcccagtatgtactggcggagcaggtgactcgtctggttcacggtgaagaaggtttacaggcggcaaaacgtattaccgaatgc ctgttcagcggttctttgagtgcgctgagtgaagcggacttcgaacagctggcgcaggacggcgtaccgatggttgagatggaaaagggcgcagac ctgatgcaggcactggtcgattctgaactgcaaccttcccgtggtcaggcacgtaaaactatcgcctccaatgccatcaccattaacggtgaaaaaca gtccgatcctgaatacttctttaaagaagaagatcgtctgtttggtcgttttaccttactgcgtcgcggtaaaaagaattactgtctgatttgctggaaag ggcccgtttaatagtgaaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagct taagtttaaaccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcc cactgtccttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaaggggg aggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatcccc ggtaatacggttatccacagaatcaggggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttt taaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcca tagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccg gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgc cgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggctt cattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaa gtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtact caaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaa agtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactg atcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatg ttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaaca aatagggggttccgcgcacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacg tgagttacgcgtcgttccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgc gccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttc ccttcctttctcgccacgttcgccagctttccccgtcaagctctaaatcgggggctcccctttagggttccgatttagtgctttacggcacctcgaccccaaa aaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactctt gttccaaactggaacaacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaa acaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgc atctcaattagtcagcaaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagt cccgccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggc cgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcgga tctgatcagcacgtgatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgc cgactaccccgccacgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgac atcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcg gcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggt tcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccgggg tgcccgccttcctggagacctccgcgccccgcaacctcccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgc gcacctggtgcatgacccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgtttt ccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataa agcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccg tcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgt gcctggcagataacttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.15 OMeY-TGA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttgaagtccgccgcgtttagccacttcgctacc cctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtga tttcccagaacacatagcgacatgcaaatattaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttgaagtccgccg cgtttagccacttcgctacccctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatag ggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaatggtgggggaaggattcgaaccttcgaagtctgtgacgg cagatttgaagtctgctccctttggccgctcgggaacccccaccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgataggg aacttataagtctctatcactgatagggattcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgata gtccattttaaaacataatttttaaaactgcaaactacccaagaaattattacttttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaatta attctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggt gccaacaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttgaagtccgccgcgtttagccacttcgctacccctccgg tgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttccca gaacacatagcgacatgcaaatattaaaaaatggaggggggacggattcgaaccgccgaacccaaagggagcggatttgaagtccgccgcgttta gccacttcgctacccctccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggattt cacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaatggtgggggaaggattcgaaccttcgaagtctgtgacggcagatt tgaagtctgctccctttggccgctcgggaacccccaccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaactta*

	taagtctctatcactgatagggattcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccat tttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattcta attatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggcgaagaagttgtccatat tgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcc tattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgac ctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccag atatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactta cggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccatt gacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggta aatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcgg ttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaa atcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctct ctggctaactagagaacccactgcttactggcttatcaccgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcga cggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaacccaagctggctagcgccaccatgg caagcagtaacttgattaaacaattgcaagagcgggggctggtagcccaggtgacggacgaggaagcgttagcagagcgactggcgcaaggccc gatcgcactcgtgtgtggcttcgatcctaccgctgacagcttgcatttggggcatcttgttccattgttatgcctgaaacgcttccagcaggcgggccac aagccggttgcgctggtaggcggcgcgacgggtctgattggcgacccgagcttcaaagctgccgagcgtaagctgaacaccgaagaaactgttcag gagtgggtggacaaaatccgtaagcaggttgccccgttcctcgatttcgactgtggagaaaactctgctatcgcggccaataattatgactggttcgg caatatgaatgtgctgaccttcctgcgcgatattggcaaacacttctccgttaaccagatgatcaacaaagaagcggttaagcagcgtctcaaccgtg aagatcaggggatttcgttcactgagttttcctacaacctgctgcagggttatagtatggcctgtttgaacaaacagtacggtgtggtgctgcaaattg gtggttctgaccagtggggtaacatcacttctggtatcgacctgacccgtcgtctgcatcagaatcaggtgtttggcctgaccgttccgctgatcactaa agcagatggcaccaaatttggtaaaactgaaggcggcgcagtctggttggatccgaagaaaaccagcccgtacaaattctaccagttctggatcaa cactgcgcgtgccgacgtttaccgcttcctgaagttcttcacctttatgagcattgaagagatcaacgccctggaagaagaagataaaaacagcggt aaagcaccgcgcgcccagtatgtactggcggagcaggtgactcgtctggttcacggtgaagaaggtttacaggcggcaaaacgtattaccgaatgc ctgttcagcggttctttgagtgcgctgagtgaagcggacttcgaacagctggcgcaggacggcgtaccgatggttgagatggaaaagggcgcagac ctgatgcaggcactggtcgattctgaactgcaacctttcccgtggtcaggcacgtaaaactatcgcctccaatgccatcaccattaacggtgaaaaaca gtccgatcctgaatacttctttaaagaagaagatcgtctgtttggtcgttttaccttactgcgtcgcggtaaaaagaattactgtctgatttgctggaaag ggcccgtttaatagtgaaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagct taagtttaaaccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcc cactgtccttccctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggg aggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccc ggtaatacggttatccacagaatcaggggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttt taaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcca tagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccg gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgc cgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggctt cattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaa gtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtact caaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaa agtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccaactg atcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatg ttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaaca aatagggttccgcgcacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacg tgagttacgcgtcgttccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgc gccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttc ccttcctttctcgccacgttcgccagctttccccgtcaagctctaaatcgggggctcccctttagggttccgatttagtgctttacggcacctcgaccccaaa aaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactctt gttccaaactggaacaacactcaaccctatctctgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattta acaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgc atctcaattagtcagcaaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagt cccgccccttaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggc cgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcgga tctgatcagcacgtgatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgc cgactaccccgccacgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgac atcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcg gcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggt tcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccgggg tgcccgccttcctggagacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgc gcacctggtgcatgacccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgtttt ccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataa agcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccg tcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgt gcctggcagataacttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.16 MbPyl-TAA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgaaaaaacgggggtgcgggagtttcacccgcctgcacggatttaaagtccgtgtgatctattcgatccacccccggtgtctctatc actgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacata gcgacatgcaaatattaaaaaacgggggtgcgggagtttcacccgcctgcacggatttaaagtccgtgtgatctattcgatccacccccggtgtctct ctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaac*

acatagcgacatgcaaatattaaaaaacggaaaccccgggaatctaacccggctgaacggatttaaagtccgttcgatctacatgatcaggtttccg
gtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggacgtcctttccacaag
atatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataatttttaaaactgcaaactacccaagaa
attattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatg
aaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggtgccaacaaaaaacgggggggtgcgggagtttcacccgcctgca
cggatttaaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc
tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacgggggggtgcgggagtttcacccgcc
tgcacggatttaaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttata
agtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacggaaaccccgggaatctaacc
cggctgaacggatttaaagtccgttcgatctacatgatcaggtttccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgat
agggaacttataagtctctatcactgatagggacgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatg
atagtccatttttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaa
attaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggggcgaagaa
gttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataat
cttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggt
ctagagcgacctactccggaatattaatagatgtacgggccgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatg
tacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgt
tacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatag
ggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgt
caatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattacc
atggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttg
ttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatat
aagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcga
attctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaatggataaaaaac
cattagatgttttaatatctgcgaccgggctctggatgtccaggactggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatac
attgaaatggcgtgtggagaccatcttgttgtgaataattccaggagttgtagaacagccagagcattcagacatcataagtacagaaaaacctgca
aacgatgtagggtttcggacgaggatatcaataattttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagtttctgctccaaa
ggtcaaaaaagctatgccgaaatcagtttcaagggctccgaagcctctggaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgta
ccttcgcctgcaaaatcaactccaaattcgtctgttcccgcatcggctcctgctccttcacttacaagaagccagcttgatagggttgaggctctcttaag
tccagaggataaaatttctctaaatatggcaaagcctttcagggaacttgagcctgaacttgtgacaagaagaaaaaacgattttcagcggctctat
accaatgatagagaagaagactacctcggtaaactcgaacgtgatattacgaaatttttcgtagaccgggggttttctggagataaagtctcctatccttatt
ccggcggaatacgtggagagaatgggtattaataatgatactgaactttcaaaacagatcttccgggtggataaaaatctctgcttgaggccaatgc
ttgccccgactctttacaactatctgcgaaaactcgataggattttaccaggcccaataaaaattttcgaagtcggaccttgttaccggaaagagtctg
acggcaaagagcacctggaagaatttactatggtgaacttctgtcagatgggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagttt
ctggactatctggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgatataatgcacggggacctggagctttcttcgg
cagtcgtcgggccagtttctcttgatagagaatggggtattgacaaaccatggataggtgcaggttttggtcttgaacgcttgctcaaggttatgcacg
gctttaaaaacattaagagggcatcaaggtccgaatcttactataatgggatttcaaccaatctgtaatagtgaaacccagctttcttgtacaaagtg
gtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctgatcagcctcgactgtgccttctagtt
gccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcatt
gtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgc
ggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccggtaatacggttatccacagaatcaggggatggaatttt
ggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctga
cagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatac
gggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaa
gggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtt
tgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggca
gcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccc
gagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaa
ctctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccaactgatcttcagcatcttttacttcaccagcgtttctgggtga
gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagc
atttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgcca
cctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgt
agaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggt
ggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccagctttccccgtcaa
gctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccat
cgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctctgtct
attcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgt
gtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccc
cagtctcccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgc
ccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgagga
ggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagtacaagcccacg
gtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactacccgccacgcgccacaccgtcgacccgga
ccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcg
gtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttcccggctggcc
gcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggc
aagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctcc
ccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgcctgaca
cgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggga
tctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttt
tttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcata
gctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtataatgtatgctatacgaa
gttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata

UCAP.17 MbPyl-TAG	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc* *tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg* *ttttaagttttataagaaaaaaaagaatatataaggctttaaagctttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag* *aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa* *gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc* *agggattggctgaaaaaacgggggtgcgggagtttcacccgcctgcacggatttagagtccgtgtgatctattcgatccacccccggtgtctctatc* *actgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacata* *gcgacatgcaaatattaaaaaacgggggggtgcgggagtttcacccgcctgcacggatttagagtccgtgtgatctattcgatccacccccggtgtct* *ctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaac* *acatagcgacatgcaaatattaaaaaacggaaacccgggaatctaacccggctgaacggatttagagtccgttcgatctacatgatcaggtttccg* *gtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggacgtcctttccacaag* *atatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataatttttaaaactgcaaactacccaagaa* *attattactttctacgtcacgtatttgtactaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatg* *aaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggtgccaacaaaaaacgggggtgcgggagtttcacccgcctgca* *cggatttagagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc* *tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacgggggtgcgggagtttcacccgcc* *tgcacggatttagagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttata* *agtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacggaaaccccgggaatctaacc* *cggctgaacggatttagagtccgttcgatctacatgatcaggtttccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgat* *agggaacttataagtctctatcactgatagggacgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatg* *atagtccatttttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaa* *attaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggcgaagaa* *gttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataat* *cttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggt* *ctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatg* *tacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgt* *tacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatag* *ggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgt* *caatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattacc* *atggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttg* *ttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatat* *aagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcg a* *attctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaatggataaaaaac* *cattagatgttttaatatctgcgaccgggctctggatgtccaggactggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatac* *attgaaatggcgtgtggagaccatcttgttgtgaataattccaggagttgtagaacagccagagcattcagacatcataagtacagaaaaacctgca* *aacgatgtagggtttcggacgaggatatcaatatttttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagtttctgctccaaa* *ggtcaaaaaagctatgccgaaatcagtttcaagggctccgaagcctctggaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgta* *ccttcgcctgcaaaatcaactccaaattcgtctgttcccgcatcggctcctgctccttcacttacaagaagccagcttgatagggttgaggctctcttaag* *tccagaggataaaatttctctaaatatggcaaagcctttcagggaacttgagcctgaacttgtgacaagaagaaaaaacgattttcagcggctctat* *accaatgatagagaagactacctcggtaaactcgaacgtgatattacgaaatttttcgtagaccggggttttctggagataaagtctcctatcctttatt* *ccggcggaatacgtggagagaatgggtattaataatgatactgaactttcaaaacagatcttccgggtggataaaaatctctgcttgaggccaatgc* *ttgccccgactctttacaactatctgcgaaaactcgataggattttaccaggcccaataaaaattttcgaagtcggaccttgttaccggaaagagtctg* *acggcaaagagcacctggaagaatttactatggtgaacttctgtcagatgggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagttt* *ctggactatctggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgatataatgcacggggacctggagctttcttcgg* *cagtcgtcgggccagtttctcttgatagagaatggggtattgacaaaccatggataggtgcaggttttggtcttgaacgcttgctcaaggttatgcacg* *gctttaaaaacattaagagggcatcaaggtccgaatcttactataatgggatttcaaccaatctgtaatagtgaaacccagctttcttgtacaaagtg* *gtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctgatcagcctcgactgtgccttctagtt* *gccagccatctgttgtttgcccctccccgtgccttccttgaccctggaagggtgccactcccactgtccttcctaataaaatgaggaaattgcatcgcatt* *gtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgc* *ggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccgttaatacggttatccacagaatcaggggatgggatttt* *ggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctga* *cagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatac* *gggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaa* *gggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtt* *tgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac* *atgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggca* *gcactgcataattctctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgacc* *gagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaa* *ctctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccccaactgatcttcagcatcttttactttcaccagcgtttctgggtga* *gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagc* *atttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgcca* *cctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgt* *agaaactatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggt* *ggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccagctttccccgtcaa* *gctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccat* *cgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctctgtct* *attcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgt* *gtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccc* *cagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgc* *ccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgagga* *ggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagtacaagcccacg* *gtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgacccgga* *ccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcg*

	gtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttcccggctggcc gcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggc aagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctcc ccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgcctgaca cgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggа tctcatgctggagttcttcgcccaccccaacttgtttattgcagcttatаatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttt tttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcata gctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtataatgtatgctatacgaa gttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.18 MbPyl-TGA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggctttaaagctttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgaaaaaacgggggtgcgggagtttcacccgcctgcacggatttgaagtccgtgtgatctattcgatccacccccggtgtctctatc actgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacata gcgacatgcaaatattaaaaaacgggggtgcgggagtttcacccgcctgcacggatttgaagtccgtgtgatctattcgatccacccccggtgtct ctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaac acatagcgacatgcaaatattaaaaaacggaaacccgggaatctaacccggctgaacggatttgaagtccgttcgatctacatgatcaggtttccg gtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggacgtcctttccacaag atatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataattttaaaactgcaaactacccaagaa attattactttctacgtcacgtatttgtactaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatg aaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggtgccaacaaaaaacgggggtgcgggagtttcacccgcctgca cggatttgaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacgggggtgcgggagtttcacccgcc tgcacggatttgaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttata agtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacggaaacccgggaatctaacc cggctgaacggatttgaagtccgttcgatctacatgatcaggtttccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggacgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatg atagtccattttaaaacataattttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaa attaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggcgaagaa gttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataat cttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggt ctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatg tacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgt tacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatag ggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgt caatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattacc atggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttg ttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatat aagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcga attctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttaatggataaaaaac cattagatgttttaatatctgcgaccgggctctggatgtccaggactggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatac attgaaatggcgtgtggagaccatcttgttgtgaataattccaggagttgtagaacagccagagcattcagacatcataagtacagaaaaacctgca aacgatgtagggtttcggacgaggatatcaataattttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagtttctgctccaaa ggtcaaaaaagctatgccgaaatcagtttcaagggctccgaagcctctggaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgta ccttcgcctgcaaaatcaactccaaattcgtctgttcccgcatcggctcctgctccttcacttacaagaagccagcttgatagggttgaggctctcttaag tccagaggataaaatttctctaaatatggcaaagcctttcagggaacttgagcctgaacttgtgacaagaagaaaaaacgattttcagcggctctat accaatgatagagaagactacctcggtaaactcgaacgtgatattacgaaatttttcgtagaccggggttttctggagataaagtctcctatccttatt ccggcggaatacgtggagagaatgggtattaataatgatactgaactttcaaaacagatcttccgggtggataaaaatctctgcttgaggccaatgc ttgccccgactctttacaactatctgcgaaaactcgataggattttaccaggcccaataaaaattttcgaagtcggaccttgttaccggaaagagtctg acggcaaagagcacctggaagaatttactatggtgaacttctgtcagatgggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagttt ctggactatctggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgatataatgcacggggacctggagctttcttcgg cagtcgtcgggccagtttctcttgatagagaatggggtattgacaaaccatggataggtgcaggttttggtcttgaacgcttgctcaaggttatgcacg gctttaaaaacattaagagggcatcaaggtccgaatcttactataatgggatttcaaccaatctgtaatagtgaaacccagctttcttgtacaaagtg gtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaaccgctgatcagcctcgactgtgccttctagtt gccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtccttctaatgaaaatgaggaaattgcatcgcatt gtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaaggggaggattgggaagacaatagcaggcatgctggggatgc ggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccggtaatacggttatccacagaatcaggggatgggatttt ggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctga cagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatac gggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaa gggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtt tgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac atgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggca gcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccc gagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaa ctctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtga gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagc atttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgcca cctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgt agaaactatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggt*

	ggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccagctttccccgtcaa gctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccat cgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctctgtct attctttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgt gtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccc cagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgc ccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgagga ggcttttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagtacaagcccacg gtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgacccgga ccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcg gtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttcccggctggcc gcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggc aagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccgggtgcccgccttcctggagacctccgcgccccgcaacctcc ccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgcctgaca cgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcgggga tctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttt tttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcata gctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtataatgtatgctatacgaa gttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.20 MmPyl-TAA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgaaaaaacgggggtgcgggagtttcacccgcctgcacggatttaaagtccgtgtgatctattcgatccacccccggtgtctctatc actgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacata gcgacatgcaaatattaaaaaacgggggtgcgggagtttcacccgcctgcacggatttaaagtccgtgtgatctattcgatccacccccggtgtct ctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaac acatagcgacatgcaaatattaaaaaacggaaacccgggaatctaacccggctgaacggatttaaagtccgttcgatctacatgatcaggtttccg gtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggacgtccttccacaag atatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataatttttaaaactgcaaactacccaagaa attattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatg aaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggtgccaacaaaaaacgggggtgcgggagtttcacccgcctgca cggatttaaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacgggggtgcgggagtttcacccgcc tgcacggatttaaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttata agtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacggaaacccgggaatctaacc cggctgaacggatttaaagtccgttcgatctacatgatcaggtttccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggacgtccttccacaagatatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatg atagtccattttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaa attaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggcgaagaa gttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataat cttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggt ctagagcgacctactccggaatattaatagatgtacgggcgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatg tacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgt tacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatag ggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgt caatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattacc atggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttg ttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatat aagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcga attctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttagccgccaccatggaa taaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacaccacgaagtctctcgaag caaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggactgcaagagcgctcaggcaccacaaatacа ggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcacaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtc gtttctgcccctaccagaacgaaaaaggcaatgccaaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaacct tctggatctaaatttcacctgcgataccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagc aactgcatccgcactggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgccccgcacttacgaagagccagactg acaggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctctctcgcagaaaa aaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacagggggttttctg gaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaa gaacttctgcctgagacccatgcttgctccaaaccttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaatagg cccatgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaa atcttgaaagcataattacggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctttggggatacccttgatgtaat gcacggagacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttcggg ctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggatttctaccaacctgtaatg atagaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaacc gctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtccttcc taataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggа agacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatcccggtaatacggt tatccacagaatcaggggatgggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatct aaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctga*

	ctccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcgtcgttccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaacccctatctcgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacgtgatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataacttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.21 MmPyl-TAG	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagctctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaaggttttaagttttataagaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgagaagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaagggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcacccagggattggctgaaaaaacgggggtgcgggagtttcacccgcctgcacggatttagagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacgggggtgcgggagtttcacccgcctgcacggatttagagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacggaaaccccgggaatctaacccggctgaacggatttagagtccgttcgatctacatgatcaggtttccggtgtttcggtgtctctatcactgatagggacgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataatttttaaaactgcaaactacccaagaaattattacttttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggggcgaagaagttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataatcttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcaccgcagggctagcgctaccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttagccgccaccatggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaatataaaacaccacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggactgcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcacaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaacctctggatctaaattttcacctgcgataccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactgacaggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctctctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacagggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaa*

	gaacttctgcctgagacccatgcttgctccaaaccttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaatagg cccatgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaa atcttgaaagcataattacggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctttggggatacccttgatgtaat gcacggagacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttcggg ctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggatttctaccaacctgtaatg atagaaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaacc gctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcc taataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggа agacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccggtaatacggt tatccacagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatct aaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctga ctccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagattt atcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagcta gagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctcc ggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggcc gcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtc attctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatca ttggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatc ttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcat actcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttc cgcgcacatttcccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcg tcgttccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcg gcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcg ccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattag ggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttgacgttggagtccacgttctttaatagtggactcttgttccaaactgg aacaacactcaacccctatctcgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaa cgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtc agcaaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgccctaact ccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctg agctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacg tgatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggct ggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata
UCAP.22 MmPyl-TGA	*gtacgtactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaaccctcatggctaacgtactaagc tctcatggctaacgtactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaaattaatataaatcagcaacttaaatagcctctaagg ttttaagttttataagaaaaaaaagaatatataaggcttttaaagcttttaaggtttaacggttgtggacaacaagccagggatgtaacgcactgag aagcccttagagcctctcaaagcaattttgagtgacacaggaacacttaacggctgacagaattagcttcacgctgccgcaagcactcagggcgcaa gggctgctccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcatagcaactttgtatagaaaagttgggtggtgaaactcaccc agggattggctgaaaaaacgggggggtgcgggagtttcacccgcctgcacggatttgaagtccgtgtgatctattcgatccacccccggtgtctctatc actgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaacacata gcgacatgcaaatattaaaaaacgggggggtgcgggagtttcacccgcctgcacggatttgaagtccgtgtgatctattcgatccacccccggtgtct ctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggatttcacgtttatggtgatttcccagaac acatagcgacatgcaaatattaaaaaacggaaaccccgggaatctaacccggctgaacggatttgaagtccgttcgatctacatgatcaggtttccg gtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtctctatcactgatagggacgtcctttccacaag atatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataatttttaaaactgcaaactacccaagaa attattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatg aaggaatcatgggaaataggccctcttcctgcccgatataatatccgcccatggtgccaacaaaaaacgggggggtgcgggagtttcacccgcctgca cggatttgaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttataagtc tctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacgggggggtgcgggagtttcacccgcc tgcacggatttgaagtccgtgtgatctattcgatccacccccggtgtctctatcactgatagggaacttataagtctctatcactgatagggaacttata agtctctatcactgatagggatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattaaaaaacggaaacccgggaatctaacc cggctgaacggatttgaagtccgttcgatctacatgatcaggtttccggtgtttcggtgtctctatcactgatagggaacttataagtctctatcactgat agggaacttataagtctctatcactgatagggacgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatg atagtccattttaaaacataatttttaaaactgcaaactacccaagaaattattacttttctacgtcacgtattttgtactaatatctttgtgtttacagtcaa attaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcctgcccgaggggggcgaagaa gttgtccatattgcacaacttttgtatacaaagttgtttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataat cttgaagttcctattccgaagttcctattctctagaaagtataggaacttcagagcgcttttgaagctagaactagtggatccgcgccaaaaccgcggt ctagagcgacctactccggaatattaatagatgtacgggccgtggatccgcgccaaaaccgcggtctagagcgacctactccggaatattaatagatg tacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgt tacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatag ggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccctattgacgt caatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattacc atggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttg ttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatat*

	aagcagagctctctggctaactagagaacccactgcttactggcttatcacctgcagggctagcgctaccggactcagatctcgagctcaagcttcga attctgcagtcgacggtaccgcgggcccgggatccaccggtatctattagaaaatcacaagtttgtacaaaaaagcaggcttagccgccaccatgga taaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaatataaaacaccacgaagtctctcgaag caaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggactgcaagagcgctcaggcaccacaaatacа ggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcacaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtc gtttctgcccctaccagaacgaaaaaggcaatgccaaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaacct tctggatctaaattttcacctgcgataccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagc aactgcatccgcactggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgccccgcacttacgaagagccagactg acaggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctctctcgcagaaaa aaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacagggggttttctg gaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaa gaacttctgcctgagacccatgcttgctccaaaccttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaatagg cccatgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacagggaaa atcttgaaagcataattacggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctttggggatacccttgatgtaat gcacggagacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggatagggggcaggtttcggg ctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggatttctaccaacctgtaatg atagaacccagctttcttgtacaaagtggtgatatctgcagaattccaccacactggactagtggatccgagctcggtaccaagcttaagtttaaacc gctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtccttcc taataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattggga agacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccggtaatacggt tatccacagaatcaggggatgggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatct aaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctga ctccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagattt atcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagcta gagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctcc ggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggcc gcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtc attctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatca ttggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatc ttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcat actcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttc cgcgcacatttccccgaaaagtgccacctgacgtctatacgctgacttgacgggacggcgcaagctcatgaccaaaatcccttaacgtgagttacgcg tcgttccactgagcgtcagaccccgtagaaactatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcg gcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcg ccacgttcgccagctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattag ggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactgg aacaacactcaaccctatctcgtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaa cgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtc agcaaccaggtgtggaaagtccccagtctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag caaccatagtcccgcccctaact ccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctg agctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctcgcgggagcttgtatatccattttcggatctgatcagcacg tgatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgcc acgcgccacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcactcgcgtcgggctcgacatcggcaaggtgt gggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcggggcggtgttcgccgagatcggcccgcgcatggcc gagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggtgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcg gcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctgg agacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatg acccgcaagcccggtgcctgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggc tggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatca caaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta gagcttggcgtaatcatggtcatagctgtttcctgtgtgaactgcaggtcggtaggtcgcgaatcgatactagtaaaacccatgtgcctggcagataa cttcgtataatgtatgctatacgaagttatggtaccgcggccgcgtagaggatctgttgatcagcagttcaacctgttgata

F. eGFP Gene Sequence for G.E.M.S. Plasmids

G.E.M.S. Plasmid	Mutation	eGFP Gene Sequence*
UTX1.01	T9TAA	*gtgagcaagggcgaggagctgttcTAAggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.02	T9TAG	*gtgagcaagggcgaggagctgttcTAGggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.03	T9TGA	*gtgagcaagggcgaggagctgttcTGAggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.04	Y39TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccaccTAAggcaagctgacccttaagttcatctgcaccaccggcaagctgccc gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgac ttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgc cgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct ggggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtga acttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacg gccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtca ccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttc aactagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.05	Y39TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccaccTAGggcaagctgacccttaagttcatctgcaccaccggcaagctgccc gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgac ttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgc cgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct ggggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtga acttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacg gccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtca ccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttc aactagtggttctggtgactacaaggacgacgacgacaagctttag*

* Only eGFP sequence has point mutations; backbone is common (*Plasmid UTX0*).

UTX1.06	Y39TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccaccTGAggcaagctgacccttaagttcatctgcaccaccggcaagctgccc gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgac ttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgc cgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct ggggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtga acttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacg gccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtca ccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttc aactagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.07	K52TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcTAActgcccg tgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgaccggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.08	K52TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcTAGctgcccg tgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgaccggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.09	K52TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcTGActgcccg tgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgaccggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.10	F99TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcTAAttcaaggacgacggcaactacaagacccgcgcc gaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.11	F99TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcTAGttcaaggacgacggcaactacaagacccgcgcc gaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt*

		cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag
UTX1.12	F99TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcTGAttcaaggacgacggcaactacaagacccgcgcc gaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg gggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.13	D117TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcTAAaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg gggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.14	D117TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcTAGaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg gggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.15	D117TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcTGAaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg gggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.16	E132TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaagTAAgacggcaacatcctg gggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.17	E132TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg*

		aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaagTAGgacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag
UTX1.18	E132TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaagTGAgacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.19	Y143TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagTAAaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.20	Y143TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagTAGaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.21	Y143TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagTGAaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.22	Y151TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtcTAAatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*

UTX1.23	Y151TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctacctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtcTAGatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.24	Y151TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctacctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtcTGAatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.25	Q157TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctacctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagTAAaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.26	Q157TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctacctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagTAGaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.27	Q157TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctacctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagTGAaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.28	V176TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctacctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcTAAcagctcgccgaccactaccagcagaacaccccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt*

		cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag
UTX1.29	V176TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcTAGcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.30	V176TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcTGAcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.31	Y182TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccacTAAcagcagaacaccccatcggcgacggc cccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatgg tcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacc accaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttca actagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.32	Y182TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccacTAGcagcagaacaccccatcggcgacggc cccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatgg tcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacc accaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttca actagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.33	Y182TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccacTGAcagcagaacaccccatcggcgacggc cccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatgg tcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacc accaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttca actagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.34	D190TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg*

		aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcTAAggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag
UTX1.35	D190TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcTAGggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.36	D190TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcTGAggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.37	Q204TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcaccTAAtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.38	Q204TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcaccTAGtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.39	Q204TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcaccTGAtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*

UTX1.40	N212TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccTAAgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.41	N212TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccTAGgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.42	N212TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccTGAgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.43	N214TAA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagTAAcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.44	N214TAG	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagTAGcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX1.45	N214TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagTGAcgcgatcacatggt*

		cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag
UTX2.23	Y39TAG-K52TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccaccTAGggcaagctgacccttaagttcatctgcaccaccggcTGActgccc gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgac ttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgc cgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct ggggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtga acttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacg gccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtca ccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttc aactagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.24	Y39TAG-F99TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccaccTAGggcaagctgacccttaagttcatctgcaccaccggcaagctgccc gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgac ttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcTGAttcaaggacgacggcaactacaagacccgcg ccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcc tggggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtga acttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacg gccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtca ccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttc aactagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.25	Y39TAG-V176TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccaccTAGggcaagctgacccttaagttcatctgcaccaccggcaagctgccc gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgac ttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgc cgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct ggggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtga acttcaagattaggcacaacatcgaggacggcagcTGAcagctcgccgaccactaccagcagaacaccccatcggcgacg gccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtca ccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttc aactagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.26	Y39TAG-Q204TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccaccTAGggcaagctgacccttaagttcatctgcaccaccggcaagctgccc gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgac ttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgc cgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct ggggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtga acttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacg gccccgtgctgctgcccgacaaccactacctgagcaccTGAtccgccctgagcaaagaccccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtca ccaccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttc aactagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.34	K52TAG-F99TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcTAGctgcccg tgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcTGAttcaaggacgacggcaactacaagacccgcgcc gaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.35	K52TAG-V176TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcTAGctgcccg tgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg*

		aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcTGAcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag
UTX2.36	K52TAG-Q204TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcTAGctgcccg tgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcaccTGAtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.45	F99TAG-V176TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcTAGttcaaggacgacggcaactacaagacccgcgcc gaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcTGAcagctcgccgaccactaccagcagaacaccccatcggcgacggc cccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatgg tcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacc accaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttca actagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.46	F99TAG-Q204TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcTAGttcaaggacgacggcaactacaagacccgcgcc gaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaac ttcaagattaggcacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcaccTGAtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*
UTX2.56	V176TAG-Q204TGA	*gtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacccttaagttcatctgcaccaccggcaagctgcccgt gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctttagccgctaccctgaccacatgaagcagcacgactt cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccg aggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg ggcacaagctggagtacaactataacagccacaacgtctatatcatggctgacaagcagaagaacggcatcaaggtgaact tcaagattaggcacaacatcgaggacggcagcTAGcagctcgccgaccactaccagcagaacaccccatcggcgacggcc ccgtgctgctgcccgacaaccactacctgagcaccTGAtccgccctgagcaaagaccccaacgagaagcgcgatcacatggt cctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggttcaactagtgggtctggtcacca ccaccatcaccatcaccaccatcacggttcaactagtggctctggtgagcagaagctgatctccgaggaggacctgggttcaa ctagtggttctggtgactacaaggacgacgacgacaagctttag*