Conserved Peptide upstream Open Reading Frames (CPuORFs): A mechanism to modulate translation in plants.

UNIVERSITY OF LEEDS

Zachary James Paling

Submitted in accordance with the requirements for the degree of PhD.

The University of Leeds

School of Biology

March 2024

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The fifth chapter of this thesis is based on work from the jointly authored publication:

Causier, B., Hopes, T., McKay, M., **Paling, Z.** & Davies, B. (2022) Plants utilise ancient conserved peptide upstream open reading frames in stress-responsive translational regulation. *Plant, Cell & Environment*, 45, 1229–1241. https://doi.org/10.1111/pce.14277. The PhD candidate created an Arabidopsis CPuORF database in this publication. All experimental work was performed by Causier, B., Hopes, T., and McKay, M. Experimental design and the manuscript is the joint work of Causier, B., and Davies, B.

Acknowledgements.

I would like to thank Prof Brendan Davies for his supervision and support since 2019. Thanks is owed to all members of the Davies lab who have a been a joy to work with. Specifically, Tas Spurgeon for assisting me with cloning and DNA extractions, Lin Rong for managing the lab and constantly providing me with confidence, Esther Lewis for guiding me through my PhD and Barry Causier for practical support. Special thanks to my co-supervisors Juan Fontana and Julie Aspden and their groups for their advice and support for chapter 4. I would also like to thank Louie Aspinall from the Astbury centre for structural biology for assisting me on electron microscopes.

Many thanks to Lord David Sainsbury and his charity, The Gatsby Charitable Foundation for funding and supporting me throughout my PhD. I will look back fondly on the training weekends and network meetings with everyone at Gatsby and mentors.

This page is too small to express my gratitude and love for Ryan Jones for his support on this journey. He lifted me up when I couldn't myself, so for that I am eternally thankful. Lastly, I would like to thank my parents Tim and Katherine Paling for their unconditional love and support. When I think of them, I think of the words of Isaac Newton.

"If I have seen further, it is by standing on the shoulders of Giants". – Isaac Newton.

Abstract

Several CPuORFs (Conserved Peptide upstream Open Reading Frames) have been shown to modulate downstream mORF (main Open Reading Frame) translation by conditionally stalling the ribosome. This thesis presents and tests a model for CPuORF function called Conditional uORF Translational Stalling (CUTS). Some CPuORFs stall the ribosome in response to a signal (repressive CUTS). The CUTS mechanism is a rapid way to modulate mORF protein levels during translation. Moreover, another subset of CPuORFs, alongside a signal, can abolish ribosome stalling and promote translation re-initiation and increase downstream mORF levels (activator CUTS). Despite partial elucidation of an *Arabidopsis* a rCUTS CPuORF through cryo-EM, the aCUTS mechanism is yet to be elucidated.

Employing luciferase assays, cryo-EM, and bioinformatics, this research identifies the functional CPuORFs in a rare multi-CPuORF transcript. Moreover, reporter assays provide evidence that the ribosome stalling mechanism of four aCUTS CPuORFs is abolished by the same single base substitutions of ribosome exit tunnel proteins (uL4z, uL16z and RACK1z). Consequently, suggesting that there is a common mechanistic framework for the ribosome stalling in aCUTS CPuORFs. Molecular data also emphasises the role of the full length CPuORF nascent peptide in aCUTS activity. Furthermore, I present the first low-resolution 3D model of an *Arabidopsis* ribosome and have developed protocols to isolate wheat germ CPuORF-stalled ribosomes. Additionally, the thesis expands a published CPuORF database and presents the eukaryotic CPuORF database (https://theeukaryoticcpuorfdatabase.github.io). Analysis of this database suggested

that CPuORFs across taxa have been conserved in regulatory transcripts and suggesting that the CPuORF CUTS mechanism fine tunes development, signalling and metabolism in response to the changing environment.

This research contributes to the broader knowledge of translational regulation in plants, highlighting the aCUTS ribosome stalling mechanism, evolutionary conservation of CPuORFs and their mORFs, and the potential of the CPuORF CUTS peptide-switch mechanism in biotechnology.

Table of contents

List of tables

List of illustrative material

List of abbreviations:

oC: Degree Celsius

Å: Angstrom

Δ: Deletion

µ: Micro

5' cap: 5'-m^7GpppN-cap

40S: small eukaryotic ribosome subunit

43S PIC: small subunit bound to initiation factors and methionine-TC

60S: large eukaryotic ribosome subunit

70S: bacterial or chroloplastic ribosome

80S: eukaryotic ribosome

ABA: Abscisic acid

aCUTS: activator CUTS

ACL5: ACAULIS 5

AdoMet: *S*-Adenosyl methionine

AdoMetDC: *S*-Adenosyl methionine decarboxylase

A. thaliana: Arabidopsis thaliana

Agrobacterium: Agrobacterium tumefaciens

bp: Base pair

CaMV: Cauliflower Mosaic Virus

CCA: cytosine-cytosine-adenosine

cDNA: complementary DNA

CK2: Casein Kinase 2

CPuORF: Conserved peptide uORF

Cryo-EM: Cryo-Electron Microscopy

CUTS: Conditional uORF Translational Stalling

DMSO: Dimethyl sulfoxide

DNA: Deoxyribonucleic acid

dNTPs: Deoxyribonucleotides triphosphates

Drosophila: Drosophila melanogaster

E. coli: Escherichia coli

eIF: eukaryotic initiation factor

eRF: eukaryotic release factor

FS: frameshift

g: Gram

GFP: Green fluorescence protein

GO: Gene ontology

GUS: β-glucuronidase

H. sapiens: Homo sapiens

ID: Identifier

IRES: Internal Ribosome Entry Sites

kDa: kilo Dalton

LB: Lysogeny Broth or Luria-Bertani broth

LHY: LONESOME HIGHWAY

LUC: Luciferase

M: Moles per decimetre (litre)

Met_i: Initiator methionine

mORF: main ORF

mRNA: Messenger RNA

MYA: Million years ago

NMD: Nonsense-mediated mRNA decay

nt: Nucleotide

ORF: Open reading frame

PABP: polyA-binding protein

PCR: Polymerase chain reaction

PHO2: Phosphate over accumulator 2

PIC: Pre-initiation Complex

PTC: peptidyl transferase center

RACK1: RECEPTOR FOR ACTIVATED C KINASE 1

rCUTS: Repressive CUTS

RISP: Re-initiation support protein

RP: ribosomal protein

RNA: Ribonucleic acid

rRNA: ribosomal RNA

SAC51: Suppressor of auculis 5 1

S6K1: Ribosomal protein S6 kinase beta-1

TAV: Trans-activator protein

TC: Tertiary complex

TOR: TARGET OF RAPAMYCIN

tRNA: transfer RNA

TRAP: Translation ribosome affinity purification

TURBS: Termination upstream ribosomal binding site

uORF: upstream ORF

mORF: main ORF

uL4z: ribosomal protein 4 z

uL16z: ribosomal protein 16 z

v/v: Volume per volume

w/v: Weight per volume

WT: Wild type

1. General introduction

Organisms have evolved regulatory mechanisms to control gene expression (1,2). Cellular and environmental signals are integrated into regulatory mechanisms to modify and fine-tune an organism's metabolism and development (2). Furthermore, gene regulation in response to environmental cues can confer a fitness advantage (2). This thesis will explore a mechanism that conditionally regulates gene expression at the translational level (3).

In 2007 Hayden and Jorgensen created a bioinformatic pipeline to compare cDNA sequences between *Arabidopsis* and rice (4). Their pipeline identified a rare subset of upstream Open Reading Frames (uORFs) that are translatable sequences found in the 5' leader sequence. The 5' leader is upstream of the main Open Reading Frame (mORF) that is the main protein product of a gene. The pipeline created by Hayden and Jorgensen identified uORFs that have been conserved at the peptide level over large evolutionary distances (4). These uORF peptides became known as Conserved Peptide upstream Open Reading Frames (CPuORFs).

A recent pipeline designed to identify plant CPuORFs found *Arabidopsis* CPuORFs that are found in moss and animal genomes. This finding potentially suggests that some CPuORFs may have an ancient origin and be of functional significance. To understand the depth of CPuORF conservation I will describe how bioinformatic pipelines extract CPuORF sequences.

Firstly, the pipelines extract uORF and mORF sequences from a transcript sequence dataset (4). The uORFs that are extracted are those that have their start and stop codons in-frame with their downstream mORF (4). Some uORFs likely form part of the final protein product and do not serve a regulatory function. To eliminate these from the dataset and identify regulatory uORFs, these pipelines calculated the uORF-mORF fusion ratio (4). This ratio is a way to calculate how likely a uORF protein will fuse to the mORF post translation. The ratio is calculated from Riboseq data of orthologous transcripts (4). Riboseq data can be used to map the locations of ribosomes on mRNA transcripts, thereby providing insights into translational activity (4). The pipelines discard uORFs above a certain threshold that are likely to have a structural relationship with the mORF protein rather than a regulatory one.

CPuORF identification pipelines then search for amino acid sequence similarities between these uORFs across different species to find sequences that have a functional role (4). This is called a homology search and results in a list of conserved uORF amino acid sequences. To confirm homology, the pipelines select for transcripts that have mORF peptide conservation in at least two other taxonomic orders (4). This step is to filter out false positives that might appear conserved due to chance or limited taxonomic scope.

The next step is to select for uORFs with peptide sequences that are undergoing purifying selection (4). This is calculated from a K_a/K_s ratio analysis of uORF amino acid sequences across taxa (4). The K_a/K_s ratio analysis measures the rate of mutations that affect the protein sequence versus those that do not affect it. A ratio less than 1 suggests that

mutations detrimental to protein function are being selectively removed, indicating a conservation of protein function across different species. The pipeline selects uORFs that have a ratio of less than 0.5 indicating that changes in amino acid sequences are being selected against, this is because they are likely negatively affecting protein function. Consequently, the pipelines have a list of conserved uORF peptides with peptide sequences that are being preserved likely because they have a functional constraint. The final step is to determine the taxonomic range of sequence conservation of these uORFs that are now characterised as CPuORFs (4). The output of this step are FASTA alignments of CPuORFs.

Surprisingly, certain *Arabidopsis* CPuORFs had homologues in species that share a last common ancestor with *Arabidopsis* over 1 billion years ago (4). For example, the CPuORF homology group 3 (HG 3) (a representative being AT3G02470) in *Arabidopsis* is found in green algae (*Ulva)* (4). HG 3 CPuORF's share a conserved amino acid sequence and an mORF that regulates polyamine biosynthesis and plant development. Polyamines such as thermospermine play critical roles across all living organisms in cell division and apoptosis. Polyamines in plants also function in root growth, flower development and stress response. Furthermore, HG 8 CPuORFs are found in species that share an ancient last common ancestor and are found upstream of the same mORF (4). HG 8 CPuORF peptide sequences are found in moss (*Physcomitrella patens*) and green algae (*Mesostigma viride*) and share the signalling mORF, macrophage inhibitory cytokine-1-like (Mic 1-like) proteins (4). Collectively, this shows that some CPuORFs may have a long evolutionary history and may suggest they are under a functional constraint.

Not only are plant CPuORF peptide sequences highly conserved but they are also rare. Bioinformatic pipelines have identified a total of *133 Arabidopsis* transcripts that contain at least one CPuORF (3). This is surprisingly small when compared to more than 13,000 transcripts that contain regular uORFs (5). For clarity, this thesis will term non-CPuORFs as regular uORFs. The key distinction being uORFs are not conserved at the peptide level where CPuORFs are. Collectively, considering the extent of their conservation and rarity, this suggests that CPuORFs are under a functional constraint.

Since their identification several plant CPuORFs have been investigated *in vitro* and/or *in vivo* to reveal that some CPuORFs conditionally modulate mORF expression (3). Transient expression assays, western blotting and ribosome profiling data suggest that CPuORFs can stall the ribosome during translation (3,6,7). Furthermore, 25 CPuORFs have been shown to modulate mORF expression conditionally in response to small metabolites, pathogen infection and abiotic stimuli (3). CPuORFs can therefore be considered to be regulatory peptide switches that modulate mORF translation in response to the external and cellular environment. However, their molecular mechanism is yet to be elucidated.

Currently, 25 CPuORFs have been functionally characterized and this has revealed three distinct modes of function (3). These distinct modes form the basis of the model for CPuORF function put forward by this thesis. The model is called Conditional uORF Translational Stalling (CUTS) (Figure 1.1). The first mode of CUTS is called the default, and this represents CPuORFs that can innately stall the ribosome without the aid of a signal. The second mode is repressive CUTS (rCUTS), these CPuORFs stall the ribosome, or stalling is enhanced, in response to a signal. 12 CPuORFs have been characterized to

have rCUTS activity. The third mode is activator CUTS (aCUTS). A CPuORF with aCUTS activity sees an increase in mORF translation upon signal application. Currently, 14 CPuORFs have been demonstrated to function via aCUTS. Interestingly, one CPuORF has been demonstrated to function through a and rCUTS in response to different signals. In this case, upon pathogen infection heat shock the ribosome is released from the nascent CPuORF peptide whereas, glycogen enhances stalling (3).

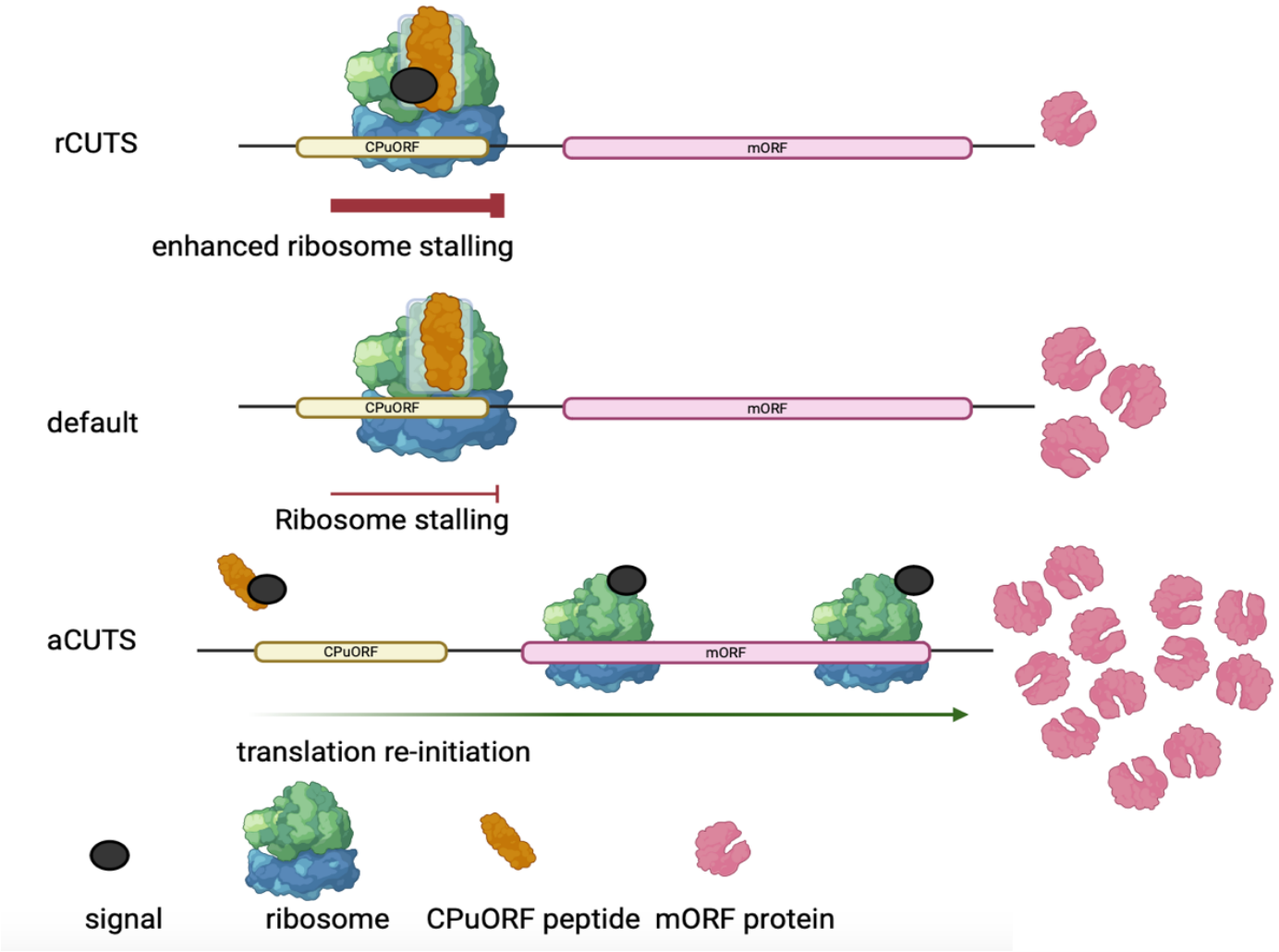

rCUTS
CPuORF
mORF
enhanced ribosome stalling
default
CPuORF
mORF
Ribosome stalling
aCUTS
CPuORF
mORF
translation re-initiation
signal
ribosome
CPuORF peptide
mORF protein

Figure 1.1. A schematic representation of the CPuORF Conditional uORF Translational Stalling mechanism (CUTS) (3,8–11)**.** The figure depicts three mature mRNA transcripts that contains a Conserved Peptide upstream Open Reading Frame (CPuORF) (pale yellow) and a downstream main Open Reading Frame (mORF) (pink). The ribosome is depicted by a green and blue protein, the CPuORF peptide is depicted as a yellow protein and the mORF protein is depicted as a pink protein. A signal that is specific to a CPuORF is depicted by a black circle. rCUTS (repressive CUTS) is the CPuORF mode of function where a signal enhances or initiates stalling by binding to the CPuORF stalled ribosome. The enhanced stalling results in low levels of mORF protein during rCUTS. The default of translation is where a CPuORF nascent peptide stalls the ribosome during translation without a signal. This results in low levels of mORF protein. Finally, aCUTS (activator CUTS) is the mode of CPuORF function where a signal prevents ribosome stalling and translation re-initiation occurs at the downstream mORF. aCUTS results in elevated levels of mORF protein when compared to the default.

CPuORFs are associated with regulatory mORFs, for example, CPuORFs are often found upstream of transcription factors (4). Furthermore, the conditional response often is associated with the biological function of the mORF. For example, sucrose responsive CPuORFs are found upstream of sucrose metabolic genes (12). Consequently, the conditional CPuORF peptide switch mechanism offers sessile plants a mechanism to fine tune gene regulation and adapt to their changing internal and external environments. The CPuORF mechanism has also been demonstrated to have its applications in crop improvement and has potential in biotechnology and agriculture (3,13). For example, the pathogen responsive CPuORF found in the TBF1 transcript was constitutively expressed in upstream of a pathogen defence gene NPR1 in rice. Transformant rice resulted in larger yields during pathogen infection without a fitness cost (3,13). § To utilise CPuORFs their molecular mechanism must be elucidated through molecular and structural biology.

The overarching aim of this project is to elucidate the CPuORF mechanism through molecular and structural biology. This first chapter will discuss what is known about CPuORF function, translation regulation, and ribosomal stalling.

1.1. Gene regulation and the central dogma of molecular biology.

CPuORFs stall the ribosome during translation, so to understand CPuORF function it is important to discuss the transfer of genetic information from DNA to proteins as detailed by the central dogma of molecular biology (3). The central dogma of molecular biology describes the flow of genetic information within biological systems (Figure 1.2). Genes are segments of DNA that encode the instructions for making a particular protein (1,14,15). DNA encoding a gene must be transcribed from DNA to RNA and then finally translated into a protein (Figures 1.2-3).

Figure 1.2. A schematic representation of the central dogma of molecular biology (15)**.** The diagram illustrates the flow of genetic information from DNA to RNA and to proteins. DNA replicates itself and can be transcribed into RNA. RNA can be reverse transcribed back into DNA. RNA is decoded and synthesised into proteins during translation.

While stored as DNA, a gene is organised into the mORF that is flanked by the 5' leader and 3' trailer sequences. It is in the 5' leader that regular uORFs and CPuORFs are found. Furthermore, upstream of the 5' leader the transcriptional promoter sequence and enhancer and silencer sites can be found (Table 1.3 A). The quality of a promoter can determine the affinity for RNA polymerase binding, thereby affecting the likelihood of the

DNA being transcribed (16). This is further modulated by transcription factors binding to enhancer and silencer sites that either aid or inhibit transcription (16). Also, downstream of the 3' trailer sequence is the terminator and potentially an enhancer or silencer site. At the level of DNA, regulatory mechanisms can control a gene's transcriptional output (RNA) (Figure 1.3).

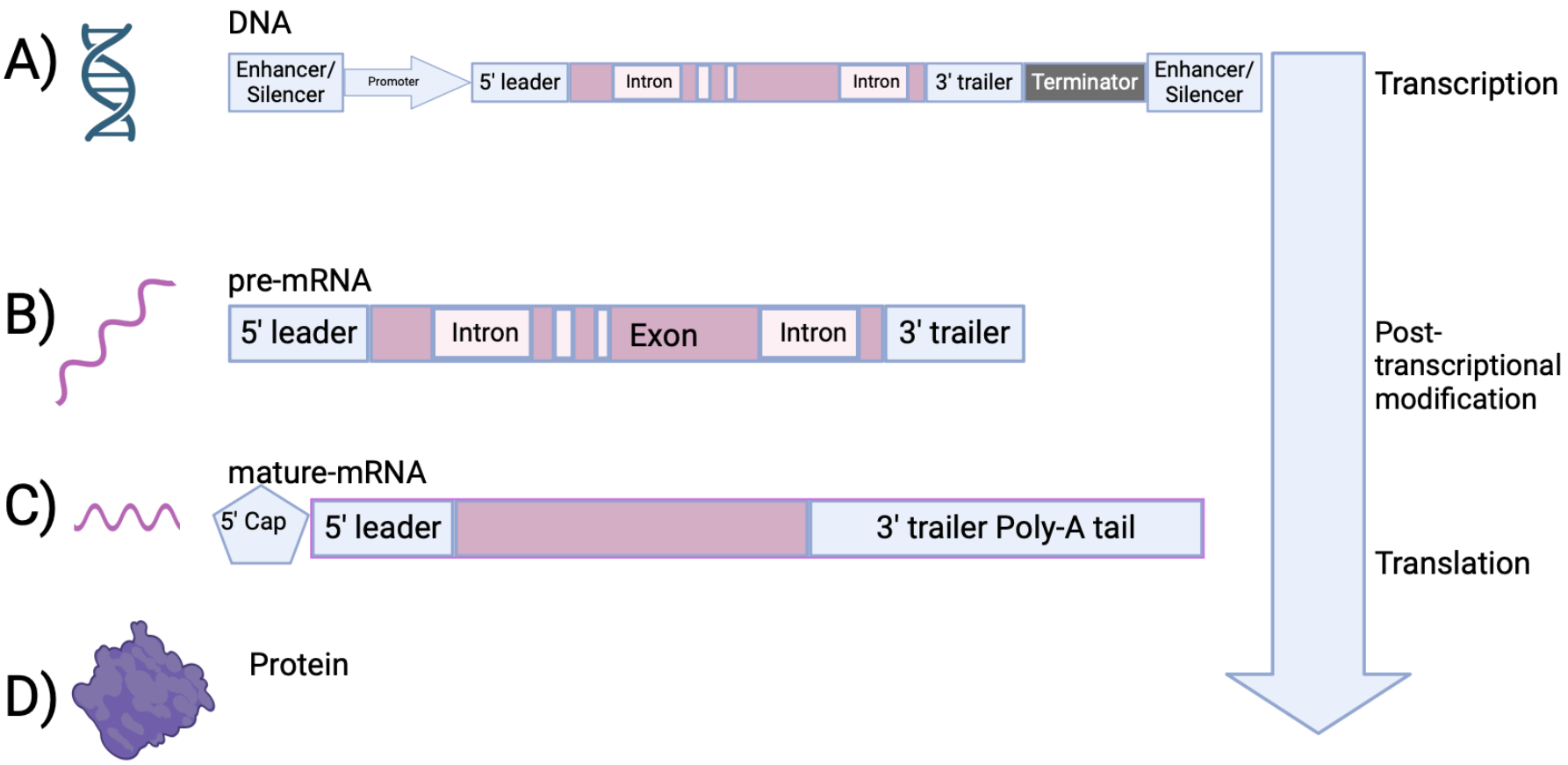

A)
DNA
Enhancer/ Silencer
Promoter
5' leader
Intron
Intron
3' trailer
Terminator
Enhancer/ Silencer
Transcription
B)
pre-mRNA
5' leader
Intron
Exon
Intron
3' trailer
Post-transcriptional modification
C)
mature-mRNA
5' Cap
5' leader
3' trailer Poly-A tail
Translation
D)
Protein

Figure 1.3. A schematic illustrating gene organisation in DNA, RNA, and proteins. The main Open Reading Frame (mORF) encodes for the main protein product of a gene. A) A gene at the level of DNA has an mORF (main Open Reading Frame), a promoter and terminator, a 5'leader and 3' trailer and enhancer/silencer sites. DNA is then transcribed into pre-mRNA. B) Pre-mRNA has the 5'leader, 3' trailer and the mORF that still has its introns. Pre-mRNA is then post-transcriptionally modified into mature mRNA. C) Mature-mRNA has been formed by removing introns, the addition of a 5'cap and a Poly-A tail. Mature-mRNA is then translated into a protein D).

Pre-mRNA is transcribed from DNA and then is further modified by splicing and post-transcriptional modifications (Figure 1.3 B) (16). Pre-mRNA becomes mature-mRNA by, the removal of introns (non-coding sequences), the addition of a 5' cap and a 3' poly(A)-tail and further modifications, such as methylation (2,16–18) (Figure 1.3 B-C). Mature mRNA is then translated into a protein (Figure 1.3 D).

During translation, mature mRNA is decoded by ribosomes and a polypeptide is synthesised from amino acids delivered by tRNAs (19). Translational regulatory mechanisms affect the likelihood of an ORF being recognised and translated into proteins (2). Proteins are crucial macromolecules in supporting life. Proteins function as enzymes, structural proteins, transport proteins, signalling proteins, receptor proteins, antibodies, motor proteins, and regulatory proteins (2). Consequently, by regulating gene expression and protein production organisms can respond to their changing environment and maintain proper functioning.

Defects in gene regulation can cause diseases such as cancer and render organisms susceptible to environmental stress (2,20). Changes within regulatory mechanisms can even lead to antibiotic resistance in bacteria (21). Thus, the study of gene regulation holds potential for advancements in medicine, crop optimisation, and biotechnology.

1.2. Translational machinery

The CPuORF mechanism functions at the translational level so to understand the CPuORF mechanism we must understand this process. Translation is the process in which a mature mRNA is translated by ribosomes to synthesise proteins (22–24). Regulation of translation is a direct way to control protein output when compared to a more indirect regulation that acts at the transcriptional level (23,25). Translation needs mature mRNA, tRNA, ribosomes, translation factors (initiation, elongation, and release factors), amino acids and energy (19).

The Ribosome

The ribosome is a ribonucleoprotein machine that catalyses protein synthesis (24,26) (Figure 1.4). Ribosomes decode genetic information encoded by mature mRNA to create the proteome.

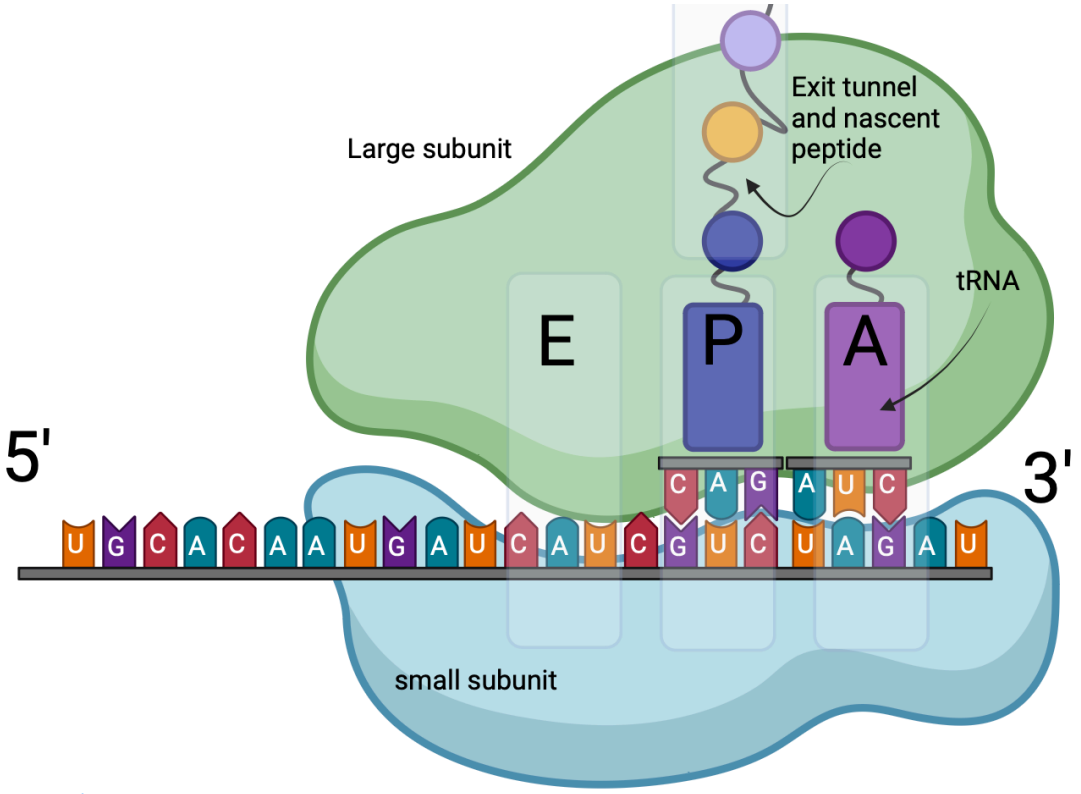


Figure 1.4. Schematic representation of a translating ribosome during the elongation phase of translation (27,28). The elongating ribosome is depicted with its smaller (40S) (blue) and large (60S) (green) subunits. The three sites of the Peptidyl Transferase Centre (PTC) and the associated E (Exit), P (Peptidyl-tRNA), and A (Aminoacyl-tRNA) sites are shown. Two tRNA molecules, represented in dark blue and purple, are illustrated within the P and A sites, respectively. The figure emphasizes the ongoing synthesis of a nascent polypeptide chain, which is progressively emerging from the ribosomal exit tunnel (light green).

Eukaryotic ribosomes are composed of two subunits: the large 60S subunit and the small 40S subunit (2,29) (Figure 1.4). The large subunit contains three rRNA species including the 28S, 5.8S and 5S rRNA and 46 ribosomal proteins (29). The small subunit consists of the 18S rRNA and 33 ribosomal proteins (29). The 28S rRNA of the large subunit forms the peptidyl transferase centre (PTC) and catalyses peptide bond formation (27,30). The small subunit binds to the mRNA and ensures correct base pairing between the mRNA codons and tRNA anticodons (27,30). The 80S ribosome forms from the joining of the two

subunits. The formation of the 80S ribosome also forms the mRNA channel and three sites of tRNA binding in the PTC (Figure 1.4). Plant genomes contain genes that encode multiple paralogues of each ribosomal protein (RP) (31).

Since plants have 2-7 paralogues for each Ribosomal Protein (RP) consequently, rice and *Arabidopsis* have 213 and 235 ribosomal proteins respectively (29). Recently, an attempt has been made to unify the nomenclature for RPs across kingdoms, considering the frequent occurrence of multiple paralogues (32). An example is presented in (Figure 1.5).

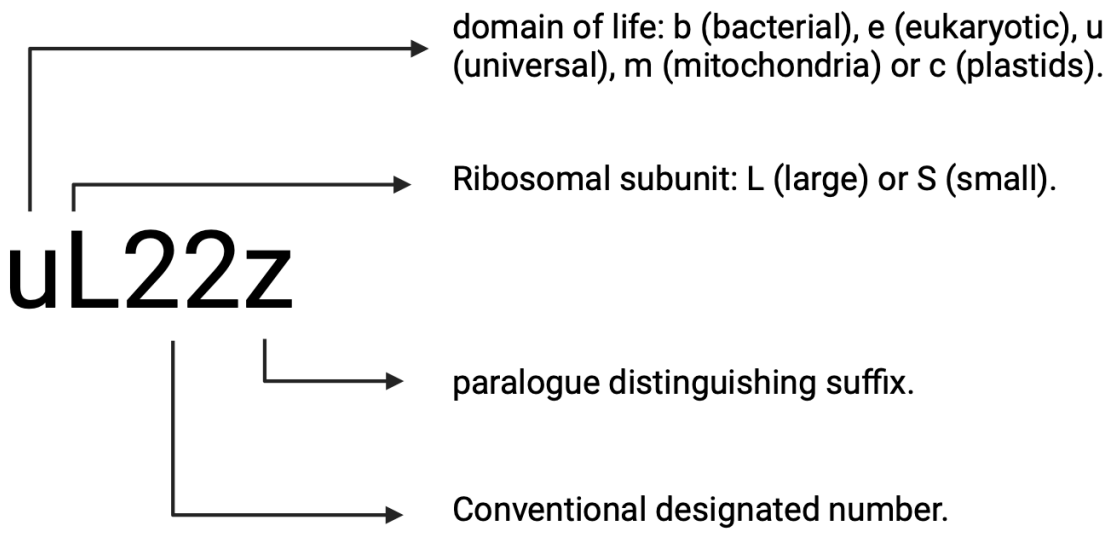


Figure 1.5. A diagram to show conventional ribosome protein nomenclature (32)**.** The first letter indicates the domain of life. Ribosomal proteins can be bacterial (b), eukaryotic (e), universal (u), mitochondrial (m) or plastids (c). The second letter indicates if the protein is recruited into the large (L) or small subunit (S). The number indicates the conventional designated number. The final letter indicates the paralogue distinguishing suffix.

The growing nascent peptide exits through the ribosome's exit tunnel. The exit tunnel is hydrophobic and negatively charged (26,33,34). The exit tunnel is around 80 Å long and 20 Å wide (35). The ribosomal exit tunnel can fit proteins from 30 to 70 amino acids depending on peptide secondary structure within the exit tunnel (36). The exit tunnel joins to the peptidyl transferase centre (PTC) in the large subunit (35) (Figure 1.4). The PTC is the catalytic heart of the ribosome where peptide bonds are formed between amino acids during translation. The exit tunnel in eukaryotes has a constriction around a third of the way from the PTC (27,35). This constriction is formed by uL22 and uL4 in eukaryotes and uL17 and uL4 in bacteria (27,33). Furthermore, the highly conserved loop of uL16 extends into the PTC in eukaryotes and interacts with the 3' end of tRNA during translation (27,33,35).

In plants, mutations in ribosomal proteins result in a variety of phenotypes, such as reduced shoot and root growth (37–39), a leaf defect known as "pointed first leaf" (2,31,40,41), developmental defects (42) and defects in dorso-ventral leaf patterning (42,43). The variability of these phenotypes suggests differing roles for ribosomal proteins and their paralogues, hinting at a regulatory mechanism facilitated by specific ribosomal proteins (2). Furthermore, some mutations in ribosomal proteins result in embryo lethality, demonstrating the necessity of a functioning ribosome (44).

Translation Factors

Translation requires extra-ribosomal proteins to mediate its different stages of translation. Translation canonically has three stages, initiation, elongation, and

termination (2). Outside of the three translation stages ribosomes can also re-initiate translation at a downstream start codon on a transcript with multiple ORFs (27,45). Translation re-initiation also uses translation initiation and release factors.

Eukaryotic Initiation Factors

Eukaryotic translation initiation factors (eIFs) interact with the ribosome to facilitate translation (Table 1.1) (19). Initiation factors have been implicated in start codon selection, ribosome biosynthesis, mRNA recruitment and processing (46,47). Initiation factors are organised into different groups, 1 to 6 (46). eIFs are single or multi-subunit protein complexes that are distinguished by numbers and Greek letters (Table 1.1) (19). Mutations in initiation factors can be embryo lethal or lead to stunted growth, impaired stress response, abnormal root hair development and impaired auxin response (2,27,46,48).

eIF	Function	TAIR	Ref
1	Plays a role in identifying the start codon and adjusting the precision during the start phase of translation. Maintains fidelity of initiation codon selection.	At4g27130, At5g54760, At5g54940, At1g54290	(2,19,49)
1A	Essential for the binding of the eIF2–GTP–Met-tRNAi Met complex to the 40S subunits to form the 43S Pre-Initiation Complex (PIC). Along with eIF1, it facilitates the scanning of the ribosome and the selection of the start codon.	At5g35680, At2g04520	(2,19,49)
2	Forms a ternary complex with GTP–Met-tRNA$_i$. eIF2 binds to met-tRNA$_i$ and transports to the 40S subunit. Subunits: α; β; γ. γ forms the core of	α (At2g40290, At5g05470); β (At5g20920, At5g01940, At3g07920); γ	(2,19,47)

	the complex and binds to both α and β.	(At1g04170, At4g18330).	
2A	Unknown role in plants; in mammals it functions to facilitate IRES mediated initiation.	At1g73180	(19)
2B	Guanosine nucleotide exchange factor that promotes GDP–GTP exchange on eIF2 function. Unknown in plants. Subunits: α; β; γ; δ; ε.	α (At1g53880, At1g72340, At1g53900); β (At3g07300); γ (At5g19485); δ (At5g38640, At1g48970, At2g44070); ε (At3g02270, At2g34970, At4g18300).	(2,19,50)
3	Participates in the formation of the pre-initiation complex and selection of the start codon, encourages the binding of mRNA and Met-tRNAi Met to the ribosome. Required for efficient re-initiation of mORF in uORF containing transcripts. 13 subunits a; b; c; d; e; f; g; h; I; j; k; l; m.	A (At4g11420); b (At5g27640, At5g25780); c (At3g56150, At3g22860); d (At4g20980, At5g44320); e (At3g57290); f (At2g39990); g (At3g11400, At5g06000); h (At1g10840); I (At2g46280, At2g46290); j (At1g66070, At5g37475); k (At4g33250); l (At5g25754, At5g25757); m (At3g02200, At5g15610).	(2,19,48,51,52)
4A	Acts as a DEAD-box ATPase and ATP-fueled RNA helicase. Facilitates the unwinding of 5' leader.	At3g13920, At1g54270	(2,19,53)
4B	RNA-binding protein that improves the helicase activity of eIF4A.	At3g26400, At1g13020	(2,19,53)
4E	Protein that binds to the cap of mRNA. Works with other eIF4 family proteins.	At4g18040, At1g29590, At1g29550	(2,19,53)

4F	Cap-binding complex that unwinds the 5' untranslated region (UTR) of mRNA. It assists in the binding of 43S pre-initiation complexes and supports ribosomal complexes during scanning. eIFiso4F. eIFiso4G interacts 4A, 4B, 5, 4E. eIFiso4E binds to iso4G and 5' cap.	Iso4G (At5g57870, At2g24050), iso4E (At5g35620)	(2,19,53)
4G	Binds to eIF4E, A, and 3, as well as PABPs (Poly A Binding Proteins) and mRNA. It enhances the helicase activity of eIF4A.	At3g60240	(2,19,53)
4H	RNA-binding protein that bolsters the helicase activity of eIF4A. It is like a fragment of eIF4B. novel cap binding protein in *Arabidopsis*.	At5g18110	(2,19,53)
5	GTPase-activating protein that triggers the hydrolysis of eIF2-bound GTP upon recognition of the start codon.	At1g77840, At1g36730	(2,17, 56)
5B	Mediates ribosomal subunit joining in a GTP-dependent manner by positioning the initiator tRNA in AUG with 1A.	At1g76810, At1g21160	(2,19)
5C	eIF5 mimic protein that regulates eIF2 function by being both a mimic and competitor for eIF5. Function unclear in plants	At5g36230, At1g65220	(19,47,54)
6	Prevents premature association of 60S and 40S subunits	At3g55620, At2g39820	(2,19)
PABP	Binds to the poly A tail on the mature mRNA. Interacts with eIF4G.	At2g23350, At4g34110, At1g22760, At1g71770, At3g16380, At1g49760, At2g36660, At1g34140, At5g65250, At5g65250	(2,19)
4E2	Function unknown in plants.	At5g18110	(19)

Table 1.1. A table detailing key eukaryotic initiation factors (eIFs) and their function in translation (2,27,30,46–48,55,56)**.**

Eukaryotic elongation and termination factors

Eukaryotic elongation and termination factors assist in their respective phases of translation. Translation elongation factors facilitate the delivery of tRNAs into the A site and translocation of the ribosome downstream on the mRNA (2,57). Furthermore, there is evidence that specific elongation factors assist in a sequence dependent manner. For example, eEF5 aids in the elongation and synthesis of poly-proline and glycine rich regions (2).

Release factors assist in the dissociation of the mRNA, tRNAs, nascent polypeptide and ribosomal subunits from each other (19,52,58).

Factor	Function	TAIR	Ref
eEF1A	Binds aminoacyl tRNA and GTP. Responsible for delivery of all aminoacyl-tRNAs to the ribosome except the initiator tRNA. Has extra-ribosomal functions in nuclear export, cytoskeleton organisation and apoptosis.	At1g07920, At1g07930, At1g07930, At1g07940, At5g60390	(19,59)
eEF1B	Recycles eEF1A-GDP. Three subunits α; β; γ. Nucleotide exchange factor. Regulates cell cycle.	α (At5g12110, At5g19510); β (At1g30230, At2g18110); γ (At1g09640, At1g57720).	(19,60)
eEF2	Translocation of ribosomes during elongation.	At1g56070, At3g12915	(19,61)
eEF5	Elongation of poly-proline/glycine regions.	AT1g26630, AT1g69410, AT1g13950	(19,62)
eRF1	Termination/peptide release. Recognises three stop codons and catalyses the hydrolysis of the peptidyl-tRNA.	At5g47880, At1g12920, At3g26618	(19,63)
eRF3	Termination/peptide release. Promotes dissociation of eRF1.	At1g18070	(19,63)
ABCE1	Involved in ribosome recycling. Family of ATPases.	At3G13640, At4g19210	(19,64)

Table 1.2. A table detailing key eukaryotic elongation and release factors and their functions in translation (2,19,59–61,63,64).

Messenger RNA

Messenger RNAs (mRNAs) are single stranded molecules of RNA that have been exported into the cytoplasm from the nucleus (2). mRNA encodes sequences that can be translated by the ribosome to synthesise protein (Figures 1.3-4). RNA is a chain of nucleotides composed of guanine, uracil adenine and cytosine (27). Unlike DNA the sugar phosphate backbone of RNA is made from ribose rather than deoxyribose. Triplet nucleotide codons, which code for a single amino acids, are complemented by of anti-

codon on transfer RNAs (tRNAs) during translation. Mature mRNA is created during transcription and post-transcriptional modification (65).

Mature mRNA contains a 5'-m^7GpppN-cap (5' cap) structure that protects the mRNA from degradation and distinguishes a transcript ready for translation (19). Scanning pre-ribosomes bind to the mRNA's 5' cap and 3' poly(A)-tail to initiate translation and produce the corresponding proteins (19). In plants, eIF4F is the 5' cap binding complex and it is composed of eIF4E and the scaffolding protein eIF4G (2). eIF4G recruits eIF4A (a helicase) alongside eIF4E, PAPB and eIF3 to circularise the mRNA (19). Upon mRNA circularization, eIF4A alongside eIF4B unwinds mRNA secondary structures and facilitates the recruitment to synthesise the 43S pre-initiation complex (19). The 43S pre-initiation complex contains the small ribosomal subunit (40S) bound by the initiation factors eIF1, eIF1A, eIF3, and the eIF2-Met-tRNAiMet-GTP ternary complex (eIF2-TC).

tRNA

Transfer RNAs (tRNAs) play a crucial role in protein synthesis by delivering corresponding amino acids encoded in mRNA triplets to the ribosome (22,23). Each tRNA must be charged with its respective amino acid by aminoacyl-tRNA synthetases (Figure 1.6). *Arabidopsis* has an abundance of tRNA genes, with 637 tRNA-like genes in the nuclear genome, 599 of which are considered potentially functional (23). In contrast, there are only 45 aminoacyl-tRNA synthetase genes, implying that a single synthetase may charge multiple tRNAs.

Each tRNA molecule consists of an L-shaped 3D structure that enables them to fit into the P and A sites of the ribosome (22,33). Eukaryotic tRNA is typically 76-90 nucleotides in length (27,33). tRNA has an acceptor stem where the amino acid is loaded onto the tRNA at the CCA amino acid attachment site (Figure 1.6). Furthermore, tRNA has a T-, D-, variable- and an Anti-codon loop that recognises the mRNA anticodon. The amino acid is loaded onto the tRNA by aminoacyl-tRNA synthetases to form aminoacyl-tRNA, binding to the 3' hydroxyl group on the CCA (cytosine-cytosine-adenosine) tail of the tRNA molecule (28). The anticodon loop contains the anticodon that complements the codon on the mRNA, leading to protein synthesis (27). The T-loop is a specialised region on the tRNA that acts as a special recognition site for the ribosome during translation (27). Finally, the D-loop is the recognition site for the aminoacyl-tRNA synthetase (27).

Interestingly, the methionine used in initiation is different to the one used in elongation. The start codon's initiator amino acid, $Met_{i,}$ is delivered by the Met-$tRNA_i$ (27) The Initiator tRNA is specific for methionine but is distinct from methionines used in translation elongation (27). Met-$tRNA_i$ is created post-transcriptionally to distinguish it from elongator methionine (27). The methionine initiator and elongator tRNAs are both charged by methionyl tRNA synthetase (66). The initiator methionine must be delivered to the P site during initiation by eIF2:GTP, whereas elongator methionine is delivered to the A site by eEF1A during elongation (66). The initiator tRNA has an affinity to bind to eIF2 and not eEF1A because of a conserved sequence change on the T-loop (59).

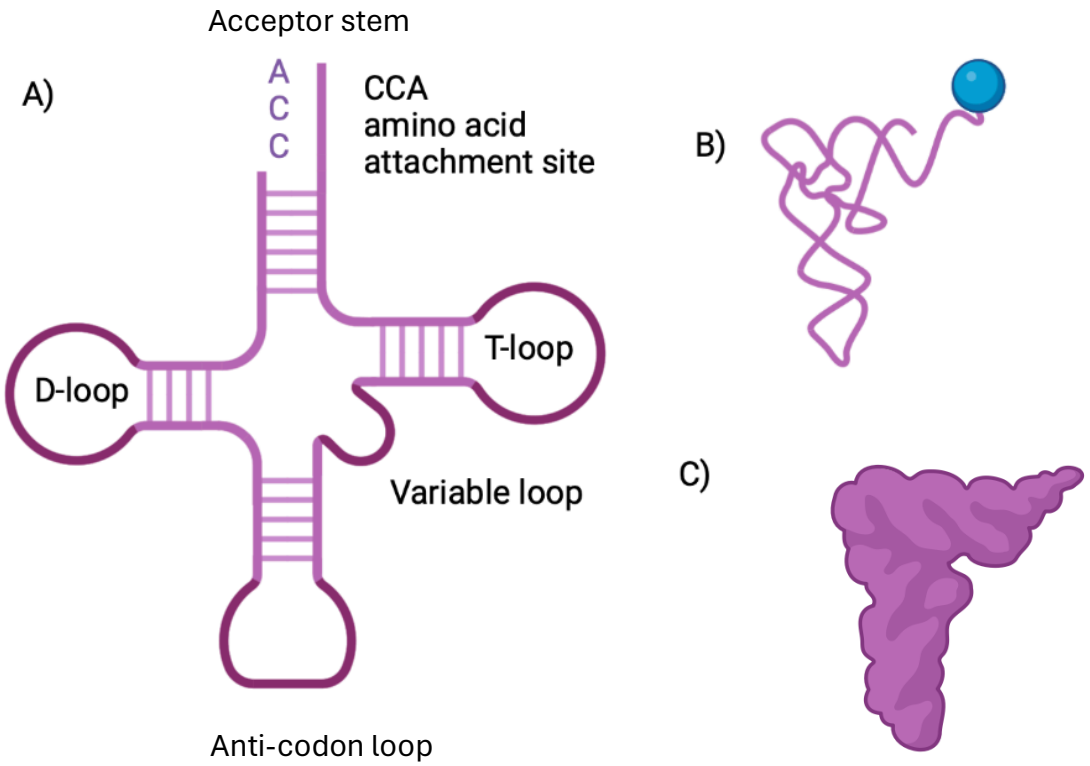


Figure 1.6. Schematic representations of tRNA structure (27)**.** A) Depicts the classic cloverleaf model of tRNA, with labels indicating the CCA site for amino acid attachment, the T-loop, the variable loop, the anticodon loop, and the D-loop. B) Represents a charged tRNA structure attached to an attached amino acid (blue circle). C) Exhibits an L-shaped structure of tRNA.

1.3. Overview of the translational processes.

Translation has three canonical steps, Initiation, Elongation and Termination. This section will discuss the general process of these steps. Furthermore, this thesis will also discuss translation re-initiation here as it requires factors from the canonical translation steps.

Translation initiation

Transient expression data suggests that the CPuORF CUTS mechanism functions during translation and this thesis will explore the general overview of this process in eukaryotes (3). Translation initiation is the main point of translational regulation in eukaryotes (2). Initiation covers the recruitment and assembly of the 80S translating ribosome and the identification of the start codon on mRNA (27,46). Translation initiation is a dynamic process encompassing interactions between rRNA, mRNA and proteins (Figure 1.7) (30). For initiation to occur, a cell needs multiple actors: mature-mRNA, eukaryotic initiation factors and proto-ribosomes (2).

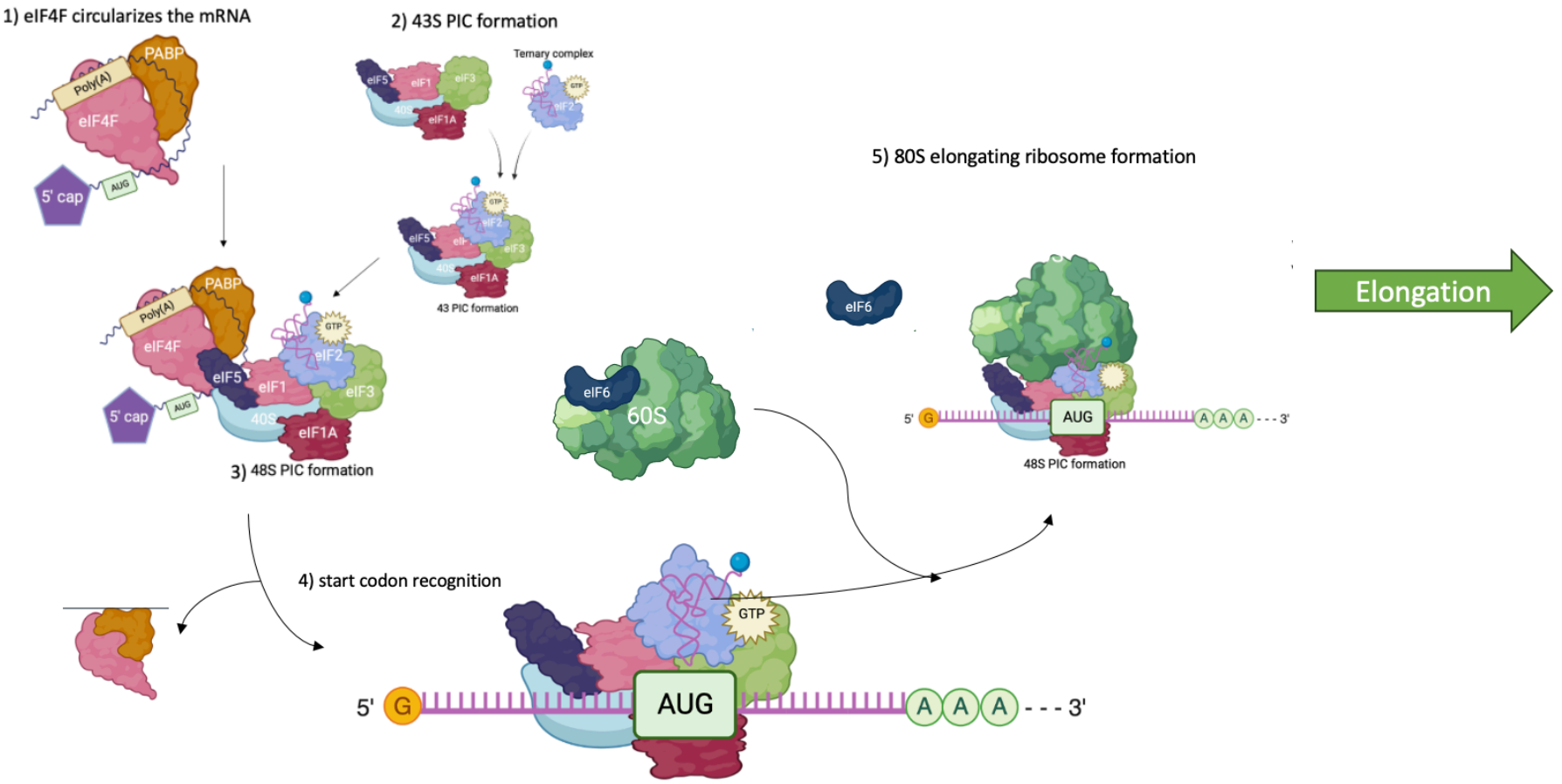

1) eIF4F circularizes the mRNA
PABP
Poly(A)
eIF4F
5' cap
AUG
2) 43S PIC formation
Ternary complex
eIF5
eIF1
eIF3
40S
eIF1A
GTP
eIF2
43 PIC formation
3) 48S PIC formation
4) start codon recognition
5) 80S elongating ribosome formation
eIF6
60S
5'
G
AUG
A A A - - - 3'
48S PIC formation
Elongation

Figure 1.7. Simplified schematic of eukaryotic translation initiation. 1) eIF4F (pink) circularises the mRNA. Mature mRNA is released into the cytoplasm. The multi protein complex eIF4F binds to the 5' cap (purple). The PABPs (Poly A Binding Proteins) (yellow), bind to the poly(A) tail (yellow box). 2) 43 PIC formation. The small 40S subunit (light blue) recruits eukaryotic initiation factors 5, 1, 1A, 3 (purple, pink, red and green respectively) and the initiator tRNA bound ternary complex (light purple). The resulting complex is the 43S Pre-initiation complex (43S PIC). 3) 48S PIC formation. The 43 PIC and eIF4F complexes bind to form the 48S PIC complex. 4) The 48S PIC scans the ribosome for an AUG start codon, upon recognition the eIF4F complex is released for 60S subunit recruitment (green). 5) 80S elongating ribosome formation. eIF6 (dark blue) is phosphorylated and dissociates from the 60S large subunit to promote its binding to the small subunit. With the 80S ribosome now formed, translation elongation can begin.

Translation initiation: forming the 48S PIC.

Before the 80S ribosome is fully formed, the 40S subunit must recruit extra factors to form the 43 PIC (pre-initiation complex) and then the 48S PIC (Figure 1.7) (19). To form the 43 PIC, the 40S subunit recruits eIFs 5, 3, 1 and 1A and the ternary complex (TC) (Figure 1.7). The TC is multi-subunit complex composing of an initiator tRNA (Met-tRNA$_i$) and GTP bound eIF2 (19). Meanwhile, the eIF4F complex binds to the 5' cap and PABPs (Poly(A)-binding proteins) bind to 3' Poly(A)-tail to circularize the mRNA (Figure 1.7). Plants have a specific eIF4F complex (eIFiso4F) that harbours plant specific proteins eIF4E and 4G(53). *Brassicaceae* also have two extra paralogues on eIF4E, 1b and 1c (19,53,67). It is suggested that plant specific initiation factors have different affinities for mRNAs and provide a mechanism to regulate specific mRNA populations at the translational level (2,19).

Translation initiation: start codon recognition.

The 43 PIC joins the eIF4F complex by binding to the 5' cap via eIF3 to form the 48S PIC otherwise known as the scanning ribosome (2,19,48,53). Scanning ribosomes move from a 5' to a 3' direction along the mRNA to find an initiation start codon (2,27). The initiator tRNA was recruited to the 48S PIC when the TC formed the 43 PIC (Figure 1.7). The initiator tRNA contains a complementary codon to the AUG start codon (19). Upon start codon-anticodon complementation eIF1, eIF1A, eIF2-GDP and eIF5B dissociates to allow for the recruitment of the 60S subunit (2,19,68). After eIF1 displacement, eIF2-GTP is hydrolysed by eIF2 and promoted by eIF5 (2,47). eIF5B GTPase is recruited to enable the large subunit recruitment to the small subunit (2,19,54). Premature joining of the large subunit is prevented by its association of eIF6 (19). eIF6 dissociates from the large subunit assisted by eIF3 upon phosphorylation of eIF6 (51). When the large subunit has been recruited to the 48S PIC, the 80S ribosome has formed and is now an elongating ribosome (Figure 1.7).

Not all start codons are optimized for recognition by the ribosome (69). The Kozak consensus sequence is the optimum sequence for translation in eukaryotes (69). The Kozak sequence can also differ across species (69). The start codon is usually AUG and the sequence around the start codon can determine the efficiency of translation initiation (69). Sub-optimal Kozak sequences can result in the ribosome bypassing a start codon and failing to initiate (69). The bypassing of start codons is called leaky scanning.

Translation elongation

Translation elongation: tRNA charging and codon recognition.

Translation elongation is a circular process in which a charged tRNA delivers a corresponding amino acid to the ribosome. The amino acid is joined to a nascent peptide and the ribosome is moved downstream on the mRNA (19). Translation elongation continues until a stop codon is recognised by the ribosome and the complete peptide is released (2,19,63). tRNA synthetases use ATP to charge tRNAs with specific amino acids (19). The charged tRNA is delivered to the ribosome by eEF1A-GTP (19). The ribosome has three sites: E (exit), P (peptidyl) and A (aminoacyl) (Figure 1.8). Post-initiation, the Met-$tRNA_i$ is in the P site and the A and E sites are vacant (27). The A site is free to receive a charged tRNA complementary to the subsequent codon from the start codon (2,19). eEF1A-GTP delivers the charged tRNA to the A site (19,59). Upon codon-anticodon recognition between the mRNA and tRNA, GTP is hydrolysed, and eEF1A-GDP is released (19,59).

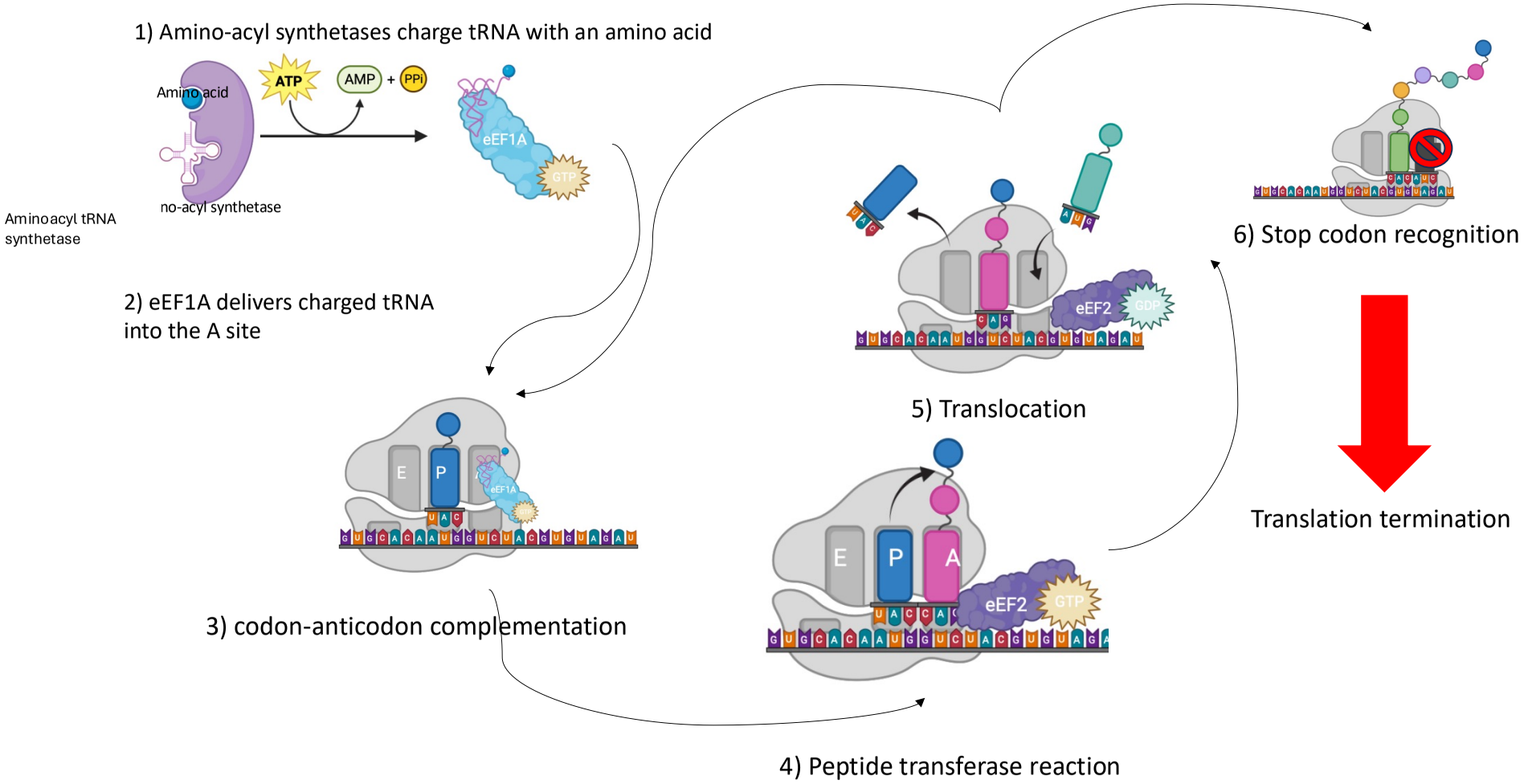

1) Amino-acyl synthetases charge tRNA with an amino acid
Amino acid
ATP
AMP + PPi
eEF1A
GTP
no-acyl synthetase
Aminoacyl tRNA synthetase
2) eEF1A delivers charged tRNA into the A site
E
P
eEF1A
3) codon-anticodon complementation
E
P
A
eEF2
GTP
4) Peptide transferase reaction
eEF2
GDP
5) Translocation
6) Stop codon recognition
Translation termination

Figure 1.8. Simplified schematic of eukaryotic translation elongation. 1) Aminoacyl-tRNA synthetases charge tRNA with an amino acid. The reaction is aided by ATP. 2) eEF1A-GTP delivers the charged tRNA into the vacant ribosomal A site. 3) codon-anticodon complementation. The mRNA codon is complemented by the tRNA anticodon ensuring the correct amino acid is added to the nascent peptide chain. 4) Peptide transferase reaction. eEF2-GTP assists in catalysing the transfer of the nascent peptide in the P site to the amino acid in the A site. 5) Translocation. Hydrolysis of eEF2-GTP to form eEF2-GDP facilitates the translocation of the ribosome downstream the mRNA. De-acylated tRNA is moved into and released from the E site. The growing peptide chain moves into the P site leaving the A site vacant. Elongation continues as the next complementary amino acid is delivered by eEF1A-GTP. 6) stop codon recognition. When a stop codon enters the A site, elongation can no longer continue, and translation termination begins.

Translation elongation: peptide transfer reaction and ribosome translocation.

The large subunit catalyses peptide bond formation between the amino acids in the P and A site (Figure 1.8). Specifically, the 28s rRNA of the large subunit catalyses the reaction to transfers the growing nascent peptide to the A site.

Once the amino acid has been bonded to the growing nascent peptide, the ribosome must translocate downstream to the next codon on the mRNA (19,57,61,70). Ribosome bound eEF2-GTP is hydrolysed to form eEF2-GDP (57,61,70). Hydrolysis moves the ribosome downstream and transfers the nascent peptide to the P site (19,57,61). The de-acylated tRNA is released from the E site and the A site is now vacant for the next complementary charged tRNA (Figure 1.8). Returning us to the start of translation elongation (Figure 1.8). Elongation continues until a stop codon enters the A site to initiate translation termination (19,63).

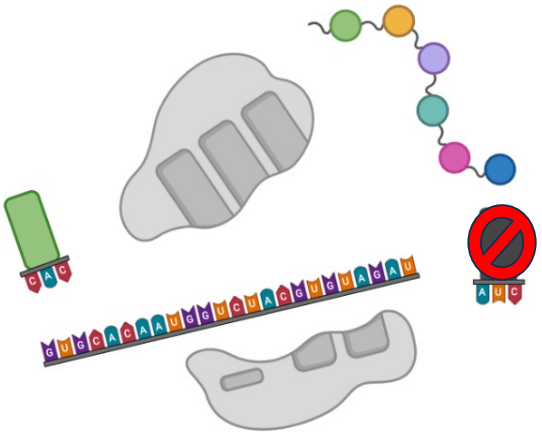
Translation termination

Stop codon recognition and peptide release.

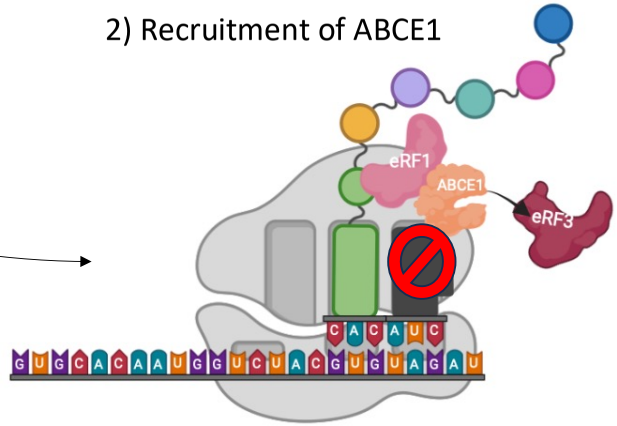
Translation termination is the process whereby the ribosome completes protein synthesis of a translatable ORF (Figure 1.9). During termination a stop codon is recognised and enters the A site post ribosome translocation (19,63). Recognition of a stop codon signals the release of the peptide, dissociation of the ribosome and associated factors and ribosome recycling (2).

When a stop codon (UAA, UGA or UAG) enters the A site, eukaryotic release factors eRF1 and eRF3-GTP forms an eRF complex which is recruited into the A-site (2,19,63). The recruitment of the eRF complex prevents the entry of further charged tRNAs (2). Interestingly, the eRF1 structure resembles that of a tRNA and may explain how it blocks the recruitment of charged tRNAs (2,19). Upon stop codon recognition, eRF3-GTP becomes hydrolysed, and eRF3-GDP is released (63). Ribosome bound eRF1 recruits the ATPase ABCE1 (ATP-binding cassette sub-family E member 1) and eIF6 to disassociate the ribosome and release the completed protein (64). The tRNA in the P site is released by eIF3, eIF1, eIF1A and eIF3j (71). Post-termination, translational machinery is released and recycled for further protein synthesis (72).

1) Stop codon recognition and release factor recruitment



2) Recruitment of ABCE1



3) Translational machinery dissociation

Figure 1.9. Simplified schematic of eukaryotic translation termination. 1) stop codon recognition and release factor recruitment. A stop codon enters the A site and recruits eRFs 1 and 3 and preventing tRNAs entering the ribosome. 2) Recruitment of ABCE1. eRF3 dissociates from its hydrolysis of its bound GTP. ABCE1 is recruited by eRF1. 3) Translational machinery dissociation. eRF factors and ABCE1 trigger peptide release and dissociation of ribosomal subunits for recycling.

***Translation re-initiation*.**

Around 50% of transcripts contains one or more translatable upstream Open Reading Frame (uORF) indicating that translation re-initiation is a common phenomenon (5). uORFs are potentially translated and are found in the 5' leader, upstream of the mORF. To translate a subsequent downstream mORF, the ribosome must remain bound to the mRNA and initiate at a downstream start codon (73,74). Therefore, translation re-initiation occurs upon the failure to dissociate translational machinery during termination and the reassembly of the scanning ribosome (74).

Investigations into translation re-initiation suggest slightly different mechanisms dependent on the length of a regular uORF (74). A short regular uORF has fewer than 10, 16 and 30 codons in yeast, plants, and mammals respectively (56,71,74–76).

Translation re-initiation: short regular uORFs.

During translation termination of a short uORF, the short duration of elongation may facilitate retention of the mRNA bound scanning ribosome and re-initiation specific factors (51,52,77,78). A short duration of elongation makes it more likely that initiation components have not dissociated from the elongating ribosome. Consequently, the machinery required for re-initiation is likely still bound to the ribosome. During translation

termination, the 80S ribosome begins to dissociate but the small subunit remains bound to the mRNA. To re-initiate translation, factors including eIF2-TC, eIF1, eIF1A, eIF3 and eIF4F have been identified in re-initiation mechanisms (58,74,78) (Tables 1.1-2). The factors collectively ensure that the small subunit is bound to the mRNA and capable of accepting an initiator tRNA (Table 1.1-2, Figures 1.7-9).

In plants eIF3h (a subunit of eIF3) has been implicated in translation re-initiation events in uORF-containing mRNAs (51,52,74,77,79). A central growth regulator called TOR activates the expression of the kinase S6K1(Ribosomal protein S6 kinase beta-1 (74,79) S6K1 phosphorylates 40S-bound eIF3h that is required for translation initiation (80). It is suggested that by S6K1 phosphorylating eIF3h the ribosome retains more initiation factors and promotes re-initiation at a downstream start codon (74).

Translation re-initiation: Long regular uORFs.

Translation re-initiation after the translation of a long uORF may require a more complex mechanism when compared to short uORF transcripts. This is because more initiation factors are likely lost during a longer translation elongation phase (74,77). Most studies have been carried out in viruses, but these may illuminate native re-initiation mechanisms. Caliciviruses produce three ORFs with the first containing a Termination Upstream Ribosome Binding Site (TURBS) (81). TURBS forms a secondary structure (stem loop) that tethers a scanning ribosome assisted by eIF3 (81,82). TURBS-eIF3 is capable of binding to a terminating ribosome and binding it to the mRNA to continue scanning and re-initiate at a downstream start codon (81,82).

Plant viruses have also implicated eIF3, TOR and specific ribosome proteins to facilitate re-initiation after the translation of a long uORF. CaMV (Cauliflower mosaic Virus) uses its trans-activator protein (TAV) to mediate translation re-initiation (74). TAV can interact and bind eIF3, TOR and a protein called RISP (re-initiation support protein) to the small or large subunit (55,73). Furthermore, TAV can bind to large subunit proteins uL13, 18 and 24 (55,71,73,74). Through TOR and S6K1's phosphorylation, TAV facilitates re-initiation by binding eIF3 to a terminating ribosome (74).

Overall, studies suggest that re-initiation is promoted by mechanisms that retain translation initiation factors during termination. These factors prevent the completion of translation termination and ribosome recycling. Factors such as eIF2-TC and eIF3 play important roles in tRNA delivery, small subunit stabilisation and mRNA binding. Other factors such as TOR, TAV and S6K1 can facilitate the retention of initiation factors during termination. There are likely multiple mechanisms in which re-initiation occurs.

1.4. mRNA structural elements modulate translation.

mRNA structural elements can also affect mORF translation output. A range of structural elements have been descripted to moderate translational output of the mORF. These features can sequester translating ribosomes and attenuate mORF translation or aid scanning ribosomes to a start codon. Structural features this thesis will discuss include the 5' cap, IRES (Internal Ribosome Entry Sites), uORFs, and Poly(A) tails (Figure 1.10) (2).

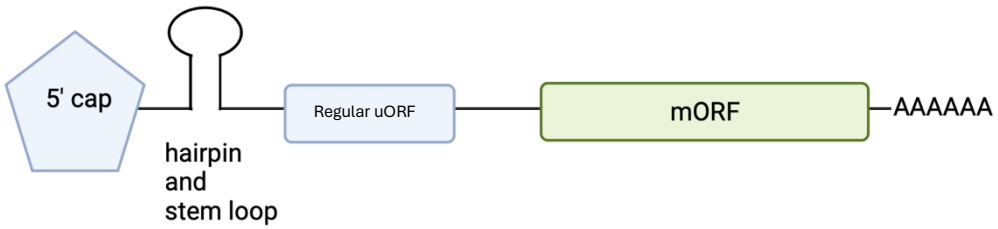


Figure 1.10. Schematic to show structural elements within mature mRNA. Translation of the mORF can be modulated by features encoded in its transcripts. The schematic shows a 5' cap, internal ribosome entry site (IRES), a hairpin loop, a regular uORF and a 3' poly(A) tail.

5' cap and poly(A) tail

The presence of a 5' cap and a 3' poly(A) tail distinguishes an mRNA molecule as being capable of translation (Figures 1.3 and 1.7). These features allow for the recognition of mRNA, facilitating the initiation of translation by the 43S Pre-Initiation Complex (43S PIC) and Poly(A)-Binding Proteins (PABPs) (2,19,71). This recognition enables the circularisation of mRNA ready for translation and the commencement of ribosome scanning.

Hairpins and ribosome shunting

Hairpins are secondary structures that can form on the mRNA and block ribosome access to a start codon as the ribosome struggles to unravel to mRNA (Figure 1.10) (2,83,84). For example, the *Arabidopsis* PHYTOENE SYNTHASE (PSY) gene and the tobacco NPT303 gene have a hairpin loop in their respective 5' leaders, that attenuate mORF expression (83–86).

***Internal Ribosomal Entry Sites*.**

Unlike other structural features, Internal Ribosomal Entry Sites (IRES) can facilitate translation initiation by enabling ribosomes to bypass scanning and cap-dependent translation (Figure 1.10) (2). IRES structures do this by recruiting 40S subunits to an internal site within the 5' leader. Currently, IRES sites have been found in plant and animal viruses and it has been suggested that IRES mechanisms facilitate the translation of specific viral ORFs under stress (87). For example, IRES-like sites have been found in heat shock protein transcripts under heat treatment (2,88). Heat treatment can lead to failures in capping the mRNA and IRES sites can bypass 5' cap mediated translation (88).

***Upstream Open Reading Frames, start codon context and leaky scanning*.**

Regular uORFs attenuate mORF expression as the ribosome initiates translation at the first codon it perceives (37,93) (Figure 1.11). Ribosomes often fail to re-initiate at a downstream subsequent start codons (74). Evidence suggests translation re-initiation on mRNAs occurs under specific conditions (45,52,73,77,80). Consequently, if a transcript has an ORF upstream of the mORF, the mORF's expression is attenuated compared to a non-uORF containing transcript (71). The presence of a uORF is common across taxa. Approximately 50% of transcripts contain at least one uORF in *Homo sapiens*, *Drosophila melanogaster* and *Arabidopsis thaliana* (4,75,90). The effect of a uORF on mORF translation is not a regulatory mechanism as the uORF is not inherently responding to stress.

Leaky scanning is a process where the ribosome does not initiate at an ORF but continues to scan the mRNA and initiates at a downstream start codon (Figure 1.11). Leaky scanning occurs because the upstream ORF has a poor start codon context, which is less likely to be recognized by a ribosome (2,73). The Kozak sequence is the optimal start codon context with AUG being the start codon. The Kozak sequences for mammals, monocots and dicots are 5'-gcc(a/g)ccAUGg-3', (5'-c(a/c)(a/g)(a/c)cAUG-3'), and (5'aaa(a/c)aAUGgcu-3', respectively (2).

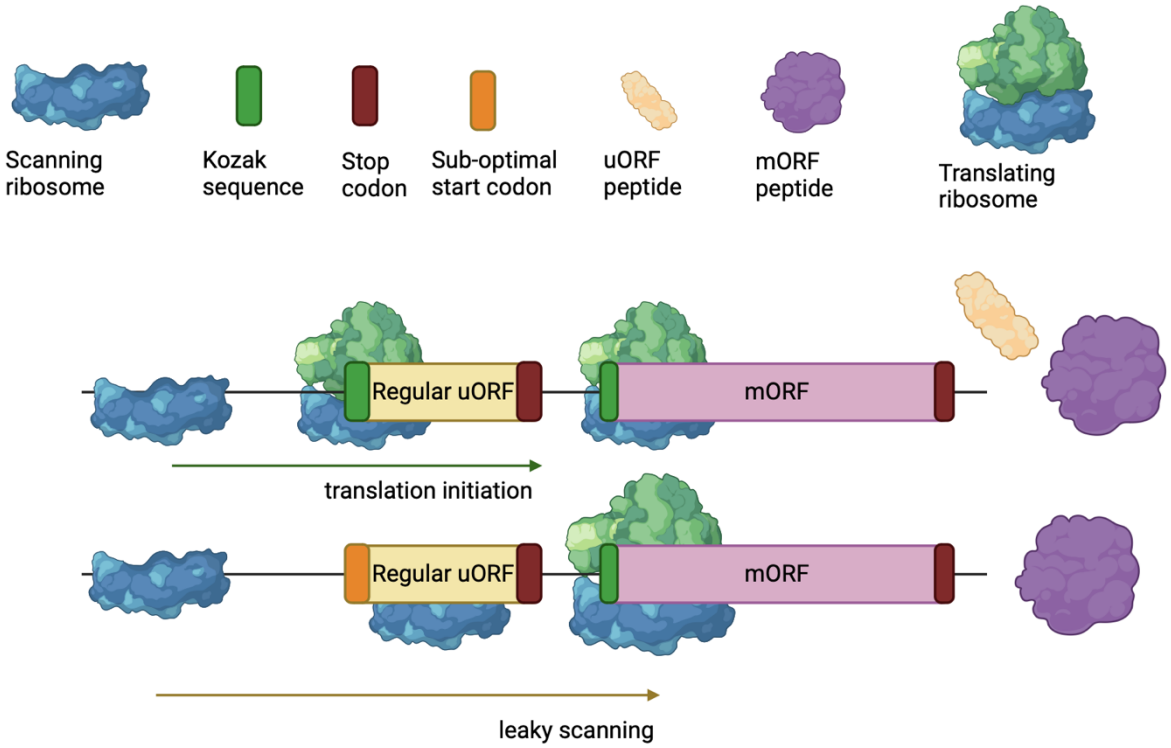


Figure 1.11: A schematic to illustrate the process of leaky scanning. The figure portrays two mature mRNA that are being scanned and translated by scanning and translating ribosomes. On the mRNA is the uORF and downstream mORF in pale blue and green respectively. The ORFs are flanked by a Kozak sequence and stop codon or a sub-optimal start codon and a stop codon. The scanning ribosome initiates translation through the recognition of the uORF Kozak sequence. The ribosomes initiate at the Kozak sequences of the uORF and mORF and produce their respective peptides. During leaky scanning the scanning ribosome bypasses the uORF sub-optimal start codon as it does not recognise the start codon for translation initiation. The scanning ribosome then initiates at the mORF start codon and translates the mORF peptide.

1.5. Plant translational regulation in response to stress.

It is important to note that there are mechanisms to regulate translation in response to stress. Plants are sessile organisms that need to regulate gene expression to adapt to their changing environments. Gene regulation can occur on a translational level to rapidly modulate protein synthesis (28). Translational control and protein synthesis is energetically demanding so there is a global translational reduction during environmental stress (2,28). This section will discuss mechanisms of translational control under stress including by the phosphorylation and composition modification of translational machinery and regulation by specific environmental factors, plant hormones and plant development (2,52).

Phosphorylation and composition modification of translational machinery.

Stress induced phosphorylation of specific eIFs and PABPs can promote or inhibit the activity of specific translation factors (28). Phosphorylation of specific translational machinery can result in differential translational outcomes (2). Stress can elicit downstream signalling cascades that phosphorylate specific translational machinery such as initiation factors. Regulation via phosphorylation does have a degree of specificity. Specific kinases, such as the eukaryotic CK2 (Casein Kinase 2), phosphorylates specific initiation factors (such as eIF2a, eIF2ß, eIF3c, eIF4B, and eIF5) (2,47,91). Modification of initiation factors can promote translation by enhancing interactions between translational components (2,47,48,51,52). Conversely, the phosphorylation of eIF2a can inhibit translation initiation and re-initiation as

phosphorylated eIF2a binds too strongly to eIF2B (2,47,50). Collectively, this constrains ribosome recycling and translation. Phosphorylation of eIF2 is likely a point of regulation by environmental factors through its phosphorylation. Phosphorylation of eIF2a is associated with a reduction of translation post nutrient starvation, UV radiation, cold treatment and in response to specific phytohormones (2).

Ribosome modification leading to specialised and heterogeneous ribosomes has been a subject of debate in recent decades (92). Initially it was hypothesised that tissue specific specialised ribosomes or heterogenous ribosomes may add another layer is translational regulation. Ribosome heterogeneity can be brought about through rRNA sequence variation, absence of specific ribosomal proteins, the exchange of ribosomal protein paralogs and the post-transcriptional or post-translational modification of ribosome components (2,92). Ribosome heterogeneity is of particular interest in plants due to significant number of ribosomal protein paralogs per ribosomal protein family (2,92).

The current consensus of the presence of specialised ribosomes remains unanswered because of the lack of evidence for evolutionary and functional conservation of specialised ribosome populations. However, there is evidence to support cytosolic ribosome heterogeneity in plants (92). Firstly, proteomic data suggests that 25% of cytosolic RPs vary in terms of mass or charge and this affects the overall composition of the 80S monosome (93). Furthermore, in *Brassica* it has been reported that there is tissue specific ribosomal protein paralog composition that may contribute to tissue differentiation and specialisation (94). Moreover, stress can regulate expression of specific paralogs. For example, UV-B stress differentially regulates uL16 paralogs in

Arabidopsis (95), and elevated sucrose induces abundance changes in specific *Arabidopsis* ribosome paralogues (96) Finally cytosolic ribosome transcripts are induced at the transcriptome level by temperature and UV-B stress (97). Overall, the presence of specialised and heterogenous ribosomes is still a matter of debate.

***Influence of environmental factors and hormones on translation*.**

Environmental signals can produce global changes in translation. For example, signals such as light promotes the translation of the circadian clock protein LHY (LONESOME HIGHWAY) (98). Furthermore, a common plant abiotic stress response is the upregulation of hormones such as ABA. These hormones can affect global translation rates by transcriptionally modulating ribosome assembly genes and thereby limiting and organism's translational capacity (99).

1.6 CPuORFs and their identification

Regulation refers to a process through which a biological system modulates the activity or expression of something in response to internal or external signals (2,71,100). There is evidence that CPuORFs are actively regulating mORF output and this is contrary to the passive attenuation of mORF expression by regular uORFs. Regular uORFs passively attenuate mORF expression by sequestering translating ribosomes from the mORF but do not seem to actively respond to signals to modulate mORF expression (2,71). This section will explore how CPuORFs were identified.

Author	Identification method	Results	Ref
Hayden & Jorgensen, 2007	Bioinformatic pipeline. Compared mORF and uORF sequences in *Arabidopsis* and rice. mORF have to be conserved between Arabidopsis and rice.	Identified 26 CPuORF homology groups. Identified 58 CPuORFs in *Arabidopsis*.	(4)
Tran, Schultz, & Baumann, 2008	Bioinformatics pipeline. Homology based approach to identify uORFs in monocots. Sequences were analysed by uORFSCAN.	Identified 4 CPuORFs in *Arabidopsis*.	(101)
Hayden & Bosco., 2008	Bioinformatics pipeline. Compared sequences between *Drosophila melanogaster* and dipteran sequences.	Identified 44 *Drosophila melanogaster* CPuORFs	(90)
Jorgensen & Dorantes-Acosta, 2012	Bioinformatics pipeline. Reanalyses cDNAs with uORF finder.	Identified 4 new homology groups and 6 CPuORFs in *Arabidopsis*.	(102)
Takahashi et al., 2012	BAIUCAS pipeline. Increased number of genomes investigated.	Identified 18 new CPuORFs.	(103)
Vaughn et al., 2012	Comparative transcriptomics pipeline to find conserved sequences in 5' and 3' UTR.	Identified 4 *Arabidopsis* CPuORFs.	(76)
Hayashi et al., 2017	Experimental validation of CPuORFs and ribosomal stalling. Toeprint analysis and transient expression assays.	Identified ribosome stalling CPuORFs.	(7)
Van der Horst et al., 2019	Bioinformatics pipeline that is not biased for AUG start codons.	Identified 29 *Arabidopsis* CPuORFs.	(104)
Takahashi et al., 2020	ESUCA pipeline for 5 plant genomes	Identified 88 novel plant CPuORF homology groups	(105)
Takahashi et al, 2020	ESUCA pipeline to include animal genomes	1373 novel and 144 known CPuORFs from 4 animal genomes	(106)

Table 1.3: Studies that have identified CPuORFs in eukaryotes. Table 3.1.1 lists papers that have identified CPuORFs in chronological order. The tables document the results of each study.

While these pipelines have proved promising in identifying CPuORFs, they have been shown to discriminate against potentially functional CPuORFs. I, as part of Causier et al., published the most extensive *Arabidopsis* CPuORF database and included *Arabidopsis* CPuORFs that have been rejected from published pipelines (3). Van der horst *et al.,* 2019's pipeline defined a CPuORF as being conserved in over 11 species (104).

Finally, initial pipelines observed differential patterns in amino acid conservation across the CPuORFs (Figure 1.12) (3,4,90,102). A subset of CPuORFs displayed conservation exclusively at the C-terminus, whereas other CPuORFs displayed conservation throughout or at the N-terminus. CPuORFs were classified based upon this pattern of conservation. The CPuORF classification system was most recently updated by Causier *et al.,* to differentiate those that displayed conservation exclusively at the N-terminus and those conserved throughout (Figure 1.12) (3). CPuORFs are classified into either Class I (C-terminal conservation), Class IIa (N-terminal conservation) or Class IIb (whole peptide conservation). CPuORF class in respect to CPuORF function will be discussed in 1.9.

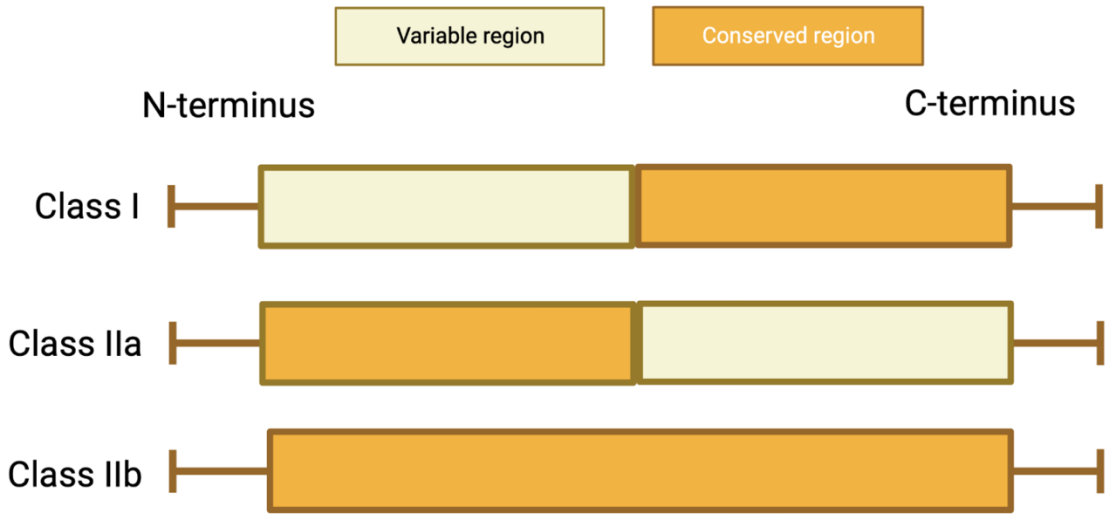


Figure 1.12. A schematic to show conservation patterns in different CPuORF classes (3)**.** CPuORFs classified into Class I, IIa and IIb. Conservation is characterised by the conservation of the N- and C- termini. Class I is conserved at the C-terminus. Class IIa is conserved at their N-termini. Class IIb is conserved throughout.

1.7. CPuORFs attenuate mORF expression by ribosomal stalling.

There is a wealth of evidence from mutagenesis and ribosome profiling that CPuORFs stall the ribosome in a sequence dependent manner (3). To lay the foundations for the CUTS mechanism and ribosome stalling this section will discuss the key evidence for CPuORFs stalling the ribosome during translation.

Mutagenic studies of CPuORF function			
Ebina et al., 2015	Investigated 16 *Arabidopsis* CPuORFs	Luciferase assays between WT, frameshift CPuORFs, and **Δ**AUG *cpuorfs*. 9/10 Class I CPuORFs and 2/6 Class II(a and b) CPuORFs show sequence dependent attenuation. Frameshift mutations as severe as **Δ**AUG. Alanine scanning suggests conserved amino acids confer stalling. Changes in arginine codon usage does not affect function. Showed non-conserved amino acids were not important in stalling in Class IIa CPuORF. Demonstrated it was the conserved region and not the stop codon.	(108)
Rahmani et al., 2009	bZIP11 CPuORF	Luciferase assay showed conserved amino acids are required for sucrose-dependent stalling. Frameshift terminated sequence dependent attenuation of mORF. This study deleted the variable N-terminal amino acids and the truncated C-terminal CPuORF protein retained its sucrose-dependent regulation of the mORF. Hence, the conserved CPuORF amino acids can function independently from the variable.	(109)
Yamashita., 2017	bZIP11 CPuORF	Western blotting with RNase treatments showed mutation in conserved amino acids (ser31) abolished sucrose-induced stalling. Mutations in arginine affect sucrose-induced stalling.	(11)
Uchiyama et al., 2014	AdoMetDC1 CPuORF	Western blots with RNase treatments showed that stalling occurred in conserved region within proximity to the stop codon.	(10)
Hayashi et alk., 2017	22 *Arabidopsis* CPuORFs.	They performed a western blot to determine if CPuORFs stall the ribosome. They investigated 22 CPuORFs and 15 of these stall the ribosome in a sequence dependent manner. Compared wild type CPuORFs to frameshift mutants. They investigated each amino acids contribution to stall the ribosome through mutagenesis and toe print analysis.	(7)
Takahashi et al., 2020	17 Human CPuORFs	*In vitro* luciferase assays in human cells compared WT to a frameshift *cpuorfs*. 8/17 human CPuORFs demonstrated to attenuate mORF in a sequence dependent manner. Lengths of functional human CPuORFs are 41, 63, 15, 70, 25, 15, 11 amino acids.	(106)
Causier et al., 2022	6 *Arabidopsis* CPuORFs, 1	Performed *in vitro* and *in planta* luciferase assays on 6 *Arabidopsis* and 1 rice CPuORF. Identified sequence dependent stalling in both species. Differences in signal response between	(3)

	rice CPuORF	divergent variable regions between Arabidopsis and rice homologues.	
Van der horst et al., 2023	bZIP11 CPuORF	Present a 3.7 Å structure of CPuORF stalled ribosome with sucrose. Evidence suggests a conserved mechanistic framework for ribosome stalling across kingdoms. Sucrose interacts with nascent peptide in a conserved metabolite pocket in the exit tunnel. Stalls during translation termination likely though sucrose affecting eRF1 function.	(110)

Table 1.4: Summary of Key Mutagenic Studies on CPuORF Function. This table outlines the significant findings from multiple studies examining the function of CPuORFs in different organisms.

Mutagenic studies that utilise *in vitro* and *in planta* transient expression assays have shown that CPuORFs can attenuate mORF expression in an amino acid sequence dependent manner (Table 1.4). Transient expression assays that use reporter proteins like, LUCIFERASE and GUS (β-glucuronidase), offer a reliable measure of a CPuORF's effect on mORF translation.

Currently, comparisons between wild type CPuORFs to frameshifted and those with a mutated start codon (ΔAUG) suggest that it is the CPuORF peptide that confers function (3). A frameshifted CPuORF has had a DNA base inserted and removed from the start and end of the protein. This induces a frameshift that has eliminated its conserved peptide sequence. A CPuORF with a mutated start codon prevents the CPuORFs translation. Reporter assays have shown that when these mutants are compared to a wild type CPuORFs there are higher levels of mORF expression when controlled by a mutant CPuORF (3). This suggests that it isn't the mRNA that is contributing to the CPuORFs

attenuation of the mORF. Consequently, mutations in the CPuORF peptide sequence should be investigated to elucidate function.

Western blotting has also proved useful in demonstrating that CPuORFs stall the ribosome (Table 1.4). Western blots can probe for ribosomal stalling by the detecting a peptidyl t-RNA fused to a tagged stalling peptide (7,111,112). The peptidyl t-RNA is approximately 20 kDa. Consequently, a stalled ribosome will have a band for an N-terminally tagged TAG-CPuORF:peptidyl-t-RNA. To demonstrate that this band represents peptidyl t-RNA, the band will no longer be observed following RNase treatment. Through this methodology, 15/22 CPuORFs were demonstrated to stall the ribosome (7). Furthermore, a band for peptidyl-tRNA is not observed when frameshift CPuORFs are investigated this way (7). Indicating that a CPuORF's ability to stall the ribosome is sequence dependent.

Specific amino acids have been suggested to be functional in different CPuORF stalling mechanisms. Mutagenesis has indicated that a serine is functional in the bZIP11 CPuORF (10,11), toe print analysis suggest that prolines around the stop codon of AdoMetDC (*S*-Adenosyl methionine decarboxylase) confer stalling (113,114), and alanine scanning mutagenesis suggest arginine is important in the bZIP11 (AT4G34590) stalling mechanism (11,108). Mutations in the amino acids of these CPuORFs result in higher levels of mORF protein being produced suggesting they are functional. Furthermore, a single mutation in bZIP11s CPuORF (serine -31) eliminates its sucrose-dependent stalling (109). When taken together, this indicates that the amino acid sequence confers CPuORF functionality.

Codon usage can affect translational efficiency, and this has been investigated in the case of five CPuORFs and their arginine residues (108). The study investigated the efficiency of the bZIP11 CPuORF sucrose mediated stalling with different codons for arginine. When using different codons for arginine there was no impact on mORF expression or stalling (108). This further demonstrates that stalling is conferred at the translational level between the nascent CPuORF peptide and ribosomal exit tunnel and not unusual codons slowing down translational efficiency.

Analysis into ribosome profiling data has further demonstrated that CPuORFs stall the ribosome during translation. *Arabidopsis* CPuORFs characteristically show ribosome occupancy enrichment at the CPuORF when compared to its mORF, indicating ribosome stalling (3,107,109,115–117). Studies have suggested that CPuORFs with exclusively N-terminal conservation stall during elongation, whereas C-terminal conserved CPuORFs stall at termination (3). Furthermore, these studies have also shown that stalling is occurring at CPuORF conserved amino acids (3,107,109,115–117).

1.8. A genetic screen implicates the ribosome exit tunnel in CPuORF-mediated stalling.

A mutagenic suppressor screen has provided evidence that a CPuORF peptide can stall the ribosome during translation by interacting with exit tunnel proteins. The screen was performed on a mutant in the thermospermine synthase gene, *acl5-1* (Figure 1.13). *acl5-1* is characterised by its over proliferated vascular tissue and semi-dwarf phenotype (37).

The screen identified four suppressor mutations, one was an early termination codon in a CPuORF and the other three were single amino acid substitutions in ribosomal exit tunnel proteins (uL16z, RACK1z and uL4z) (Figure 1.10-11) (8,37,118). All four mutations partially rescued the mutant phenotype suggesting that CPuORF function may have a relationship with the ribosome exit tunnel (Figure 1.11).

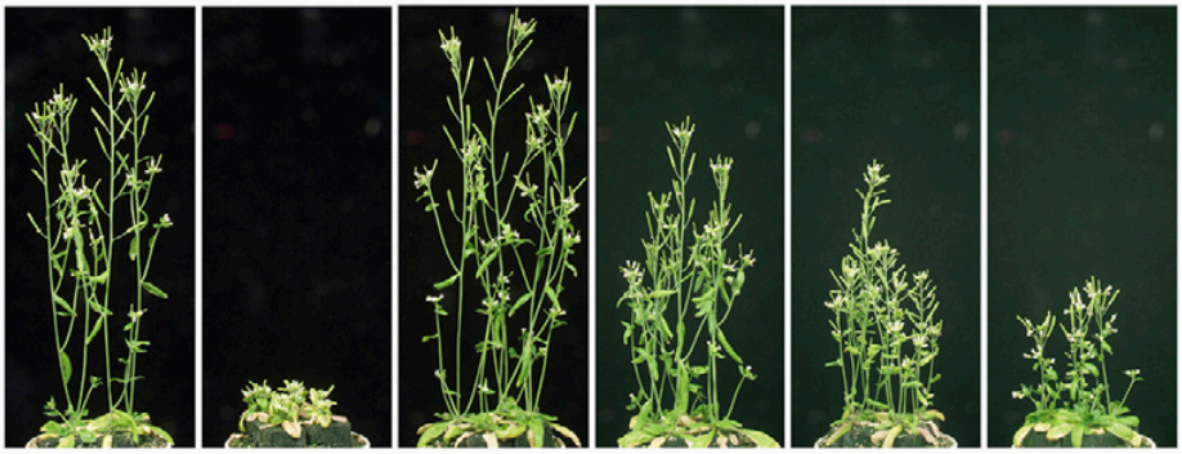


Figure 1.13. Adapted from Imai *et al.*, 2016 to show *acl5-1* dwarf phenotype and suppressor line mutant phenotypes. From left to right, wild type (WT) *Arabidopsis* phenotype, *acl5-1* dwarf phenotype, *acl5-1 cpuorf40* mutant line rescues *acl5-1* phenotype, *acl5-1 ul16z/rack1z/*ul4z partially rescues the acl5-1 phenotype.

To fully explain the implications of the genetic screen we must first introduce the *SAC51 (SUPPRESSOR OF AUCULIS 5-1)* transcript, the *SAC51* thermospermine responsive 5'leader and *ACL5* (Figure 1.14). Firstly, ACL5 converts spermine to thermospermine and the SAC51 mORF is a transcription factor that inhibits the expression of *ACL5* (37,118). Moreover, the expression of the SAC51 mORF is regulated by the CPuORFs in the 5'leader of *SAC51* (3). *In vivo* investigations into the *SAC51* transcript have demonstrated that under low levels of thermospermine, SAC51 expression is attenuated by the CPuORFs in the 5' leader of *SAC51* (3,8). Furthermore, under elevated levels of thermospermine

SAC51 expression is activated resulting in higher levels of the SAC51 mORF protein (3,8). It has been suggested that thermospermine prevents the ribosome from stalling at the CPuORFs. Moreover, as the ribosome does stall, it re-initiates at the downstream *SAC51* mORF. The expression of SAC51 can then inhibit ACL5 expression, reducing the production of thermospermine from spermine by ACL5 (119). This mechanism could be described as a feedback loop to modulate thermospermine biosynthesis through the thermospermine responsive *SAC51* CPuORFs (Figure 1.14).

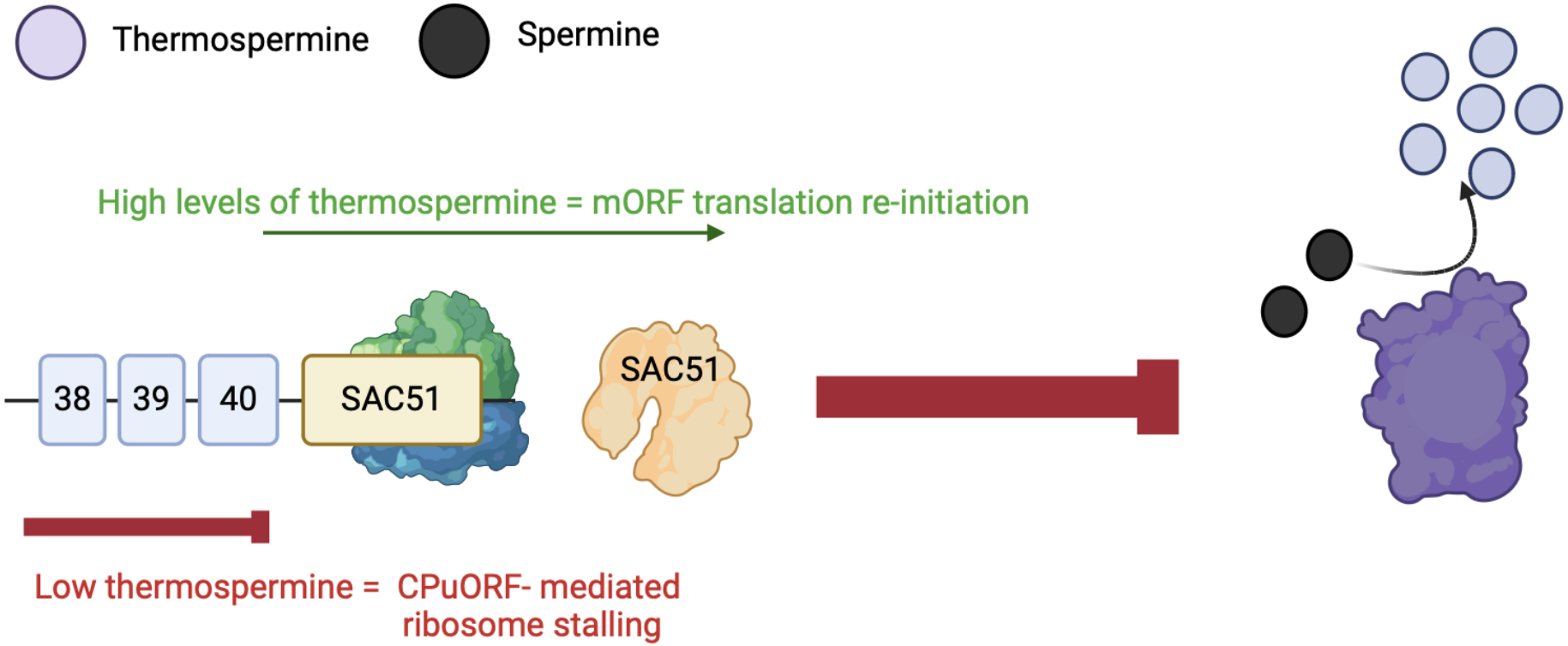

Thermospermine
Spermine
High levels of thermospermine = mORF translation re-initiation
38
39
40
SAC51
SAC51
Low thermospermine = CPuORF- mediated ribosome stalling

Figure 1.14. A diagram to illustrate the ACL5:SAC51 thermospermine feedback loop. The SAC51 transcript has three CPuORFs in its 5'leader (CPuORF 38, CPuORF 39 and CPuORF 40). Under low thermospermine conditions the *SAC51* CPuORFs inhibit SAC51 mORF expression by stalling the ribosome. This attenuates the expression of the SAC51 mORF that inhibits the expression of ACL5. In the absence of the SAC51 mORF protein, ACL5 converts spermine to thermospermine and creating high thermospermine conditions. Thermospermine and spermine is represented by a black and yellow circle, respectively. Under high levels of thermospermine, the ribosome does not stall at the SAC51 CPuORFs and the SAC51 mORF protein is translated. The production of the SAC51 mORF protein then inhibits ACL5 expression, lowering thermospermine conditions.

The suppressor screen by Imai *et al.,* investigated a loss of function *acl5-1* mutant that had a severe dwarf phenotype (Figure 1.13) (37,118). Interestingly, a suppressor mutation of the *acl5-1* phenotype was found in the third CPuORF in the *SAC51* transcript called CPuORF 40 (Figure 1.14) (37). At this juncture it is important to note, the *SAC51* transcript is of particular interest because it harbours multiple CPuORFs (Figure 1.14). Most CPuORF transcripts contain a single CPuORF however, there are 6 transcripts in *Arabidopsis* that contains one or more (3). Four of those transcripts are *SAC51* and *SAC-LIKE* proteins, *SAC51* (AT5G64340) (three CPuORFs), and *SACL1* (AT5G09460) (three CPuORFs), *SACL2* (AT5G50010) (three CPuORFs) and *SACL3* (AT1G29950) (two CPuORFs) (3). The three CPuORFs in SAC51 are called CPuORF 38, CPuORF 39 and CPuORF 40 in a 5' to 3' direction. Multi-CPuORF transcripts are very rare when you consider 0.5 % of *Arabidopsis* transcripts contain a CPuORF and 4.5 % of CPuORF-containing transcripts have multiple CPuORFs (3). The SAC51 CPuORFs have been reported to have aCUTS activity, meaning that in response to thermospermine more mORF protein is produced (Figure 1.1) (3,8). It is currently unknown how the SAC51 CPuORFs individually or jointly confer aCUTS activity. However, the genetic screen described above showed that an early STOP codon in the third CPuORF of SAC51 (CPuORF 40) rescues the *acl5-1* mutant phenotype (Figure 1.14-15) (37).

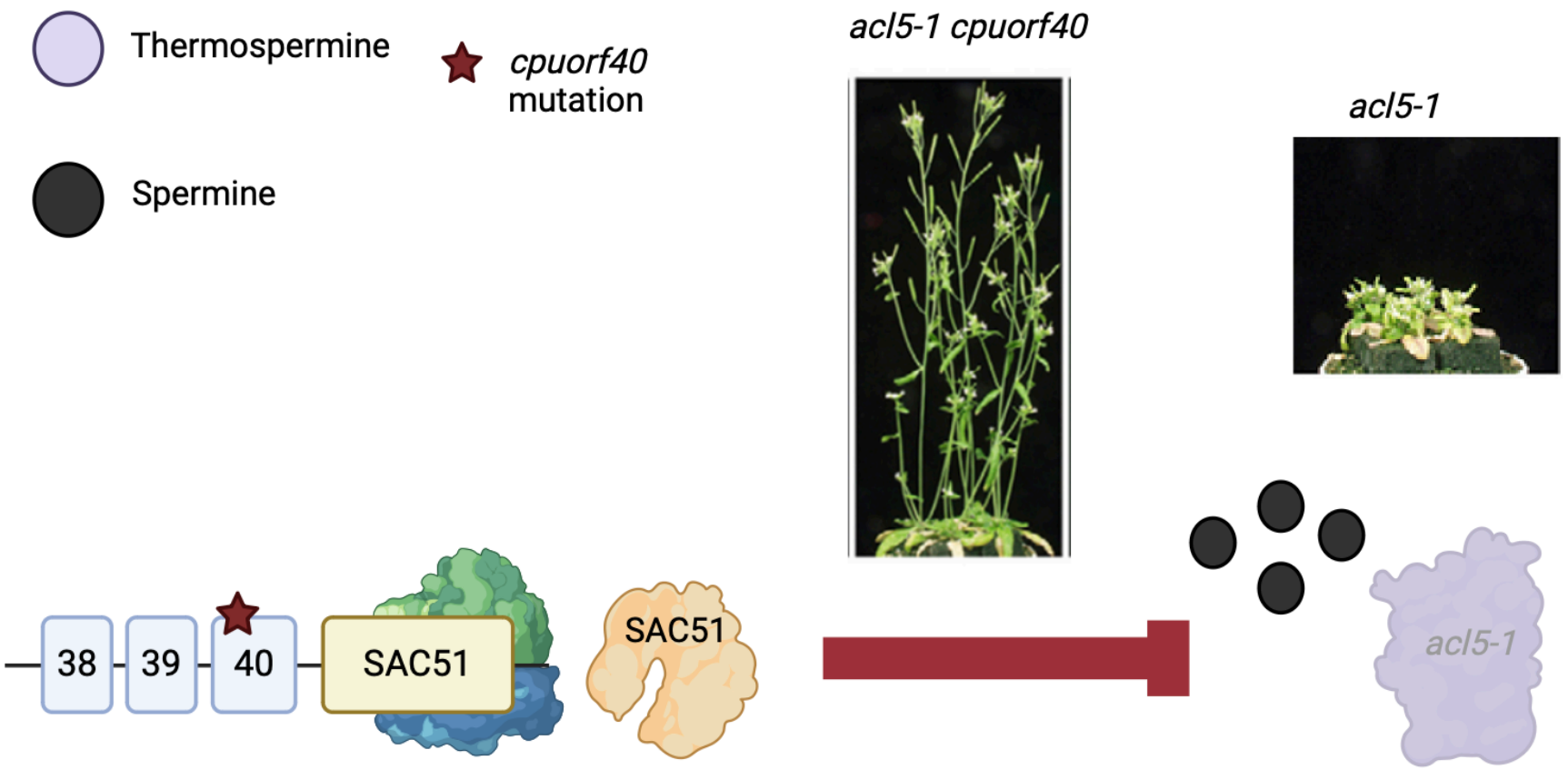

Thermospermine
cpuorf40 mutation
Spermine
acl5-1 cpuorf40
acl5-1
38
39
40
SAC51
SAC51
acl5-1

Figure 1.15. A diagram to show how a mutation in the SAC51 CPuORF 40 rescues the *acl5-1* phenotype. The diagram displays the SAC51 transcript with its three CPuORFs (38, 39 and 40) and the early stop mutation in CPuORF 40 (red star). The *acl5-1* mutant is a loss of function mutation of the thermospermine synthase gene ACL5. Thermospermine is a polyamine that ACL5 synthesises from spermine. In the *acl5-1* and *acl5-1 cpuorf40* backgrounds there is no ACL5 to convert spermine into thermospermine. In the *acl5-1 cpuorf40* background, the mutation prevents the ribosome from stalling at SAC51's CPuORFs. Therefore, the ribosome is then able to translate the SAC51 mORF that rescues the *acl5-1* phenotype. The mutation in CPuORF 40 bypasses the need for thermospermine to alleviate ribosomal stalling.

So how does a mutation in a SAC51 CPuORF 40 rescue the *acl5-1* phenotype? Evidence suggests that some CPuORFs can attenuate mORF expression through stalling the ribosome during translation (Table 1.4). Previous investigations into the regulation of the SAC51 mORF suggest that under elevated levels of thermospermine, the ribosome fails to stall on its CPuORFs (3,8). Consequently, under elevated levels of thermospermine, the ribosome can initiate translation at the SAC51 mORF start codon and produce the SAC51 mORF protein (Figure 1.14-15) (3,8). In the *acl5-1* mutant there are lower levels of thermospermine and therefore, thermospermine is not available to prevent the ribosome from stalling at the SAC51 CPuORFs (Figure 1.15) (119). This prevents the SAC51 mORF from being expressed to contribute to a wild type phenotype. However, in the *acl5-1 cpuorf40* line, the truncated CPuORF 40 cannot facilitate ribosome stalling and allowing the translation of the SAC51 mORF to rescue the *acl5-1* phenotype (Figure 1.15).

The next question is, how does the expression of the SAC51 mORF protein contribute to the rescuing of the *acl5-1* phenotype? To address this, it is important to explain how the SAC51 mORF functions post translation in regulating plant vasculature (Figure 1.16) (119). The SAC51 mORF protein competes with LHY (LONESOME HIGHWAY) to form a heterodimer with TMO5 (AT3G25710) (Figure 1.16) (119). The SAC51:TMO5 and

TMO5:LHY heterodimers are bHLH transcription factors that compete to regulate plant vasculature (119). The over-representation of TMO5:LHY causes the excessive proliferation of vascular tissue whereas, SAC51:TMO5 counteracts TMO5:LHY targets (119).

So how does a loss of function mutant in *ACL5* affect the SAC51:TMO5 and TMO5:LHY balance? In the *acl5-1* mutant there are lower levels of thermospermine and consequently, lower levels of the SAC51 mORF protein being produced due to the ribosome stalling at its CPuORF (Figure 1.16) (119). This results in less SAC51 being available and a larger proportion of the TMO5:LHY heterodimer. This imbalance contributes to the dwarf phenotype and over-proliferated vascular tissue (119). Whereas, in the *acl5-1 cpuorf40-d* suppressor line, the ribosome does not stall at the SAC51 CPuORFs and resulting in more SAC51 mORF being produced to re-establish a wild type balance of the SAC51:TMO5 and TMO5:LHY heterodimers (119), therefore, rescuing the *acl5-1* phenotype. It is unknown how a combination of the SAC51 CPuORFs contribute to the regulation of *SAC51* but this evidence suggests that CPuORF 40 is functional.

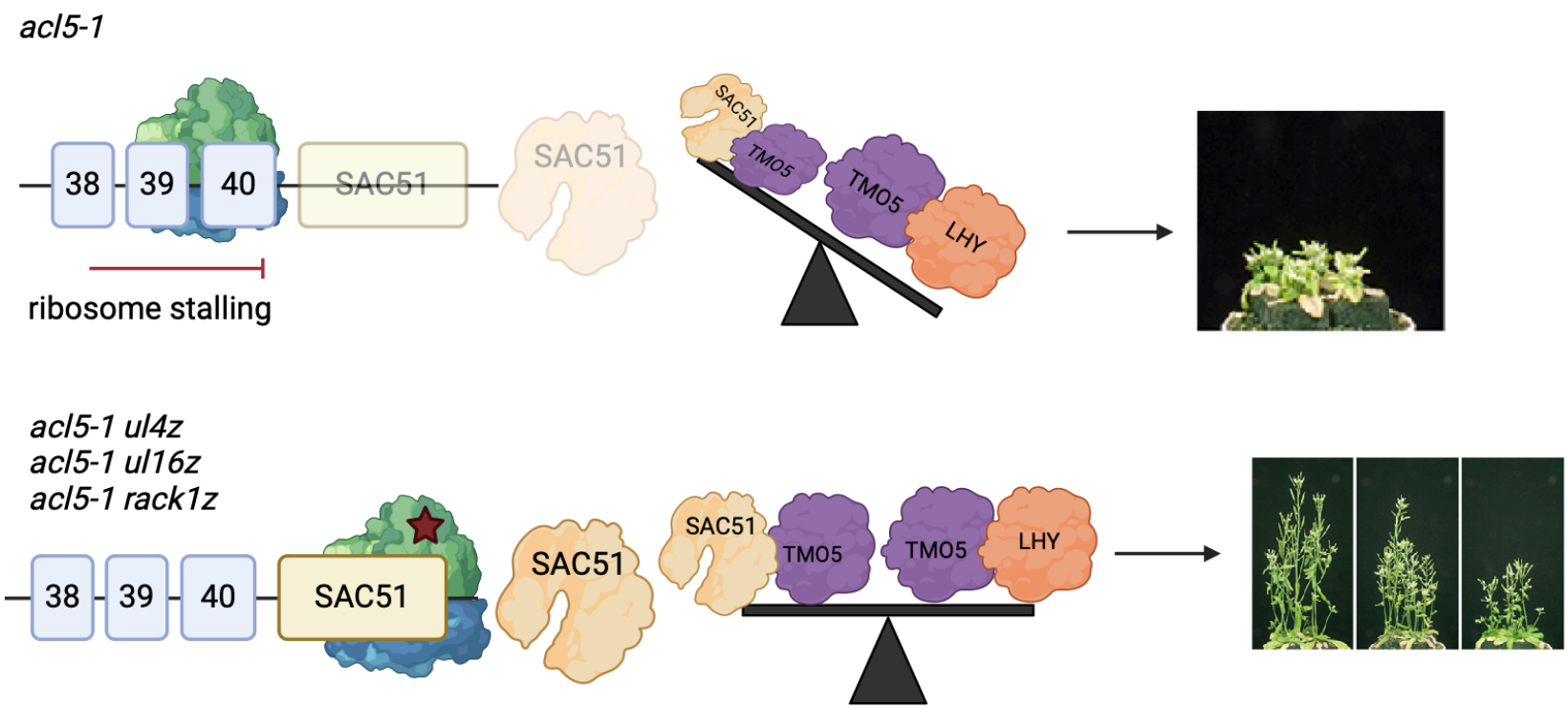

acl5-1
38
39
40
SAC51
SAC51
ribosome stalling
SAC51
TMO5
TMO5
LHY
acl5-1 ul4z
acl5-1 ul16z
acl5-1 rack1z
38
39
40
SAC51
SAC51
SAC51
TMO5
TMO5
LHY

Figure 1.16: A diagram to show the effect of the *acl5-1* and *acl5-1 cpuorf40-d* on the relative abundance of SAC51:TMO5 and TMO5:LHY. In the *acl5-1* background the ribosome stalls at the CPuORFs of SAC51 as there is relatively low thermospermine due to the absence of ACL5 function. This results in an imbalance of TMO5:LHY when compared to SAC51:TMO5. By have more TMO5:LHY heterodimers this results in the over proliferation of vascular tissue resulting in the dwarf phenotype. In the *acl5-1 cpuorf40-d* the ribosome does not stall at SAC51's CPuORFs and causing the expression of the SAC51 mORF protein. The production of SAC51 mORF re-establishes the wild type balance between the two heterodimers and contributes to the rescuing of the *acl5-1* phenotype.

Three further *acl5*-1 suppressor mutations were identified and these shed light on the mechanism of CPuORF-mediated stalling as suggested by CPuORF 40(37,118) (Figure 1.17-18). The ribosome mutations are in the proteins uL4z, uL16z and RACK1 (RECEPTOR FOR ACTIVATED C KINASE 1) (figure 1.13, 1.17) (8,37,118). uL4 is an exit tunnel protein that forms a constriction within the tunnel alongside uL22 (32,118). uL16 is a protein that extends into the PTC and assists tRNA delivery to the PTC (95,120,121). RACK1 is a scaffold protein between the large and small subunit, found around the exit of the exit tunnel (122,123). RACK1 is a eukaryotic specific protein that stabilises stalled and collided ribosomes (124). Notably, all three proteins have been implicated in ribosome stalling mechanisms of known ribosome arrest peptides (to be discussed in section 1.11). Moreover, cryo-EM data indicates that these proteins interact with nascent peptides capable of stalling the ribosome during translation (20,124). Thus, the evidence suggests that the suppressor screen has potentially hit ribosome residues likely involved in the CPuORF stalling mechanism.

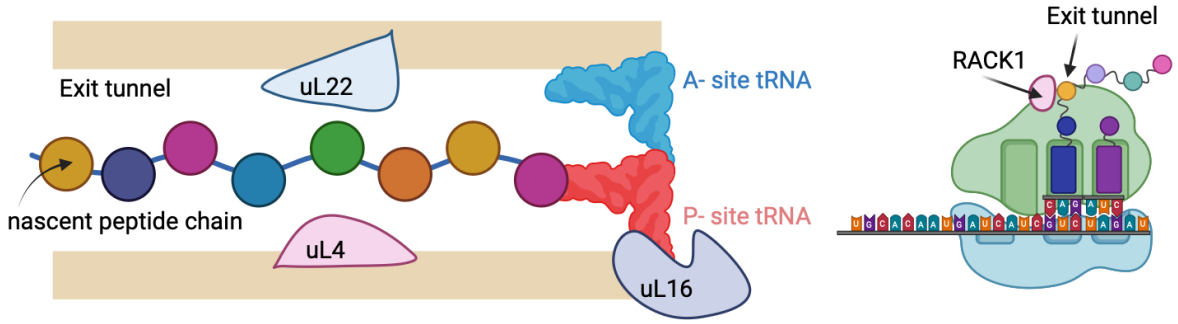
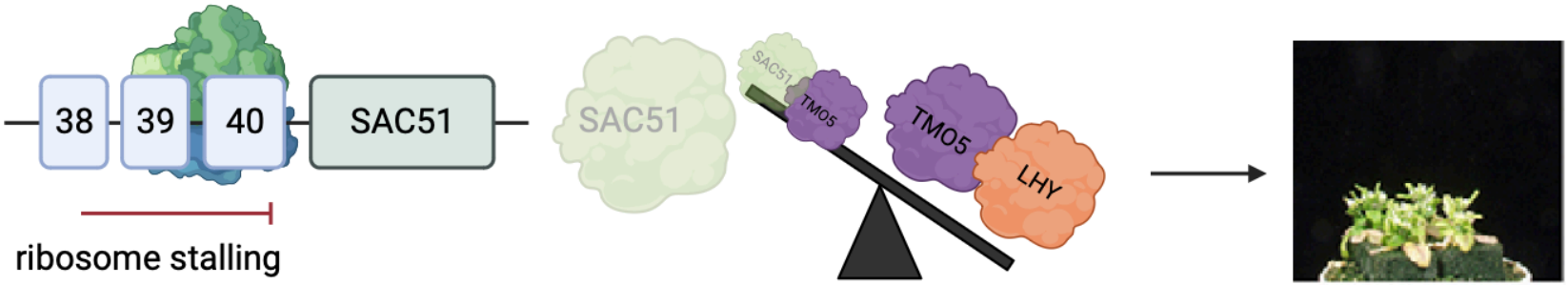


Figure 1.17. A schematic to show the location of the ribosomal proteins identified by Imai *et al.*, 2006 (37)**.** On the left is a diagram of the ribosomal exit tunnel. The nascent chain extends into the exit tunnel away from the tRNA sites and PTC. The nascent peptide chain is represented by coloured circles on a chain. The A and P tRNA sites are represented by blue and red tRNAs, respectively. uL4 (red) forms a constriction in the exit tunnel with uL22 (blue). uL16 is found around the PTC and has a finger-like projection into the PTC. Finally on the right, RACK1 (pink) is located around the exit of the exit tunnel. The RACK1 protein is overlayed on a translating ribosome.

Building upon the model proposed for how a mutation in CPuORF 40 suppresses the *acl5-1* phenotype, mutations in the ribosome exit tunnel may also prevent CPuORF led ribosomal stalling (Figure 1.17-18). The suppressor mutations identified in *ul4z-d, ul16-d* and *rack1-d* potentially alleviates stalling by the CPuORFs upstream of SAC51 mORF. Furthermore, as the ribosome cannot stall at those CPuORFs this facilitates the production of the SAC51 mORF protein under the low thermospermine *acl5-1* background. Consequently, SAC51 can compete with LHY to form a heterodimer with TMO5 and contribute to the partial rescue of the *acl5-1* phenotype (Figure 1.18) (119).

acl5-1



acl5-1 ul4z
acl5-1 ul16z
acl5-1 rack1z

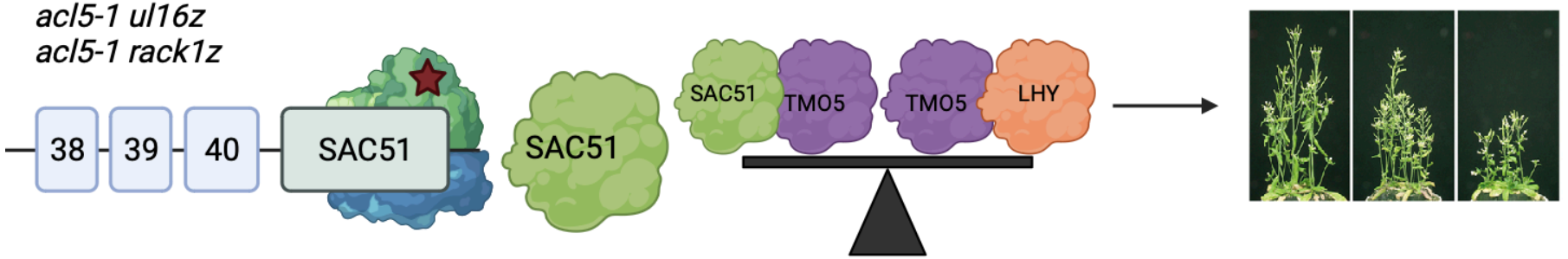


Figure 1.18. A diagram to show the effect of *acl5-1 ul4z, acl5-1 ul16z* and *acl5-1 rack1z* on the relative abundance of SAC51:TMO5 and TMO5:LHY. As in figure 1.16, the acl5-1 mutant causes a larger proportion of TMO5:LHY and the *acl5-1* dwarf phenotype. The introduction of a mutation in *ul4z, ul16z* and *rack1z* prevents the ribosome from stalling at the SAC51 mRNA. The SAC51 mORF protein is then produced and re-establishes the balance between SAC51:TMO5 and TMO5:LHY. Consequently, in *acl5-1 ul4z, acl5-1 ul16z* and *acl5-1 rack1z* backgrounds there is a partial rescue of the *acl5-1* dwarf phenotype.

Overall, molecular studies (Table 1.4) and the *acl5-1* suppressor screen (Figures 1.13-18) offers compelling evidence of a relationship between CPuORFs, the ribosomal exit tunnel, and ribosomal stalling (37,118,119).

1.9. CPuORFs conditionally modulate mORF expression.

This thesis has already introduced the evidence suggesting that CPuORFs respond to signals and conditionally stall the ribosome. Currently, there are twenty-five *Arabidopsis* CPuORFs that have been functionally characterized (Table 1.5), demonstrating their ability to conditionally modulate mORF expression (3,8–11,107,108,113,117,125–127). These CPuORFs respond to a diverse range of signals, including small metabolites, pathogen infections, and abiotic treatment (Table 1.5). Collectively, studies have identified distinct modes that CPuORFs can conditionally modulate mORF expression. CPuORFs can attenuate mORF expression in a sequence dependent manner by stalling the ribosomes (Table 1.4). Some CPuORFs stall or further stall the ribosome with the aid of a signal (Table 1.5) (rCUTS). Whereas other CPuORFs will increase mORF expression in response to a signal likely through facilitating translation re-initiation (Table 1.5) (aCUTS). This following section will discuss what is currently known about signal responsive CPuORFs and their aCUTS and rCUTS activity.

The CPuORF signal response is specific.

In vivo reporter assays have indicated that CPuORFs respond specifically to their respective signals (Table 1.5). CPuORF specificity has been demonstrated by measuring translation output in response to a wide range of signals from small metabolites to abiotic signals (3). Furthermore, specificity has also been investigated against similar metabolites. TBF1 and bZIP11 CPuORFs respond to galactinol, and sucrose and studies have shown that these CPuORFs respond specifically to these chemicals and not chemically similar saccharides (11,126–128). Demonstrating the resolution at which CPuORFs respond to their signals.

aCUTS CPuORFs				
Paper	CPuORF	signal	Notes	ref
Imai et al	SAC51	Thermospermine	GUS assay indicated 5'leader is responsive to thermospermine	(37)
Zhu et al.,	TBF1	Heat shock	GUS assays indicated heat shock releases attenuation of mORF by TBF1's CPuORF	(129)
Xu et al.,	TBF1	Pathogen shock	In planta luciferase assays show pathogen infection releases mORF attenuation. Utilized TBF1's CPuORF to enhance plant defence response to pathogen infection.	(128)
Pajerowska-Mukhtar et al.,	TBF1	Pathogen infection	GUS assays, sucrose gradients and western blots show pathogen infection increases mORF expression via CPuORF. Suggested phenylalanine starvation alleviates stalling.	(9)
Ishitsuka et al.,	SAC51 CPuORFs and SAC51-like	Thermospermine, heat	GUS assay indicated CPuORF 40 is functional. Thermospermine responsiveness is retained in *Brassica oleracea*, rice and soy. Western blots showed SAC51-like CPuORFs are heat responsive.	(8)
Li et al.,	WNK8	ABA	Western blot, phenotypic data, GUS stains and ribosome profiling data indicate CPuORF stalls the ribosome and is responsive to ABA. from transformant plants. Suggested relationship with RACK1z	(117)
Causier et al.,	6 *Arabidops*	Heat shock, mannitol	Luciferase assay show *Arabidopsis* and rice CPuORFs stall the ribosome. Implicated non	(3)

	is CPuORFs, 1 rice CPuORF	and thermosperm ine.	conserved amino acids in receptor-like signal response.	
Xu et al.,	TBF1	Pathogen infection	Utilises signal responsive CPuORFs to make pathogen resistant rice.	(13)

rCUTS CPuORFs				
Paper	CPuORF	signal	Notes	ref
Wiese et al .,	bZIP11	Sucrose	GUS assays demonstrate sucrose dependent stalling. Indicated conserved amino acids confer sucrose dependent stalling and not variable amino acids.	(130)
Hanfrey et al .,	AdoMetDC	Polyamines	Transient expression assays in tobacco.	(131)
Hanson et al .,	bZIP11 and HG1	Sucrose	CPuORF Homology group 1 are all sugar responsive	(127)
Ivanov et al .,	AdoMetDC	Polyamine	Transient expression assays and toeprint analysis. Penultimate proline always conserved.	(113)
Rahmani et al.,	bZIP11	Sucrose	Transient expression assays demonstrate sucrose mediated stalling via a CPuORF. Variable N-terminus doesn't confer function.	(109)
Ma et al.,	bZIP11	Sucrose	Overexpression of bZIP11, sucrose doesn't rescue growth inhibition phenotype.	(126)
Zhu et al.,	TBF1	Galactinol	GUS assays galactinol specificity and mediated repression. Demonstrates a self-contained regulatory unit and retains functionality outside native context. Conserved region is responsive to galactinol.	(129)
Uchiyama-kadokura et al.,	AdoMetDC	Polyamines	Western blots indicate conserved region is functional. Ribosomes stall with a serine in the P site and stop codon in the A site.	(10)
Guerrero-gonzalez et al.,	PAO2	Polyamines	Histochemical GUS stains demonstrated polyamine induced repression of mORF.	(132)
Laing et al.,	GGP	Ascorbate	LUC assays demonstrate GGP CPuORF is responsive to ascorbate in *Arabidopsis* and in Kiwi. Identified conserved region and leucine confer functionality.	(125)
Guerrero-gonzalez et al.,	PAO2	Polyamines	PAO2 conditional response demonstrated in onion cells.	(133)
Yamashita., 2017	bZIP11 CPuORF	Sucrose	Western blots showed the conserved region confers functionality. Stop codon placement is irrelevant to function. Investigated arginine codon usage. Alanine scanning assay	(11)

			implicated arginine as a functional amino acid.	
Bazin et al.,	CRF10, ARF4, PHO2	Phosphorous	Ribosome profiling and ribo-seq data under different phosphorous concentrations induced stalling at the CPuORF.	(107)
Alatorre de cobos et al.,	XPL1	Phosphocholine	GUS assay to phosphocholine and choline repressed mORF translation via CPuORF	(134)
Van der horst et al., 2023	bZIP11	Sucrose	Cryo-EM structure of sucrose-CPuORF stalled ribosome. Sucrose stalls the ribosome by interacting with a conserved metabolite pocket in the exit tunnel. Sucrose interacts with the exit tunnel and peptide and stalls the ribosome during translation termination. Metabolite pocket may be utilised for other signals such as the bacterial TnaC tryptophan dependent stalling. Suggesting a conserved mechanism in which the ribosome is a metabolite sensor.	(110)

Table 1.5. Summary of key studies identifying CPuORFs that conditionally express their downstream mORF. The table has two sections 1) Studies that have identified a signal that increases mORF protein expression via a CPuORF (aCUTS). 2) Studies that have identified a signal that represses mORF protein expression via a CPuORF (rCUTS).

aCUTS

Transient expression assays and western blotting have demonstrated aCUTS CPuORF functionality (Table 1.5). Furthermore, studies detailing this mode of CPuORF regulation have identified a wide range of signals from metabolites, abiotic signals, and pathogen infection (Table 4.2). The ability to conditionally activate and increase mORF translation has been demonstrated *in vitro* and *in vivo*. The CPuORF conditional activator response has been validated in *Arabidopsis*, rice, soybean, onion showing this mechanism is functional beyond *Arabidopsis* (3,8,9,13,37,117).

Transient expression assays of homologous aCUTS CPuORFs in *Arabidopsis* and rice suggest a role for non-conserved amino acids in signal sensing (3). *Arabidopsis* CPuORFs found within the same homology group (HG17) that share the same C-terminus but have divergent N-termini, respond to different signals (3). These signals are mannitol (simulating drought or water limitation) and heat shock. Interestingly, homologous CPuORFs in rice when expressed in *Arabidopsis* did not respond to either mannitol or heat shock despite evidence that the homologous rice CPuORF is capable of sequence dependent stalling (3). If the rice CPuORF does respond to a signal, it has yet to be identified.

rCUTS

Transient expression assays, western blotting, cryo-EM and ribosome profiling have demonstrated rCUTS CPuORF functionality. rCUTS CPuORFs can cause or enhance a CPuORFs repression of its downstream mORF (Table 1.5). Interestingly, all repressive CPuORFs identified to date respond to metabolites.

Bazin *et al.,* have demonstrated the utilization of ribosome profiling data in identifying functional CPuORFs (107) (Table 1.5). By measuring ribosome occupancy at CPuORFs in different conditions (high and low phosphorous concentrations), ribosome profiling data can identify conditional stalling events (31,107). Specifically, ribosome profiling data can be utilised to functionally characterise CPuORFs by their signal response. Importantly, ribosome profiling data can identify functional CPuORFs that respond to agronomic or commercially useful stimuli.

Van der horst *et al*., have presented a 3.7 Å structure of a bZIP11 CPuORF stalled ribosome upon the application of sucrose (110). The map suggests that sucrose stalls the ribosome alongside the conserved residues and stop codon of the bZIP11 nascent CPuORF peptide (110). Furthermore, sucrose seems to be sequestered by a conserved pocket that has been reported to bind to tryptophan in bacterial translation arrest peptides like TnaC (135–137). This pocket may be utilised by other repressive CPuORFs to stall the ribosome (20,110,135). It is possible that this pocket may function as a metabolite sensor to conditionally regulate gene expression at the translational level. Interestingly, sucrose induced stalling occurs during termination upon recognition of a stop codon(110). Sucrose binding inhibits the activity of eRF1 during termination, causing the breakdown of translation termination and ribosome stalling (110). This study is the first comprehensive investigation into how repressive CPuORFs can sense signals.

TBF1 a dual CUTS CPuORF.

Finally, the CPuORF upstream of TBF1 is unique as it is the only CPuORF that has been functionally characterised to use all modes of CPuORF function (9,128,129). It has been shown to attenuate mORF expression in a sequence dependent manner, repress mORF translation (by galactinol) and activate mORF translation (by pathogen infection and heat shock) (9,128,129). It is yet to be validated If other CPuORFs can both activate and repress mORF expression or if it is specific to the TBF1 CPuORF. TBF1 is a master molecular switch that induces a change from a vegetative to a defence metabolism in plants (9). TBF1's distinct mechanism may have arisen from a selective pressure evolutionarily to accommodate the complex regulatory demands of its mORF.

1.10. CPuORF classification and a potential modular nature.

CPuORF classification may indicate of a modular organisation of CPuORF function. Plant CPuORFs were first classified by Hayden and Jorgensen, into Class I and Class II (Figure 1.12). CPuORFs demonstrated different patterns of conservation that suggested the potential for deviating mechanisms (4). Class I CPuORFs have exclusive C-terminal conservation, while Class II CPuORFs comprised of CPuORFs with both exclusive N-terminal conservation and those with conservation throughout (both N- and C-termini) (4).

This classification system may provide insights into a modular nature of a selection of CPuORFs. Specifically, the conserved amino acids confer the nascent CPuORF peptide's ability to stall the ribosome during translation. This is evidenced by investigations into ribosome profiling data by Causier *et al.,* and reporter gene assays (3,10,107,117) (Table 1.5). Causier *et al*., correlated ribosome stalling events occurring on the conserved regions of Class I, IIa and IIb CPuORFs. Reporter gene assays have demonstrated in several CPuORFs that when conserved amino acids are mutated this disrupts a CPuORFs attenuation of the mORF resulting in a higher production of mORF protein.

Regarding signal sensing, Causier *et al.,* showed that a selection of homologous aCUTS CPuORFs (CPuORFs that have the same conserved region) can all stall the ribosome during translation but respond specifically to different signals (3). The study investigated two aCUTS CPuORFs that responds to specifically to heat shock and another aCUTS CPuORF that responds to water limitation (3). The fact that these three CPuORFs all share

the same conserved region but can respond to completely different abiotic signals may suggest that the variable amino acids have a role in signal sensing (3). If this is the case, aCUTS CPuORF function could potentially be modular and have potential in biotechnology to create bespoke regulatory peptides.

To elucidate the significance the position of amino acid conservation in CPuORFs the molecular mechanisms amongst CPuORFs classes should be investigated by molecular and structural techniques. A comparative analysis of the structures of all three CPuORF classes and their stalled ribosomes could determine whether the ribosomes stall exclusively at the conserved regions. This can be further investigated by analysing ribosome profiling data and correlating ribosome occupancy to conserved amino acids.

1.11. Known ribosome stalling and translation arrest mechanisms.

Conditionally stalling the ribosome during translation is not a mechanism unique to plant CPuORFs. In fact, various peptides across bacteria and eukaryotes have been demonstrated to stall the ribosome during translation through structural biology and cryo-EM (20). Translation arrest peptides have been shown to interact with conserved ribosomal exit tunnel proteins and conserved rRNA species (20,111,135,138,139). A review of cryo-EM data of stalled ribosomes has indicated there is an ancient mechanistic framework to stall the ribosome during translation (20,110). This section will discuss the evidence for a common mechanistic framework for ribosome stalling through exit tunnel rRNA nucleotides and proteins, exit tunnel electrostatic charge and the

conserved exit tunnel metabolite pocket implicated in a CPuORF signal sensing mechanism.

A comparative analysis performed by Ito and Chiba suggested that arrest peptides stall the ribosome by interacting with common exit tunnel elements including exit tunnel proteins, exit tunnel rRNA nucleotides, the exit tunnel constriction, and the PTC (Figure 1.19) (20,110). In both eukaryotes and bacteria, arrest peptides interact with a constriction a third of the way through the tunnel away from the PTC (20,111,138–140) (Figure 1.19). The constriction is formed by uL4 and uL22 in bacteria and uL4 and uL17 in eukaryotes. In addition, 23S rRNA has also the same conserved nucleotides (A2058, A751, U2609, A2062 and U2585) that have been shown to interact with the arrest peptides across kingdoms (20,111,135,138–140).

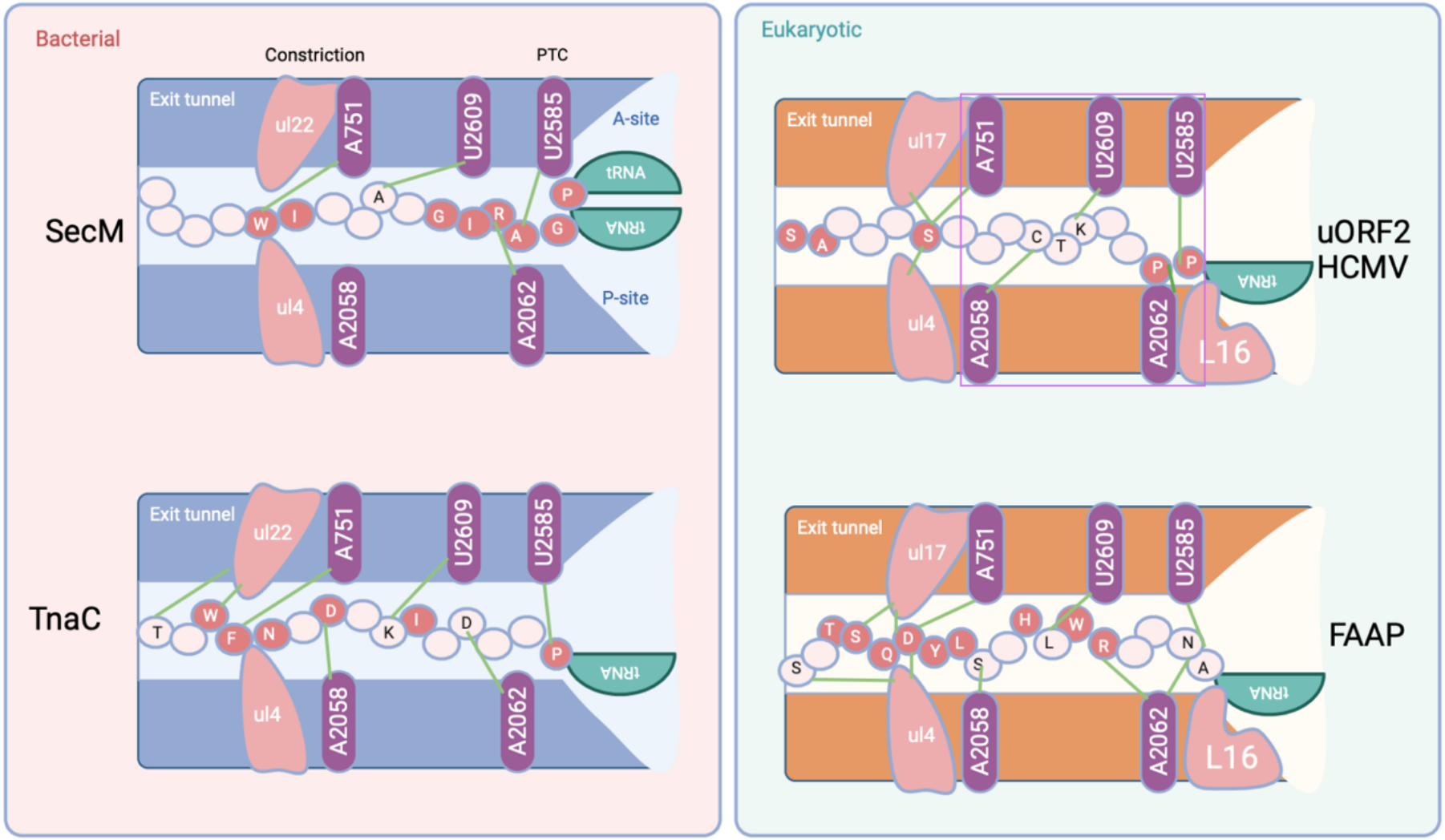

Bacterial
Constriction
PTC
Exit tunnel
ul22
A751
U2609
U2585
A-site
tRNA
SecM
W
I
A
G
I
R
A
P
G
tRNA
ul4
A2058
A2062
P-site
TnaC
T
W
F
N
D
K
I
D
P
Eukaryotic
Exit tunnel
ul17
A751
U2609
U2585
S
A
S
C
T
K
P
P
tRNA
uORF2
HCMV
ul4
A2058
A2062
L16
Exit tunnel
ul17
A751
U2609
U2585
T
S
S
Q
D
Y
L
S
H
L
W
R
N
A
tRNA
FAAP
ul4
A2058
A2062
L16

Figure 1.19. A schematic adapted from Ito and Chiba to represent interactions between bacterial and eukaryotic arrest peptides as determined by cryo-EM and molecular studies (20,111,135,139)**.** Ribosomal proteins are illustrated in pink, rRNA nucleotides in purple, and amino acids in either red (indicating evidence for loss of function mutations) or white. Green lines represent interactions between amino acids and ribosomal components. Ribosomal features such as the peptidyl transfer centre (PTC), A and P sites, and the constriction tunnel formed by uL4 and uL22 in bacteria and uL17 in eukaryotes are labelled. The position of tRNA during stalling is also indicated. Panels showcase specific arrest peptides: bacterial SecM and TnaC from *E. coli* (red panel), and eukaryotic uORF2 from human *Cytomegalovirus* and FAAP from *N. crassa* (green panel). This schematic is not to scale, and the nascent peptide structure is simplified, not representing actual secondary structures during stalling in the exit tunnel. *E. coli* numbering is used for rRNA nucleotide numbers.

Specific exit tunnel proteins have been implicated in arrest peptide and CPuORF mechanisms. A summary of ribosomal proteins that have been identified in the stalling mechanisms of arrest peptides can be found in Table 1.6.

The exit tunnel constriction: uL4 and uL22.

The constriction formed by uL4 and uL22 is known to interact with nascent arrest peptides and including nascent peptides that can stall the ribosome during translation (Table 1.6) (Figure 1.19, Table 1.6) (20,38). Cryo-EM and mutagenic investigations into the constriction have evidenced its role in translation kinetics and regulation by interacting with nascent translation arrest peptides. For example, mutating residues of uL4 and uL22 that extend into the exit tunnel to form the constriction results in less ribosome stalling (20,139,141–143). Furthermore, ribosome footprinting has demonstrated that the exit tunnel constriction is affected by abiotic signals such as heat shock(144), Specifically, under heat shock, eEF1A and heat shock proteins narrow the constriction between uL4 and uL22 and causes more ribosome stalling events (142).

uL16

uL16 is a ribosome protein that extends into the exit tunnel within proximity to the PTC (33,120,142). Cryo-EM data has demonstrated that this protein is involved in the ribosome stalling mechanisms of an *Arabidopsis* CPuORF (bZIP11), the fungal arginine attenuator peptide and the uORF2 of the human cytomegalovirus (110,139). Furthermore, mutagenesis suggests an important role for this exit tunnel protein in translation, ribosome integrity, and tRNA delivery to the PTC (20,37,118,120,121,138,139) (Table 1.6). Overall, structural biology and mutagenesis supports uL16's role in stalling regulation.

RACK1

The final ribosome protein we shall discuss is the highly conserved eukaryotic scaffold protein RACK1 (122,123). Cryo-EM data has suggested RACK1 facilitates stalling by stabilising collided ribosomes (124). When ribosomes stall on the mRNA this can cause ribosomes to stack and collide. Stalled ribosomes are stabilised by RACK1 interacting with ribosome proteins from the collided ribosome (124) When considering RACK1 was identified by the genetic screen of ACL5, this evidence may suggest that the *rack1-d* mutation from the Imai *et al* screen knocked out RACK1's ability to stabilise stalled ribosomes (Figure 1.18). Thus, overcoming ribosomal stalling on the SAC51 transcript (37,118). Overall, this is compelling evidence for RACK1's involvement in ribosome stalling by CPuORFs and arrest peptides.

Ribosome stalling and the exit tunnel.

Ribosome stalling in CPuORFs is sequence dependent, so how do amino acids confer stalling? Alongside cryo-EM, mutagenesis has explored the role of specific amino acids of arrest peptides. Amino acids such as arginine, proline, phenylalanine, and serine have been identified as functional in arrest peptides (135,138,139,145–147). The amino acids interact with exit tunnel ribosomal proteins and rRNA nucleotides (Figure 1.19). Mutagenesis has confirmed structural studies to show that they confer ribosomal stalling. It is unknown whether the biochemical properties and/or the position of the conserved amino acid confers function. Bhushan et al., suggested that proline's poor ability to form peptidyl bonds, compared to alanine and phenylalanine, may contribute ribosome stalling or pausing (139).

Another hypothesis is that hydrophobic and electrostatic interactions in the exit tunnel contribute to ribosome stalling. The exit tunnel is hydrophilic, and it is postulated that changes to the exit tunnel environment by nascent peptide charge and signals can initiate or alleviate stalling (148). Exit tunnel charge has been shown to affect known ribosome stalling peptides such as SecM (70,145,148). Investigations into SecM's function included an experiment where a hydrophobic sequence was fused to the N-terminus of SecM and it disrupted its ability to stall (70,145,148). To further investigate electrostatic interactions and stalling bioinformatic tools could quantify charge within the exit tunnel of known stalled ribosome cryo-EM maps.

Finally, cryo-EM studies of the sucrose responsive bZIP11 CPuORF, which stalls the ribosome during translation, has revealed a similar stalling mechanism to the bacterial TnaC (20,110). This study may provide evidence for a common mechanism as to how the ribosome exit tunnel functions as a metabolite sensor to regulate translation. Although the published map of the bZIP11-stalled ribosome is unable to resolve the interactions between the CPuORF and the entire exit tunnel, the map does illuminate a conserved mechanism to conditionally stall the ribosome (110). TnaC and bZIP11 CPuORF stall the ribosome through a conserved metabolite sensor pocket within the exit tunnel (110). Tryptophan and sucrose respectively interact with the peptides during translation. Stalling occurs when the stop codon is in the A- site (110). In TnaC, stalling causes the inhibition of eRF1 function, but it is unclear if eRF1 is recruited in bZIP11 stalling (110). In any case, the metabolite pocket enables tryptophan and sucrose to interact with exit tunnel proteins, rRNA nucleotides and the arrest peptide/CPuORF (110). Overall, suggesting that CPuORF stalling mechanism is conserved and analogous to that of reported stalling peptides.

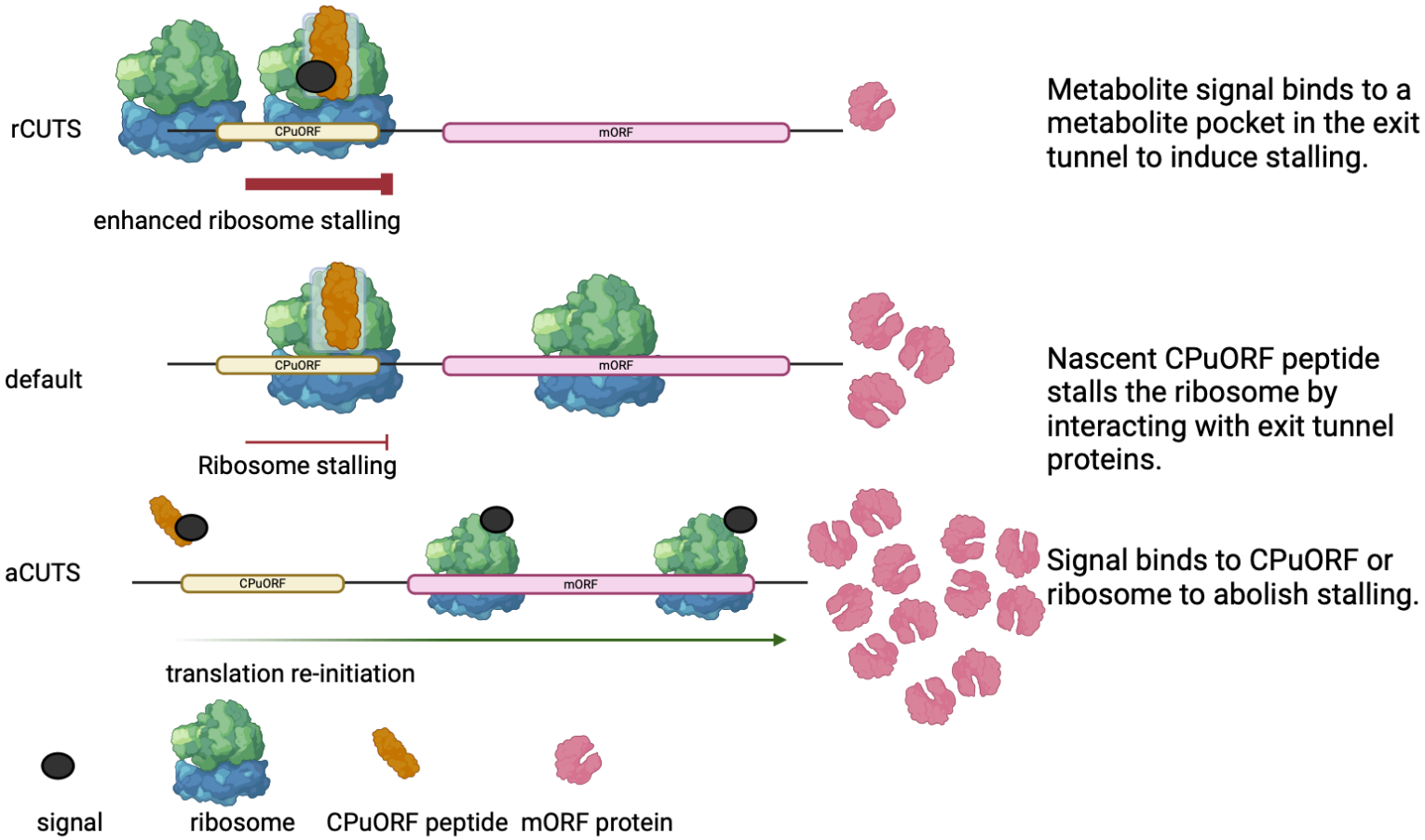
Protein	Notes	Ref
uL4	Forms the exit tunnel constriction alongside uL22. Heat shock, heat shock proteins and eEF1A narrows the constriction and causes a breakdown of elongation kinetics, causing global ribosome stalling or pausing. Mutations of residues that are within the exit tunnel can terminate stalling.	(20,37,111,118,138–144,149–152)
uL22	Forms the exit tunnel constriction alongside uL4. Mutations within exit tunnel residues can release stalling. Mutations within this protein can abolish stalling by SecM. Implicated in the rCUTS CPuORF (bZIP11) sucrose-dependent stalling mechanism.	(20,110,111,135,138–143,149,151,152)
uL16z	Located by the PTC and has a finger that extends into the exit tunnel. Implicated in eukaryotic stalling mechanisms of the fungal arginine attenuator peptide, uORF2 of the human cytomegalovirus and the plant CPuORF bZIP11. Mutations can lead to impaired ribosome integrity and the breakdown of translation.	(20,37,118,120,121,138–140,142,143,149,150,152)
RACK1	A scaffold protein between the large and small ribosome subunits and is located on top of uS17. RACK1 functions both on and off the ribosome, participating in various signal transduction pathways by interacting with signal molecules. Acting as a docking site for translation initiation factors and kinases, RACK1 can stimulate overall protein synthesis rates and control translation of specific mRNA subsets. RACK1 stabilises stalled and collided ribosome species and may be involved in ribosome quality control.	(18,20,99,118,122,123,151,153–161)

Table 1.6. Overview of ribosome proteins involved in ribosome stalling mechanisms. Table outlining the functional roles, structural features, and interactions of key ribosomal proteins implicated in ribosome stalling and associated cellular processes.

Structural biology is an effective discipline to elucidate the molecular mechanisms of ribosomal arrest peptides and CPuORFs. Cumulatively with molecular data, there is precedent to suggest that CPuORFs stall the ribosome and interact with proteins and rRNAs within the exit tunnel (3,37,110,118). Lastly, screens have been successful in identifying functional components within the exit tunnel, screens could potentially be used to identify downstream signal factors within the CPuORF mechanism (37,118).

1.12. Conditional uORF translational stalling (CUTS).

This thesis presents a model of CPuORF function known as Conditional uORF Translational Stalling (CUTS). CUTS is characterized by three modes: CPuORF-led ribosomal stalling (default), Repressive CUTS (rCUTS), and Activator CUTS (aCUTS) (Figure 1.20).



Metabolite signal binds to a metabolite pocket in the exit tunnel to induce stalling.

Nascent CPuORF peptide stalls the ribosome by interacting with exit tunnel proteins.

Signal binds to CPuORF or ribosome to abolish stalling.

Figure 1.20. A schematic representation of the CPuORF Conditional uORF Translational Stalling mechanism (CUTS) (3,8–11)**.** Repressive CUTS (rCUTS) illustrates when a specific signal enhances or triggers ribosomal stalling by binding to a conserved metabolite pocket in the exit tunnel, resulting in reduced mORF protein production compared to control conditions. The default: A CPuORF nascent peptide inherently stalls the ribosome during translation independent of a external signal. The nascent peptide likely forms interactions with exit tunnel proteins rRNA nucelotides, leading to decreased mORF protein production when compared to control conditions. Activator CUTS (aCUTS) demonstrates how a signal terminates ribosomal stalling by the nascent CPuORF peptide. The signal may bind to the ribosome and/or CPuORF peptide and enabling the ribosome to re-initiate translation at the downstream mORF and promoting mORF protein production.

Transient expression assays, western blots and ribosomal profiling data have revealed that CPuORFs can attenuate mORF expression in a sequence dependent manner by stalling the ribosome during translation (Table 1.3). Stalling likely occurs through interactions with conserved exit tunnel and nascent CPuORF peptide residues during translation (37,118). Not all CPuORFs inherently stall the ribosome during translation (11,126) (Table 1.3).

Repressive CUTS (rCUTS) is the mechanism in which CPuORFs repress mORF expression (Table 1.5, Figure 1.20). CPuORF repression occurs upon the detection of a specific signal. The signal can either cause ribosomal stalling or enhance a CPuORF's repression of its mORF (Figure 1.20). Signals likely induce or strengthen interactions between the nascent CPuORF peptide and exit tunnel. Evidence has shown that specific conserved amino acids confer an rCUTS CPuORF's conditional response (11) (Table 1.5). Furthermore, cryo-EM studies have shown that conserved metabolite pockets that can sequester a signal and induce stalling during translation termination by interacting with exit tunnel proteins, rRNA nucleotides and the nascent CPuORF peptide(110). All the currently characterised rCUTS CPuORFs respond exclusively to metabolites (Table 1.5).

However, it is unknown whether other mechanisms and/or signals can elicit rCUTS with uncharacterised CPuORFs.

Activator CUTS (aCUTS) is a mechanism whereby a signal terminates ribosomal stalling and facilitates translation re-initiation at a downstream start codon (Table 1.5, Figure 1.20). Studies indicate that conserved amino acids confer the ability to stall the ribosome and non-conserved amino acids confer aCUTS signal response (3). Notably, the aCUTS CPuORFs respond to a wide range of signals from metabolites, pathogens, and abiotic signals (Table 1.5). Despite current efforts, the aCUTS signal response is possibly induced by a conformational change by the CPuORF peptide and the weakening of the interactions between the nascent CPuORF and exit tunnel proteins (37,118) (Figures 1.10-1.20).

The CPuORF mechanism is yet to be elucidated. Future studies should generate high resolution cryo-EM structures of CPuORF-stalled ribosomes. A comparative approach across multiple CPuORFs should be undertaken to cover the breadth of diversity in CPuORF function. A comparative analysis should cover CPuORFs from all classes, types of signal response, mode of CUTS and stalling capabilities.

1.13. The CUTS mechanism and applications in biotechnology and agriculture.

Currently, only two studies demonstrate the applications of CPuORFs in agriculture. The TBF1 5'leader was utilized to generate pathogen resistant rice (13). Pathogen infection terminates ribosomal stalling by TBF1's CPuORF, facilitating translation re-initiation at

the mORF (9,13). The TBF1 CPuORF was placed upstream of the master immune regulatory gene NPR1 (AT1G64280) in *Arabidopsis* and rice (13). Remarkably, the resulting lines demonstrated an enhanced pathogen response without fitness cost (13). Plants utilizing TBF1's CPuORF demonstrated improved metrics for lesion length without a reduction in height in the laboratory and field, when compared to the wild type. In another example, CRISPR/Cas9 was utilized to edit a sucrose responsive CPuORF (bZIP1) in tomato (162). By editing the CPuORF sequence this affected the expression of its downstream transcription factor mORF and its targets (162). The resulting lines had altered composition of sweet and bitter amino acids without phenotypic costs (162). Overall, these studies demonstrate the potential of CPuORF and crop manipulation.

CPuORFs are found across taxa including crop species. Various agronomic signals have been shown to modulate mORF expression via the CPuORF CUTS mechanism (Table 1.5; Figure 1.20). Furthermore, CPuORFs can modulate mORF expression through repression or activation (both in the case of TBF1). Overall, the diversity of CPuORF stalling and signal response offers scope for crop enhancement with native regulatory features. Upon the characterisation of the CUTS molecular mechanism, bespoke CPuORFs could be synthesized. Bespoke CPuORFs could utilise the receptor like activity observed in aCUTS CPuORFs. Receptors for agricultural chemicals or signals have the potential to be fused to a stalling module of a CPuORF.

In the field of biotechnology and research, the inducible switch nature of CPuORFs and the CUTS mechanism can be leveraged in a multitude of ways. Transient expression assays have shown that it is a rapid switch mechanism to control any downstream ORF

(Tables 1.4-5). The CUTS mechanism could be used for the study of protein function and interactions, development of novel biosensors, investigation of signal transduction pathways, and in the design of controlled gene expression systems for synthetic biology. Finally, the diversity of CPuORFs offers researchers a multitude of CPuORFs that function in different ways in response to different signals (Table 1.5).

In summary, the CPuORF and CUTS mechanism have potential applications in biotechnology and agriculture (Table 1.5, Figure 1.20). Harnessing this potential could result in significant advancements in crop optimization and biotechnological research, paving the way for novel solutions to current challenges.

1.14. Aims and Objectives.

The overarching aim of this thesis is to elucidate the CPuORF CUTS mechanism using molecular and structural biology. Furthermore, this thesis will explore CPuORF DNA and amino acid sequences to illuminate functional modules.

Investigating the CPuORF CUTS Mechanism Using Molecular Techniques

- **Aim**: To understand the mechanisms by which CPuORFs stall the ribosome during translation and how they respond to specific signals.
- **Objectives**:
    - Create constructs placing CPuORF and mutant CPuORFs in an artificial context upstream of a luciferase reporter gene.

- Investigate the effect of CPuORFs and their mutations on mORF translation both *in vitro* and *in planta*.
- Analyse the interaction within *Arabidopsis* wild type and ribosome mutant backgrounds to delineate the stalling effect of CPuORFs.

Structural Analysis of the CPuORF Mechanism

- **Aim**: To perform a detailed structural analysis to understand how CPuORF peptides interact with ribosomal exit tunnels.
- **Objectives**:
  - Purify ribosomes from *Arabidopsis* and wheat germ extract for cryo-EM analysis.
  - Purify stalled ribosomes from wheat germ extract using an N-terminal tag for cryo-EM analysis.
  - Identify interacting residues between the CPuORF nascent polypeptide and ribosomal exit tunnel, offering insights into the structural interplay.

Creation and Investigation of an Expanded CPuORF Database

- **Aim**: To expand and investigate a comprehensive CPuORF database to elucidate functional patterns and commonalities across different species (Arabidopsis, Drosophila, and *Homo sapiens*).
- **Objectives**:
  - Develop an expanded CPuORF database, capturing detailed information across different species.
  - Investigate common sequences that could be deemed functional.

- Analyse ribosome profiling data to discover evidence of stalling in CPuORF-containing transcripts compared to regular uORF-containing transcripts.

Together, this will present a comprehensive study of CPuORFs, focusing on their molecular behaviour, structural interactions, and a broad-based examination across species. The findings aim to deepen our understanding of CPuORFs, shedding light on their essential role in gene regulation and offering a rich resource for future research in the field.

Chapter 2. Materials and methods

2.1. Computer resources.

2.1.1 Bioinformatic tools.

Genome databases	TAIR NCBI UniProt	https://www.arabidopsis.org https://www.ncbi.nlm.nih.gov https://www.uniprot.org CLU
Sequence alignment	ClustalX	http://www.clustal.org/clustal2/
DNA sequence analysis and primer design.	SnapGene	https://www.snapgene.com
Luminescence analysis	NightOwl by Bertholdt technologies	https://www.berthold.com/en/
Sequence analysis	LOGO analysis	https://weblogo.berkeley.edu/logo.cgi
Gene ontology analysis	GOrilla REVIGO	https://cbl-gorilla.cs.technion.ac.il http://revigo.irb.hr
Ribosome profiling data	GWIPSviz Ribogalaxy UCSC browser	https://gwips.ucc.ie https://ribogalaxy.genomicsdatascience.ie https://genome.ucsc.edu
R packages	Devtools Seqinr biocmanager	https://cran.r-project.org/web/packages/devtools/index.html https://cran.r-project.org/web/packages/seqinr/index.html https://cran.r-project.org/web/packages/BiocManager/index.html

Table 2.1. Bioinformatic resources used in this thesis. The table details the bioinformatic tools used and their purpose. In the final column is a link to the website for the respective tool.

2.1.2. Database creation and analysis.

Data from bioinformatic studies (Table 1.3) were compiled into an Excel database (Figure 5.1). The database has been made publicly available at

https://theeukaryoticcpuorfdatabase.github.io. Data included DNA and protein sequence, homology group, gene identifier, and CPuORF class. Mode of CUTS was compiled from published literature (Table 1.5).

2.1.3. Pipelines.

2.1.3.1. CPuORF classification pipeline.

An R pipeline was created to classify CPuORF peptide sequences based upon conservation distribution patterns (Figure 5.2). C terminal conserved CPuORFs are class I, N-terminal conserved CPuORFs are Class IIa and CPuORFs with conservation throughout are Class IIb. CPuORF alignments were provided by Takahashi et al (105,106). The R packages used include devtools, seqinr and BioManager (Table 2.1). As input, CPuORF peptide alignments in FASTA format were used. The pipeline stored peptide sequence data as a matrix and gave every amino acid a conservation score. Next, the CPuORF peptide matrix was split in half to represent the C and N-termini. The conservation scores for each respective half were summed and a ratio between the N and C-termini was generated to determine the relative conservation between the C and N termini. CPuORFs that were conserved at the C or N terminus would be classed as I and IIa respectively. A CPuORF conserved throughout would be classified as IIb.

2.2. General growth media.

The following growth media were made accordingly. 1) LB media: tryptone 10 g/l, yeast extract 5 g/l, NaCl 10 g/l and Agar 15 g/l (solid media only). 2) 0.5 MS media: 2.2 g/l (Sigma: M5524), 1 % sucrose and Agar 15 g/l (solid media only). Treatments were applied with the

addition of either 0.05 mM thermospermine, 300 mM mannitol, 37 ºC incubation and 560 mM sucrose.

2.3. Plant material and growth conditions.

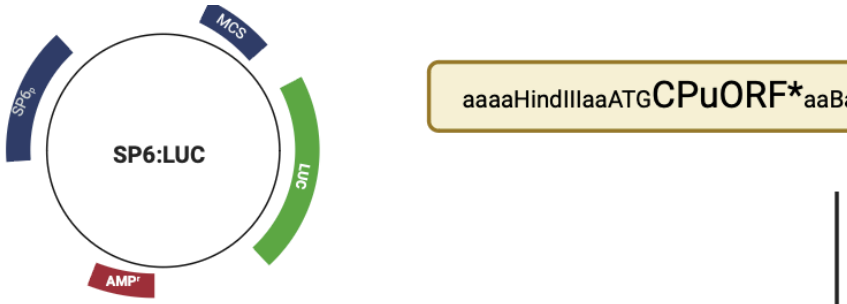
Arabidopsis thaliana lines were grown on peat-free soil or on 0.5 MS plates (with or without appropriate treatments as described) at 21°C, under long days (16h light/8h dark). *Arabidopsis* lines investigated in this study are from *Landsberg erecter* (*Ler*) and Columbia-0 (*Col-0*) backgrounds. Also, RPL18:GFP:HIS:HA tagged *Arabidopisis* lines and ribosome mutant lines (*ul16z, rack1z* and *ul4z*) from the Takahashi group (known formerly as *sac52-d, sac53-d* and *sac56-d*) (37,118,163).

2.4. Cloning.

2.4.1. Construct design and cloning plan.

Luciferase assays were performed *in vitro* (Promega wheat germ extract) and *in vivo* (*Agrobacterium* transformed *Arabidopsis*) systems (Figures 2.1-2). Wild type and mutant CPuORF sequences were synthesised by Genewiz (https://www.genewiz.com/) or IDT (https://eu.idtdna.com/). Oligonucleotides and primers (Sequences are found in section 2.4.2) were synthesised. CPuORFs were synthesised with flanking restriction sides HindIII and BamHI (Figure 2.1-2). CPuORFs and the vector supplied by Promega's wheat germ extract kit (SP6:LUC) (Figure 2.1) were digested using Thermo Fisher fast digest enzymes (Thermo Fisher: FD0504 and FD0054, respectively). The digested CPuORF fragments were ligated into SP6:LUC with t4 DNA ligase (Thermo Fisher EL0011). 2 µl of ligated DNA was transformed into electrocompetent *E. coli* cells (Transgen CD201-02). *E. coli* was incubated overnight on LB + 100 ug/µl ampicillin media. Transformant colonies

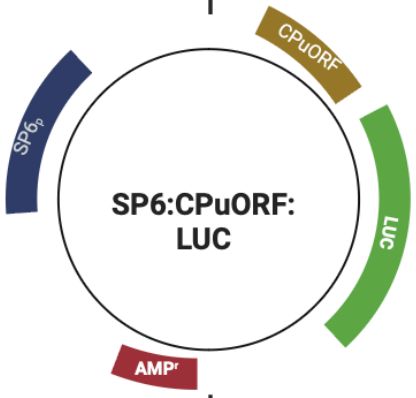
were selected and incubated in LB + 100 ug ampicillin broth overnight at 37 ºC. Overnight cultures were mini prepped using EasyPure® Plasmid MiniPrep Kit (Transgen EM101-02) and eluted in 40 µl of elution buffer. DNA concentration was measured with the Thermo Fisher's nanodrop Lite (ND-LITE-PR) and sequenced using Genewiz sanger sequencing. DNA sequences were aligned and confirmed using Snapgene. Glycerol stocks with 50% glycerol were made and stored at -70 ºC.



aaaaHindIIIaaATGCPuORF*aaBamHIaaaa

1) Wild type and mutant CPuORF DNA sequences are synthesised.

2) CPuORF double stranded DNA fragments and SP6:LUC plasmid DNA are digested with restriction enzymes. The CPuORF DNA fragment is ligated into a multiple cloning site (MCS) upstream of a luciferase reporter gene in SP6:LUC. The ligated plasmid is transformed into E. coli. Successful transformants are selected via Ampicillin resistance, colony PCR and sequencing.



3) Used for *in vitro* luciferase assays and western blotting analysis.

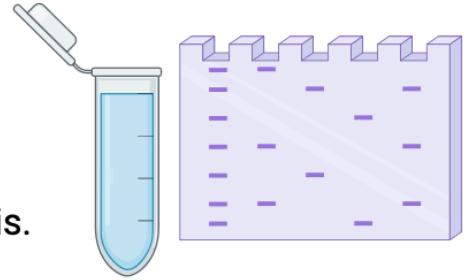
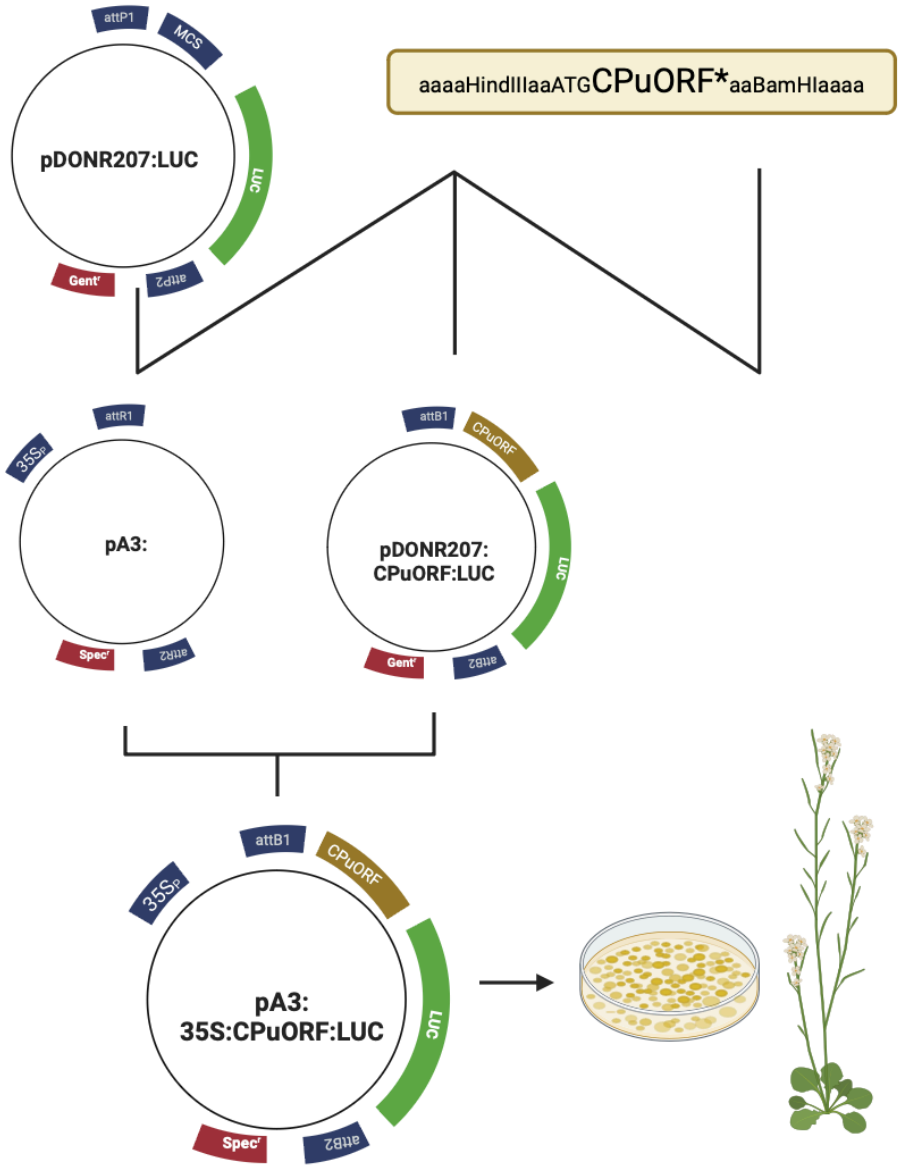


Figure 2.1. Simplified schematic to show cloning plan for *in vitro* luciferase assays and western blotting. Synthesised CPuORF DNA sequences and SP6:LUC plasmid DNA were digested with restriction enzymes to insert a CPuORF upstream of a *LUCIFERASE* reporter gene via a multiple cloning site (MCS). Ligated plasmids were transformed into chemically competent E. coli and successful transformants were selected by ampicillin resistance. Successful clones were confirmed by colony PCR, restriction digest and sequencing. Plasmid DNA was extracted from liquid culture using the EasyPure® Plasmid MiniPrep Kit (Transgen EM101-02). Plasmid DNA was used as a template to generate capped mRNA for i*n vitro* analysis.

For *in planta* systems CPuORFs were amplified by PCR to be flanked by EcoRI and XhoI or HindIII or BamHi restriction sites (Figure 2.2). A attB1:MCS:LUC:AattB2 DNA fragment was synthesised by IDT and gateway cloned into pDONR207 to create the pDONR207:LUC plasmid (Figure 2.2). Positive transformants were selected for by gentamycin resistance (10 ug/µl) on LB plates. Successful transformants were identified by colony PCR and confirmed by Genewiz sanger sequencing. CPuORFs were ligated into MCS as detailed in Methods 2.4.1. Successful pDONR207:CPuORF:LUC were then Gateway cloned into the entry vector pALLIGATOR3 (Figure 2.2) plasmids with LR Clonase II (11791020). Successful transformants were selected by (75 ug/µl) spectinomycin resistance and confirmed by gel electrophoresis or colony PCR. 100 ng of entry vector plasmid DNA was transformed into electrocompetent CS301 *Agrobacterium*. Successful *Agrobacterium* transformations were identified through antibiotic resistance of 10 ug/µl gentamicin, 75 ug/µl spectinomycin and 50 ug/µl rifampicin. Glycerol stocks were stored at -70 ºC. pA3:35S:CPuORF:LUC in *Agrobacterium* was grown as a liquid culture. *Arabidopsis* was transformed using the floral dip method (Methods 2.5) and luciferase assays were performed on T1 leaves.



1. Wild type and mutant CPuORF DNA sequences are synthesised.
2. CPuORF DNA fragments and pDONR207:LUC plasmid DNA were digested with restriction enzymes. The Digested CPuORF sequence is then ligated into the plasmids MCS (multiple cloning site).
3. Ligated plasmids were transformed into E. coli and successful transformants were selected by gentamycin resistance, restriction digest, colony PCR and sequencing.
4. Transformant pDONR207:CPuORF:LUC DNA is purified from liquid culture. CPuORF:LUC is LR gateway cloned into the entry vector pA3 (pALLIGATOR3) to generate pA3:35S:CPuORF:LUC through *E.coli* transformation and spectinomycin resistance (as in step 3).
5. Confirmed pA3:35S:CPuORF:LUC plasmid DNA is transformed into Electrocompetent *Agrobacterium*.
6. *Arabidopsis* plants are transformed via the floral dip method.
7. Successful transformants are selected by GFP seed coat selection.

→ Luciferase assays on T1 leaves.

Figure 2.2. Simplified schematic to show cloning plan for *in vivo* luciferase assays. Synthesised CPuORF DNA sequences and pDONR207:LUC plasmid DNA were digested with restriction enzymes to insert a CPuORF upstream of a *LUCIFERASE* reporter gene via a multiple cloning site (MCS). Ligated plasmids were transformed into chemically competent *E. coli* and successful transformants were selected by gentamycin resistance. Successful clones were confirmed by colony PCR, restriction digest and sequencing. Plasmid DNA was extracted from liquid culture using the EasyPure® Plasmid MiniPrep Kit (Transgen EM101-02). The CPuORF:LUC fragment was then LR gateway cloned into the entry vector pALLIGATOR3 (pA3). Doing this would place the CPuORF:LUC fragment under the control of a 35S constitutive promoter. Successful pA3:35S:CPuORF:LUC clones were selected for by spectinomycin resistance and confirmed by colony PCR. Confirmed entry vector clones were then transformed into electro competent *Agrobacterium*. Liquid cultures of confirmed pA3:35S:CPuORF:LUC in *Agrobacterium* were used to transform Arabidopsis using the floral dip method. Successful *Arabidopsis* transformants were selected through GFP seed coat selection. T1 plants and leaves were used for luciferase assays.

2.4.2. Primers and sequences.

attB1F	Primer	aaaaGGGGACAAGTTTGTACAAAAAAGCAGGCT	Colony PCR
attB2R	Primer	aaaaGGGGACCACTTTGTACAAGAAAGCTGGGT	Colony PCR
attL1F	Primer	aaaaTCGCGTTAACGCTAGCATGGATCTC	Colony PCR
attL2R	Primer	aaaaGTAACATCAGAGATTTTGAGACAC	Colony PCR
HindIIIF	Primer	aaaaGAAAAAGCTTAACAGACCACCATG	Ligation and colony PCR
BamHIR	Primer	aaaaGAAAAATTGGGGATCGTCA	Ligation and colony PCR
38F	Primer	aaaaGAAAGAATTCACAGCATGTGTATTGCCGTATACCG	Ligation PCR
38R	Primer	aaaaGAAAAAGCTAGAGTATCACACGGCAATACAG	Ligation PCR
39F	Primer	aaaaGAAAGAATTCCGGTATATTATAAACACTTTTGAAAATG	Ligation PCR
39R	Primer	aaaaaGAAAAAGCTTAGCTGATTCTTTGTTCACAATGG	Ligation PCR
40F	Primer	aaaaGAAAGAATTCTCAAGATTTCCTCAAAGTTATGG	Ligation PCR
40R	Primer	aaaaGAAAAAGCTTACGTGACTGGTTTGAGTAGAAGTC	Ligation PCR
39deleteF	Primer	aaaaTTTCAAAAGTGTTTATAATATACC	SDM PCR
39deleteR	Primer	aaaaTAGACTAAAAGCTCACTC	SDM PCR
Att-MCS:LUC-att	Gene fragment	acaagtttgtacaaaaaagcaggctaGAATTCaCTCGAGaAAGCT TaGGATCCaggtaccatctagaaTGGTTGGTGACATTTCGCTG CATTTTTCAATCTGTGATTGTTTTTCGTTCGTCTTTCTTTTACT ATTTTCTCGAAAAGGACACAAGAAGTATTGCATTCACTCA GTTGAGCAACTTAACAATCGTGTTGTACTTTTTGAAGTTCC CTTGAGCTAAACTGCTAAGAGCATGGAAGACGCCAAAAA CATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAG ATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGA GATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGC ACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAA ATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGG CTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACT CTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATC GGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAA CGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACC GTGGTGTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGA	Backbone

		ACGTGCAAAAAAAGCTCCCAATCATCCAAAAAATTATTAT CATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATG TACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGA ATACGATTTTGTGCCAGAGTCCTTCGATAGGGACAAGACA ATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGC CTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGA GATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATC ATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCA CGGTTTTGGAATGTTTACTACACTCGGATATTTGATATGTGG ATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTT TCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCT GCTGGTGCCAACCCTATTCTCCTTCTTCGCCAAAAGCAC TCTGATTGACAAATACGATTTATCTAATTTACACGAAATTGC TTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAG CGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAA GGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTA CACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAA AGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATA CCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACTG TGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAA TCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGAT GGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACG AACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATTAA GTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATC CATCTTGCTCCAACACCCCAACATCTTCGACGCAGGTGT CGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCG CCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACG GAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACA ACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGA CGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAA GAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGC GGAAAGATCGCCGTGTAAacccagctttcttgtacaaagtggt	
SAC51	sequence	TCTTTAGAAATATTTCGTAAAAAAAAATTCTCCGTTTAGACA TTATTGTTCGTTGGTCACTTTCATCTCTCTTATCCCTTTGCTT TTCCATTGATTCAGCTCAAAAAAAAACAAAAAAGTCTCATC GTTTTCTCTGGATCTGACCGAAAGATCATTCCTTTCTAAGA TACTAAAGTAGGTTTGAGTGATCAGAATCAATTAGAGTGTG TATAAGAGGAAGAGAAAGAGAAAACAGCATGTGTATTGCC GTATACCGTAAAGTTTTGAGCTTGAATCTGTATTGCCGTGT GATACTGTAGATTGTTTCACCTTTGAAGCAACACCTTCAAG AGTCAAATCCTTTTCTTTCTTATTCATCTTCCCTTCTTTTTCG GTATATTATAAACACTTTTGAAAATGGGATCACAACATTTCT ATGTGTACCATTGTGAACAAAGAATCAGCTAGACTAAAAG CTCACTCTCCTTCCACAAACAAAACTTTCCTCCTTTTGGTT TTGGTATCTGTCTACTTTCTTCAATCAAGATTTCCTCAAAGT TATGGTGTGCCAATCACCTGGTAAGACAAGATTTCGAGGA TTGAAGTACGAGACCGGAAACGCTAACGAATCAACAATT GTAGTGAGAGTTATTGAATGCTATCAACCGATGGATAATTG CCAGGCTGAGTACTTCAGACTTCTACTCAAACCAGTCAC GTAGTTGGTTGGTGACATTTCGCTGCATTTTTCAATCTGTG ATTGTTTTTCGTTCGTCTTTCTTTTACTATTTTCTCGAAAAGG ACACAAGAAGTATTGCATTCACTCAGTTGAGCAACTTAAC AATCGTGTTGTACTTTTTGAAGTTCCCTTGAGCTAAACTGC TAAGAGC	CPuORF
eIF5	Sequence	GAAAAAGCTTAACAGACCACCATGATGTCTGAACAGGCT TCTCTCTTTCAATCATCTATCCGGCTTTTAGGAGAGTACTG TCCTATATCTAGGAAAAAGTCCACTCATAAAGTTGGTGGTT	CPuORF

		TGCGTTTTGAGGAGGACATCACGAGCTGTATTGGATTTCA GTTTATGCAACTGAGTGATTATCACTGCCTCTTCtgaCGGA TCCCCAA	
Heat	Sequence	GAAAAAGCTTAACAGACCACCATGATGGATCTTTTGATGA TACAGAGACGTGAATCGAGAAGGGTCTGGTTAGGTGTAC AAAATTCAATAGTTCTCCTCGTGGGTGGCAACCATACCTC CTCCAGGTTTTGCACTCGCtgaCGGATCCCCAA	CPuORF
ROJ	Sequence	GAAAAAGCTTAACAGACCACCATGATGGTGGATCGAACG GTGAGAATGGGGACGGAGGTGAAGTCGATTAGGGTTCC AAGAAGGGTCTGGTTTGGATCACAAATATCTATAGTTCTCC TCGTGGGTGGAAACCATACCTCCTCCAGGTTTTGTACTC GCtgaCGGATCCCCAA	CPuORF
bZIP11	Sequence	GAAAAAGCTTAACAGACCACCATGTACCCATACGATGTTC CTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATG CAATGTCTCCAATAATACTCAGTGAGATCTTTCTCTCTGGG TTTATGTTAAATTCCACAATCAGGCGCAGGACCCATCTTG TTCAATCTTTCTCTGTTGTCTTCCTTTACTGGCTTTACTATGT CTCAtgaCGGATCCCCAA	CPuORF
TBF1	Sequence	GAAAAAGCTTAACAGACCACCATGTACCCATACGATGTTC CTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATG CAATGGAAGAAACCAAACGAAACTCCGATCTTCTCCGTT CTCGTGTTTTCCTCTCTGGCTTTTATTGCTGGGATTGGGAA TTTCTCACCGCTCTCTTGCTTTTTAGTTGCtgaCGGATCCC CAA	CPuORF
PAO2	Sequence	GAAAAAGCTTAACAGACCACCATGTACCCATACGATGTTC CTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATG CAATGAATTTGTTTGGAAGCCTGATTTTTTCATCTTCGATTT CGAGCCACTTCACGACTTGTCTTCTCATCTTTATTCTTCATT TCTCTCCCCAGCTTAATAGTAATAATAATAATAATAGTAA TCGTAATTCCAATAATCTCGAGAGTAATCGTTTGATCAATtg aCGGATCCCCAA	CPuORF
HisHisHA: HEAT	Sequence	GAAAAAGCTTAAAGTTGTCACTTTATTATTACTTCACTTGAG CTTCTCTCAAGATCACAGATCTCTCGTTCCTTCTGCTCCC AACTCTCTACCCAGATTCCACCTTTACCTTACGAAACCCT AATCCCCAAAATCAGCTTAAAGTTTCAATCTTTGTTGAGAA GAAGAGAGTTCGTCGTCTTAGTATGTAGCTAGATCTCAAG GAAAATACGAATGGTTCGATCAATTGGGTAGACAACAAAA AAATGCATCACCATCACCATCACCATCACCATCACCATC ACTACCCATACGATGTTCCAGATTACGCTGATCTTTTGATG ATACAGAGACGTGAATCGAGAAGGGTCTGGTTAGGTGTA CAAAATTCAATAGTTCTCCTCGTGGGTGGCAACCATACCT CCTCCAGGTTTTGCACTCGCTAAATCTCTCTTCTTCTTCTG CCTAGAAATCTTCTCTTGAAATCAAACTCTTCAGTTCTTGTT TTAACTACTCGGATCCCCAA	Purify stalled ribosome

Table 2.2. Primers and sequences used within this thesis. Details of the synthesised gene fragments and primers used in this study. Final column details the purpose for each sequence or primer.

2.4.3. PCR reactions

PCR was conducted using Q5® High-Fidelity DNA Polymerase (NEB M0491) in reaction volumes of either 25 µl or 50 µl. The reaction mixture consisted of 1X Q5 Reaction Buffer, 200 µM dNTPs, 0.5 µM of both forward and reverse primers, a variable amount of template DNA (less than 1,000 ng), and 0.02 U/µl Q5 High-Fidelity DNA Polymerase. Reactions were assembled on ice and thermocycled with initial denaturation at 98°C for 30 seconds, followed by 25-35 cycles of 98°C for 5-10 seconds, 50-72°C for 10-30 seconds, and 72°C for 20-30 seconds/kb, and a final extension at 72°C for 2 minutes.

2.4.4. DNA agarose gel electrophoresis

DNA gel electrophoresis was performed using 1x TAE agarose (Sever Biotech Ltd.), which was dissolved by heating in a microwave. Ethidium bromide was added to the buffer to a final concentration of 0.5 µg/µl for gel staining. Before loading, the DNA samples were mixed with the appropriate loading buffer. The samples were run alongside the Generuler 1kb Plus (Thermo Fisher SM1331) molecular weight ladder for 25 minutes using a Bio-Rad gel tank and power pack at 100 V. Following electrophoresis, the gel was visualized with a UV transilluminator and photographed with a Syngene imaging system.

2.5. *Arabidopsis Agrobacterium* transformation.

Arabidopsis plants were grown to flowering stage under long days and clipped to encourage secondary bolts. *Agrobacterium tumefaciens* carrying the gene of interest were prepared and resuspended to OD600 = 0.8 in 5% sucrose solution, with 0.05%

Silwet L-77 added before dipping. The above-ground parts of the plants were dipped in the *Agrobacterium* solution for 2-3 seconds and then placed under high humidity for 16-24 hours. Plants were then grown normally, and dry seeds were harvested. Transformants were selected using GFP selection.

2.6. Luciferase assays

The *in vitro* translation reactions were carried out with capped 6 ug of RNA from specific plasmid constructs. mRNA was synthesised using the HiScribe RNA synthesis kit (NEB E2070S) and capped with (NEB S1407S). These were processed with the TNT Uncoupled Wheat Germ Kit (Promega L4130) for two hours at 25 °C. Luciferase activity was evaluated by combining 10 µl from each translation reaction with an equivalent volume of LUC assay mixture containing 0.5mM luciferin and 0.01% Triton X-100 (w/v). This mixture was placed in an opaque container and luciferase activity was detected and quantified using the NightShade plant imaging system (Berthold LB985). Specific settings included exposure times of 2s, a 4x4 binning pattern, high gain, and a slow reading mode. The photon count rate was measured for every sample and averaged across the triplicates. *In vitro* luciferase reactions were optimised for RNA concentration and duration (figures 2.3-4).

For the translation reactions *in planta*, mature and healthy leaves were carefully selected from transgenic plant rosettes and treated with LUC assay buffer (0.5mM luciferin, 0.01% (w/v) Triton X-100) at room temperature for 2 minutes. Luminescence detection and quantification were performed under the same conditions as used for *in vitro* systems.

The resulting cps (counts per second) per area (mm^2) of leaf tissue was computed and the average was taken across all replicates.

2.7. Isolating the *Arabidopsis* monosome

To purify ribosomes from plant tissue two methods were employed 1) Centrifugation and a sucrose cushion or 2) Co-immunoprecipitation with a tagged ribosomal protein.

2.7.1 Ribosome purification by ultracentrifugation and a sucrose cushion

Arabidopsis wild type plant tissue was flash frozen with liquid nitrogen and pulverized with a glass homogenizer. Homogenised tissue was added to a sterile beaker. Twice the volume of tissue of a polysome extraction buffer was added. The extraction buffer contained 0.4 M Tris (pH 9.0), 0.2 M KCl, 0.025 M EGTA, 0.035 M MgCl2, 50 mg/mL cycloheximide, and 50 mg/mL chloramphenicol. The mixture was allowed to thaw on ice with occasional mixing. The mixture was again homogenized using a glass homogenizer.

After homogenization, the samples were subjected to an initial centrifugation at 4°C, 16,000 x g for 15 minutes using a Beckman J2-21 high-speed centrifuge fitted with a JA-20 rotor at 11,500 rpm. The pellet was discarded, and the supernatant was filtered through Miracloth (Merck 475855). The supernatant was ultracentrifuged through a 1.75 M sucrose cushion (0.4 M Tris (pH 9), 0.2 M KCl, 0.005 M EGTA, 0.035 M $MgCl_2$, 1.75 M sucrose, 5 mM DTT, 50 ug/mL cycloheximide, 50 ug/mL chloramphenicol)in a Beckman ultracentrifuge 2 XL-I machine, using a 50.2 Ti rotor. Ultracentrifugation was performed at 116,000 g for 18 hours at 4°C to isolate the ribosomes.

The pellet containing the ribosomes was carefully collected and resuspended in a wash buffer. The wash buffer composition included 0.2 M Tris (pH 9.0), 0.2 M KCl, 0.025 M EGTA, 0.035 M MgCl2, 5 mM DTT, 1 mM PMSF, 50 mg/mL cycloheximide, 50 mg/mL chloramphenicol, and 20 U/mL RNAsin. The isolated ribosomes were investigated via a sucrose gradient, western blot, or cryo-electron microscopy.

2.7.2 Co-immunoprecipitation.

Arabidopsis TRAP tissue that harbours a FLAG tag on uL18 was prepared through homogenization and centrifugation as in Methods 2.8. In preparation for immunoprecipitation, magnetic α-FLAG M2 agarose beads (Sigma M8823) were suspended, transferred, and washed the wash buffer described in Methods 2.8. After homogenization, TRAP tissue suspended in a Polysome extraction buffer (PEB) (0.2 M Tris, pH 9.0, 0.2 M KCl, 0.025 M EGTA, 0.035 M $MgCl_2$, 1% Detergent mix (20% Brij-35, 20% Triton X-100, 20% octylphenyl-polyethylene glycol, 20% Tween 20), 1% socium deoxycholate, 1% Polyoxyethylene 10 tridecyl ether, 5 mM Dithiothreitol, 1 mM Phenylmethylsulfonyl fluoride, 50 ug/mL cycloheximide, 50 ug/mL chloramphenicol and 0.5 mg/mL RNasIn). was pelleted through centrifugation at 4°C, 16,000 x g for 15 minutes using a Beckman J2-21 high-speed centrifuge fitted with a JA-20 rotor at 11,500 rpm. The pellet was removed leaving the supernatant that contained ribosomes (clarified extract).

The clarified extract was added to α-FLAG M2 agarose beads in a 15 mL plastic Falcon tube. The beads and clarified extract were incubated for 2 hours at 4°C with gentle

shaking. The tagged ribosomes should be bound to the beads. The beads were washed using a wash buffer as detailed in Methods 2.8. using a magnet to separate the beads. After a total of four washes, the ribosomes were eluted by incubating the beads with wash buffer containing 200 ng/mL of FLAG3 peptide (A36806) and 20 U/mL RNAsin for 30 minutes at 4°C with shaking. This supernatant was transferred to a new tube after centrifugation. To separate the 80S ribosome, the eluate was spun through a sucrose gradient by ultracentrifugation at 4°C and 237,000 g (50,000 rpm, SW55.1 rotor) for 1.5 hours.

2.7.3. Sucrose gradients

To isolate the 80S monosome from subunits and polysomes, the eluate from Methods 2.8. was ultracentrifuged through a sucrose gradient. Sucrose gradients were prepared for 5 mL tubes for the SW55 rotor. Step gradients were prepared in Seton 5022 tubes as per Table 2.3. On the day of use, the gradients were removed from the freezer and thawed in a 37°C incubator for 1 hour, then cooled at 4°C for 1–1.5 hours.

Sucrose %	2 M Sucrose (mL)	10 x Sucrose Salts (mL)	Sterile water (mL)	Chloramphenicol and cycloheximide (µl)	Volume per gradient (mL)
60	44	5	1	5	0.75
45	59.5	7.5	18	7.5	1.5
30	33	7.5	34.5	7.5	1.5
15	11	5	34	5	0.75

Table 2.3. Table to show how sucrose gradient stock solutions were made and volume per gradient of each sucrose layer.

300 µl of sample was loaded onto the sucrose gradients and were spun at 237,000 x *g* for 1.5 hours at 4 ºC in an SW55 rotor. The resulting gradients were fractionated by a Density Gradient Fractionation System. 80S monosome peak fractions were isolated and investigated via cryo-EM or western blotting. The fractions containing the 80S monosome were selecting from the sucrose gradient profile.

2.7.4. Western blots.

Samples for western blot analysis were diluted in 2x Llamelli SDS sample buffer (BioRad #1610737) with 2-mercaptoethanol (Sigma M6250). Samples were then boiled at 95 ºC for 3 minutes. Samples were separated using BioRad 4-20% gradient gels (BioRad #4561094). Samples were run alongside a protein ladder (Thermo Fisher LC5800) at 200 V until appropriate separation. Transfer was performed using the Trans-Blot Turbo Transfer System with Trans-Blot Turbo Mini 0.2 µm PVDF Transfer Packs (BioRad 1704156). The membrane was incubated for 1 hour at room temperature with 5% milk + 1% TBS (Tris buffered saline solution (Fisher 10776834). After blotting, the membranes were incubated overnight at 4 °C with their primary antibodies at a dilution of 1:1000 in 5% milk + TBS. The primary antibodies used were anti-60S Ribosomal Protein L37 RPL37 Antibody (BosterBio A10535), and RPS6A 40S ribosomal protein S6-1 (Argisera AS19 4292).

Post primary antibody incubation the membranes were incubated with a secondary antibody for 90 minutes. The secondary antibody was suspended in 5% milk + TBS at a dilution of 1:1000. The secondary antibody was a HRP conjugated Anti-Rabbit antibody

(BosterBio BA1054). Finally, the membranes were washed for 1 hour with TBS + 1% TWEEN and imaged with SuperSignal (Thermo Fisher 34578) and Syngene (www.syngene.com). For the anti-60S Ribosomal Protein L37 RPL37 Antibody (BosterBio A10535), and RPS6A 40S ribosomal protein S6-1 (Argisera AS19 4292), these antibodies were incubated overnight. The secondary antibody was incubated for 1 hour and 30 minutes, washed for 1 hour, and imaged with SuperSignal (Thermo Fisher 34578) and Syngene (www.syngene.com).

2.8. Isolating stalled CPuORF ribosomes.

To isolate CPuORF stalled ribosomes, CPuORFs were tagged with an N-terminal HisHisHA tag (110,139). Capped RNA was synthesized using the SP6 HiScribe RNA synthesis kit. A 1 mL *in vitro* translation reaction was loaded with 80-200 ug of capped RNA and translated for 1 hour at 25°C. Post translation, the reaction was ultracentrifuged through a sucrose cushion (0.4 M Tris (pH 9.0), 0.2 M KCl, 0.005 M EGTA, 0.035 M MgCl2, 1.75 M Sucrose and 5 mM DTT). Samples were centrifuged at 4°C, 170,000 g for 3 hours (50,000 rpm, TY 70Ti rotor), or alternatively at 116,000 g (35,000 rpm, TY 70Ti rotor) overnight, using a 50.2 Ti rotor. The resulting pellet was resuspended in RNC buffer (50 mM Tris-HCl (pH 7.0), 250 mM KOAc, 25 mM Mg(OAc)2, and 5 mM 2-mercaptoethanol). To the resuspended ribosome pellet, 60 µL of magnetic anti-His beads (Introvogen 10104D) were added and incubated for 30 minutes at 4 °C. The beads were then washed four times with RNC buffer + 0.01% TWEEN. Stalled ribosomes were eluted using RNC buffer with 200 mM imidazole and the eluate was pelleted again through a sucrose cushion. The resulting pellet contained the stalled ribosomes that were eluted in cryo-

EM buffer (50 mM Tris (pH 8), 150 mM NaCl, and 10 mM MgCl2) and aliquoted for cryo-EM and western blotting analysis.

2.9. Cryo-electron Microscopy.

Ribosome and stalled ribosomes were prepared as described in Methods 2.8-9. Samples were suspended in a buffer suitable for cryo-EM. The buffer consisted of 50 mM Tris (pH 8), 150 mM NaCl and 10 mM MgCl2.

2.9.1 Negative stains and TEM imaging.

Negative staining of ribosomes was conducted using the side blotting protocol (164). Staining was performed on in-house carbon-coated grids. Grids were glow discharged using PELCO easiGlow discharged for 30 seconds. Following the staining procedure, the samples were analysed using the Tecnai T12 Transmission Electron Microscope equipped with a Gatan 4000 CCD camera.

2.9.2. Making cryo-EM grids.

Cryo-EM grids were prepared using lacey carbon films on 400 mesh copper grids (Sigma GS166-4H). They were glow discharged for 30 seconds using the easiGlow. Grids were loaded with 3 µl sample and plunge frozen using the Vitrobot mK IV system. The Vitrobot was calibrated to a temperature of 4°C and a relative humidity of 95%. Grids were loaded with 3.5 µl of sample. The vitrobot was set to a blot time of 6 seconds with a blot force of 6. Following blotting, the grids were immediately plunged into liquid ethane to flash-freeze the sample. Grids were then clipped and stored in liquid nitrogen.

2.9.3. Cryo-EM data collection.

Data for cryo-electron microscopy was acquired using the FEI Titan Krios 2 transmission electron microscope. Data was recorded on a Falcon 4 direct electron detector with an accelerating voltage of 300 kV. The acquisition parameters had a pixel size of 0.86 Å, nominal magnification of 96k, and a spot size of 7, illuminating an area of 0.56 µm. Dose calculations were performed to ensure optimal electron exposure, with a square pixel dimension of $Å^2$, dose per physical pixel per second of 5, and a dose per $Å^2$ per second of 6.76. The exposure time was set at 5.15 seconds, yielding a total dose per $Å^2$. A dose per hardware frame of 0.028 e-/$Å^2$ and dose per frame of 0.8 e-/$Å^2$ was achieved, with the grouping of 28 frames.

2.9.4. cryo-EM data analysis.

Cryo-EM image processing was carried out on-the-fly using RELION 3.1.2 (165). Motion correction and CTF estimation was performed using MOTIONCORR and Niko Grigorieff's ctffind 4.1, respectively (165,166). Motion correction was employed to address blurring and distortions caused by movements within the data frames, while CTF correction aimed to reduce electron aberrations arising from the electron microscope. Particles were selected using the open-source crYOLO particle picker for cryo-EM (165,167). Particle coordinates were subsequently imported into RELION 3.1.2 (165). Particles picked by crYOLO were binned by a factor of 5.

Datasets underwent a 2D classification to eliminate low-quality and unwanted particles. Following 2D classification the data set underwent 3D classification. These

classifications utilized an in-house reference map of the wheat germ extract ribosome provided by Juan Fontana and were executed over 25 iterations using RELION 3.1.2. 3D models were examined using ChimeraX (168). Particles extracted from refined 3D classification datasets underwent a refine3D job to generate high resolution models.

3. A molecular analysis of the CPuORF CUTS mechanism.

3.1. Introduction

3.1.1 Plant CPuORFs and the CUTS mechanism

A selection of *Arabidopsis* CPuORFs have been shown to conditionally stall the ribosome during translation (Figure 1.13). Bringing together the wider literature and in house data, the Davies Lab proposed the CUTS model (Figure 1.13). The CUTS model is based upon the following observations:

1) Some CPuORFs cstall the ribosome during translation either innately (default mode) or in response to a signal (rCUTS) (Table 1.5 and Figure 1.13).

2) Other CPuORFs promote mORF translation in response to a signal (Table 1.5 and Figure 1.13).

3) CPuORFs respond to a range of signals, including small metabolites, abiotic conditions, and infections (Table 1.5).

4) The CPuORF response is specific and not a generalized stress response (3).

3.1.2. Applications of CUTS in agriculture and biotechnology.

Climate change is causing reductions in crop yields (169). The Food and Agriculture Organization (FAO) calculate that we will need to produce 70 % more food to feed a global population of 9.3 billion by 2050 (169). Constraints on land availability and increasing degradation exacerbate the difficulties in increasing crop yields.

A strategy to improve crop output is through genetic modification. The CPuORF CUTS mechanism has already shown promising results in the lab and field. For example, the TBF1 CPuORF was utilized to develop pathogen-resistant rice without a fitness cost (13). This study inserted a pathogen responsive CPuORF upstream of NPR1 (AT1G64280). NPR1 is a key regulator of the pathogen defence system. By controlling the expression of a pathogen defence gene with a pathogen responsive CPuORF, transgenic plants exhibited fewer lesions and less pest damage (13).

In research, the CUTS mechanism could be used in a similar way to known peptide switches (for example, β-oestradiol switches, phosphorylation switches, GTPase switches, and calcium-binding switches) (170,171). Hence, CPuORFs can assist in the exploration of the conditional expression of specific genes. The diversity amongst CPuORFs in terms of signal response and mode of CUTS could offer a library of regulatory peptides to suit specific research aims.

3.1.3. Understanding the CPuORF CUTS mechanism: Current knowledge and unanswered questions.

Using *in vitro* and *in planta* transient expression systems studies have been able to demonstrate the conditional effect of CPuORFs on mORF translation (Table 1.5). Evidence points to the role of conserved amino acids in enabling CPuORFs to stall the ribosome (Tables 1.4-5) (3,11,126). This has been shown by mutagenic investigations into the CPuORF peptide sequences and their effect on the attenuation of an artificial

downstream reporter gene such as luciferase (6,109,130). Alanine scanning mutagenesis has shown that when conserved amino acids are changed there are greater levels of mORF expression when compared to changes affecting less conserved amino acids.

The conservation patterns of CPuORF sequences can be classified into three classes. Class I CPuORFs are conserved at the C-terminus, Class IIa CPuORFs are conserved at the N-terminus and Class IIb are conserved throughout (Figure 1.12). In the case of Class I and Class IIa the CPuORFs could be described as organised into a conserved region and a variable region. The CPuORF molecular mechanism for stalling the ribosome and signal sensing is yet to be elucidated. As detailed in the *Arabidopsis* CPuORF database in Causier et al., there is currently evidence that 23 CPuORFs conditionally stall the ribosome during translation (3). Moreover, Causier et al., also correlated ribosome stalling sites at the conserved CPuORF amino acids of different CPuORF classes (3). Therefore, it could be suggested that the conservation patterns observed could confer aspects of the CUTS mechanism. For example, the conserved region/amino acids confer ribosome stalling and the variable region/amino acids confer signal sensing. If CPuORF function is organised this way, it could be described as modular. Understanding if CPuORFs have a modular nature would allow different CPuORF domains to be swapped to create bespoke CPuORFs.

To create bespoke CPuORFs and determine how CPuORFs conditionally control translation of downstream mORFs, the molecular mechanism needs to be resolved. Alongside molecular biology, structural biology could elucidate the CPuORF molecular mechanism. Currently, one Cryo-EM map has been published of a the bZIP11 rCUTS

CPuORF stalled ribosome (20). The structures of the bZIP11 CPuORF and known translation arrest peptides from other model systems (Figure 1.19) suggest that ribosome stalling happens through interactions between nascent peptide and the ribosome exit tunnel (20) A comparison of cryo-EM data of stalled ribosomes suggests there are commonalities between stalling mechanisms. Nascent stalling peptides stall by interacting with the constriction in the exit tunnel (uL4 and uL17), rRNA nucleotides and the PTC. Furthermore, the bZIP11 rCUTS CPuORF stalls the ribosome if sucrose is sequestered into a conserved metabolite pocket in the exit tunnel (110). The same exit tunnel metabolite pocket has been shown to bind tryptophan and the chemically similar drug PF846 in rCUTS mechanisms across bacteria and eukaryotes. This evidence may suggest that other CPuORFs may also sense their metabolite signals via this pocket in the exit tunnel.

The question remains, how do CPuORFs respond specifically to a signal using the same pocket? The specificity is likely conferred by the specific biochemistry of the nascent peptide aligned in the tunnel during stalling. This is supported by transient expression assays demonstrating that CPuORF function is sequence dependent (Table 1.4). However, CPuORFs also respond to abiotic signals, and it is harder to explain how the ribosome could detect drought or heat shock through a metabolite pocket. It is therefore possible that these triggers are detected via a downstream signal.

3.1.4. Aims

This chapter will investigate the CPuORF CUTS mechanism using molecular biology. To elucidate the mechanism, this study will largely focus on a selection of CPuORFs previously investigated by the Davies lab (SAC51 (AT5G64340), eIF5 (AT1G36730), HEAT (AT3G53400) and ROJ (AT5G03190)) and others to cover CPuORF diversity in class and signal response (Table 3.1) (3). These CPuORFs are all aCUTS and are either Class I or Class IIa (Table 3.1). Furthermore, this study will also investigate bZIP11 which is a Class I CPuORF and functions through rCUTS, TBF1 which works both via aCUTS and rCUTS and PAO2 which is aCUTS and class IIb (Table 3.1). Each of these CPuORFs respond specifically to various signals from small metabolites, abiotic conditions, and pathogen infections (biotic).

Aim: To elucidate the CPuORF CUTS mechanism using *in vitro* and *in vivo* luciferase assays.

- Design and synthesize constructs that position both wild-type and mutant CPuORFs upstream of a luciferase reporter gene.
- Evaluate the influence of mutations in the CPuORF and ribosome on ribosomal stalling, signal sensing and mORF translation.

CPuORF	TAIR	HG	Class	Length aa	Mode of CUTS	Signal	Notes	Ref
SAC51 (CPuORFs 38,39,40)	AT5G64340	15.1, 15.2, 15.3	IIb, IIa, I	20, 16, 53	a	Thermospermine	Contains three CPuORFs in its 5' leader. Imai et al., suppressor screen suggests CPuORF 40 is functional. Unknown how the three CPuORFs respond to thermospermine to modulate mORF expression.	(3,8,37,118)
eIF5 (CPuORF 19)	AT1G36730	7a	IIa	57	a	Drought	Conserved over 0.5 billion years ago since bryophytes and angiosperms diverged. Poor start codon context.	(3)
HEAT (CPuORF 46)	AT3G53400	17	I	38	a	Heat shock	Conserved in fern species (350 million years of evolution). Homologous CPuORFs within HG17 have the same conserved C-terminal amino acids but have divergent N-terminal amino acids. HG17 CPuORFs respond to different abiotic signals while the sequences only diverge at the N-termini. Rice homologues although stalls the ribosome, does not respond to same stimuli tested from HG17. mORF encodes a methyltransferase.	(3)
ROJ (CPuORF 47)	AT5G03190	17	I	45	a	Drought	See above.	(3)
bZIP11 (CPuORF 2)	AT4G34590	1	I	42	r	Sucrose and hypoxia	Only stalls in the presence of sucrose. Sucrose response is conferred by conserved amino acid not variable N-termini. However, less conserved amino acids also confer stalling. Mutation in serine31 abolishes stalling. mORF encodes a transcription factor within the sucrose response network. Cryo-EM map has elucidated conditional stalling mechanism.	(11,12,126,127)
TBF1 (CPuORF 49)	AT1G36730	18	I	36	both	Galactinol (r), Heat and Pathogens (a)	Currently the only CPuORF reported to function via both modes of CUTS. CPuORF is found upstream of an mORF important in the transition from a vegetative to a defence mechanism. CPuORF functions outside native context. Utilized in crop improvement.	(9,128,129)
PAO2 (CPuORF 17)	AT2G43020	6	IIb	55	r	Polyamines	CPuORF shown to be functional outside of *Arabidopsis* (onion). Conserved region and specifically a leucine residue confer functionality. mORF encodes a polyamine oxidase	(132,133)

Table 3.1. Overview of CPuORFs investigated in this study. The table details each CPuORF's mORF, TAIR number, homology group (HG), class, amino acid length, mode of CUTS (either repression (r), activation (a), or both), signal response, key observations, and references.

3.2. Results

3.2.1. Identifying CPuORFs 38 and 40 as the functional SAC51 CPuORFs.

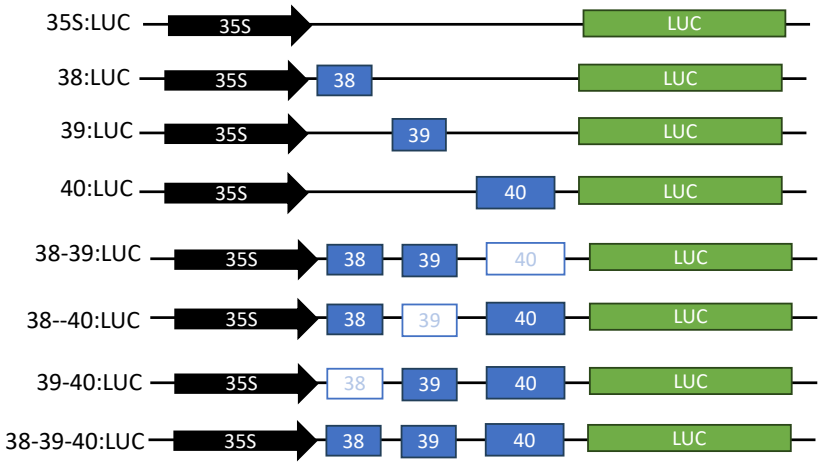
This chapter is an investigation into the CPuORF molecular mechanism and will use molecular biology to determine the functional significance of the CPuORF amino acid sequence. 23 different CPuORFs have been shown to respond to diverse signals. This thesis investigated four CPuORFs that were already being examined in the lab (SAC51, eIF5, HEAT and ROJ). To add increased CPuORF diversity this thesis will also investigate the CPuORFs found in TBF1, bZIP11 and PAO2 (as described in section 3.1.4 (Table 3.1). The PAO2 (Polyamine oxidase 2) mORF functions in polyamine metabolism and harbours a polyamine responsive CPuORF. SAC51 has multiple CPuORFs in its 5' leader (CPuORFs 38, 39 and 40) and it is unknown which single or combination of CPuORFs confer its reported thermospermine aCUTS activity (3). Consequently, this first section will explore the SAC51 transcript and determine which combination of SAC51 CPuORFs confer stalling and thermospermine responsiveness.

To determine the functional CPuORFs in the SAC51 transcript a selection of each SAC51 CPuORF and CPuORF combinations were placed upstream of an artificial luciferase mORF (Figure 3.1 A). By putting luciferase expression under the control of different CPuORFs luciferase activity can be measured to quantify levels of stalling under control conditions. Furthermore, luciferase activity quantified under elevated thermospermine

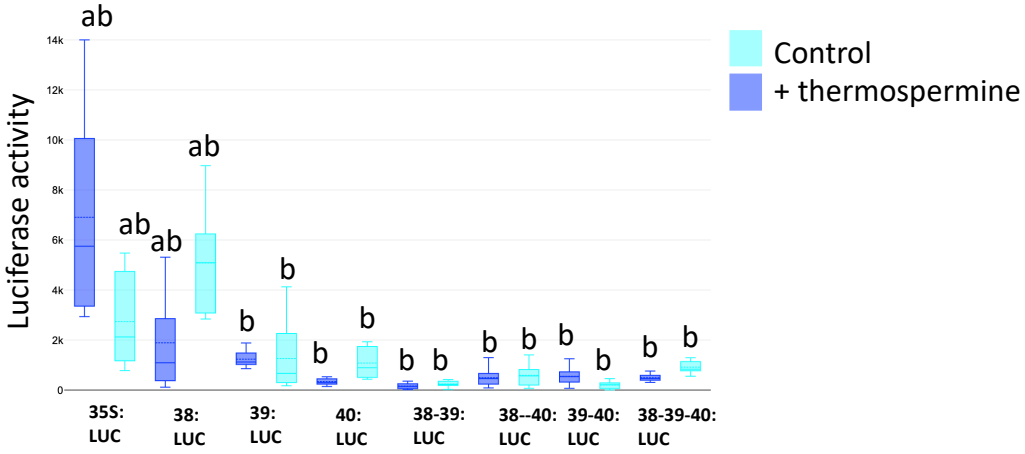
levels to test CPuORF signal response. By comparing different luciferase levels as controlled by various SAC51 CPuORFs under control and elevated thermospermine conditions this experiment could identify the CPuORFs responsible for thermospermine aCUTS activity. The constructs driven by a constitutive 35S promoter and expressed in wild type *Arabidopsis*. Luciferase data collected from T1 leaves post treatment incubation. The decision to investigate >10 independent T1's for each construct due to practical efficiency of rapidly assessing transgene effects without developing stable homozygous lines. By testing multiple T1's this would help mitigate the impact of transgene integration variability.

Three single SAC51 CPuORF constructs were made 38:LUC, 39:LUC and 40:LUC (Figure 3.1 A). Furthermore, three constructs were generated that contained two CPuORFs 38-39:LUC, 38--40:LUC and 39-40:LUC. 39 was removed by site directed mutagenesis to create 38--40:LUC. 38-39-40:LUC was effectively representative of the native SAC51 5' leader that will served as a positive control. Finally, 35:LUC will acted as a negative control as it does not have any CPuORFs upstream of the luciferase mORF.

A)


35S:LUC
38:LUC
39:LUC
40:LUC
38-39:LUC
38--40:LUC
39-40:LUC
38-39-40:LUC
35S
38
39
40
LUC

B)


Luciferase activity
Control
+ thermospermine
ab
ab
ab
ab
b
b
b
b
b
b
b
b
b
35S: LUC
38: LUC
39: LUC
40: LUC
38-39: LUC
38--40: LUC
39-40: LUC
38-39-40: LUC

C)

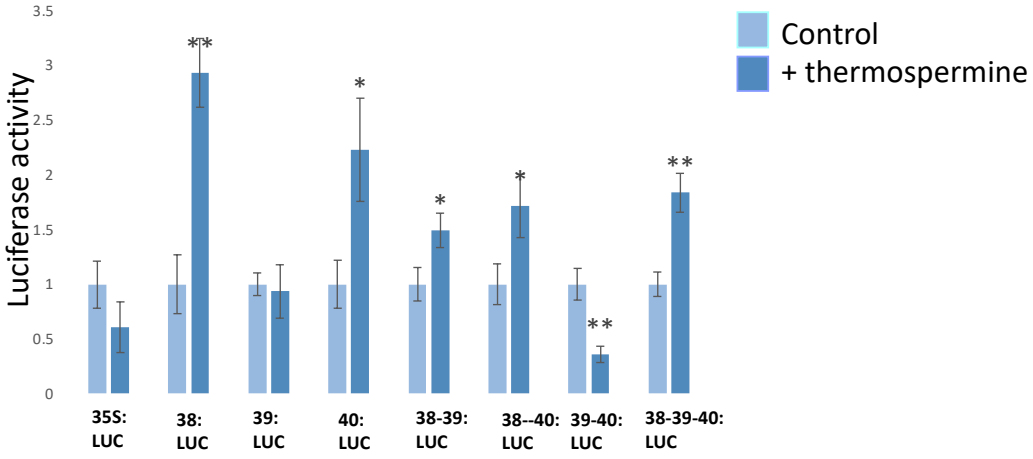

Luciferase activity
Control
+ thermospermine
35S: LUC
38: LUC
39: LUC
40: LUC
38-39: LUC
38--40: LUC
39-40: LUC
38-39-40: LUC

Figure 3.1. SAC51's functional CPuORFs are CPuORF38 and CPuORF40. A) Schematic representation of the construct design. The 35S promoter is symbolized by a black arrow. The CPuORFs are shown as blue boxes, each labelled with the specific SAC51 CPuORF. Faded boxes represent a CPuORF that has not been included in the construct. The luciferase reporter gene is denoted by a green box. Constructs were transformed into wild type *Arabidopsis* lines by agrobacterium B) A box and whisker plot to identify functional SAC51 CPuORFs. Control conditions (0.5 MS) is illustrated in light blue, while the results from thermospermine conditions (0.5 MS combined + 0.05 mM thermospermine) are in dark blue. Luciferase activity was measured from T1 leaves and is calculated by "luciferase activity(cps)/area of leaf". Statistical significance is determined through an ANOVA Tukey test and statistical groups are labelled on the box and whisker plot. Statistical significance is determined through a t- test and is marked by asterisks, with one asterisk (*p<0.05) indicating a p-value less than 0.05 and two asterisks (**p<0.01) a p-value less than 0.01. Error bars are the standard error of the mean. An ANOVA Tukey test was used to determine significance and cluster datasets. The number of replicates for each group ranges from 8-22. C) A bar chart to show fold change between elevated thermospermine and control conditions. Data is normalized to control conditions between each construct.

When evaluating the 35S:LUC negative control lines, there was no change in mORF translation under thermospermine conditions (Figure 3.1 B and C). This is likely because elevated thermospermine has been reported to stress plants and reduce mORF expression. On the other hand, a student's T test indicated there is a significant increase in mORF expression in 38-39-40:LUC lines under elevated levels of thermospermine when compared to control conditions as previously reported (3,8). This line represents the native SAC51 5' leader and there is a fold increase of 1.8 in luciferase activity under elevated levels of thermospermine. Specifically, this is an increase of an average 498.42 cps/mm^2 to 916.02 cps/mm^2 from control to elevated thermospermine conditions. Consequently, the 38-39-40:LUC line has retained its reported thermospermine aCUTS activity. Moreover, the fact that the 35S:LUC lines demonstrated an average decrease in mORF expression upon thermospermine application, highlights the role of the SAC51 5'leader (38-39-40) in conditionally modulating mORF expression.

Considering the individual CPuORF constructs (38:LUC, 39:LUC, and 40:LUC), 38:LUC and 40:LUC retain the SAC51 thermospermine response as thermospermine causes a fold change of 2.9 and 2.2 respectively. This is similar to the 1.8 fold change observed in the native SAC51 5'leader. Conversely, there is no change between control and thermospermine conditions in 39:LUC plants. Overall, this indicates that CPuORFs 38:LUC and 40:LUC can independently modulate mORF expression in response to thermospermine.

Two combinations of double SAC51 CPuORFs (38:39:LUC and 38:40:LUC) also retain the aCUTS activity as seen in 38-39-40:LUC (Figure 3.1). 38:39:LUC and 38:40:LUC demonstrate an increase in mORF translation in response to thermospermine. However, this is not observed in the 39:40:LUC plants. CPuORF 39 is not responsive to thermospermine but still may stall the ribosome during translation. Consequently, CPuORF 39 being the first CPuORF in 39:40:LUC plants may sequester ribosomes, either by stalling or the failure of the ribosome to re-initiate at the downstream CPuORF 40, potentially diluting CPuORF 40's conditional response in these lines.

An ANOVA Tukey test revealed some interesting results. The data indicates that CPuORF 40 more closely resembles the 38-39-40:LUC construct, which acts as the native 5' leader. The ANOVA does not differentiate the aCUTS response in the 38-39-40 construct or the double CPuORF constructs as they all cluster in group "b". Interestingly, CPuORF 38 appears to be the least capable of ribosome stalling, which might be expected given

that it is the shortest CPuORF. CPuORF 38 under control and thermospermine treatment cluster in group “ab” alongside the 35S:LUC control plasmid. Despite the findings from the ANOVA, a Student's t-test suggests significant increases in luciferase activity for both CPuORF 38 and CPuORF 40, indicating that they are indeed capable of an aCUTS response.

Nominal luciferase activity reveals higher mORF translation levels for CPuORFs 38 and 39 under control conditions compared to CPuORF 40 and the native SAC51 leader. Specifically, CPuORFs 38:LUC and 39:LUC averaged 1583.1 and 1346.1 cps/mm^2 respectively, whereas 40:LUC, 38:39:LUC, 38:40:LUC, 39:40:LUC and the native SAC1 (38-39-40:LUC) have averages of 484.3, 162.1, 559.7, 545.9, and 498.42 cps/mm^2 respectively. 38:39:LUC has a lower mORF translation than its single CPuORF counterparts, this may be a result of multiple ORFs in 38:39:LUC. Moreover, this may indicate CPuORF 40 is more successful at stalling the ribosome compared to 38 and 39. Furthermore, CPuORF 40 is capable of stalling the ribosome to similar levels as 38-39-40:LUC that contains multiple ORFs. This may suggest that CPuORF 40 is crucial in modulating mORF expression *in vivo*. Finally, the Imai *et al.,* suppressor screen identified a mutation in CPuORF 40 that rescues the semi-dwarf *acl5-1* mutant phenotype (Figures 1.13-16), and data in Figure 3.1 provides evidence that CPuORF 40 is the functional SAC51 CPuORF. Thus, this thesis will investigate CPuORF 40 as a representative from the SAC51 transcript in further experiments.

3.2.2. CPuORFs are self-contained regulatory units that attenuate mORF expression in a sequence dependent manner.

In vitro translation systems provide a convenient and efficient method to test the effects of CPuORF peptide sequences on the translation of a downstream reporter, such as luciferase. To optimize the experimental conditions for *in vitro* investigations, preliminary experiments were conducted to determine the optimal RNA concentration and incubation time for observing the effects of different CPuORF peptides on mORF translation (Figures 3.2-3).

The first investigation aimed to identify the optimal RNA concentration for in vitro translation. Various RNA concentrations were tested to assess their impact on luciferase activity (Figure 3.2). The results showed that mORF translation was statistically similar when the wheat germ extract (WGE) was loaded with 1-6 µg/µl of RNA. Consequently, 3 µg/µl of RNA was used in subsequent experiments.

Next, the optimal incubation time for *in vitro* translation reactions was determined by comparing the luciferase activity of an SP6p:CPuORF:LUC mRNA construct to an SP6p:LUC control plasmid (Figure 3.3). The results indicated that LUC expression from the SP6p:LUC construct remained consistent regardless of incubation time. However, for the SP6p:CPuORF:LUC construct, there was no significant difference in LUC levels compared to the control within the first 30 minutes of the reaction. The greatest difference in luciferase levels between the CPuORF-containing mRNA and the control mRNA was observed after 60 minutes of incubation. Therefore, subsequent *in vitro*

assays were incubated for 60 minutes to achieve the most pronounced effects of CPuORFs on mORF translation.

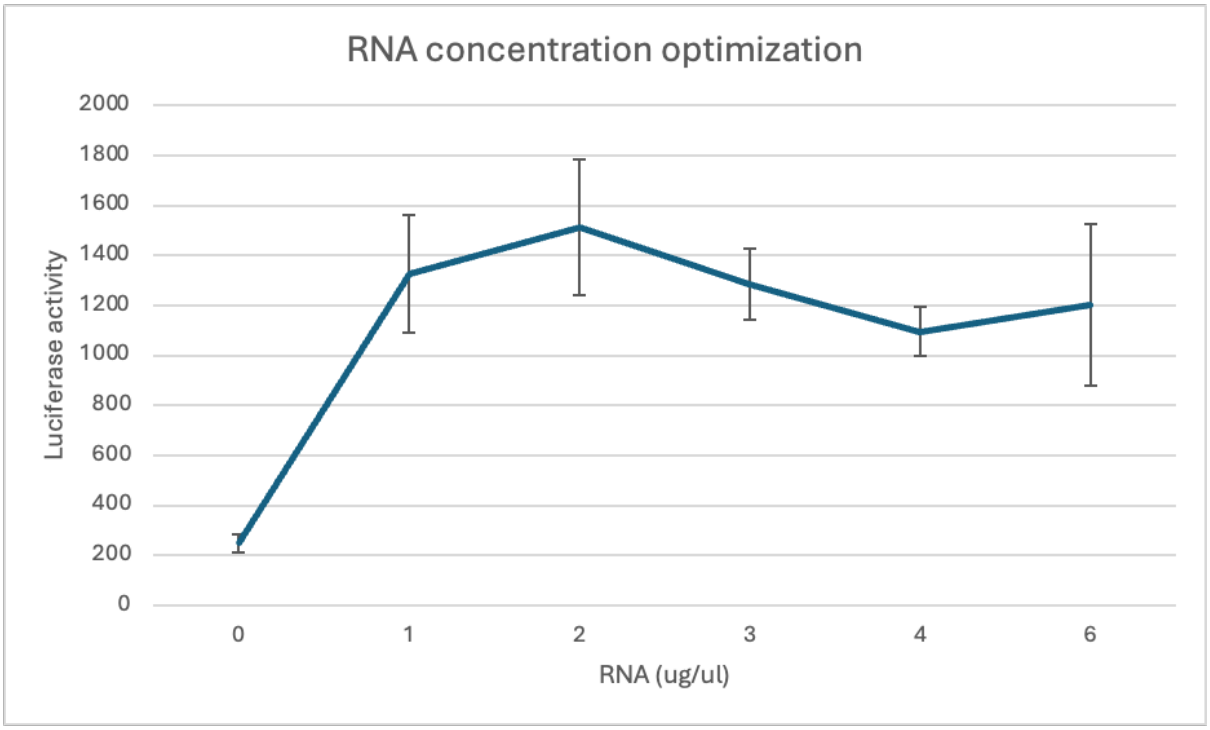


Figure 3.2. Luciferase assay to identify appropriate RNA concentration for *in vitro* transient expression assays. To optimise and test RNA concentration a the ROJ CPuORF (HG17) was used upstream of a luciferase reporter gene in the SP6:LUC plasmid. SP6:ROJ:LUC was transcribed into RNA using the HiScribe SP6 RNA kit. 0-6 ug of RNA was loaded into a 50 µl *in vitro* wheat germ extract translation reaction. Luciferase activity was detected using the Bertholdt Nightshade *in vivo* luminescence detection system. Three values were collected from each reaction. Data is presented as a line graph. 6 ug of RNA were used in follow up experiments.

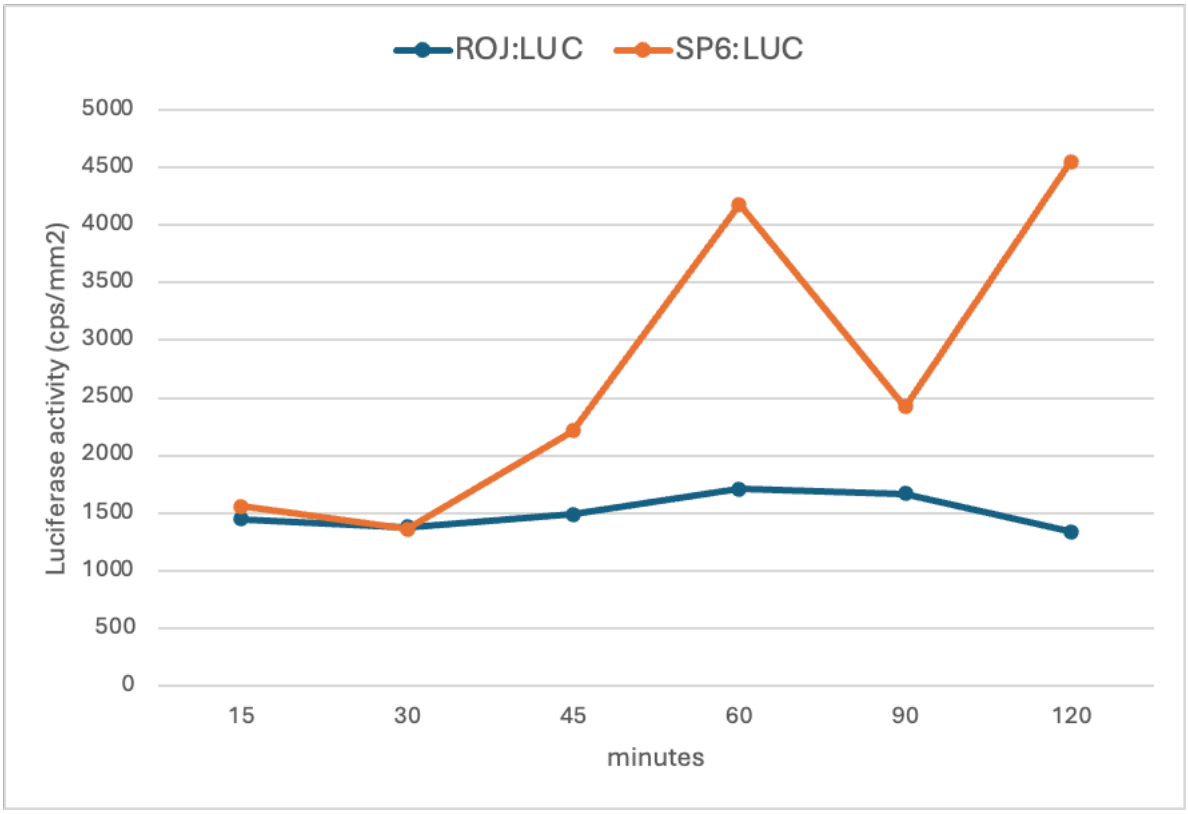


Figure 3.3. Time course luciferase assay. 6 ug of RNA was loaded into a 50 µl *in vitro* translation reaction using the Promega wheat germ extract. RNA was made for the SP6:LUC and SP6:ROJ:LUC constructs. A reading for luciferase activity was recorded at 15, 30, 45, 60, 90 and 120 minutes. Future experiments utilised an incubation time of 2 hours as it achieved a significant separation between the SP6:LUC control construct and the CPuORF containing construct.

Previous *in vitro* studies have investigated the effect of frameshift mutations on CPuORF's ability to modulate mORF expression (6,7,106). By inserting and deleting DNA bases (a frameshift mutation) in a CPuORF sequence the peptide sequence can change drastically. The resulting frameshift mutant has a new peptide sequence when compared to the wild type. Furthermore, the frameshift mutation has abolished the conserved nature of the wild type CPuORF. Previous *in vitro* and *in vivo* studies have shown that mORF expression is higher when a frameshift mutation has induced in a CPuORF when

compared to the wild type peptide sequence (6,7,106). Consequently, this thesis explored frameshift mutations in *in vitro* wheat germ extract translation system.

Unmodified and frameshifted versions of four CPuORFs were inserted via a multiple cloning site upstream of the luciferase reporter gene. The Four CPuORFs were CPuORF 40, eIF5, HEAT and ROJ (Figure 3.4 B). Frameshift mutations were a result of an insertion or deletion of a DNA nucleotide at the beginning and end the CPuORF peptide. This would result in a peptide with a completely different peptide sequence of the same length as the unmodified and wild type version. Sequences were edited to retain the same peptide length, stop codon and Kozak sequence. Finally, the CPuORFs were outside of their native 5'leader and context and placed into a common context.

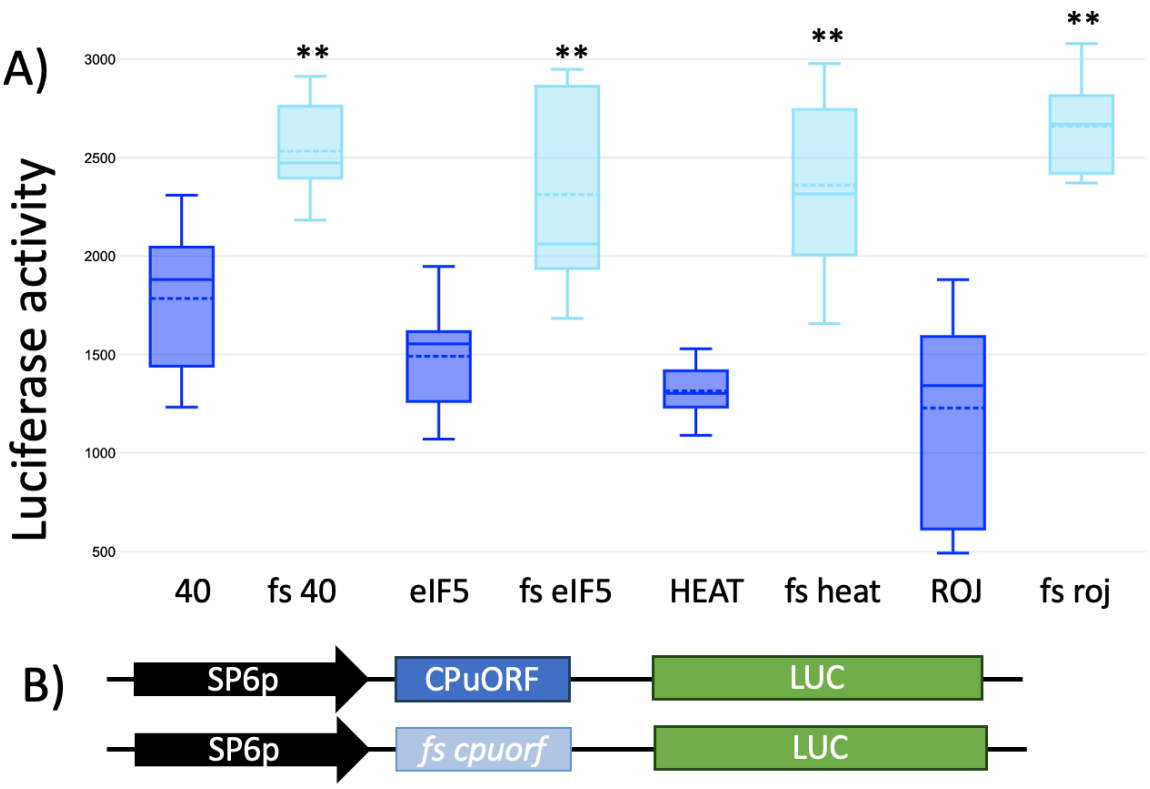


Figure 3.4. CPuORFs attenuate mORF expression in an amino acid sequence dependent manner outside their native context. A) I*n vitro* luciferase activity of WT and frameshifted (fs) CPuORFs using the Promega wheat germ extract system. The data is presented as a box and whisker plot. The Dark blue bars represent WT CPuORFs, while light blue bars indicate frameshift variants. Asterisks denote statistically significant differences between WT and frameshift variants (**p<0.01). The *in vitro* data consists of 6-9 replicates per CPuORF. B) Schematics illustrate construct designs. The median and mean are depicted as a solid and dotted lines respectively.

WT CPuORF:LUC and frameshifted (fs) variants fscpuorf:LUC mRNA's (Figure 3.4 B) were translated in an uncoupled *in vitro* translation system. mORF expression is quantified by luciferase activity. In all cases, frameshift mutations caused an increase in mORF translation when compared to their respective wild types (Figure 3.4. A). The CPuORFs investigated are CPuORF 40, eIF5, HEAT and ROJ. Frameshift mutations in these CPuORFs resulted in a fold change of 1.45, 1.55, 1.69, and 2.17 respectively. Statistical analysis using the Student's t-test confirmed the significant differences between the frameshifts and their respective WTs (**p<0.01). The data demonstrates that the CPuORF attenuation of its mORF translation is sequence dependent and not because the CPuORF is a structural feature within the mRNA. Explicitly this data is suggesting that the amino acid sequence and not mRNA structure or sequence confer the CPuORF's stalling ability. This is highlighted by the fact that minimal changes in the frameshifted CPuORF mRNA sequence resulted in significant increases in mORF translation. Moreover, it may suggest that the CPuORF peptide sequence affect translation by ribosome stalling. Furthermore, this data demonstrates that CPuORFs can attenuate mORF translation outside their native context and are self-contained units that act in cis.

3.2.3. The CPuORFs peptide switch response is specific and not a general stress response.

CPuORFs have been demonstrated to attenuate mORF expression in a sequence dependent manner outside of their native context (Figure 3.4). However, *in vitro* systems are not suitable to test signal responsiveness because abiotic signals may be detected by downstream signalling factors that are absent in commercial *in vitro systems*. To test this a range of 35S:CPuORF:LUC constructs were transformed into wild type *Arabidopsis*. T1 leaves were subjected to a range of conditions and then luciferase activity was measured from each leaf and normalized for leaf area. Transgenic leaves were incubated in control conditions (0.5 MS or at room temperature (blue)) and subjected to a range of treatments (0.5 MS + 0.05 mM thermospermine (orange), + 300 mM mannitol (yellow) to simulate drought, + heat shock at 37 ºC (red) and + 560 mM sucrose (green)). Signal concentrations and incubation periods were taken from previously reported studies (Table 1.5). Chemical stimuli (thermospermine, mannitol and sucrose) were incubated overnight, whereas temperature and heat shock stimuli were incubated for 8 hours as per Causer *et al* (3).

As a negative control 35S:CPuORF:LUC data was compared to 35S:LUC. In 35S:LUC the mORF is not proceeded by a CPuORF (Figure 3.5 B). Wild type CPuORFs were compared to frameshift or 'scrambled' CPuORFs (pale colours) where available. Frameshift mutations were created as in 3.2. and are available for CPuORF 40 and eIF5. Scrambled CPuORFs are mutants that have had conserved amino acids rearranged, abolishing the CPuORF conserved peptide sequence. Scrambled variants were created for HEAT and ROJ.

In the literature, the CPuORFs upstream of eIF5, HEAT, ROJ, bZIP11, TBF1 and PAO2 have been shown to respond to drought (mannitol), heat shock, drought, sucrose, heat shock and polyamines, respecitvely. CPuORF 40 was shown to respond to thermospermine in section 3.2.1. (Figure 3.1). All CPuORFs work through aCUTS except for the rCUTS bZIP11 and PAO2. In aCUTS CPuORFs the application of their respective signal will increase mORF expression whereas, rCUTS CPuORFs repress and lower mORF expression (Figure 1.20). The luciferase assay was used to determine whether these CPuORFs are necessary and sufficient to retain their signal response outside their native context.

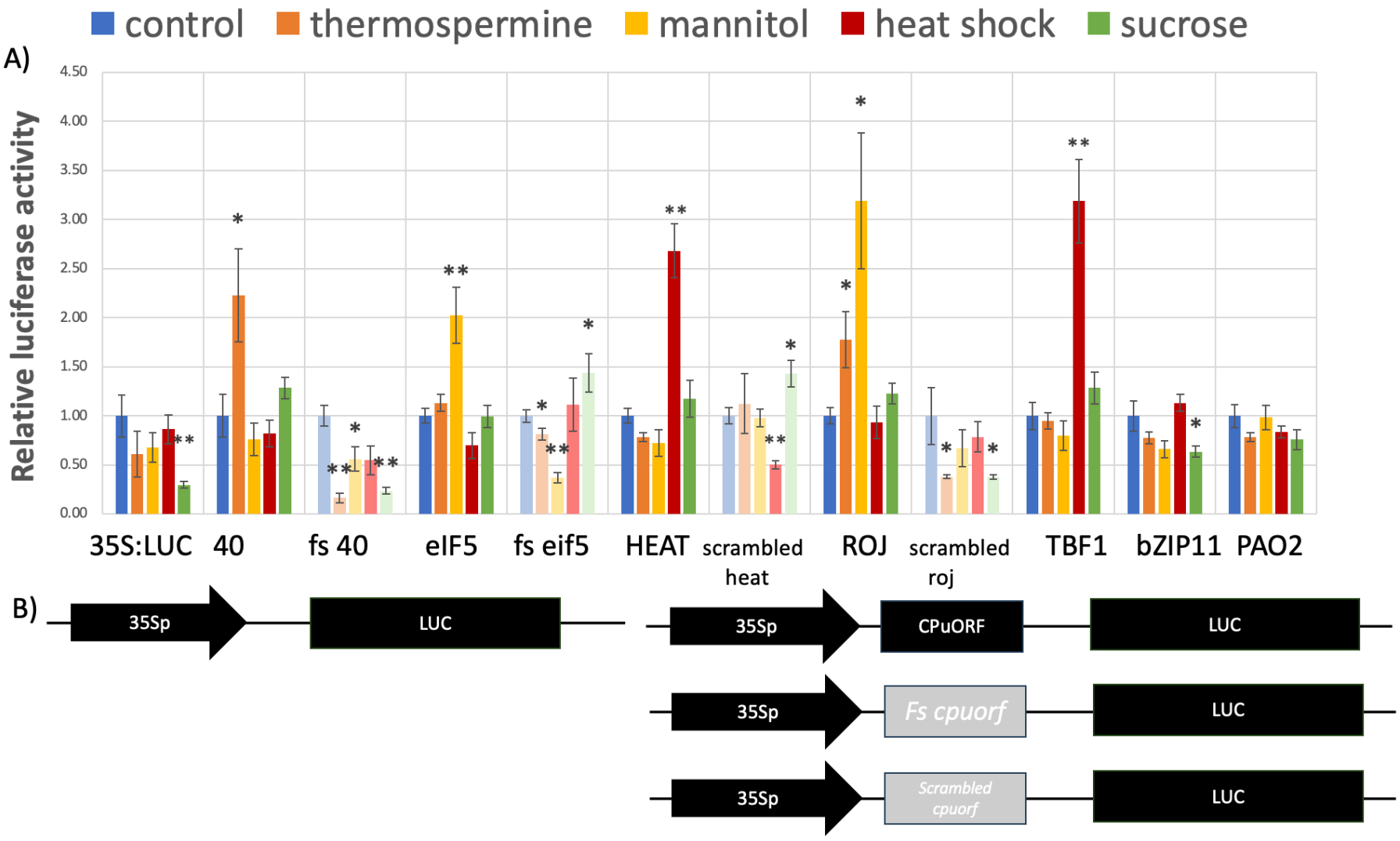

control
thermospermine
mannitol
heat shock
sucrose
A)
Relative luciferase activity
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
35S:LUC
40
fs 40
eIF5
fs eif5
HEAT
scrambled heat
ROJ
scrambled roj
TBF1
bZIP11
PAO2
B)
35Sp
LUC
35Sp
CPuORF
LUC
35Sp
Fs cpuorf
LUC
35Sp
Scrambled cpuorf
LUC

Figure 3.5. CPuORFs respond to specific signals. A) A bar chart presenting relative luciferase activity that normalizes the effect of a signal on mORF translation to control conditions for each transgenic line. The stimuli that have been applied are: 0.5MS + 0.05 mM thermospermine (orange), + 300 mM mannitol (yellow), + heat shock at 37°C (red), + 560 mM sucrose (green). Control conditions (either 0.5 MS or room temperature) are represented in blue. Metabolite treatments were applied overnight, while heat shock and RT conditions were applied for 8 hours. Frameshift and scrambled CPuORF mutants are presented in pale colours where available. For comparison, the 35S:LUC transgenic line was used as a reference control. Statistical significance between the signal and control conditions was assessed using a T-test, with error bars denoting the standard error of the mean. The statistical analysis was a four-way comparison comparing WT untreated against WT treated, fs untreated and fs treated B) Schematic representation of the 35S:LUC, CPuORF:LUC and frameshift and scrambled CPuORF construct design. Data is collected from 10-29 replicates.

Treatments on the 35S:LUC control lines all caused a slight decrease in mORF translation (Figure 3.5). When compared to untreated condition, elevated thermospermine, mannitol, heat shock and sucrose caused a reduction in mORF translation by 39, 32, 14 and 70% respectively. In contrast, lines where LUC is preceeded by CPuORF 40, eIF5, HEAT, TBF1 and bZIP11 all retain their reported conditional response outside their native context. For example, only elevated thermospermine elicited a statistically significant fold change (2.23 fold) in CPuORF40:LUC lines. A similar story is observed in eIF5:LUC where only drought simulation caused a statistically significant increase in luciferase activity (2 fold). Heat shock caused a statistically significant increase in HEAT:LUC and TBF1:LUC lines (2.68 fold and 3.19 fold respectively). Elevated thermospermine caused a significant fold increase of 1.22 in ROJ:LUC lines, but this was less than the increase caused by mannitol application.

PAO2 has been shown to respond to the polyamines spermidine and putrescine but not thermospermine (132,133). Elevated spermidine and putrescine were not tested here

however, thermospermine was. There was no significant difference between mORF expression under control conditions compared to elevated thermospermine in PAO2:LUC lines.

ROJ:LUC demonstrated an unreported signal response when tested (Figure 3.5). ROJ showed an increase in mORF translation in response mannitol (as reported) but, unusually, also to thermospermine. These signals resulted in a fold change compared to control conditions of 3.2 and 1.7 respectively.

In comparison to the wild type lines, frameshifted and scrambled CPuORFs do not show the increase in mORF translation (Figure 3.5). In fact, the general reduction in mORF translation mirrors what is observed in the control 35S:LUC lines, which lacks any CPuORF.

When considering mORF expression generally decreased under treatment in 35S:LUC and frameshifted/scrambled CPuORFs this highlights the regulation by CPuORF in response to their signals (Figure 3.5). Furthermore, the data supports the claim that the CPuORF peptide sequence is necessary and sufficient to conditionally modulate mORF expression outside their native context. Furthermore, the fact that each CPuORF modulates mORF expression specifically to a signal indicates that the CPuORF signal response is specific and not a generic stress response.

3.2.4. Mutations in the ribosome abolish CPuORF mediated ribosome stalling.

Ribosome profiling data, cryo-EM data and western blotting has indicated that CPuORFs stall the ribosome during translation (Table 1.4). Furthermore, a genetic screen has functionally implicated CPuORF 40 and three ribosomal exit tunnel proteins (uL4z, uL16z and RACK1z) (Figures 1.13-18). Mutations in CPuORF 40 and uL4z, uL16z and RACK1z suppressed an *Arabidopsis* mutant phenotype in a thermospermine synthase gene (*ACL5*). Interestingly, these ribosomal proteins are found in the exit tunnel and evidence indicates that they have a role in the stalling mechanism of CPuORF 40 (Figures 1.13-18). To further explore these mutations, this thesis used publicly available cryo-EM data to map these proteins in the exit tunnel (Figure 3.6).

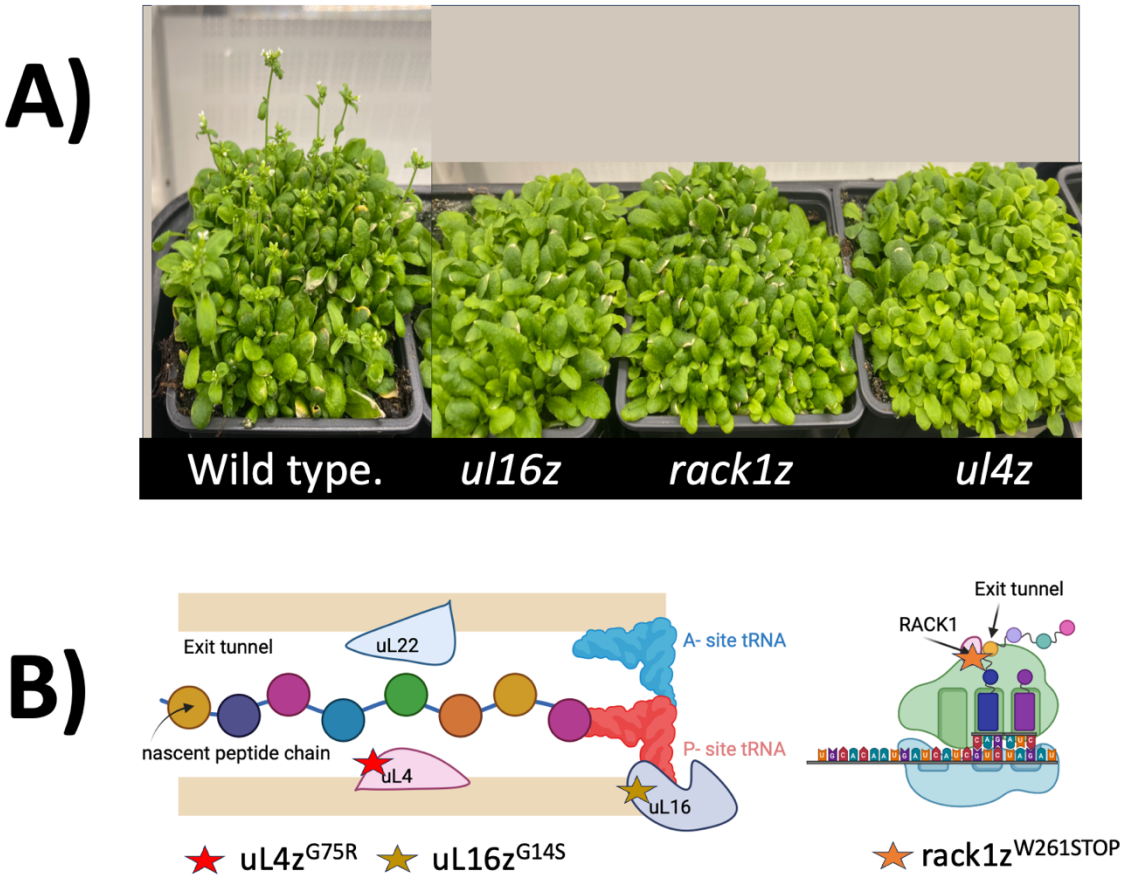


Figure 3.. Late flowering phenotypes in *uL16z, rack1z* and *uL4z* and diagrammatic representation of the location of the mutations. A) Late bolting and flowering phenotypes of *uL16z, rack1z* and *uL4z*. Mutant lines are grown alongside wild type *Arabidopsis*. The wild type has bolted and begun flowering, while, the ribosome mutants are still vegetative. B) A diagrammatic representation of the location of the mutations in *uL16z, rack1z* and *uL4z*. The ribosome mutations *uL16z*^{G14S} (yellow star), *rack1z*^{W261STOP} (orange star) and *uL4z*^{G75R} (red star) were located on a publicly available cryo-EM map of a translating wheat germ ribosome (PDB: 474E). For clarity the mutations have been depicted schematically. The *uL4z*^{G75R} (red star) mutations is a substitution of a glycine for an arginine on the exit tunnel constriction formed by uL4 and uL22. To be explicit, the mutation is found within the exit tunnel on uL4 at the constriction tunnel interface. The *uL16z*^{G14S} (yellow star) mutation is found on the finger like projection of uL16 into the exit tunnel and PTC. The mutation is found in proximity to the P-site tRNA and 28S rRNA. The mutation is a substitution of a glycine for a serine. The *rack1z*^{W261STOP} (orange star) is a tyrosine for a stop codon. The early stop mutation truncates the RACK1 protein by a third.

To map the mutations this thesis investigated a publicly available cryo-EM map (PDB: 474E) of a translating wheat germ ribosome (Figure 3.6 C-D). The mutation in uL16z is in proximity to the PTC and the 28S rRNA. Specifically, the mutation is within proximity to 28s rRNA^{U1128} and the protein is in proximity to the P-site tRNA (Figure 3.6). uL16 is

important in the delivery of tRNAs to the PTC and bridging the subunits. The mutation in RACK1 introduces an early termination codon that truncates the protein by a third. RACK1 stabilises collided stalled ribosomes and is a scaffold between ribosome subunits. The severe mutation may lead to the degradation of RACK1, or it may be recruited into the ribosome and is possibly dysfunctional. The deletion of the C-terminal end of RACK1 (cyan) may interfere with RACK1's ability to interact with and stabilise stalled ribosomes. Finally, the uL4a^{G75R} mutation is found at the constriction between uL4 and uL22. This constriction within the exit tunnel has been reported to interact with nascent peptides that stall the ribosome during translation (20). Furthermore, this constriction is important in the kinetics and translocation of the ribosome nascent peptide during translation. Because of its location it is highly probable that this mutation affects an arrest peptide's ability to stall the ribosome during translation.

Known ribosome mutations in plants can lead to a wide range of phenotypes from embryo lethality to no phenotype (2). Furthermore, mutations in ribosome proteins may cause pointed leaves, reduced root and shoot growth and increased ploidy in cells (2). The mutations investigated here are not lethal and can be described as showing late flowering and reduced shoot growth as described in Imai et al., (37,118) As the mutant phenotypes are not severe it is likely the mutations do not affect ribosome integrity.

The suppressor screen in Imai *et al.,* has potentially implicated these mutations *uL16z^{G14S}, rack1z^{W261STOP}* and *uL4z$^{G75R.}$* in the stalling mechanism of SAC51 CPuORFs (37,118). The suppressor screen identified a mutation in the third and final CPuORF (CPuORF 40) in the 5' leader of SAC51 (37,118). This thesis has provided evidence that

CPuORF 40 is the functional *SAC51* CPuORF, as it retains thermospermine-responsive aCUTS activity, as reported for the native *SAC51* 5'leader (Figure 3.1). Collectively, the evidence suggests that the ribosome mutations identified by Imai *et al.*, *uL16z*^{G14S}, *rack1z*^{W261STOP} and *uL4z*^{G75R} disrupt the ribosome stalling mechanism in the SAC51 5' leader, specifically CPuORF 40. To test this hypothesis 35S:LUC, 35S:38-39-40:LUC (will be referred to as SAC51) and 35S:40:LUC were transformed into the ribosome mutant and a wild type lines. If the stalling mechanism is disrupted by the ribosome mutations, then *LUCIFERASE* mORF expression will be higher in those backgrounds compared to the wild type. mORF expression is quantified via a luciferase assay. 35S:LUC serves as a control for the effect of the mutations on transcripts not under the control of a CPuORF. Furthermore, the CPuORFs found in the eIF5, HEAT, ROJ, bZIP11 and TBF1 will be explored if their stalling mechanisms respond to the ribosome mutations. Moreover, frameshift (fs 40 and fs eif5) and scrambled (scrambled heat and scrambled roj) mutants from Figure 3.4 were also tested in the ribosome mutant lines.

Data collected from wild type lines is represented by dark blue, *uL16z* (orange), *RACK1z* (grey), and *uL4z* (yellow). Frameshift or scrambled CPuORF mutants are represented in pale colours. Luciferase data was collected from transgenic T1 leaves and normalized to the wild type background. Asterisks denote statistically significant differences as determined by T-tests: **p<0.01.

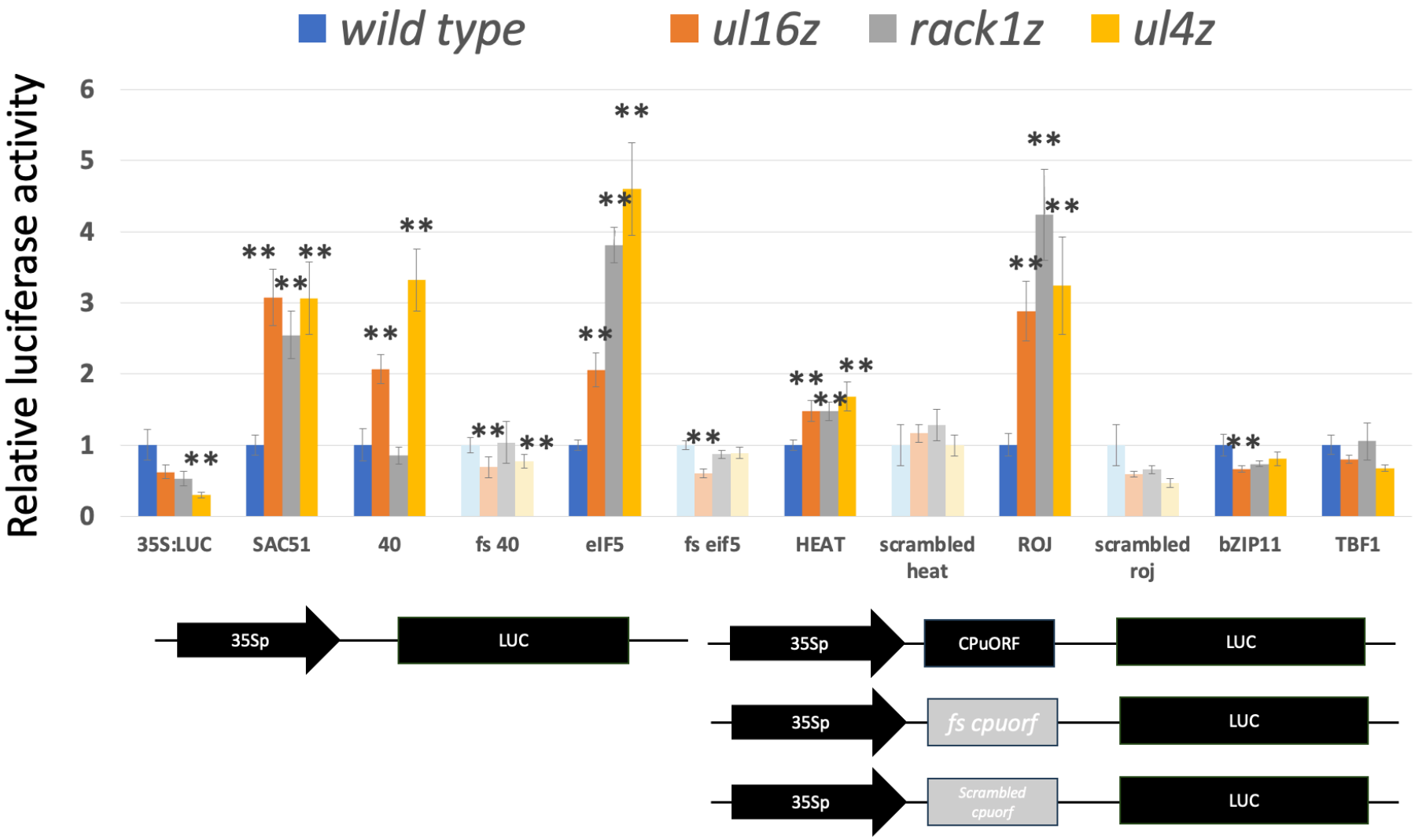

wild type
ul16z
rack1z
ul4z
Relative luciferase activity
0
1
2
3
4
5
6
35S:LUC
SAC51
40
fs 40
eIF5
fs eif5
HEAT
scrambled heat
ROJ
scrambled roj
bZIP11
TBF1
35Sp
LUC
35Sp
CPuORF
LUC
35Sp
fs cpuorf
LUC
35Sp
Scrambled cpuorf
LUC

Figure 3.7. Specific ribosome mutations abolish CPuORF-led ribosome stalling. A) The bar chart displays luciferase data collected from transgenic *Arabidopsis* leaves with transformations of either 35S:CPuORF:LUC constructs or control constructs (35S:LUC, frameshift, or scrambled CPuORF mutants). Data is normalized to the wild type background. Different ribosomal mutations are indicated: uL16 (orange), RACK1z (grey), and uL4z (yellow). Frameshift and scrambled mutants are represented in pale colours. Error bars represent the standard error of the mean (SEM). Asterisks denote statistically significant differences, determined by Student's t-test: **p<0.01, *p<0.05. B) A schematic representation of the constructs used is provided below the graph. N ranges from 8-27 replicates.

When examining the 35S:LUC control lines, mutations in the ribosome led to a general reduction in luciferase activity (Figure 3.7). Mutations in uL16z, RACK1z, and uL4z resulted in reductions in luciferase activity by 38%, 47%, and 71%, respectively. Notably, the mutation in the constriction tunnel protein, uL4z, caused a statistically significant decrease in mORF translation in the control 35S:LUC construct. This data suggests that ribosome mutations may cause a general decrease in translation initiation efficiency.

Notably in SAC51 there is a significant increase in mORF expression in all three ribosome mutant lines when compared to the wild type (Figure 3.7). The mutations in uL16z, RACK1z, and uL4z resulted in a significant fold change of 3.1, 2.5 and 3.1 respectively. These results indicate that all three ribosome mutations lead to an increase in mORF expression when compared to the wild type background. Furthermore, this could suggest that the ribosome mutations have weakened or abolished ribosome stalling by the CPuORFs in the SAC51 5'leader. This is strengthened by the fact that in 35S:LUC lines the ribosome mutations cause a general decrease in mORF expression when compared to the wild type background.

Remarkably, the aCUTS CPuORFs eIF5, HEAT and ROJ also respond to the same ribosome mutations as they elicit a significant increase in mORF expression when compared to their respective wild types (Figure 3.7). For eIF5, mutations in uL16z, RACK1z, and uL4z resulted in a significant fold change of 2.1, 3.8 and 4.6 respectively. For HEAT, the mutations caused a significant fold change of 1.5, 1.4 and 1.7 respectively. For ROJ, ribosome mutations resulted in a significant fold change of 2.9, 4.2 and 3.2 respectively. Consequently, the results suggest that in the case of SAC51, eIF5, HEAT and ROJ, their stalling mechanisms are all disrupted by the same ribosome mutations. This suggests that a common mechanism between the stalling mechanisms of these CPuORFs. For CPuORF mutations in uL16z, RACK1z, and uL4z resulted in a fold change of 2.1, 0.8 and 3.1 respectively. Meaning that there was a significant increase in mORF expression in *ul16z* and *ul4z* but not in *rack1z*. This is surprising as the suppressor screen identified a suppressor mutation in CPuORF 40 alongside the ribosome mutants, translation data from this thesis (Figure 3.1) has indicated CPuORF 40 is functional and other aCUTS CPuORFs with different peptide sequences respond to all three ribosomal mutations. This implies that another *SAC51* CPuORF might be responding to the mutation in RACK1Z. The translation assay in 3.2.1. also indicated that CPuORF 38 is functional, and this may be the *SAC51* CPuORF that is responsive to this mutation.

Outside the realm of aCUTS CPuORFs, the rCUTS bZIP11 CPuORF did not react to any tested ribosomal mutation (Figure 3.6). In fact, mutations in the ribosome led to a general reduction in mORF translation, akin to what was seen in the 35S:LUC and frameshift/scrambled mutants. These mutations resulted in decreases of 34%, 27%, and 19% in bZIP11, respectively (Figure 3.5). An explanation for this is that only aCUTS

CPuORFs are expected to stall without a signal and they are prevented from stalling by the three mutants. However, rCUTS CPuORFs are not expected to stall without a signal, so disruption of stalling should not be observed.

A similar observation is made upon investigation of the dual CUTS TBF1 CPuORF. Ribosome mutant lines expressing the TBF1:LUC construct do not elicit an increase in mORF translation when compared to the wild type background. This may suggest that the TBF1 stalling mechanism is independent from other aCUTS CPuORFs despite its reported aCUTS activity (9,128,129).

For the frameshift and scrambled CPuORFs, ribosomal mutations caused a decrease in mORF translation compared to the WT (Figure 3.7). This is because in the ribosome mutations are preventing the CPuORF stalling the ribosome during translation.

Overall, the data supports the hypothesis that there are commonalities between ribosome stalling mechanisms. This study has shown that in four CPuORFs, each of which respond by aCUTS to a different signal, stalling is disrupted by a set of specific single amino acid substitutions in ribosome exit tunnel proteins, indicating a common stalling mechanism between aCUTS CPuORFs.

3.2.5. Ribosomal mutations abolish stalling and signal sensing.

Mutations in the ribosome lead to an increase in mORF translation in aCUTS CPuORFs (Figure 3.7). To determine whether these ribosomal mutations completely eliminate stalling or impair signal sensing, a luciferase assay was conducted in T1 plants

transformed with 35S:CPuORF:LUC in both wild type and ribosomal mutant backgrounds. T1 leaves were subjected to control conditions (0.5 MS or room temperature) and to treatment conditions specific to the aCUTS CPuORFs under investigation (Figure 3.8). Data for CPuORF40, eIF5, HEAT, and ROJ are presented in a box and whisker plot (Figure 3.8). Statistical significance between control and treatment conditions for each background was assessed using a student's t-test, with **$p<0.01$ and *$p<0.05$ denoting significance levels.

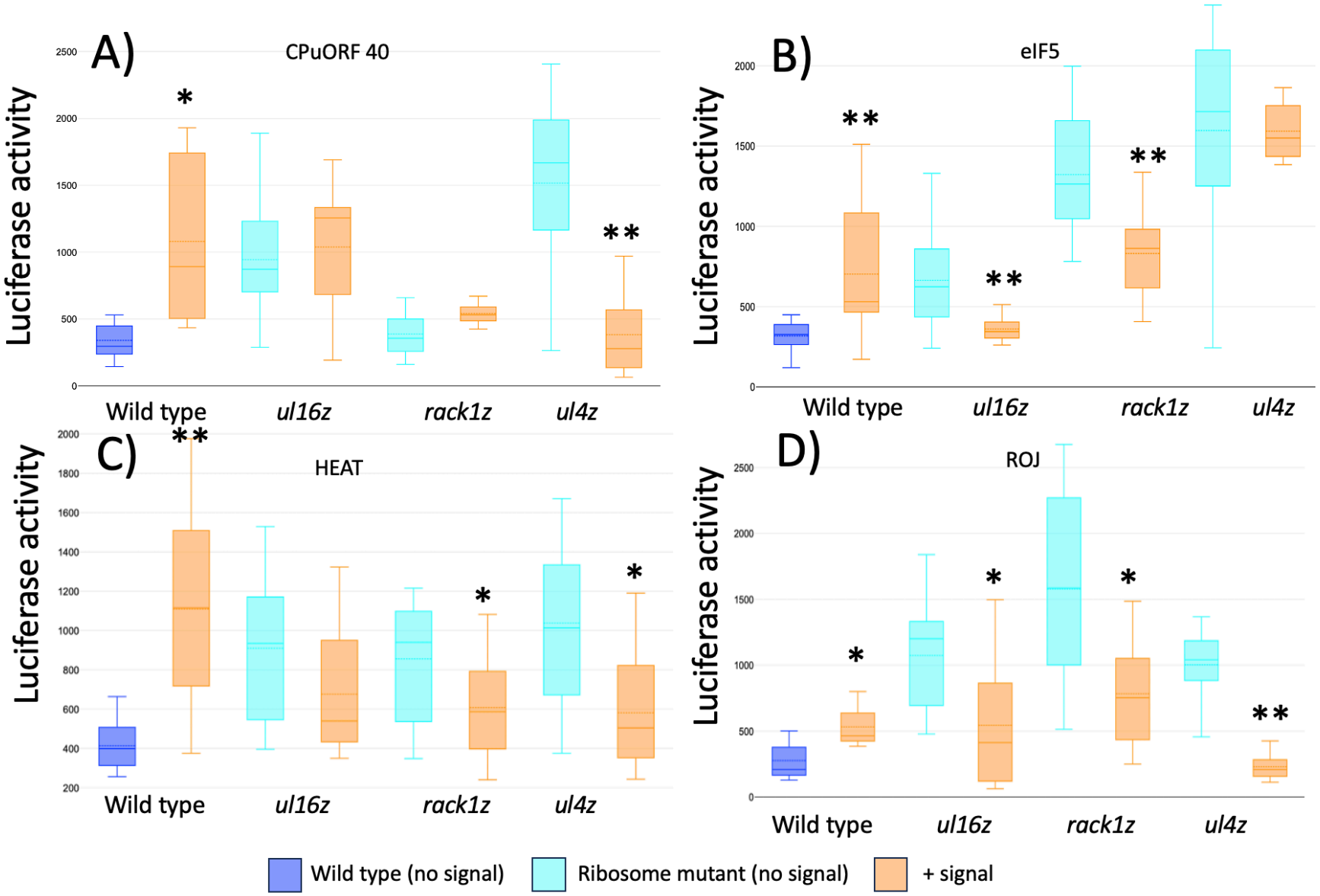

A)
CPuORF 40
Luciferase activity
2500
2000
1500
1000
500
0
*
**
Wild type
ul16z
rack1z
ul4z
B)
eIF5
Luciferase activity
2000
1500
1000
500
0
**
**
**
Wild type
ul16z
rack1z
ul4z
C)
HEAT
Luciferase activity
2000
1800
1600
1400
1200
1000
800
600
400
200
**
*
*
Wild type
ul16z
rack1z
ul4z
D)
ROJ
Luciferase activity
2500
2000
1500
1000
500
0
*
*
*
**
Wild type
ul16z
rack1z
ul4z
Wild type (no signal)
Ribosome mutant (no signal)
+ signal

Figure 3.8. Ribosomal mutations abolish the conditional aCUTS response. Four box and whisker plots illustrate the effects of ribosomal exit tunnel mutations on aCUTS CPuORFs, assessed via a luciferase assay using a luciferase reporter gene. The results are displayed across four box and whisker plots labelled A-D, each representing data for CPuORF40 (A), eIF5 (B), HEAT (C), and ROJ (D). Experiments were conducted both in wild type backgrounds (Ler) and in ribosomal mutants (*ul16z, rack1z,* and *ul4z*). The plots distinguish between data collected under control conditions (0.5 MS or room temperature, shown in dark blue for the wild type and light blue for ribosomal mutants) and upon the application of a treatment specific to each CPuORF (0.5 MS + 0.05 mM thermospermine for CPuORF40, +300 mM mannitol for eIF5, and 37°C heat shock for both HEAT and ROJ, all shown in orange). Statistical significance was determined using a t-test to compare signal application and control conditions for each genotype. The aim was to determine if a signal elicits the aCUTS response in mutant lines, in comparison to the wild type. Asterisks indicate the levels of significance: * for $p<0.05$ and ** for $p<0.01$.

When treatments were applied in the CPuORF:LUC ribosomal mutant backgrounds, there was either a statistically significant decrease in mORF expression or no statistical difference in mORF expression (Figure 3.8). Signals resulted in a notable reduction in mORF translation relative to control conditions for CPuORF 40 (ul4z), eIF5 (ul16z and rack1z), HEAT (rack1z and ul4z), and ROJ (ul16z, rack1z, and ul4z) (Figure 3.13). Notably, no CPuORF exhibited its typical aCUTS response in a ribosomal mutant background. This suggests that the ribosomal mutation either completely disrupts stalling or impedes the signal sensing aCUTS mechanism. If there is still some level of stalling, it seems that the CPuORF cannot conditionally adjust translation. In frameshifted and scrambled CPuORF variants and 35S:LUC, signals generally caused a reduction in mORF expression as observed under treatment conditions in the ribosome mutant lines. Therefore, the similar mORF expression profiles might indicate that the ribosomal mutations have abolished stalling and rendered the CPuORFs effect on mORF expression to that of a regular uORF.

3.2.6. *In vitro* investigations suggest that any mutation in the CPuORF weakens stalling.

The molecular mechanism by which CPuORFs stall the ribosome during translation has yet to be elucidated. Mutations in the ribosome can weaken or abolish CPuORF-mediated ribosomal stalling (Figure 3.7-8). The CPuORF classification system may indicate that aCUTS CPuORFs have a modular nature (Figure 1.12). Molecular and ribosomal profiling data suggest that the conserved amino acids play a role in stalling the ribosome, while the divergent N-terminus is responsible for sensing signals.

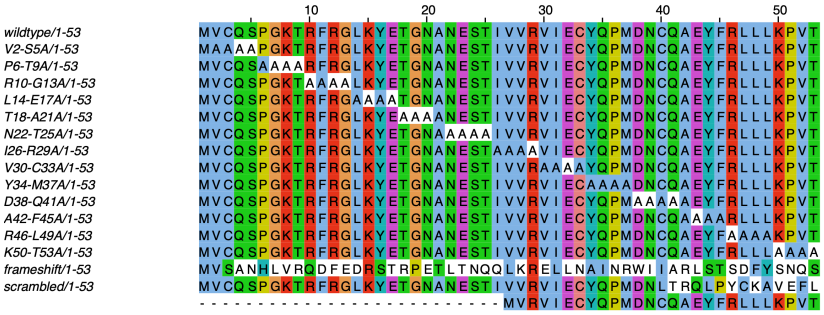
To investigate this, wild type and mutant CPuORFs were inserted upstream of a luciferase reporter gene. This study took an alanine scanning approach and substituted 4 consecutive amino acids sequentially throughout each CPuORF peptide. Evidence suggest that ribosome stalling is sequence dependent and conferred by the CPuORF peptide sequence (Figure 3.4). Therefore, by taking an alanine scanning approach, mutagenesis can map functional domains across the CPuORF peptide. Furthermore, by taking an alanine scanning approach, this study can test if there is a correlation between conserved residues and ribosomal stalling and if variable amino acids confer signal sensing. For example, Class I CPuORFs have a conserved C-terminus and a variable N-terminus. If CPuORFs contain a conserved stalling domain, mutations of the C-terminus would result in reduced stalling and higher levels of luciferase mORF expression. An example of the nomenclature for the alanine scanning mutants is “V2-S5A”. For concise nomenclature, in this example V and S refer to the range of amino acids from positions 2 to 5 have been substituted for alanine, respectively. The final A represents the change to

alanine for all four amino acids. Data for alanine scanning mutants will be identified by light blue whereas, wild type data is shown in dark blue.

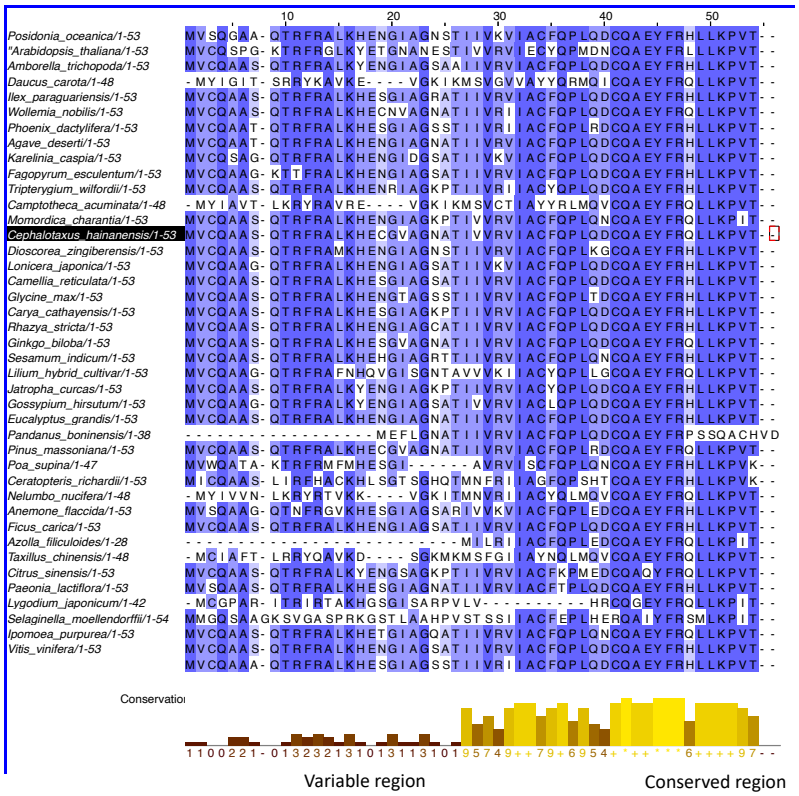
Alongside alanine scanning mutagenesis, larger mutations to investigate the CPuORFs peptide sequence were produced and investigated. These mutants include the previously used frameshift variant. Frameshift mutations have inserted an extra DNA base to change the entire peptide sequence with minimal changes to the RNA sequence. Furthermore, the frameshift mutants have been edited to maintain the wild type CPuORF peptide length. Edits also included removing any early stop codons and replacing the native stop codon, post frameshift. By frameshifting the CPuORF DNA sequence, this changes the CPuORF peptide sequence to test if CPuORF function is conferred by the peptide. Also, a mutant called "scrambled" was investigated and in this mutant the conserved amino acids of the CPuORF were rearranged. Notably, this mutant maintains the same length and exact amino acid composition as the wild type but the order of the amino acids is changed. Additionally, C- or N-terminal deletion mutants were generated to remove variable amino acids and leaving the conserved region for each respective CPuORF. The CPuORFs examined include CPuORF40, HEAT, ROJ, and eIF5, which are classified as Class I, I, I, and IIa, respectively. For Class I CPuORFs, the C-terminus encompasses the conserved region, while for eIF5, the N- and C-termini represent its conserved and variable regions, respectively. These mutants aim to assess whether the conserved regions of the CPuORFs are critical for their aCUTS function. These mutants will be called whole-peptide mutants (frameshift, scrambled, C-terminus and N-terminus) and the data for these will be shown in orange.

Initially, alanine scanning and whole peptide mutants were investigated in a commercially available uncoupled wheat germ extract kit (Promega) detailed in Methods 2.6. The effect of these mutations on mORF expression will be evaluated alongside a wild type CPuORF to determine their effect on stalling. If there are significant increases in mORF expression this will imply that the mutation has weakened or abolished the CPuORF peptides ability to stall the ribosome. As *in vitro* assays cannot test signal sensing the results from this investigation will inform what CPuORFs to explore *in vivo*. The CPuORFs to be investigated in this way will be Class I CPuORFs CPuORF 40, ROJ, HEAT and Class IIa eIF5. These CPuORFs and their mutant variants will be driven by an SP6 promoter and will be upstream a luciferase reporter gene. A luciferase assay will be performed on 6-9 replicates.

A)



B)



Class I

Figure 3.9. ClustalX alignments for CPuORF40 and the mutants investigated in this study and CPuORF40 conservation. A) An alignment for the wild type CPuORF 40 amino acid sequence alongside the sequences for the alanine scanning mutants and whole peptide mutants. The unchanged amino acids are highlighted in yellow. B) An alignment of CPuORF 40 peptide conservation across taxa. Below the ClustalX conservation score is plotted below. CPuORF 40 is a Class I CPuORF, and its conserved region (dark yellow) and variable regions (pale yellow) have been labelled.

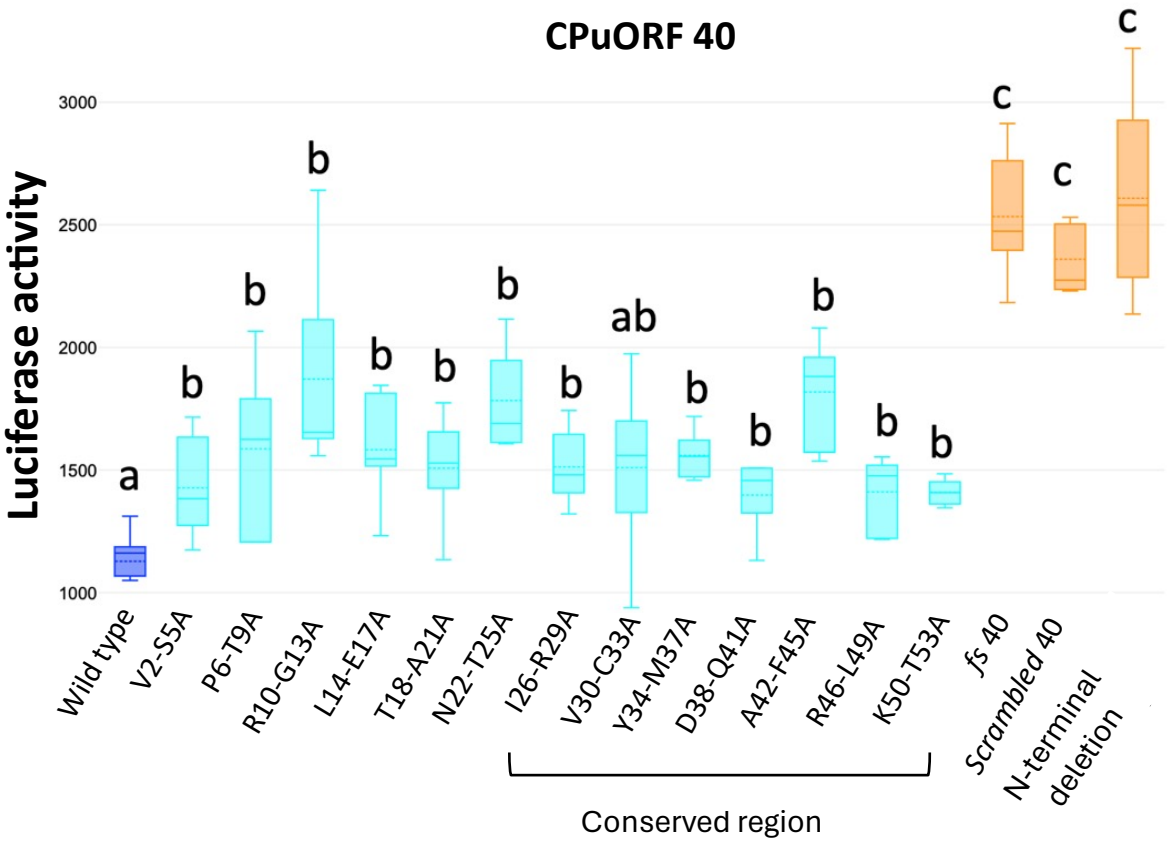


Figure 3.10. A box and whisker plot to show any mutation in CPuORF 40 weakens its ability to stall the ribosome. An *in vitro* luciferase assay was performed on CPuORF40 and its alanine and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. An ANOVA Tukey test was used to determine significance and cluster datasets. The wild type appears in dark blue, alanine scanning mutants are in light blue, and full peptide mutants are in orange. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 6-9 replicates. The alanine scanning mutations that hit the conserved region have been identified below (I26-T53).

The impact of wild type CPuORF40 and its mutants on mORF translation was examined using an *in vitro* luciferase assay (Figure 3.9-10). ClustalX alignments of CPuORF40 and the mutants investigated in this study and CPuORF 40 peptide conservation have been provided in Figure 3.9. Furthermore, the conserved and variable regions have been

labelled. Luciferase data is illustrated as a box and whisker plot and an ANOVA Tukey test was performed to determine statistical significance and cluster data sets (Figure 3.10).

For CPuORF 40, all mutations resulted in a statistically significant increase in mORF translation when compared to the wild type (Figure 3.10). Whole peptide mutations (orange) displayed the most substantial increase in mORF translation when compared with the alanine scanning (light blue) mutants. Specifically, frameshift, scrambled, and the N-terminal deletion (conserved region) mutations caused fold changes of 2.22, 2.13, and 2.29, respectively when compared to the wild type. This contrasts with the alanine scanning mutants, which only increased between 1.2 and 1.6 times when compared to the wild type. This disparity suggests that whole peptide mutants may have completely abolished ribosome stalling. Whereas the alanine scanning mutants partially abolish ribosome stalling.

All alanine scanning mutations elicited a significant increase of mORF translation when compared to the wild type except for V30-C33A. Although V30-C33A is statistically similar to the other alanine scanning mutations, it is also statistically similar to the wild type. Consequently, V30-C33A clusters with both the wild type and the alanine scanning mutants. It is surprising that there is no obvious difference in effect of alanine scanning between conserved and variable regions of CPuORFs.

The N-terminal deletion mutant, which retains only the conserved region, does not independently induce ribosome stalling. The scrambled mutant, even with its intact variable region but rearranged conserved amino acids, also fails to induce ribosome

stalling. These results highlight the importance of the whole peptide sequence in its native order for ribosomal stalling. Furthermore, there is no statistical difference between the effect of a mutation in the conserved C-terminus and the variable N-terminus. CPuORF40 is a class I CPuORF, if it had a modular nature, it would be expected N-terminal deletions would affect stalling more severely. In fact, the alanine scanning mutations eliciting the highest average luciferase activity are R10-G13A and A42-F45A. R10-G13A and A42-F45A are mutations in the N- and C-termini respectively.

Collectively, these findings emphasize the influence of the full length and primary structure of the CPuORF peptide in conferring ribosomal stalling. Moreover, the observation that any group of four alanine substitutions within the CPuORF affects stalling suggests that the residues that the interactions responsible for ribosome stalling are distributed throughout the CPuORF peptide.

Mutations in ROJ CPuORF show a similar outcome to those of CPuORF40 (Figures 3.9-12). All mutations assessed in ROJ led to an increase in mORF translation (Figure 3.12). Notably, the frameshift, scrambled, and N-terminal deletion mutants prompted fold changes of 1.85, 1.96, and 1.83, respectively, slightly higher than the 1.2-to-1.65-fold increments elicited by the alanine scanning mutants. This again might indicate that the whole peptide mutations have fully abolished stalling, whereas the alanine scanning mutations partially affect stalling. The conserved C-terminal amino acids alone are insufficient for ribosome stalling. Similarly, the scrambled mutant, despite preserving its variable region, failed to maintain WT levels of mORF translation. This highlights the importance of the full length CPuORF peptide to stall the ribosome.

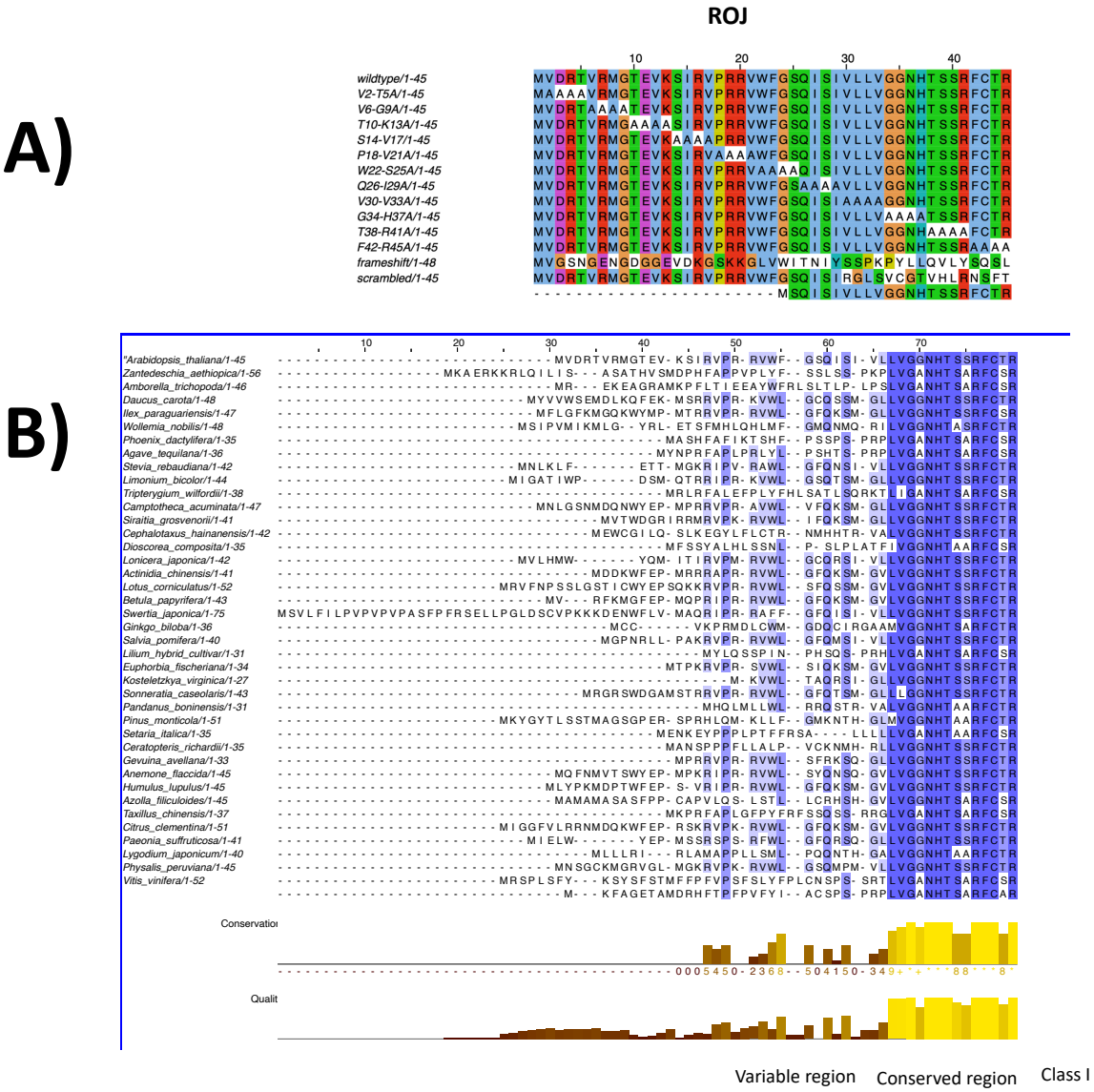


Figure 3.11. ClustalX alignments for ROJ and the mutants investigated in this study and ROJ conservation. A) An alignment for the wild type ROJ amino acid sequence alongside the sequences for the alanine scanning mutants and whole peptide mutants. The unchanged amino acids are highlighted in yellow. B) An alignment of the ROJ CPuORF peptide conservation across taxa. Below the ClustalX conservation score is plotted below. ROJ is a Class I CPuORF, and its conserved region (dark yellow) and variable regions (pale yellow) have been labelled.

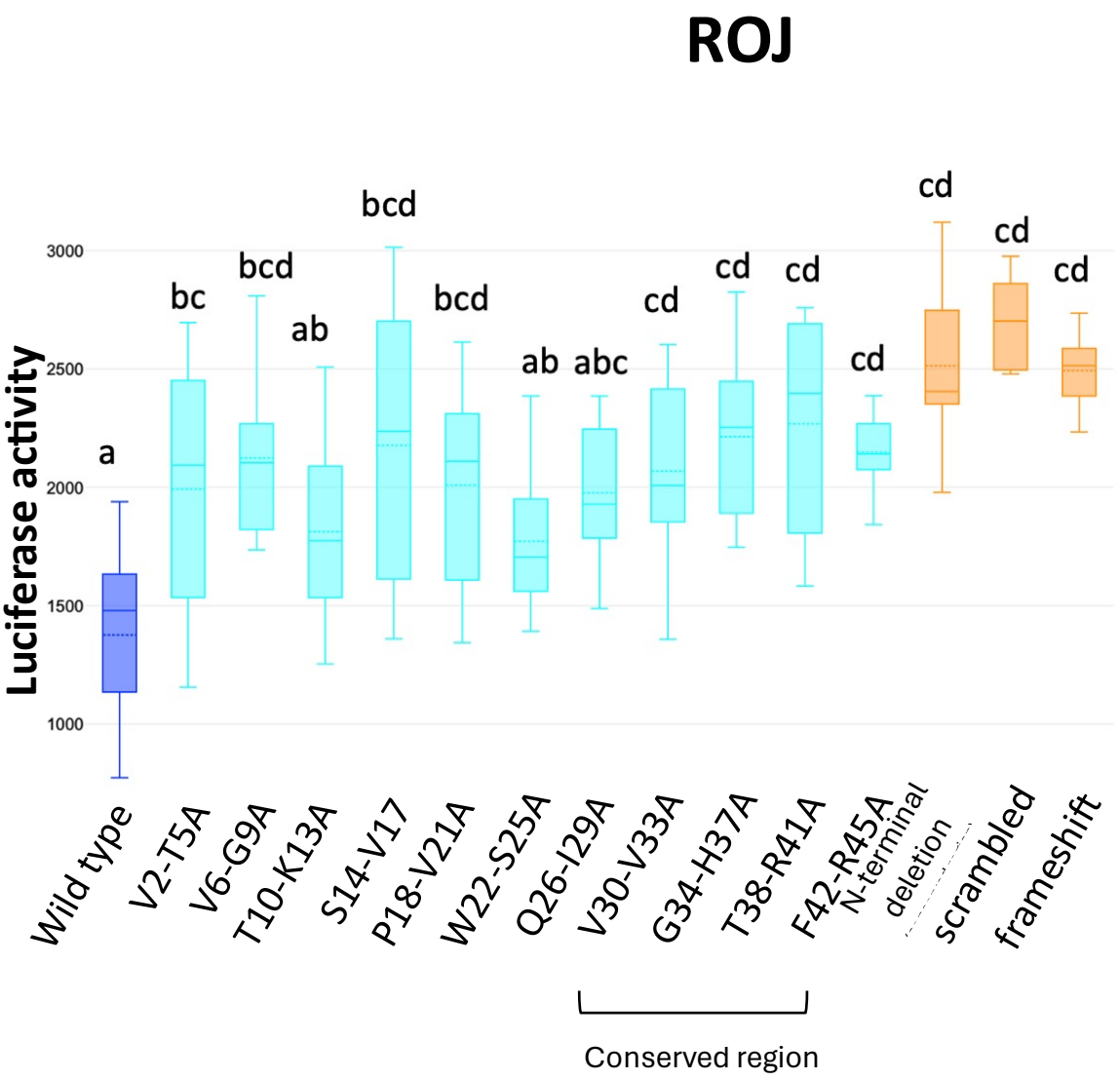


Figure 3.12. A box and whisker plot to show any mutation in ROJ weakens its ability to stall the ribosome. An *in vitro* luciferase assay was performed on ROJ and its alanine and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. An ANOVA Tukey test was used to determine significance and cluster datasets. The wild type appears in dark blue, alanine scanning mutants are in light blue, and full peptide mutants are in orange. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 6-9 replicates. The alanine scanning mutations that hit the conserved region have been identified below (V30-R45).

Considering the alanine scanning mutations, the ANOVA Tukey test suggests that there are differences between the effect of a mutation in conserved C-terminal residues as opposed to the mutations in the variable N-terminus (Figure 3.12). Mutations in the conserved region are more likely to a greater increase in mORF translation when compared to those in the N-terminus (Figure 3.12). The full peptide mutants cluster together in groups c and d and cluster with the conserved region C-terminal mutations (V30-V33A, G34-H37A, T38-R41A and F42-R45A). This may suggest that the variable amino acids T10-K13A, W22-S25A and Q26-I29A have a weaker influence on ribosome stalling than the conserved region. Consequently, in regards to ROJ that could be interpreted to support the hypothesis that CPuORF function is modular .

A similar analysis of mutations in the Class I HEAT CPuORF supports the findings from Class I CPuORFs 40 (Figures 3.9-14). All mutations to the HEAT CPuORF led to a significant increase in mORF translation (Figure 3.14). Whole peptide mutations brought about the highest increase in mORF translation compared to the alanine scanning mutants, but statistically cluster with most alanine scanning mutants. This is true except for the C-terminal I22-L25A and R34-R38A mutations that exclusively cluster in group b. This is surprising as HEAT is a class I CPuORF, and it is expected that conserved residues confer stalling. Interestingly, the ANOVA statistical analysis for HEAT is similar to CPuORF 40 despite being in the same homology group at CPuORF 40. Therefore, this may indicate that the hypothesis of a CPuORF modular organisation may be limited to specific CPuORFs.

Fold changes for the frameshift, scrambled, and N-terminal deletion (conserved region) mutations were 1.76, 1.97, and 2.07, respectively. The whole peptide mutations means and medians are all greater than the alanine scanning mutations. This may suggest that the alanine scanning mutations may represent an intermediate effect on stalling compared to the whole peptide mutants. Overall, *in vitro* data indicates the importance of the full length CPuORF in maintaining ribosome stalling.

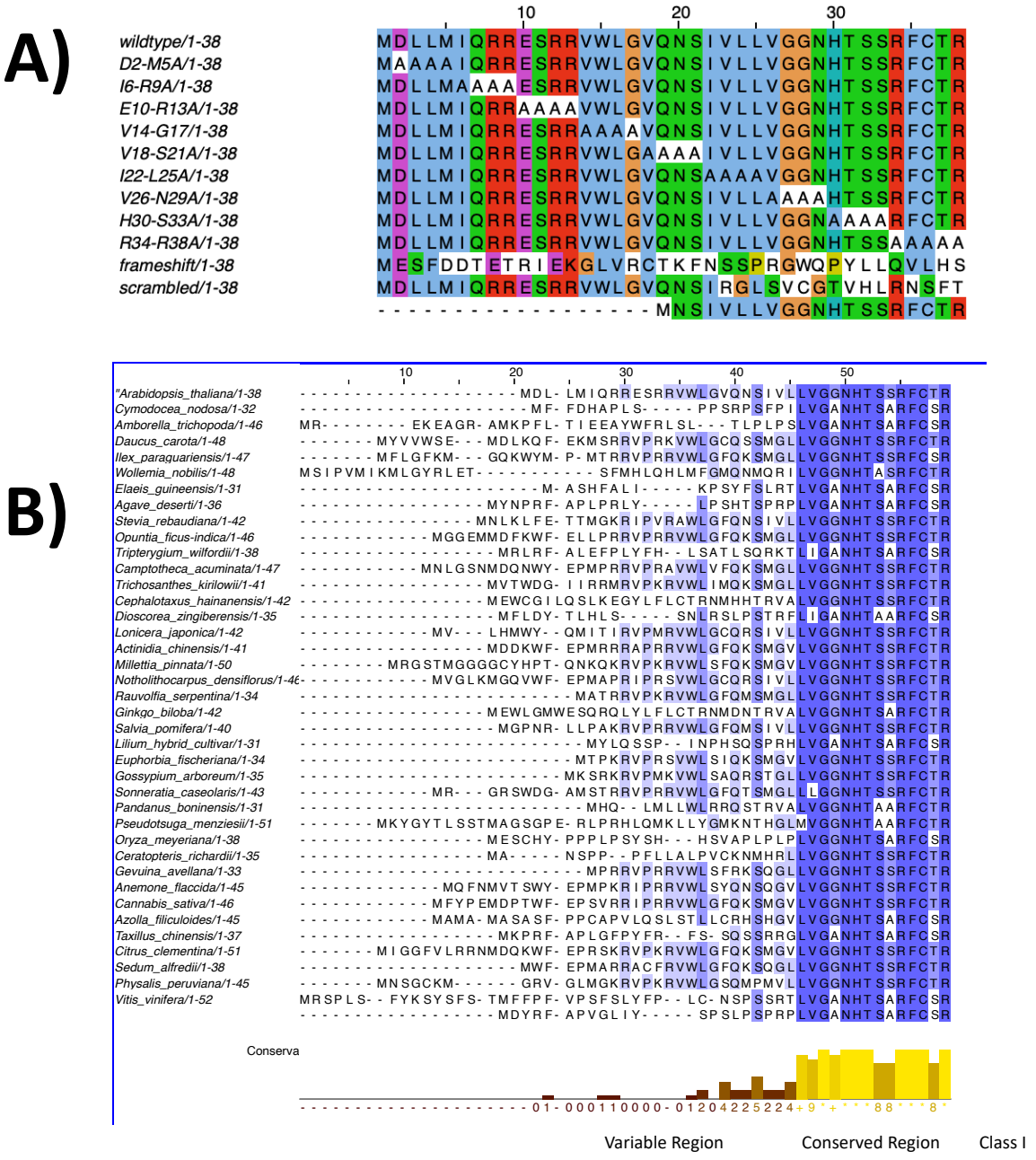


Figure 3.13. ClustalX alignments for HEAT and the mutants investigated in this study and HEAT CPuORF conservation. A) An alignment for the wild type HEAT amino acid sequence alongside the sequences for the alanine scanning mutants and whole peptide mutants. The unchanged amino acids are coloured in yellow. B) An alignment of the HEAT CPuORF peptide conservation across taxa. Below the ClustalX conservation score is plotted below. HEAT is a Class I CPuORF, and its conserved region (dark yellow) and variable regions (pale yellow) have been labelled.

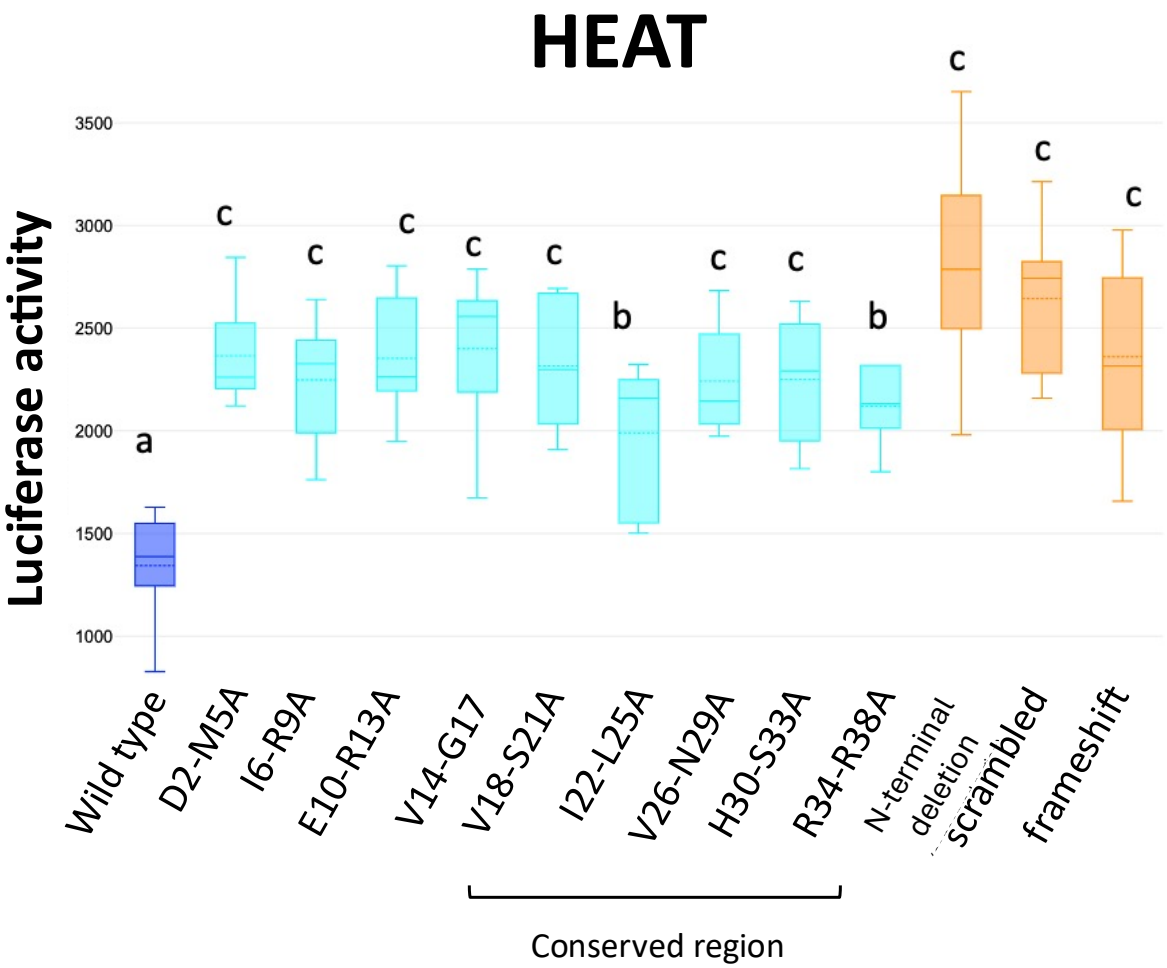


Figure 3.14. A box and whisker plot to show any mutation in HEAT weakens its ability to stall the ribosome. An *in vitro* luciferase assay was performed on ROJ and its alanine and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. An ANOVA Tukey test was used to determine significance and cluster datasets. The wild type appears in dark blue, alanine scanning mutants are in light blue, and full peptide mutants are in orange. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 6-9 replicates. The alanine scanning mutations that hit the conserved region have been identified below (V18-R38).

The eIF5 CPuORF is class IIa in contrast to the CPuORFs 40, ROJ and HEAT CPuORFs just examined. Despite this, alanine scanning mutations lead to an intermediatory increase in mORF translation when compared to the whole peptide mutants (Figures 3.15-16). The whole peptide mutant that elicited the greatest increase in mORF translation was the C-terminal deletion (variable region) line. This mutant just contained the conserved region

(C-terminal deletion) and caused an average increase of 92 % in mORF translation when compared to the wild type. Clustering with the C-terminal deletion lines are the rest of the whole peptide mutants including the variable N-terminal deletion line, the fameshift, and the scrambled variants. These mutants caused an increase 77, 55 and 47 % respectively. All whole peptide mutants are statistically higher than the wild type suggesting they have abolished ribosome stalling.

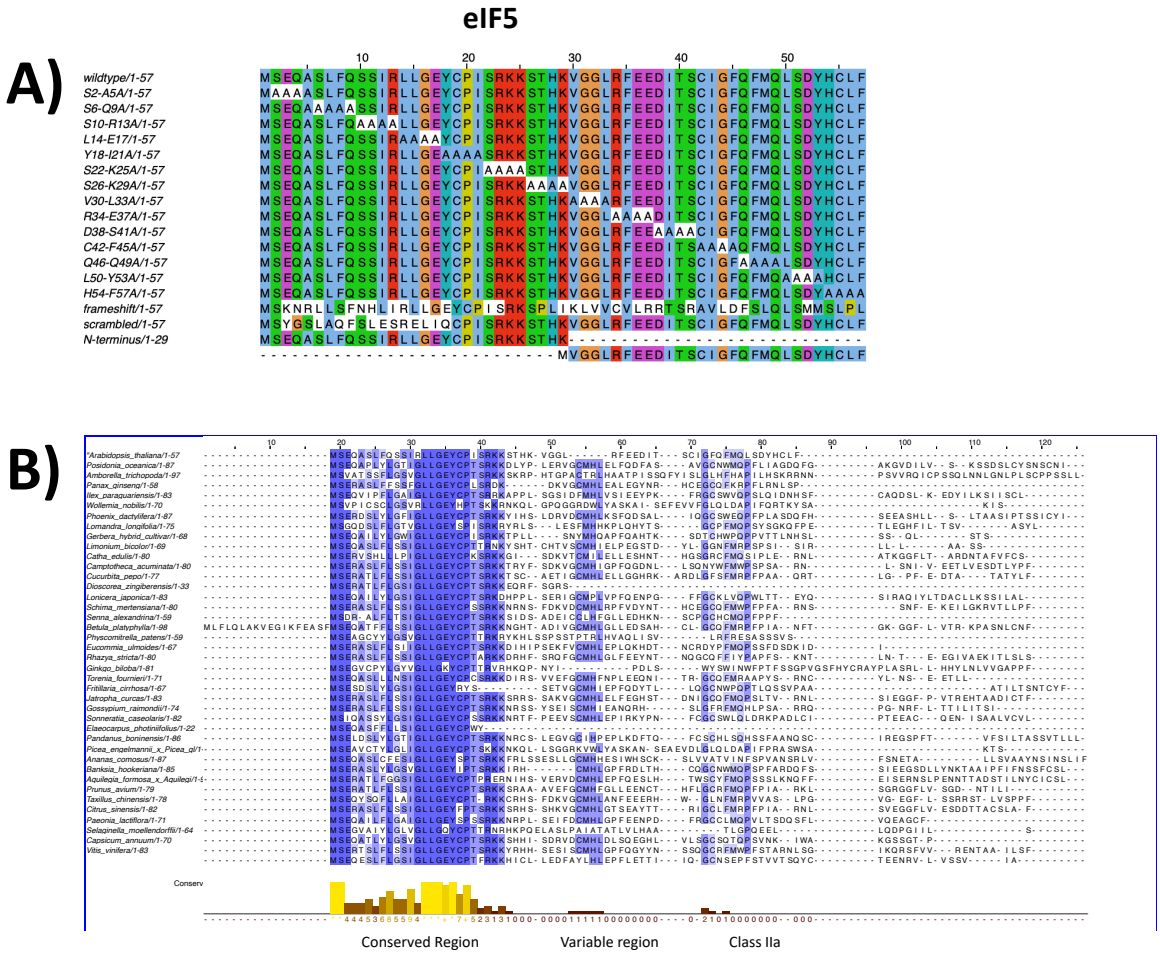


Figure 3.15. ClustalX alignments for eIF5 and the mutants investigated in this study and eIF5 CPuORF conservation. A) An alignment for the wild type eIF5 amino acid sequence alongside the sequences for the alanine scanning mutants and whole peptide mutants. Unchanged amino acids are highlighted in yellow. B) An alignment of the eIF5 CPuORF peptide conservation across taxa. Below the ClustalX conservation score is plotted below. eIF5 is a Class IIa CPuORF, and its conserved region (dark yellow) and variable regions (pale yellow) have been labelled.

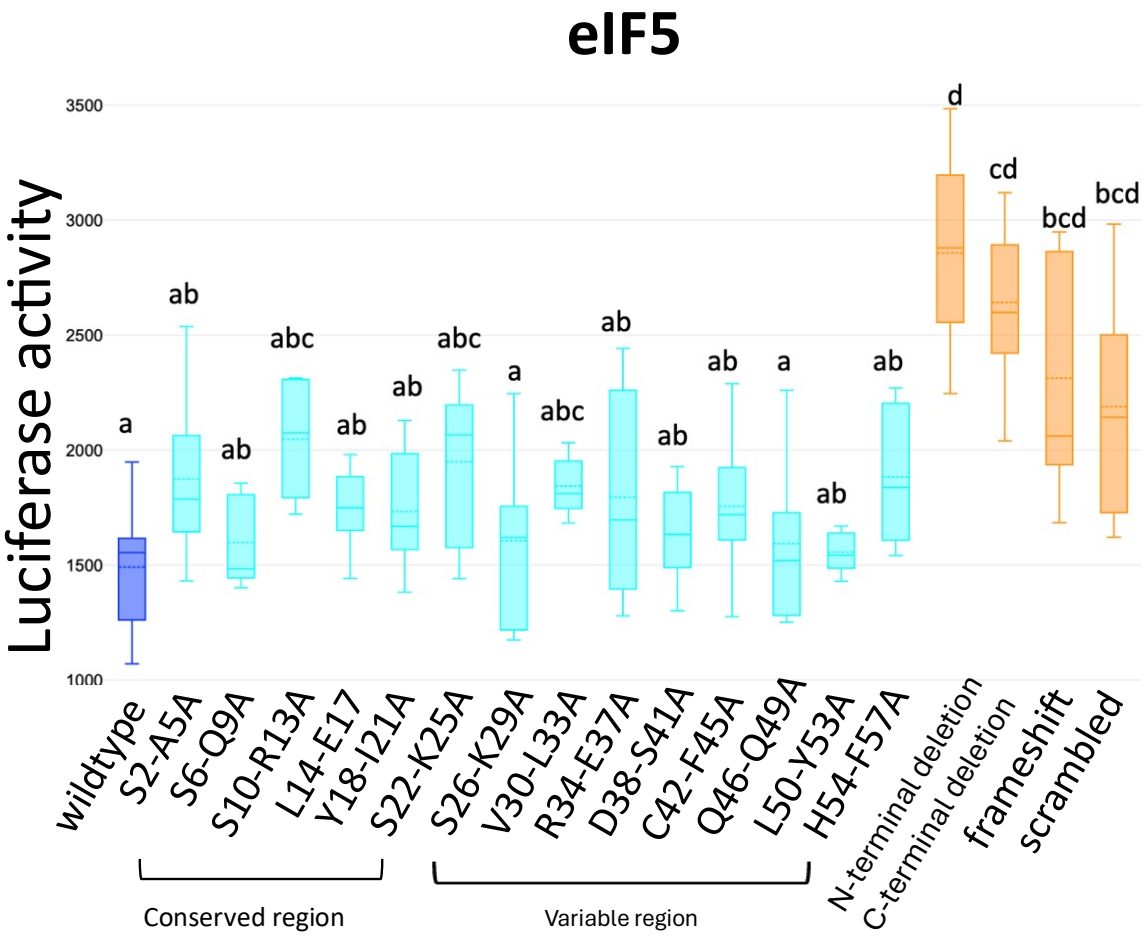


Figure 3.16. A box and whisker plot to show any mutation in eIF5 weakens its ability to stall the ribosome. An *in vitro* luciferase assay was performed on eIF5 and its alanine and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. An ANOVA Tukey test was used to determine significance and cluster datasets. The wild type appears in dark blue, alanine scanning mutants are in light blue, and full peptide mutants are in orange. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 6-9 replicates. The alanine scanning mutations that hit the conserved region have been identified below (S2-K25).

Collectively, *in vitro* investigations into four aCUTS CPuORFs from Class I and IIa (CPuORF 40, ROJ, HEAT and eIF5) have revealed similar findings (Figures 3.9-16). Specifically, whole peptide mutations that either truncate or abolish the conserved

amino acid sequence result in the largest increased in mORF expression when compared to the wild type. Furthermore, the effect of alanine scanning mutations on mORF expression tend to be intermediate. Moreover, statistical analysis does not suggest that there is a clear effect on mORF expression when the conserved region is mutated compared to the variable region. Overall, *in vitro* investigations suggest that the whole CPuORF peptide confers stalling and there is no specific domain that is primarily responsible for stalling domains (Figures 3.9-16).

3.2.7. *In vivo* investigations do suggest CPuORFs have a modular nature.

In vitro analysis into CPuORF mutants indicated that the whole peptide is required for ribosomal stalling (Figures 3.9-16). To validate these findings two aCUTS CPuORFs were selected for analysis *in vivo*. The CPuORFs selected were the Class IIa CPuORF eIF5 and the Class I CPuORF HEAT. Their respective alanine scanning and whole peptide mutants were transformed into wild type *Arabidopsis* by agrobacterium transformation. A luciferase assay was performed on a range of 7-28 T1 leaves.

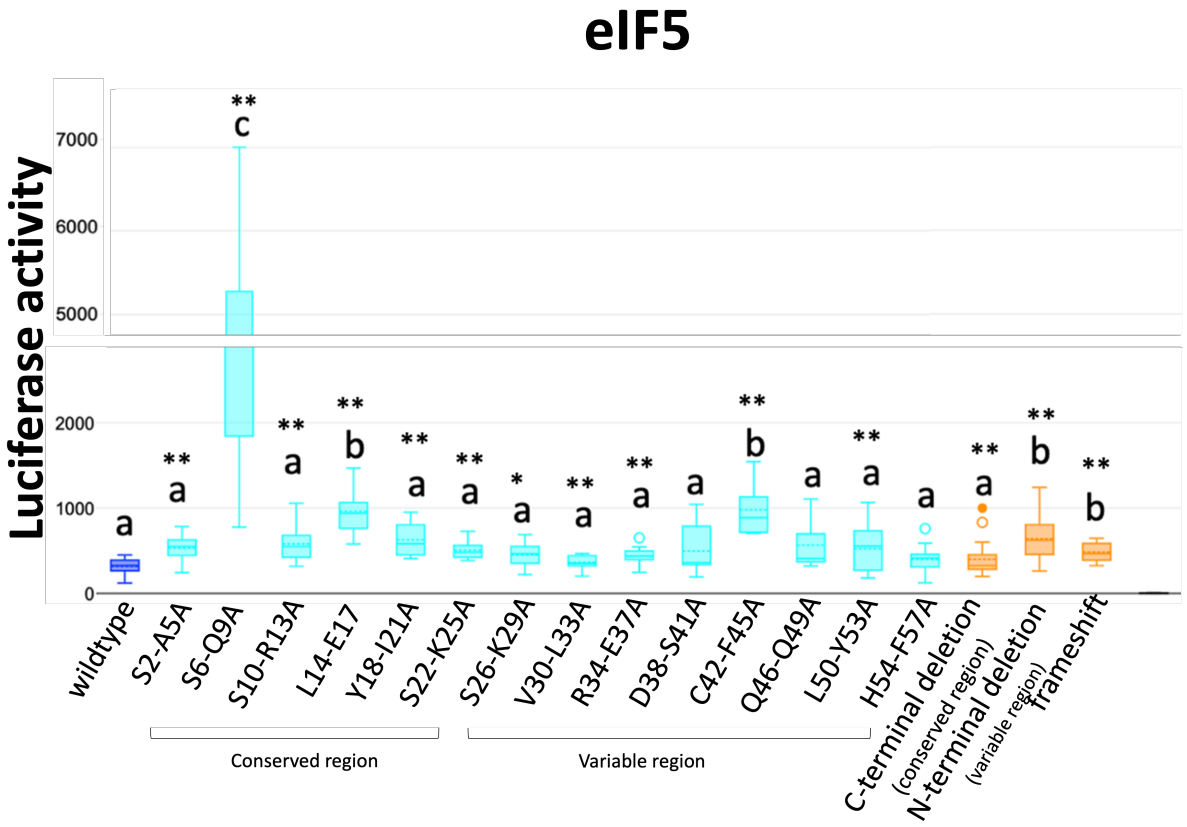


Figure 3.17. A box and whisker plot to show S6-Q9A, L14-E17A and C42-F45A confer ribosomal stalling in eIF5. An *in vivo* luciferase assay was performed on T1 leaves of eIF5 and its alanine and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. An ANOVA Tukey test was used to determine significance and cluster datasets. Asterisks denote statistically significant differences, determined by Student's t-test: **p<0.01, *p<0.05. The wild type appears in dark blue, alanine scanning mutants are in light blue, and full peptide mutants are in orange. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 7-28 replicates. The alanine scanning mutations that hit the conserved region have been identified below (S2-K25).

In contrast to the *in vitro* data, *in planta* data identifies important residues that confer stalling in the eIF5 CPuORF (Figure 3.17). In *vivo,* the S6-Q9A, L14-E17A and C42-F45A mutations elicit a statistically significant increase in mORF translation when compared to the wild type, causing fold increases of 10.1, 2.7 and 2.8 when compared to the wild type, respectively. Notably, the S6-Q9A mutations elicits a particularly large increase in mORF translation that is not observed elsewhere. This may indicate the importance of

these N-terminal residues. When compared to L14-E17A and C42-F45A this may indicate these mutations are not as severe as S6-Q9A. Surprisingly, the C-terminal deletion mutant (contains the conserved region) is statistically similar to the wild type. eIF5 is conserved at its N-terminus, and when investigated *in vitro* (C-terminal deletion) it was unable to elicit stalling. This may suggest the conserved regions of CPuORFs may be able to retain stalling in their native species. This is further supported when the variable region (N-terminal deletion) is compared to the wild type (Figure 3.15). The data for the variable region (N-terminal deletion) suggests that it cannot stall the ribosome as it showed higher levels of luciferase activity when compared to the WT. Overall, *in vivo* data unlike *in vitro* data suggests that the conserved region (C-terminal deletion) of eIF5 can independently stall the ribosome but not the variable region.

Overall, *in vivo* and *in vitro* investigations into eIF5's CPuORF demonstrate that mutations in the CPuORF peptide weaken its ability to stall the ribosome (Figure 3.15-19). Furthermore, when tested *in vivo*, the conserved region (C-terminal deletion) retains its ability to stall. Finally, *in vivo* alanine scanning mutations have identified functional eIF5 residues (S6-Q9A, L14-E17A and C42-F45A). These mutations are found in the conserved (S6-Q9A, L14-E17A) and variable region (C42-F45A).

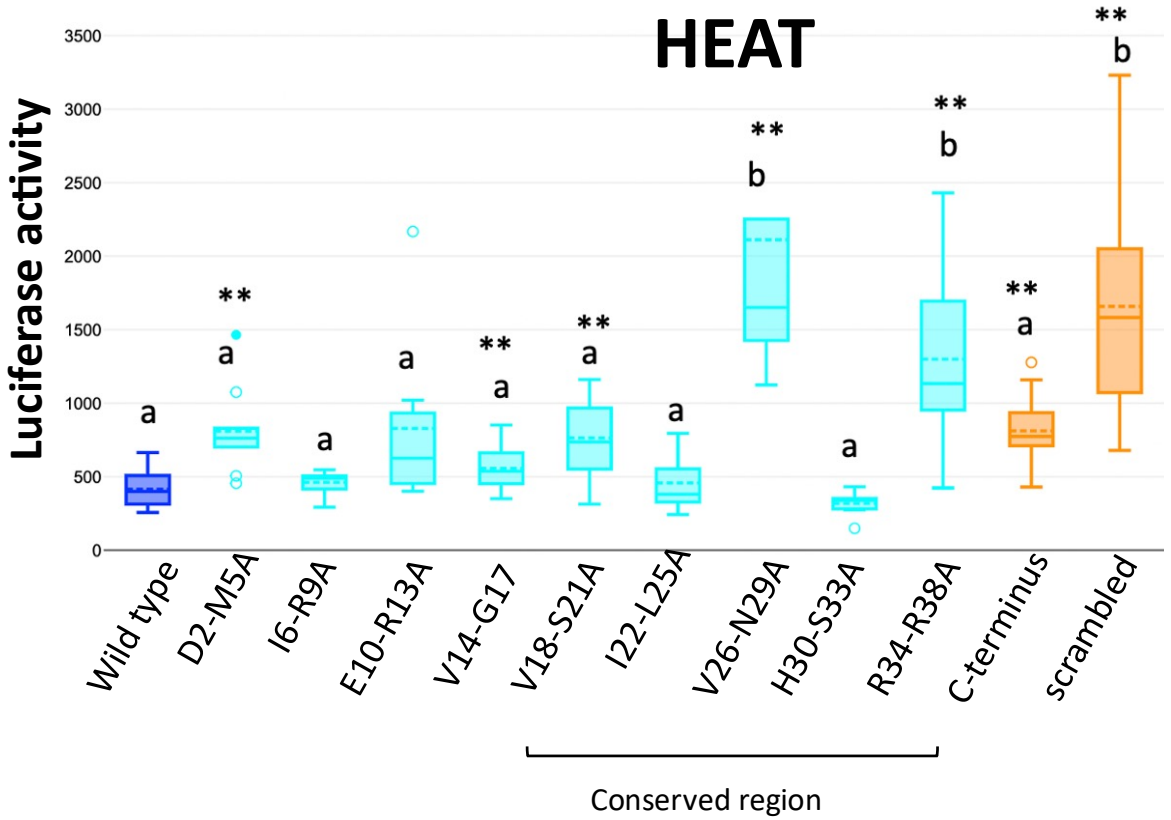


Figure 3.19. A box and whisker plot to show V26-N29A and R34-R38 ribosomal stalling in HEAT. An *in vivo* luciferase assay was performed on T1 leaves of HEAT and its alanine and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. An ANOVA Tukey test was used to determine significance and cluster datasets. Asterisks denote statistically significant differences, determined by Student's t-test: **p<0.01, *p<0.05. The wild type appears in dark blue, alanine scanning mutants are in light blue, and full peptide mutants are in orange. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 7-28 replicates. The alanine scanning mutations that hit the conserved region have been identified below (V18-R38).

Similar to the eIF5 CPuORF, the *in planta* data for the HEAT CPuORF is not in agreement with *in vitro* data (Figure 3.16-19). A Tukey test of *In planta* data suggests that the only alanine scanning mutations to release stalling are the C-terminal mutations of the conserved V26-N29A and R34-R38. These two mutations elicit an average fold increase from the wild type of 5.1 and 3.1 respectively, suggesting that these amino acids are required for ribosomal stalling during translation. This is supported by the fact that they

cluster with the scrambled mutant where stalling has been abolished. The remaining alanine scanning mutations (D2-M5A, I6-R9A, E10-R13A, V14-G17A, V18-S21A, I22-L25A and H30-S33A) elicit fold increases of 2.0, 1.1, 2.0, 1.3, 1.8, 1.1 and 0.7, respectively. Although these alanine scanning mutations (except H30-S33A) cause an increase in mORF translation they are not statistically different from the wild type under a Tukey test. However, D2-M5A, V14-G17A and V18-S21A are statistically different under less conservative statistical tests such as a student's T-test. Overall, when tested *in vivo* V26-N29A and R34-R38 appear to hit crucial residues.

Surprisingly, when tested *in vivo*, the N-terminal deletion mutant appears to retain ribosome stalling (Figure 3.16). This only contains the conserved C-terminus of HEAT and is statistically similar to the wild type but not the frameshift under a Tukey test. The Tukey test may suggest the conserved region can maintain stalling independently.

Collectively, *in planta* data may support the idea of a modular nature for the HEAT CPuORF as its C-terminus may retain stalling and alanine scanning mutations (V26-N29A and R34-R38) in the conserved C-terminus elicit a clear increase in mORF translation unlike the N-terminus. This pattern is also reflected in the Class IIa CPuORF eIF5 when investigated *in vivo*. Ultimately, the differences between *in vivo* and *in vitro* studies may indicate that there are limitations to investigating *Arabidopsis* CPuORF peptides outside of *Arabidopsis* systems.

3.2.8. Most CPuORF mutants lose their ability to sense signals.

In vivo data suggests that the conserved region largely confers ribosome stalling. If this is the case, this could suggest that the variable region could confer signal sensing. To test this, eIF5 and HEAT CPuORFs and their mutants were investigated in wild type *Arabidopsis* (Figures 3.17-18). Wild type and mutant lines were subjected to control conditions and under elevated mannitol and heat shock respectively. A luciferase assay was performed on T1 leaves and each data set has a range of 7-28 replicates. If the variable region confers signal sensing it is expected that these mutations would not result in increased mORF expression, as observed in the wild type.

The eIF5 wild type CPuORF peptide demonstrates an aCUTS response to water limitation or mannitol (Figure 3.19) (3). Upon application of mannitol, eIF5:LUC lines observe a statistically significant fold increase in luciferase activity of 3.1. Most alanine scanning mutants and all whole peptide mutants lose the eIF5 aCUTS response under elevated mannitol. Most mutant lines observe no significant change between control conditions and elevated mannitol. In fact, under elevated mannitol a students T-test suggested that S2-A5A, S6-Q9A, L14-E17, S26-K29A and the conserved region (C-terminal deletion) C-terminus all show a significant decrease in mORF expression. However, when assessed by an ANOVA Tukey test only S6-Q9A, L14-E17, and the C-terminus deletion's effect were statistically significant.

Furthermore, the conserved N-terminus (C-terminal deletion) has lost its signal sensing ability despite evidence to suggest that it may retain stalling (Figure 3.15). None of the

whole peptide mutants display aCUTS activity, suggesting a requirement for a full-length peptide. Despite this, some mutant lines retain aCUTS activity indicating that these alanine scanning mutations have no effect on eIF5's aCUTS activity. These lines are Y18-I21A, V30-L33A and D38-S41A. The lines elicit a fold increase of 3.3, 1.5 and 2.4 respectively under elevated mannitol compared to control conditions. The ANOVA Tukey test suggest that Y18-I21A and D38-S41A were significant. Overall, this evidence cannot conclude if aCUTS CPuORF are organised into a conserved stalling and variable signal sensing module.

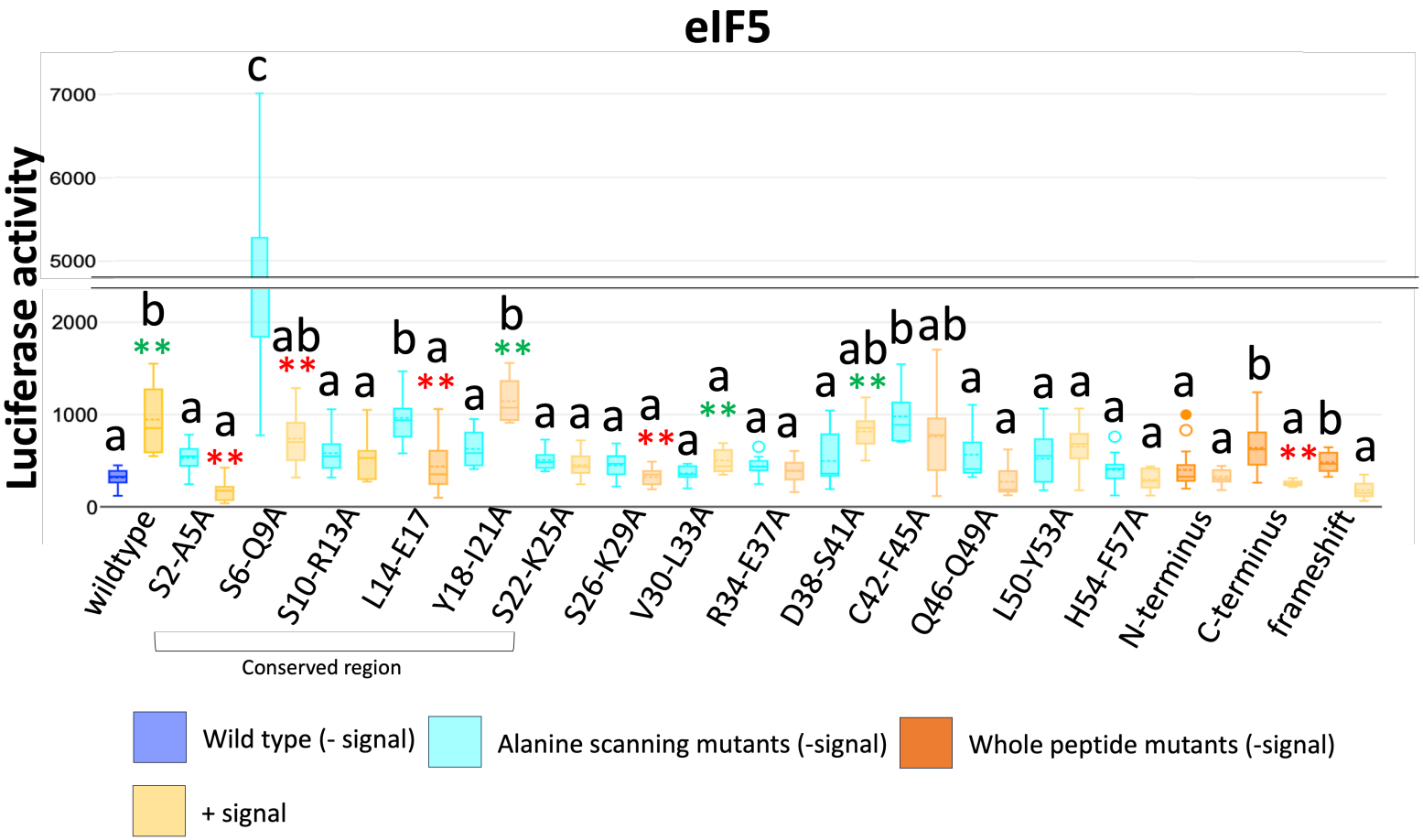

eIF5
Luciferase activity
7000
6000
5000
2000
1000
0
wildtype
S2-A5A
S6-Q9A
S10-R13A
L14-E17
Y18-I21A
S22-K25A
S26-K29A
V30-L33A
R34-E37A
D38-S41A
C42-F45A
Q46-Q49A
L50-Y53A
H54-F57A
N-terminus
C-terminus
frameshift
Conserved region
Wild type (- signal)
Alanine scanning mutants (-signal)
Whole peptide mutants (-signal)
+ signal

Figure 3.19. All mutations except for Y18-I21A, V30-L33A and D38-S41A abolish the eIF5 CPuORF aCUTS signal sensing mechanism. An *in vivo* luciferase assay was performed on T1 leaves of eIF5 and its alanine scanning and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. An ANOVA Tukey test was used to determine significance and cluster datasets. Asterisks denote statistically significant differences, determined by Student's t-test: **p<0.01, *p<0.05. The T-test was performed between -signal and +signal data. Red asterisks indicate a statistically significant decrease in mORF expression in the mutant line under + signal conditions compared to – signal. Green asterisks indicate a statistically significant increase in mORF expression under + signal conditions compared to – signal conditions (Green indicates aCUTS activity). Here – signal refers to 0.5 MS and + signal refers to 0.5 MS + 300 mM Mannitol. Mannitol was used to simulate drought. Signal treatments were incubated overnight. The wild type appears in dark blue, alanine scanning mutants are in light blue, full peptide mutants are in orange and corresponding + signal data is in yellow. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 7-28 replicates. The alanine scanning mutations that hit the conserved region have been identified below (S2-K25).

The HEAT wild type CPuORF peptide demonstrates an aCUTS response to heat shock (Figure 3.20). Upon application of heat shock, HEAT:LUC lines observe a statistically significant fold change of 2.7. Most alanine scanning mutants and all whole peptide mutants lose their ability to modulate mORF translation (aCUTS) in response to heat shock. Most mutant lines observe no significant change between control conditions and heat shock. Post-heat shock, V14-G17A, V18-S21A, V26-N29A, H30-S33A and the N-terminal deletion (conserved region) show a significant decrease in mORF expression when assessed by a student's T-test. However, they clustered in group a along with the untreated counterparts and the wild type. Furthermore, the conserved C-terminus has lost its aCUTS ability despite evidence to suggest it may retain stalling (Figure 3.16). None of the whole peptide mutants display aCUTS activity, indicating the requirement for a full-length peptide. Despite this, a single mutant line retains aCUTS activity, indicating that this alanine scanning mutation has no effect on HEAT's aCUTS activity. Interestingly I22-

L25A is found within the conserved region and, under heat shock , it elicits a fold increase of 2.4 compared to control conditions similar to the wild type. These results indicate that signal sensing is conferred via the whole CPuORF peptide. When taken with the data for eIF5 (Figure 3.19), the data suggests that stalling is confined to the conserved region, but signal sensing is conferred by the entire peptide.

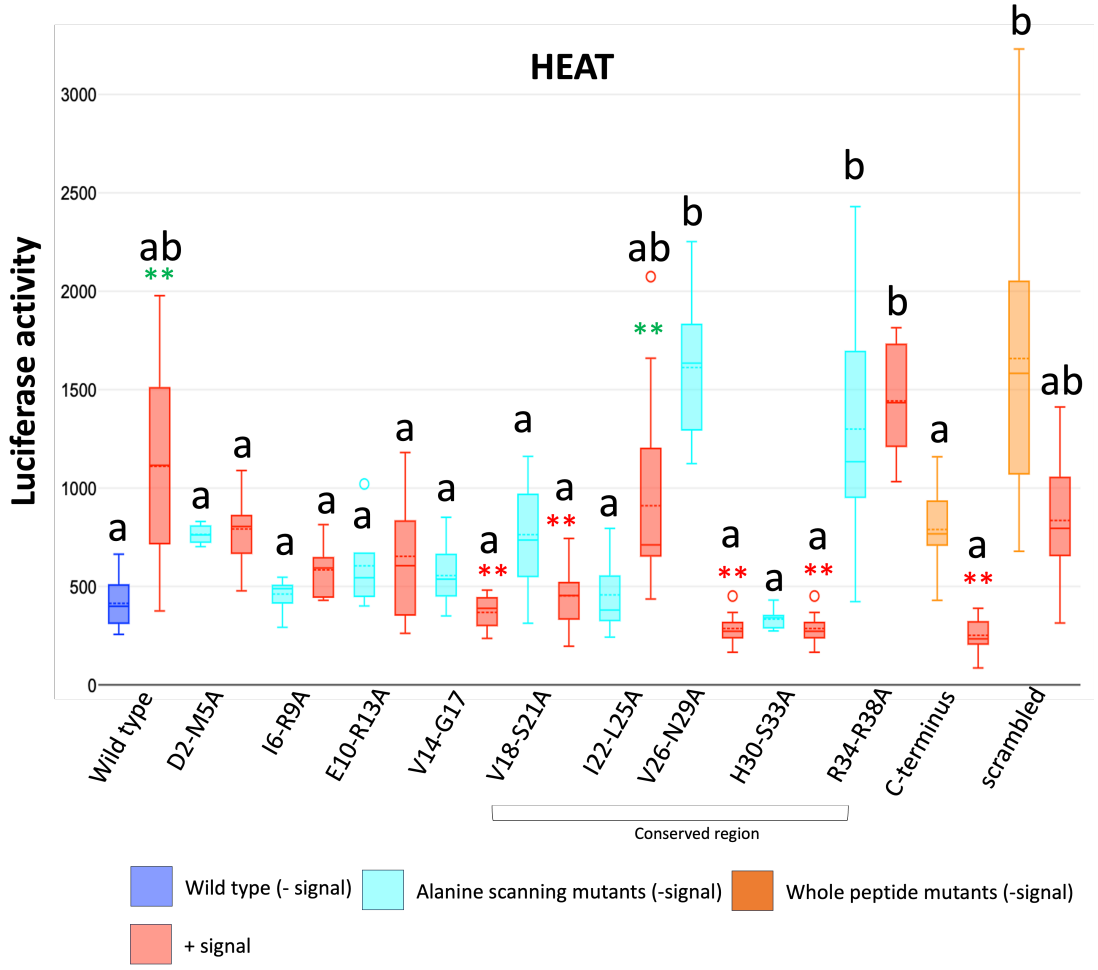


Figure 3.20. All mutations except for I22-L25A abolish the HEAT CPuORF aCUTS signal sensing mechanism. An *in vivo* luciferase assay was performed on T1 leaves of HEAT and its alanine and full-peptide mutants. Luciferase activity (cps/area) for each line has been plotted as a box and whisker plot. . An ANOVA Tukey test was used to determine significance and cluster datasets. Asterisks denote statistically significant differences, determined by Student's t-test: **p<0.01, *p<0.05. The T-test was performed between -signal and +signal data. Red asterisks indicate a statistically significant decrease in mORF expression in the mutant line under + signal conditions compared to – signal. Green asterisks indicate a statistically significant increase in mORF expression under + signal conditions compared to – signal conditions (Green indicates aCUTS activity). Here – signal refers to room temperature and + signal refers to 37 ºC heat shock. Signal treatments were incubated overnight. The wild type appears in dark blue, alanine scanning mutants are in light blue, full peptide mutants are in orange and corresponding + signal data is in red. The median and mean are depicted as a solid and dotted lines respectively. Each data set has 7-28 replicates. The alanine scanning mutations that hit the conserved region have been identified below (V18-R38).

3.3. Discussion

This study aimed to elucidate the CPuORF CUTS mechanism (Figure 1.20) through molecular biology. By comparing wild type CPuORFs to a range of CPuORF mutants, this study demonstrated that CPuORFs act as peptide switches that conditionally modulate the translation of downstream mORFs (Figure 3.4-7). Specifically, the CPuORFs studied here conditionally modulate mORF expression by stalling the ribosome during translation (Figure 3.4-7). By utilizing *Arabidopsis* mutants in ribosomal exit tunnel proteins, this study showed that aCUTS CPuORFs respond to the same specific mutations in ribosome exit tunnel proteins (*uL16z*^{G14S}, *rack1z*^{W261STOP} and *uL4z*^{G75R}), suggesting a common mechanistic framework for ribosome stalling (Figure 3.4-5). Finally, ribosome mutations not only abolished stalling but completely abolished aCUTS activity that includes stalling and signal sensing (Figure 3.8).

Various CPuORF peptide mutations and their effect on mORF expression were investigated through an *in vitro* translation assay (Figure 3.9-16). This study took an alanine scanning approach by substituting four consecutive amino acids with four alanine to identify functional CPuORF residues. Alongside these mutations frameshifted, scrambled and N/C-terminal deletion variants were made. *In vitro* data suggested that whole peptide mutations completely abolish ribosome stalling activity for all four aCUTS CPuORFs. Furthermore, alanine scanning mutations caused an intermediate disruption of ribosome stalling. Moreover, statistical analysis indicated that variable and conserved amino acids generally play an equal role in ribosome stalling.

Conversely, when two aCUTS CPuORFs with completely different conservation patterns (N-terminally conserved Class IIa eIF5 and C-terminally conserved Class I HEAT CPuORFs) are investigated *in vivo*, mutagenesis suggests that ribosome stalling is conferred by specific amino acids in their respective conserved regions (Figures 3.16-20). Moreover, when signal sensing is tested in CPuORF mutants *in vivo* data could not locate specific amino acids that confer this functionality (Figures 3.17-18).

Finally, this thesis investigated the three CPuORFs in the 5' leader of *SAC51*. The organisation of the *SAC51* transcript is rare, considering it is one of only five transcripts in *Arabidopsis* that contains multiple CPuORFs. Luciferase assays under control and elevated thermospermine conditions provided evidence that CPuORFs 38 and 40 are thermospermine responsive (Figure 3.1).

3.3.1. CPuORFs are self-contained regulatory units that function in cis.

Plants, being sessile organisms, have evolved mechanisms to adapt to their changing environments. The CPuORF CUTS mechanism offers a rapid regulatory means to modulate mORF expression to specific treatment (Tables 1.4-6, Figures 1.20, 3.3-8). This thesis has demonstrated that CPuORFs respond to specific signals to modulate mORF expression (Figure 3.5). Moreover, this study has shown that by placing the CPuORF peptide sequence upstream of an artificial mORF in an artificial context, the CPuORF peptide retains its regulatory capacity (Figure 3.4-6). Consequently, the various CPuORF CUTS mechanisms have potential as peptide switches in biotechnology and crop improvement.

3.3.2. CPuORFs 38 and 40 are the functional CPuORFs in the SAC51 transcript.

There has been a handful of studies that have investigated the regulatory mechanism of the thermospermine responsive SAC51 5' leader. A genetic screen and GUS assays have pointed towards CPuORF 40 as being the functional SAC51 CPuORF (8,37,118). To identify the SAC51 CPuORF(s) that confer thermospermine responsive aCUTS activity, this thesis explored the effect of single and double combinations of the three SAC51 CPuORFs on luciferase activity (Figure 3.1). The results indicated that CPuORF 38 and CPuORF 40 can independently retain the reported SAC51 5'leader thermospermine responsive aCUTS mechanism. Interestingly, mORF translation levels in CPuORF38:LUC are much higher than CPuORF40:LUC and the wild type SAC51 transcript with all three CPuORFs. Furthermore, mORF expression levels in CPuORF40:LUC and the wild type

SAC51:LUC transcripts are similar. Indicating that CPuORF40 is more capable of attenuating mORF expression compared to CPuORF38. Moreover, CPuORF40 is as successful in stalling the ribosome as the three CPuORF containing SAC51 transcript. In any case, CPuORF 38 and CPuORF 40 confer the reported SAC51 aCUTS activity.

3.3.3. A common CPuORF stalling mechanism.

Initially, CPuORFs were identified by peptide sequence conservation between *Arabidopsis* and rice. The fact that this rare class of uORFs has been conserved at the amino acid level for over 250 million years indicates that these peptides are under a functional constraint. CPuORF peptides been classified into three categories, based on the positions of their conserved residues (Figure 1.12). CPuORFs have generally been considered to have a conserved and a variable region, and the position of the conserved region has been used to assign each CPuORF to a category (I, IIa or IIb, see Figure 1.12). Thus, one hypothesis was that the conserved region amino acids confer ribosome stalling and the variable region amino acids confer signal sensing (3). In this model, the classification system, CPuORFs could be described as having a conserved stalling domain and a variable signal sensing domain. To test this hypothesis, this thesis placed WT and modified CPuORFs upstream of a luciferase reporter gene in wild type and ribosome mutant backgrounds (Figure 3.6-8). The modifications that this thesis explored included a comprehensive alanine scanning series, covering all the amino acids in four CPuORFs as well as more disruptive mutations (called “whole peptide mutations”), where the CPuORF peptides were more radically altered. Luciferase expression was quantified to determine the effect of WT and mutated CPuORFs on mORF expression and

to correlate that to conserved and variable CPuORF amino acids (Figures 3.9-20). The results indicate that there is a common mechanistic framework through which all tested aCUTS CPuORFs stall the ribosome (Figure 3.6-8). In addition, *in vivo* data suggests that stalling is conferred by the conserved region, but questions remain about the functional organisation of signal sensing which does not seem to be restricted to the variable domain (Figure 3.9-20).

A significant finding is that all the CPuORFs tested in this study, that utilise the aCUTS mechanism are sensitive to the same ribosomal exit tunnel mutations, suggesting a common stalling mechanism (Figure 3.5-8). Four unrelated aCUTs CPuORFs were constitutively expressed *in planta*, as part of the 5' leader of a luciferase-encoding RNA. These CPuORFs were investigated in a Ler wild type background as well is in three different Ler backgrounds containing single base substitution mutations affecting exit tunnel proteins uL16z, uL4z and RACK1z (Figure 3.7) (37,118). These three proteins are located within, or associated with, the exit tunnel (Figure 3.7). Previously published structural evidence has demonstrated that *uL16* and *uL4* stall the ribosome with known eukaryotic ribosome arrest peptides (Table 1.6) (20,111,138,139).

This study investigated CPuORFs 40, eIF5, HEAT and ROJ, all of which have been shown to utilise the aCUTS mechanism (3,8), *in vivo* (Figure 3.4-8). In all three of the ribosome mutant backgrounds, the luciferase mORF showed increased translation for all the tested aCUTS CPuORFs, indicating that ribosome stalling at all these CPuORFs is sensitive to same mutations. The exception to this is CPuORF 40, where no increase in mORF expression was observed in the *rack1z* background. It remains possible that that

a different CPuORF in the *SAC51* 5' leader sequence, such as CPuORF 38, is sensitive to this mutation. As a control, the 35S:LUC plasmid was transformed into the same ribosomal protein mutant backgrounds. In this case, a decrease in *LUC* mORF translation was observed, when compared to the wild type background. This observation suggests that the ribosomal protein mutations generally have a negative influence on translation. However, the extent of their impact may vary depending on the specific ribosomal protein involved and the context of the mRNA being translated, indicating a need for further investigation to explore the variable effects of RP mutants, treatments, and translation. The decrease in mORF translation observed in the 35S:LUC ribosome mutant lines is replicated in the frameshift 40, frameshift eIF5, scrambled HEAT and scrambled ROJ.

Outside of aCUTS CPuORFs, this study also investigated the rCUTS bZIP11 and a/rCUTS TBF1 CPuORFs in these ribosome mutant backgrounds (Figure 3.7). Unexpectedly, both bZIP11 and TBF1 CPuORFs demonstrated a decrease in mORF expression in the ribosome mutant lines when compared to the wild type background. bZIP11 CPuORF requires sucrose to stall the ribosome and this study did not investigate mORF expression under elevated sucrose (110). Therefore, it is likely that under 0.5 MS conditions there is no stalling at the bZIP11 CPuORF and any effect negative of the ribosomal protein mutants on stalling would therefore be undetectable. On the other hand, TBF1 functions via both aCUTS and rCUTS and there was also no evidence for a loss of stalling at this CPuORF in any of the ribosomal protein mutant backgrounds (9,128,129). This may suggest that the aCUTS/rCUTs stalling mechanism of TBF1 fundamentally differs from the mechanism utilised by the exclusively aCUTS CPuORFs.

This thesis located the three ribosomal protein mutations within a cryo-EM map of the translating wheat germ ribosome (Figure 3.6). The uL4^{G75R} mutation is located in the exit tunnel, on the constriction between uL4 and uL22. Mutations in this constriction have been reported to abolish or enhance stalling (Figures 1.13-18, Table 1.6) (20,137–139,148,149). The substitution of a non-polar glycine to an amphipathic arginine could affect the charge within the exit tunnel and between it the nascent CPuORF. This substitution could also severely affect the local structure of the constriction, since glycine has the smallest side chain whereas arginine has a large and complex side chain. Consequently, the substitution could disrupt the local biochemical and/or structural requirements necessary to elicit ribosome stalling during translation.

Cryo-EM data has demonstrated that RACK1 stabilises stalled and collided ribosomes by interacting with adjacent ribosomes and forming disomes (Table 1.6) (124). The RACK1z^{W261STOP} mutation introduces a termination codon, preventing the expression of the complete peptide (Figure 3.5). When known ribosome stalling aCUTS CPuORFs are constitutively expressed in the RACK1z^{W261STOP} background, there are higher levels of mORF translation when compared to wild type lines (Figure 3.5). Suggesting that the RACK1z^{W261STOP} mutation disrupts ribosome stalling. A defective RACK1, produced in the RACK1z^{W261STOP} mutant, could reduce ribosome stalling events and consequently, reduce the attenuation of mORF translation by effected stalling peptides.

The ability of uL16^{G14S} to abolish ribosome stalling is less easy to explain (37). uL16 has been implicated in stalling and the CPuORF rCUTS mechanisms (Table 1.6, Figure 3.5) (20,95,120,121,139). Furthermore, mutants in uL16 indicate it plays a role in delivering

tRNAs into the PTC and acting in the interface between the ribosomal subunits. uL16 is found in proximity to the P-site tRNA (95,120,139). Arrest peptides can stall during translation termination or elongation (20). Furthermore, stalling events occur when there is a specific conformation of tRNAs during termination/elongation within the PTC (20). The substitution of glycine for a polar serine in uL16z may disrupt function. Glycine has a small side chain whereas serine contains a hydroxymethyl group. This could lead to a local change within the ribosome and have effects on the PTC and the binding of translation elongation and release factors (2). Due to the proximity to the PTC, it is possible that this mutation could abolish stalling.

All aCUTS CPuORFs tested responded to the same set of three ribosomal protein mutations, leading to an increase in mORF translation, indicating a reduction in CPuORF-mediated ribosomal stalling (Figure 3.5). It is unknown whether these residues directly interact with the nascent CPuORF peptide, but this seems plausible in two cases. This is the first study to demonstrate that common ribosome mutations affect the function of multiple aCUTS CPuORFs (Figure 3.5). Furthermore, this study showed that these mutations specifically affect the expression of mORFs associated with WT CPuORFs and not frameshifted or scrambled CPuORFs (Figure 3.5). Peptides that are capable of stalling may stall the ribosome in a state that still retains factors that facilitate re-initiation. Future investigations should take a comparative approach to generate high resolution structures of CPuORF stalled ribosomes using cryo-electron microscopy. By comparing CPuORF-stalled ribosome to non-stalled ribosomes, studies can elucidate the stalling mechanism. Cryo-EM will identify interactions between the CPuORFs, ribosome exit tunnel proteins and rRNA species that could be responsible for stalling across different

mechanisms. Furthermore, structural biology can identify translation factors and protein conformations that aid in stalling the ribosome and facilitating re-initiation.

Current structural studies have elucidated the stalling mechanisms on non-CPuORF peptides such as SecM, the fungal arginine attenuator peptide (FAAP), uORF2 of the human cytomegalovirus and the TnaC peptide from *E. coli* (20,111,135–137,139). Data from these studies and cryo-EM data of the rCUTs bZIP11 CPuORF support the idea of an ancient mechanistic framework to stall the ribosome at CPuORFs (110). Similarities between these mechanisms include interactions between exit tunnel proteins (including the constriction), rRNA nucleotides and a signal sensing pocket (20,111,135–139).

The signal sensing pocket is interesting as it has been implicated in two rCUTS mechanisms (*Arabidopsis* bZIP11 CPuORF and the *E.coli* TnaC) that respond to structurally similar signals (sucrose and tryptophan, respectively). This suggests that a peptide-based mechanism to stall ribosomes conditionally, in response to small metabolites, has ancient origins (110). Further cryo-EM analysis of different CPuORFs, including CPuORFs that work by both aCUTS and rCUTs and known arrest peptides could elucidate more signal sensing mechanisms.

3.3.4. *In vivo* investigations suggest that the conserved region confers ribosome stalling.

To test the hypothesis that the conserved region confers stalling and the variable region confers signal sensing, this study investigated a range of CPuORF mutants (Figure 3.9-

20). Initially, *in vitro* investigations into four aCUTS CPuORFs from Class I and Class IIa suggested that any mutation in the CPuORF peptide weakens its ability to attenuate mORF expression (Figure 3.9-20). This implied that the entire CPuORF peptide, regardless of conservation, conferred stalling. Furthermore, statistical analysis did not reveal an enhanced role in stalling for either the conserved or variable regions. In contrast, *in vivo* investigations suggested that stalling was conferred by the conserved regions in a Class I and Class IIa CPuORF (Figure 3.9-20). As a consequence, the *in vitro* investigations into CPuORF function are not fully consistent with *in vivo* data using the same constructs. The CPuORF peptide sequences investigated here are from *Arabidopsis* and they were investigated*, in vitro*, using a wheat germ cell-free translation system (139). The differences observed between the wheat germ system and stably transformed *Arabidopsis* may be attributed to differences between the wheat ribosome and the *Arabidopsis* ribosome. Alternatively, the differences between the *in vivo* and *in vitro* data could reflect missing factors in the purified in vitro translation system.

Whole peptide mutations including frameshift, scrambled and C/N-terminal mutations indicate that the full-length peptide is important to maintain stalling and aCUTS activity (Figures 3.17-20). For example, the conserved regions of HEAT and eIF5 (N-terminal and C-terminal deletion respectively) independently maintain mORF expression like the full-length wild type CPuORF *in vivo* (Figures 3.17-20). However, upon application of treatment the N-terminal and C-terminal deletions of HEAT and eIF5 respectively, do not show an increase in luciferase activity (aCUTS activity) as seen in the wild type CPuORF (Figures 3.17-20). Therefore, indicting the necessity for the full-length peptide and variable amino acids for aCUTS activity.

3.3.4. Concluding remarks.

This study has shed light on the stalling mechanisms of aCUTS CPuORFs. It has emphasised their role as self-contained regulatory units and suggested that there could be a common stalling mechanism for this class of CPuORF. This study has also shown that CPuORF conserved residues likely confer ribosome stalling in their native species. Furthermore, this study has demonstrated that the variable region does not solely confer signal sensing. Future research should utilise genetic screens and structural biology for a detailed insight into the molecular mechanisms of CPuORF function.

Chapter 4: A structural analysis of the CPuORF mechanism

4.1. Introduction

Chapter 3 of this thesis investigated the CPuORF CUTS mechanism through molecular biology (Figure 1.20). Luciferase assays provided evidence that there is a common framework that aCUTS CPuORFs stall the ribosome (Figures 3.4-5). Chapter 3 demonstrated, when aCUTS CPuORFs are constitutively expressed *in vivo* in *Arabidopsis* ribosome mutant backgrounds, there are higher levels of mORF expression when compared to CPuORFs expressed in wild type backgrounds. Furthermore, *in vivo* luciferase assays of mutant CPuORFs suggest that the CPuORF conserved region confers ribosome stalling (Figures 3.15-18).

Although molecular studies within chapter 3 have advanced our understanding of CPuORF function, they have not elucidated the full molecular mechanism. Structural biology and specifically cryo-electron microscopy (cryo-EM) has proved to be a powerful technology that has provided insight into the mechanistic underpinnings of stalled ribosomes (20,110). Consequently, this chapter will discuss how structural biology can be utilised to investigate ribosome structure and function, ribosome stalling mechanisms and how to produce plant ribosome and CPuORF stalled ribosomes for cryo-EM.

4.1.1. Structural insights into the ribosome.

Structural biology began providing insights into ribosome organization in the 1970s (172). Early studies using X-ray crystallography provided unprecedented insights into the bacterial ribosome, including the catalytic centre (the PTC). Subsequent studies have investigated eukaryotic ribosomes by utilising similar techniques and new technologies such as cryo-electron microscopy. Over the past few decades structural biology has provided high resolution maps and structures of bacterial, yeast, plant, and human ribosomes (173–176). Recent maps have achieved global resolutions of 1.9, 2.2, 2.05 and 1.55 Å of human, plants, yeast, and bacterial ribosomes (173–176). Comparative analyses and accurate mapping for ribosomal proteins and rRNA species has provided insights into ribosome function (20,172–176). Significant findings about ribosome structure include that tRNA-binding sites are primarily composed of rRNA, while ribosomal proteins play significant roles in decoding and stabilizing tRNAs (33).

Structural biology has also provided insights into the peptide exit tunnel during translation, the vestibule through which the nascent peptide exits the ribosome during translation (20,142,177). Studies have shown the eukaryotic peptide exit tunnel is approximately 80 Å in length and 10-20 Å wide (178). The tunnel is hydrophilic and possess an overall electronegative potential (148). The tunnel is mainly composed of rRNA (23S for bacteria and 28S for archaea and eukaryotes) but it is also lined by ribosomal proteins like uL4 and uL22 (148). A comparative study of 20 cryo-EM maps and X-ray structures across bacteria, archaea, and eukaryotes revealed distinct differences in their ribosomal tunnels (179). Bacterial tunnels are larger than those in eukaryotes,

with archaea having an intermediate-sized tunnel (179). The structure of uL4 varies between eukaryotes and bacteria. In eukaryotes and bacteria, uL4 and uL22 create a constriction within the exit tunnel. Furthermore, in eukaryotes an extension of the uL4 creates a second constriction site within the exit tunnel (179). The study also highlights that the part of the tunnel closest to the PTC is more conserved and possesses a positive charge compared to the exit tunnel (179).

Plant-Specific Ribosome Features and Their Implications

Recent advancements in structural biology have shed light on the unique characteristics of plant ribosomes (173–176). An analysis of an atomic structure of a wheat ribosome has facilitated a comparative study with yeast and human ribosomes, uncovering distinctive features in plant ribosomes (173–176). Notably, plant ribosomes exhibit specific rRNA modifications within the exit tunnel. Moreover, there is divergence between rRNA modifications between monocots and dicots suggesting the exit tunnel is a site for taxa specific adaptions (92,97,174). Furthermore, the uL4 protein that forms the constriction in the exit tunnel, is longer in plants compared to other eukaryotes (174).

What could be the advantage for plant specific ribosome adaptions? The hypothesis that the unique features of plant ribosomes are a result of their sessile nature (180). Sessile organisms have a selective pressure for rapid gene expression and regulation in response to environmental changes. This is evidenced by multiple isoforms for eukaryotic initiation factors, multiple paralogues of ribosomal proteins, extensions in certain ribosomal proteins (such as uL4, uL42, uL6, eL19, uS2, uS5, and eS10), and a high density of

chemical modifications on plant rRNA (92,93,174,180). Understanding these plant-specific ribosome features is crucial for our study, as they may underpin the mechanisms by which CPuORFs influence ribosome stalling in plants. This knowledge not only contributes to our comprehension of plant biology but also enhances our understanding of the nuanced mechanisms of gene regulation in these organisms.

4.1.2. Structural biology and stalled ribosomes

Cryo-Electron Microscopy (Cryo-EM) has been a pivotal tool in achieving high-resolution visualization of complex interactions within stalled ribosomes (20). The electron microscope setup includes an electron source, a series of these electromagnetic lenses, and an image detecting system (181). Stalled ribosomes are purified (methods of ribosome purification are discussed in the subsequent section) and prepared on a carbon grid (181–183). To preserve their near-native state, the ribosomes are rapidly vitrified by plunging the loaded grid into liquid ethane (181). A key component of Cryo-EM is the use of electromagnetic lenses, which function similarly to optical lenses in a light microscope (173,181–183). These lenses are essential for deflecting the path of negatively charged electron particles using a magnetic field (181). Through Cryo-EM and image processing software, detailed 3D reconstructions of ribosomes are generated from various 2D projections (181).

Cryo-EM studies have provided critical insights into how peptides stall ribosomes (20). Common aspects of stalling include disruption of the PTC and interactions between the exit tunnel and nascent peptide (20,135,138,141). Specifically, interactions between the

constriction formed by proteins uL4 and uL22 and rRNA nucleotides have been associated with ribosomal stalling during translation (20). Stalling can be induced between the nascent peptide and exit tunnel independently but in other cases in combination with co-factors such as metabolites (110).

Specific stalling mechanisms have been elucidated through structural biology. For example, the bacterial SecM and TnaC peptides utilise consecutive positively charged amino acids and polyproline motifs (111,127,135,141,147,184). Proline motifs hinder translation and induce stalling as proline is a poor substrate for peptide bond formation due to its non-optimal conformation to the PTC (111,127,135,141,147,184). Polyproline and positively charge amino acids can hinder A-site tRNA binding and induce tRNA dissociation from the P-site (111,127,135,141,147,184). Leading to ribosome stalling during translation. Cryo-EM has demonstrated that arrest peptides can cause conformational changes within the exit tunnel to induce stalling (111,138,139). For example, the fungal arginine attenuator peptide (FAAP) induces ribosome stalling in the presence of high concentrations of arginine (139). Stalling by the FAAP nascent peptide is achieved via direct interaction of the nascent peptide and constriction in the exit tunnel causing conformational changes in the PTC and preventing its activity (139). The human cytomegalovirus uORF2, nascent peptide forms an alpha helix in the exit tunnel, between the PTC and tunnel constriction (139). This helix prevents nascent chains from being translocated and released by disrupting PTC function and stalling the ribosome (139).

Alongside structural biology, mutagenesis can aid in elucidating stalling mechanisms. Mutagenesis of the arrest peptide TnaC has identified critical residues required for

stalling (136,181). Suggesting that these residues affect positioning of the TnaC nascent peptide in the exit tunnel, preventing stalling interactions between the nascent peptide and exit tunnel. A 3.8 Å map identified that the stalled ribosomes possess tRNA at the P-site and the TnaC peptide directly contacts the ribosome exit tunnel wall (20,135,136). Crucially, the ribosome exit tunnel has two tryptophan binding pockets (110,135,136). The binding of tryptophan stabilises TnaC in the exit tunnel, resulting in conformational changes in the PTC to induce stalling (20,135,136).

In the context of CPuORFs, cryo-EM has provided insights into the mechanism of ribosome stalling, particularly in relation to the bZIP11 transcript in plants (110). This research focused on how an rCUTS CPuORF from the bZIP11 transcript senses sucrose levels and subsequently stalls the ribosome (11,110,126,127). Using a wheat germ *in vitro* translation system, the study demonstrated that elevated sucrose levels lead to ribosome stalling at the stop codon of the bZIP11 CPuORF, akin to the mechanism seen in bacterial TnaC (110,135,136).

Cryo-EM analysis of the bZIP11 CPuORF stalled ribosome achieved a global resolution of 3.7 Å (110). The bZIP11 CPuORF stalls in response to elevated sucrose levels but previously it was unknown how the CPuORF/ribosome senses this signal. Remarkably, the results revealed a distinctive density in the ribosomal exit tunnel, corresponding to the binding site for sucrose (110). This binding is crucial for the stalling mechanism and occurs at a specific pocket within the ribosome that also aligns with the binding sites of other stalling-inducing molecules like tryptophan in bacterial TnaC mechanism and the drug-like molecule PF846 in humans (110,135,136). This suggests a conserved

mechanism across different organisms where specific molecules interact with the nascent peptide and the ribosome to modulate protein synthesis. Furthermore, the study identified a conserved ribosomal RNA nucleotide and His^{134} of uL22 is potentially crucial for the interaction with sucrose (110). The proximity of the sucrose molecule to this nucleotide and uL22 residue indicates a potential hydrogen bond formation, playing a key role in the stalling process. Additionally, the location of this density, approximately 24 Å from the PTC, corresponds to the conserved C-terminus of the nascent peptide previously shown to be vital for stalling (110). Lastly, the study suggested that the stalling mechanism affects the PTC and release factor function in a similar manner to the TnaC peptide (110). Considering the evolutionary distance between the *Arabidopsis* bZIP11 CPuORF and bacterial TnaC arrest peptide may utilise a similar framework to stall the ribosome, this suggests that aspects of stalling may have evolved before plants and bacteria diverged (110). A schematic to represent the bZIP11 CPuORF mechanism as detailed in van der horst et al can be found in figure 4.1.

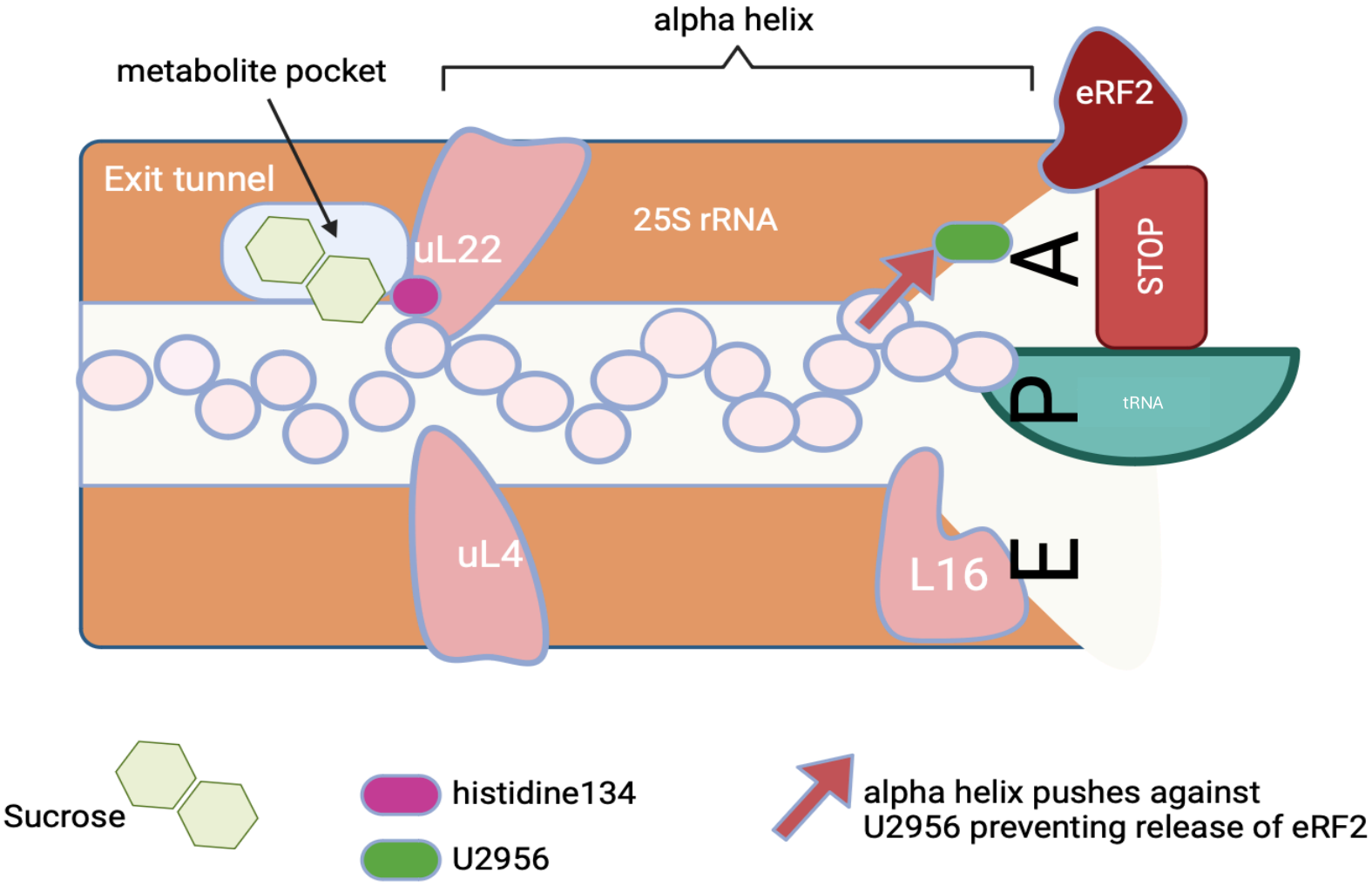

alpha helix
metabolite pocket
eRF2
Exit tunnel
25S rRNA
uL22
STOP
A
P
tRNA
uL4
L16
E
Sucrose
histidine134
U2956
alpha helix pushes against
U2956 preventing release of eRF2

Figure 4.1. A schematic to show the sucrose mediated stalling mechanism of the rCUTS bZIP11 CPuORF. The schematic shows a cross section of the exit tunnel with the nascent CPuORF peptide forming. To the right of the diagram is the PTC (peptidyl transferase centre) with the E, P and A tRNA sites labelled. The P site has a tRNA whereas the E and A sites are free of tRNA. The A site contains the STOP codon of the bZIP11 CPuORF mRNA. The nascent chain forms an alpha helix immediately after the PTC. The alpha helix pushes against the 25S rRNA against a conserved uracil residue. It is hypothesised that pushing against this uracil prevents the release of the eukaryotic release factor 2 (eRF2) thus inducing stalling. Sucrose binds to a conserved metabolite pocket that has been shown to bind to tryptophan in the bacterial TnaC stalling mechanism. The binding of sucrose likely interacts with a conserved histidine (His134) of the uL22 protein. uL22 forms part of the constriction within the exit tunnel with uL4. These interactions likely disrupt the kinetics and translocation of the nascent peptide in the exit tunnel therefore, inducing stalling.

4.1.3. Producing samples of plant stalled ribosomes.

This thesis is interested in *Arabidopsis* CPuORFs that stall the ribosome during translation. Structural biology and cryo-EM can produce high resolution maps and structures of plant stalled ribosomes (20). Recently, the Nicotiana tabacum 80S ribosome was solved by cryo-EM to a global resolution of 2.2 Å (174). This structure included two tRNAs, the decoded mRNA and the nascent peptide chain (174). The data provided insights into the molecular underpinnings of the cytosolic translation process in plants (174). Insights gained from this study include the role of metal ions, specifically potassium, in stabilizing the decoding center and facilitating correct mRNA and tRNA positioning. Furthermore, the study identified plant specific chemical modifications that function in decoding during translation.

Currently, alongside the aCUTS stalling mechanism the plant model organism *Arabidopsis thaliana* ribosome is yet to be solved. Therefore, producing samples for cryo-

EM of the *Arabidopsis* ribosome and CPuORF stalled ribosomes could provide useful insights into ribosome stalling and ribosome evolution (174).

The study that produced the 2.2 Å tobacco ribosome purified ribosomes by harvesting aerial tissue and homogenizing it, post liquid nitrogen freezing (174). The ribosomes were extracted from the tissue using an extraction buffer and then underwent centrifugation (174). The resulting purified ribosomes were suspended in a buffer appropriate for cryo-EM, loaded onto a grid, and then plunge-frozen in liquid ethane to preserve the ribosome particles in their native state (174). Data from the grids was captured using electron microscopy, with subsequent processing of particle movies through dedicated software: RELION for cryo-EM data processing and 3D reconstruction, cryoSPARC for its efficiency with large datasets, and SPIDER for image processing (165,174). Models were built with software called Coot and UCSF Chimera and using sequence data from NCBI. These methods can be replicated to produce *Arabidopsis thaliana* ribosome sample for cryo-EM investigations (Figure 4.2).

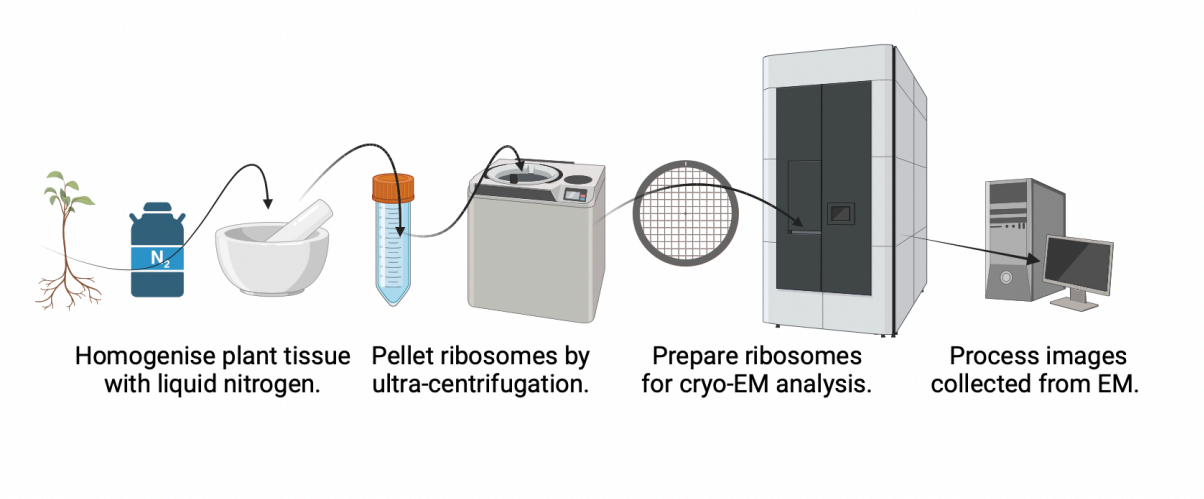


Figure 4.2. A schematic showing the general protocol to generate high resolution structures/maps of ribosomes from plant tissue.

Van der Horst et al., 2023 produced a high-resolution structure of an rCUTS CPuORF stalled ribosome (110). This paper purified ribosomes from an *in vitro* wheat germ extract rather than harvesting plant tissue (110). This study set up a wheat germ reaction with the addition of elevated sucrose and translated capped mRNA of the bZIP11 CPuORF sequence (110). To purify the ribosomes, the group fused an N-terminal His tag to the CPuORF sequence (110). Post-translation the ribosomes were pulled out with anti-His beads (110). Despite reaching a global resolution of 3.7 Å the resulting map could not resolve all the interactions between the CPuORF nascent peptide and exit tunnel, although they did manage to provide evidence for the signal sensing and stalling mechanism (110). Consequently, similar methods using a commercially available wheat germ extract kit and an N-terminal tag will be used in an attempt to produce the first aCUTS CPuORF stalled ribosomes (Figure 4.3).

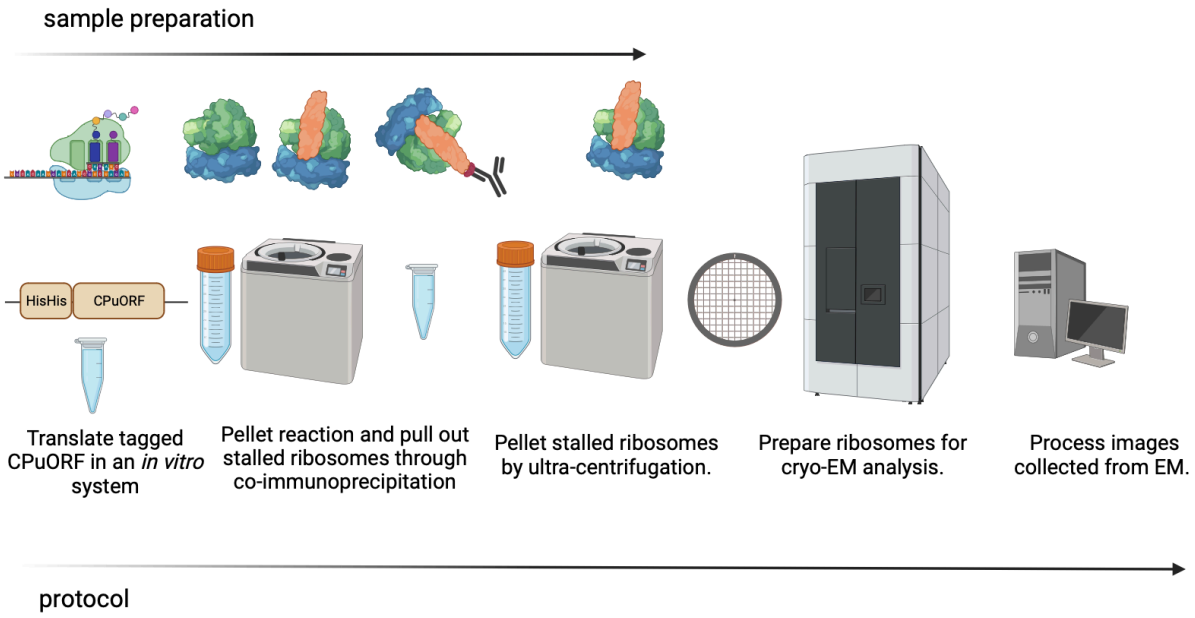


Figure 4.3. A schematic to show the general protocol to generate high resolution structures/maps of stalled ribosomes from *in vitro* systems. Above the protocol is a schematic representation of the types of ribosomes within the sample during the protocol. At the start of the protocol, the ribosomes translate the tagged CPuORF. The first ultracentrifugation step pellets the stalled ribosomes and non-stalled ribosomes. Co-immunoprecipitation is employed to pull out the stalled ribosomes. Stalled ribosomes are then pelleted again through ultracentrifugation leaving a sample of just stalled ribosomes.

4.1.4. Aims

Since there is currently no published model of the *Arabidopsis* ribosome, this study initially aimed to generate the first structure/model of the *Arabidopsis* ribosome. Following this, high resolution structures of *Arabidopsis* CPuORF-stalled ribosomes will enable us to identify the interactions between the CPuORF nascent peptide and the ribosome (exit tunnel proteins, rRNA and/or translation factors). Of note, currently no aCUTS CPuORF stalled ribosomes have been published. Alternatively, I considered employing ribosomes from a commercially available wheat germ extract to understand the mechanism of action of aCUTS, as this sample was successfully employed to unravel

the rCUTS mechanism for bZIP11 CPuORF, fungal arginine attenuator peptide and uORF4 human cytomegalovirus peptide (110,111,139).

Objectives:

1. ***Arabidopsis* Ribosome Purification:** Purify ribosomes via ultracentrifugation through sucrose cushions or co-immunoprecipitation via a tagged ribosomal protein (uL18:FLAG).
2. **Generate a high-resolution structure of the *Arabidopsis* ribosome:** Use cryo-EM single particle analysis to build an atomic structure of the *Arabidopsis* monosome.
3. **Purify a CPuORF stalled ribosome from *Arabidopsis* or wheat germ extract:** Purify CPuORF stalled ribosomes from an N-terminal His tag as demonstrated by van der Horst et al., (110)
4. **Generate and analyse a high-resolution model of a CPuORF stalled *Arabidopsis* or wheat germ ribosome:** Using cryo-EM single particle analysis to build a 3D map and atomic model of an aCUTS stalled ribosome.

4.2. Results

4.2.1. Sucrose gradient purification and co-immunoprecipitation resulted in the purification of *Arabidopsis* ribosomes for EM analysis.

There currently is no published map/structure of the model organism *Arabidopsis thaliana* ribosome. Initial experiments were focused on purifying *Arabidopsis* ribosomes from plant tissue. The published literature suggested two approaches to purify ribosomes from plant tissue: 1) purification by centrifugation and a sucrose gradient and 2) co-immunoprecipitation by a tagged ribosomal protein (uL18). Both methods were adapted from Reynoso et al., which successfully purified *Arabidopsis* ribosomes for ribosome profiling (33). *Arabidopsis* seedlings were grown, flash-frozen with liquid nitrogen, and then homogenized and extracted as outlined in methods 2.7. Seedling tissue was chosen as per Reynoso et al, suggesting it yielded the more ribosomes than mature tissue. Ribosome purification using a 1 M sucrose cushion and ultracentrifugation. Following this, the ribosome pellet formed through the sucrose cushion is resuspended and then ultra-centrifuged through a 60-15% sucrose gradient. Sucrose gradient purification facilitates the isolation of the 80S monosome from free subunits and polysomes. In contrast, the co-immunoprecipitation method employs *Arabidopsis* lines known as TRAP (Translational Ribosome Affinity Purification), which have a FLAG tag on uL18 (33). TRAP lines were generated through *Agrobacterium* mediated transformation.

Initially, a commercial *in vitro* wheat germ extract, translating a control LUC RNA, was ultracentrifuged through the 60-15% sucrose gradients to determine where the 80S

monosome peak falls under the ultra-centrifuge conditions (Figure 4.4). The wheat germ 80S ribosome post ultra-centrifugation is separated at 26 mm, as determined by absorbance at 254 nm, corresponding to approximately 30 % sucrose, as described in the literature (163).

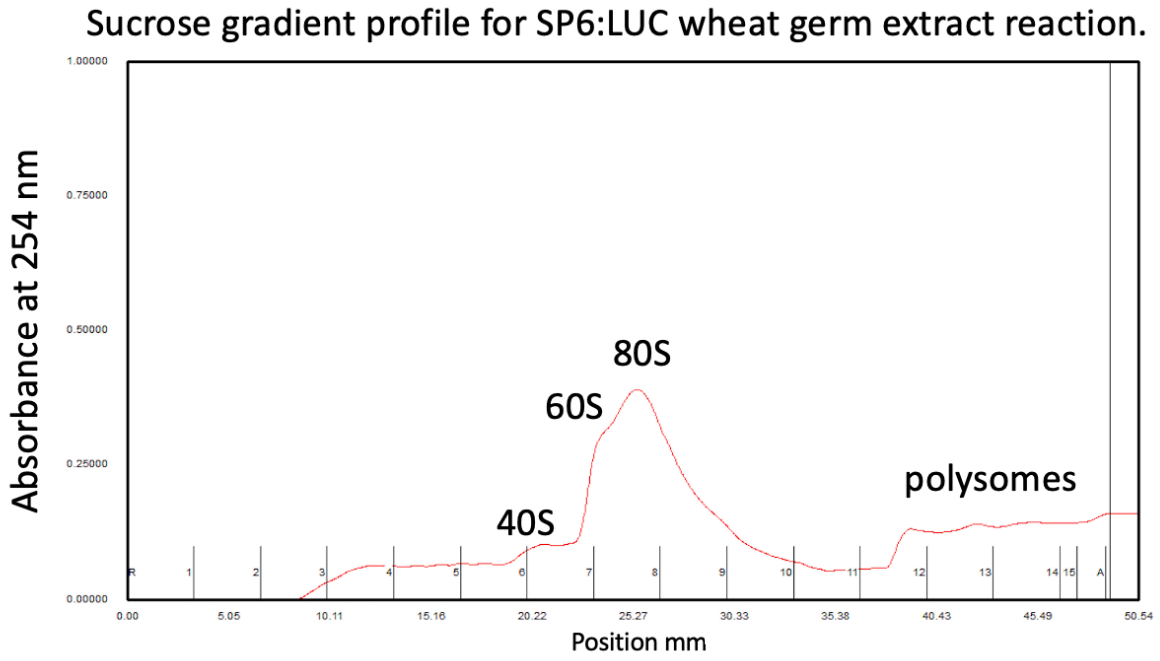


Figure 4.4. A sucrose gradient profile of the wheat germ extract *in vitro* reaction. To determine where a plant 80S monosome would separate in our sucrose gradients at an absorbance at 254 nm, an *in vitro* translation reaction was ultra-centrifuged with our spin conditions. The wheat germ extract translated SP6:LUC mRNA as a control. The X-axis is the distance the proteins separate from the top of the tube. The 40S, 60S, 80s and polysomes are labelled. The 80S peak falls approximately at 26 mm as determined by absorbance at 254 nm.

Following the initial investigation into the wheat germ ribosome, *Arabidopsis* ribosomes were purified by co-immunoprecipitation using a FLAG tag fused to uL18 (163). Ribosomes were purified from harvested seedling tissue that was homogenised by liquid nitrogen and a pestle and mortar. The harvested tissue was then suspended in a ribosome extraction buffer. Using anti-FLAG magnetic beads, ribosomes were pelleted,

and released from the beads using a competing FLAG peptide. The beads were removed via a magnet and the purified ribosomes were ultra-centrifuged through a sucrose gradient to isolate the 80S monosome (Figure 4.5).

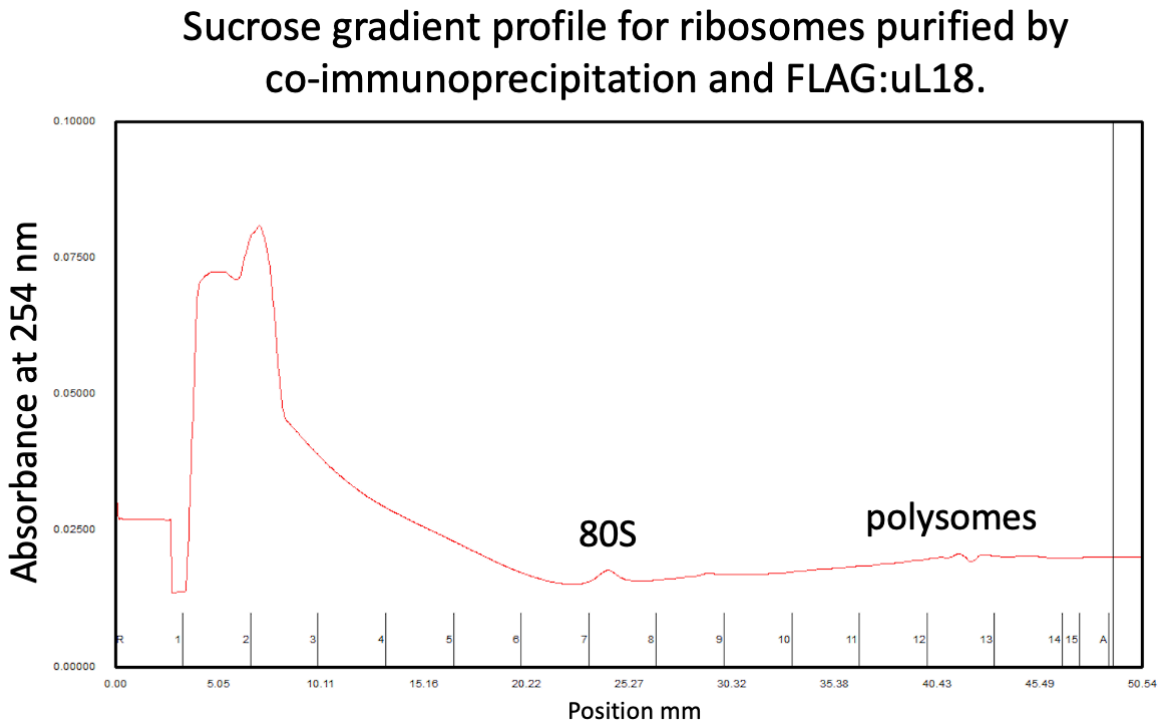


Figure 4.5. A sucrose gradient profile of the *Arabidopsis* ribosomes purified by co-immunoprecipitation and ultra-centrifugation through a sucrose gradient. *Arabidopsis* 35S:FLAG:uL18 seedling tissue was homogenised with liquid nitrogen and resuspended in a ribosome extraction buffer. Ribosomes were pulled out of the extract using anti:FLAG magnetic beads and the FLAG tagged uL18. The ribosomes were eluted off the beads using a competing FLAG peptide. The eluted 80S ribosomes were then purified through a sucrose gradient and ultra-centrifugation to separate the 80S ribosome from sub-units and polysomes. The peaks corresponding to the 80S and polysomes have been labelled on the gradient profile. The 80S peak falls approximately at 25 mm as determined by absorbance at 254 nm. The fractions labelled 6-9 were selected for clean-up and cryo-EM analysis.

Post ultracentrifugation a small peak is observed at 25 mm as determined by an absorbance of 254 nm (Figure 4.5). The absorbance of the peak has a value of approximately 0.02 as opposed to the absorbance of 0.35 from the 80S wheat germ ribosome (Figures 4.4-5), indicating that purifying ribosomes by co-immunoprecipitation

has a low yield and may not be suitable from cryo-EM analysis. The relatively small size of the peak at 25 mm suggests a low success rate in purifying the 80S monosome. Interestingly, free subunits are not observed, as they were in the wheat germ ribosome sucrose gradient. This is due to immunoprecipitation of uL18, a large subunit protein, not purifying free small subunits, although large subunits and subuits that have dissociated from 80S ribosomes should still be present. Overall, immunoprecipitation could only produce small amounts of 80S ribosomes for cryo-EM analysis.

The 80S fraction was collected and concentrated down by centrifugation. A buffer exchange was performed to remove the sucrose for a buffer appropriate for EM. The resulting 80S monosome fraction was negative stained for EM imaging using the T-12 microscope to assess sample quality (Figure 4.6). The EM images show particles that are consistent with published EM images of ribosomes. The particles are approximately 25 μm, indicating that ribosomes were successfully purified using immunoprecipitation. There are some small particles in the sample that are not indicative of an 80S ribosome and are likely free subunits that have dissociated from an 80S ribosome. The yield is significantly lower than published studies, which may indicate that this is a poor sample for cryo-EM.

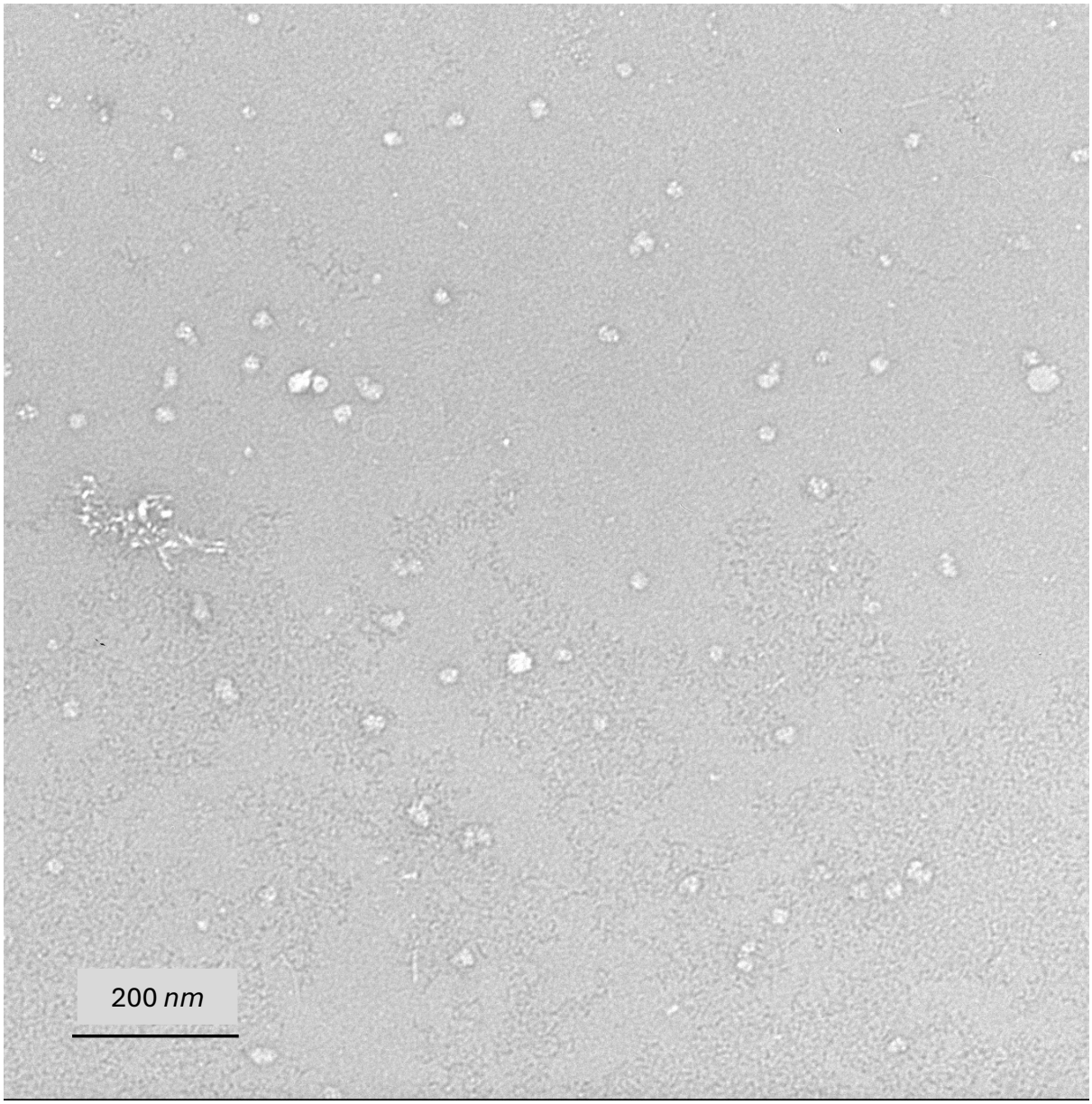


Figure 4.6. Negative stain EM images of *Arabidopsis* ribosomes purified by co-immunoprecipitation and ultra-centrifugation through a sucrose gradient. Purified *Arabidopsis* ribosomes were concentrated down by centrifugation. A buffer exchange was performed to resuspend ribosomes in a buffer appropriate from cryo-EM. 5 uL of sample was negatively stained and examined using the T-12 electron microscope at 120 kV.

Purification through a sucrose cushion and a sucrose gradient may overcome issues in yield as ultracentrifugation relies on the mass of particles to purify rather than protein-antibody binding affinity. *Arabidopsis* wild type seedling tissue was harvested and

homogenised through liquid nitrogen. The frozen tissue was suspended in a ribosome extraction buffer and then ultra-centrifuged through a 1 M sucrose cushion. This resulted in a ribosome pellet that was resuspended and then ultracentrifuged through a sucrose gradient (Figure 4.7).

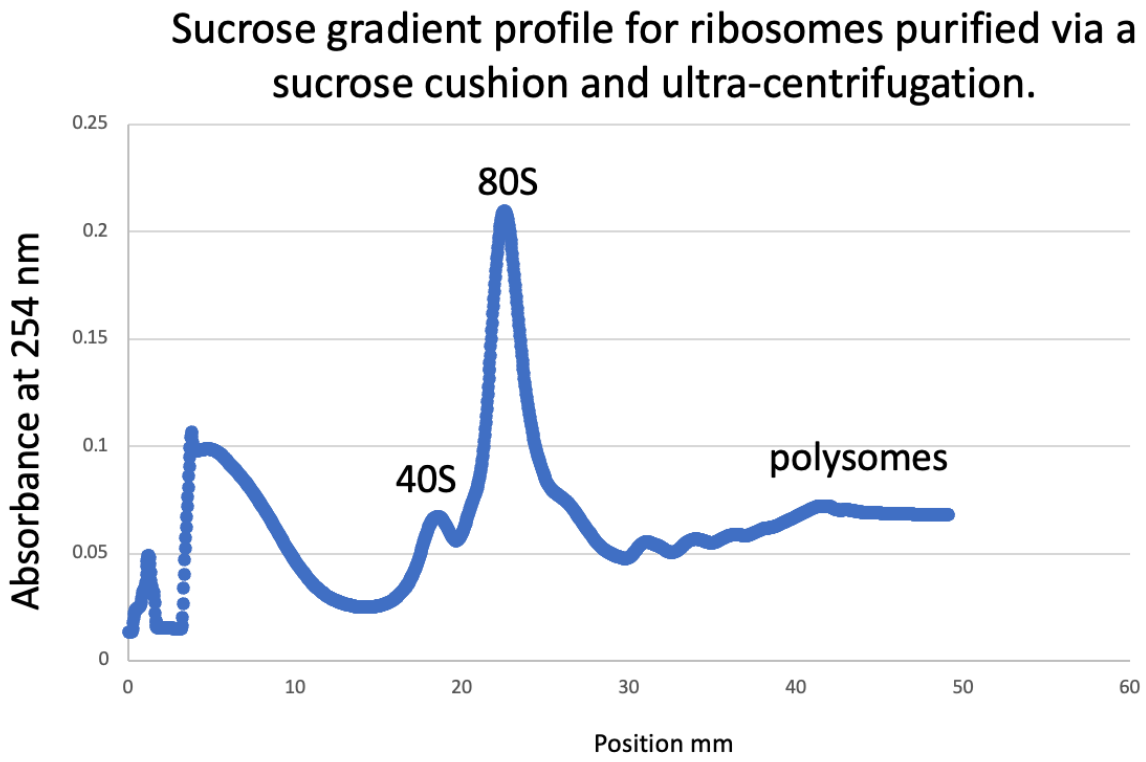


Figure 4.7. A sucrose gradient profile of the *Arabidopsis* wild type ribosomes purified through a sucrose cushion and ultra-centrifugation through a sucrose gradient. *Arabidopsis* wild type seedling tissue was homogenised with liquid nitrogen and resuspended in a ribosome extraction buffer. Ribosomes were pelleted through a 1 M sucrose cushion and ultra-centrifugation. The pellet was resuspended and the 80S ribosomes were separated by a sucrose gradient and ultra-centrifugation. The 80S and polysomes have been labelled on the gradient profile. The 80S peak falls approximately at 24 mm as determined by absorbance at 254 nm.

The resulting sucrose gradient profile at 254 nm is typical of reported sucrose gradients and similar to the sucrose gradient of the wheat germ ribosome (Figure 4.4, 4.7). The 80S peak can be found at approximately 24 mm after the smaller 60S peak and before the polysomes. The relative absorbance of the 80S peak is over 0.2 and is 10-fold greater than ribosomes purified via immunoprecipitation (Figures 4.5, 4.7). Therefore, this indicates this method of purification is more successful than by immunoprecipitation. Consequently, this method was employed to produce *Arabidopsis* monosome samples for cryo-EM analysis.

The resulting 80S monosome fraction was prepared for EM imaging to assess sample quality (Figure 4.8). The EM images show particles that are consistent with published EM images of ribosomes. The particles are approximately 25 μm, indicating that ribosomes were successfully purified by ultra-centrifugation by a sucrose cushion and sucrose gradient. There are some small particles in the sample that are not indicative of an 80S ribosome. They are likely free subunits that have dissociated from an 80S ribosome. Compared to the negative stain of ribosomes purified by immunoprecipitation, the sample prepared by purification by sucrose cushion and gradient gave a better yield of 80S monosomes, supporting the idea that this sample is more suitable for cryo-EM analysis.

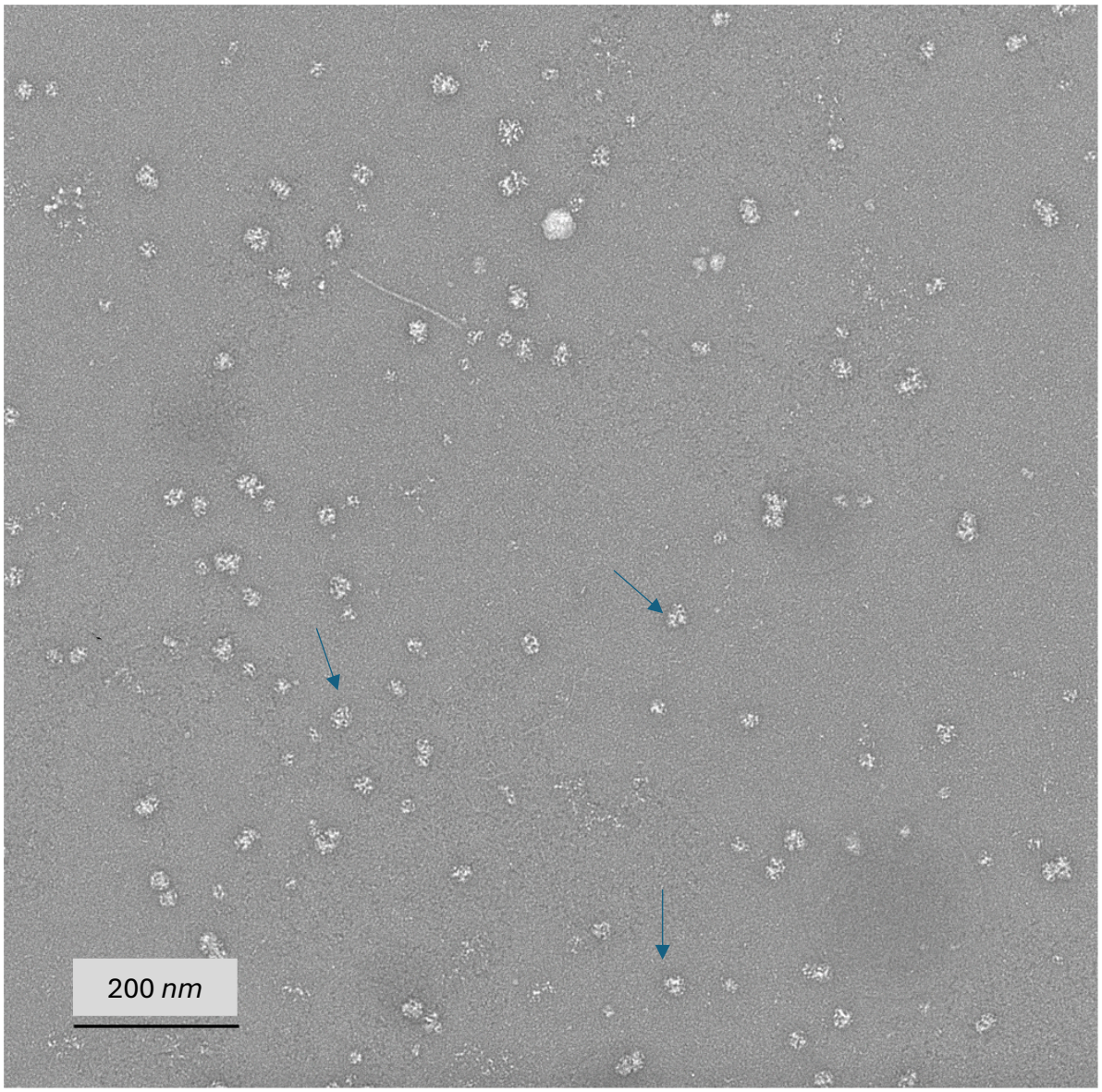


Figure 4.8. Negative stain EM images of *Arabidopsis* ribosomes purified by ultra-centrifugation through a sucrose cushion and sucrose gradient. *Arabidopsis* ribosomes purified in 4.5 were concentrated down by centrifugation. A buffer exchange was performed to resuspend ribosomes in a buffer appropriate from cryo-EM. 5 uL of sample was negatively stained and examined using the T-12 electron microscope at 120 kV. Scale bar is provided. Arrows indicate typical ribosomes.

4.2.2. Cryo-EM image processing of crYOLO auto-picked particles results in a 15.2 Å low resolution model of the Arabidopsis ribosome.

To generate the first model of the *Arabidopsis* 80S monosome, single particle cryo-EM analysis was performed on ribosomes purified through a sucrose cushion and gradient (Figures 4.7-8). An electron microscope operating at 300 kV and equipped with a Falcon III camera was used for this, and a dataset of 21,587 micrographs were obtained (Figure 4.9).

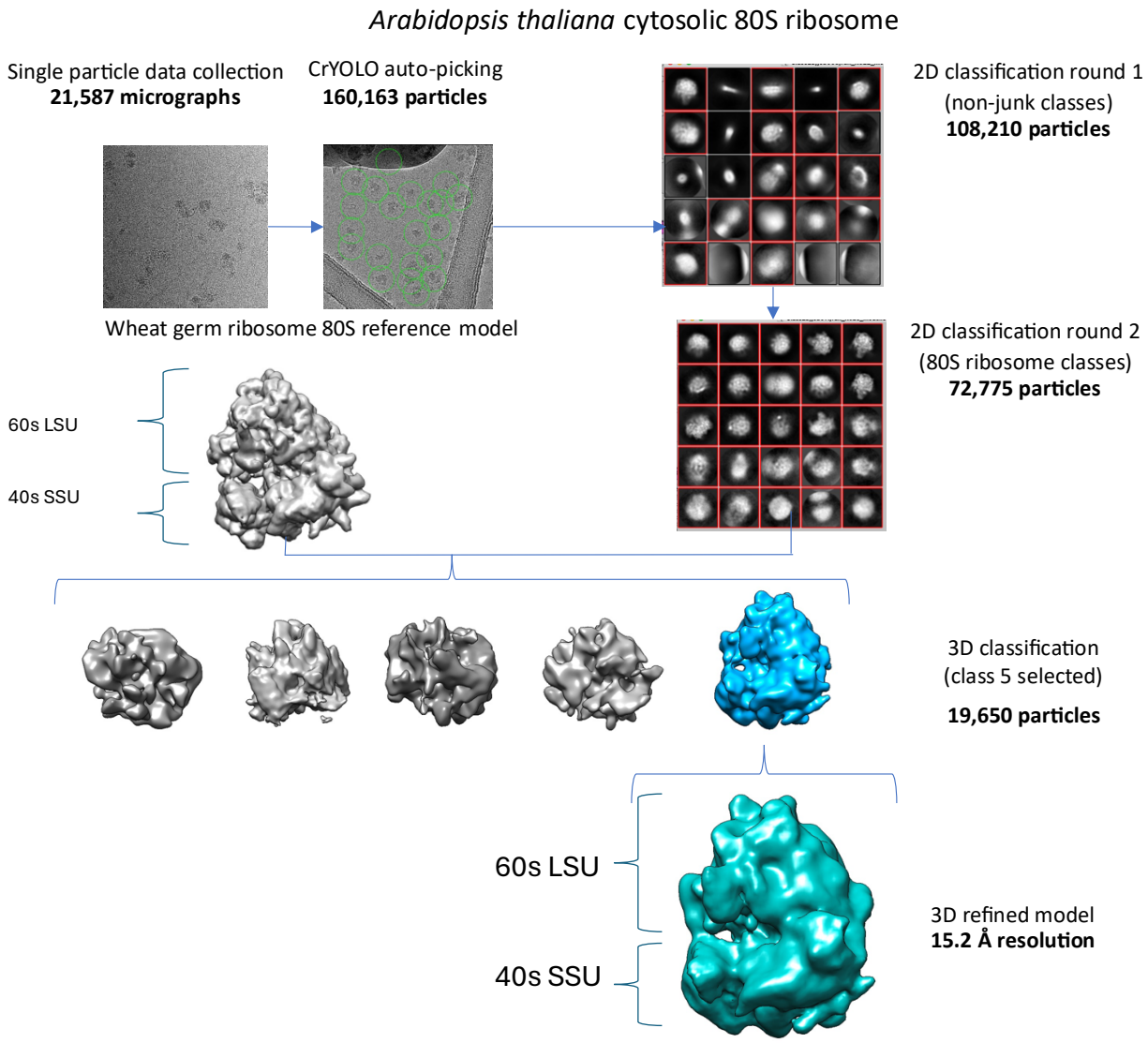


Figure 4.9. Schematic illustration of image processing of cryo-EM data using RELION v3.1.2 and crYOLO auto picked particles. From 21,587 micrographs, 160,163 particles were picked using the crYOLO auto-picking software. 2 rounds of 2D classification were carried out to remove junk particles, resulting in 72,775 particles submitted to 3D classification. From the 5 3D classes generated, only one class resembled the wheat germ 80S ribosome. The remaining 4 classes are likely ribosome subunits. The only ribosome 3D class (19,650) was selected for 3D refinement. 3D refinement of these particles resulted in a 15.2 Å resolution. Representative 2D classes selected and taken forward are highlighted in red boxes. 3D classes taken forward are coloured in blue. Figure made with UCSF Chimera.

Given the negative stain EM confirmed the presence of *Arabidopsis* ribosomes in our samples we aimed to obtain a structure of these ribosomes by cryo-EM. A grid loaded

with *Arabidopsis* ribosomes purified using a sucrose cushion were examined via cryo-EM. Data from the grid was captured using an electron microscope as described in methods 2.9. A total of 21,587 movies were recorded and processed in Relion (165). The data underwent motion correction and CTF correction using CTFFIND (165,166). Particle picking was performed by crYOLO resulting in 160,163 particles (167). After particle extraction, the particles were binned by 5 to expedite image processing, yielding 3.7 Å/pixel (original pixel size was 0.74 Å). 2D classification was conducted to filter out undesired particles. Classes consistent with an 80S monosome were selected, leaving 43,695 particles for 3D classification. 2 rounds of 3D classification were then performed using a 3D map from wheat germ ribosome filtered at 60 Å resolution. The first iteration produced a single class closely resembling a ribosomal subunit (Figure 4.8).

The remaining particles were further classified, resulting in a single 3D class that aligned well with the WGE ribosome reference. This left 19,650 particles, which underwent 3D refinement to produce a 3D map of the *Arabidopsis* ribosome with a resolution of ~15.2 Å (Figure 4.2.3 C). When overlaid on the wheat germ ribosome, the model is consistent with a known eukaryotic 80S ribosome. However, challenges emerged when unbinning particles for another refine-3D process. This step struggled to align the particles, hindering further resolution. Binning in cryo-electron microscopy refers to the process of combining pixel data to reduce the resolution of images thereby, increasing the signal-to-noise ratio therefore, speeding up computational tasks. Unbinning, on the other hand, involves reverting the particle data to its original, higher resolution by separating the combined pixels. By initially binning particles this facilitated the removal of low resolution particles and avoiding excessive computational resources.

One possible explanation for the limitations in resolution improvement is related to the sample preparation process. Residual sucrose from the buffer, if not entirely removed, could have interfered with the clarity and contrast of the images, affecting the alignment accuracy. Sucrose, commonly used in buffer solutions for ribosome purification, can contribute to background noise in cryo-EM imaging if not eliminated during the washing steps. Furthermore, the particle density on the grid was suboptimal. A lower particle density can lead to difficulties in finding sufficient particles with adequate orientation diversity, which is essential for accurate 3D reconstruction. Both factors - residual sucrose and particle density - could have cumulatively contributed to the alignment challenges observed, thus impeding the enhancement of resolution in our *Arabidopsis* ribosome model.

In conclusion, I present a low-resolution model of the *Arabidopsis* monosome. Plans to improve resolution through another data collection were cancelled as another group generated a model of 2 Å therefore, efforts were reorientated to generating a CPuORF stalled ribosome.

4.2.3. Sucrose gradient purification and co-immunoprecipitation resulted in the purification of CPuORF stalled wheat germ extract ribosomes for EM analysis.

Previous studies have purified stalled ribosomes through an arrested peptide via an N-terminal tag (110,111). To purify an aCUTS CPuORF stalled ribosome, the heat shock responsive CPuORF46 (HEAT) was N-terminally his tagged and cloned downstream of an

SP6 promoter (HisHisHA:HEATCPuORF) as in Van der Horst *et al* (110). Capped mRNA was translated by the Promega wheat germ translation kit. The *in vitro* reaction (WGE) was then pelleted by ultracentrifugation to form a ribosome pellet (Methods 2.8, Figure 4.10). The first ribosome pellet (SC1) would contain both vacant ribosomes and stalled ribosomes. The pellet is then resuspended and His-purified using magnetic anti-His beads to pull out the stalled ribosomes from the vacant ribosomes. His-purified aCUTS CPuORF stalled ribosomes were then eluted with 200 mM imidazole. The eluted ribosomes were then purified through a second sucrose cushion (SC2) by ultra-centrifugation that resulted in a pellet exclusively containing stalled ribosomes. The final pellet was resuspended in a buffer suitable from cryo-EM.

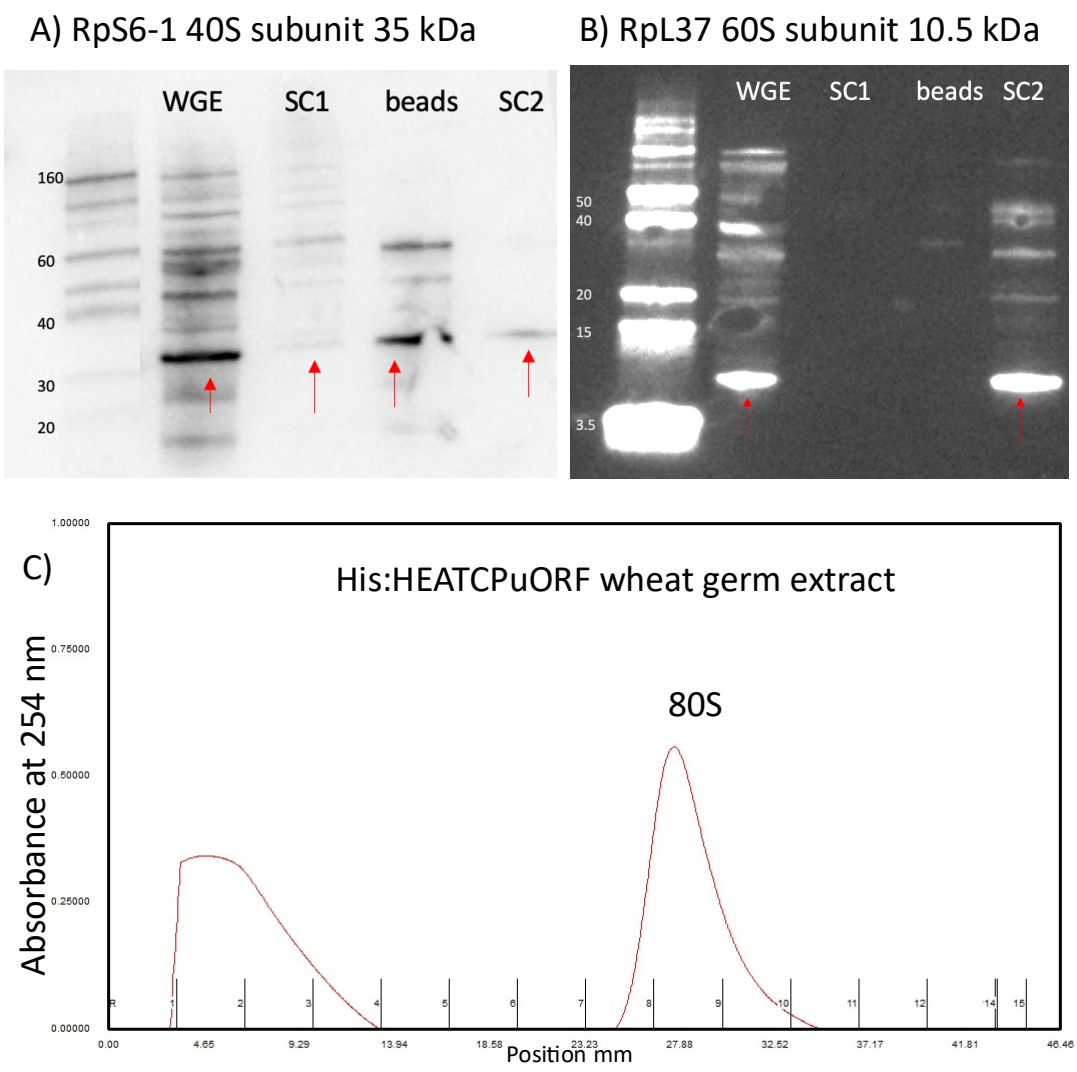


Figure 4.10. Western blots and sucrose gradient profile for ribosome purification verification. Aliquots from various stages of the His-purification process were immunoblotted to detect A) small and B) large ribosomal subunits. Samples were taken from an *in vitro* translation reaction (WGE). Samples correspond to the first sucrose cushion (SC1), which contains both stalled and non-stalled ribosomes; the ribosomes that remained on the beads after elution (beads); and the second sucrose cushion with his-purified CPuORF stalled ribosomes (SC2). The small subunit RpS6-1 and the large subunit RpL37 are ~35 kDa and ~10.5 kDa, respectively. Bands corresponding to these proteins are marked with red arrows. Sucrose gradient profile from the wheat germ translation assay. A distinct peak, at ~27 mm with an absorbance of 0.55 a.u corresponds to the 80S ribosome.

To confirm the presence of the large and small ribosome subunits, a western blot was performed during different stages of the His-purification protocol (Figure 4.10). Aliquots were taken from the *in vitro* translation reaction post translation (WGE), the resuspended

ribosome pellet after the first sucrose cushion (SC1), the beads post elution (beads), and finally, the resuspended pellet containing just stalled ribosomes after the second sucrose cushion (SC2). The beads aliquot represents stalled ribosomes that did not elute. The western blots were performed using antibodies for the small subunit (RpS6-1) and the large subunit (RpL37) (1).

The presence of the large and small subunits was confirmed by immunoblotting in the initial translation reaction (WGE) and the final sucrose cushion (SC2). Immunoblotting also indicated that the elution step could be improved due to the presence of the small subunit on the beads. Interestingly, there was no band for the large subunit in the first sucrose cushion (SC1). Immunoblotting detected the presence of the small and large subunits downstream (SC2). Finally, an aliquot from the WGE was purified through a sucrose gradient and concentrated by using 100,000 kDa centrifugal filters (Figure 4.10 C). The sucrose gradient profile demonstrated a singular peak at 27 mm corresponding to the 80S ribosome. However, the gradient profile did not contain peaks indicating the presence of free ribosomal sub-units or polysomes.

To check the presence of ribosomes and the suitability of the sample for single particle cryo-EM samples from the sucrose cushion 2 (SC2) were screened by negative stain EM (Methods 2.8-10; Figure 4.11). EM analysis showed low abundance of ribosomes (Figure 4.11).

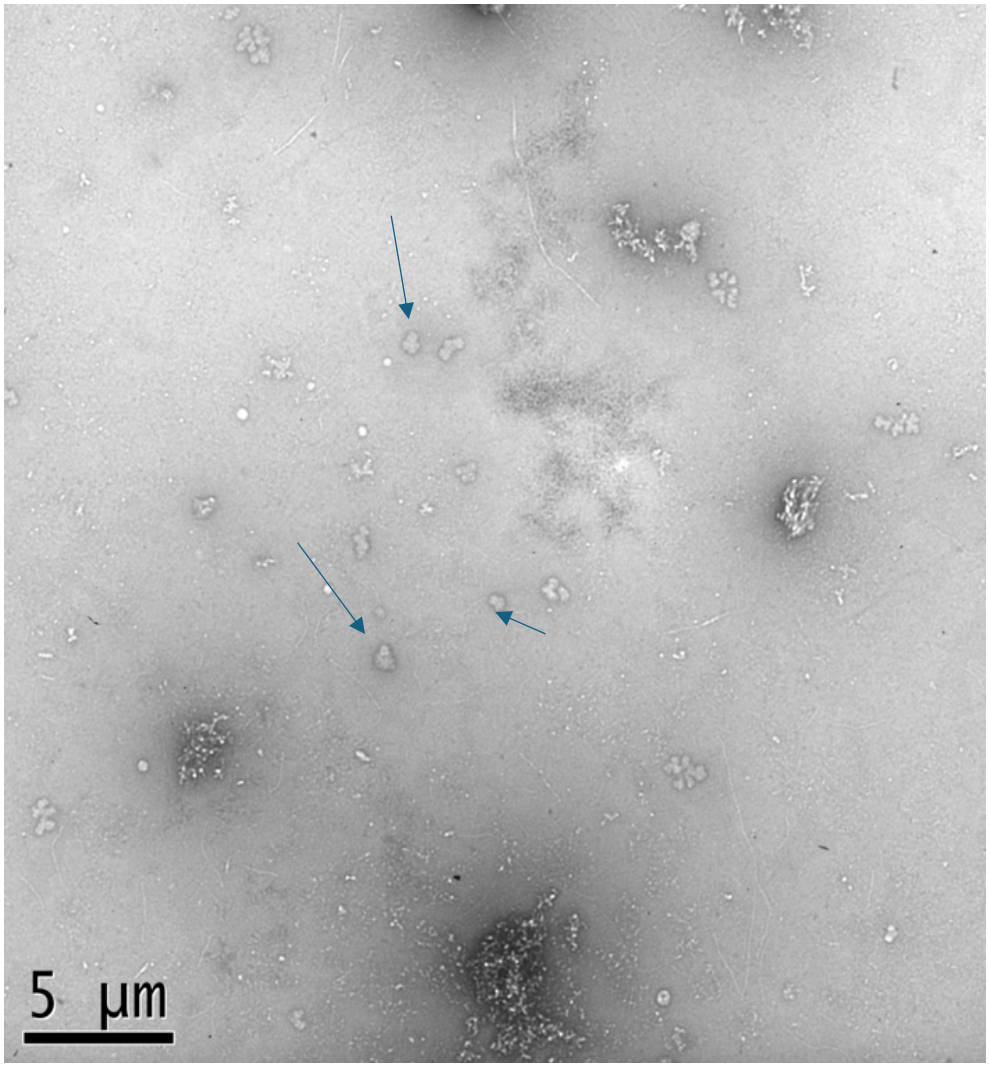


Figure 4.11. Negative Stain EM Images of his-Purified Ribosomes. Samples from the *in vitro* translation reaction, both before and after His purification, were stained and observed using T-12 microscope. The image depicts His-purified CPuORF stalled ribosomes. Scale bars are included. Arrows indicate examples of ribosomes.

4.2.4. Cryo-EM image processing of crYOLO auto-picked particles results in low resolution models of vacant ribosomes.

Sample containing ribosomes and stalled ribosomes were analysed through cryo-EM to generate a model of a CPuORF stalled ribosome. Figure 4.12 A) provides an example of the sample. There is good contrast between the ribosomes and the background suggesting optimal imaging conditions for data collection. The grid atlas indicates a

gradient in ice thickness across the grid commonly observed in plunge-frozen grids (Figure 4.12).

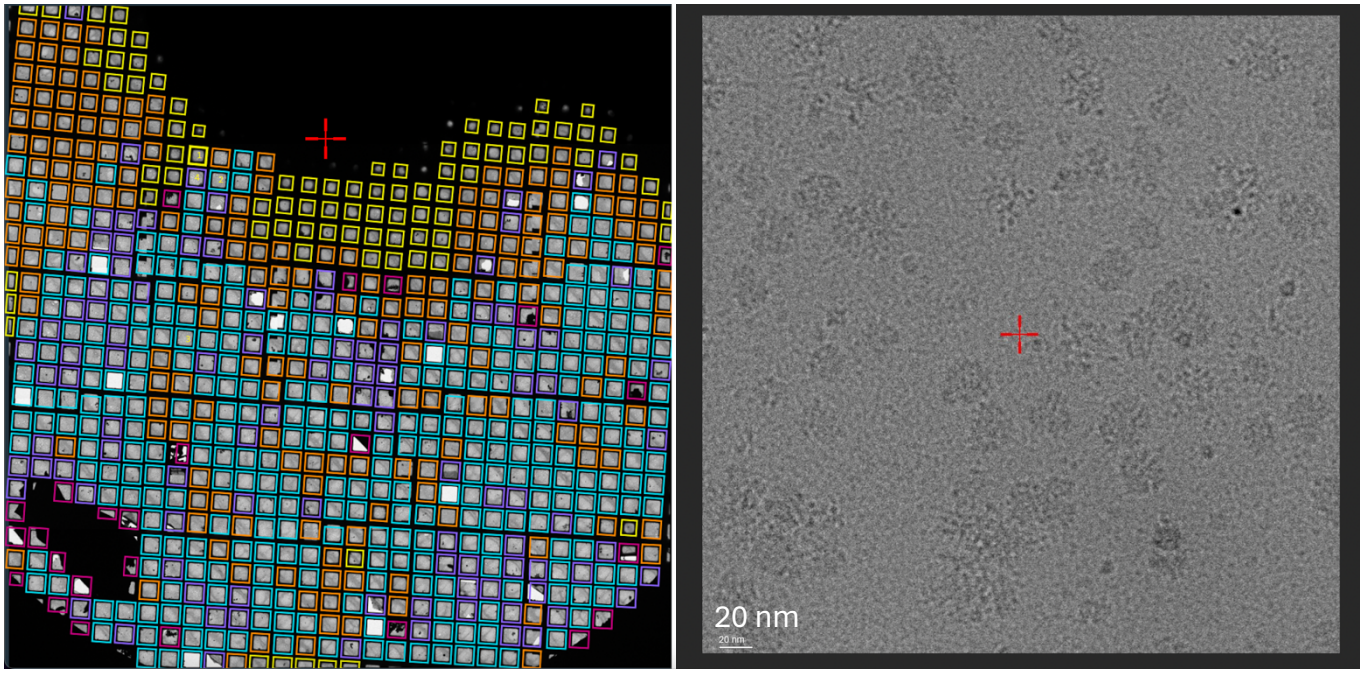


Figure 4.12. Cryo-EM screening image of a grid loaded with a sample from the *in vitro* translation reaction and the grid atlas. A) Grid atlas of the same sample. Similar squares in terms of ice thickness are grouped by colour. B) Screening image at 96,000 x magnification taken on an electron microscope operated at 300 kV. Scale bar is provided at the bottom left corner.

8259 movies of micrographs were collected. Following data collection, the data set was subjected to motion correction and CTF correction. Motion correction fixes blurring and distortions by movements within the frames of the data. CTF correction estimates the defocus deployed during data collection to compensate for electron aberrations caused by the electron microscope. crYOLO is a selective particle picking software, which employs machine learning to avoid picking junk particles compared to program used within RELION. An example of three crYOLO particle picked micrographs is provided, demonstrating crYOLO's success as it picked particles and not grid edges (Figure 4.13).

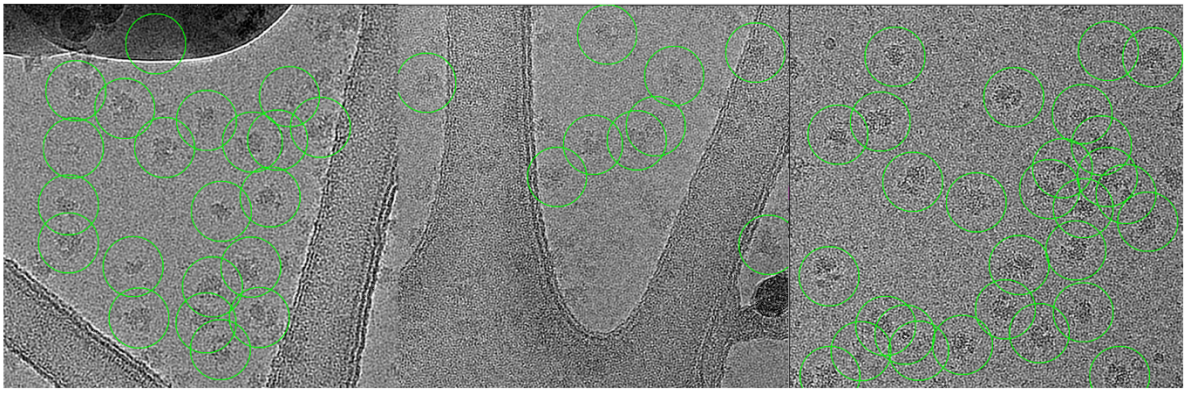
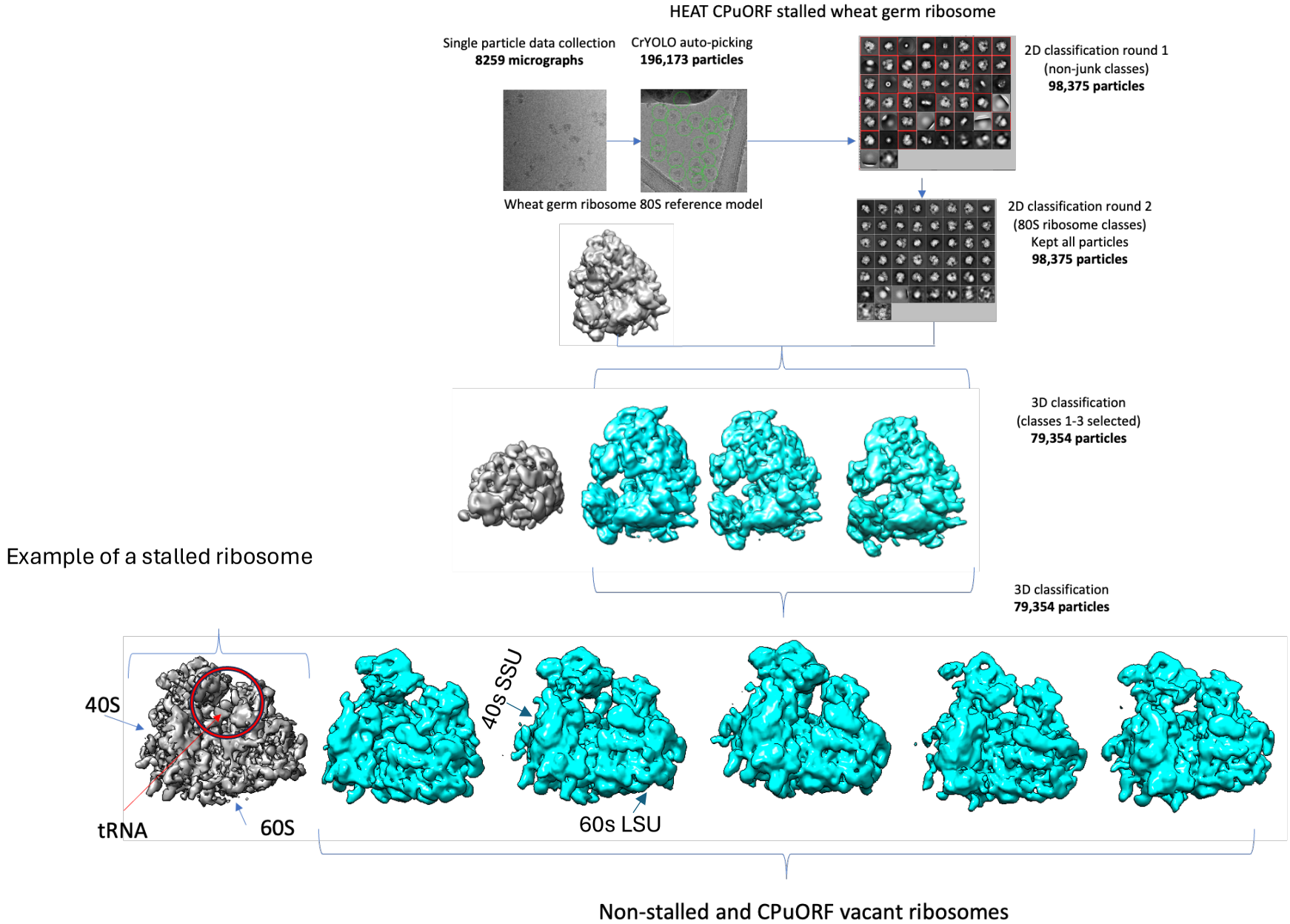


Figure 4.13. Micrographs with particles picked by crYOLO. Three micrographs are provided from the data collection of the *HEAT* CPuORF stalled wheat germ extract sample. Pixel size is 0.82 Å. Box size is 500 and indicated by a green circle. Picking threshold was 0.1.

To generate the first model of an aCUTS CPuORF stalled ribosome, single particle cryo-EM analysis was performed on wheat germ ribosomes that have translated SP6:HisHEATCPuORF mRNA (Figures 4.10). An electron microscope operated at 300 kV and equipped with a Falcon III camera was used for this, and a dataset of 8259 micrographs was obtained (Figure 4.14).



Example of a stalled ribosome

Figure 4.14. Schematic illustration of image processing of cryo-EM data using RELION v3.1.2 and crYOLO auto picked particles of a HEATCPuORF stalled wheat germ sample. From 8259 micrographs, 196,173 particles were picked using the crYOLO auto-picking software. 2 rounds of 2D classification were carried out to remove junk particles, resulting in 98,375 particles submitted to 3D classification. From the 4 3D classes generated, three classes resembled a Wheat germ 80S ribosome (79,354 particles). The remaining class is likely a ribosome subunit. A second round of 3D classification was performed and 5 classes of 80S ribosomes were generated. This indicates that all non-80S particles have been removed. All five classes mRNA channel were investigated to see if tRNA is still bound to the ribosome. Bound tRNA would indicate a stalled ribosome. For reference the fungal arginine attenuator stalled wheat germ ribosome (EMD-176) has been provided. The fungal arginine attenuator peptide stalled ribosome and 5 3D models of the HEATCPuORF stalled wheat germ ribosome were orientated to reveal the mRNA channel. The 40s and 60s subunits are labelled on the fungal arginine attenuator peptide stalled ribosome reference. The tRNA is labelled with a red arrow. The red circle highlights the mRNA channel. There is a tRNA in the reference model however, all mRNA channels appear vacant in our dataset. Therefore, indicating the particles in this sample are not stalled. Representative 2D classes selected and taken forward are highlighted in red boxes. 3D classes taken forward are coloured in blue. Figure made with UCSF Chimera.

Particles picked by crYOLO were binned by 5 resulting in a pixel size of 4.1 Å/pixel. The dataset of 196,173 particles was subject to 2D classification to remove low quality or unwanted particles (Figure 4.14). The first round of 2D classification removed particles that did not resemble an 80S ribosome, which corresponded to half of the particles initially selected (Figure 4.14). The removed classes may resemble ribosome subunits, degraded ribosomes, and grid edges. The proportion of ribosome subunits was surprising as the sucrose gradient profile suggested they were not present in the sample (Figure 4.10). This may suggest that the stalling mechanism was not stable. After removing these classes another 2D classification was performed on the remaining 98,375 particles. The remaining classes resembled published ribosomes and all particles were taken through to 3D classification (Figure 4.14).

98,375 particles were subjected to two rounds of 3D classification using a wheat germ ribosome filtered to 60 Å as a reference (Figure 4.14). The first round of 3D revealed that 23.36% of the input particles corresponded to a ribosomal subunit. Again, this was interesting when considering the sucrose gradient profile as this suggested the absence of free subunits (Figure 4.10). The remaining ribosome particles were selected and subject to a second round of 3D classification into 5 classes. The 5 classes 3D models were investigated further using to determine if any class contains a stalled ribosome (Figure 4.14). The mRNA channel of the five classes was examined for the presence of tRNA. If the mRNA channel was occupied this would indicate that a nascent peptide is occupying the exit tunnel. For reference, the fungal arginine attenuator peptide is provided. Examination of the mRNA channel indicates that the sample does not contain stalled ribosomes, or at least this method of data processing was unable to find a stalled ribosome population.

4.3. Discussion.

The aim of this chapter was to use structural biology to elucidate the aCUTS CPuORF ribosome stalling mechanism. Previous studies that have utilised single particle cryo-EM analysis have elucidated eukaryotic and prokaryotic ribosome stalling mechanisms (20,110,111,139). Recently, cryo-EM has also provided insights into the sucrose dependent rCUTS stalling mechanism of the *Arabidopsis* bZIP11 CPuORF (110). The long-term goal of this project is to solve the structure of an *Arabidopsis* CPuORF stalled ribosome. To achieve this, this study first began to generate high resolution models/maps of the *Arabidopsis* cytosolic ribosome. Surprisingly, *Arabidopsis* is one of the few model organisms where the ribosome is yet to be resolved. This thesis presents a low-resolution

model of an *Arabidopsis thaliana* cytosolic ribosome (Figure 4.9). Part way through this project a group achieved (yet to be published) a global resolution of approximately 2 Å of an *Arabidopsis* cytosolic ribosome, so remaining efforts of this chapter were spent solving the structure of the HEAT CPuORF stalled wheat germ ribosome. Ultimately, this thesis was unable to generate stable samples of the aCUTS HEAT CPuORF stalled ribosomes.

4.3.1. Producing higher resolution models of the *Arabidopsis* ribosome.

An initial aim of this project was to produce the first high resolution map/structure of the model organism *Arabidopsis thaliana* cytosolic ribosome. This thesis presents a low-resolution model of the ribosome at a global resolution of 15.2 Å (Figure 4.9). Furthermore, this study developed two protocols that successfully purified ribosomes from plant tissue by immunoprecipitation and ultra-centrifugation through a sucrose cushion and gradient (Figure 4.4-9).

To generate high-resolution models of the *Arabidopsis* ribosome, a high yielding and high-quality sample must be loaded onto a grid. When examining the 3D classifications of the *Arabidopsis* ribosome (Figure 4.9), the 2D projections of the 3D classifications do not match the quality of those for the stalled wheat germ ribosome (Figure 4.9-14). Good 2D projections of 3D models exhibit a clear black background contrasting with the white protein. High contrast between the background and protein indicates the quality of the data and potential to generate high resolution images. Poor contrast could indicate non-optimal ice thickness, buffer suitability or limitations of pre-processing (Figure 4.9). The

software used to pre-process the *Arabidopsis* ribosome data set was the same as for the wheat germ extract, thus indicating that sample preparation was the problem.

There was a problem with particle alignment in refine 3D jobs post un-binning of particles (Figure 4.9). When performing a refine 3D job on binned particles a resolution of 15.2 Å was achieved. When subjecting unbinned particles to a refine 3D job, the particles would not align consistently to the reference. The aim of binning the particles is to make to data sets 'smaller' to speed up data processing at the cost of losing data and a lower resolution. When unbinning particles, a higher resolution is supposed to be achieved as there is now more data for each particle. When including more data, the particles are unable to align resulting in a low-resolution model. This overall raises questions regarding the quality and integrity of the sample. It implies that the enhanced unbinned data set, rather than contributing to a higher resolution, exacerbated the alignment of particles. Overall, failure to align particles during the refine 3D job and poor contrast in 2D and 3D classifications suggests the sample is of low quality to be resolved by cryo-EM.

Publishing a high-resolution ribosome map of the model organism *Arabidopsis thaliana* would contribute significantly to the existing body of literature in plant molecular biology. This is of particular interest in plant sciences as plant ribosomal proteins have multiple paralogues (2). As plant ribosome composition may diverge, the *Arabidopsis* ribosome map would be an essential reference for comparative studies (151,172,173,185). By examining the differences and similarities in ribosomal structures, researchers can gain a deeper understanding of evolutionary biology and the functional diversification of ribosomes and protein synthesis in plants and eukaryotes.

This thesis was unable to produce a sample suitable to generate a high-resolution map of the *Arabidopsis* cytosolic monosome. However, Alan Warren's group at Cambridge have generated a high-resolution map of *Arabidopsis* ribosomes by purifying ribosomes from root tissue this may reduce sample heterogeneity as it excludes the 70S chloroplastic ribosome from polluting the sample (186). Because of this, this approach may have led to the acquisition of higher-quality samples, thereby facilitating more accurate cryo-EM analyses. Future experiments could include testing the efficacy of seedling-based extraction of *Arabidopsis* ribosomes. To avoid replicating efforts future cryo-EM studies could resolve the *Arabidopsis* chloroplastic ribosome, *Arabidopsis* polysomes and *Arabidopsis* stalled ribosomes.

4.3.2. Purifying CPuORF-stalled ribosomes.

This thesis attempted to generate a high-resolution structure of an aCUTS CPuORF (CPuORF46) to complement the published rCUTS CPuORF (110) (Figures 4.12-14). To purify a stalled aCUTS CPuORF ribosome, I utilised an N-terminal His tag and a purification protocol as demonstrated by van der Horst et al., (110). Western blotting and negative staining demonstrated the presence of ribosomes in the final sample post his-purification and two ultra-centrifugation steps (Figure 4.10). However, poor yield and grid quality of His-purified stalled ribosomes suggested that this sample was not appropriate for cryo-EM imaging (Figure 4.11). As a consequence, the unpurified *in vitro* translation reaction was used (Figure 4.12-14). These grids would harbour a population of vacant ribosomes and CPuORF stalled ribosomes. To resolve the structure of a stalled ribosome,

this would require computationally removing the population of stalled ribosomes during 3D classification (Figure 4.12-14) (187–190).

Surprisingly, 3D classification revealed a significant proportion of particles were a ribosomal subunit (Figure 4.14). The sucrose gradient of the *in vitro* translation reaction suggested the absence of free ribosomal subunits (Figure 4.10). Collectively, this may suggest that the stalling mechanism was not stable and future investigations could utilise cycloheximide to fix and stabilise the ribosomes post-translation. Cycloheximide is utilised in Bhushan's protocol to resolve the rCUTS fungal arginine attenuator peptide but not in van der Horst *et al*., (110,139), which used its signal (sucrose) to stabilise the stalling mechanism to produce a structure of the rCUTS plant CPuORF. The decision to not use cycloheximide was out of concern that cycloheximide may create noise as it may produce populations of stalled ribosomes that have been stalled due to cycloheximide rather than the arrest peptide.

Cryo-EM investigations into arrest peptides and their stalling mechanisms have utilised N-terminal tags to isolate and purify stalled ribosomes (110,111,139). Comparative investigations of cryo-EM structures of a CPuORF and non-CPuORF arrest peptide mechanisms suggest an ancient mechanistic framework for ribosome stalling (20,110,111,135,139). The commonalities between the reported mechanism include: 1) the nascent peptide interacts with exit tunnel proteins and rRNA nucleotides. 2) Nascent peptide interacts with the tunnel constriction formed by uL4 and uL22. 3) Nascent peptide interferes with PTC function. 4) metabolite responsive arrest peptides utilise a conserved metabolite pocket to conditionally stall the ribosome. 5) The biochemistry

between nascent peptide amino acids, exit tunnel residues and metabolite interfere with exit tunnel charge to induce stalling. Currently, there is only one published study into a plant CPuORF's (bZIP11) stalling mechanism (110). Although a high-resolution structure of the bZIP11 CPuORF stalled ribosome was obtained, the interactions within the exit tunnel were not fully resolved. This highlights the difficulties in capturing detailed structures despite achieving a resolution of 3.7 Å (110).

Future comparative cryo-EM studies could elucidate plant CPuORF stalling mechanisms (20). By comparing high resolution maps and structures of CPuORF stalled ribosomes key residues in stalling and signal sensing can be identified. Current studies have identified a conserved metabolite pocket in the exit tunnel that facilitates the conditional stalling of eukaryotic and bacterial arrest peptides (110). Comparative studies have identified common ribosome exit tunnel interactors in arrest mechanisms. Conserved rRNA nucleotides and exit tunnel proteins (uL4, uL22, uL16) have been implicated in multiple arrest mechanisms. In the previous chapter, mutations in conserved exit tunnel proteins (uL4, uL16 and RACK1) have been shown to affect the aCUTS stalling mechanism (Figure 3.6-8), collectively, suggesting a conserved mechanistic framework to stall and conditionally stall the ribosome.

Chapter 5: An analysis of a CPuORF database.

5.1. Introduction

5.1.1. CPuORF databases

Databases serve as valuable community resources by centralising information. They can streamline research and standardise information. For example, uORFdb (191) is a eukaryotic uORF literature database. When considering CPuORF databases, two currently exist uORFLIGHT (192) and The *Arabidopsis* CPuORF database (3). uORFLIGHT is a plant uORF database that facilitates the exploration of uORF variation in *Arabidopsis thaliana* and rice (192). uORFLIGHT has a specific section for *Arabidopsis* CPuORFs and their rice homologues, but this is incomplete(192).

The *Arabidopsis* CPuORF database is a complete record of currently identified *Arabidopsis thaliana* CPuORFs and evidence for their functionality (3). The *Arabidopsis* CPuORF database contains information such as the gene identifier, CPuORF homology group, the mORF, mode of CUTs, CPuORF class, length in amino acids, intercistronic distance, amino acid sequence, and identification method (3). This database has catalogued 150 confirmed and 'low confidence' *Arabidopsis* CPuORFs (3). An updated eukaryotic CPuORF database will be made public at https://theeukaryoticcpuorfdatabase.github.io.

5.1.2. Sequence analysis.

CPuORFs stall the ribosome in a sequence dependent manner suggesting it is the peptide sequence that confers function (Figure 5.2, Table 1.4-5). The evidence for CPuORF function is covered extensively in chapter 1. Molecular and cryo-EM data suggests that there is a correlation between conserved amino acids and stalling mechanisms in CPuORFs and other known arrest peptides (Table 1.4-5) (20,110,111,135,139). CPuORF sequence analysis, through bioinformatic tools, could reveal insights into the CPuORF CUTS mechanism. Stalling and signal sensing is potentially conferred through specific biochemistry or electrostatic charge within the exit tunnel (20,110,148). Furthermore, non-optimal codon usage and specific amino acids have been reported to slow down or halt translation (193). Moreover, bioinformatic tools such as LOGO sequence analysis have been demonstrated to identify functional residues (194). A sequence analysis into CPuORF peptide sequences may identify common motifs that confer stalling or identify residues of interest to explore further.

5.1.3. CPuORFs identified in D. melanogaster and H. sapiens.

Several plant CPuORFs have been functionally characterised *in vivo* and *in vitro* however, experimental data for CPuORFs outside of plants can only be found in one study (106). A pipeline was developed (ESCUA) that can identify CPuORFs (105,106) in 8 steps. 1) Data preparation from genome and transcriptome databases. 2) uORF sequences were extracted from 5'leader regions. 3) The fusion ratios between regular-uORFs and mORFs were calculated to select for uORF peptides that have been conserved under a functional constraint. 4) Find homologous proteins using a tBLASTn search. Regular-uORFs that

have BLAST hits from other species are selected for further analysis. 5) mORF sequence analysis to find homologous genes. 6) a K_a/K_s analysis is performed to select for regular-uORFs that have been conserved at the peptide level. 7) Determination of the taxonomic range for CPuORF conservation. Those with an taxonomic range in three different orders classified as CPuORFs. Finally, 8) Manual validation of identified CPuORFs.

Initially, the pipeline was applied this to 5 angiosperm genomes *Arabidopsis thaliana*, tomato *(Solanum lycopersicum),* poplar (*Populus trichocarpa*), grape (*Vitis vinifera*) and rice (*Oryza sativa*). Through this methodology 89 novel CPuORF families (homology groups - HGs) were identified. Subsequently, the Takahashi group applied the same pipeline to four animal genomes, fruit flies (*Drosophila melanogaster),* zebrafish (*Danio rerio*), chicken (*Gallus gallus*), and humans (*Homo sapiens*) (106). By using the same pipeline on animal genomes, the ESUCA pipeline identified 1373 novel animal CPuORF HGs.

Post-ESCUA 17 human CPuORFs were selected for *in vitro* investigation (106). They performed a luciferase assay that compared wild type and frameshifted CPuORF's effect on their downstream mORF (LUC). 9 out of 17 frameshifted CPuORFs investigated displayed higher luciferase levels than their wild type counterparts. This suggests that these 9 human CPuORFs attenuate mORF expression in a sequence dependent manner. This data reflects frameshift mutants observed on *Arabidopsis* CPuORFs and implies that the wild type's attenuation of the mORF is a consequence of ribosome stalling (3,6,7,106,110) (Figures 3.7-18). Consequently, this study serves as the only evidence that non-plant CPuORFs stall the ribosome during translation.

5.1.3. Aims.

The overarching goal of this chapter is to broaden our understanding of the occurrence, conservation, and potential function of CPuORFs across eukaryotic organisms.

Aims:

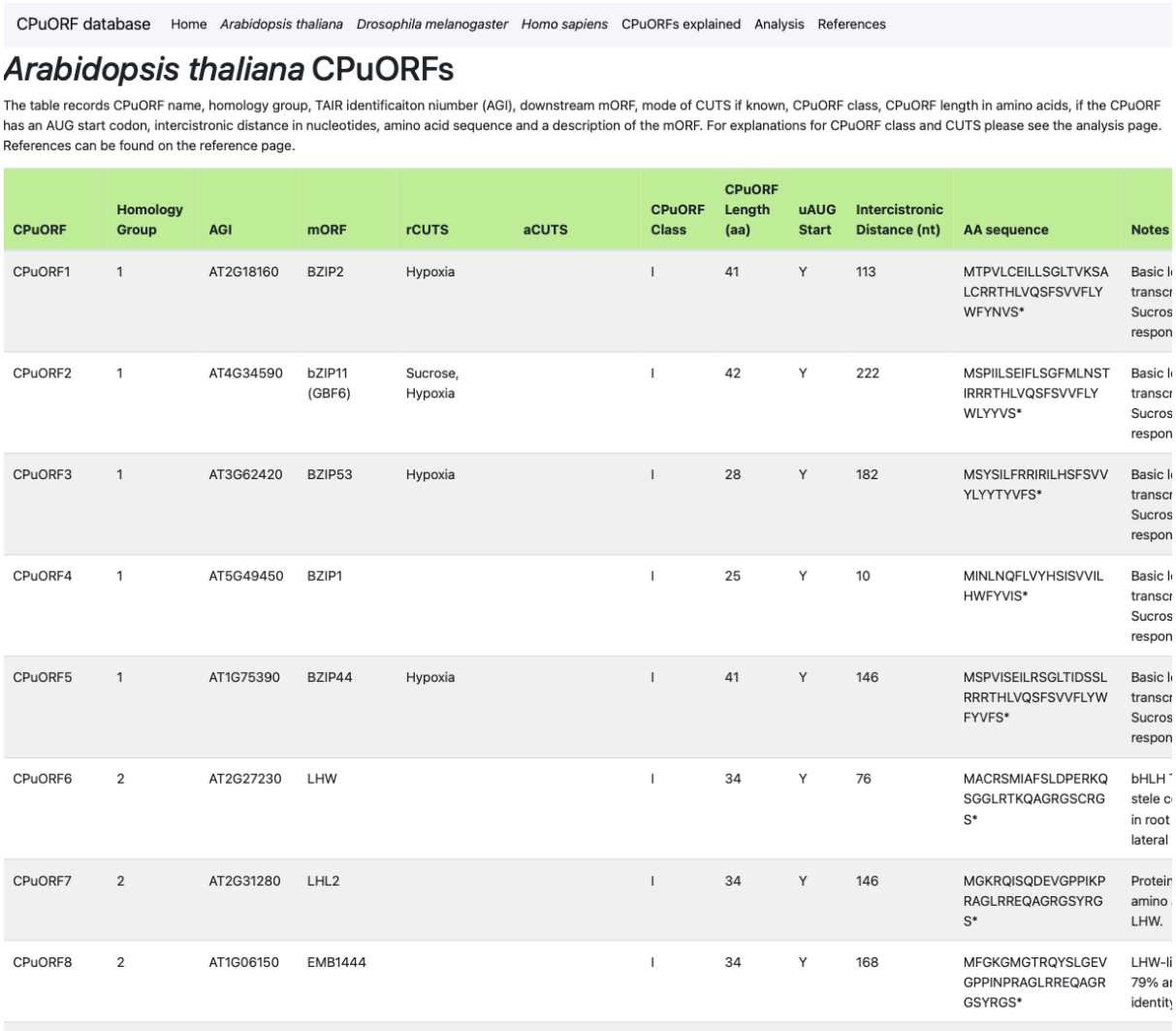
- ***Establish a comprehensive eukaryotic CPuORF Database:*** This database will facilitate future investigations into the potential roles in gene expression and applications of CPuORFs.
- ***Analyse DNA and amino acid sequences of CPuORFs:*** Investigating conservation of CPuORF DNA and amino acid sequences can potentially reveal functional domains.

5.2. Results

5.2.1. An updated eukaryotic CPuORF database.

Initially, this thesis aimed to produce an expanded CPuORF database (Figure 5.1). The previously published *Arabidopsis* CPuORF database was updated to include CPuORFs identified in *Drosophila melanogaster* and *Homo sapiens* (Figure 5.1) (3). The database detailed information such as gene identifier, CPuORF homology group, the mORF, mode of CUTS, CPuORF class, length in amino acids, intercistronic distance, amino acid sequence and identification method. It also records whether CPuORFs have been functionally characterized. The database has logged 150 confirmed and 'low confidence' *Arabidopsis* CPuORFs, 57 *Drosophila melanogaster* CPuORFs and 1093 *Homo sapiens* CPuORFs. An R pipeline was written to classify CPuORFs (Figure 5.1-2). The classification

pipeline is the first standardized method for classifying CPuORFs based on the location of their conserved amino acids (Figure 5.2). The database will serve as a community resource to aid research in using CPuORFs in different taxa.



CPuORF database Home *Arabidopsis thaliana* *Drosophila melanogaster* *Homo sapiens* CPuORFs explained Analysis References

Arabidopsis thaliana CPuORFs

The table records CPuORF name, homology group, TAIR identificaiton niumber (AGI), downstream mORF, mode of CUTS if known, CPuORF class, CPuORF length in amino acids, if the CPuORF has an AUG start codon, intercistronic distance in nucleotides, amino acid sequence and a description of the mORF. For explanations for CPuORF class and CUTS please see the analysis page. References can be found on the reference page.

CPuORF	Homology Group	AGI	mORF	rCUTS	aCUTS	CPuORF Class	CPuORF Length (aa)	uAUG Start	Intercistronic Distance (nt)	AA sequence	Notes
CPuORF1	1	AT2G18160	BZIP2	Hypoxia		I	41	Y	113	MTPVLCEILLSGLTVKSA LCRRTHLVQSFSVVFLY WFYNVS*	Basic l transcr Sucros respon
CPuORF2	1	AT4G34590	bZIP11 (GBF6)	Sucrose, Hypoxia		I	42	Y	222	MSPIILSEIFLSGFMLNST IRRRTHLVQSFSVVFLY WLYYVS*	Basic l transcr Sucros respon
CPuORF3	1	AT3G62420	BZIP53	Hypoxia		I	28	Y	182	MSYSILFRRIRILHSFSVV YLYYTYVFS*	Basic l transcr Sucros respon
CPuORF4	1	AT5G49450	BZIP1			I	25	Y	10	MINLNQFLVYHSISVVIL HWFYVIS*	Basic l transcr Sucros respon
CPuORF5	1	AT1G75390	BZIP44	Hypoxia		I	41	Y	146	MSPVISEILRSGLTIDSSL RRRTHLVQSFSVVFLYW FYVFS*	Basic l transcr Sucros respon
CPuORF6	2	AT2G27230	LHW			I	34	Y	76	MACRSMIAFSLDPERKQ SGGLRTKQAGRGSCRG S*	bHLH stele c in root lateral
CPuORF7	2	AT2G31280	LHL2			I	34	Y	146	MGKRQISQDEVGPPIKP RAGLRREQAGRGSYRG S*	Protein amino LHW.
CPuORF8	2	AT1G06150	EMB1444			I	34	Y	168	MFGKGMGTRQYSLGEV GPPINPRAGLRREQAGR GSYRGS*	LHW-li 79% ar identity

Figure 5.1. An updated eukaryotic CPuORF database. The eukaryotic CPuORF database is a community resource including data from *Arabidopsis thaliana, Drosophila melanogaster* and *Homo sapiens*. The initial *Arabidopsis* CPuORF database has been expanded to incorporate non-plant CPuORFs. The database logged information such as gene identifier, CPuORF homology group, mORF name and details, mode of CUTS (if known), CPuORF class, length, sequence, and identification method. The full eukaryotic database can be found at https://theeukaryoticcpuorfdatabase.github.io.

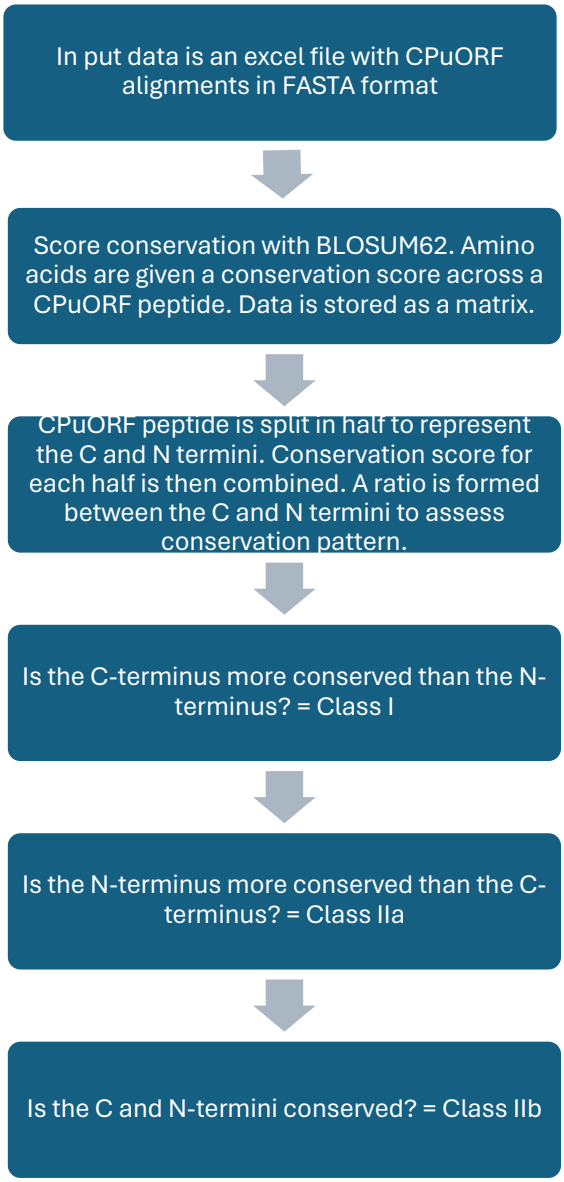


Figure 5.2. Pipeline design to classify CPuORFs. An R pipeline was created to classify CPuORFs based upon their conservation patterns. CPuORFs are classified by the location of the conserved regions. C terminal conserved CPuORFs are class I, N-terminal conserved CPuORFs are Class IIa and CPuORFs with conservation throughout are Class IIb. CPuORF alignments were provided by Takahashi et al (105,106). R packages used include devtools, seqinr and BioManager. The output was an excel file with the alignment and CPuORF class.

5.2.2. CPuORFs are longer than regular-uORFs in *Arabidopsis* and *Human* than in *Drosophila*.

Evidence from chapter three and the literature suggests that CPuORFs stall the ribosome during translation (Tables 1.4-6, Figures 3.2-7) (6,7,110). The average length of the peptide tunnel is approximately 40 aa (148,149). Furthermore, *in vitro* and *in planta* data from chapter 3 suggest that the full length CPuORF peptide is required to stall the ribosome (Figure 3.9-20). Evidence from *Arabidopsis* evidence suggests that the full length CPuORF peptide is important, therefore, we asked if CPuORF length is consistent across taxa (Figure 5.3). CPuORF length in amino acids in *A. thaliana, D. melanogaster* and *H. sapiens* was sourced from the eukaryotic CPuORF database (Figure 5.1). Peptide length data of 106,462 *Arabidopsis* regular-uORFs was collected from the uORFLIGHT database (5). Human and fly uORF data were taken from Ensembl (https://www.ensembl.org).

CPuORF length data across taxa reveals that *A. thaliana* CPuORFs are statistically similar in length to *D. melanogaster* and humans (Figure 5.3, Table 5.1). They have a mean length of 41.52, 69.24 and 22.78 aa, respectively. D. melanogaster CPuORFs however are significantly longer than human CPuORFs. For regular uORFs, flies and *Arabidopsis* regular uORFs are statistically similar with average lengths of 36.77 and 15.8 aa, respectively. On the other hand, human regular uORFs are significantly longer with an average length of 51.58 aa.

CPuORF length data is presented as a boxen plot to clearly to visualise data variability of large datasets (as is the case for regular uORF data) (Figure 5.3). The width of each boxen reflects the interquartile range of the level, indicating the variability of the data. Interestingly, there is relatively high variability in *D. melanogaster* CPuORFs and this isn't reflected in plants or human CPuORFs. As expected, due to the large sample size, regular uORF data in humans is substantial. Unexpectedly, despite the large sample size, plant and fly uORFs do not show variability to the same extent. This variability is also reflected in the standard deviation values in Table 5.1. These differences between regular-uORF lengths among *Arabidopsis, Drosophila*, and humans could potentially be influenced by distinct aspects of gene organization within each taxa. For example, in humans gene organization is complex, featuring intricate regulatory elements and a higher prevalence of alternative splicing (2).

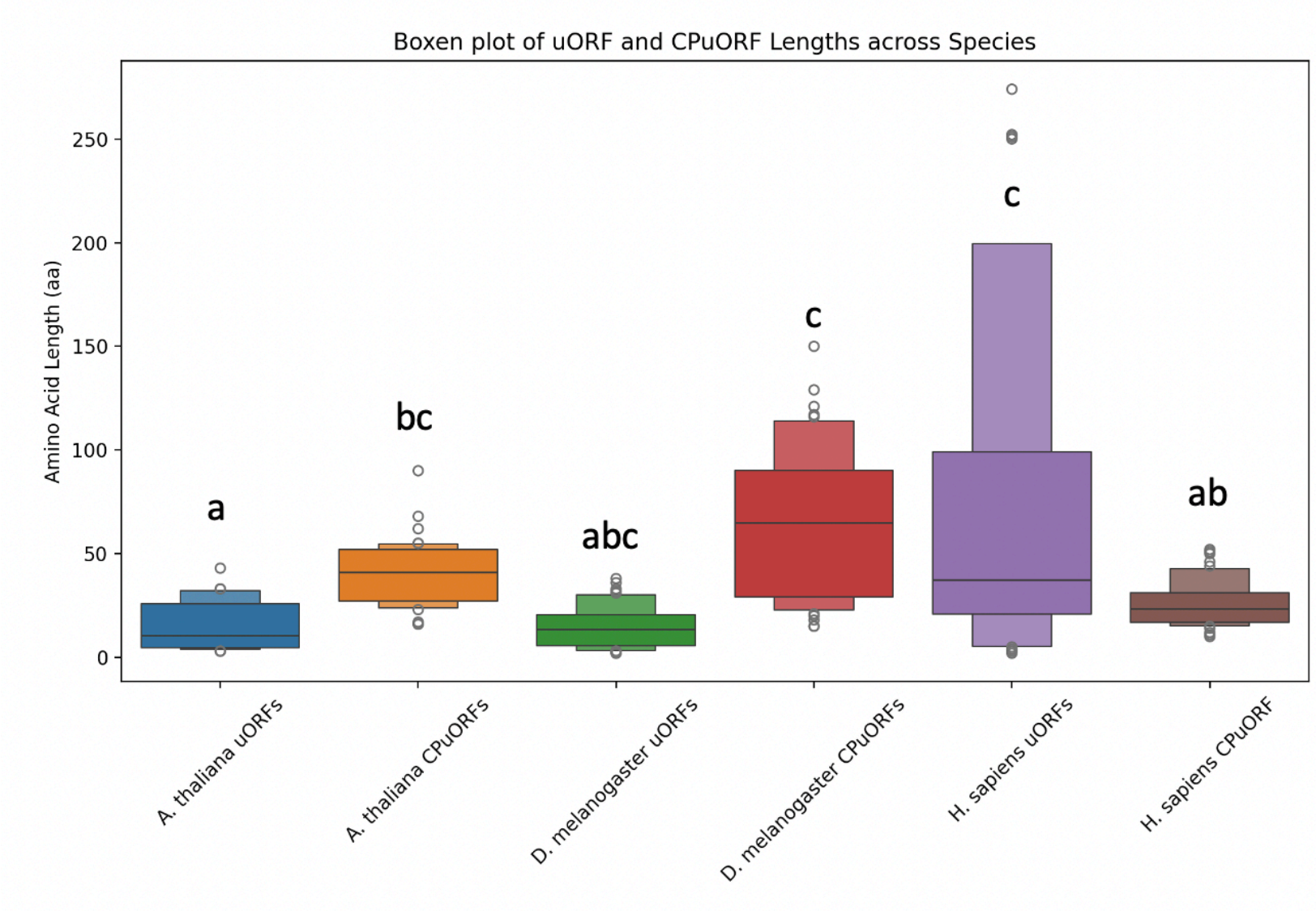

Boxen plot of uORF and CPuORF Lengths across Species
Amino Acid Length (aa)
0
50
100
150
200
250
a
bc
abc
c
c
ab
A. thaliana uORFs
A. thaliana CPuORFs
D. melanogaster uORFs
D. melanogaster CPuORFs
H. sapiens uORFs
H. sapiens CPuORF

Figure 5.3: Boxen Plot Illustrating the Length Distribution of CPuORFs and regular-uORFs Across Species. This boxen plot represents the length distribution of conserved upstream open reading frames (CPuORFs) and upstream open reading frames (uORFs) in *Arabidopsis thaliana* (blue), *Drosophila melanogaster* (orange), and *Homo sapiens* (purple). The plot provides a detailed visualization of the distribution, with each shaded box indicating the spread of the data between percentiles in a stepwise fashion, allowing for better visualization of larger datasets. The median lengths are marked by the lines within the boxes. Outliers are depicted as individual circles outside of the whiskers. Statistical groupings are denoted by letters 'a', 'b', and 'c', indicating significant differences in CPuORF lengths with group 'a' representing the shortest and group 'c' the longest. The groupings were calculated using an One-way ANOVA (ANalysis Of VAriance) with post-hoc Tukey HSD (Honestly Significant Difference) Test. The data for CPuORF and regular uORF lengths in *Arabidopsis* were sourced from the eukaryotic CPuORF database and uORFLIGHT database. Regular uORF data for humans and flies were extracted from Ensembl.

	mean	n	SD
A. thaliana regular-uORFs	15.84	106,456	14.84
A. thaliana CPuORFs	41.52	133	18.51
D. melanogaster rregular-uORFs	36.77	1,048,576	73.60
D. melanogaster CPuORFs	69.24	56	37.68
H. sapiens regular-uORFs	51.58	2,422,998	86.08
H. sapiens CPuORFs	22.78	1092	10.34
Class I	46.74	38	23.01
Class IIa	37.33	15	18.64
Class IIb	45.04	80	27.57

Table 5.1: Summary Statistics of uORF and CPuORF Lengths Across Species. This table presents a summary of the mean lengths and standard deviations (SD) for upstream open reading frames (uORFs) and conserved upstream open reading frames (CPuORFs) in *Arabidopsis thaliana*, *Drosophila melanogaster*, and *Homo sapiens*. Additionally, the table includes data for Class I, Class IIa, and Class IIb *Arabidopsis thaliana* CPuORFs. The lengths are measured in amino acids, and the values are derived from the eukaryotic CPuORF database and the uORFLIGHT database for *Arabidopsis*, and from Ensembl for human and fly data. Sample size is indicated by n.

Overall, peptide length in CPuORFs varies across taxa suggesting that CPuORF peptide length is not a conserved feature across taxa. Moreover, despite plant CPuORFs having

an average length that extends the exit tunnel, this isn't reflected in flies or humans. This data does have some interesting findings such as the relative uniformity of plant a fly uORF length when compared to humans.

CPuORF peptide length data does not suggest that CPuORF length is conserved across taxa (Figure 5.3, Table 5.1). However, data in chapter 3 highlights the importance of a full length CPuORF peptide in ribosome stalling (Figures 3.7-18). Previously, Causier et al., correlated CPuORF conservation with ribosome occupancy and suggested C-terminally conserved CPuORFs (Class I and IIb) stall during translation termination whereas, N-terminally conserved class IIa stalls during elongation (3). Therefore, I investigated CPuORF length across CPuORF class.

Within *Arabidopsis thaliana*, Class I CPuORFs are the longest with a mean length of 46.74 aa (SD of 23.01), followed by Class IIb (mean length of 45.04 aa, SD of 27.57), and Class IIa (mean length of 37.33 aa, SD of 18.64) (Figure 5.4). Despite this, all groups are statistically similar in length.

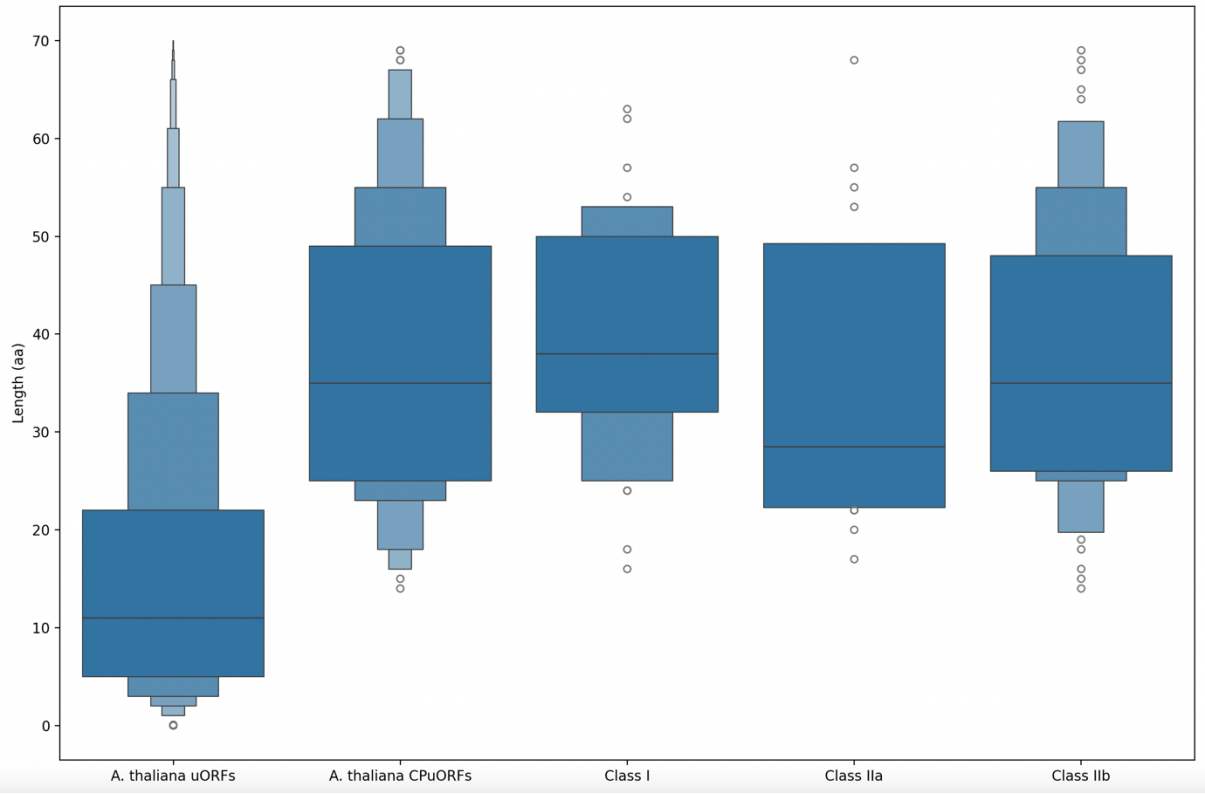


Figure 5.4: Boxenplot of Peptide Lengths for regular-uORFs and CPuORFs in *Arabidopsis thaliana* and Classified CPuORFs. This boxenplot illustrates the distribution of peptide lengths for upstream open reading frames (uORFs) and conserved upstream open reading frames (CPuORFs) in *Arabidopsis thaliana*, along with classified CPuORF lengths (Class I, Class IIa, and Class IIb). Each box shows the interquartile range (IQR) of peptide lengths, with the median indicated by the horizontal line within the box. Outliers are shown as individual circles. The lengths are measured in amino acids (aa).

5.2.3. Human CPuORFs are 10-fold more common that *Arabidopsis thaliana* and Human CPuORFs.

In 2021 around 0.5 % of *Arabidopsis* transcripts were reported to contain at least one CPuORF (3). Therefore, this thesis asked the question if CPuORFs occur at the same rate across taxa (Table 5.1). Rate of CPuORF containing transcripts was calculated as a percentage (Table 5.2). Data was extracted from the eukaryotic CPuORF database (Figure 5.1) and Ensembl (https://www.ensembl.org). The data showed a notable difference in CPuORF occurrence between *H. sapiens* and *Arabidopsis* and *Drosophila*. Human

CPuORFs occur at a rate of 5.2 % whereas, CPuORFs occur at a rate of 0.51 and 0.41 % in *Arabidopsis* and *Drosophila* respectively. The variation in CPuORF prevalence between humans, *Arabidopsis,* and *Drosophila* could be partly explained by differences in their gene structures, particularly the 5' leader sequences (2). In humans, 5' untranslated regions (5' UTRs) are typically longer and more complex than in *Arabidopsis* and *Drosophila*. Furthermore, when considered alongside the prevalence of alternative splicing events in humans collectively this could suggest there are more opportunities for the occurrence of CPuORFs (2).

The substantial difference between CPuORF prevalence could suggest potential differences in regulatory roles and importance of CPuORFs in humans when compared to *Arabidopsis* and *Drosophila*.

	#CPuORFs	#genes	%
Arabidopsis thaliana	133	25500	0.51
Drosophila melanogaster	56	13601	0.41
Homo sapiens	1092	21,000	5.2

Table 5.2. **A table of CPuORF rarity to show Human CPuORFs are 10x more common than *Drosophila* and *Arabidopsis* CPuORFs.** Data was collected from the eukaryotic CPuORF database (Figure 5.1).

5.2.4. Analysis of CPuORF sequences failed to identify functional components.

Evidence suggests that the conserved CPuORF amino acid sequences confer ribosome stalling (Table 1.4-6, Figures 3.15-18). Therefore, I compared average occurrence for each amino acid between *A. thaliana* CPuORFs and regular uORFs. Data was extracted from the eukaryotic CPuORF database and uORFLIGHT (Figure 5.1) (5) Moreover, Cryo-EM data of the bZIP11 CPuORF suggests that sucrose binding to the exit tunnel metabolite pocket alters the biochemistry of the exit tunnel (110). Other stalling mechanisms suggest that electrostatic potential could play a role in stalling (148). Therefore, data comparing amino acid occurrence could illuminate the CPuORF mechanism.

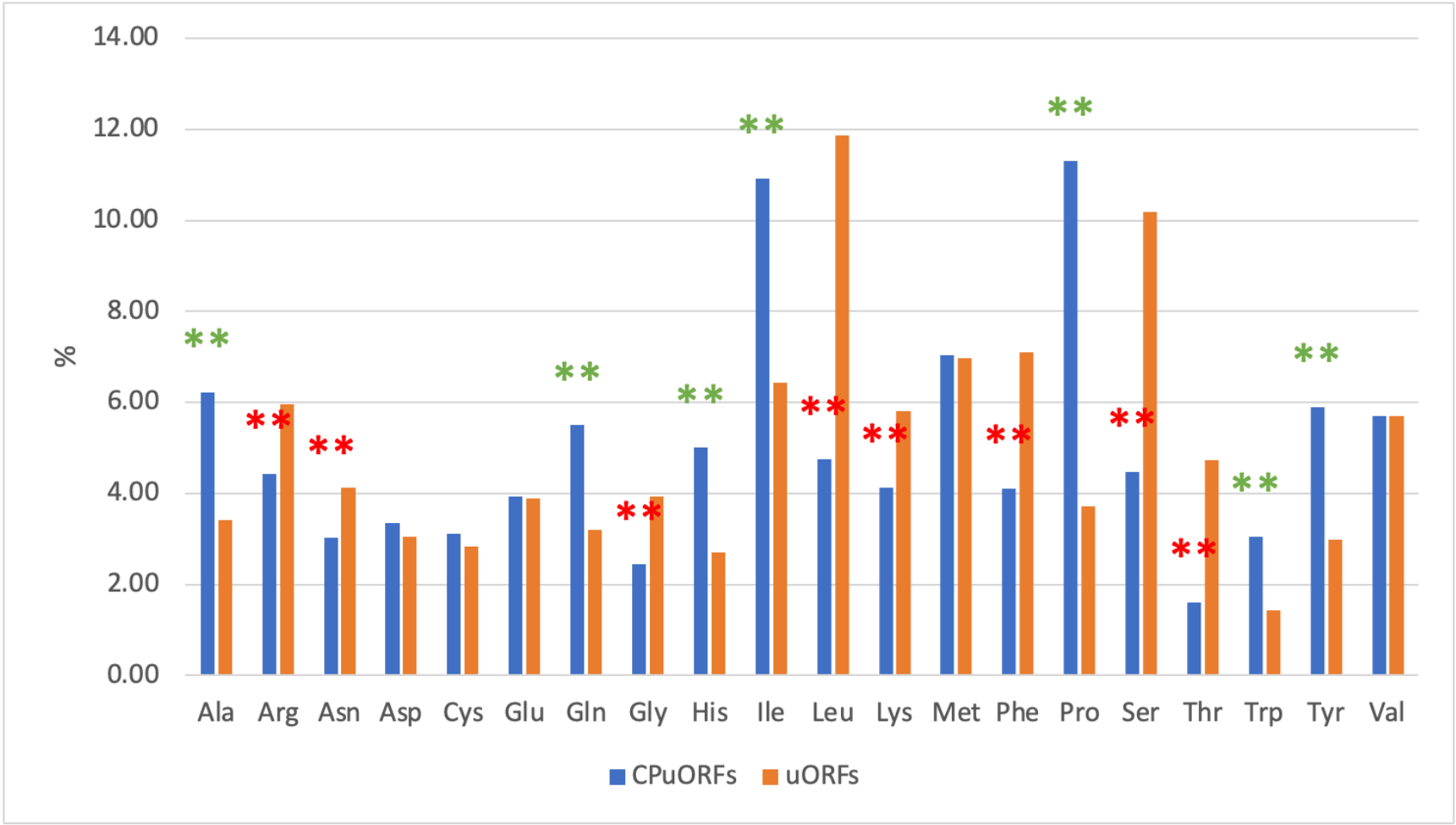

14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
%
Ala Arg Asn Asp Cys Glu Gln Gly His Ile Leu Lys Met Phe Pro Ser Thr Trp Tyr Val
CPuORFs
uORFs

***Figure 5.5*. Bar chart representing the percentage occurrence of each amino acid in conserved upstream open reading frames (CPuORFs) and upstream open reading frames (uORFs) for *Arabidopsis thaliana*.** The CPuORF data is shown in blue, while the regular uORF data is depicted in orange. Statistically significant differences between CPuORFs and regular-uORFs are indicated by asterisks above the bars, with green asterisks denoting a higher percentage in CPuORFs and red asterisks a higher percentage in regular-uORFs. Data for amino acid occurrence was calculated based on the total counts from the eukaryotic CPuORF database and uORFLIGHT database. Statistical significance was determined using Fisher's exact test.

A. thaliana CPuORFs have statistically higher rates of occurrence for alanine, glutamine, histidine, Isoleucine, proline, tryptophan, and tyrosine when compared to regular uORFs (Figure 5.5). These amino acids occur at a fold increase of 1.8, 1.7, 2.1, 1.7, 3, 2.1, and 2.0, respectively. In contrast, arginine, asparagine, glycine, leucine, lysine, phenylalanine, serine, and threonine statistically occur at lower rates in CPuORFs when compared to regular uORFs. These amino acids occur at a fold decrease of 0.7, 0.7, 0.6, 0.4, 0.7, 0.6, 0.4 and 0.3, respectively. These results may indicate that the amino acids that are occur at higher rates in plant CPuORFs may have been selected for their role in ribosome stalling. To test the significance of these amino acids in ribosome stalling molecular and structural biology should be employed.

Further investigation was carried out via LOGO consensus sequence analysis on *Arabidopsis* CPuORFs (Supplementary file 1). The analysis explored differences between CPuORF class and aligned from both the C and N-termini to account for potential alignments in the exit tunnel during translation. Ultimately, no clear consensus sequence emerged from the analysis. A similar outcome occurred when investigating peptide charge (Supplementary file 1).

Lastly, Codon usage has been shown to contribute to translation efficiency and ribosome stalling (193,198,199). Therefore, codon usage of *Arabidopsis* CPuORFs was compared to the preferential codon usage in *Arabidopsis* (199). Relative Synonymous Codon Usage (RCSU) values are utilized to measure the frequency of specific codons used for a given amino acid relative to the expected frequency if all synonymous codons for that amino acid were used equally. A RCSU score of 1.6 and 0.6 would indicate the overrepresentation or underrepresentation of a specific codon. CPuORF Codon usage data was sourced from the eukaryotic CPuORF database and the average *A. thaliana* RCSU values were sourced from O'Connell, M.J., Doyle, A.M., Juenger, T.E. *et al*, (Figure 5.1) (200).

Upon investigation, only the leucine amino acid demonstrated a different preferential pattern in CPuORFs when compared to the *Arabidopsis* average. Specifically, the CTC codon was overrepresented in CPuORFs with an RCSU value significantly above 1.6, while the CTA codon was underrepresented with an RCSU value below 0.6 (Table 5.3). This pattern was distinct when compared to the broader *Arabidopsis* genome, which does not exhibit a similar bias for these codons in coding for leucine.

	Leucine					
	CTA	CTC	CTG	CTT	TTA	TTG
CPuORFs	0.32	1.68	0.62	1.48	0.72	1.17
Arabidopsis	0.63	1.03	0.63	1.55	0.81	1.34

Table 5.3. A table to show that *Arabidopsis* CPuORFs have an over representation of CTC codons for leucine. Relative synonymous codon usage (RSCU) of *Arabidopsis* CPuORF sequences were compared to the standard Arabidopsis RCSU. RSCU values higher than 1.6 and lower than 0.6 indicate overrepresented and underrepresented codons, respectively. Overrepresented and underrepresented codons are highlighted in green and red, respectively. Of all codons and amino acids investigated the CPuORF leucine codons CTA and CTC are significantly over or underrepresented when the *Arabidopsis* average is not.

Notably, amino acid composition analysis found that leucine (L) is underrepresented in A. *thaliana* CPuORFs when compared to regular uORFs (Figure 5.5). Despite these collective results, the various bioinformatic tools applied to investigate CPuORF sequences failed to provide conclusive insights into CPuORF functionality.

5.3 Discussion

This chapter aimed to establish a comprehensive eukaryotic CPuORF database and investigate CPuORF DNA and peptide sequences to elucidate CPuORF function (Figures 5.1-6, Tables 5.1-2). Furthermore, this study used bioinformatic tools to investigate CPuORF DNA and peptide sequences, but these techniques and findings were not informative with respect to function (Figures 5.6, Table 5.2). This study did establish that CPuORFs are found at a higher rate in humans than those in *Arabidopsis* and *Drosophila* (Table 5.1), suggesting that CPuORFs have an expanded role in Humans.

5.3.1 The eukaryotic CPuORF database.

This thesis aimed to create a eukaryotic CPuORF database to facilitate research and provide a central repository for conditional regulatory uORF peptides. The efforts by uORFlight and Causier et al., to compile a CPuORF database provided a foundation for further expansion (3,192). While these databases encompassed monocots and dicots, the eukaryotic CPuORF database presented in this thesis extends the range to a broader variety of taxa (Figure 5.1). The eukaryotic database logs 57 and 1072 CPuORFs in *Drosophila melanogaster* and *Homo sapiens* respectively. Furthermore, the eukaryotic CPuORF database has included information on CPuORF class, CPuORF homology group, amino acid sequence and associated mORFs for *Drosophila melanogaster* and *Homo sapiens*. Consolidation of this information could prove to be a useful resource to the scientific community to facilitate bioinformatic, comparative and functional studies.

Alongside the eukaryotic CPuORF database is the first standardized R pipeline to classify CPuORFs based upon the location of their conserved amino acids (Figure 5.2). The three CPuORF classes, class I, IIa and IIb were introduced by Causier et al., (3). Each class reflects different conservation patterns across the CPuORF. The availability of a R pipeline will assist further CPuORF classification. Causier et al., have provided evidence to show that the ribosome stalls during termination in CPuORFs with C-terminal conservation (3). Conversely, CPuORFs with exclusively N-terminal conservation stall during elongation (3). The classification pipeline could potentially identify specific CPuORFs to investigate gene regulation in different phases of translation.

The eukaryotic CPuORF database builds upon the *Arabidopsis* CPuORF database by recording literature that provides evidence for the mode of CPuORF function (activator/repressive CUTS). The CUTS mechanism has potential applications in biotechnology and agriculture, making this database valuable for research. *Arabidopsis* CPuORFs respond to a range of signals, from small metabolites to environmental conditions and pathogen infection (3,8–11,125,201,202). The database provides a centralised place to amalgamate information on responsive CPuORFs. Currently, CPuORFs that conditionally express their mORF (by CUTS) are only characterised *in Arabidopsis*. Previous studies have shown that CPuORFs can be utilised in crop optimization (9). This database details 23 CPuORFs and their signals that could be further investigated in crop improvement and biotechnology. As more research is done on human and fly CPuORFs, and are shown to conditionally stall the ribosome, this database should be updated.

However, the eukaryotic CPuORF database is currently limited in its use of ribosome profiling data and lacking in confirmed sequence-dependent mORF attenuation data (7,106). The utility of the database would be significantly enhanced by incorporating ribosome profiling data, which offers a high-resolution view of ribosome occupancy on mRNAs and can be a powerful indicator of CPuORF activity. Additionally, documented instances of sequence-dependent mORF attenuation from mutant studies provide a direct link between CPuORF sequences and their functional consequences (7,106). Finally, the inclusion of CPuORFs from crop species would enhance the database's utility for crop improvement research.

5.3.2. CPuORF length diverges across taxa suggesting it is not a conserved feature of CPuORFs.

This thesis aimed to investigate CPuORF peptide properties that may confer function. Since the average *Arabidopsis* CPuORFs is approximately the same length as the ribosome peptide exit tunnel, this thesis asked if this was a conserved feature across taxa. The ribosomal exit tunnel can fit proteins from 30 to 70 amino acids depending on peptide structure (36). Ergo, the average CPuORF length in all three taxa would fit inside the exit tunnel (Figure 5.5). Taken together, CPuORF length may be an important feature in CPuORF function and potentially conserved.

When comparing average lengths of CPuORFs to regular-uORFs across taxa, plant and fly CPuORFs are longer than regular uORFs and human CPuORFs are shorter than regular uORFs (Figure 5.3-4). Furthermore, *Arabidopsis* CPuORFs are statistically similar to fly and human CPuORFs but fly CPuORFs are statistically longer than human CPuORFs. Overall, there isn't statistical uniformity between CPuORFs across the taxa studied here, and therefore, suggesting CPuORF length is not a conserved feature.

Plant CPuORFs conditionally stall the ribosome to modulate mORF expression (3,8–11,107,108,113,126,127,201,202). However, it remains uncertain whether a specific peptide length is required for ribosome stalling. For instance, human CPuORFs as short as 11 amino acids can attenuate mORF expression in a sequence-dependent manner (106). This is surprising considering known stalling peptides are much longer (*Arabidopsis* CPuORFs) (36). This might indicate that stalling can occur even when the

nascent protein is only partially within the exit tunnel upstream from the E-site. This is supported by in planta data in chapter 3. Truncated CPuORFs that have deleted the variable regions of the HEAT and eIF5 CPuORFs maintained similar mORF levels as the wild type (Figures 3.16-18). However, it is unclear if these small human CPuORFs can conditionally stall the ribosome as the truncated plant CPuORFs had lost this aspect of the CUTS mechanism.

When considering different classes of *Arabidopsis* CPuORFs, the data shows that all three are of a similar length to the average *Arabidopsis* CPuORFs (44 amino acids) (Figure 5.4). Classes with C-terminal conservation (Class I and IIb) are slightly longer, while CPuORFs with exclusively N-terminal conservation (Class IIa) are the shortest. Interestingly, ribosomal stalling at translation elongation is observed in Class IIa, while C-terminal-conserved classes exhibit stalling at translation termination (3). While these differences exist, the mean lengths are not statistically different enough to imply distinct mechanistic functions based on length alone.

5.3.3. Greater annotation of the human genome may result in the identification of more human CPuORFs.

Plant CPuORFs are very rare and this thesis was interested in determining whether conserved uORF peptides are found as infrequently across taxa. Interestingly, data from the eukaryotic CPuORF database suggests that human CPuORFs occur are ten times more frequent than *in D. melanogaster* or *A. thaliana (*Table 5.2). This result is surprising as the CPuORF transcript data was generated using the same pipelines (ESUCA) (105,106). The higher prevalence of CPuORFs in the human genome could be attributed

to several factors unique to human genomic organization and alternative splicing (2). These genomic features, coupled with more in-depth and focused genomic annotations driven by medical research interests, likely contribute to the apparent abundance of CPuORFs observed in human transcripts.

5.3.4. Bioinformatic tools do not identify functional domains of CPuORF sequences.

Evidence suggests that the conserved peptide sequences confer CPuORF function. Therefore, this thesis investigated CPuORF amino acid and DNA sequences in an attempt to find functional motifs for ribosome stalling. *Arabidopsis* and human CPuORFs have been observed to attenuate mORF expression in a sequence-dependent manner (7,106,108). *Arabidopsis* CPuORFs are noted for their ability to conditionally modulate mORF expression (3). These factors suggest that CPuORF peptides across eukaryotes may have the ability to stall the ribosome during translation. Bioinformatic tools were utilised to explore amino acid composition, codon usage, protein charge and consensus sequence analysis (Figures 5.3-4, Supplementary file 1). Investigations into amino acid occurance between CPuORFs and regular uORFs showed higher rates for alanine, glutamine, histidine, Isoleucine, proline, tryptophan, and tyrosine (Figure 5.5). On the other hand, the data observed significant lower rates of arginine, asparagine, glycine, leucine, lysine, phenylalanine, serine, and threonine in plant CPuORFs. Despite this, this data isn't useful without experimental data.

During these investigations, CPuORF class and alignments from the N- and C-termini were considered, to account for potential alignments in the ribosomal exit tunnel during

ribosome stalling. Codon usage has been reported to affect translation efficiency (184,198,199). This study identified a specific leucine codon that is overrepresented in CPuORF peptide sequences compared to the *Arabidopsis* average (Table 5.3). This could indicate that this specific codon confers stalling and should be further investigated. Despite these investigations, no sequences or amino acids of interest were identified.

The overrepresentation of a specific leucine codon in CPuORF sequences, compared to the *Arabidopsis* average, merits further investigation, particularly considering the unique properties of leucine (Table 5.3). Over-representation of leucine could indicate that its hydrophobic properties may affect translational dynamics (184,198,199). Consequently, changes in peptide biochemistry can influence nascent chain and translational kinetics. For example, leucine's hydrophobic nature may enhance interactions with the hydrophobic regions within the tunnel and nascent CPuORF peptide (184,198,199). To further elucidate the CPuORF stalling mechanism(s), a comparative analysis of CPuORF-stalled ribosome cryo-EM structures could provide further insights (111,139,140). Additionally, a CPuORF-wide alanine scanning mutagenesis approach could be deployed to identify key stalling amino acids, focusing particularly on leucine and its impact on translation. Discovering specific stalling domains and understanding how they leverage the properties of amino acids like leucine could be utilized in biotechnology to create bespoke regulatory proteins.

As investigations into finding a CPuORF stalling sequence failed, it is unlikely there is a universal or common CPuORF peptide sequence that stalls the ribosome. It is more likely

that there are multiple CPuORF peptide conformations that can stall the ribosome. To further elucidate the CPuORF stalling mechanism(s), a comparative analysis of CPuORF-stalled ribosome cryo-EM structures could provide further insights (111,139,140). A CPuORF-wide alanine scanning mutagenesis approach could be deployed to identify key stalling amino acids. Finding stalling domains could be utilised in biotechnology to create bespoke regulatory proteins.

Chapter 6. General discussion

Several plant CPuORFs have been shown to conditionally modulate downstream mORF translation *in vitro* and *in vivo* (3,6,7,106). Previous studies into the CPuORF mechanism have functionally characterised CPuORF function by placing a CPuORF upstream of a reporter mORF like LUCIFERASE (3,6,7,106). Such reporter gene assays can quantify the effect of mutations to the CPuORF amino acid sequence on the translation of downstream mORFs under different conditions. The mechanism this thesis has termed "Conditional uORF Translational Stalling (CUTS)" describes how CPuORFs conditionally stall the ribosome during translation, attenuating mORF expression via repressive CUTS (rCUTS) or activator CUTS (aCUTS) (Figure 1.20) (3). Recently, cryo-EM data of the rCUTS bZIP11 CPuORF partially elucidated its signal sensing and ribosome stalling mechanism (110). This thesis aimed to elucidate the CPuORF aCUTS mechanism through luciferase assays, cryo-EM and bioinformatics.

This thesis, through luciferase assays 1) has identified the functional CPuORFs in the SAC51 transcript (Figure 3.1). 2) Showed that CPuORFs are self-contained regulatory peptides that respond specifically to signals (Figure 3.4-5). 3) Provided evidence that there is a common mechanistic framework between four aCUTS CPuORFs (Figure 3.7-8). 4) Demonstrated the importance of the full length CPuORF peptide i*n vivo* and *in vitro* (Figures 3.9-20). 5) Presented the first low resolution A. thaliana ribosome and developed two methods to isolate ribosomes for cryo-EM analysis (Figures 4.1-14). 6) Published the most comprehensive *Arabidopsis* CPuORF database as part of Causier., et al (3) and made publicly available an expanded eukaryotic CPuORF database (Figure 5.1).

6.1. CPuORFs 38 and 40 are the functional SAC51 CPuORFs.

The Davies group has previously investigated the SAC51 transcript and its three CPuORFs (CPuORFs 38, 39 and 40) (Figure 3.1) (3,37,118). It has been reported by us and in the literature that the SAC51 CPuORFs have aCUTS activity and are responsive to thermospermine (3). By leveraging reporter assays, chapter 3 established that CPuORFs 38 and 40 independently exhibit thermospermine-responsive aCUTS activity, a novel finding that extends our comprehension of CPuORF-mediated gene regulation in the SAC51 transcript (Figure 3.1) (3,37,118).

A genetic screen in a thermospermine synthase mutant (*acl5-1*) identified a suppressor mutation (early STOP codon) in the third CPuORF in the SAC51 transcript (CPuORF 40) (37,118). The STOP codon introduced in CPuORF40 partially rescued the *acl5-1* mutant phenotype. CPuORF 40 is in the 5'leader of the SAC51 (SUPPRESSOR OF AUCULIS 5 1) transcript and the SAC51 mORF protein inhibits the expression of ACL5 (37,118). It has been suggested that the mutation in CPuORF 40 rescued the *acl5-1* mutant phenotype as it alleviated the stalling mechanism of CPuORF, and therefore, its attenuation of the SAC51 mORF (37,118). Despite this evidence, it is unclear how the three CPuORFs in the SAC51 transcript contribute to the thermospermine responsive aCUTS mechanism (37,118).

To determine the functional SAC51 CPuORFs, constructs were made that placed single and double combinations of the SAC51 CPuORFs upstream of a luciferase reporter gene (Figure 3.1). The effect that these CPuORFs had on mORF expression was quantified in

control and elevated thermospermine conditions (Figure 3.1). The luciferase levels of these mutant constructs were compared to the wild type SAC51 5' leader and a luciferase mORF that was not under the control of a CPuORF.

The results showed that CPuORFs 38 and 40 retained their reported thermospermine responsive aCUTS activity independently (Figure 3.1). Consequently, CPuORF 40 was chosen for further analysis as it has been shown to be functional, unlike CPuORF 38 it retained stalling to similar levels of the native SAC51 5'leader control and because it was identified in a genetic screen (37,118). Moreover, this is the first data demonstrating that these CPuORFs retain their thermospermine aCUTS function in an artificial 5'leader.

The differential responsiveness of CPuORFs 38 and 40 to thermospermine, compared to CPuORF 39, warrants further investigation into the SAC51 CPuORFs (Figure 3.1). Future studies could explore how altering CPuORF positions within the SAC51 5'leader affects their regulatory function. The results in chapter 3 have identified that CPuORF 38 and 40 are independently function however, it is still unclear why three CPuORFs were conserved in frame and upstream of the SAC51 mORF. Moreover, it is still unclear how these three CPuORFs collectively function to modulate mORF translation.

6.2. CPuORFs are peptide switches that conditionally fine tune gene regulation, development, and metabolism.

CPuORF 40 and three other aCUTS CPuORFs (eIF5, HEAT and ROJ) were investigated in an in vitro cell free extract outside of their native 5'leaders, in a common context (Figure 3.4). The effect of the wild type CPuORF sequence on downstream LUCIFERASE

translation was measured and compared to that of frameshifted variants. This showed that frameshift mutations led to an increase in mORF translation for all four CPuORFs (Figure 3.4).

These results show that frameshift CPuORFs are less successful in attenuating luciferase activity than their wild type counterparts (Figure 3.4). The effect of frameshift mutations on mORF translation is similar in nature in experiments that have compared wild type CPuORFs and ΔAUG variants (3). Mutating the start codon prevents the translation of the CPuORF, resulting in higher levels of mORF translation as observed in frameshift mutants (3). Investigations into the modulation of mORF translation by wild type and ΔAUG variants was useful comparing the effects of wild type CPuORFs to frameshifted variants is a better control (3). Regular uORFs do attenuate mORF translation as they sequester ribosomes from the mORF therefore, comparing wild type and ΔAUG CPuORFs the study doesn't compare the effect of CPuORF-ribosome stalling on mORF expression (2). In fact, this trend would be replicated if the study compared a wild type regular uORF to a ΔAUG regular uORF variant. Therefore, in chapter 3 a better control to study the effect of the CPuORF peptide on mORF expression is to compare it to a frameshifted version that has the same length, different peptide sequence and acts as a 'regular uORF' (Figure 3.4)

Overall, these results (alongside Causier et al.,) show that the CPuORF attenuation of mORF translation is conferred by the peptide sequence and not the mRNA as has been reported for regular uORFs (2,3). After demonstrating that CPuORFs are self-contained and attenuate mORF translation in a sequence dependent manner, I moved to test signal responsiveness *in vivo* (Figure 3.5) (3).

As part of Causier et al., we demonstrated that CPuORFs, in their native contexts, respond specifically to their signals (3). Using five aCUTS CPuORFs, one rCUTS CPuORF and a CPuORF that has been reported to function via aCUTS and rCUTS, our data showed that these CPuORFs retain their specific signal response in our common backbone. A luciferase assay on T1 *Arabidopsis* leaves compared activity of wild type CPuORFs under elevated thermospermine, mannitol, heat shock and sucrose compared to control conditions (Figure 3.5). The CPuORFs investigated modulated mORF translation as reported in the literature. Collectively, this demonstrates that CPuORFs act as self-contained peptide-switches that conditionally modulate any downstream mORF.

CPuORFs have been identified through pipelines that select for uORF peptide sequences that have been conserved (4,105). Data presented in chapter 3 and in the literature show that several plant CPuORFs can conditionally stall the ribosome during translation (3,6,7) (Figures 3.2-3). Upon application of their respective signals, the CPuORFs found in the SAC51, eIF5, HEAT, ROJ, TBF1 and bZIP11 modulated mORF translation in response to a specific signal (Figure 3.5). The aCUTS CPuORFs tested (CPuORF 40, eIF5, HEAT, ROJ and TBF1), mORF translation increased by a fold change of 2.23, 2, 2.68, 3.2 and 3.19, respectively. On the otherhand, the rCUTS CPuORF bZIP11 under elevated sucrose conditions caused a decrease in mORF translation by 35% (Figure 3.5). This fold increase or decrease in mORF translation was only elicited by their respective and reported signals. The exception to this rule is the CPuORF in the ROJ transcript as elevated thermospermine caused a statistically significant increase in mORF translation by 22% (Figure 3.5). Beside this, all other CPuORFs tested saw a general decrease in mORF

translation when exposed to other signals (Figure 3.5). This reflects the data for the 35S:LUC control lines as all signal applications caused a reduction in LUC translation and suggesting this effect is likely due to stress causing a global decrease in translation (Figure 3.5) (2).

Moreover, the mORFs that the signals and CPuORFs regulate often function in the metabolism of the signal that the CPuORF is responsive to (12,110). For example, the SAC51 CPuORFs regulate an mORF (*SAC51)* that inhibits a thermospermine synthase protein (37) (Figures 1.13-18). Taken together this may suggest that some CPuORFs have been conserved to regulate networks associated with their respective signals. Jointly, GO data, conservation patterns and luciferase data suggest that the CPuORF peptide switch mechanism has been conserved to rapidly regulate and fine tune regulatory mORFs that function in developmental, signalling, and metabolic networks (Figures 3.2-3) (3).

6.3. A common mechanistic framework for stalling the ribosome during translation.

A previously published genetic screen implicated the SAC51 CPuORF 40 mechanism and specific ribosomal exit tunnel proteins (37). The screen revealed that a mutation in CPuORF 40 disrupts the SAC51 ribosome stalling mechanism, a finding that is supported by experiments described here (Figure 3.1). The same genetic screen also showed that single amino acid substitutions in uL4z, uL16z (uL4z^{G75R} and uL16z^{G14S}) and the creation of a premature stop codon in RACK1z (RACK1^{W261STOP}) disrupt ribosome stalling at the SAC51 CPuORFs (Figures 13-18). In the light of this evidence from the genetic screen,

efforts were taken to explore these ribosome mutations on ribosome stalling mediated by other CPuORFs (Figure 3.1).

The locations of the ribosome mutations were determined using publicly available cryo-EM data for the wheatgerm ribosome, showing that the mutations in uL4z and uL16z are found in the peptide exit tunnel (Figure 3.6-7). In contrast, the mutation in RACK1z is predicted to produce a truncated protein. Cryo-EM data suggests that the role of RACK1z is to stabilise stalled ribosomes that have collided on a transcript during translation (124). To test if these mutations affect ribosome stalling at CPuORF 40 and other a CUTS CPuORFs, *Arabidopsis* wild type plants and the three ribosome mutant lines were transformed to constitutively express CPuORF:LUC. Remarkably, a luciferase assay showed that ribosome stalling was reduced for all tested aCUTS CPuORFs (SAC51, eIF5, HEAT and ROJ) in all three mutant lines (Figures 3.6-7).

Overall, these findings suggest that there is common framework for CPuORF-mediated ribosome stalling (Figures 3.6-8). Encouragingly, the literature also supports these findings. Cryo-EM data of peptides that stall the ribosome during translation, but are not CPuORFs, shows peptide interactions with the same exit tunnel proteins (20,111,139). Furthermore, cryo-EM data elucidating the bZIP11 CPuORF mechanism exhibits similarities to the stalling and signal sensing mechanism of an arrest peptide from *E. coli* (110,135–137). Taking this together suggests that the mechanism of ribosome stalling is ancient in origin.

6.4. The full length CPuORF peptide is required for function.

Luciferase data utilising ribosome mutations provided convincing evidence that CPuORFs can stall the ribosome during translation (Figure 3.4-8). Furthermore, as this suggested a common framework to stall the ribosome, we wanted to explore CPuORF peptide sequences and their role in stalling and signal sensing. The initial hypothesis was that the CPuORF peptide was organized into different functional domains. This hypothesis was guided by the CPuORF classification system first devised by Hayden and Jorgensen and then adapted by Causier *et al.,* to include 3 CPuORF classes (Figure 1.12) (3,4). Causier *et al.,* utilised ribosome profiling data to correlate the position of ribosome stalling events with the CPuORF regions that contain conserved amino acids (3). Furthermore, there is evidence in the literature to show that mutating conserved amino acids in CPuORFs abolishes their ribosome stalling activity (109). Therefore, to test the functional associations of the conserved and variable amino acids, I took an alanine scanning approach to measure the effects of mutations on mORF translation and map their effects to the conserved and variable amino acids. *In vitro* data suggested that mutating four consecutive amino acids at any position in the CPuORF peptide abolishes ribosome stalling (Figures 3.7-14). Moreover, CPuORF variants that have had their variable regions deleted were unable to maintain stalling *in vitro*. As *in vitro* data did not show any difference in luciferase activity between mutating conserved and variable amino acids and the conserved region could not independently stall the ribosome, the data suggested that the full-length peptide is required for function.

In contrast to the *in vitro* data, *in vivo* data provided evidence that conserved amino acids have a more pronounced role in stalling the ribosome at two aCUTS CPuORFs (eIF5 and HEAT) (Figure 3.17-20). When tested *in vivo*, mutations in conserved amino acids in the CPuORFs of eIF5 (S6-Q9A and L14-E17A) and HEAT (V26-N29A and R34-R38) abolished ribosome stalling and led to an increase in mORF translation compared to their wild types (Figures 3.15-16).

Previously, in vitro, and in vivo experiments pointed to a modular organisation of CPuORF function (3,4). Specifically, the data suggested that conserved amino acids conferred stalling whereas, variable amino acids conferred signal sensing (3,4). With this model under control conditions, I anticipated that mutating conserved amino acids would increase mORF translation due to disrupted stalling, whereas mutations in the variable regions would not affect mORF translation as they confer signal sensing. Conversely, I expected that mutations in variable amino acids under signal treatment would abolish signal sensing and therefore, stalling would be maintained and there would be no increase in mORF translation. Surprisingly, *in vitro* data of all aCUTS CPuORFs tested indicated that the full length peptide contributed to stalling regardless of conservation (Figure 3.9-20).

However, when tested in vivo evidence suggested that certain conserved residues confer stalling when tested in *Arabidopsis* (Figure 3.15-20). These discrepancies likely stem from differences between the wheat germ *in vitro* system and transformant *Arabidopsis* lines. Unlike the wheat germ system, *Arabidopsis* contains a full array of cellular co-factors, auxiliary proteins, and specific ribosomal protein isoforms, as well as additional

regulatory elements and plant-specific post-translational modifications. These components are crucial for the intricate regulatory functions of CPuORFs and are absent in the wheat germ system. Moreover, the in vivo cellular environment of Arabidopsis provides a complex network of interacting signals and molecular pathways that significantly influence CPuORF activity. Overall, this highlights the limitations of investigating CPuORFs *in vitro*.

aCUTS CPuORFs can stall the ribosome and respond to signals (Figure 3.5). To test the effect of eIF5 and HEAT CPuORF peptide mutations on signal sensing luciferase assays were performed *in vivo* (Figures 3.17-20). The results suggested that some CPuORF peptide mutations did not knock out eIF5 and HEAT's ability to sense mannitol and heat shock, respectively. These mutations were Y18-I21A, V30-L33A and D38-S41A for eIF5 and I22-L25A in HEAT (Figures 3.17-20). These mutations in eIF5 that did not knock out aCUTS can be located in both the conserved and variable regions whereas, the mutations in heat is confined to the conserved region.

Additionally, we investigated mutant variants of the eIF5 and HEAT CPuORFs, where the variable amino acids were specifically removed. This resulted in truncated peptides that retained only the conserved amino acid sequences of each CPuORF. For eIF5, the deletions were made at the C-terminal end, whereas for HEAT, the deletions occurred at the N-terminal end (Figures 3.17-20). Under control conditions both truncated peptides elicited a statistically similar levels of LUC activity compared to their wild types. When tested under elevated mannitol and heat shock the truncated proteins did not cause an increase in mORF translation as observed in the wild type (Figures 3.17-20). As these

peptide sequence only contained eIF5 and HEAT's conserved regions, the data suggests that the conserved regions can stall the ribosome but have lost their aCUTS signal sensing activity (Figures 3.17-20) (3).

Collectively, the data suggests that the conserved region can stall the ribosome independently and this function could be described as modular (Figures 3.17-20). However, the data for signal sensing is less clear. The data suggests that the conserved region alone cannot respond to signals without the full-length peptide including the variable amino acids. However, alanine scanning data suggests that variable and conserved amino acids play a role in signal sensing.

6.5. The first low resolution model of an *Arabidopsis* ribosome and producing models of CPuORF-stalled ribosome.

Data from chapter 3 emphasised the advantages of investigating *Arabidopsis* CPuORF function using *Arabidopsis* ribosomes (Figures 3.7-20). Furthermore, although luciferase assays provided insights into CPuORF function they did not fully resolve the CUTS mechanism. Recently, cryo-EM data of translational arrest peptides and the plant bZIP11 CPuORF has been effective in elucidating individual stalling and signal sensing mechanisms and comparative analysis have identified similarities between stalling mechanisms (110). Therefore, subsequent efforts were made to produce high resolution models of the unresolved *Arabidopsis* ribosome and CPuORF-stalled ribosomes. In this thesis, I developed protocols to isolate ribosomes for cryo-EM from Arabidopsis tissue, by either centrifugation through a sucrose cushion or through immunoprecipitation (Figures 4.1-14). Initial experiments, involving isolating Arabidopsis ribosomes through a

sucrose cushion, provided the first low-resolution model of the Arabidopsis thaliana ribosome (Figure 4.9). However, recognizing the parallel success of another research group in producing a 2 Å resolution model, collaboration was deemed more efficient than duplicative efforts.

Producing a map of CPuORF-stalled ribosomes in *Arabidopsis* would require the full published map of the *Arabidopsis* ribosome that is yet to be made publicly available. Therefore, given that the stalling mechanisms of the *Arabidopsis* bZIP11 CPuORF, the fungal arginine attenuator peptide and human cytomegalovirus were elucidated by cryo-EM using a wheat germ extract, we aimed to produce a CPuORF-stalled ribosome map using wheat germ.

Wheat germ was chosen despite reservations of investigating *Arabidopsis* CPuORF function outside of *Arabidopsis* based on *in vitro* and *in vivo* luciferase data from chapter 3 (Figures 3.7-20). *In vitro* data for these CPuORFs (CPuORF 40, eIF5, HEAT and ROJ) suggests that wild type CPuORFs can stall in wheat germ (Figure 3.4). However, these efforts were unable to produce a stable sample of a CPuORF-stalled ribosome.

6.6. The eukaryotic CPuORF database.

As part of Causier et al., I generated the most comprehensive *Arabidopsis* CPuORF database (3). To build on this published database I expanded it to include CPuORFs from flies and humans to form the eukaryotic CPuORF database (Figure 5.1). This database can be used to find regulatory peptides to be used in biotechnology and agricultural

studies. Molecular investigations detailed in Chapter 3 demonstrated that the CPuORF mechanism is sequence dependent and there are commonalities in the stalling mechanisms between CPuORFs (Figures 5.2-6). Therefore, I set out to analyse CPuORF and regular uORF sequence data to find functional domains. Moreover, sequence length data across taxa suggested that CPuORF length was not a conserved CPuORF feature and that human CPuORFs are 10x more frequent that fly and plant CPuORFs (Figures 5.2-6).

6.7. An updated CUTS mechanism.

In chapter 1 this thesis presented the CUTS (Conditional uORF Translational Stalling) mechanism and utilised luciferase assays, cryo-EM, and bioinformatics to test the model and elucidate CPuORF function (Figure 6.1). Luciferase assays suggested that specific exit tunnel proteins, that have previously been implicated in other ribosome stalling mechanisms, potentially also function in four aCUTS CPuORFs (SAC51, eIF5, HEAT, ROJ) (Figure, 3.4-5) (139). These proteins form the constriction within the exit tunnel (uL4z) and the peptidyl transferase centre (uL16z) (148,149,179). Moreover, luciferase assays have provided evidence for the role of RACK1 in the CPuORF aCUTS mechanism (Figure 3.6-8). Four aCUTS CPuORFs were constitutively expressed in an *Arabidopsis* mutant RACK1 background and mORF translation was higher compared to the wild type background (Figure 3.6-8). Cryo-EM data in humans has demonstrated RACK1's role in stabilising ribosome stalling therefore, the data may suggest that RACK1 also stabilises stalled ribosome in the aCUTS CPuORF mechanisms (Figure 3.6-8*)*. As the data suggests that a mutant RACK1 results abolishes ribosome stalling in these aCUTS CPuORFs. Finally, *In*

vivo luciferase assays also suggested that a full length CPuORF peptide confers its ability to stall the ribosome and sense signals (Figure 3.7-20).

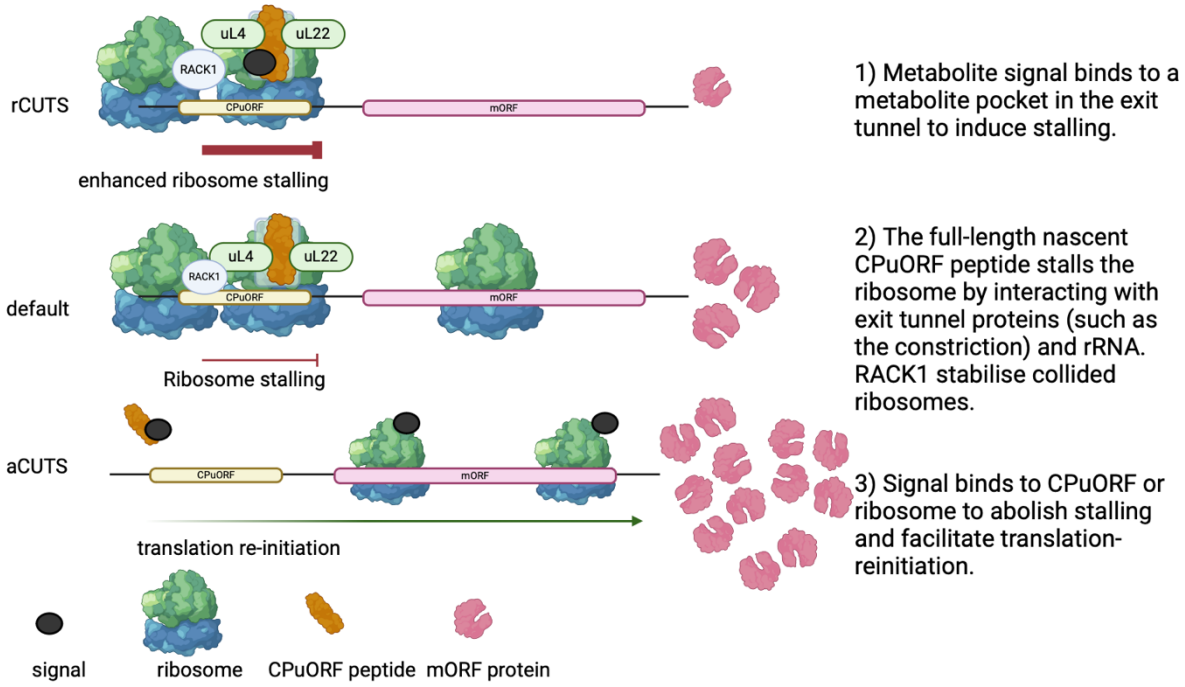


Figure 6.1. An updated model of CPuORF function. Repressive CUTS (rCUTS) illustrates when a specific signal enhances or triggers ribosomal stalling via a nascent CPuORF. Resulting in reduced mORF protein production compared to control conditions. Cryo-EM data suggests that the ribosome senses metabolites through a conserved pocket to stall the ribosome. Upon signal sensing, translation termination is disrupted. The default: A CPuORF nascent peptide inherently stalls the ribosome during translation independent of an external signal. Consequently, leading to decreased mORF protein production when compared to control conditions. Comparative cryo-EM data of arrest peptides suggest that this is facilitated by interactions with exit tunnel proteins, rRNA and disrupting PTC function. Finally, this thesis has demonstrated that RACK1 may contribute to ribosome stalling by stabilising collided ribosomes. Activator CUTS (aCUTS) demonstrates how a signal terminates ribosomal stalling by the nascent CPuORF peptide. Enabling the ribosome to re-initiate translation at the downstream mORF and promoting mORF protein production. The literature suggests that translation reinitiation occurs because of the retention of initiation factors, signals may facilitate the retention of initiation factors.

6.8. Future directions

Comparative cryo-EM study to elucidate common stalling mechanism.

Chapter 4 presents the first low-resolution model of an *Arabidopsis* ribosome, alongside attempts to solve the structure of a CPuORF-stalled ribosome using cryo-EM. The long-term aim of this project was to elucidate the interactions between a nascent CPuORF peptide and the ribosome. In chapter 3 I provided evidence that four aCUTS CPuORFs use a common framework to stall the ribosome (Figure 3.6-8). Alongside cryo-EM data of a rCUTS CPuORF that also shares commonalities with a bacterial translation arrest peptide, the evidence is convincing that there could be a common stalling mechanism (110,135–137).

A comparative cryo-EM study that generated high resolution CPuORF-stalled ribosomes using a diverse set of CPuORFs and arrest peptides could find common interactors in stalling. By selecting a diverse set of CPuORFs from varying classes, modes of CUTS, and signals to produce high-resolution cryo-EM models to elucidate their CUTS mechanisms. This will allow comparison of CPuORF peptide interactions with the ribosomal exit tunnel and enhance our understanding of the peptide-exit tunnel biochemistry and signal-induced stalling.

A genetic screen can identify functional residues and elucidate signal sensing mechanism.

Cryo-EM is best suited for elucidating ribosome stalling and rCUTS CPuORF mechanisms as aCUTS ribosomes are not stalled post signal application (Figure 6.1) (3). Moreover,

three aCUTS CPuORFs investigated in this thesis respond to abiotic signals and current examples of conditional stalling detail a metabolite pocket in the exit tunnel facilitates conditional stalling (Figure 3.4-5) (3). Therefore, how do CPuORFs respond to abiotic signals? Cryo-EM's limitations prompt us to consider a genetic screen approach. By constitutively expressing a CPuORF upstream of a GFP reporter in stable homozygous lines, and subjecting seedlings to EMS screens under elevated signal conditions, we can discern mutations impacting the CUTS mechanism. This approach would be particularly valuable for abiotic signal responsive CPuORFs, as a genetic screen could identify genes that have been upregulated under treatment, or it might identify parts of the exit tunnel or CPuORF that bind so a metabolite signal as is the case for the bZIP11 CPuORF (110). Furthermore, cryo-EM could identify translation factors that might be employed in the signal sensing and stalling mechanisms.

Ribosome profiling data can identify stalling proteins.

In chapter five, amino acid sequence data and bioinformatic tools were unable to provide insight into the mechanistic underpinnings of the CUTS mechanism. Causier et al., used ribosome profiling data to correlate ribosome stalling events with regions of CPuORFs (3). Ribosome profiling data and ribosome occupancy can be mapped to the codon level (3,115,143). Pipelines can be developed that analyse the sequence context around the stalling events to identify critical residues for stalling. Furthermore, pipelines to identify CPuORFs could screen for functional peptides by filtering for transcripts that have ribosome profiling data indicating that they stall the ribosome during translation (107). Or identify functional peptides regardless of conservation (107).

6.9. Concluding remarks.

In summary, this thesis has explored the mechanisms by which CPuORFs modulate gene expression in plants utilising molecular, structural and bioinformatic techniques. This thesis has highlighed the sequence-dependent nature of this regulation and the importance of the full-length CPuORF peptide. I have provided supporting evidence for a common framework underpinning the CPuORF-mediated translational stalling (CUTS) mechanism. In the absence of a resolved CPuORF-stalled ribosome structure, this thesis presents the first low-resolution model of an *Arabidopsis* ribosome and established a comprehensive eukaryotic CPuORF database.

7.References

1. **Akash., S, Fateha Arefine., M.** A short review on central dogma of molecular biology. IJAAR. Vol. 6, (2021).
2. **Merchante C, Stepanova AN, Alonso JM.** Translation regulation in plants: an interesting past, an exciting present and a promising future. Plant Journal. (2017).
3. **Causier B, Hopes T, McKay M, Paling Z, Davies B.** Plants utilise ancient, conserved peptide upstream open reading frames in stress-responsive translational regulation. Plant Cell Environ. (2022).
4. **Hayden CA, Jorgensen RA.** Identification of novel conserved peptide uORF homology groups in *Arabidopsis* and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. BMC Biol. (2007).
5. **Niu R, Zhou Y, Zhang Y, Mou R, Tang Z, Wang Z, et al.** UORFlight: A vehicle toward uORF-mediated translational regulation mechanisms in eukaryotes. Database. (2020).
6. **Ebina I, Takemoto-Tsutsumi M, Watanabe S, Koyama H, Endo Y, Kimata K, et al.** Identification of novel *Arabidopsis thaliana* upstream open reading frames that control expression of the main coding sequences in a peptide sequence-dependent manner. Nucleic Acids Res. (2015).
7. **Hayashi N, Sasaki S, Takahashi H, Yamashita Y, Naito S, Onouchi H.** Identification of *Arabidopsis thaliana* upstream open reading frames encoding peptide sequences that cause ribosomal arrest. Nucleic Acids Res. (2017).
8. **Ishitsuka S, Yamamoto M, Miyamoto M, Kuwashiro Y, Imai A, Motose H, et al.** Complexity and conservation of thermospermine-responsive uorfs of Sac51 family genes in angiosperms. Front Plant Sci. (2019).
9. **Pajerowska-Mukhtar KM, Wang W, Tada Y, Oka N, Tucker CL, Fonseca JP, et al.** The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. Current Biology. (2012).
10. **Uchiyama-Kadokura N, Murakami K, Takemoto M, Koyanagi N, Murota K, Naito S, et al.** Polyamine-responsive ribosomal arrest at the stop codon of an upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA decay in *Arabidopsis thaliana*. Plant Cell Physiol. (2014).

11. **Yamashita Y, Takamatsu S, Glasbrenner M, Becker T, Naito S, Beckmann R.** Sucrose sensing through nascent peptide-meditated ribosome stalling at the stop codon of *Arabidopsis* bZIP11 uORF2. FEBS Lett. (2017).

12. **Van Der Horst S, Filipovska T, Hanson J, Smeekens S.** Metabolite control of translation by conserved peptide uORFs: The ribosome as a metabolite multisensor. Plant Physiol. (2020).

13. **Xu G, Yuan M, Ai C, Liu L, Zhuang E, Karapetyan S, et al.** UORF-mediated translation allows engineered plant disease resistance without fitness costs. Nature. (2017).

14. **Cooper S.** The central dogma theory of molecular biology. CBIR, (1981).

15. **Li GW, Xie XS**. Central dogma at the single-molecule level in living cells. Nature. (2011).

16. **Hager GL, McNally JG, Misteli T.** Transcription Dynamics. Molecular Cell. (2009).

17. **Latchman DS.** Transcription Factors: An Overview. Int. J. Biochem. (1997).

18. **Soemedi R, Cygan KJ, Rhine CL, Glidden DT, Taggart AJ, Lin CL, et al.** The effects of structure on pre-mRNA processing and stability. Methods, Academic Press Inc. (2017).

19. **Browning KS, Bailey-Serres J.** Mechanism of Cytoplasmic mRNA Translation. Arabidopsis Book. (2015).

20. **Ito K, Chiba S.** Arrest peptides: Cis-acting modulators of translation. Annual Review of Biochemistry. (2013).

21. **Nelde A, Flötotto L, Jürgens L, Szymik L, Hubert E, Bauer J, et al.** Upstream open reading frames regulate translation of cancer-associated transcripts and encode HLA-presented immunogenic tumor antigens. Cellular and Molecular Life Sciences. (2022).

22. **Dutt S, Parkash J, Mehra R, Sharma N, Singh B, Raigond P, et al.** Translation initiation in plants: roles and implications beyond protein synthesis. Biologia Plantarum. (2015).

23. **Merchante C, Stepanova AN, Alonso JM.** Translation regulation in plants: an interesting past, an exciting present and a promising future. Plant Journal. (2017).

24. **Horiguchi G, Van Lijsebettens M, Candela H, Micol JL, Tsukaya H.** Ribosomes and translation in plant developmental control. Plant Science. (2012).

25. **Hager GL, McNally JG, Misteli T.** Transcription Dynamics. Molecular Cell. (2009).

26. **Ramakrishnan., V.** Ribosome Structure and the Mechanism of Translation. Cell press. (2002).

27. **Dutt S, Parkash J, Mehra R, Sharma N, Singh B, Raigond P, et al.** Translation initiation in plants: roles and implications beyond protein synthesis. Biologia Plantarum. (2015).

28. **Hershey JWB, Sonenberg N, Mathews MB.** Principles of Translational Control: An Overview. Cold Spring Harb Perspect Biol. (2012).

29. **Lan T, Xiong W, Chen X, Mo B, Tang G.** Plant cytoplasmic ribosomal proteins: an update on classification, nomenclature, evolution and resources. Plant Journal. (2022).

30. **Hashem Y, Frank J.** The Jigsaw Puzzle of mRNA Translation Initiation in Eukaryotes: A Decade of Structures Unraveling the Mechanics of the Process. Annu Rev Biophys. (2018).

31. **Juntawong P, Girke T, Bazin J, Bailey-Serres J. Translational** dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*. Proc Natl Acad Sci U S A. (2014).

32. **Scarpin MR, Busche M, Martinez RE, Harper LC, Reiser L, Szakonyi D, et al.** An updated nomenclature for plant ribosomal protein genes. American Society of Plant Biologists. (2023).

33. **Wilson DN, Cate JHD.** The structure and function of the eukaryotic ribosome. Cold Spring Harb Perspect Biol. (2012).

34. **Schmeing TM, Ramakrishnan V.** What recent ribosome structures have revealed about the mechanism of translation. Nature. (2009).

35. **Zimmerman RA, Dahl-Berg AE, Nissen P, Hansen J, Ban N, et al.** The structural basis of ribosome activity in peptide bond synthesis. Science. (2000).

36. **Lange S, Franks WT, Rajagopalan N, Döring K, Geiger MA, Linden A, et al.** Structural analysis of a signal peptide inside the ribosome tunnel by DNP MAS NMR. Sci Adv. (2016).

37. **Imai A, Hanzawa Y, Komura M, Yamamoto KT, Komeda Y, Takahashi T.** The dwarf phenotype of the Arabidopsis ac15 mutant is suppressed by a mutation in an upstream ORF of a bHLH gene. Development. (2006).

38. **Ito T, Kim GT, Shinozaki K.** Disruption of an Arabidopsis cytoplasmic ribosomal protein S13-homologous gene by transposon-mediated mutagenesis causes aberrant growth and development. Plant Journal. (2000).

39. **Szakonyi D, Byrne ME.** Ribosomal protein L27a is required for growth and patterning in Arabidopsis thaliana. Plant Journal. (2011).

40. **Horiguchi G, Van Lijsebettens M, Candela H, Micol JL, Tsukaya H.** Ribosomes and translation in plant developmental control. Plant Science. (2012).

41. **Casanova-Sáez R, Candela H, Micol JL.** Combined haploinsufficiency and purifying selection drive retention of RPL36a paralogs in Arabidopsis. Sci Rep. (2014).

42. **Liu H, Ren D, Jiang L, Li X, Yao Y, Mi L, Chen W, Mo A, Jiang N, Yang J, Chen P, Ma H, Luo X, Lu P.** A Natural Variation in *PLEIOTROPIC DEVELOPMENTAL DEFECTS* Uncovers a Crucial Role for Chloroplast tRNA Modification in Translation and Plant Development. Plant Cell. (2020).

43. **Creff A, Sormani R, Desnos T.** The two Arabidopsis RPS6 genes, encoding for cytoplasmic ribosomal proteins S6, are functionally equivalent. Plant Mol Biol. (2010).

44. **Tzafrir I, Pena-Muralla R, Dickerman A, Berg M, Rogers R, Hutchens S, et al.** Identification of genes required for embryo development in arabidopsis. Plant Physiol. (2004).

45. **Ryabova LA, Pooggin MM, Hohn T.** Viral Strategies of Translation Initiation: Ribosomal Shunt and Reinitiation. Prog Nucleic Acid Res Mol Biol. (2002).

46. **Sonenberg N, Dever TE**. Eukaryotic translation initiation factors and regulators. Current Opinion in Structural Biology. (2003).

47. **Kimball SR.** Eukaryotic initiation factor eIF2. Int J Biochem Cell Biol. (1999).

48. **Zhou F, Roy B, Dunlap JR, Enganti R, von Arnim AG.** Translational Control of Arabidopsis Meristem Stability and Organogenesis by the Eukaryotic Translation Factor eIF3h. PLoS One. (2014).

49. **Lomakin IB, Kolupaeva VG, Marintchev A, Wagner G, Pestova T V.** Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. Genes Dev. (2003).

50. **Wortham NC, Proud CG.** EIF2B: Recent structural and functional insights into a key regulator of translation. Biochem Soc Trans. (2015).

51. **Hronová V, Mohammad MP, Wagner S, Pánek J, Gunišová S, Zeman J, et al.** Does eIF3 promote reinitiation after translation of short upstream ORFs also in mammalian cells? RNA Biology. (2017).

52. **Schepetilnikov M, Dimitrova M, Mancera-Martínez E, Geldreich A, Keller M, Ryabova LA.** TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. EMBO Journal. (2013).

53. **Rogers GW, Richter NJ, Lima WF, Merrick WC.** Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. Journal of Biological Chemistry. (2001).

54. **Jennings MD, Pavitt GD.** eIF5 is a dual function GAP and GDI for eukaryotic translational control. Small GTPases. (2010).

55. **Sanfaçon H.** Plant translation factors and virus resistance. Viruses. (2015).

56. **Kozak M.** Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene. (2005).

57. **Noller HF, Lancaster L, Zhou J, Mohan S.** The ribosome moves: RNA mechanics and translocation. Nature Structural and Molecular Biology. (2017).

58. **Wang XQ, Rothnagel JA.** 5'-Untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. Nucleic Acids Res. (2004).

59. **Mateyak MK, Kinzy TG.** eEF1A: Thinking outside the ribosome. Journal of Biological Chemistry. (2010).

60. **Le Sourd F, Boulben S, Le Bouffant R, Cormier P, Morales J, Belle R, et al.** eEF1B: At the dawn of the 21st century. Biochimica et Biophysica Acta - Gene Structure and Expression. (2006).

61. **Spahn CMT, Gomez-Lorenzo MG, Grassucci RA, Jørgensen R, Andersen GR, Beckmann R, et al.** Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. EMBO Journal. (2004).

62. **Dever TE, Gutierrez E, Shin BS**. The hypusine-containing translation factor eIF5A. Critical Reviews in Biochemistry and Molecular Biology. (2014).

63. **Zhouravleva G, Frolova L, Le Goff X, Le Guellec R, Inge-Vechtomov S, Kisselev L, et al.** Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. EMBO Journal. (1995).

64. **Rodnina M V.** Protein synthesis meets ABC ATPases: New roles for Rli1/ABCE1. EMBO Reports. (2010).

65. **Cramer P.** Organization and regulation of gene transcription. Vol. Nature. (2019).

66. **Kolitz SE, Lorsch JR.** Eukaryotic initiator tRNA: Finely tuned and ready for action. FEBS Letters. (2010).

67. **Patrick RM, Mayberry LK, Choy G, Woodard LE, Liu JS, White A, et al.** Two arabidopsis loci encode novel eukaryotic initiation factor 4E isoforms that are functionally distinct from the conserved plant eukaryotic initiation factor 4E. Plant Physiol. (2014).

68. **Chaudhuri J, Si K, Maitra U**. Function of eukaryotic translation initiation factor 1A (eIF1A) (formerly called eIF-4C) in initiation of protein synthesis. Journal of Biological Chemistry. (1997).

69. **Kozak M.** Point Mutations Define a Sequence Flanking the AUG Initiator Codon That Modulates Translation by Eukaryotic Ribosomes. Cell. (1986).

70. **Ismail N, Hedman R, Schiller N, Von Heijne G.** A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. Nat Struct Mol Biol. (2012).

71. **Von Arnim AG, Jia Q, Vaughn JN.** Regulation of plant translation by upstream open reading frames. Plant Science. (2014)

72. **Young DJ, Guydosh NR.** Rebirth of the translational machinery: The importance of recycling ribosomes. BioEssays. (2022).

73. **Ryabova LA, Pooggin MM, Hohn T.** Translation reinitiation and leaky scanning in plant viruses. (2006).

74. **Gunišová S, Hronová V, Mohammad MP, Hinnebusch AG, Valášek LS**. Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. FEMS Microbiol Rev. (2018).

75. **Calvo SE, Pagliarini DJ, Mootha VK**. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. PNAS. (2009).

76. **Vaughn JN, Ellingson SR, Mignone F, Von Arnim A.** Known and novel post-transcriptional regulatory sequences are conserved across plant families. RNA. (2012).

77. **Dong Y, Ryabova LA.** Do plants drive translation reinitiation to dodge nonsense-mediated decay? J Exp Bot. (2023).

78. **Hinnebusch AG.** Translational regulation of GCN4 and the general amino acid control of yeast. Annual Review of Microbiology. (2005).

79. **Schepetilnikov M, Ryabova LA.** Auxin signaling in regulation of plant translation reinitiation. Frontiers in Plant Science. (2017).

80. **Roy B, Vaughn JN, Kim BH, Zhou F, Gilchrist MA, Von Arnim AG.** The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames. RNA. (2010).

81. **Luttermann C, Meyers G.** The importance of inter- and intramolecular base pairing for translation reinitiation on a eukaryotic bicistronic mRNA. Genes Dev. (2009).

82. **Pöyry TAA, Kaminski A, Connell EJ, Fraser CS, Jackson RJ.** The mechanism of an exceptional case of reinitiation after translation of a long ORF reveals why such events do not generally occur in mammalian mRNA translation. Genes Dev. (2007).

83. **Álvarez D, Voß B, Maass D, Wüst F, Schaub P, Beyer P, et al**. Carotenogenesis is regulated by 5‘UTR-mediated translation of phytoene synthase splice variants. Plant Physiol. (2016).

84. **Walter MH, Stauder R, Tissier A.** Evolution of root-specific carotenoid precursor pathways for apocarotenoid signal biogenesis. Plant Science. (2015).

85. **Kirk MM, Stark K, Miller SM, Müller W, Taillon BE, Gruber H, Schmitt R, Kirk DL.** regA, a Volvox gene that plays a central role in germ-soma differentiation, encodes a novel regulatory protein. Development. (1999).

86. **Hulzink RJM, De Groot PFM, Croes AF, Quaedvlieg W, Twell D, Wullems GJ, et al.** The 5'-untranslated region of the ntp303 gene strongly enhances translation during pollen tube growth, but not during pollen maturation. Plant Physiol. (2002).

87. **Miller WA, Wang Z, Treder K.** The amazing diversity of cap-independent translation elements in the 3'-untranslated regions of plant viral RNAs. Biochemical Society Transactions. (2007).

88. **Dinkova TD, Zepeda H, Martínez-Salas E, Martínez LM, Nieto-Sotelo J, Sánchez De Jiménez E.** Cap-independent translation of maize Hsp101. Plant Journal. (2005).

89. **Jackson RJ, Hellen CUT, Pestova T V.** The mechanism of eukaryotic translation initiation and principles of its regulation. Nature Reviews Molecular Cell Biology. (2010).

90. **Hayden CA, Bosco G.** Comparative genomic analysis of novel conserved peptide upstream open reading frames in Drosophila melanogaster and other dipteran species. BMC Genomics. (2008).

91. **Dennis MD, Person MD, Browning KS.** Phosphorylation of plant translation initiation factors by CK2 enhances the in Vitro interaction of multifactor complex components. Journal of Biological Chemistry. (2009).

92. **Martinez-Seidel F, Beine-Golovchuk O, Hsieh YC, Kopka J.** Systematic review of plant ribosome heterogeneity and specialization. Frontiers in Plant Science. (2020).

93. **Chang IF, Szick-Miranda K, Pan S, Bailey-Serres J.** Proteomic characterization of evolutionarily conserved and variable proteins of arabidopsis cytosolic ribosomes. Plant Physiol. (2005).

94. **Whittle CA, Krochko JE.** Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in brassica napus. Plant Cell. (2009).

95. **Falcone Ferreyra ML, Casadevall R, Luciani MD, Pezza A, Casati P.** New evidence for differential roles of L10 ribosomal proteins from Arabidopsis. Plant Physiol. (2013).

96. **Hummel M, Cordewener JHG, de Groot JCM, Smeekens S, America AHP, Hanson J.** Dynamic protein composition of Arabidopsis thaliana cytosolic ribosomes in response to sucrose feeding as revealed by label free MS E proteomics. Proteomics. (2012).

97. **Sáez-Vásquez J, Delseny M.** Ribosome biogenesis in plants: From functional 45S ribosomal DNA organization to ribosome assembly factors. Plant Cell. (2019).

98. **Kim JY, Song HR, Taylor BL, Carré IA.** Light-regulated translation mediates gated induction of the Arabidopsis clock protein LHY. EMBO J. (2003).

99. **Guo J, Wang S, Valerius O, Hall H, Zeng Q, Li JF, et al.** Involvement of Arabidopsis RACK1 in protein translation and its regulation by abscisic acid. Plant Physiol. (2011).

100. **Roy B, Arnim AG von.** Translational Regulation of Cytoplasmic mRNAs. Arabidopsis Book. (2013).

101. **Tran MK, Schultz CJ, Baumann U.** Conserved upstream open reading frames in higher plants. BMC Genomics. (2008).

102. **Jorgensen RA, Dorantes-Acosta AE.** Conserved peptide upstream open reading frames are associated with regulatory genes in angiosperms. Front Plant Sci. (2012).

103. **Takahashi H, Takahashi A, Naito S, Onouchi H.** BAIUCAS: A novel BLAST-based algorithm for the identification of upstream open reading frames with conserved amino acid sequences and its application to the Arabidopsis thaliana genome. Bioinformatics. (2012).

104. **Van Der Horst S, Snel B, Hanson J, Smeekens S.** Novel pipeline identifies new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid sequences in the 5 ' leader of mRNAs in Arabidopsis thaliana. RNA. (2019).

105. **Takahashi H, Hayashi N, Hiragori Y, Sasaki S, Motomura T, Yamashita Y, et al.** Comprehensive genome-wide identification of angiosperm upstream ORFs with peptide sequences conserved in various taxonomic ranges using a novel pipeline, ESUCA. BMC Genomics. (2020).

106. **Takahashi H, Miyaki S, Onouchi H, Motomura T, Idesako N, Takahashi A, et al.** Exhaustive identification of conserved upstream open reading frames with potential translational regulatory functions from animal genomes. Sci Rep. (2020).

107. **Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, Bailey-Serres J.** Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. Proc Natl Acad Sci U S A. (2017).

108. **Ebina I, Takemoto-Tsutsumi M, Watanabe S, Koyama H, Endo Y, Kimata K, et al.** Identification of novel Arabidopsis thaliana upstream open reading frames that control expression of the main coding sequences in a peptide sequence-dependent manner. Nucleic Acids Res. (2015).

109. **Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, Hanson J.** Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. Plant Physiol. (2009).

110. **van der Horst S, Englmeier R, Hanson J, Smeekens S, Förster F.** Sucrose-mediated translational stalling involves a conserved ribosomal pocket. Biorxiv. (2023).

111. **Bhushan S, Hoffmann T, Seidelt B, Frauenfeld J, Mielke T, Berninghausen O, et al.** SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. PLoS Biol. (2011).

112. **Gerashchenko M V., Gladyshev VN.** Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res. (2014).

113. **Ivanov IP, Atkins JF, Michael AJ.** A profusion of upstream open reading frame mechanisms in polyamine-responsive translational regulation. Nucleic Acids Res. (2009).

114. **Ivanov IP, Shin BS, Loughran G, Tzani I, Young-Baird SK, Cao C, et al.** Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. Mol Cell. (2018).

115. **Hsu PY, Calviello L, Wu HYL, Li FW, Rothfels CJ, Ohler U, et al.** Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. PNAS. (2016).

116. **Hou CY, Lee WC, Chou HC, Chen AP, Chou SJ, Chen HM.** Global analysis of truncated RNA ends reveals new insights into Ribosome Stalling in plants. Plant Cell. (2016).

117. **Li Z, Fu Y, Shen J, Liang J.** Upstream open reading frame mediated translation of wnk8 is required for aba response in arabidopsis. Int J Mol Sci. (2021).

118. **Kakehi JI, Kawano E, Yoshimoto K, Cai Q, Imai A, Takahashi T.** Mutations in ribosomal proteins, RPL4 and RACK1, suppress the phenotype of a thermospermine-deficient mutant of arabidopsis thaliana. PLoS One. (2015).

119. **Vera-Sirera F, De Rybel B, Úrbez C, Kouklas E, Pesquera M, Álvarez-Mahecha JC, et al.** A bHLH-Based Feedback Loop Restricts Vascular Cell Proliferation in Plants. Dev Cell. (2015).

120. **Pollutri D, Penzo M.** Ribosomal Protein L10: From Function to Dysfunction. Cells. (2020).

121. **Falcone Ferreyra ML, Casadevall R, Luciani MD, Pezza A, Casati P.** New evidence for differential roles of L10 ribosomal proteins from Arabidopsis. Plant Physiol. (2013).

122. **Adams DR, Ron D, Kiely PA.** RACK1, A multifaceted scaffolding protein: Structure and function. Cell Communication and Signaling. (2011).

123. **Su J, Xu J, Zhang S.** RACK1, scaffolding a heterotrimeric G protein and a MAPK cascade. Trends in Plant Science. (2015).

124. **Narita M, Denk T, Matsuo Y, Sugiyama T, Kikuguchi C, Ito S, et al.** A distinct mammalian disome collision interface harbors K63-linked polyubiquitination of uS10 to trigger hRQT-mediated subunit dissociation. Nat Commun. (2022).

125. **Laing WA, Martínez-Sánchez M, Wright MA, Bulley SM, Brewster D, Dare AP, et al.** An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in arabidopsis. Plant Cell. (2015).

126. **Ma J, Hanssen M, Lundgren K, Hernández L, Delatte T, Ehlert A, et al.** The sucrose-regulated Arabidopsis transcription factor bZIP11 reprograms metabolism and regulates trehalose metabolism. New Phytologist. (2011).

127. **Hanson J, Hanssen M, Wiese A, Hendriks MMWB, Smeekens S.** The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE1 and PROLINE DEHYDROGENASE2. Plant Journal. (2008).

128. **Zhu X, Li Y, Fang W, Kusano T.** Galactinol is involved in sequence-conserved upstream open reading frame-mediated repression of Arabidopsis HsfB1 translation. Environ Exp Bot. (2018).

129. **Zhu X, Thalor SK, Takahashi Y, Berberich T, Kusano T.** An inhibitory effect of the sequence-conserved upstream open-reading frame on the translation of the main open-reading frame of HsfB1 transcripts in Arabidopsis. Plant Cell Environ. (2012).

130. **Wiese A, Elzinga N, Wobbes B, Srneekens S.** A conserved upstream open reading frame mediates sucrose-induced repression of translation. Plant Cell. (2004).

131. **Hanfrey C, Elliott KA, Franceschetti M, Mayer MJ, Illingworth C, Michael AJ.** A dual upstream open reading frame-based autoregulatory circuit controlling polyamine-responsive translation. Journal of Biological Chemistry. (2005).

132. **Guerrero-González ML, Rodríguez-Kessler M, Jiménez-Bremont JF.** UORF, a regulatory mechanism of the Arabidopsis polyamine oxidase 2. Mol Biol Rep. (2014).

133. **Guerrero-González M de la L, Ortega-Amaro MA, Juárez-Montiel M, Jiménez-Bremont JF.** Arabidopsis Polyamine oxidase-2 uORF is required for downstream translational regulation. Plant Physiology and Biochemistry. (2016).

134. **Fulgencio Alatorre-Cobos, Alfredo Cruz-Ramírez, Celine A. Hayden, Claudia-Anahí Pérez-Torres, Anne-Laure Chauvin, Enrique Ibarra-Laclette, Erika Alva-Cortés, Richard A. Jorgensen, Luis Herrera-Estrella**. Translational regulation of *Arabidopsis XIPOTL1* is modulated by phosphocholine levels via the phylogenetically conserved upstream open reading frame 30. *Journal of Experimental Botany*. (2012).

135. **Gong F, Ito K, Nakamura Y, Yanofsky C**. The mechanism of tryptophan induction of tryptophanase operon expression: Tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNA Pro. PNAS. (2001).

136. **Emmanuel JS, Sengupta A, Gordon ER, Noble JT, Cruz-Vera LR**. The regulatory TnaC nascent peptide preferentially inhibits release factor 2-mediated hydrolysis of peptidyl-tRNA. Journal of Biological Chemistry. (2019).

137. **Martínez AK, Shirole NH, Murakami S, Benedik MJ, Sachs MS, Cruz-Vera LR.** Crucial elements that maintain the interactions between the regulatory TnaC peptide and the ribosome exit tunnel responsible for Trp inhibition of ribosome function. Nucleic Acids Res. (2012).

138. **Bhushan S, Gartmann M, Halic M, Armache JP, Jarasch A, Mielke T, et al.** α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. Nat Struct Mol Biol. (2010).

139. **Bhushan S, Meyer H, Starosta AL, Becker T, Mielke T, Berninghausen O, et al.** Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. Mol Cell. (2010).

140. **Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, et al.** Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. Proc Natl Acad Sci U S A. (2010).

141. **Nakatogawa H, Ito K.** The Ribosomal Exit Tunnel Functions as a Discriminating Gate the protein is the nascent polypeptide in the cytoplasm. The translation of SecM is subject to elongation arrest at a site close to the C terminus. Cell. (2002).

142. **Klinge S, Voigts-Hoffmann F, Leibundgut M, Ban N.** Atomic structures of the eukaryotic ribosome. Trends in Biochemical Sciences. (2012).

143. **Han Y, Gao X, Liu B, Wan J, Zhang X, Qian SB.** Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. Cell Res. (2014).

144. **Shalgi R, Hurt JA, Krykbaeva I, Taipale M, Lindquist S, Burge CB.** Widespread Regulation of Translation by Elongation Pausing in Heat Shock. Mol Cell. (2013).

145. **Seidelt B, Innis CA, Wilson DN, Gartmann M, Armache JP, Villa E, et al.** Structural insight into nascent polypeptide chain-mediated translational stalling. Science (2009).

146. **Dever TE, Ivanov IP, Sachs MS.** Conserved Upstream Open Reading Frame Nascent Peptides That Control Translation. Annual Review of Genetics. (2002).

147. **Muto H, Nakatogawa H, Ito K.** Genetically Encoded but Nonpolypeptide Prolyl-tRNA Functions in the A Site for SecM-Mediated Ribosomal Stall. Mol Cell. (2006).

148. **Lu J, Kobertz WR, Deutsch C.** Mapping the Electrostatic Potential within the Ribosomal Exit Tunnel. J Mol Biol. (2007).

149. **Wilson DM, Li Y, LaPeruta A, Gamalinda M, Gao N, Woolford JL.** Structural insights into assembly of the ribosomal nascent polypeptide exit tunnel. Nat Commun. (2020).

150. Fritz, S., and K. Boris-Lawrie. The RNPs of eukaryotic translation control. Trends in Cell & Molecular Biology. (2015).

151. **Sitron CS, Brandman O.** Detection and Degradation of Stalled Nascent Chains via Ribosome-Associated Quality Control. annurev-biochem. (2020).

152. **Halic M, Becker T, Pool MR, Spahn CMT, Grassucci RA, Frank J, et al.** Structure of the signal recognition particle interacting with the elongation-arrested ribosome. RNA. (2004).

153. **Zhang X, Jain R, Li G.** Roles of rack1 proteins in fungal pathogenesis. BioMed Research. (2016).

154. **Wolf AS, Grayhack EJ.** Asc1, homolog of human RACK1, prevents frameshifting in yeast by ribosomes stalled at CGA codon repeats. RNA. (2015).

155. **Wang W, Wang X, Wang X, Ahmed S, Hussain S, Zhang N, et al.** Integration of RACK1 and ethylene signaling regulates plant growth and development in Arabidopsis. Plant Science. (2019).

156. **Rollins MG, Shasmal M, Meade N, Astar H, Shen PS, Walsh D.** Negative charge in the RACK1 loop broadens the translational capacity of the human ribosome. Cell Rep. (2021).

157. **Sengupta J, Nilsson J, Gursky R, Spahn CMT, Nissen P, Frank J.** Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM. Nat Struct Mol Biol. (2004).

158. **Sundaramoorthy E, Leonard M, Mak R, Liao J, Fulzele A, Bennett EJ.** ZNF598 and RACK1 Regulate Mammalian Ribosome-Associated Quality Control Function by Mediating Regulatory 40S Ribosomal Ubiquitylation. Mol Cell. (2017).

159. **Kuroha K, Akamatsu M, Dimitrova L, Ito T, Kato Y, Shirahige K, et al.** Receptor for activated C kinase 1 stimulates nascent polypeptide-dependent translation arrest. EMBO Rep. (2010).

160. **Zaher HS, Green R**. Quality control by the ribosome following peptide bond formation. Nature. (2009).

161. **Buskirk AR, Green R.** Ribosome pausing, arrest and rescue in bacteria and eukaryotes. Vol. 372, Philosophical Transactions of the Royal Society B: Biological Sciences. (2017).

162. **Nguyen NH, Bui TP, Le NT, Nguyen CX, Le MTT, Dao NT, et al.** Disrupting Sc-uORFs of a transcription factor bZIP1 using CRISPR/Cas9 enhances sugar and amino acid contents in tomato (Solanum lycopersicum). Planta. (2023).

163. **Reynoso MA, Juntawong P, Lancia M, Blanco FA, Bailey-Serres J, Zanetti ME**. Translating ribosome affinity purification (TRAP) followed by RNA sequencing technology (TRAP-SEQ) for quantitative assessment of plant translatomes. In: Plant Functional Genomics. (2015).

164. **Scarff CA, Fuller MJG, Thompson RF, Iadaza MG.** Variations on negative stain electron microscopy methods: Tools for tackling challenging systems. Journal of Visualized Experiments. (2018).

165. **Scheres SHW.** RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. (2012).

166. **Rohou A, Grigorieff N.** CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol. (2015).

167. **Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, et al.** SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun Biol. (2019).

168. **Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al.** UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Science. (2021).

169. **Ranganathan, J., Waite, R., Searchinger, T., & Hanson, C.** How to Sustainably Feed 10 Billion People by 2050. World Resources Institute. (2018).

170. **Buchwald G, Hostinova E, Rudolph MG, Kraemer A, Sickmann A, Meyer HE, et al.** Conformational Switch and Role of Phosphorylation in PAK Activation. Mol Cell Biol. (2001).

171. **Andres J, Blomeier T, Zurbriggen MD.** Synthetic switches and regulatory circuits in plants. Plant Physiol. (2019).

172. **Mitra K, Frank J.** RIBOSOME DYNAMICS: Insights from Atomic Structure Modeling into Cryo-Electron Microscopy Maps. Annu Rev Biophys Biomol Struct. (2006).

173. **Wang YH, Dai H, Zhang L, Wu Y, Wang J, Wang C, et al.** Cryo-electron microscopy structure and translocation mechanism of the crenarchaeal ribosome. Nucleic Acids Res. (2023).

174. **Smirnova J, Loerke J, Kleinau G, Schmidt A, Bürger J, Meyer EH, et al.** Structure of the actively translating plant 80S ribosome at 2.2 Å resolution. Nat Plants. (2023).

175. **Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, et al.** Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. PNAS. (2010).

176. **Holvec S, Barchet C, Lechner A, Fréchin L, Nimali S, De Silva T, et al.** Structure of the human 80S ribosome at 1.9 Å resolution-the molecular role of chemical modifications and ions in RNA. Biorxiv. (2023).

177. **Wilson DN, Arenz S, Beckmann R.** Translation regulation via nascent polypeptide-mediated ribosome stalling. Current Opinion in Structural Biology. (2016).

178. **Zimmerman RA, Dahl-Berg AE, Nissen P, Hansen J, Ban N, et al.** The structural basis of ribosome activity in peptide bond synthesis. Science. (2000).

179. **Duc KD, Batra SS, Bhattacharya N, Cate JHD, Song YS.** Differences in the path to exit the ribosome across the three domains of life. Nucleic Acids Res. (2019).

180. **Byrne ME.** A role for the ribosome in development. Trends in Plant Science. (2009).

181. **Orlova E V., Saibil HR.** Structural analysis of macromolecular assemblies by electron microscopy. Chemical Reviews. (2011).

182. **Murata K, Wolf M.** Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochimica et Biophysica. (2018).

183. **Milne JLS, Borgnia MJ, Bartesaghi A, Tran EEH, Earl LA, Schauder DM, et al.** Cryo-electron microscopy - A primer for the non-microscopist. FEBS Journal. (2013).

184. **Charneski CA, Hurst LD.** Positively Charged Residues Are the Major Determinants of Ribosomal Velocity. PLoS Biol. (2013).

185. **Frank J.** Cryo-electron microscopy as an investigative tool: The ribosome as an example. BioEssays. (2001).

186. **Bieri P, Leibundgut M, Saurer M, Boehringer D, Ban N.** The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. EMBO J. (2017).

187. **Abid Javed, Tomasz Wlodarski, Anaïs. M.E. Cassaignau, Lisa D. Cabrita, JohnChristodoulou, Elena V. Orlova.** Visualising nascent chain dynamics at the ribosome exit tunnel by cryo-electron microscopy. bioRxiv. (2019).

188. **Gong X, Qian H, Zhou X, Wu J, Wan T, Cao P, et al.** Structural insights into the Niemann-Pick C1 (NPC1)-mediated cholesterol transfer and ebola infection. Cell. (2016).

189. **White HE, Ignatiou A, Clare DK, Orlova E V.** Structural Study of Heterogeneous Biological Samples by Cryoelectron Microscopy and Image Processing. BioMed Research International. (2017).

190. **Aspden J, Fontana J, Bretman A.** Dissecting the mechanisms of translational regulation by neuronal ribosomes. (2022).

191. **Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A.** UORFdb - A comprehensive literature database on eukaryotic uORF biology. Nucleic Acids Res. (2014).

192. **Niu R, Zhou Y, Zhang Y, Mou R, Tang Z, Wang Z, et al.** UORFlight: A vehicle toward uORF-mediated translational regulation mechanisms in eukaryotes. Database. (2020).

193. **Rubio A, Ghosh S, Mülleder M, Ralser M, Mata J.** Ribosome profiling reveals ribosome stalling on tryptophan codons and ribosome queuing upon oxidative stress in fission yeast. Nucleic Acids Res. (2021).

194. **Yihan Dong, Lyubov A Ryabova**. Do plants drive translation reinitiation to dodge nonsense-mediated decay?, *Journal of Experimental Botany*. (2023).

195. **Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z.** GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. (2009).

196. **Babaria K.** Using Treemaps to Visualize Gene Ontologies. (2001).

197. **Supek F, Bošnjak M, Škunca N, Šmuc T.** Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One. (2011).

198. **Qian W, Yang JR, Pearson NM, Maclean C, Zhang J.** Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet. (2012).

199. **Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, et al.** Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. Mol Cell. (2015).

200. **O'Connell MJ, Doyle AM, Juenger TE, Donoghue MT, Keshavaiah C, Tuteja R, et al.** In Arabidopsis thaliana codon volatility scores reflect GC3 composition rather than selective pressure. BMC Res Notes. (2012).

201. **Alatorre-Cobos F, Cruz-Ramírez A, Hayden CA, Pérez-Torres CA, Chauvin AL, Ibarra-Laclette E, et al.** Translational regulation of Arabidopsis XIPOTL1 is modulated by phosphocholine levels via the phylogenetically conserved upstream open reading frame 30. J Exp Bot. (2012).

202. **Ribone PA, Capella M, Arce AL, Chan RL**. A uORF represses the transcription factor AtHB1 in aerial tissues to avoid a deleterious phenotype. Plant Physiol. (2017).

203. **Tomczak A, Mortensen JM, Winnenburg R, Liu C, Alessi DT, Swamy V, et al.** Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. Sci Rep. (2018).