

**ASCORBATE-SENSING uORF CONTROLS GDP-L-
GALACTOSE PHOSPHORYLASE (GGP) mRNA
TRANSLATION**

by
Leda Eleanor Lotspeich-Cole

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
May 2020

© 2020 Leda E. Lotspeich-Cole
All Rights Reserved

Abstract

According to the scanning model of eukaryotic translation initiation, ribosomes bind near the 5' cap of an mRNA and then initiate translation at the first start codon they encounter as they progress down the mRNA. Upstream open reading frames (uORFs) can regulate translation initiation at downstream start codons. Recently, a uORF in the GDP-L-galactose phosphorylase (*GGP*) mRNA was reported to mediate autoregulation of ascorbate, or Vitamin C, production. Ascorbate is an essential enzymatic cofactor and antioxidant in plants and animals. Ascorbate concentrations are tightly regulated in plants through post-transcriptional repression of *GGP*, the major enzyme controlling ascorbate biosynthesis. Key attributes of the noncanonical *GGP* uORF include: (1) initiation at either a near-cognate start codon or an AUG codon in poor context; (2) it encodes a 32-65 residue peptide containing conserved motifs; and (3) ribosome profiling indicates the presence of an elongation pause at the uORF C-terminus. These attributes suggest that the *GGP* uORF is a novel upstream Conserved Coding region (uCC), a new paradigm of translational control involving a uORF whose initiation at a weak start site is triggered by queuing of scanning ribosomes behind an elongating ribosome paused at a conserved amino acid motif within the uCC. We hypothesize that ascorbate promotes pausing of translation elongation in the uORF to enhance uORF translation and to repress translation of *GGP*. To assess ascorbate mediated *GGP* repression by the uORF, we examined reporter constructs in mammalian cell lines and in rabbit reticulocyte lysate *in vitro* translation systems. Alignment of 20 algae species' *GGP* uORFs shows absolute conservation of five amino acid motifs. Consistent with the *GGP* uORF being a uCC, conservation of all five motifs and their spacing is essential for regulation of *GGP* mRNA translation by ascorbate *in vivo*

and *in vitro*, and the elongation pause is specific to ascorbate. Ribosome profiling from *Chlamydomonas reinhardtii* and *Solanum lycopersicum* showed an elongation pause at the C-terminal glycine repeat. In support of the queuing model, reducing ribosome loading by inserting a cap-proximal stem-loop in the GGP mRNA abolished ascorbate specific regulation of *GGP* expression. Taken together, these data indicate that the *GGP* uORF is a uCC. We propose that ascorbate interacts with conserved residues of the uCC nascent peptide within the ribosome exit tunnel, causing an elongation pause that regulates GGP mRNA translation.

Thesis Committee

Dr. Sarah Woodson, Ph.D. Department of Biophysics, CMDB/JHU
Dr. Richard Maraia, M.D. Section on Molecular and Cell Biology, NICHD/NIH
Dr. Karen Beemon, Ph.D. Department of Biology, CMDB/JHU
Dr. Thomas Dever, Ph.D. Advisor, Section on Protein Biosynthesis, NICHD/NIH

Readers

Dr. Thomas Dever
Dr. Sarah Woodson

Preface and Acknowledgements

First and foremost, I thank my family for their support during this journey. Without the support of my immediate family, I would not have been able to complete my Ph.D. while having a young family. My grandparents encouraged my interest in science and my grandmother provided immense support during the first years of my graduate studies. Unfortunately, she is not alive to see me graduate, but I was able to tell her that I was wrapping up my experiments and writing. My parents have helped me throughout the years by providing support for my education and by helping at home so that I could work later and know that my daughter was in good hands. I also need to thank my husband for his support (and pressure) to complete my Ph.D. He provided technical and scientific support as well as keeping my eye on the prize when I felt burned out. Finally, I need to thank my daughter for reminding me that science, although consuming, is a job and that there is a lot to life outside the lab. She also reminded me that what I do is awesome and that even transferring fluids between containers can be fun and “super cool”.

I am grateful to my advisor, Dr. Thomas Dever, for his guidance and support over the years. He helped me feel enthusiastic about starting my third and final project and his love of science helped me get through the grind of experiment after experiment. I would not have had the motivation to keep at it without his support and understanding.

I also want to thank the Dever and Hinnebusch lab members who assisted me throughout the years. Chune Cao was an amazing source of technical knowledge as well as plant clippings and friendship. Dr. Byung Sik-Shin patiently shared protocols and years of knowledge with me as did Joo-Ran Kim. The current and past post-doctoral fellows in the lab shared technical knowledge, reagents, and career advice with me. Fan Zhang

provided reagents and technical support. Dr. Ivaylo Ivanov provided guidance on reporter development and expertise on the *GFP* uCC. And finally, Dr. Alan Hinnebusch and the other members of the lab meeting super group provided feedback on experimental design and my research during meetings.

Table of Contents

Front Matter	i
Abstract	i
Committee	iv
Preface and Acknowledgements	v
Table of Contents	vii
List of Tables	xiii
List of Figures	xv
1. Introduction	18
1.1. Regulation of gene expression	19
1.2. mRNA features and RNA stability	24
1.3. Protein biosynthesis	29
1.4. Upstream open reading frames and translation of main open reading frames	44
1.5. uORF-encoded nascent polypeptide-mediated translational stalling	55
1.6. Polyamines and uORF mediated autoregulatory circuits	63
1.7. Regulation of protein expression by upstream conserved coding motifs	70
1.8. Plant L-ascorbic acid biosynthesis and GDP-L-galactose phosphorylase	75
1.9. Is the GGP motif a uCC that regulates ascorbate biosynthesis?	84
2. Materials and Methods	86
2.1. Plasmid constructs	87

2.1.1. Molecular techniques for cloning the reporters	87
2.1.2. Primers used in the study	92
2.1.3. The synthesized G-blocks	95
2.1.4. Generation of the uCC-FLuc fusion <i>in vitro</i> translation reporters (Table 2.2)	103
2.1.5. Generation of the uORF <i>in vitro</i> translation reporters (Table 2.3)	108
2.1.6. Generation of the <i>in vitro</i> translation FLuc controls (Table 2.4) ..	110
2.1.7. Generation of the uCC-regulated <i>Renilla</i> reporters for mammalian transfection experiments (Table 2.5).....	113
2.1.8. Generation of the uCC-FLuc fusions for mammalian transfection experiments (Table 2.6 and Table 2.7)	117
2.1.9. Generation of the phRL based reporters containing alanine substitutions within the uCC (Table 2.8)	119
2.1.10. Generation of the alanine insertions within the uCC for the uCC- FLuc fusion reporters used in mammalian transfection experiments (Table 2.9)	121
2.1.11. Generation of the stem-loop uCC FLuc reporters for mammalian transfection experiments (Table 2.10)	122
2.1.12. Generation of the phRL based stem-loop uCC RLuc reporters for mammalian transfection experiments (Table 2.11)	124
2.1.13. Generation of the phRL based Trip12-extended uCC reporters for mammalian transfection experiments (Table 2.12)	125

2.1.14. Generation of the Trip12-extended uCC FLuc fusion reporter for mammalian transfection experiments (Table 2.13)	127
2.1.15. Other reporters for mammalian transfection	128
2.1.16. Peptidyl-tRNA release assay templates (Table 2.15)	128
2.2. Mammalian cell transfection procedures	130
2.3. Dual luciferase assay	133
2.4. Detection of eIF2α phosphorylation	136
2.5. qPCR quantification of luciferase mRNA levels	139
2.6. <i>In vitro</i> transcription of mRNAs	142
2.7. <i>In vitro</i> translation of mRNAs	146
2.8. Western blot detection of translation products	147
2.9. Peptidyl-tRNA release assay	149
3. Results	152
3.1. A noncanonical uORF in the plant GGP mRNA contains highly conserved amino acid motifs	153
3.1.1. Significance and rationale for study	153
3.1.2. The GGP uORF contains highly conserved motifs	154
3.2. GDP-L galactose phosphorylase (GGP) mRNA translation is regulated by ascorbic acid in a uORF-dependent manner in mammalian cells	158
3.2.1. Significance and rationale for study	158
3.2.2. GDP-L galactose phosphorylase (GGP) mRNA translation is regulated by ascorbic acid in a uORF-dependent manner	159

3.2.3. Conservation of the motifs is essential for regulation	164
3.2.4. The GGP uORF is specifically sensitive to ascorbic acid <i>in vivo</i> .	168
3.2.5. Treatment of mammalian cells with ascorbic acid does not trigger phosphorylation of eIF2 α	173
3.3. L-ascorbic acid enhances translation of the ACG-initiated GGP uORF in mammalian cell culture (<i>in vivo</i>)	177
3.3.1. Significance and rationale for study	177
3.3.2. Ascorbic acid controls GGP uCC translation <i>in vivo</i>	178
3.3.3. Conservation of the GGP uCC motifs is essential for initiation regulation	181
3.3.4. Conservation of the motif spacing is essential for regulation of uCC translation	188
3.4. Ascorbate regulates uCC translation in rabbit reticulocyte lysate (<i>in vitro</i>)	194
3.4.1. Significance and rationale for study	194
3.4.2. Optimization of RRLs for <i>in vitro</i> translation	195
3.4.3. Ascorbate regulates uCC translation in rabbit reticulocyte lysate (<i>in vitro</i>)	201
3.4.4. The uCC motifs are essential for ascorbate inhibition of uCC- FLuc translation	206
3.4.5. uCC-FLuc mRNA translation is inhibited specifically by L- ascorbate	209

3.5. Ascorbate regulation of the uCC is mediated by ribosome pausing	
during translation elongation	215
3.5.1. Significance and rationale for study	215
3.5.2. uCC regulation is due to ribosome pausing during translation	215
3.5.3. Ribosome profiling in plants reveals ribosome pausing on the conserved C-terminal glycine repeat of the uCC	223
3.6. Ribosome queuing on the uCC is essential for enhancement of	
initiation at the uCC start site	228
3.6.1. Significance and rationale for study	228
3.6.2. Lengthening the GGP uCC impairs ascorbate regulation in mammalian cell culture (<i>in vivo</i>); however, the extension may interfere with ascorbate sensing	230
3.6.3. Reducing ribosome loading with stem loop structures impairs GGP regulation in mammalian cell culture (<i>in vivo</i>)	235
4. Discussion	243
4.1. The first described uCC: control of <i>AZINI</i> through a polyamine	
regulated uCC	244
4.2. GDP-L galactose phosphorylase (GGP) mRNA was an ideal	
candidate for regulation by a uCC that responds to L-ascorbic acid	248
4.3. How does the GGP uCC detect changes in L-ascorbic acid	
concentration?	253
4.4. Are there other uCCs and what is the advantage of a uCC?	259
4.5. What can we learn from the uCC model?	269

5. eIF2γ mutations and MEHMO syndrome	274
5.1. Introduction to the role of eIF2γ in translation	275
5.2. Materials and methods	282
5.2.1. <i>GCD11</i> mutant plasmids	282
5.2.2. Yeast transformation	283
5.2.3. Generation of <i>GCD11</i> mutant yeast strains	285
5.2.4. β -galactosidase assay	288
5.3. Results	291
5.4. Discussion and conclusions	303
6. References	313
6.1. Appendix: Glossary of abbreviations	314
6.2. Bibliography	319
7. Curriculum vitae	369

List of Tables

Table 2.1. Primers used in the study	92
Table 2.2. The <i>in vitro</i> translation uCC-FLuc fusion reporters used in the study .	108
Table 2.3. <i>In vitro</i> translation uORF reporters generated for study	110
Table 2.4. The <i>in vitro</i> translation FLuc controls used for RRL optimization	112
Table 2.5. The RLuc reporters used in <i>in vivo</i> assays to study uCC regulation ...	117
Table 2.6. The p2Luc based uCC-FLuc fusions used in <i>in vivo</i> assays	118
Table 2.7. The p2Luc based uCC substitution-FLuc fusions used in <i>in vivo</i> assays	119
Table 2.8. The phRL reporters containing uCC substitutions	120
Table 2.9. The p2Luc based FLuc reporters containing alanine insertions within the uCC	122
Table 2.10. The p2Luc based FLuc reporters containing the GGP uCC regulating FLuc and a stem-loop to reduce ribosome loading	124
Table 2.11. The phRL based RLuc reporters containing the GGP uCC regulating RLuc and a stem-loop to reduce ribosome loading	125
Table 2.12. The phRL based Trip12-uCC extensions used in <i>in vivo</i> assays	127
Table 2.13. The p2Luc based Trip12-uCC-FLuc fusion control	127
Table 2.14. The phRL based AZIN-AAP controls used in <i>in vivo</i> assays (Fig. 2.4D)	128
Table 2.15. The FLAG tagged peptidyl-tRNA stalling assay templates	129
Table 2.16. Components of the FLuc Assay Buffer	134
Table 2.17. Components of the RLuc Assay Buffer	135

Table 2.18. Composition of the <i>in vitro</i> transcription reaction	143
Table 2.19. Composition of the ribomax buffer	144
Table 2.20. Composition of a Promega Flexi RRL reaction	146
Table 3.1. Relative reporter mRNA levels – related to Fig. 3.2.	162
Table 3.2. Relative reporter mRNA levels – related to Fig. 3.3.	167
Table 3.3. Relative reporter mRNA levels – related to Fig. 3.4.	171
Table 3.4. Relative reporter mRNA levels – related to Fig. 3.6.	181
Table 3.5. Relative reporter mRNA levels – related to Fig. 3.7.	187
Table 3.6. Relative reporter mRNA levels – related to Fig. 3.8.	192
Table 3.7. Relative reporter mRNA levels – related to Fig. 3.21.	242
Table 5.1. The yeast <i>GCD11</i> plasmids used in this study	283
Table 5.2. Stock 10X LiOAc solution	284
Table 5.3. The 44% PEG stock solution	284
Table 5.4. The yeast strains generated in this study	287
Table 5.5. β -galactosidase assay lacZ reporters	288
Table 5.6. Composition of the β -galactosidase assay breaking buffer	290
Table 5.7. Composition of the β -galactosidase assay Z-buffer	290

List of Figures

Figure 1.1 Scanning model of eukaryotic translation initiation	31
Figure 1.2. Regulated translation reinitiation controls <i>GCN4</i> and <i>ATF4</i> expression	49
Figure 1.3. Polyamines and uORF-mediated autoregulatory circuits	64
Figure 1.4. A conserved uCC regulates <i>AZIN</i> expression through <i>eIF5A</i>	71
Figure 1.5. The L-ascorbate biosynthetic pathways in plants and animals	78
Figure 1.6. The L-ascorbic acid biosynthetic enzyme GGP is negatively regulated L-ascorbic acid via a proposed uCC	81
Figure 2.1. Sequence information for the <i>A. thaliana</i> GGP uCC	97
Figure 2.2. Templates used for <i>in vitro</i> transcription of mRNAs used in RRL assays	104
Figure 2.3. Templates used for <i>in vitro</i> transcription of FLuc mRNAs used to optimize RRL assays	111
Figure 2.4. Plasmid reporters used in mammalian cell assays	114
Figure 3.1. The GGP uCC is highly conserved across plants	156
Figure 3.2. GGP mRNA translation is regulated by ascorbate in a uCC-specific manner <i>in vivo</i>	160
Figure 3.3. Conservation of the uCC amino acid motifs is essential for ascorbate regulation	165
Figure 3.4. Regulation of GGP is specific to ascorbate and ascorbate does not regulate AAP	169

Figure 3.5. Ascorbate-P does not regulate GGP mRNA translation through eIF2 α phosphorylation	175
Figure 3.6. Ascorbate controls GGP uCC translation <i>in vivo</i>	179
Figure 3.7. Conservation of the motifs is essential for ascorbate control of GGP uCC translation <i>in vivo</i>	184
Figure 3.8. Conservation of motif spacing is essential for regulation by ascorbate	190
Figure 3.9. Determination of the linear range for incubation time and amount of FLuc mRNA translated in RRL	197
Figure. 3.10. Optimization of RRL magnesium concentration	200
Figure 3.11. Ascorbate regulates uCC elongation in RRL	202
Figure 3.12. The reduction in FLuc levels in the presence of ascorbate is due to a uCC-specific reduction in uCC-FLuc translation <i>in vitro</i>	204
Figure 3.13. Conservation of the uCC motifs is essential for ascorbate control of GGP uCC elongation <i>in vitro</i>	208
Figure 3.14. Chemical structures of L-ascorbic acid and chemical analogs	211
Figure 3.15. The uCC-specific reduction in uCC-FLuc expression is unique to ascorbate	213
Figure 3.16. uCC regulation is due to ribosome pausing during translation	218
Figure 3.17. Ribosome profiling reveals a ribosome stall at the C-terminal glycine residue of the last conserved motif in the GGP uCC	225
Figure 3.18. Lengthening the GGP uCC leads to a loss of regulation <i>in vivo</i>	231

Figure 3.19. Inserting <i>Trip12</i> sequences to lengthen the GGP uCC impairs ascorbate sensing	234
Figure 3.20. Insertion of a SL into the 5' leader of a GGP-Rluc reporter in a phRL vector (CMV promoter) did not significantly impact ascorbate regulation	237
Figure 3.21. Insertion of a SL into the 5' leader of a GGP-Rluc reporter under the control of an SV40 promoter impairs ascorbate regulation <i>in vivo</i>	240
Figure 4.1. Proposed model of <i>GGP</i> regulation by the uCC and L-ascorbic acid	249
Figure 4.2. <i>GGP</i> uCC and other metabolite-sensing nascent peptides.....	254
Figure 5.1. Putative MEHMO mutation S108R maps to an unstructured loop in eIF2 γ	278
Figure 5.2. Schematic of the eIF2 γ (yeast GCD11) mutations generated for this study	292
Figure 5.3. Mutations at residue D167 of yeast eIF2 γ do not impair cell growth	295
Figure 5.4. Mutations at residue D167 of yeast eIF2 γ do not affect <i>GCN4</i> expression	297
Figure 5.5. Mutations at residue D167 of yeast eIF2 γ do not alter translation start site fidelity	300

1. Introduction

1.1. Regulation of gene expression

Eukaryotic genomes contain tens of thousands of protein coding genes. Humans are estimated to have 20,000-25,000 protein coding genes and ten times as many total mRNA transcripts (Salzberg, 2018). Gene regulation allows cells to control which genes in the cell's DNA are expressed, at what time, and at what level. Regulation of gene expression is essential for cells to adapt to changing stresses and for the development of different cell types in complex organisms. In eukaryotes, gene expression involves many steps and regulation can occur at any of these steps. Some of the steps that can be regulated are at the level of chromatin accessibility and transcription, at RNA processing, RNA stability, translation, and at protein activity and stability.

In eukaryotes, DNA is packaged into chromatin inside the cell nucleus (Kornberg, 1974). Transcription of DNA into RNA is a key regulatory point for many genes. Sets of transcription factor proteins bind to specific DNA elements in or near a gene and promote or repress its transcription in response to cellular signals (Matsui et al., 1980). The number of promoters targeted by any given sequence-specific regulator ranges from just a few to several hundred, or even several thousand in complex organisms like metazoans (Venters and Pugh, 2009). The transcriptional activators then recruit transcriptional coactivators to modify the local chromatin environment and promote the assembly of the transcription pre-initiation complex, which is composed of general transcription factors and RNA Polymerase II (Pol II) (Roeder, 1996). After Pol II is recruited to a promoter, transcription proceeds through initiation, elongation, and termination. Unlike bacterial mRNAs, eukaryotic mRNAs require processing in the nucleus before they are fully translatable. The mRNAs are capped, introns are spliced out, and the 3' end is modified (Proudfoot et al.,

2002). mRNA nuclear export is directed and coordinated with processing and assembly into messenger ribonucleoprotein particles (mRNPs). During capping, splicing, and polyadenylation, many RNA-binding and modifying proteins are recruited to the transcripts including the transcription-export (TREX) complex, which is essential for mRNA export (Carmody and Wentz, 2009; Cheng et al., 2006; Kohler and Hurt, 2007; Masuda et al., 2005). TREX acts as a quality control mechanism and also appears to have gene regulatory effects (Kohler and Hurt, 2007). In higher eukaryotes the TREX complex is poorly recruited to mRNAs that lack splicing components or the 5' cap (Carmody and Wentz, 2009; Cheng et al., 2006). mRNA export utilizes Nxf1-Nxt1, which is not dependent on the Ran-GTP gradient (Herold et al., 2000), and is recruited to the mRNP through TREX (Carmody and Wentz, 2009). After export from the nucleus, the mRNA can be translated in the cytosol or directed to the endoplasmic reticulum (ER) if the encoded protein is destined for secretion.

The stability of an mRNA molecule in the cytosol affects how many times it can be translated and thus how many proteins will be produced per mRNA. The steady-state level of any mRNA in a cell is a balance between its synthesis via transcription and its degradation by the mRNA decay machinery. Multiple decay pathways have been characterized for mRNAs and different mRNAs can be degraded by different distinct pathways, but the same mRNA can also be degraded by different pathways depending on cellular conditions (Guhaniyogi and Brewer, 2001). The rate at which an mRNA is decayed is dependent on *cis*-acting elements within the mRNA and *trans*-acting factors that bind the elements. mRNA stability is also regulated through translation to modulate gene expression and as a quality control mechanism (Shoemaker and Green, 2012; Stoecklin

and Muhlemann, 2013). The role of mRNA features and translation in mRNA stability will be discussed in more detail in the next section.

Protein synthesis is an essential component of the gene expression pathway and is a key element in its control. Regulation of translation is essential to most cellular processes and is critical to maintain homeostasis in the cell and the organism. In general, the synthesis rate of a protein is proportional to the concentration and translational efficiency of its mRNA (Hershey et al., 2012). The majority of eukaryotic mRNAs have long half-lives (>two hours) (Raghavan et al., 2002), so rapid regulation of the cellular levels of the proteins they encode is achieved by controlling mRNA translational efficiency and protein degradation rates rather than the mRNA level. Translational control is the only way to regulate gene expression under some conditions such as during early viral infection, before viral transcription has started and the RNAs packaged in the virus are the sole source of viral protein synthesis (Walsh et al., 2013), and in cells lacking active transcription such as oocytes and reticulocytes (Richter and Lasko, 2011). Protein synthesis also accounts for a large proportion of a cells' energy budget, especially for cells that are growing or are biosynthetically active. In addition, because protein synthesis is integrated with all aspects of cellular metabolism, dysregulation of translation contributes to a variety of disease states including cancers and neurological diseases (Le Quesne et al., 2010). Translation and translational control will be discussed in more detail in a later section.

After synthesis, proteins are still subject to regulation. Proteins can undergo regulated modifications that affect the activity or behavior of the protein, some proteins are regulated by small molecules or protein-protein interactions, and proteins undergo degradation. Post-translational modifications (PTMs) are mostly catalyzed by enzymes that

recognize specific target sequences in specific proteins. The most common post-translational modifications are cleavage of precursor proteins, formation of disulfide bonds, or covalent addition or removal of low-molecular-weight groups, such as a phosphoryl group. As of 2015, over 200 different types of post-translational modifications have been identified (Duan and Walther, 2015). Post-translational modifications are fundamental in regulating protein folding, interactions, and the functional state of the protein. Post-translational modifications, such as phosphorylation, are reversible by the action of specific deconjugating enzymes. In the case of phosphorylation by phosphatases. The balance between modifying and reversing the modification allows for rapid and energy efficient control of protein function; control by protein degradation and *de novo* synthesis are much more time and energy intensive.

Many enzymes are controlled by conformational changes that alter catalytic activity. Frequently, the conformational change is the result of small molecule binding. This type of regulation is common in biosynthetic pathways, where the end product of the pathway inhibits the enzyme that catalyzes the first step in its synthesis by feedback inhibition (Alves and Savageau, 2000). This ensures an adequate supply of the end product while preventing synthesis of excess product.

Finally, protein expression can be regulated through proteolysis, or protein degradation. A common means of regulated protein degradation involves post-translational conjugation of ubiquitin to select Lys residues on target proteins (Hershko et al., 1983). Degradation signals within the protein are referred to as degrons and can be activated or inactivated via covalent modifications or sequestration (Ravid and Hochstrasser, 2008). Ubiquitination involves three main steps and types of enzymes: ubiquitin activation by

ubiquitin-activating enzymes (E1s); conjugation , or transfer of ubiquitin from E1 to E2, by ubiquitin-conjugating enzymes (E2s); and ligation by ubiquitin ligases (E3s). Degradation is also regulated by the activity, expression, and localization of the E3 ligases that recognize substrates of the ubiquitin system and conjugate ubiquitin to them (Varshavsky, 2005). There are over 500 distinct E3 ligases in the average mammal (Varshavsky, 2005). After ubiquitination, the protein is processively degraded into short peptides by the ATP-dependent 26S proteasome.

1.2. mRNA features and RNA stability

Eukaryotic genomes encode large numbers of RNases, many with overlapping activities, making redundancy a general feature of RNA degradation pathways (Houseley and Tollervey, 2009). With a few exceptions, mutating a single RNA degradation enzyme does not result in a complete block to RNA degradation, indicating that multiple enzymes are capable of targeting the same RNAs. In a wild-type cell, different mRNAs can be degraded by distinct pathways while a given mRNA can also be degraded by different, seemingly redundant pathways, depending on cellular conditions (Guhaniyogi and Brewer, 2001).

As discussed above, the processing of pre-mRNAs includes the addition of a 5' cap and the modification of the 3' end by the addition of a non-templated poly(A) tail. In addition to being required for efficient translation, the 5' cap and poly(A)-tail confer resistance to non-specific degradation in the cytoplasm by exonucleases (Stoecklin and Muhlemann, 2013). Regulated decay of mRNAs is influenced by components of the mRNA including the 5'-cap, the 5'-untranslated region (5'-UTR), the protein coding region, 3'-UTR, and the 3'-poly(A) tail, and *trans*-acting factors that target these specific features (Guhaniyogi and Brewer, 2001).

Transcripts generated by RNA Pol II can be degraded in the nucleus or in the cytosol. In the nucleus, capping, transcription termination, and mRNA export act as a quality control mechanism. During transcription, pre-mRNAs are capped with an N⁷-methylated guanosine linked to the first nucleotide of the RNA by a reverse 5' to 5' triphosphate linkage (Ramanathan et al., 2016). The m⁷G cap is necessary for cap-dependent initiation of protein synthesis and also acts as a quality control mechanism. The

mRNA cap protects against 5' to 3' exonuclease cleavage and is an identifier for recruiting protein factors for pre-mRNA splicing (Fresco and Buratowski, 1996; Inoue et al., 1989), polyadenylation (Flaherty et al., 1997), and nuclear export (Nojima et al., 2007; Ramanathan et al., 2016). In mammalian cells, a trifunctional enzyme, DXO/Dom3Z, specifically de-caps an unmethylated cap and degrades the 5' monophosphate-terminated RNA with its 5' to 3' exonuclease activity (Jiao et al., 2013).

Pol II transcription termination involves multiple steps and features: the poly(A) signal, pre-mRNA cleavage (at the poly(A) site or downstream), pre-mRNA degradation, and specific terminator elements (Porrua et al., 2016; West et al., 2008). Utilizing the human β -globin gene, West et al. found that transcription of the poly(A) signal renders Pol II termination competent. As Pol II synthesizes the β -globin transcript through and beyond the terminator element, the mRNA is cotranscriptionally cleaved, and the 3' portion is degraded by a 5'-3' exonuclease that releases Pol III from the DNA template (West et al., 2008). Termination can occur hundreds or thousands of nucleotides downstream from the poly(A) signal, so cleavage of the mRNA occurs near the poly(A) signal. Nearly every known mRNA contains a polyadenylation signal sequence AAUAAA, or variants, 10-30 bases upstream of the cleavage/polyadenylation site (Colgan and Manley, 1997). Cleavage of the 3' end is coupled to poly(A) addition by a large complex of multisubunit proteins. In yeast, the processes are mediated by the cleavage and polyadenylation factor (CTF) that is organized into three modules that each mediate a key function: a nuclease module that cleaves the transcript; a polymerase module that binds other components including poly(A) polymerase; and a phosphatase that regulates 3' end processing (Stewart, 2019).

At steady state, the length of the poly(A) tail on a cytoplasmic mRNA varies between approximately 70 nt in *Saccharomyces cerevisiae* and over 200 nt in mammalian cells (Wahle and Winkler, 2013). The poly(A) tail serves as a binding site for poly(A)-binding protein (PABP). PABP has four RNA recognition motifs (RRM1-4) and the first RRM (1-2) bind the poly(A) tail and are thought to enhance translation by binding the eukaryotic initiation factor (eIF) eIF4G (Machida et al., 2018; Safaee et al., 2012). PABP also has binding sites for other molecules important for translation and metabolism of RNA including for Paip2, which inhibits translation by competing with poly(A) RNA and eIF4G for binding to PABP (Karim et al., 2006).

Correct nuclear processing and recruitment of export factors target an mRNA for export to the cytosol. Defects in mRNP assembly prevent the association of export factors, and capping and splicing are important for the recruitment of the TREX complex (Carmody and Wente, 2009; Cheng et al., 2006; Zhou et al., 2000). Export failure leads to degradation of the mRNA by 5' and/or 3' exonucleases (Houseley and Tollervey, 2009).

After export to the cytosol the mRNA can be degraded as part of translation quality control or by 3'-5' decay or 5'-3' decay. mRNA stability is regulated through translation to modulate gene expression and as a quality control mechanism (Shoemaker and Green, 2012; Stoecklin and Muhlemann, 2013). Ribosomes that arrest during translation, usually on aberrant mRNAs, trigger mRNA quality control pathways that lead to mRNA decay. In eukaryotes, examples of aberrant translation events include premature stop codons in the middle of an open reading frame (ORF) that trigger nonsense-mediated decay (NMD) (Amrani et al., 2004); translation of the poly(A) tail at the end of the mRNA due to lack of a stop codon triggers non-stop decay (Frischmeyer et al., 2002); and kinetic traps, such as

hairpins or truncations that prevent further translation, trigger no-go decay (Doma and Parker, 2006). In each case, the ribosome and associated factors are thought to ‘sense’ the defect in the mRNA and recruit machinery to begin the process of mRNA decay, proteolysis of the nascent peptide (NC), and recycling of the stalled ribosome (Radhakrishnan and Green, 2016).

In the cytoplasm, mRNAs undergo progressive deadenylation at rates that are specific for each mRNA species (Houseley and Tollervey, 2009). The mRNP contains information that defines its lifespan. The best characterized stability/instability elements are AU-rich elements (AREs), which are bound by proteins that extend or shorten mRNA lifespan in response to cellular signals (Houseley and Tollervey, 2009), and miRNA recognition sites (Wahle and Winkler, 2013). Stoeklin et al. found that ARE-mediated mRNA decay occurs in a 5’-3’ direction in a processing body that includes decapping enzymes, the 5’-3’ exonuclease Xrn1, and Lsm1-7 (Stoeklin et al., 2006). The majority of cytoplasmic deadenylation is associated with the Ccr4-Not and Pan2-Pan3 complexes (Wahle and Winkler, 2013). Following deadenylation, the mRNA can be degraded in a 3’-5’ direction by the exosome. Deadenylation is a trigger for mRNAs decapping, shutting down mRNA translation and activating mRNA degradation by Xrn1 (Franks and Lykke-Andersen, 2008). Cytoplasmic decapping is catalyzed by Dcp2 in conjunction with a coactivator Dcp1 and is stimulated by decapping enhancer proteins. The Edc family of enhancers recruit Dcp2 cap-binding and catalysis (Franks and Lykke-Andersen, 2008). Pat1 and Dhh1 appear to activate decapping at least partly by repressing translation. Dhh1 is hypothesized to promote destabilization of the eIF4F-mRNA cap complex (Coller and Parker, 2005; Franks and Lykke-Andersen, 2008). Dhh1 is also proposed to be a sensor of

ribosome speed, targeting mRNAs with slow moving ribosomes for degradation (Radhakrishnan et al., 2016).

The eukaryotic RNA exosome has nuclear and cytoplasmic forms that are defined by unique subunit compositions that interact with factors in these compartments (Zinder and Lima, 2017). In *S. cerevisiae* the cytoplasmic exosome includes a nine-subunit core (Exo9) that interacts with Dis3 to form a 10-subunit complex. The Exo9 core lacks catalytic activity, while Dis3 catalyzes endoribonuclease and processive 3'-5' exoribonuclease activities (Dziembowski et al., 2007; Lebreton et al., 2008; Liu et al., 2006; Zinder and Lima, 2017). Despite appearing redundant with the 5'-3' decay pathways, the cytoplasmic exosome contributes to translation-dependent mRNA surveillance.

1.3. Protein biosynthesis

Translation of a eukaryotic mRNA to produce a protein consists of four steps: initiation, mediated by initiation factors; elongation, mediated by elongation factors; termination, mediated by release factors; and recycling. Protein synthesis is largely regulated at the initiation stage rather than during elongation or termination (Jackson et al., 2010). Regulation of initiation allows rapid and reversible control of gene expression, along with spatial control. During translation initiation, an elongation-competent 80S ribosome is assembled on an mRNA at the initiation codon, which is base paired with the anticodon loop of an initiator transfer RNA (Met-tRNA^{Met}) in the ribosomal peptidyl (P) site. Initiation requires at least twelve eukaryotic initiation factors and has three main steps (Hinnebusch, 2014; Hussain et al., 2014; Jackson et al., 2010). The first step is the formation of a 43S pre-initiation complex (PIC); the second step is binding of the PIC to an mRNA - forming a 48S PIC - and selection of the start codon; and the third step is the joining of a 60S subunit to the 48S PIC to form an elongation competent 80S ribosome.

Marilyn Kozak lists five structural features of an mRNA that influence translation initiation: the m⁷G cap; the context nucleotides flanking the start codon; the position of the start codon relative to the 5' end of the mRNA (whether it is the first start codon) (discussed in a later section); secondary structure upstream and downstream from the start codon; and the length of the leader (Kozak, 1991b). As discussed above, the m⁷G cap is ubiquitous and serves as a determinant of mRNA quality control as well as increasing translation efficiency (Kozak, 1991b; Ramanathan et al., 2016). Interestingly, a guanylated, unmethylated cap is sufficient to protect an mRNA from 5'-exonuclease but N-7 methylation is essential for efficient translation (Horikami et al., 1984; Kozak, 1991b). The

m⁷G cap facilitates translation and scanning by recruiting the cap-binding complex eIF4F, which attaches to the cap to activate the mRNA for 43S PIC attachment (Hinnebusch, 2014). The cap binding complex eIF4F is comprised of the cap binding protein eIF4E, eIF4G, and the RNA helicase eIF4A. eIF4G is a scaffold with binding domains for mRNA, eIF4E, eIF4A, PABP, eIF3 (mammals), and eIF1 and -5 (budding yeast) (Hinnebusch, 2014; Hinton et al., 2007). The eIF4G scaffold allows assembly of a stable circular mRNA-protein complex through interactions with eIF4E and PABP, which bind the 5' and 3' ends, respectively, of the mRNA (Hinton et al., 2007). eIF4G allosterically activates eIF4A and recruits it to the cap, where eIF4A locally unwinds secondary structure in the mRNA (Hinnebusch, 2014). mRNA unwinding along with eIF4G interactions with eIF3, eIF5, or eIF1 promote 43S PIC attachment to the 5' end of the mRNA (Hinnebusch, 2014).

Kozak proposed the scanning model of eukaryotic translation initiation in 1978 (Kozak, 1978). In this model, the 40S subunit, along with initiation factors, binds the mRNA near the 7-methyl guanosine cap at the 5' end and moves primarily in a 5' to 3' direction until it encounters an AUG codon (Fig. 1.1) (Kozak, 1978). Particular sequences immediately surrounding the AUG codon (the context nucleotides) enhance selection of the AUG codon by the scanning PIC, but a 5' proximal AUG codon that deviates sufficiently from the optimum context will be bypassed in an event referred to as leaky scanning (Hinnebusch, 2014). In her 1984 paper, Kozak proposed the 'Kozak consensus sequence', CC(A/G)CCAUGG (A/G represents A or G and AUG represents the initiating codon) (Kozak, 1984). Kozak examined the effects of point mutations in the nucleotides flanking the AUG codon by measuring the yield of a cloned preproinsulin gene in transfected cells, and found that mutations within the sequence can modulate protein

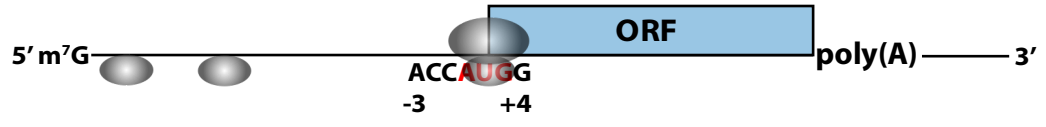


Figure 1.1. Scanning model of eukaryotic translation initiation

Figure 1.1. Scanning model of eukaryotic translation initiation

A 43S pre-initiation complex consisting of a 40S ribosomal subunit with bound eIF2–GTP–Met–tRNA_i^{Met} ternary complex and other initiation factors binds near the 5' m⁷G cap of an mRNA and then scans down the mRNA in search of a start codon. Selection of the start codon is mediated by base-pairing interactions between the anticodon of the tRNA_i^{Met} in the preinitiation complex and an AUG codon in the mRNA. Typically, the ribosome initiates translation at the most 5' proximal start codon it encounters during scanning; however, the sequence context of the start codon influences the efficiency of start codon selection with a purine at the -3 position and a G at +4 relative to the A of the AUG codon having the greatest impact on start codon selection.

expression over a 20-fold range (Kozak, 1986b). A purine at the -3 position and a G at the +4 position were found to be the most important for AUG codon selection, and Kozak confirmed that ACCAUGG is the optimal sequence for initiation by eukaryotic ribosomes (Fig. 1.1) (Kozak, 1986b). However, only 0.2% of vertebrate genes contain the precise 'Kozak consensus sequence' (Cavener and Ray, 1991), and the preferred nucleotide sequence around the start codon varies among different species (Nakagawa et al., 2008). For example, an early study performed prior to completion of plant genome sequences, identified the most common sequences for monocots as GCGGC(A/C)(A/G)(A/C)CAUGGCG and AAAAAAA(A/C)AAUGGCU for dicots (Joshi et al., 1997). More recent studies reveal that, while preferred nucleotide sequences are diverse among species, patterns of nucleotide bias roughly reflect evolutionary

relationships between species; and there are strong biases for particular nucleotides at certain positions (Nakagawa et al., 2008). There is a strong bias for A/G at position -3, A/C at position -2, and C at position +5 across all species. Nakagawa et al. also found that the signal for these particular nucleotides at the -3, -2, and +5 positions was especially high in genes with high expression levels, suggesting that the nucleotides are responsible for regulating translation initiation (Nakagawa et al., 2008).

In the past, protein coding sequences were traditionally defined as open reading frames that begin with a universal AUG start codon and terminate with one of three stop codons; however, it has been known since the 1980s that translation can initiate at codons other than AUG, although at lower efficiency (Clements et al., 1988; Kearse and Wilusz, 2017; Peabody, 1987, 1989; Zitomer et al., 1984). Most cases of alternate initiation occur at near-cognate codons that differ from AUG by only one nucleotide. The prevalence of near-cognate initiation was revealed by ribosome profiling when thousands of novel potential non-AUG initiation events were identified (Ingolia et al., 2009). The efficiency of initiation varies by codon, with CUG generally being the most efficient, followed by GUG, ACG, and AUU (Clements et al., 1988; Ivanov et al., 2010b; Kearse and Wilusz, 2017; Peabody, 1989; Wei et al., 2013); however, efficiency varies especially in *in vitro* assays where the method of preparation and ionic concentrations vary (Kearse and Wilusz, 2017). In a study of near-cognate initiation efficiency using luciferase reporters in HEK293T cells, CUG has an initiation efficiency of about 19%, GUG of 9%, UUG of 2%, ACG of 7%, and AUU of 3% compared to an AUG codon (Ivanov et al., 2010b).

The length and structure of the 5'-UTR also influence translation initiation efficiency (Hinnebusch, 2014; Kozak, 1991a, b). Very short 5'-UTRs increase leaky

scanning and reduce translational efficiency. In yeast, shortening the 5'-UTR of *PGK1* from 45 to 21 nucleotides reduced translational efficiency by 50% (Hinnebusch, 2011; van den Heuvel et al., 1989). Using wheat germ extract and mRNAs containing two AUG codons in-frame with the chloramphenicol acetyltransferase (CAT) coding sequence, Kozak demonstrated that as the distance of the first AUG codon from the m⁷G cap was decreased from 32 to 3 nucleotides, the yield of protein from the first AUG codon decreased, with a corresponding increase in production from the second AUG codon (Kozak, 1991a). When the leader length was only three nucleotides to the first AUG codon, the majority of protein was produced from AUG #2, at six nucleotides similar levels of protein were produced from AUG #1 and AUG #2, and at 12 nucleotides about twice as much protein was produced from AUG #1 compared to AUG #2. At 32 nucleotides, leaky scanning was almost completely absent and only trace amounts of protein were initiated at AUG #2 (Kozak, 1991a). Similar effects are exploited in yeast to permit translation of multiple proteins from the same *MOD5* mRNA initiated either inefficiently at a cap-proximal AUG codon or at the second AUG codon by ribosomes that skip the first AUG codon (Hinnebusch, 2011; Slusher et al., 1991). The *MOD5* isoform initiated from the first AUG codon is imported into the mitochondria, while protein initiated from the second AUG codon functions in the cytoplasm (Gillman et al., 1991; Najarian et al., 1987; Slusher et al., 1991).

Stable secondary structures impede the progression of the PIC along the 5'-UTR, possibly because RNA hairpins cannot thread through the 40S mRNA entry channel (Hinnebusch, 2011). Kozak studied the effect of RNA hairpins on translation initiation using plasmids encoding preproinsulin in mammalian cell culture (Kozak, 1986a). Hairpin

structures of varying stability were inserted upstream of the AUG initiator codon and protein production was monitored. Kozak found that a stable hairpin ($\Delta G = -50$ kcal/mol) reduced protein yield by 85-95% and that the hairpin inhibited translation even when it was placed in the middle of the 5'-UTR where it did not involve the cap or the AUG codon (Kozak, 1986a). In summary, in yeast and mammals, mRNAs with high translational efficiencies are expected to contain a 5'-UTR of about 32 nucleotides or greater; lack stable secondary structure, extra AUG codons and near-cognate codons in an optimum context; and to contain the functional AUG start codon in an optimum context (Hinnebusch, 2011).

Beginning translation requires a pool of separated ribosomal subunits generated by recycling post-termination ribosomal complexes (post-TCs) (Jackson et al., 2010). Recycling begins with termination upon recognition of a stop codon in the aminoacyl (A) site of the 80S ribosome by eukaryotic release factors (eRF) eRF1 and GTP-bound eRF3. GTP is hydrolyzed and eRF3 dissociates, leaving eRF1 in the A site ready to hydrolyze the peptidyl-tRNA in the P site (Pisarev et al., 2007; Pisarev et al., 2010). Termination and recycling are coupled through ABCE1 (mammals)/Rli1 (yeast), which participates in peptide release and separating the 80S ribosome into a free 60S subunit and a 40S subunit bound to deacylated tRNA and mRNA (Dong et al., 2004; Pisarev et al., 2007; Young et al., 2015a). Next, the deacylated tRNA is removed from the ribosome with dissociation of the 40S-mRNA complex in a process mediated by Ligatin and MCT-1/DENR (Jackson et al., 2012; Skabkin et al., 2010).

Before translation of an mRNA can begin, multiple complexes are assembled on and off the mRNA. The ternary complex (TC) is essential for initiation and is composed of eIF2, GTP, and Met-tRNA_i^{Met}. Scanning requires an open 43S PIC conformation that is

stabilized by eIF1 and eIF1A binding to the small ribosomal (40S) subunit along with eIF5 and eIF3 (Hinnebusch, 2014). The open PIC is preloaded with Met-tRNA_i^{Met} by binding the ternary complex and forming the 43S PIC. As described above, attachment of the PIC to the mRNA is facilitated by the eIF4F complex, comprised of the cap-binding protein eIF4E, eIF4G, the RNA helicase eIF4A, and PABP (Hinnebusch, 2011).

After attachment to an mRNA, the 43S PIC scans the mRNA in search of an AUG codon, or less frequently a near-cognate codon, in a suitable context. Scanning by the PIC involves two linked processes: the first is unwinding of 5'-UTR secondary structures, and the second is ribosomal movement along the mRNA. Scanning a 5'-UTR with even weak secondary structures requires ATP and the factors eIF4A, eIF4G, and eIF4B along with additional DEAD box helicases (Hinnebusch, 2011, 2014; Jackson et al., 2010). Using *in vitro* translation of mRNAs with different 5'-UTR lengths, Berthelot et al. estimated that scanning occurs at about eight bases per second (Berthelot et al., 2004). Using the same method, Bethelot et al. and Vassilenko et al. also found that scanning exhibits a strong 5' to 3' directional bias (Berthelot et al., 2004; Hinnebusch, 2014; Jackson et al., 2010; Vassilenko et al., 2011).

An essential part of the scanning process is the capacity for the 43S PIC to bypass AUG codons in poor sequence contexts, as well as near-cognate triplets, in the 5' leader, so that the correct AUG start codon on the mRNA is selected. The PIC scans until it encounters an AUG codon in a proper sequence context, using complementarity with the anticodon of the initiator tRNA to select an AUG codon when it enters the P site of the 40S subunit (Hinnebusch, 2011). Base pairing of the Met-tRNA_i^{Met} anticodon with the AUG start codon triggers a halt to scanning and initiates selection of the translation start site

(Cigan et al., 1988a; Hinnebusch, 2014). In yeast *in vitro* assays, single-base substitutions at the second and third position of the AUG triplet reduce ternary complex affinity for the PIC/unstructured-mRNA complex by 10-50 fold (Kolitz et al., 2009). The reduction in affinity is attributed to a slower conformational change of the PIC to a more stable complex following the initial association between the ternary complex and the PIC. The conformational change corresponds to the switch from the open (scanning) PIC to the closed (non-scanning) PIC conformation, which increases the stability of ternary complex binding to reconstituted 43S/mRNA complexes upon AUG codon recognition (Hinnebusch, 2014; Kolitz et al., 2009; Passmore et al., 2007).

Studies by Hussain et al. and Llacer et al. give insight into how structural changes allow start codon selection through a change from the open to the closed PIC conformation (Hussain et al., 2014; Llacer et al., 2015). The binding of eIF1 and the eIF1A C-terminal tail to the 40S subunit stabilize the open PIC conformation that allows codon sampling during scanning (Hussain et al., 2014; Llacer et al., 2015; Saini et al., 2010). Llacer et al. found that eIF2 connects eIF1 and eIF1A with the Met-tRNA_i^{Met} on the 40S head, and that eIF3 appears to form connections between the mRNA entry and exit channels of the 40S subunit and the P site (Llacer et al., 2015). Hussain et al. found that eIF1 and eIF1A binding rotates the 40S head, likely facilitating ternary complex binding in the open PIC position where the initiator Met-tRNA_i^{Met} is not fully inserted into the 40S P site (Hussain et al., 2014). During scanning, eIF5 stimulates GTP hydrolysis by eIF2, but completion of the reaction with release of inorganic phosphate (P_i) is blocked by the presence of eIF1 in the complex. AUG codon specificity is ensured by an energetic penalty imposed by movement of the tRNA anticodon stem loop (ASL) to permit codon recognition that can normally

only be fully compensated for by a perfect codon:anticodon duplex (Hussain et al., 2014). In addition to moving the anticodon stem-loop, base pairing between the tRNA_i^{Met} and an AUG codon causes conformational changes in the PIC that likely weaken eIF1 association with the 40S subunit and trigger eIF1 release from the PIC. The disassociation of eIF1 from the 40S subunit leads to a closed, scanning-arrested conformation of the 40S subunit and to P_i release from eIF2-GDP (Algire et al., 2005; Hussain et al., 2014). eIF2-GDP and eIF5 leave the PIC and eIF5B-GTP binds to the 40S subunit and accelerates the rate of 60S subunit joining forming an 80S initiation complex (Hinnebusch, 2014). The 80S initiation complex contains Met-tRNA_i^{Met} in the P site with its anticodon base paired to the AUG start codon of the mRNA, and the ribosome is ready to begin the elongation phase of protein synthesis.

Following initiation, an 80S ribosome is poised on an mRNA ready to begin elongation, with the anticodon of Met-tRNA_i^{Met} in the P site base-paired with the start codon (Dever and Green, 2012). The second codon of the open reading frame is in the A site of the ribosome waiting for the cognate aminoacyl-tRNA to bind. Eukaryotic elongation factor (eEF) eEF1A binds aminoacyl-tRNA in a GTP-dependent manner and directs the tRNA to the ribosome A site. Codon recognition by the tRNA triggers GTP hydrolysis by eEF1A, which releases the factor and enables the aminoacyl-tRNA to be accommodated into the A site. Following accommodation of the aminoacyl-tRNA into the A site, peptide bond formation with the P-site peptidyl-tRNA occurs. The ribosomal peptidyl transferase center (PTC) consists primarily of conserved ribosomal RNA (rRNA) elements on the large ribosomal subunit that position the aminoacyl-tRNA substrates for catalysis.

During peptide bond formation, ratcheting of the ribosomal subunits triggers movement of the acceptor end of the P and A site tRNAs into the exit (E) and P sites, respectively, while the anticodon loops remain bound in the P and A sites such that the tRNAs adopt hybrid P/E and A/P states (Behrmann et al., 2015; Budkevich et al., 2011; Dever et al., 2018; Dever and Green, 2012; Korostelev et al., 2008; Laurberg et al., 2008; Moazed and Noller, 1989). Translocation of the tRNAs into the E and P sites is facilitated by the GTPase eEF2. Binding of eEF2-GTP stabilizes the hybrid state and promotes translocation. Hydrolysis of GTP enables the release of eEF2 from the post-translocation complex (Dever and Green, 2012). In the post-translocation state, a deacylated tRNA occupies the E site and the peptidyl-tRNA is in the P site. The A site is empty and is available for the next aminoacyl-tRNA in complex with eEF1A. Following GTP hydrolysis, eEF1A-GDP is released from the ribosome. The eukaryotic factor eEF1B acts as a guanine exchange factor (GEF) and catalyzes the displacement of GDP from eEF1A enabling GTP to bind, re-forming active eEF1A-GTP.

In addition to the elongation factors described above, the factor eIF5A is required for translation elongation. eIF5A was originally denoted as an initiation factor because it was characterized as stimulating the transfer of methionine from Met-tRNA_i^{Met} in the 80S initiation complex to the aminoacyl-tRNA analog puromycin, an assay that monitors the formation of the first peptide bond (Dever and Green, 2012; Kemper et al., 1976). eIF5A contains a conserved post-translational modification, where a conserved lysine residue is modified to hypusine in a two-step process. First, an N-butylamine moiety is transferred from spermidine to the ϵ -amino group of a specific lysine side chain to form deoxyhypusine, then a hydroxylation reaction completes the modification. eIF5A promotes

translation elongation, and by utilizing hydroxyl radical probing experiments, Gutierrez et al. localized eIF5A near the E site of the ribosome with the hypusine residue next to the acceptor stem of the P-site tRNA (Gutierrez et al., 2013). It is proposed that eIF5A stimulates the peptidyl transferase activity of the ribosome and facilitates the reactivity of poor substrates like proline (Gutierrez et al., 2013).

Codon optimality can determine elongation rates because of the non-uniform decoding rate of each of the 61 codons that encode a pool of 20 amino acids (Hanson and Coller, 2018; Sabi and Tuller, 2014). Multiple codons code for a single amino acid and these codons are recognized distinctly by the ribosome; therefore, individual codons can be optimal or non-optimal depending how efficiently the appropriate cognate tRNA can be selected from the cytoplasmic pool of tRNAs. Highly expressed genes show codon bias, disproportionately utilizing codons that are overrepresented in the transcriptome, that correlates with high tRNA levels, but codon optimality can also influence protein folding and translation fidelity (Hanson and Coller, 2018; Roth, 2012).

Sørensen and Pedersen used radio-labelled amino acid incorporation assays to show that codon identity affects the translation elongation rate (Hanson and Coller, 2018; Sorensen and Pedersen, 1991). The elongation rate for a particular codon is dependent on the tRNA pool because a ribosome will wait longer for a rare cognate tRNA to enter the ribosomal A-site. The kinetic cost to elongation is particularly high when the cognate tRNA species is rare compared with near-cognate tRNAs (Chu et al., 2011; Koutmou et al., 2015). Yu et al. used ribosome profiling and biochemistry experiments in *Neurospora crassa*, an organism that shows an unusually strong codon preference, to compare the overall rates of luciferase mRNA translation in a cell-free translation system to demonstrate that the rate

of ribosome elongation is impacted by codon choice (Koutmou et al., 2015; Yu et al., 2015). The luciferase mRNAs containing primarily optimized or de-optimized codons were translated, and the differences in the time until the first luciferase signal was detected was used as a metric of the ribosome elongation rate (Yu et al., 2015). Using ribosome profiling, they observed differential ribosome occupancy on luciferase reporter mRNA regions that were optimized or de-optimized for codon usage (Yu et al., 2015). Elongation rate also plays an important role in co-translational protein folding (Harding et al., 1999; Koutmou et al., 2015; Purvis et al., 1987; Thanaraj and Argos, 1996). Optimizing codons (Yu et al., 2015) or increasing rare tRNA abundance (Zhang et al., 2009) increases protein production, but in some cases decreases the production of functional protein. In addition to directly influencing elongation rates, codon usage can influence mRNA stability by targeting degradation factors to slow moving ribosomes (Radhakrishnan et al., 2016).

Changes in tRNA pools help maintain cellular homeostasis. tRNAs are aminoacylated by aminoacyl-tRNA synthetases that attach the appropriate amino acid to the correct tRNAs by catalyzing the esterification of the specific cognate amino acid or its precursor to its compatible cognate tRNAs. Twenty different aminoacyl tRNA synthetases, one for each natural amino acid, catalyze the aminoacylation of tRNAs (Ibba and Soll, 2004). Changes in the levels of tRNA pools and abundance of non-optimal codons in an mRNA play critical roles in promoting selective mRNA translation during amino acid starvation (Koutmou et al., 2015; Saikia et al., 2016). Selective aminoacylation of rare tRNA isoacceptors was first shown in *Escherichia coli* (Dittmar et al., 2005; Elf et al., 2003), and Saikia et al. demonstrated the same effect in human cell culture (Saikia et al., 2016). Using ribosome profiling, global codon usage analysis, quantitative tRNA arrays,

and luciferase reporter assays Saikia et al. studied differential mRNA translation in amino acid starved cells (Saikia et al., 2016). They found that the pools of aminoacylated tRNAs are altered in response to amino acid starvation to favor the translation of mRNAs involved in protein recycling.

Translation termination occurs upon recognition of a stop codon in the ribosomal A-site, followed by hydrolysis of the ester bond of the peptidyl-tRNA located in the P site, and release the nascent chain (Jackson et al., 2012). The stop codons (UAA, UGA, or UAG) are recognized by eukaryotic release factor eRF1, which binds to the A site in complex with GTP-bound eRF3. eRF1, a tRNA shaped protein with three domains (Dever and Green, 2012; Song et al., 2000), is responsible for high-fidelity stop codon recognition and peptidyl-tRNA hydrolysis, and its function is enhanced by the translational GTPase eRF3 (Dever and Green, 2012). The eRF1 N-terminal domain recognizes all three stop codons, the middle domain is functionally analogous to the acceptor stem of a tRNA and extends into the PTC to promote peptide release, and the C-terminal domain facilitates interactions with eRF3 (Song et al., 2000).

During termination, eRF1 and eRF3 bind to the A-site of a pre-termination complex as an eRF1-eRF3-GTP ternary complex. GTP is hydrolyzed and eRF3 dissociates, leaving eRF1 in the A site ready to hydrolyze the peptidyl-tRNA in the P site (Pisarev et al., 2007; Pisarev et al., 2010). Conformational changes enable the eRF1 middle domain to enter the PTC, and the peptidyl-tRNA is hydrolyzed following a nucleophilic attack of water on the ester carbonyl group of peptidyl-tRNA (Jackson et al., 2012). Termination and recycling are temporally coupled through the action of ABCE1 (mammals)/Rli1 (yeast), a member of the ATP-binding cassette protein family, that participates in peptide release and separating

the 80S ribosome into a free 60S subunit and a 40S subunit bound to deacylated tRNA and mRNA (Dong et al., 2004; Pisarev et al., 2007; Young et al., 2015a). ABCE1/Rli1 is critical for release and contributes to peptide release independently of eRF3 through eRF1 (Shoemaker and Green, 2011).

Following peptide release, the 80S ribosome is still bound to the mRNA along with the now deacylated tRNA in the P site, and likely eRF1. The presence of eRF1 accounts for a 2-nt toe-print shift in post-termination complexes (Jackson et al., 2012; Pisarev et al., 2007; Pisarev et al., 2010). ABCE1/Rli1 stimulates efficient dissociation of the 60S subunit, leaving mRNA and deacylated tRNA bound to the 40S subunit (Shoemaker and Green, 2011; Young et al., 2015a). The deacylated tRNA is removed from the ribosome with dissociation of the 40S-mRNA complex in a process mediated by Ligatin and MCT-1/DENR (Jackson et al., 2012; Skabkin et al., 2010). In some cases, the partial dissolution of the complex can allow for “re-initiation” under the correct circumstances, enabling the translation of more than one open reading frame on an mRNA. Reinitiation is discussed in more detail below, but Dong et al. found that affinity-tagged Rli1 copurifies with eIF2, eIF3, and eIF5, perhaps linking ribosome recycling with initiation (Dong et al., 2004; Nurenberg and Tampe, 2013).

1.4. Upstream open reading frames and translation of main open reading frames

Many genes contain at least one short upstream open reading frame (uORF) that is less than about 30 codons long. Approximately 50% of mammalian genes and 13% of yeast genes contain a uORF (Calvo et al., 2009; Lawless et al., 2009; Resch et al., 2009). Estimates of plant genes containing uORFs range from 11% to 60% (Hayden and Jorgensen, 2007; Pesole et al., 2000). After translating the uORF, some ribosomes resume scanning and are capable of reinitiating at downstream sites. The process is not well understood, but termination at uORF stop codons probably proceeds conventionally, and the 60S subunit is released followed by deacylated tRNA, but some 40S subunits remain on the mRNA and resume scanning (Jackson et al., 2010).

Multiple factors influence reinitiation efficiency including the time to translate the uORF, the distance between the uORF termination codon and the main open reading frame (mORF) start site, and the availability of ternary complex. The probability of rescanning and re-initiation decreases rapidly with increasing uORF length or if the uORF includes stable RNA secondary structures that cause pausing during elongation (Jackson et al., 2010; Kozak, 2001). Luukkonen *et al.* used the HIV type 1 *tat* mRNA, which contains the *tat* open reading frame followed by *rev* and *nef* open reading frames, to examine how open reading frame length impacts reinitiation at a second open reading frame (Luukkonen et al., 1995). With the native mRNA, only Tat is produced due to lack of reinitiation at the *rev* and *nef* open reading frames (Luukkonen et al., 1995; Schwartz et al., 1990; Schwartz et al., 1992). Normalizing to Rev expression from an mRNA lacking the *tat* uORF, Luukkonen *et al.* showed that a uORF length of 75 nt (25 codons) reduced Rev expression to 60%. Increasing the nt length from 102 nt (34 codons) to 111 nt (37 codons) decreased

expression from 40% to 20% and at 120 nt (40 codons) Rev expression was undetectable. In a similar study, Kozak used an *in vitro* translation system to study how uORF length impacts downstream CAT expression. When normalized to an mRNA without a uORF, even a 39-nt (13 codon) uORF reduced CAT expression to 30% (Kozak, 2001). In results mirroring Lukkonen, at 99 nt (33 codons) CAT expression was reduced to about 20% (Kozak, 2001; Luukkonen et al., 1995). In addition to examining uORF length, Kozak also examined the effect of elongation time on reinitiation (Kozak, 2001). A viral pseudoknot that had been shown to cause a transient delay in elongation was inserted into the uORF or into the mORF. Inserting the pseudoknot into the CAT mORF had little effect on CAT expression, but inserting it into the uORF strongly inhibited downstream translation of CAT (Kozak, 2001). The control uORF, a 17 codon uORF without secondary structure, allowed expression of CAT (Kozak, 2001). The fact that both increased uORF length and impeding elongation with RNA secondary structure reduces the efficiency of rescanning indicates that the time taken to translate the uORF is crucial.

The inverse relationship between reinitiation and length of the uORF/translation time could be related to the loss of initiation factors (Kozak, 1987, 1992, 2001; Luukkonen et al., 1995). Some of the initiation factors which promoted initiation at the uORF might remain after subunit joining, and could be stochastically released during the elongation and termination phases (Jackson et al., 2012; Kozak, 1987). At the time Kozak proposed the loss of initiation factors, there was no obvious way to identify which factors might be essential for reinitiation; however, alternate mechanisms of initiation were discovered later, such as IRES-dependent initiation, that require only a subset of initiation factors. By examining mRNAs with alternate initiation mechanisms and systems lacking specific

initiation factors, researchers have been able to deduce which factors contribute to reinitiation after translating a uORF. Pöyry et al. used a series of mRNAs based on triose phosphate isomerase-CAT (TPI-CAT) (Poyry et al., 2004). The reporter consists of a 5'-UTR and the first 38 codons of TPI placed immediately upstream of the CAT coding sequence. A termination codon was introduced at codon 2 and reinitiation could occur at codon 14 (Met14) (Poyry et al., 2004). In an *in vitro* coupled transcription-translation system, the reinitiation efficiency was approximately 40% of the primary initiation at codon Met1 (Poyry et al., 2004). Efficient reinitiation was only seen when the mRNA was translated by a scanning-dependent mechanism or by initiation dependent on the encephalomyocarditis virus IRES (Jackson et al., 2012; Poyry et al., 2004). No reinitiation was observed when the primary initiation event was dependent on the classical swine fever virus IRES, which initiates independently of eIFs 4A, 4B, 4E, and 4G; or the cricket paralysis virus IRES, which is completely independent of all eIFs (Jackson et al., 2012; Poyry et al., 2004). Pestova and Kolupaeva used an mRNA with an unstructured 5'-UTR consisting of CAA repeats that can be translated at lower efficiency in the absence of any of the members of the eIF4 group of initiation factors (Pestova and Kolupaeva, 2002). Reinitiation was observed when the mRNA was translated using rabbit reticulocyte lysate (RRL) containing all of the initiation factors, but not in an eIF4G-depleted system (Pestova and Kolupaeva, 2002); however, reinitiation did occur when a recombinant eIF4G p50 fragment that interacts with both eIF3 and eIF4A was added (Poyry et al., 2004). These observations led to the proposal that reinitiation requires the persistence of the eIF4G-eIF3-40S subunit chain of interactions (Jackson et al., 2012; Poyry et al., 2004). This model was initially met with skepticism because the prevailing view was that all factors dissociated

from the ribosome at or just before subunit joining. The notion of factor dissociation was based on the observation that no initiation factors could be detected co-sedimenting with 80S initiation complexes (Benne and Hershey, 1978; Unbehauen et al., 2004); however, the length of time required for these experiments and the stringent conditions might have caused factor dissociation. With this caveat, the results cannot rule out a brief persistence of the factor-ribosome interactions for the length of time it would take to translate a short uORF (Jackson et al., 2012). Bohlen et al. developed a technique that they term selective 40S footprinting to visualize initiating 40S ribosomes on endogenous mRNAs *in vivo* (Bohlen et al., 2019). They discovered that eIF3B, eIF4G1, and eIF4E remain on elongating ribosomes with a decay half-life of about 12 codons. Bohlen et al.'s findings provide a rationale for how ribosomes can reinitiate translation after translating small uORFs.

The efficiency of re-initiation increases with increasing distance between the upstream stop codon and the downstream AUG codon, with a maximum efficiency reached around a distance of 80 nt in mammalian cells (Kozak, 1987). This led to the hypothesis that the re-scanning 40S subunits are incompetent for re-initiation because they lack ternary complex, but a new ternary complex can be acquired during scanning. Accordingly, this means that the critical distance between the stop codon and the reinitiation site is dependent on the availability of eIF2-ternary complex, which is a regulatory target (Dever et al., 1992; Dever et al., 1995; Hinnebusch, 2005; Jackson et al., 2012; Sonenberg and Hinnebusch, 2009). Two well-studied examples of uORFs regulating mORF translation through the availability of eIF2-ternary complex are the genes *GCN4* in yeast and *ATF4* in mammals (Abastado et al., 1991; Dever et al., 1995; Hinnebusch, 2005; Vattam and Wek, 2004).

Both the *GCN4* and *ATF4* mRNAs are regulated through reversible phosphorylation of eIF2 in response to environmental stressors, in particular in response to amino acid starvation or endoplasmic reticulum stress (Fig. 1.2) (Harding et al., 1999; Hinnebusch, 2005). As discussed in more detail above, eIF2 is responsible for binding Met-tRNA_i^{Met} to the 40S ribosomal subunit in an early step in the translation initiation pathway, as part of the ternary complex formed by eIF2, Met-tRNA_i^{Met}, and GTP. During translation initiation the GTP is hydrolyzed to GDP, and an eIF2-GDP binary complex is released. eIF2 has a 100-400-fold higher affinity for GDP than GTP making the GEF eIF2B essential for recycling eIF2-GDP back to active eIF2-GTP (Dever et al., 1995; Pain, 1986). Phosphorylation of the eIF2 α subunit at serine 51 inhibits the exchange of GDP for GTP (Dever et al., 1995; Nika et al., 2001; Rowlands et al., 1988). In addition, due to competitive inhibition of eIF2B, phosphorylated eIF2 also prevents eIF2B from exchanging GDP for GTP on unphosphorylated eIF2 (Rowlands et al., 1988). In all cells examined, eIF2 is present at higher levels than eIF2B; therefore, substoichiometric phosphorylation of eIF2 is sufficient to abolish eIF2B activity (Dever et al., 1995). Without eIF2B, eIF2 cannot be returned to its active GTP bound state, and initiation rates are decreased because of the lack of available ternary complex (Sonenberg and Hinnebusch, 2009). A reduction in ternary complex decreases global translation, but increases the translation of a select number of mRNAs containing short uORFs including the mRNAs encoding the stress response factors yeast *GCN4* and mammalian *ATF4*, *GADD34*, and *CHOP* (Abastado et al., 1991; Dever et al., 1995; Hinnebusch, 2005; Lee et al., 2009; Palam et al., 2011; Vattem and Wek, 2004; Young and Wek, 2016; Young et al., 2015b).

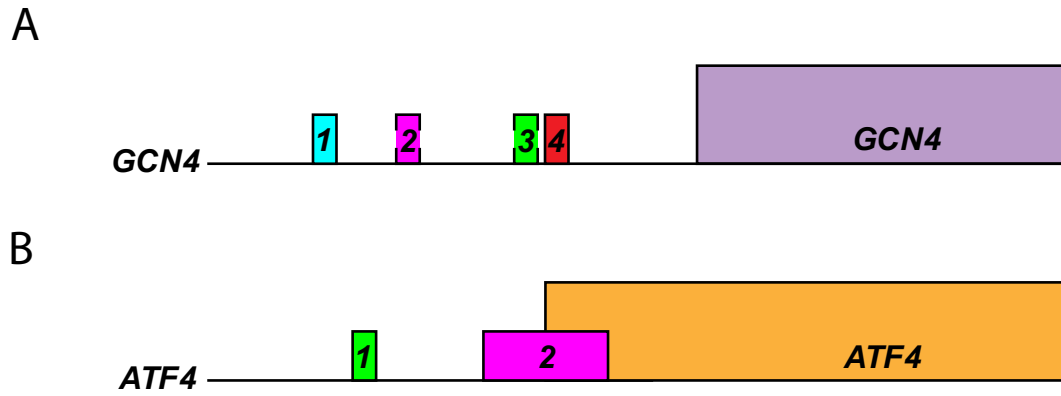


Figure 1.2. Regulated translation reinitiation controls *GCN4* and *ATF4* expression

Figure 1.2. Regulated translation reinitiation controls *GCN4* and *ATF4* expression

(A) Translation of the *S. cerevisiae GCN4* mRNA is regulated by reversible phosphorylation of eIF2 in response to amino acid stress. Four short uORFs regulate the flow of ribosomes to the *GCN4* open reading frame and control *GCN4* expression in response to ternary complex levels. Under normal growth conditions, eIF2 is not phosphorylated and ternary complex levels are high. Scanning ribosomes translate uORF1 and then resume scanning. These ribosomes readily reacquire a ternary complex and then reinitiate translation at the inhibitory uORFs 3 or 4. Translation of these latter uORFs causes ribosomes to disengage from the mRNA and *GCN4* is not expressed. Under stress conditions, eIF2 is phosphorylated leading to decreased levels of ternary complex. Following translation of uORF1, ribosomes scan a longer time, and thus a longer distance, before reacquiring a ternary complex. The ribosomes scan past uORFs 2-4 without reinitiating and then reinitiate translation at the *GCN4* open reading frame leading to increased *GCN4* expression. (B) Translation of metazoan *ATF4* mRNA is also regulated by reversible phosphorylation of eIF2 as part of the integrated stress response. Two uORFs regulate *ATF4* expression, with uORF2 overlapping out-of-frame with *ATF4*. Under normal conditions, eIF2 is not phosphorylated and ternary complex levels are high. Scanning ribosomes translate uORF1 and then resume scanning. The ribosomes readily reacquire a ternary complex and then translate uORF2, preventing initiation at the *ATF4* start site. Under stress conditions, phosphorylation of eIF2 leads to reduced ternary complex levels. After translating uORF1, ribosomes scan a longer time before reacquiring a ternary complex. The ribosomes scan past the uORF2 start codon without initiating, then

reacquire a ternary complex and reinitiate translation at the *ATF4* start codon, leading to increased *ATF4* expression.

Yeast *GCN4* is a transcriptional activator that allows *S. cerevisiae* to respond to amino acid starvation through the general amino acid control (GAAC) response (Fig. 1.2A). As part of GAAC, *GCN4* stimulates the transcription of more than 30 amino acid biosynthetic genes in 12 pathways, as well as genes encoding aminoacyl-tRNA synthetases and pathway specific activators (Hinnebusch, 2005). Because yeast can synthesize all 20 amino acids, GAAC is not triggered by growth in minimal medium, but GAAC is triggered by antimetabolites, mutations of a biosynthetic enzyme, or by media containing amino acid imbalances that elicit starvation by feedback inhibition of enzymes in shared pathways (Hinnebusch, 2005; Niederberger et al., 1981). During GAAC, the protein kinase (PK) *GCN2* is activated by a buildup in uncharged tRNA and then phosphorylates eIF2 α to enhance translation of *GCN4* (Hinnebusch, 2005; Lanker et al., 1992; Wek et al., 1995). The *GCN4* mRNA contains four uORFs in the 5' leader. The uORFs are very short with uORFs 1, 3 and 4 containing four codons and uORF2 containing three codons (Fig. 1.2A) (Hinnebusch, 1984; Mueller and Hinnebusch, 1986). The first and fourth uORFs are sufficient for nearly wild-type regulation and have different effects on *GCN4* expression (Hinnebusch, 1997). If uORF1 is left intact but the other uORFs are mutated, *GCN4* translation is reduced by 50%; however, if only uORF4 is left intact, *GCN4* translation is reduced to about 1% under both amino acid starvation and non-starvation conditions (Hinnebusch, 1997; Mueller and Hinnebusch, 1986). This indicates that uORF1 is necessary to overcome the translational barrier imposed by uORF4 through reinitiation

(Hinnebusch, 1997). During translation of the *GCN4* mRNA, the 43S PIC binds near the cap, scans the mRNA, and initiates translation at the 5' proximal uORF1 (Abastado et al., 1991; Cigan et al., 1988b; Hinnebusch, 1997). After synthesis of the short uORF1 peptide, some 40S subunits, along with essential initiation factors, retain association with the mRNA and can reinitiate translation downstream if they acquire ternary complex. Translation re-initiation on the *GCN4* mRNA is regulated by the availability of eIF2-ternary complexes (Dever et al., 1995). In unstressed cells, ternary complexes are readily available and a scanning ribosome will rapidly reacquire a ternary complex following termination at uORF1. The scanning ribosome with bound ternary complex will reinitiate translation at uORF2, 3, or 4 and then dissociate from the mRNA, failing to translate *GCN4* (Hinnebusch, 1997). In contrast, under amino acid starvation conditions, phosphorylation eIF2 α reduces available ternary complex. The scanning 40S ribosomal subunits fail to acquire Met-tRNA^{Met} as part of the ternary complex and bypass uORFs 2-4. Many of these subunits will acquire a ternary complex while scanning between uORF4 and *GCN4*, allowing them to initiate translation at the *GCN4* start site (Dever et al., 1992; Hinnebusch, 1997). As part of the GAAC, increased eIF2 α phosphorylation and *GCN4* levels contribute to a program of gene expression that coordinates amino acid and purine metabolism, salvages nutrients, represses general protein synthesis, and extends yeast lifespan (Hinnebusch, 2005; Mittal et al., 2017; Vattam and Wek, 2004).

ATF4 is a master regulator in metazoan cells that is crucial for the adaptation to stresses by increasing expression of genes involved in metabolism, nutrient import, and countering oxidative stress as a key mediator of the Integrated Stress Response (ISR) (Fig 1.2B) (Harding et al., 2000; Harding et al., 2003; Jiang et al., 2004; Young and Wek, 2016).

In mammals, four protein kinases are known to link distinct upstream stress signals to eIF2 α phosphorylation (Taniuchi et al., 2016). The four kinases respond to different stresses: PKR-like eukaryotic initiation factor 2 α kinase (PERK) is activated by misfolded proteins in the endoplasmic reticulum (Bertolotti et al., 2000; Harding et al., 1999); general control nonderepressible 2 (GCN2) is activated by uncharged tRNAs, allowing cells to adapt to amino acid starvation (Dever et al., 1992; Zhang et al., 2002); heme-regulated inhibitor (HRI) is activated during heme deficiency (Chen and London, 1995; Chen et al., 1991); and protein kinase R (PKR) is activated during viral infection (Levin et al., 1980; Meurs et al., 1990). Harding et al. studied the ATF4-mediated integrated stress response by treating mouse fibroblasts with or without *Atf4* or *Perk* with the drug tunicamycin to induce ER stress (Harding et al., 2003). Some of the genes most affected by ER stress that were dependent on ATF4 include *Herp*, an ER stress inducible member of a secretory pathway; *Atf3*, a stress response transcription factor; *Cpo*, important for mitochondrial function and redox/detoxification; many genes involved in translation, amino acid import, and metabolism including tRNA synthetases (Harding et al., 2003).

The *ATF4* mRNA contains two uORFs (Fig. 1.2B). The 5' proximal uORF1 is only three codons while uORF2 is 59 codons and overlaps out-of-frame (OF) with the *ATF4* coding region (Lu et al., 2004; Vattam and Wek, 2004). Like *GCN4* uORF1, the *ATF4* uORF1 promotes downstream translation reinitiation (Lu et al., 2004; Vattam and Wek, 2004). In unstressed cells, eIF2-GTP is abundant and ribosomes scanning downstream of uORF1 efficiently reinitiate at the uORF2 start codon and are thus unable to translate the *ATF4* coding region because the *ATF4* start codon is a considerable distance upstream of the uORF2 stop codon. During cell stress, eIF2 α phosphorylation results in inhibition of

eIF2B and reduced recycling of eIF2-GDP leading to reduced levels of eIF2-GTP and ternary complexes. The increased time required for scanning ribosomes to acquire a ternary complex allows the scanning ribosome to bypass the uORF2 start codon, acquire a ternary complex, and reinitiate translation at the *ATF4* start codon (Vattem and Wek, 2004).

1.5. uORF-encoded nascent polypeptide-mediated translational stalling

Interactions between the nascent chain and the ribosome exit tunnel can modulate the translation rate and induce translational stalling to regulate gene expression. The active site for peptide bond formation is the PTC located in a cleft on the intersubunit side of the large ribosomal subunit (Hussain et al., 2014; Llacer et al., 2015; Rodnina, 2013; Simonovic and Steitz, 2009). As the nascent chain is synthesized, it traverses a tunnel within the large subunit and exits at the solvent side. Mounting evidence has revealed that instead of being a passive channel for the nascent chain, the ribosomal tunnel plays a role in early protein folding events as well as regulation of translation (Wilson and Beckmann, 2011). Proteolytic and early cryogenic electron microscopy (cryo-EM) studies show that, in an extended conformation, nascent chains require approximately 30 amino acids to span the ribosomal exit tunnel (Malkin and Rich, 1967; Sabatini and Blobel, 1970); however, subsequent cryo-EM studies of ribosome-nascent chain complexes have revealed more compacted or α -helical-like nascent chain conformations within distinct regions of the ribosome exit tunnel (Bhushan et al., 2010).

As described in more detail in the previous section, peptide bond formation at the PTC involves the accurate placement of the substrates to allow nucleophilic attack of the α -amino group of the aminoacyl-tRNA in the A site onto the carbonyl-carbon of the peptidyl-tRNA in the P site (Rodnina, 2013; Simonovic and Steitz, 2009). The rate of this nucleophilic attack varies for each amino acid (Wohlgemuth et al., 2008) with some amino acids or combinations of amino acids disfavoring peptide bond formation to such an extent that they can slow translation or promote translational arrest. In eukaryotes, consecutive stretches of positively charged amino acids have been shown to induce translational arrest

(Chiabudini et al., 2014; Lu and Deutsch, 2008; Wilson et al., 2016). Proline-containing motifs also promote ribosome stalling during translation elongation and termination (Gutierrez et al., 2013) because proline is both a poor A-site acceptor of the peptidyl moiety (Johansson et al., 2011; Pavlov et al., 2009) as well as a poor donor when located in the P site (Wohlgemuth et al., 2008).

Due to differences in bacterial and eukaryotic translation, the mechanism of regulation by ribosome stalling differs in bacteria versus eukaryotes. Bacteria have coupled transcription and translation that is not possible in eukaryotes because in eukaryotes transcription occurs in the nucleus while translation occurs in the cytoplasm. In addition, unlike eukaryotic mRNAs, bacterial mRNAs can be polycistronic and contain multiple open reading frames on a single mRNA. Instead of having a uORF that represses mORF translation, bacteria utilize ribosome stalling to mask (del Valle et al., 2019; Gong and Yanofsky, 2002a) or unmask (Arenz et al., 2014; Sarker and Oliver, 2002) regulatory domains or binding sites within the mRNA. One such regulator is Rho, a bacterial protein that terminates transcription by binding to transcription terminator pause sites on exposed sections of RNA.

Regulatory nascent chain mediated translational stalling can be categorized into three types: bacterial transcription antitermination via translational stalling, bacterial translation induction via translation elongation stalling, and eukaryotic translation repression via translation stalling (Wilson et al., 2016). TnaC is an example of induction via transcription antitermination mediated by translation termination stalling. TnaC is a peptide with an arrest sequence that stalls the ribosome dependent on the presence of L-tryptophan. The stalling leads to upregulation of the expression of tryptophanase (TnaA)

and a tryptophan-specific permease (TnaB) (Gong and Yanofsky, 2002b) that are encoded together with the TnaC peptide in the *tna* operon in the order TnaC-TnaA-TnaB. A spacer between *tnaC* and *tnaA* contains Rho binding sites that are no longer accessible upon L-tryptophan dependent stalling, resulting in anti-termination of transcription and induction of *tnaA* and *tnaB* in the presence of free L-tryptophan. Cryo-EM analysis of a TnaC-stalled ribosome confirmed that the nascent TnaC peptide and the ribosome exit tunnel cooperate to monitor free L-tryptophan levels (Bischoff et al., 2014; Wilson et al., 2016). Critical residues of the nascent chain were found in a defined conformation in the ribosome exit tunnel establishing multiple contacts to the tunnel wall from the PTC all the way down to the uL22/uL4 constriction (Bischoff et al., 2014). Along with the peptide, two L-tryptophan molecules were present in the tunnel, in binding pockets formed by the ribosome exit tunnel and the nascent chain (Bischoff et al., 2014). Stabilization of the nascent chain in the presence of L-tryptophan is communicated allosterically to the PTC, where a C-terminal proline residues together with induction of an unfavorable geometry of the PTC for release factor binding and peptide transfer results in stalling (Bischoff et al., 2014; Wilson et al., 2016).

Two examples of translation induction via translation elongation stalling are SecM and ErmCL. SecM (secretion monitor) is a peptide that stalls ribosomes during synthesis, upregulating the SecA-dependent pathway in *E. coli* (McNicholas et al., 1997). SecA is an ATPase-dependent motor that helps secretory and outer membrane proteins cross the bacterial cytoplasmic membrane. SecM is encoded by the 5'-end of the bicistronic *secM-secA* mRNA that also encodes SecA. The intergenic region of the *secM-secA* mRNA can form a stem-loop (SL) secondary structure, masking the translation initiation Shine-

Dalgarno sequence of *SecA* and limiting its expression (Kiser and Schmidt, 1999; Zhang et al., 2015). Impairment of SecM secretion increases the duration of the SecM stall (Sarker and Oliver, 2002), allowing the bacterium to sense secretion defects. Increasing the SecM stall destabilizes the *secM-secA* mRNA stem-loop structure and increases the exposure of the *SecA* Shine-Dalgarno, enabling ribosomes to bind and translate SecA. Unlike eukaryotic uORFs, SecM is large, 170 amino acids in length, and contains a 17-amino-acid arrest sequence that interacts with components of the ribosome exit tunnel. Interactions between the arrest sequence and the ribosome exit tunnel lead to an inhibitory conformation of the PTC (Zhang et al., 2015), slowing peptide bond formation, especially because a proline-tRNA is in the A-site of the PTC and proline has a naturally slow peptide bond formation rate (Pavlov et al., 2009).

The *Staphylococcus aureus* ErmCL leader peptide senses macrolide antibiotics and induces expression of a downstream resistance gene ErmC, encoding a macrolide resistance methyltransferase. Similar to the *SecM-SecA* mRNA, the *ErmCL-ErmC* mRNA has a stem-loop secondary structure that sequesters the ribosome-binding site and the AUG start codon of the *ErmC* open reading frame (Arenz et al., 2014; Horinouchi and Weisblum, 1980). In the presence of subinhibitory concentrations of erythromycin, ribosomes translating *ErmCL* stall, leading to a change in the mRNA secondary structure. The change in mRNA secondary structure exposes the ribosome-binding site and the start codon of the *ermC* gene, allowing expression of ErmC. Cryo-EM structures show that erythromycin binds in the ribosome exit tunnel and that the ErmCL nascent chain directly senses the drug in the tunnel (Arenz et al., 2014). The interaction between the ErmCL nascent chain and erythromycin in the ribosome exit tunnel induces allosteric conformational changes at the

PTC, preventing the stable binding and accommodation of the A-site tRNA and inducing translational arrest.

Unlike in bacteria where stalling can induce expression of a mORF, in eukaryotes mORF translation is repressed via translation stalling in a uORF. A well-studied example of uORF-mediated translational control is from human cytomegalovirus (hCMV) (Chang et al., 1989; Degnin et al., 1993; Schleiss et al., 1991). The 5' leader of the hCMV UL4 mRNA encoding the glycoprotein gp48 contains three upstream AUG (uAUG) codons that initiate uORF1, uORF2, and uORF3 (Schleiss et al., 1991). Mutational studies revealed that the 22-residue uORF2 represses gp48 translation during early stage hCMV infection (Janzen et al., 2002; Schleiss et al., 1991). The uORF2 start codon is in poor context (GUGaugC) suggesting that ribosomes will frequently leaky scan past the AUG codon without initiating translation. Even with leaky scanning, uORF2 is repressive (Cao and Geballe, 1995). Putting uORF2 upstream of β -galactosidase (β -gal) reduced β -gal expression 10-fold (Cao and Geballe, 1995). When the uORF2 start codon context was optimized, expression of β -gal dropped to background levels (Cao and Geballe, 1995). Mutational analysis of uORF2 identified residues 21 and 22 as critical for the inhibitory activity of the uORF with the PP-stop motif impairing translation termination and leading to persistence of the uORF2 peptidyl-tRNA linkage with the proline-tRNA in the P site and the termination codon is in the A site. (Bhushan et al., 2010; Degnin et al., 1993; Janzen et al., 2002). In a cryo-EM structure, the hCMV stalling peptide was essentially extended, but there were interactions with the ribosome exit tunnel constriction that appeared to alter the PTC (Bhushan et al., 2010). The repressive nature of uORF2 despite the poor context of the AUG start codon suggests that either the stall is very long preventing subsequent

scanning ribosomes from leaky scanning to the UL4 start codon or that there are additional factors that increase initiation at uAUG2.

An additional well described stalling uORF-encoded peptide is the fungal arginine attenuator peptide (AAP). The *N. crassa arg-2* mRNA, and the related *S. cerevisiae CPAI* mRNA, encode a 23-residue and 24-residue peptide, respectively, upstream of an arginine biosynthetic enzyme subunit (Luo et al., 1995; Werner et al., 1987). The mORF encodes the small subunit of carbamoyl-phosphate synthetase A (CPSase A) that binds glutamine, the physiological nitrogen donor of the CPSase, and transfers its amide nitrogen group to a larger subunit that catalyzes the synthesis of carbamoyl phosphate from ammonia (Pierard and Schroter, 1978; Werner et al., 1987; Werner et al., 1985). Expression of *CPAI* and *arg-2* is repressed by arginine at the translation level in a manner dependent on the *cis* action of the leader peptide encoded in the 5' region of the mRNA (Delbecq et al., 1994; Luo et al., 1995; Werner et al., 1987). The *CPAI* and *arg-2* uORFs initiate at an AUG codon in poor context allowing frequent leaky scanning. In fact, removing the uORF abolishes repression of *CPAI* expression, but does not increase the level of *CPAI* expression (Wang et al., 1999; Werner et al., 1987). Mutating a single highly conserved aspartate residue, D12 in *arg-2*, abolishes arginine specific repression of the downstream gene (Freitag et al., 1996; Werner et al., 1987). In the presence of high arginine concentrations, ribosomes translating the AAP stall at the stop codon; however, the stop codon is not a prerequisite for stalling, suggesting that the stall is due to impaired elongation (Fang et al., 2000; Wang et al., 1998; Wei et al., 2012). In a cryo-EM study, the C terminus of the AAP appears to compact adjacent to the PTC. At the tunnel constriction,

the AAP nascent chain appears to be stabilized and interacts with ribosomal proteins L4 and L17 altering the PTC, which could lead to stalling (Bhushan et al., 2010).

There are several features that most nascent chain staller have in common. First, in all the cases described above, the geometry of the PTC is perturbed in order to achieve the stall. Three residues of the PTC, large rRNA nucleotides A2602, U2585, and U2506 (numbered according to *E. coli*) are particularly prone to be affected due to critical contributions to peptide bond formation and/or termination (Wilson et al., 2016). Second, proline residues are critical in many, but not all arrest sequences presumably due to their already poor reaction kinetics and restricted geometry (Johansson et al., 2011; Wilson et al., 2016). Third, all described staller establish specific contacts to the wall of the ribosome tunnel; however, the critical regions appear to be concentrated on the upper half of the tunnel extending from the PTC to the uL22/uL4 constriction (Wilson et al., 2016). Finally, allosteric relay systems are proposed to transmit the molecular recognition events from the tunnel to the PTC through the nascent chain or via rRNA nucleotides comprising the tunnel (Wilson et al., 2016).

There are also variations in arrest sequences that allow versatility. First, different arrest sequences stall elongation or termination with the critical amino acid stretch ranging from as short as three amino acids (Sothiselvam et al., 2014) to as long as 99 amino acids (Ishii et al., 2015)(Wilson, 2016). Second, stalling during elongation can occur at a distinct single site, as in TnaC or ErmCL (Arenz et al., 2014; Bischoff et al., 2014; Gong and Yanofsky, 2002a, b), or over an entire stretch of amino acids as for SecM (Sarker and Oliver, 2002; Zhang et al., 2015). Third, there is no consensus and no common contact

pattern between the nascent chain and the ribosome tunnel with some arrest sequences developing secondary structure within the tunnel (Wilson et al., 2016).

1.6. Polyamines and uORF mediated autoregulatory circuits

Polyamines are small organic polycations that interact with negatively charged molecules such as RNA and DNA through their linearly distributed positive charges. In addition to binding nucleic acids, the polyamine spermidine is essential for the post-translational hypusine modification of eIF5A (Park et al., 1981) and for maintaining the inward rectification of potassium ion channels (Lopatin et al., 1995). Polyamines are required for cell growth, survival, and proliferation (Minois et al., 2011), and their elevated levels in cancer cells make them a therapeutic target (Wang and Casero, 2006).

Polyamines are synthesized in a highly regulated multistep process (Fig. 1.3). The first step is the production of putrescine from ornithine by ornithine decarboxylase (ODC) (Fig. 1.3A). ODC is regulated by the inhibitory protein ODC antizyme (OAZ), which is itself regulated by antizyme inhibitor (AZIN) (Ivanov et al., 2010a). Interestingly, AZIN is a non-functional paralog of ODC (Murakami et al., 1996). Spermidine is generated from putrescine by the addition of an aminopropyl group transferred from decarboxylated S-adenosylmethionine (dcAdoMet) by spermidine synthase (Pegg, 2009). Spermine is generated in the same way by the addition of an aminopropyl group to spermidine by spermine synthase (Fig. 1.3A). The dcAdoMet is generated by S-AdoMet decarboxylase (AdoMetDC). Spermine and spermidine can be acetylated by spermine/spermidine acetyltransferases leading to catabolism and/or cellular export (Pegg, 2009).

The ability of uORFs to act as sensors for regulatory circuits and to amplify signals probably explains why they are common in most polyamine biosynthetic pathway mRNAs (Fig. 1.3B) (Ivanov et al., 2010a). As an essential biochemical, polyamine biosynthesis is tightly regulated. AdoMetDC is a key regulated enzyme in the polyamine biosynthetic

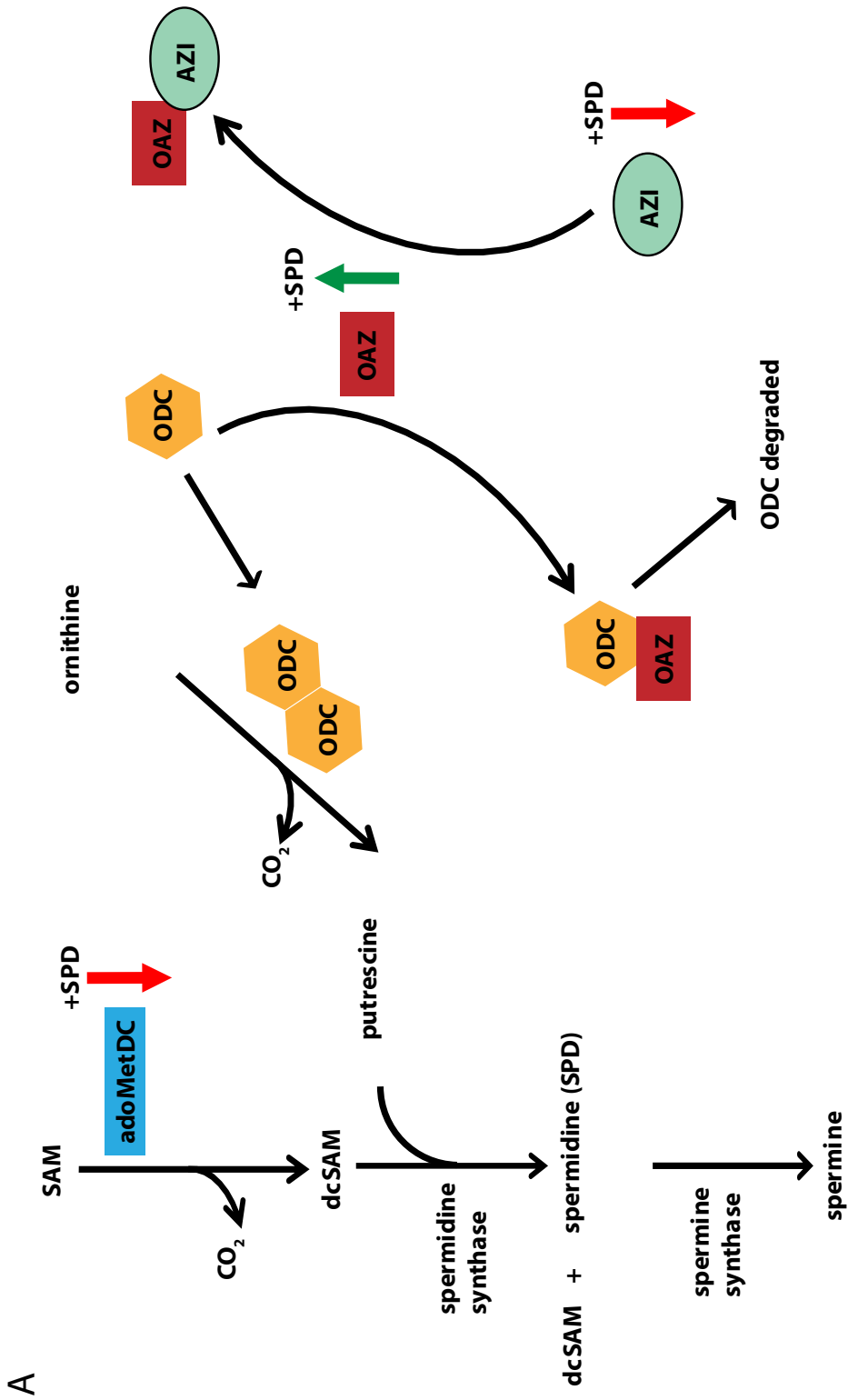


Figure 1.3. Polyamines and uORF-mediated autoregulatory circuits

B

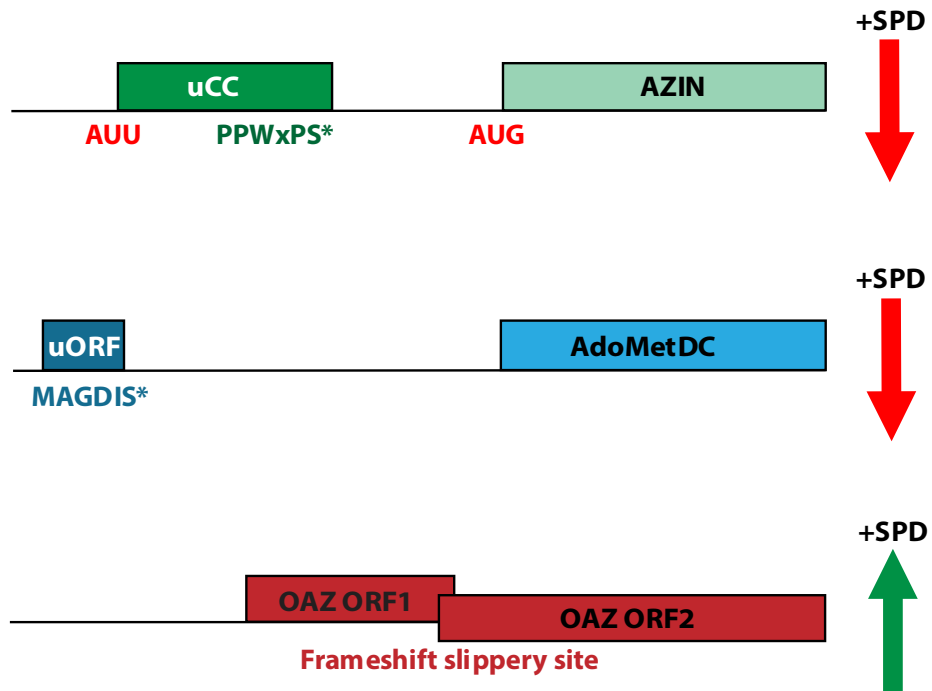


Figure 1.3. Polyamines and uORF-mediated autoregulatory circuits

Figure 1.3. Polyamines and uORF-mediated autoregulatory circuits

(A) Cartoon representation of polyamine biosynthesis in vertebrates and (B) the accompanying translational regulation through uORFs. ODC, ornithine decarboxylase; OAZ, ornithine decarboxylase antizyme; AZIN, antizyme inhibitor; AdoMetDC, S-adenosylmethionine decarboxylase; SpmSyn, spermine synthase. Putrescine, spermidine, and spermine are polyamines that promote OAZ expression by stimulating programmed ribosomal frameshifting. The polyamines cause elongating ribosomes to pause during translation of uORFs in the AZIN and AdoMetDC mRNAs to repress their expression.

pathway and AdoMetDC is feedback-controlled at translation by spermidine and spermine by a uORF (Law et al., 2001). The uORF encodes a short peptide of the sequence MAGDIS; the aspartic acid residue at position 4 is essential, the isoleucine at position 5 can only be substituted for valine, and the peptide must have an amino acid at position 6, so length is important (Law et al., 2001). In the model proposed by Law et al., a ribosome binds the AdoMetDC mRNA at the 5' cap and begins scanning. The AdoMetDC uORF has a high rate of leaky scanning because the uORF AUG start codon is only 14 nucleotides from the 5' cap (Ruan et al., 1996). Very short leader sequences of less than 32 nucleotides increase the rate of leaky scanning with a 50% increase in leaky scanning seen at a distance of 12 nucleotides compared to 32 nucleotides (Kozak, 1991a). In the case of the AdoMetDC uORF, most scanning ribosomes skip over the uORF start codon without initiating. Occasionally, the scanning ribosomal complex recognizes the start codon of the uORF and initiates translation. At low polyamine levels, the ribosome translates the uORF,

terminates, and then presumably either disengages from the mRNA or resumes scanning allowing subsequent ribosomes to access the AdoMetDC reading frame. In the presence of high polyamine levels, the ribosome translating the MAGDIS uORF fails to terminate properly and instead stalls at the stop codon. Ribosome stalling only 35-nts from the cap prevents additional ribosomes from loading onto the AdoMetDC mRNA and thus inhibits AdoMetDC expression (Law et al., 2001).

AdoMetDC is just one example of translational control regulating polyamine levels in cells. In most species, the OAZ mRNA contains an open reading frame that is regulated via a programmed frame-shift (Ivanov et al., 2010a). OAZ is a negative regulator of polyamines that functions by binding to ODC monomers, preventing the formation of functional ODC homodimers (Matsufuji et al., 1990). The bound OAZ presents ODC for ubiquitin-independent degradation by the 26S proteasome (Li and Coffino, 1992; Murakami et al., 1992). In most cases, OAZ is encoded by two partially overlapping open reading frames, orf1 and orf2. The important biochemical properties are encoded by the downstream orf2 that can only be translated after initiation at the upstream orf1. A proportion of ribosomes that initiate translation at the orf1 start site switch to the +1 reading frame when the orf1 stop codon is in the A site and then proceed to decode orf2 to synthesize functional antizyme (Matsufuji et al., 1995). The ribosomal frameshift is controlled by polyamine levels in the cell; at low polyamine levels, few ribosomes change frame, decreasing the synthesis of the negative regulator. Elevated polyamine levels increase the proportion of ribosomes that switch frame and synthesize functional OAZ. OAZ, in turn, promotes ODC turnover leading to a reduction in polyamine levels in the cell.

In addition to translational control of AdoMetDC and antizyme, a third regulator of polyamine biosynthesis is regulated by a type of uORF that we classify as an upstream conserved coding region (uCC). AZIN positively regulates polyamine levels by negatively regulating OAZ. AZIN is a homolog of ODC that has a higher affinity for antizyme than ODC. Thus, AZIN can sequester the intracellular pool of antizyme, leading to increased levels of ODC, higher ODC activity, and higher polyamine levels. A uCC is a uORF-like element that encodes a peptide with high amino acid sequence conservation. The uCC typically lacks an AUG start codon and instead initiates with a near-cognate start codon or an AUG codon in poor sequence context. In most cases, the AZIN uCC starts at a AUU codon, but UUG and ACG initiation codons have also been identified (Ivanov et al., 2010a).

Bioinformatic analysis identified uCCs in the 5' leaders of AZIN homologs in animals, and ODC homologs in mushrooms, Zygomycota, and the Pezizomycota subphylum of Ascomycota (Ivanov et al., 2010a). A distinguishing feature of this uCC is that in nearly all cases, translation appears to initiate at a non-AUG codon. There is little amino acid sequence conservation across the uCC from different phyla, except that almost all AZIN uCCs encode two adjacent proline residues at or very close to the C terminus.

The AZIN uCC represses expression of the mORF (Ivanov et al., 2008). Ivanov et al. generated RLuc reporters where the mouse AZIN 5' leader was upstream of RLuc and transfected the reporters into cells with low polyamines, due to pre-treatment with the ODC inhibitor difluoromethylornithine (DFMO), or with high polyamines, supplemented with spermidine. Experiments using reporters with the wild-type uCC sequence showed that spermidine enhanced repression. Ivanov et al. saw a 6.5-fold repression of RLuc in cells treated with spermidine compared to polyamine-depleted cells (Ivanov et al., 2008). In the

mouse AZIN based reporter, the initiation codon was experimentally verified to be an absolutely conserved AUU codon present in good sequence context (Ivanov et al., 2008). Mutating the AUU initiation codon to a non-initiating UUU codon abolished polyamine-induced repression of the mORF (Ivanov et al., 2008).

The ability of the uCC to repress AZIN expression is dependent on the uCC C-terminal amino acid sequence. A mutant where the last 10 codons of the uCC were out-of-frame led to a complete loss of polyamine-dependent repression (Ivanov et al., 2008). Interestingly, spermidine enhanced initiation at the uCC AUU start codon (Ivanov et al., 2008). In polyamine depleted cells, an AUU initiation codon was 18% as efficient as an AUG codon in directing expression of a uCC-FLuc fusion, while in polyamine supplemented cells the AUU initiation codon was 54% as efficient as the AUG codon (Ivanov et al., 2008). Taken together, the loss of polyamine regulation by the out-of-frame control and the enhanced uCC translation by polyamines, suggest that both the non-AUG initiation and the C-terminal amino acid sequence of the uCC are key regulatory components of the system.

1.7. Regulation of protein expression by upstream conserved coding motifs

At the time this project was begun, the only described uCC was, as described above, in the AZIN mRNA (Fig. 1.4). In a paper describing the regulation of AZIN expression, Ivanov et al. (2018) proposed that uCCs have three key characteristics that make them regulatory: they have a weak start site; the uCCs are conserved at the amino acid level, not at the nucleotide level, and contain highly conserved motifs that are essential for their regulatory effect through ribosomal pausing; and finally, that the ribosomal pause leads to ribosome queuing that both blocks downstream scanning, and enhances initiation at the weak start site of the uCC (Ivanov et al., 2018).

The uCC weak start site is essential because it allows ‘leaky scanning’, which is necessary for downstream mORF translation. During ‘leaky scanning’, scanning ribosomes skip potential weak start sites and instead initiate downstream at an AUG codon in good context. A weak start site can be a near-cognate codon that differs from AUG by a single nucleotide change as in the AUU codon of the AZIN uCC (Fig. 1.4) (Ivanov et al., 2010b). Finally, proximity of the start codon to the 5’ cap can affect initiation efficiency, which is the case with the AdoMetDC uORF encoding the MAGDIS peptide.

In addition to the poor initiation site, the AZIN uCC is conserved at the amino acid level not at the nucleotide level and contains a very highly conserved amino acid sequence motif that is essential for its regulatory effect. The AZIN uCC contains a PPW motif that requires eIF5A for its synthesis (Fig. 1.4) (Ivanov et al., 2018). High polyamine levels interfere with eIF5A activity, causing an elongation defect and the translating ribosome pauses on the PPW motif.

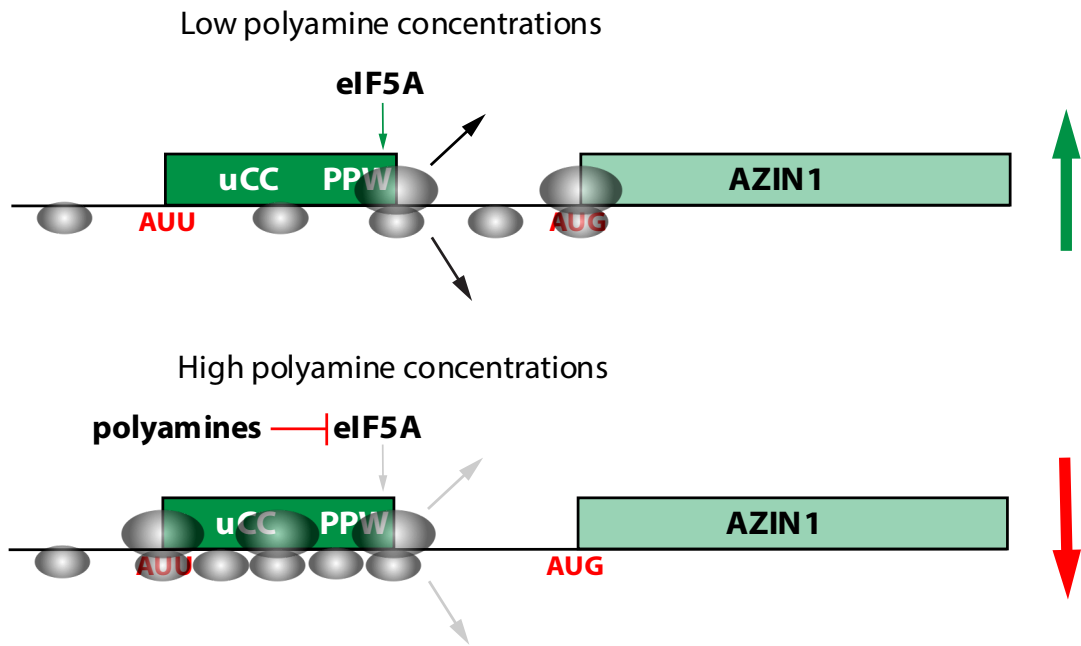


Figure 1.4. A conserved uCC regulates AZIN expression through eIF5A

Figure 1.4. A conserved uCC regulates AZIN expression through eIF5A

The AZIN uCC regulates AZIN expression in response to polyamine concentrations. The uCC initiates at a near-cognate AUU start codon that allows leaky scanning. A conserved PPW motif near the 3' end of the uCC controls translation elongation in response to polyamine levels. Under low polyamine conditions, most ribosomes leaky scan over the uCC without translating and initiate translation at the AZIN start codon. Any ribosomes that initiate translation at the uCC start codon will efficiently translate the PPW motif with the assistance of eIF5A. Following termination at the uCC stop codon, these ribosomes disengage from the mRNA, enabling subsequent ribosomes to leaky scan the uCC and synthesize AZIN. Under high polyamine conditions, eIF5A activity is inhibited. Any ribosomes that initiate translation at the uCC start codon pause at the PPW motif. Subsequent scanning ribosomes, most of which leaky scan past the AUU codon at the start of the uCC, form a queue behind the paused elongating ribosome. When the queue reaches the AUU start codon of the uCC, the scanning ribosome will spend a longer time over the near cognate start site, providing greater opportunity for initiation at this weak start codon. The increased initiation at the uCC start codon will provide more elongating ribosomes to reinforce the elongation pause on the PPW motif, resulting in enhanced uCC translation and decreased translation of the AZIN open reading frame.

A unique feature of uCCs is that part of the regulatory effect is via an elongation or termination defect that generates a ribosome queue. In their 2018 paper, Ivanov et al. show enhancement of initiation at the weak AZIN uCC AUU start site in the presence of

polyamines is dependent on ribosome queuing (Fig. 1.4) (Ivanov et al., 2018). Ribosome profiling using HEK293T cells grown under low polyamine conditions (treated with DFMO) or high polyamines (supplemented with 2 mM spermidine), showed that in the presence of high polyamines, elongating ribosomes pause at the PPW motif with the Trp codon in the A site. Reporter assays demonstrated that the initiation enhancement in the presence of high polyamines was dependent on the conserved motifs within the uCC. Queuing of scanning ribosomes has not been determined directly, but queuing has been deduced experimentally. Ivanov et al. reduced queuing potential using the drug eIF4E/eIF4G interaction inhibitor 1 (4EGI-1). The drug impairs ribosome loading on mRNAs by disrupting the association of eIF4G with eIF4E (Sekiyama et al., 2015). HEK293T cells were transfected with reporters containing the AUU initiated AZIN uCC or an out-of-frame control uCC upstream of *Renilla* luciferase (RLuc). The cells were supplemented with spermidine and the effect of reducing ribosome loading was observed by normalizing to the out-of-frame control. Reducing queuing potential with 4EGI-1 de-repressed RLuc expression (Ivanov et al., 2018). The observation that impairing ribosome loading leads to reduced uCC translation and derepression of AZIN synthesis is consistent with a model in which the elongation pause leads to a queue of subsequent scanning and translating ribosomes. The ribosome queue could increase the time that a scanning ribosome spends in the vicinity of the weak start site, increasing the probability of initiation on the uCC.

In the uCC model, under permissive conditions leaky scanning allows the majority of ribosomes to scan past the weak uCC start site and initiate translation downstream at the mORF start site. Thus, the mORF is translated under ‘standard’ conditions. Occasionally a

scanning ribosome will initiate at the weak start site of the uCC, the frequency of initiation will depend, in part, on the strength of the start codon and context nucleotides. Ribosomes that translate the uCC will likely disengage from the mRNA after termination due to the length of the uCC coding region. When the concentration of polyamines is elevated, the occasional ribosome that initiates translation of the uCC will pause on the PPW motif because high polyamines interfere with eIF5A function. Biochemical studies showed that polyamines competitively inhibit eIF5A function on the ribosome. The paused ribosome in the uCC prompts queuing of subsequent translating and scanning ribosomes (many of which scan past the uCC start codon without initiating). When the queue builds to a sufficient length, a scanning 40S subunit will be poised in the vicinity of the uCC start codon and will thus spend more time traversing over the start codon, increasing the chances for initiation at the weak start site of the uCC. In this way, the elongation pause generates a feedback loop which increases translation of the uCC to reinforce the pause, leading to decreased translation of the mORF (Fig. 1.4).

We believe that the AZIN uCC is just one example of a uCC that regulates a biosynthetic pathway. Current research indicates that the AAP in the *N. crassa arg-2* and *S. cerevisiae CPA1* mRNAs also has the characteristics of a uCC rather than a uORF. Moreover, in reviewing the literature we found that a uORF present in the plant GDP-L-galactose phosphorylase (GGP) mRNA, encoding an enzyme in the ascorbic acid biosynthesis pathway, has similarities to a uCC. As first described by Laing et al., the uORF regulates GGP mRNA translation in response to L-ascorbic acid (Fig. 1.2) (Laing et al., 2015).

1.8. Plant L-ascorbic acid biosynthesis and GDP-L-galactose phosphorylase

L-ascorbic acid, or vitamin C, is an essential biochemical in plants and animals, and a similar compound, D-erythroascorbic acid, is found in yeast (Hancock et al., 2000). In plants and animals, ascorbate is an important electron donor, acting as an antioxidant and an enzymatic co-factor for fifteen mammalian enzymes (Padayatty and Levine, 2016). ascorbate is a co-factor for two monooxygenases, 12 dioxygenases, and one amine oxidase. These enzymes are essential for the synthesis of collagen, carnitine, and neurotransmitters among other functions (Naidu, 2003).

Most plants and animals synthesize ascorbate from D-glucose or D-galactose; however, a small number of animal species have lost the ability to synthesize ascorbate. These species include guinea pigs, fruit eating bats, and apes, including humans, because of the absence of L-gulonolactone oxidase (Burns, 1957; Nishikimi and Yagi, 1991). L-gulonolactone oxidase is the last enzyme in the animal ascorbate biosynthetic pathway and catalyzes the reaction of L-gulono-1,4-lactone with oxygen to generate L-xylo-hex-3-gulonolactone and hydrogen peroxide (Linster and Van Schaftingen, 2007) (Fig. 1.5). L-xylo-hex-3-gulonolactone then spontaneously converts to ascorbate. Animals that cannot produce their own ascorbate, must acquire ascorbate through their diet. Ascorbate is widely available in fresh fruits and vegetables, in particular in fruits such as oranges, lemons, watermelon, strawberries, and in vegetables such as broccoli and cabbage. Ascorbate is also the mostly widely used vitamin supplement throughout the world (Naidu, 2003).

A severe ascorbate deficiency produces the fatal disease scurvy, which can only be cured by ascorbate. Scurvy was a severe problem for large armies, Northern European cities, and sea voyages. Victims typically died after two to three months of a diet lacking

ascorbate. In 1753, James Lind showed that scurvy could be cured by citrus fruits in the first controlled clinical trial (Lind, 1953; Padayatty and Levine, 2016). Ascorbic acid was isolated by Albert Szent-Gyorgyi in 1928 and shown to be the antiscorbutic factor by Szent-Gyorgyi and King in 1932 (King and Waugh, 1932; Svirbely and Szent-Gyorgyi, 1932, 1933), for which Szent-Gyorgyi was awarded the 1937 Nobel Prize in Physiology or Medicine.

In humans, ascorbate is absorbed through the small intestine and transported via SVCT1, while the oxidized form of ascorbic acid, dehydroascorbic acid (DHAA), is likely transported via GLUT2 and then reduced to ascorbic acid within the cell (Padayatty and Levine, 2016). Ascorbate is water soluble and accumulates against a concentration gradient to millimolar concentrations gradient in tissues. The majority of ascorbate is found in the liver, brain, and skeletal muscle of humans. In the human liver, ascorbate is present in the 600-900 μM range but specialized cells like lymphocytes have ascorbic acid concentrations exceeding 3.8 mM (Padayatty and Levine, 2016).

Ascorbate is found in all the cell compartments of higher plants and the average ascorbate concentration is 2-25 mM, with concentrations being highest in chloroplasts (Zhang, 2013). In plant photosynthetic tissues, ascorbate concentrations are regulated according to demand. Leaf ascorbate concentrations increase under high light intensity when photosynthesis drives the production of reactive oxygen species (Bartoli et al., 2006; Dowdle et al., 2007; Gao et al., 2011; Yabuta et al., 2007). The response to light is driven by changes in gene expression for several L-galactose pathway genes. Ascorbate concentrations are tightly regulated in green tissues such as leaves because of the maintenance of the redox state; however, fruits can act as skin organs and accumulate

ascorbate at much higher levels (Cruz-Rus et al., 2012). In non-photosynthetic tissues such as fruit, ascorbate concentrations are relatively stable but can vary between cultivars due to changes in gene expression (Bulley et al., 2009). Studies of ascorbate concentrations in fruits over time indicate that ascorbate is synthesized in the green tissues and is then transported to, and stored in, the fruits (Rasanu, 2005). In the strawberry plant, the roots have the lowest ascorbate concentrations (1 $\mu\text{mol/gFW}$), followed by stems (3.5 $\mu\text{mol/gFW}$), leaf (5.5 $\mu\text{mol/gFW}$), flower (6 $\mu\text{mol/gFW}$), and fruit (10 $\mu\text{mol/gFW}$) (Cruz-Rus et al., 2011). The acerola fruit accumulates ascorbate in concentrations as high as 1400 mg/100 g fresh fruit. In strawberries, ascorbate concentrations increased from 20 mg/100 g fresh fruit in green fruits to 300 mg/100 g fresh ripe fruit (Cruz-Rus et al., 2011).

In animals only one pathway for ascorbate biosynthesis has been described (Fig. 1.5); however, several alternative pathways for ascorbate biosynthesis and metabolism have been discovered in plants (Wheeler et al., 1998; Zhang, 2013). The Smirnoff or D-mannitol/L-galactose pathway was proposed in 1998 (Wheeler et al., 1998), and appears to be the predominant pathway used by plants (Zhang, 2013) (Fig. 1.5). The other pathways include galacturonate pathway (Agius et al., 2003), the myoinositol pathway (Lorence et al., 2004), and a pathway utilizing gulose is also proposed to function in plants although it has not been experimentally validated (Wolucka and Van Montagu, 2003; Zhang, 2013) (Fig. 1.5). The myoinositol pathway and proposed gulose pathway indicate that parts of the animal pathway could also operate in plants, as the pathways all converge on L-gulonolactone, one step upstream of ascorbate. In contrast, the D-mannitol/L-galactose

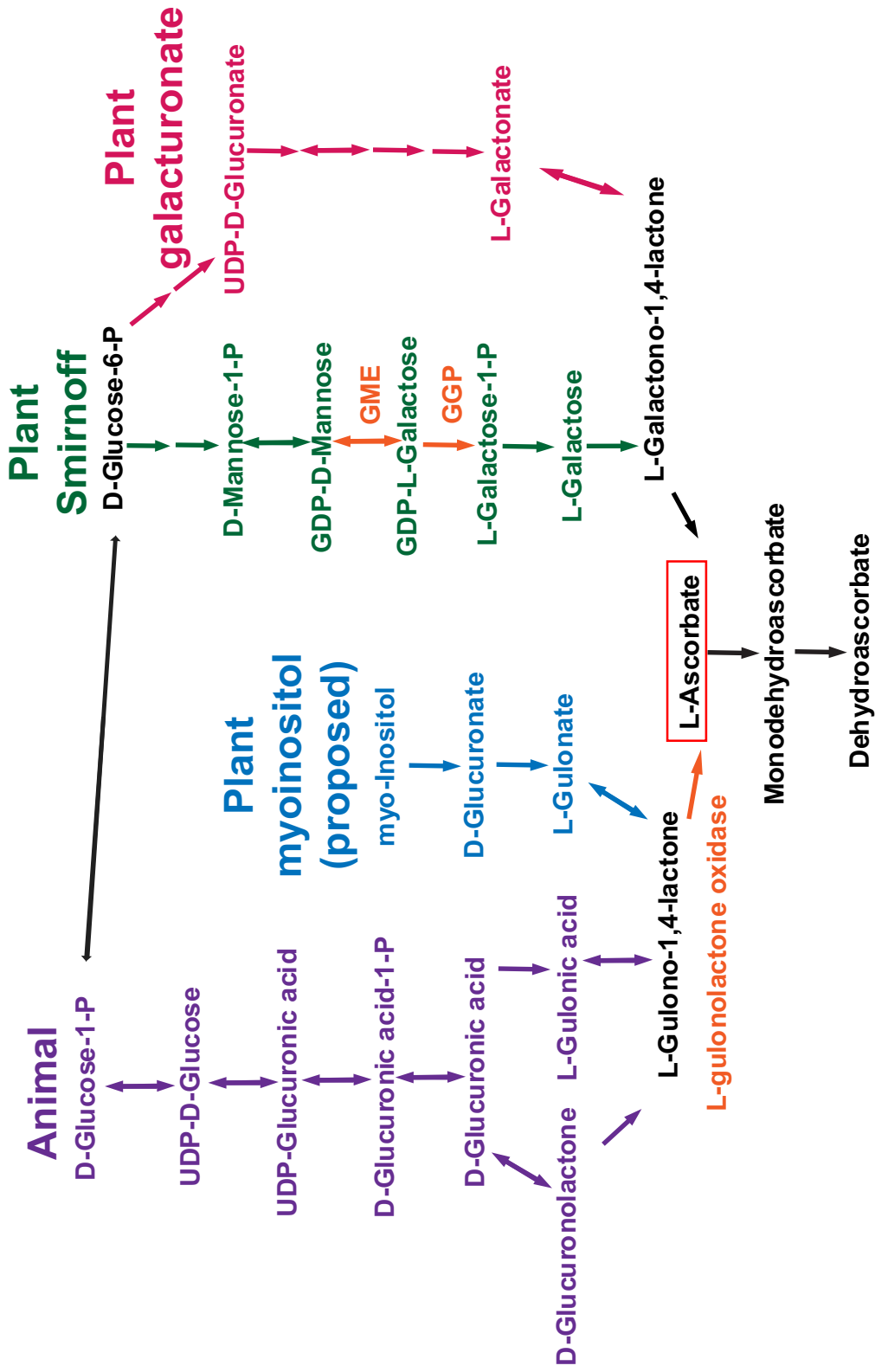


Figure 1.5. The L-ascorbate biosynthetic pathways in plants and animals

Figure 1.5. The L-ascorbate biosynthetic pathways in plants and animals

The ascorbic acid biosynthetic pathways in plants and animals converge on L-ascorbic acid through L-gulonono-1,4-lactone (animals and possibly plants) and L-galactono-1,4-lactone (plants). The animal pathway is shown in purple with the enzyme L-gulonolactone oxidase highlighted in orange. The proposed plant myoinositol pathway is shown in blue. The predominant plant biosynthetic pathway, the Smirnoff pathway, is shown in green and the secondary plant galacturonate pathway is shown in pink. The plant enzymes GME and GGP are shown in orange.

pathway and galacturonate pathway converge on L-galactono-1,4-lactone, which is not present in animals.

In the D-mannitol/L-galactose pathway, ascorbate is synthesized from the precursor of D-glucose over more than ten enzymatic reaction steps (Zhang, 2013). The first six steps are responsible for the biosynthesis of the nucleotide sugar GDP-D-mannose, which is a precursor for cell wall polysaccharides and glycoproteins as well as the substrate for ascorbate biosynthesis. The last four enzymatic steps after synthesis of GDP-L-galactose are unique to ascorbate biosynthesis. GDP-L-galactose phosphorylase (GGP) is the first enzyme in this latter half of the pathway; and converting GDP-L-galactose to L-galactose 1-P is the first committed step in the D-mannitol/L-galactose ascorbate biosynthetic pathway.

The enzyme GGP, encoded by the gene *GGP* or *VTC2*, is the control point of ascorbate biosynthesis in plants, and overexpression of the kiwifruit *GGP* gene resulted in

a four-fold increase in ascorbate accumulation (Bulley et al., 2009; Dowdle et al., 2007; Zhang, 2013). No other single gene in the L-galactose pathways has a significant effect on ascorbate levels (Bulley et al., 2012), but simultaneously increasing the expression of GDP mannose epimerase (GME) and GGP in *Arabidopsis* leads to a synergistic seven-fold increase in ascorbate (Bulley et al., 2009; Zhang, 2013). During fruit development, *GGP* and *GME* transcript levels parallel ascorbate concentration (Bulley et al., 2009). Consistent with the rate controlling effects of GGP, *GGP* transcript levels show a strong circadian rhythm in leaves, with peak expression before dawn, likely in preparation for photosynthesis (Dowdle et al., 2007).

GGP (*VCT2*) is of great interest to plant biologists because of its role in directly influencing ascorbate concentration in plant tissues. GGP was hypothesized to possibly control gene transcription because a GGP-YFP fusion protein was detected in the nucleus as well as the cytosol (Muller-Moule, 2008). In 2015, Laing et al. published that ascorbate, or a metabolite closely connected to ascorbate, downregulated the translation of the GGP mRNA (Laing et al., 2015). High concentrations of ascorbate reduced GGP expression via a feedback loop, reducing ascorbate concentrations (Fig. 1.6). The regulation of GGP expression occurred at the translational level and was mediated by a conserved noncanonical uORF in the 5'UTR of the *GGP* mRNA. Alignment of uORFs from diverse plant species showed that the amino acid sequence was more highly conserved than the nucleotide sequence. Consistent with this observation, synonymous mutations of codons within the uORF did not affect regulation of the downstream mORF (Laing et al., 2015). For the synonymous mutation study, Laing et al. synthesized a *GGP* uORF fragment from *Arabidopsis* (*VTC2*), changing each codon to a synonymous codon except for the initial

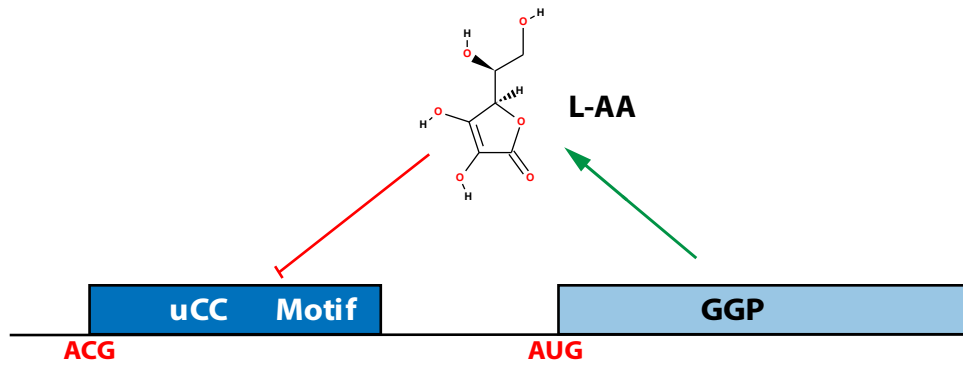


Figure 1.6. The L-ascorbic acid biosynthetic enzyme GGP is negatively regulated by L-ascorbic acid via a proposed uCC

Figure 1.6. The L-ascorbic acid biosynthetic enzyme GGP is negatively regulated by L-ascorbic acid via a proposed uCC

In plants, the GGP enzyme is the control point in the ascorbate biosynthetic pathway. GGP expression is negatively regulated by ascorbate (L-AA) through a conserved non-canonical uORF that we propose is a uCC.

codon. This altered 38% of the uORF nucleotides. The synthesized fragment was cloned into a GGP-LUC reporter and then transiently transfected into *Nicotiana benthamiana*; luciferase activity was measured after 7 days. To generate high ascorbate conditions, the leaf was transfected with a plasmid containing *GGP* under the control of the strong 35S promoter; to lower ascorbate concentrations, the native *GGP* mRNA was targeted by RNAi. Laing et al. also reported that analysis of previously published ribosome profiling studies by Liu et al. (2013) and Juntawong et al. (2014) showed evidence of a higher than average ribosome density on the uORF and stalling during translation of the uORF (Juntawong et al., 2014; Laing et al., 2015; Liu et al., 2013). Two possible mechanisms were proposed for how the *GGP* uORF regulates *GGP* mORF translation: first, ascorbate was proposed to interact directly with the uORF peptide causing the ribosome to stall and thus impairing ribosome scanning to the downstream mORF AUG start site (Laing et al., 2015); second, ascorbate was proposed to directly enhance noncanonical initiation at the near cognate start codon of the *GGP* uORF, and then the uORF-encoded peptide was proposed to cause the ribosome to stall constitutively (Laing et al., 2015). The authors

preferred the later model where ascorbate directly promotes initiation at the ACG start codon of the uORF and then the uORF peptide causes ribosome stalling.

1.9. Is the GGP motif a uCC that regulates ascorbate biosynthesis?

We hypothesize that the noncanonical uORF discovered by Laing et al. (2015) is actually a uCC. Based on the Laing et al. data, the *GGP* uCC exhibits the three hallmark features of a uCC. First, the uORF sequences published by Laing et al. (2015) show conservation of a near-cognate ACG codon, located upstream of the conserved region of the uORF, that could serve as a translation start site for the uORF. Recall that the AZIN uCC utilizes an AUU start site that allows leaky scanning through the uCC and translation of the downstream mORF; perhaps the ACG codon in the *GGP* uORF is functioning in a similar manner. Second, there is absolute conservation of amino acid sequence motifs within the uORF-encoded peptide, but the nucleotide sequence varies. In particular, nucleotide substitutions at the third position of codons, the wobble nucleotide, indicate that the uORF peptide sequence is conserved, but not the mRNA sequence. This conservation is consistent with the conservation of the PPW motif in the AZIN uCC. Finally, Laing et al. (2015) showed evidence of high ribosome density in the *GGP* uORF, consistent with translation of the uORF. Moreover, a prominent ribosome footprint was noted towards the 3' end of the uORF, indicating the presence of a translation stall site. Similarly, Ivanov et al. showed that ribosomes accumulate on the PPW and PS-stop motifs on the AZIN uCC in the presence of higher polyamine concentrations (Ivanov et al., 2018). In the case of the AZIN uCC, the ribosome pausing leads to queuing of subsequent scanning ribosomes and enhanced initiation at the weak upstream start site of the uCC. We propose that a similar mechanism might control translation of the *GGP* uORF in response to ascorbate levels.

The goal of my project was to test our hypothesis that the *GGP* noncanonical uORF is in fact a uCC that regulates GGP expression in response to ascorbate. In support of our

hypothesis, I show that translation of the GGP mRNA is regulated by ascorbate in a uORF-dependent manner and that the conserved motifs in the uORF are essential for regulation. I also provide evidence that the spacing between the conserved motifs in the uORF is critical for ascorbate regulation of GGP mRNA translation. Finally, by manipulating ribosome loading on the mRNA, I provide evidence that ribosome queuing is necessary for ascorbate regulation of GGP mRNA translation by the uORF.

2. Materials and Methods

2.1. Plasmid constructs

2.1.1. Molecular techniques for cloning the reporters

Unless otherwise specified, polymerase chain reactions (PCRs) were performed using Invitrogen Platinum SuperFi DNA Polymerase (ThermoFisher Scientific). For a 25 μ l reaction: 5 μ l of SuperFi 5X buffer, 0.5 μ l 10 mM dNTP mix, 1.25 μ l forward primer, 1.25 μ l reverse primer, 1 μ l template DNA diluted 1:500 in water, 0.25 μ l Platinum SuperFi DNA Polymerase, and water to 25 μ l. A 3-step thermocycler protocol was used: initial denaturation at 98°C for 30 s, 30 cycles of denaturing at 98°C for 5 s, annealing at a varying temperature for 10 s, extension at 72°C for varying time, and finally a final extension at 72°C for 5 minutes and a hold at 4°C. The annealing temperature was determined for each primer pair using the T_m calculator at www.thermofisher.com/tmcalculator and the extension time was calculated based on product size using 15s/kb.

For plasmid and PCR digests, 1 μ g of plasmid, or the full PCR reaction, was digested in 50 μ l according to New England Biolabs' (NEB) recommendations: 5 μ l of the appropriate buffer, 1 μ g of plasmid or the PCR reaction, 2 μ l of each restriction enzyme, and water to 50 μ l. The reaction was incubated at 37°C for 30 minutes to overnight. Plasmids were gel purified using the Qiagen QiaexII Gel Extraction Kit. PCR product digestions were heat-inactivated at the appropriate temperature or were column purified with the Roche High Pure PCR Purification Kit if the enzymes could not be heat-inactivated. For plasmid digests with only one restriction enzyme, 1 μ l of Shrimp Alkaline Phosphatase (rSAP) (NEB) was added during the digestion.

For gel purifications using the QiaexII Gel Extraction Kit, the plasmid was run on an 1-2% agarose gel and the band of interest was cut out of the gel and placed in a 1.5 ml microcentrifuge tube. Depending on the size of the DNA fragment, QX1 buffer and water were added. For fragments under 4 kb 750 μ l of QX1 was added, for fragments over 4 kb an additional 500 μ l of water was also added. QIAEXII was resuspended by vortexing and 10 or 30 μ l was added to the DNA. The solution was incubated at 50°C for 10 minutes with shaking. The sample was sedimented by centrifugation at 17,900xg for 30 s and the supernatant removed. The pellet was washed with 500 μ l of QX1 and the pellet was resuspended by mixing on a vortex. The sample was sedimented again as above. After removing the supernatant, the pellet was washed by resuspending in 500 μ l Buffer PE twice. The pellet was air-dried for 15 minutes and resuspended in water or tris-EDTA (TE) buffer. The resuspended pellet was centrifuged and the supernatant containing the purified DNA was transferred to a clean tube.

For Roche High Pure PCR Purification, the PCR reaction volume was adjusted to 100 μ l with water, then 500 μ l of Binding Buffer was added. The solution was transferred to a column and centrifuged at max for 1 minute. The flow through was discarded and 500 μ l of Wash buffer was added, and the column was centrifuged again. The flow through was discarded and 200 μ l of Wash buffer was added, and the column was centrifuged again. The column was transferred to a clean tube and 50 μ l of Elution Buffer was added to the column. The column was centrifuged at max for 1 minute to elute the DNA.

Site-directed mutagenesis was done using the QuikChange II XL Site-Directed Mutagenesis kit from Agilent. For a 50 μ l reaction: 5 μ l of 10x reaction buffer, 1 μ l of plasmid template, 1.25 μ l of primer 1, 1.25 μ l of primer 2, 1 μ l of dNTP mix, 3 μ l of

QuikSolution, and water to 50 μ l. Then 1 μ l of *PfuUltra* HF DNA polymerase was added. The following thermocycler parameters were used: initial denaturation at 95°C for 1 min, 16 cycles of denaturation at 95°C for 50 s, annealing at 60°C for 50 s, and extension at 68°C for 1 minute per kb of plasmid size. After cycling, the reaction was placed on ice for 2 minutes and then treated with 1 μ l of DpnI and digested at 37°C for 1 hour. The DNA was used to transform XL10-Gold Supercompetent cells by thawing the cells on ice. For each transformation, 45 μ l of cells were transferred to a prechilled 14 ml Falcon tube and 2 μ l of β -ME was added. The cells were incubated on ice for 10 minutes with swirling and then 2 μ l of the DpnI treated DNA was added. The reactions were incubated on ice for 30 minutes and then heat shocked in a 42°C water bath for 30 s. The cells were incubated on ice for 2 minutes and then 0.5 ml of SOC outgrowth medium (NEB) was added to each tube and the tubes were incubated at 37°C with shaking at 250 rpm for 1 hour. The transformants were plated on Luria Broth Agar-Ampicillin (100 mg/ml) (LB-Amp) plates using 250 μ l on each of two plates and incubated overnight at 37°C.

For NEBuilder HiFi Assembly reactions, PCR products or synthesized inserts were designed according to NEB's recommendations. The inserts were designed to have a 20-40-nt overlap with either the second insert or the vector with a T_m equal to or greater than 48°C. PCR products generated from plasmids were digested with DpnI by adding 1 μ l of 10X CutSmart Buffer (NEB), 1 μ l of DpnI (NEB), and 8 μ l of PCR reaction and then incubating at 37°C for 30 minutes. The DpnI was heat-inactivated by incubating at 80°C for 20 minutes. PCR products generated from synthesized DNA were not treated with DpnI. To quantify the amount of DNA in each sample, 1 μ l of insert and 1 μ l of prepared

vector was run on an agarose gel. The band intensity and relative size of each fragment were used to determine the volume of each to add to the assembly reaction to achieve the recommended ratio for the assembly reaction, for 2-3 fragment assembly a vector: insert ratio of 1:2, or for 4-6 fragment assembly a vector to insert ratio of 1:1. In the reactions, 0.5 μl of vector was used for a 10 μl reaction volume and 5 μl of NEBuilder HiFi DNA Assembly Master Mix was added to the DNA. The volume was adjusted to 10 μl with water and the reaction was incubated at 50°C for 15 minutes for a 2-3 fragment assembly or 60 minutes for a 4-6 fragment assembly. After incubation the samples were placed on ice for transformation. To transform NEB 5-alpha competent *E. coli* (C29871/H), the chemically competent cells were thawed on ice and then 2 μl of the chilled assembly product was added to the cells and mixed in by gentle flicking. The mixture was placed in ice for 30 minutes then heated shocked at 42°C for 30 s. The tube was transferred back onto ice for 2 minutes. Then 950 μl of SOC media was added to the tube and the tube was incubated at 37°C for 1 hour with shaking at 250 rpm. The LB-Amp (100 mg/ml) plates were warmed to 37°C and 100 μl of the cells was plated, and the plates were incubated overnight at 37°C or for 2-3 days at room temperature (RT).

The NEB Quick Ligation Kit was used for ligation if HiFi Assembly was not used. The relative amounts of DNA were determined as above, and a molar ratio of 1:3 vector to insert was used. For a 20 μl reaction, 1 μl of vector was used, and the insert was scaled accordingly. For a 20 μl reaction, 10 μl of Quick Ligase Reaction Buffer was added to the DNA fragments and water was added to 20 μl , then 1 μl of Quick Ligase was added. The reaction was mixed and incubated at RT for 5 minutes. The reaction was then placed on ice before 2 μl was used to transform 50 μl of competent cells as above for HiFi Assembly.

Plasmids used for sequencing or as templates for PCR were prepared using the Zyppy Plasmid Miniprep Kit (Zymo). A 3 ml *E. coli* culture was grown overnight in LB-Amp (100 mg/ml). After growing overnight, the culture was transferred to a 1.5 ml tube and centrifuged for 1 minute to pellet the cells. The supernatant was removed, and the pellet was resuspended in 500 µl of water. After resuspending the cell pellet, 100 µl of 7X Lysis Buffer was added and the cells were mixed by inverting 4-6 times until the solution was blue. Then 350 µl of cold Neutralization Buffer was added, and the solution was mixed until it was white. The tube was centrifuged at 11,000-16,000 xg for 2 minutes. The clear supernatant was transferred into the Zymo-Spin IIN column and centrifuged at max for 15 s. The flow through was discarded and 200 µl of Endo-Wash Buffer was added to the column. The column was centrifuged at max for 15 s and 400 µl of Zyppy Wash Buffer was added to the column. The column was centrifuged for 30 s and then transferred to a clean 1.5 ml tube. After transferring to a clean tube, 30 µl of Zyppy Elution Buffer was added to the column, or 30 µl of water was added if the plasmid would be used for sequencing. The column was incubated for 1 minute at RT and then centrifuged at max for 15 s to elute the DNA. The plasmid concentration was determined using a NanoDrop ND-1000 Spectrophotometer.

2.1.2. Table 2.1. Primers used in the study

ID	Name	Sequence
P1	p2lucACG_F	5'-CTAGGCTTTTGCAAAAAGCTTTAGGTGATCGGAATCACGGC-3'
P2	p2lucATG_F	5'-CTAGGCTTTTGCAAAAAGCTTTAGGTGATCGGAATCATGGC-3'
P3	p2lucBamHI_uCC_R	5'-CTTCGAGCTCAGGGAAGTTGAAGGATCCAAACCTAAAGGAAAAGAA G-3'
P4	p2lucBamHI_FS_R	5'-CTTCGAGCTCAGGGAAGTTGAAGGATCCAAACCTAAAGGAAAAGA AG-3'
P5	NheIStopACGuCC_F	5'-GGCTAGCTAGGTGATCGGAATCACGGCTATACACGGGATATCACG-3'
P6	NheIStopATGuCC_F	5'-GGCTAGCTAGGTGATCGGAATCATGGCTATACACGGGATATCACG-3'
P7	NheIStopNoStartuCC_F	5'-GGCTAGCTAGGTCAACGGAAAAAAGCTATACACGGGATATCACG-3'
P8	UTRrenAvaI_R	5'-CTCGGGGTCGTACACCTTGAAGCCATGGTGGCAAAAGACAAATTC GAAGAG-3'
P9	FSuCC R2	5'-CTAAAAACCTAAAGGAAAAGAAG-3'
P10	FSuCC F2	5'-ACTTCTTTTCCTTTAGGTTTTTAGGAGTTAGGGTTTGTAGTG-3'
P11	phRL_hifi_AvaI_R	5'-GTGATCATGCGTTTTCGCTCGGGGTCGTACACCTTGG-3'
P12	Trip12AAAAAA F	5'-AGGCTAGCTAGTTAAAAAAACCGAGTAAAAGTGGTGGCTC-3'
P13	modAAAAAAuCC_R	5'-GCTTTTTTCCGGTCGACCTGCGTTTCTTGGTGGACTTTC-3'
P14	Trip12ATCACG_F	5'-AGGCTAGCTAGTTAATCACGCCGAGTAAAAGTGGTGGCTC-3'
P15	modATCACGuCC_R	5'-GCCGTGATTCCGGTCGACCTGCGTTTCTTGGTGGACTTTC-3'
P16	modAAAAAAuCC_F	5'-CAGGTCGACCGAAAAAAGCTATACACGGGATATCACGG-3'
P17	modATCACGuCC_F	5'-CAGGTCGACCGAATCACGGCTATACACGGGATATCACGG-3'
P18	HF_NheI_Trip12_F	5'-CTTAATACGACTCACTATAGGCTAGCTAGTTAA-3'
P19	ATGSubsAmpHiFi_F	5'-CTTAATACGACTCACTATAGGCTAGCTAGGTGATCGGAATCATGGCT AT-3'
P20	ACGSubsAmpHiFi_F	5'-CTTAATACGACTCACTATAGGCTAGCTAGGTGATCGGAATCACGGCT AT-3'
P21	uCCSubs_R	5'-CCTAAAGGAAAAGAAGTTAAAGCTGG-3'
P22	uCC5UTR_F	5'-CCAGCTTTAACTTCTTTTCCTTTAGG-3'
P23	phRL_Hifi_NheI_F	5'-CTTAATACGACTCACTATAGGCTAGCTAGGTGATCGGAAT-3'
P24	uCC_ACG_ATG_F	5'-GAACAGGTGATCGGAATCATGGCTATACACGGGATATC-3'
P25	uCC_ACG_ATG_R	5'-GATATCCCGTGTATAGCCATGATTCCGATCACCTGTTC-3'
P26	uCCint_ATG_AAG_F	5'-GGGTGTTAGCTCACAAGTCCATATTGTCCGACAG-3'
P27	uCCint_ATG_AAG_R	5'-CTGTCGGACAATATGGACTTGTGAGCTAACACCC-3'
P28	Frag1_F	5'-CGAGCTCGGTACCTAATACG-3'
P29	uCC_R	5'-CCTTTCTTTATGTTTTTGCGTCTTCGTCGAC-3'

P30	Fusion SalI LucF	5'- CTTTAACTTCTTTTCCTTTAGGTTTGTGCGACGAAGACGCCAAAAACAT AAAG-3'
P31	LucPolyA50_HindIII_p uc19_R	5'- GCTATGACCATGATTACGCCAAGCTTCACGTTGTAACGACGGCCA GTG-3'
P32	FSDel_R	5'-CCACGTCTCCTGCGGCAAAG-3'
P33	FS_Del F	5'-CTTTGCCGAGGAGGACGTGG-3'
P34	FSInsertion_R	5'-CCTTTCTTTATGTTTTTGGCGTCTTCGTCGACAAAACCTAAAGG-3'
P35	CPA1AmpF	5'- CGACTCACTATAGCAAGAAG-3'
P36	CPA1_D13N_R	5'- GATATGTAGTTTTTGGCAGGTG-3'
P37	CPA1_D13N F	5'- CACCTGCCAAAACCTACATATC-3'
P38	uCCAmpR	5'- CCTTTCTTTATGTTTTTGGCGTCTTC-3'
P39	uCC_ATC_TTC_ACG TCG F	5'-GGGAACAGGTGATCGGATTCTCGGCTATACACGGGATATC-3'
P40	uCC_ATC_TTC_ACG TCG R	5'-GATATCCCGTGTATAGCCGAGAATCCGATCACCTGTTCCC-3'
P41	FSuCC_R2	5'-CTAAAAACCTAAAGGAAAAGAAG-3'
P42	FSuCC_F2	5'- CTTCTTTTCCTTTAGGTTTTTAG-3'
P43	ACCAUG_R2	5'- CCTTTCTTTATGTTTTTGGCGTCTTCGTCGACCATGGTTTAGTTAATCA GTTCCCCGGATTTTTTC-3'
P44	UTRrenAvaI_R	5'- CTCGGGGTCGTACACCTTGAAGCCATGGTGGCAAAAGACAAATTC GAAGAG-3'
P45	FSuCC_R3	5'- CTAAAAACCTAAAGGAAAAGAAG-3'
P46	FSuCC_F3	5'- ACTTCTTTTCCTTTAGGTTTTTAGGAGTTAGGGTTTGTAGTG-3'
P47	phRL_hifi_AvaI_R	5'- GTGATCATGCGTTTGCCTGCTCGGGGTCGTACACCTGG-3'
P48	Trip12AAAAAA F	5'-AGGCTAGCTAGTTAAAAAAACCGAGTAAAACTGGTGGCTC-3'
P49	modAAAAAAuCC_R	5'-GCTTTTTTTCCGGTCGACCTGCGTTTCTTGGTGGACTTTC-3'
P50	gp48_F	5'-CCCCGCTAGCCGCCGAATAC-3'
P51	gp48_R	5'-ATGCGGATCCGAAGTCCAAATTTTAAG-3'
P52	uCCStall_R	5'-ATGCGGATCCGAAGTCCAAATTC-3'
P53	EcoRI uCC F	5'-GACTCGAATCCGCCGTAATACGACTCAC-3'
P54	HindIII pramod_R	5'-GAGTCAAGCTTGAAGTCCAAATTC-3'
P55	Pramod_FLUC_F	5'- CGACTCACTATAGCAAGAAGCTAGCGCAACAACAACAAACATTTGC- 3'
P56	uCC_SalI_FLUC_R	5'- CCTTTCTTTATGTTTTTGGCGTCTTCGTCGACAAAACCTAAAGGAAAAG AAGTTAAAGC-3'
P57	OF_SalI_FLUC_R	5'- CCTTTCTTTATGTTTTTGGCGTCTTCGTCGACAAAACCTAAAGGAAAA GAAGTTAAAGC-3'
P58	occ1096	5'-GCTATTCCAGAAGTAGTGAGGAGGC-3'
P59	occ1098	5'-GTTGACGCAAATGGGCGTAGGCGTGT-3'
P60	uCCseq_F	5'-GATCGGTGCGGGCCTTTCGC-3'
P61	uCCseq_R	5'-GTTACCTCGATATGTGCATCTG-3'

P62	SLseq F	5'-CCAGTTCGCCCATTCTCCG-3'
P63	SLseq R	5'-CTTCAGACGGCAGAGCTCCAC-3'
P64	HindIIIp2lucTrip12 F	5'- GCAAAAAGCTTTAGGTGATCGGAATCATGCCGAGTAAAACTGGTGGC TC-3'
P65	p2lucBamHI-uCC R	5'- CTTCGAGCTCAGGGAAGTTGAAGGATCCAAACCTAAAGGAAAAGAA G-3'

2.1.3. The synthesized G-blocks

The *Arabidopsis thaliana* *GGP* uCC contains an out-of-frame internal uORF that is not conserved in ancient plants (Fig. 2.1A). The internal uORF was disrupted changing the ATG start codon to AAG, creating an H13Q amino acid substitution in the uCC sequence (Fig. 2.1B).

G-block 1: *GGP* uCC wild-type sequence with *GGP* uCC 5' leader for insertion into pUC19. KpnI, NheI, and SalI restriction sites are indicated in red. The T7 promoter is highlighted in yellow, the *GGP* uCC is underlined, the internal uORF start site is shown in cyan, and firefly luciferase (FLuc) is highlighted in green.

```
GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAACCCACATCAAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTCGGGTTTCGTTTTAGCTGTAATCTTTTTTTCCGGCGTTTCGATTTGAAAAAT
CCGGGGAACAGGTGATCGGAATCACGGCTATACACGGGATATCACGGGGTGTAGCTCACATGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCGTCTGA
AGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTGTCGACGAAGACGCCAAAAACATAAAGAAAGGCCCGGCCAT
```

G-block 2: *GGP* uCC wild-type sequence with *GGP* uCC 5' leader, the uCC, and the region between the uCC stop codon and the *GGP* start site for the uORF reporter in pUC19. KpnI, NheI, and SalI restriction sites are indicated in red. The T7 promoter is highlighted in yellow, the *GGP* uCC is underlined, the internal uORF start site is shown in cyan, and FLuc is highlighted in green.

```
GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAACCCACATCAAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTCGGGTTTCGTTTTAGCTGTAATCTTTTTTTCCGGCGTTTCGATTTGAAAAAT
CCGGGGAACAGGTGATCGGAATCACGGCTATACACGGGATATCACGGGGTGTAGCTCACATGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCGTCTGA
```

AGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTTAGGAGTTAGGGTTTGTAGTGTTTTTCTTCTTCTTTTTTTGGTGCTCTTGAATCGCTTTTTTCTTG
GGGAAGTTTTTTCTTTTGCTCTTCGAAATTTGTCTTTTTTTGAGAATGGTCGACGAAGACGCCAAAAACAT
AAAGAAAGGCCCGGCGCCAT

G-block 3: *GFP* uCC with an AUG start and alanine substitutions G52A, G53A, G54A (GGG/3A). KpnI, NheI, and the SalI restriction sites are indicated in red, the T7 promoter is highlighted in yellow, the uCC is underlined, and modifications to the uCC are shown in fuchsia. FLuc is highlighted in green.

GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAACCCACATCAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTCGGGTTTCGTTTTAGCTGTAATCTTTTTTTCCGGCGTTTCGATTTGAAAAAAT
CCGGGGAACAGGTGATCGGAATCATGGCTATACACGGGATATCACGGGGTGTAGCTCACAAGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCGTCTGA
AGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGCTGCAGCTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTGTCGACGAAGACGCCAAAAACATAA

G-block 4: *GFP* uCC with an AUG start and alanine substitutions P44A, S45A, D46A (PSD/3A). KpnI, NheI, and SalI are indicated in red, the T7 promoter is highlighted in yellow, the uCC is underlined, and modifications to the uCC are shown in fuchsia. FLuc is highlighted in green.

GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAACCCACATCAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTCGGGTTTCGTTTTAGCTGTAATCTTTTTTTCCGGCGTTTCGATTTGAAAAAAT
CCGGGGAACAGGTGATCGGAATCATGGCTATACACGGGATATCACGGGGTGTAGCTCACAAGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCGTCTGA
AGGCGGCAGCGCTGCAGCTCTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTGTCGACGAAGACGCCAAAAACATAA

A 1 M A I H G I S R G V S S H V H I V R Q K
 1 ACGGCTATACAGGGATATCACGGGGTGTTAGCTCACATGTC **CAIAT** TGTC **CCGACAGAAAG**
 21 G C L I E T N P L P H G G R G A L P S E
 61 **GGT** **TGTTTTAA** TCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCCGTCTGAA
 41 G G S P S D L L F L A G G S S F N F F
 121 GCGGCAGCCCTTCCGATCTCCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTT
 61 S F R F *
 181 TCCTTTAGGTTT**TAG**

B 1 M A I H G I S R G V S S Q V H I V R Q K
 1 ATGGCTATACAGGGATATCACGGGGTGTTAGCTCA**CAA**GTCCATATTGTC **CCGACAGAAAG**
 21 G C L I E T N P L P H G G R G A L P S E
 61 GGT**TGTTTTAA**TCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCCGTCTGAA
 41 G G S P S D L L F L A G G S S F N F F
 121 GCGGCAGCCCTTCCGATCTCCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTT
 61 S F R F
 181 TCCTTTAGGTTT

C 1 M A I H G I S R G V S S Q V H I V R Q K
 1 ATGGCTATACAGGGATATCACGGGGTGTTAGCTCA**CAA**GTCCATATTGTC **CCGACAGAAAG**
 21 G C L I E T N P L P Q E D V E L C R L K
 61 GGT**TGTTTTAA**TCGAAACTAATCCCTTTGCCGCAGGAGGACGTGGAGCTCTGCCCGTCTGAAAG
 41 A A A L P I S S F S P V A V P A L T S F
 121 GCGGCAGCCCTTCCGATCTCCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTT
 61 P L G F
 181 CCTTTAGGTTT

Figure 2.1. Sequence information for the *A. thaliana* GGP uCC

Figure 2.1. Sequence information for the *A. thaliana* GGP uCC

The *A. thaliana* GGP uCC was used to generate the reporters used in this study. (A) The uCC contains an internal out-of-frame uORF that is not highly conserved shown in orange. (B) The internal out-of-frame uORF was disrupted by a nucleotide substitution in the underlined His codon of the uCC, generating an H13Q substitution. (C) The control out-of-frame uCC (referred to as OF) was generated by removing one nucleotide from the CAC codon at H31 (underlined) generating a frameshift. The reading frame was restored by inserting a T (purple) immediately upstream of the stop codon. The start codons are shown in green (either AUG or ACG depending on the reporter), changes to the sequence are shown in purple, and the stop codon is shown in red.

G-block 5: GGP uCC with an AUG start and alanine substitutions G41A, G42A (GG/2A). KpnI, NheI, and Sall are indicated in red, the T7 promoter is highlighted in yellow, the uCC is underlined, and modifications to the uCC are shown in fuchsia. FLuc is highlighted in green.

```
GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTTGTGAAAAAACCCACATCAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTCGGGTTTCGTTTTAGCTGTAATCTTTTTTCCGGCGTTTCGATTTGAAAAAT
CCGGGGAACAGGTGATCGGAATCATGGCTATACACGGGATATCACGGGGTGTAGCTCACAAAGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCGTCTGA
AGCTGCTAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTGTCGACGAAGACGCCAAAAACATAAA
```

G-block 6: GGP uCC with an AUG start and alanine substitutions P38A, S39A (PS/2A). KpnI, NheI, and Sall are indicated in red, the T7 promoter is highlighted in yellow, the

uCC is underlined, and modifications to the uCC are shown in fuchsia. FLuc is highlighted in green.

GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAAACCCACATCAAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTGCGGGTTCGTTTTAGCTGTAATCTTTTTTTCCGGCGTTTCGATTTGAAAAAAT
CCGGGGAACAGGTGATCGGAATCATGGCTATACACGGGATATCACGGGGTGTAGCTCACAAGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGGCTGCAGA
AGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTGTCGACGAAGACGCCAAAAACATAAA

G-block 7: *GFP* uCC with an AUG start and alanine substitutions G32A, G33A, R34A, G35A (GGRG/4A). KpnI, NheI, and Sall are indicated in red, the T7 promoter is highlighted in yellow, the uCC is underlined, and modifications to the uCC are shown in fuchsia. FLuc is highlighted in green.

GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAAACCCACATCAAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTGCGGGTTCGTTTTAGCTGTAATCTTTTTTTCCGGCGTTTCGATTTGAAAAAAT
CCGGGGAACAGGTGATCGGAATCATGGCTATACACGGGATATCACGGGGTGTAGCTCACAAGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGCTGCTGCAGCTGCTCTGCCGTCTGA
AGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTGTCGACGAAGACGCCAAAAACATAAA

G-block 8: *GFP* uCC with an AUG start and alanine substitution R34A. KpnI, NheI, and Sall are indicated in red, the T7 promoter is highlighted in yellow, the uCC is underlined, and modifications to the uCC are shown in fuchsia. FLuc is highlighted in green.

GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
ATTCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAAACCCACATCAAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTGCGGGTTCGTTTTAGCTGTAATCTTTTTTTCCGGCGTTTCGATTTGAAAAAAT
CCGGGGAACAGGTGATCGGAATCATGGCTATACACGGGATATCACGGGGTGTAGCTCACAAGTCCATATT

GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGAGCTGGAGCTCTGCCGTCTGA
AGGCGGCAGCCCTTCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGTT
TGTCGACGAAGACGCCAAAAACATAAA

G-block 9: *GGP* uCC with an AUG start and alanine substitutions P38A, S39A, E40A, G41A, G42A, S43A, P44A, S45A, D46A (9A). KpnI, NheI, and SalI are indicated in red, the T7 promoter is highlighted in yellow, the uCC is underlined, and modifications to the uCC are shown in fuchsia. FLuc is highlighted in green.

GGTACCTAATACGACTCACTATAGCAAGAAGCTAGCTTCGTATCATCAAAAAACACCTCAAAGAATTATTC
AATCAGGCATCTTCTCAAATTTTTGTTTGTGAAAAAACCCACATCAAAAGATCTCTCATTATTCGTTTC
GTTTCTGCTGTTTTGAGTGTCCGGTTCGTTTTAGCTGTAATCTTTTTTCCGGCGTTCGATTTGAAAAAT
CCGGGAACAGGTGATCGGAATCATGGCTATACACGGGATATCACGGGGTGTAGCTCACAAGTCCATATT
GTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCGCCGC
CGCTGCTGCTGCAGCAGCACTCCTCTTTCTCGCCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGT
TTGTCGACGAAGACGCCAAAAACATAAA

G-block 11: *GGP* uCC with S11A substitution highlighted in cyan for insertion into p2luc. HindIII and BamHI are indicated in red, the uCC is underlined, and FLuc is highlighted in green.

AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTTTAGGTGATCGGAATCACGGCTATA
CACGGGATATCACGGGGTGTTCCTCACAAGTCCATATTGTCCGACAGAAGGGTTGTTTAATCGAAACTAA
TCCTTTGCCGCACGGAGGACGTGGAGCTCTGCCGTCTGAAGGCGGCAGCCCTTCGATCTCCTCTTTCTCG
CCGGTGGCGGTTCCAGCTTTAACTTCTTTTCCTTTAGGTTTGGATCCTTCAACTCCCTGAGCTCGAAGAC
GCCAAAAACATAAAGA

G-block 12: *GGP* uCC with an ACG start and an alanine insertion three codons upstream of the GGRG motif (1A@P28) for insertion into p2luc. The alanine codon insertion is shown in cyan. HindIII and BamHI are indicated in red, the uCC is underlined, and FLuc is highlighted in green.

GGAGGCCTAGGCTTTTGC AAAAAGCTTTAGGTGATCGGAATCACGGCTATACACGGGATATCACGGGGTGT
TAGCTCACAAGTCCATATTGTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTGCTTTGCCGCACGGAG
GACGTGGAGCTCTGCCGTCTGAAGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGC
TTTAACTTCTTTTCCTTTAGGTTTGGATCCTTCAACTTCCTGAGCTCGAAG

G-block 13: *GGP* uCC with an ACG start and three alanine codons inserted three codons upstream of the GGRG motif (3A@P28) for insertion into p2luc. The alanine codon insertions are shown in cyan. HindIII and BamHI are indicated in red, the uCC is underlined, and FLuc is highlighted in green.

GGAGGCCTAGGCTTTTGC AAAAAGCTTTAGGTGATCGGAATCACGGCTATACACGGGATATCACGGGGTGT
TAGCTCACAAGTCCATATTGTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTGCTGCCGCTTTGCCGC
ACGGAGGACGTGGAGCTCTGCCGTCTGAAGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGT
TCCAGCTTTAACTTCTTTTCCTTTAGGTTTGGATCCTTCAACTTCCTGAGCTCGAAG

G-block 14: *GGP* uCC with an ACG start and alanine insertion between P and D (1A@p44) for insertion into p2luc. The alanine insertion is shown in cyan. HindIII and BamHI are indicated in red, the uCC is underlined, and FLuc is highlighted in green.

GGAGGCCTAGGCTTTTGC AAAAAGCTTTAGGTGATCGGAATCACGGCTATACACGGGATATCACGGGGTGT
TAGCTCACAAGTCCATATTGTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGAC
GTGGAGCTCTGCCGTCTGAAGGCGGCAGCCCTGCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGTTCCAGC
TTTAACTTCTTTTCCTTTAGGTTTGGATCCTTCAACTTCCTGAGCTCGAAG

G-block 15: *GGP* uCC with an ACG start and three alanine codons inserted between P and D (3A@p44) for insertion into p2luc. The alanine codon insertions are shown in cyan. HindIII and BamHI are indicated in red, the uCC is underlined, and FLuc is highlighted in green.

GGAGGCCTAGGCTTTTGC AAAAAGCTTTAGGTGATCGGAATCACGGCTATACACGGGATATCACGGGGTGT
TAGCTCACAAGTCCATATTGTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTGCCGCACGGAGGAC
GTGGAGCTCTGCCGTCTGAAGGCGGCAGCCCTGCTGCCGCTTCCGATCTCCTCTTTCTCGCCGGTGGCGGT
TCCAGCTTTAACTTCTTTTCCTTTAGGTTTGGATCCTTCAACTTCCTGAGCTCGAAG

G-block 16: *GFP* uCC tagged with FLAG. The T7 promoter is highlighted in yellow, FLAG is highlighted in green, a linker region added to increase the protein molecular weight is highlighted in cyan. The uCC is underlined, and modifications to the uCC are shown in fuchsia. *NheI* and *BamHI* are indicated in red.

CCCCGCTAGCCGCCGTAATACGACTCACTATAGGCAACAACAACAACATTTGCTTCTGACACAACCTGTGT
 TCACTAGCAACCTCAAACAGACACCATGGACTACAAAGACCACGACGGTGATTATAAAGATCACGACATCG
 ATTACAAGGACGACGACGACAAGTCCAAGGAGCCGCTTCGGCCACGGTGCCGCCCATCAACGCCACCCTG
 GCTGTGGAGAAGGAGGGCTGCCCGTGTGCATCACCGTCAACACCACCATCTGTGCCGGCTACTGCCCCAC
 CGCAACCCGCGTGCTGCAGGGGGTCTGCCGGCCCTGCCTCAGGTGGTGTGCAACTACCGCCGGTCCGTAA
 CCCACCGTATTCTTACCGTTCGGATTGCCCAAGATCAAGTGGGCGCATACTATCAGCAACCAGGTCAGCAG
 AACGCCACCTGGATTGTGCCACCAGGGCAGTATTTTCGACGAGGTGATAACCGCGACAACAGTGCCGATTC
 ACGGTATTGGGGCTTTGTGCCTGAAGCTAACCTGGTAGGCAGGGCAACTGCCATCTGGGCTATACACGGGA
 TATCACGGGGTGTTAGCTCACAAGTCCATATTGTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTG
 CCGCACGGAGGACGTGGAGCTCTGCCGTCTGAAGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGG
 CGGTTCCAGCTTTAACTTCTTTTCCTTTAGGTTTTAGAAATTTGGACTTCGGATCCGCAT

G-block 17: Out-of-frame (OF) *GFP* uCC tagged with FLAG. The T7 promoter is highlighted in yellow, FLAG is highlighted in green, a linker region added to increase the protein molecular weight is highlighted in cyan. The uCC is underlined, and modifications to the uCC are shown in fuchsia. One nucleotide was removed from the CAC codon to change the codon to CAG and shift the frame of the conserved region. One codon was added at the end of the uCC to put the stop back in-frame. *NheI* and *BamHI* are indicated in red.

CCCCGCTAGCCGCCGTAATACGACTCACTATAGGCAACAACAACAACATTTGCTTCTGACACAACCTGTGT
 TCACTAGCAACCTCAAACAGACACCATGGACTACAAAGACCACGACGGTGATTATAAAGATCACGACATCG
 ATTACAAGGACGACGACGACAAGTCCAAGGAGCCGCTTCGGCCACGGTGCCGCCCATCAACGCCACCCTG
 GCTGTGGAGAAGGAGGGCTGCCCGTGTGCATCACCGTCAACACCACCATCTGTGCCGGCTACTGCCCCAC
 CGCAACCCGCGTGCTGCAGGGGGTCTGCCGGCCCTGCCTCAGGTGGTGTGCAACTACCGCCGGTCCGTAA

CCCACCGTATTCTTACCGTTCCGATTGCCCAAGATCAAGTGGGCGCATACTATCAGCAACCAGGTCAGCAG
 AACGCCACCTGGATTGTGCCACCAGGGCAGTATTTTCGCAGCAGGTGATAACCGCGACAACAGTGCCGATTC
 ACGGTATTGGGGCTTTGTGCCTGAAGCTAACCTGGTAGGCAGGGCAACTGCCATCTGGGCTATACACGGGA
 TATCACGGGGTGTAGCTCACAAAGTCCATATTGTCCGACAGAAGGGTTGTTTAATCGAAACTAATCCTTTG
 CCGCAGGAGGACGTGGAGCTCTGCCGTCTGAAGGCGGCAGCCCTTCCGATCTCCTCTTTCTCGCCGGTGGC
 GGTTCCAGCTTTAACTTCTTTTCTTTAGGTTTTTAGAATTTGGACTTCGGATCCGCAT

G-block 18: hCMV uORF2 UL4/gp48 stall control tagged with FLAG. The T7 promoter is highlighted in yellow, FLAG is highlighted in green, a linker region added to increase the protein molecular weight is highlighted in cyan. UL4 is underlined. *NheI* and *BamHI* are indicated in red.

CCCCGTAGCCGCGGTAATACGACTCACTATAGGCAACAACAACAACATTTGCTTCTGACACAACCTGTGT
 TCACTAGCAACCTCAAACAGACACCATGGACTACAAAGACCACGACGGTGATTATAAAGATCAGCAGATCG
 ATTACAAGGACGACGACGACAAGTCCAAGGAGCCGCTTCGGCCACGGTGCCGCCCATCAACGCCACCCTG
 GCTGTGGAGAAGGAGGGCTGCCCGTGTGCATCACCGTCAACACCACCATCTGTGCCGGCTACTGCCCCAC
 CGCAACCCGCGTGTGCAGGGGGTCTGCCGGCCCTGCCTCAGGTGGTGTGCAACTACCGCCGGTCCGTAA
 CCCACCGTATTCTTACCGTTCCGATTGCCCAAGATCAAGTGGGCGCATACTATCAGCAACCAGGTCAGCAG
 AACGCCACCTGGATTGTGCCACCAGGGCAGTATTTTCGCAGCAGGTGATAACCGCGACAACAGTGCCGATTC
 ACGGTATTGGGGCTTTGTGCCTGAAGCTAACCTGGTAGGCAGGGCAACTGCCATCTGGGAACCGCTGGTGC
 TGAGTGCGAAAAAACTGAGCAGCCTGCTGACCTGCAAATATATTCTCCTTAAAATTTGGACTTCGGATCC
 GCAT

2.1.4. Generation of the uCC-FLuc fusion *in vitro* translation reporters (Table 2.2)

To generate plasmid pC5568 for *in vitro* transcription (Fig. 2.2A), a gene fragment was synthesized and cloned into the plasmid pCR2.1 vector backbone, G-block #1 (Eurofins Genomics) (Fig. 2.1B). G-block #1 was sub-cloned into the pUC19 backbone vector between *KpnI* and *HindIII* using a rapid ligation, as described above, and correct insertion was verified by sequencing with M13 universal primers M13F(-40) and

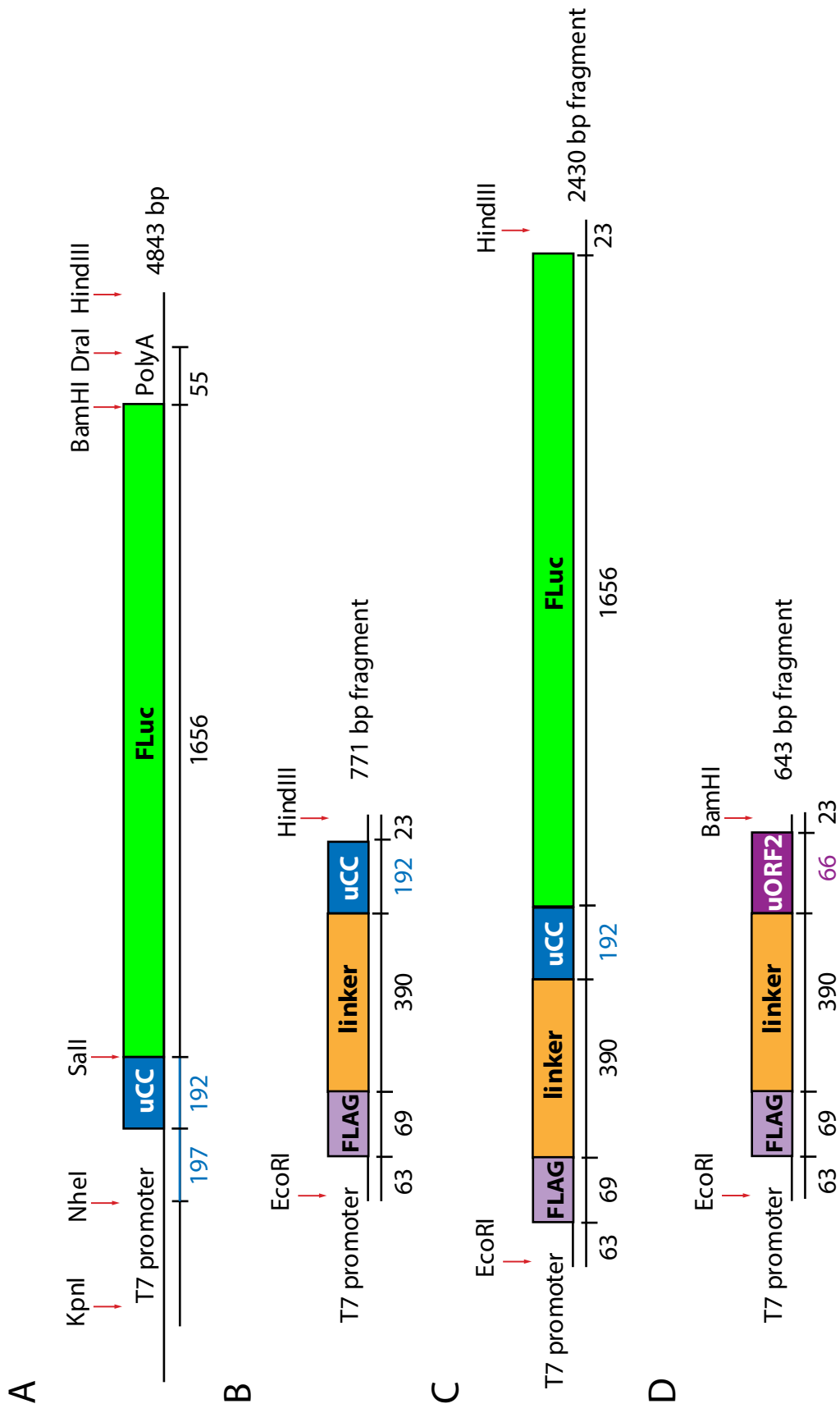


Figure 2.2. Templates used for *in vitro* transcription of mRNAs used in RRL assays

Figure 2.2. Templates used for *in vitro* transcription of mRNAs used in RRL assays

Plasmid or synthesized dsDNA G-block fragments were used as templates for *in vitro* transcription of mRNAs used in the *in vitro* translation assays. *GGP* uCC specific sequences are indicated in blue and all templates include a T7 promoter for transcription.

(A) To test the *GGP* uCC response to ascorbate, the uCC was fused in-frame to the FLuc coding region followed by 22 nts and a template encoded poly(A) tail. (B) For the peptidyl-tRNA release assay templates, FLAG was fused upstream of a 390-bp linker fragment encoding residues 21-81 from the human chorionic gonadotropin beta subunit and fused to the *GGP* uCC. (C) The reporter FLAG-linker-uCC fusion was fused in-frame to FLuc. (D) The same FLAG-linker element was fused in frame to the sequence of uORF2 from the hCMV UL4/gp48 mRNA, which is known to cause stalling at translation termination.

M13R(-40). The resulting plasmid was named pC5560. To generate a uCC AUG start site (ATG), site-directed mutagenesis, as described above, was performed on plasmid pC5560 using primers uCC_ACG_ATG_F and uCC_ACG_ATG_R. Mutagenesis was confirmed by sequencing using M13 universal primers M13F(-40) and M13R(-40) and the plasmid was named pC5563. Because *A. thaliana* contains an internal out-of-frame non-conserved uORF (Fig. 2.1A), site-directed mutagenesis was used to change the internal start site from ATG to AAG using primers uCCint_ATG_AAG_F and uCCint_ATG_AAG_R. Mutagenesis was confirmed by sequencing using M13 universal primers M13F(-40) and M13R(-40).

The resulting plasmid was used as a template for PCR amplification of the uCC. The insert was generated with homology to pUC19 and a KpnI site upstream of the T7 promoter to generate a homologous 5' end for HiFi Assembly. At the 3' end, the uCC stop codon was replaced by a Sall site, followed by homology to FLuc with the start site removed. The PCR was performed as described above using primers Frag1_F and uCCampR to generate the uCC fragment including homology to pUC19, a KpnI site, the T7 promoter, the uCC with no stop codon followed by a Sall site, and FLuc without a start codon.

The FLuc fragment with a polyA tail was generated from plasmid pC365 (LucPolyA50) by PCR as described above. Primers were used to add a Sall site and to remove the FLuc start codon and to add pUC19 homology to the 5' end after the HindIII site. The primers used for the PCR reaction were Fusion_Sall_LucF and LucPolyA50_HindIII_puc19_R.

For the HiFi Assembly reaction, the fragments were treated as described above, and pUC19 was digested with KpnI and HindIII and gel purified. The three fragments were assembled as described above and correct insertion was confirmed by sequencing using M13 universal primers M13F(-40) and M13R(-40) and the plasmid was named pC5568 (Fig. 2.2A).

The AUG uCC OF reporter pC5867 was generated using fusion PCR (Fig. 2.1C, Fig. 2.2A). Using primers Frag1_F and FSDel_R and plasmid pC5568 as a template, one nucleotide was deleted in the codon H31Q, CAC to CA shown in purple, directly upstream of the first conserved motif. This shifted the reading from +0 to +1. The second fragment was generated using primers FS_Del_F and FSInsertion_R. Primer FS_Del_F added 21

bases of homology to the first fragment and FSInsertion_R added a single T nucleotide, shown in purple, at the end of the uCC coding region to restore the FLuc reading frame before the SalI restriction site (Fig. 2.1C). The fragments were mixed 1:1 and fused by PCR using primers Frag1_F and FSInsertion_R. The resulting PCR product was cloned into digested plasmid pC5568 using HiFi Assembly as described above. Correct assembly was confirmed by sequencing with uCCseq_F and uCCseq_R.

The reporters containing amino acid substitutions were generated using synthesized G-blocks #3-9 (LifeSct). The fragments were designed with homology for insertion between NheI and SalI in the existing plasmid pC5568 vector. Plasmid pC5568 was digested with NheI and SalI and gel purified. The synthesized fragments were inserted into the vector using NEBuilder HiFi DNA Assembly Master Mix (NEB). Correct insertion was confirmed by sequencing with primers uCCseq_F and uCCseq_R.

The reporters containing amino acid substitutions were generated using synthesized G-blocks #3-9 (LifeSct). The fragments were designed with homology for insertion between NheI and SalI in the existing pC5568 vector (Fig. 2.2A). Plasmid pC5568 was digested with NheI and SalI and gel purified. The synthesized fragments were resuspended in water to 50 ng/ μ l and inserted into the digested vector by HiFi Assembly as described above. Correct insertion was confirmed by sequencing using primers uCCseq_F and uCCseq_R.

Table 2.2. The *in vitro* translation uCC-FLuc fusion reporters used in the study

<u>uCC-FLuc substitutions</u>	<u>Stock ID</u>	<u>Source</u>
uCC	pC5568	This study
OF	pC5867	This study
GGG/3A	pC5731	This study
PSD/3A	pC5733	This study
GG/2A	pC5734	This study
PS/2A	pC5736	This study
GGRG/4A	pC5737	This study
R34A	pC5738	This study
9A	pC5740	This study

Table 2.2. The uCC-FLuc fusions for *in vitro* translation all initiate at an AUG codon.

2.1.5. Generation of the uORF *in vitro* translation reporters (Table 2.3)

To generate the uORF reporters for *in vitro* transcription, a gene fragment was synthesized as G-block #2 and cloned into the pCR2.1 vector backbone (Eurofins Genomics). G-block #2 was sub-cloned into the pUC19 backbone vector between KpnI and HindIII using Quick Ligation, as described above, and correct insertion was verified by sequencing using M13 primers and the plasmid was named pC5559.

To generate an AUG uCC reporter, the ACG start site was mutated to ATG by site-directed mutagenesis, as described above, using plasmid pC5559 and primers uCC_ACG_ATG_F and uCC_ACG_ATG_R. Mutagenesis was confirmed by sequencing

using M13 universal primers M13F(-40) and M13R(-40). Because *A. thaliana* contains an internal out-of-frame non-conserved uORF, site-directed mutagenesis was used to change the internal start site from ATG to AAG using primers uCCint_ATG_AAG_F and uCCint_ATG_AAG_R (Fig. 2.1B). Mutagenesis was confirmed by sequencing using M13 universal primers M13F(-40) and M13R(-40) and the plasmid was named pC5566. To generate the ACG uCC reporter, the internal start site was changed from ATG to AAG for the pC5559 plasmid, using the primer pair described above for plasmid pC5566 resulting in plasmid pC5570. To generate a uORF reporter without a start codon, the adjacent ATC and ACG codons were changed to TTC and TCG, respectively, using site directed mutagenesis as described above. Plasmid pC5559 was used as a template with primers uCC_ATC_TTC_ACG_TCG_F and uCC_ATC_TTC_ACG_TCG_R for plasmid pC5574. Mutagenesis was confirmed by sequencing using M13 primers and the plasmid was used as a template for mutagenesis of the internal start site as described for the other reporters.

The AUG uCC OF reporter was generated by fusion PCR using the existing plasmid pC5867 as a template. The first fragment was generated using plasmid pC5867 as a template and Frag1_F and FSuCC_R2 to add a T nucleotide and restore the uCC stop codon frame (Fig. 2.1C). The reporter ACG uCC OF, pC5954, was generated using fusion PCR with plasmid pC5570 as the template and to generate NS uCC OF, plasmid pC5955, plasmid pC5574 was used as the template. The second fragment contained the 5'UTR of the *GGP* uCC and was amplified by PCR using plasmid pC5570 as a template and primers FSuCC_F2 and uCCamp_R. The 5'UTR fragment had 23 bases of homology to the uCC fragment and they were mixed 1:1 and amplified by PCR with primers Frag1_F and

uCCamp_R. The product was cloned into plasmid pC5570 that was digested at KpnI and Sall using HiFi Assembly. Correct assembly was confirmed by sequencing with uCCseq_F and uCCseq_R primers.

Table 2.3. *In vitro* translation uORF reporters generated for study

<u><i>In vitro</i> translation uORF reporters</u>	<u>Stock ID</u>	<u>Source</u>
AUG uCC	pC5566	This study
ACG uCC	pC5570	This study
NS uCC	pC5574	This study
AUG uCC OF	pC5953	This study
ACG uCC OF	pC5954	This study
NS uCC OF	pC5955	This study

Table 2.3. The *in vitro* translation uORF reporters were used as templates to generate *in vivo* reporters

2.1.6. Generation of the *in vitro* translation FLuc controls (Table 2.4)

In order to test start site fidelity in the *in vitro* translation system, FLuc reporters were generated with a *GGP* 5' leader containing stop codons in each of the three reading frames just upstream of the FLuc start site (Fig. 2.3). The pC5568 reporter was digested with NheI and Sall and gel purified to remove the uCC and uCC leader sequence. PCR was used to generate DNA fragments where the FLuc start codon was varied: ATG, ACG, ATT, or AAA. The Frag1_F primer was used as the forward primer for all four reactions while the reverse primer was varied to change the start site. For an AUG start

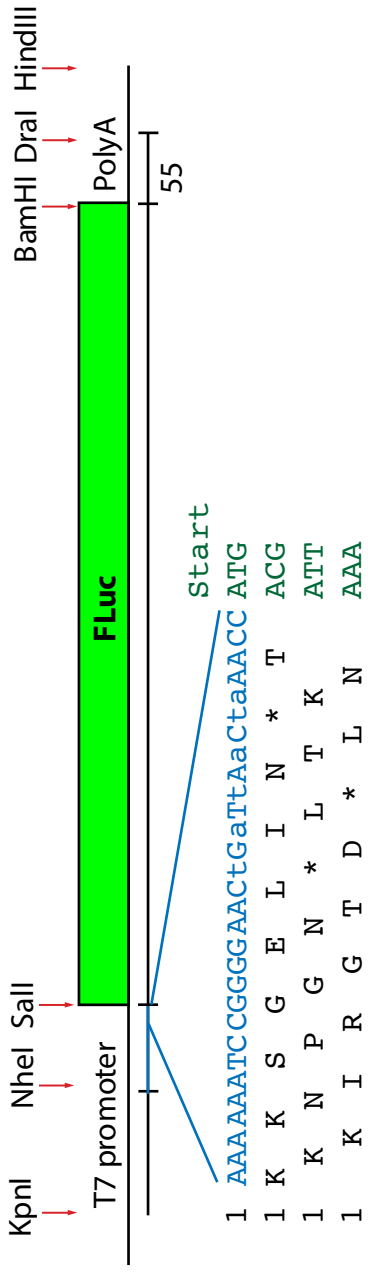


Figure 2.3. Templates used for *in vitro* transcription of FLuc mRNAs used to optimize RRL assays

Figure 2.3. Templates used for *in vitro* transcription of FLuc mRNAs used to optimize RRL assays

Plasmids were generated as templates for *in vitro* transcription of FLuc. The reporters contained a T7 promoter upstream of a 200-nt segment of the *GFP* leader located immediately upstream of the uCC start codon (blue). The leader sequence was altered (lowercase nts) to introduce stop codons in all three reading frames (black). To optimize initiation fidelity, three alternate start codons were substituted for ATG (green).

site ACCAUG_R2 was used, for an ACG start ACCACG_R2 was used, for an AUU start ACCAUU_R2 was used, and for a AAA start ACCAAA_R2 was used. The PCR products were cloned into plasmid pC5568 using HiFi Assembly as described above. Correct insertion was verified by sequencing using uCCseq_F and uCCseq_R primers.

Table 2.4. The *in vitro* translation FLuc controls used for RRL optimization

<u>FLuc controls</u>	<u>Stock ID</u>	<u>Source</u>
3 stops ACC ATG FLuc	pC5851	This study
3 stops ACC ACG FLuc	pC5852	This study
3 stops ACC ATT FLuc	pC5853	This study
3 stops ACC AAA FLuc	pC5854	This study

Table 2.4. The *in vitro* translation FLuc controls (Fig. 2.3).

2.1.7. Generation of the uCC-regulated *Renilla* reporters for mammalian transfection experiments (Table 2.5)

To study whether the *GGP* uCC will regulate a downstream mORF, the phRL plasmid (Promega) was used as a vector (Fig. 2.4A). The phRL vector contains the CMV immediate early enhancer/promoter followed by a T7 promoter for *in vitro* transcription. A *NheI* restriction site is between the T7 promoter and the RLuc start codon and an *AvaI* site is 9 codons into RLuc. The simian virus 40 (SV40) late polyadenylation signal is downstream of RLuc.

To generate the phRL based plasmids for mammalian transfection, fragments were generated by PCR using plasmid pC5566 as a template and primer pairs that encoded the uCC with the appropriate start codon (Fig. 2.1B). For an ACG start codon, primers *NheI*StopACGuCC_F and UTRren*AvaI*_R were used to generate a DNA fragment with *NheI* and *AvaI* restriction sites. phRL was digested with *NheI* and *AvaI* and gel purified. The DNA fragment was digested with *NheI* and *AvaI* and heat-inactivated at 80°C. The DNA fragment was cloned into phRL using Quick Ligase, as described above. Correct insertion was confirmed by sequencing using primers occ1098 and phRLuCC_R. The plasmid was named pC5914. To generate a uCC reporter with an AUG start codon, primers *NheI*StopATGuCC_F and UTRren*AvaI*_R were used to generate the DNA fragment for insertion by ligation, as described above, and the plasmid was named pC5915. To generate a reporter with no start codon, primers *NheI*StopNoStartuCC_F and UTRren*AvaI*_R were used to generate the DNA fragment that was cloned into phRL as described above. The reporter was named pC5916.

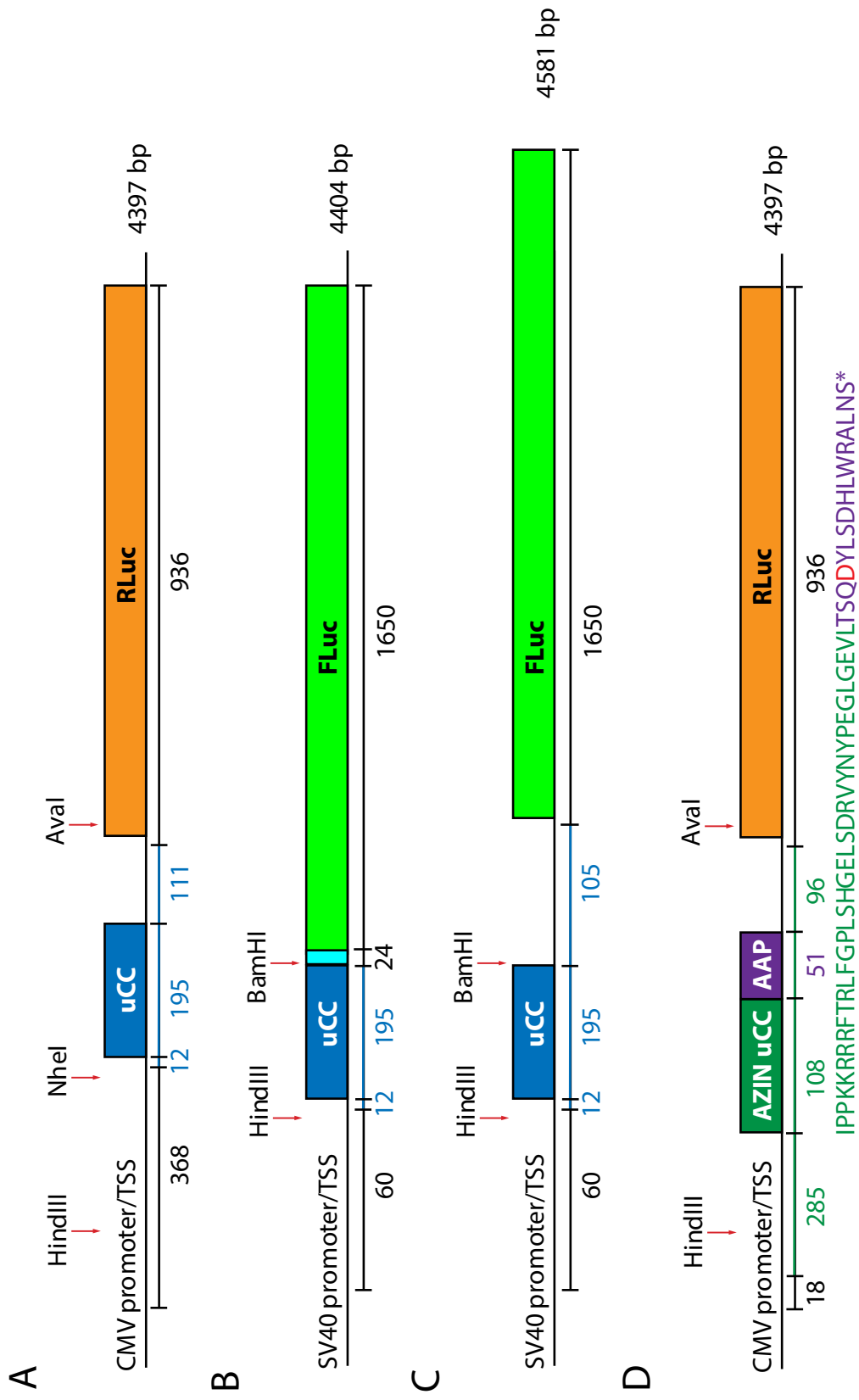


Figure 2.4. Plasmid reporters used in mammalian cell transfection

Figure 2.4. Plasmid reporters used in mammalian cell assays

Luciferase reporters were generated to observe ascorbate specific effects on *GGP* uCC translation. (A) To observe regulation of *GGP* expression in mammalian cells, a phRL based vector was generated with the CMV promoter/transcription start site (TSS) where *GGP* was replaced with RLuc. The reporter contained 12-nts of the *A. thaliana* *GGP* uCC leader sequence, the 195-nt uCC, and the 111-nt region between the uCC stop codon and the *GGP* start site (*GGP* specific regions depicted in blue). Restriction sites used for cloning are indicated with red arrows. (B) To observe translation of the *GGP* uCC in mammalian cells, a p2luc uCC-FLuc fusion was generated with the SV40 promoter/transcription start site. The reporter contained 12-nts of the *A. thaliana* *GGP* uCC leader sequence, the 195-nt uCC, a 24-bp linker region with restriction sites, and FLuc fused in-frame. (C) To reduce ribosome loading, a stem-loop (SL) was inserted upstream of the uCC in a p2luc based reporter with the SV40 promoter/transcription start site. The reporter contained 12-nts of the *A. thaliana* *GGP* uCC leader sequence, the 195-nt uCC, the 105-nt region between the uCC stop codon and the *GGP* start site, and FLuc in place of *GGP*. (D) In order to test specificity to ascorbate and Arg, *AZIN*-AAP fusion reporters previously generated by Ivanov were used. The reporters were generated from phRL with the CMV promoter/transcription start site. The reporters contain 285-nt of the mouse *AZIN1* uCC leader sequence upstream of 108-nt of the *AZIN1* uCC fused to 51-nt of the *N. crassa* *arg-2* AAP, followed by 96-nt of the *AZIN1* 5'UTR between the *AZIN1* uCC stop codon and the *AZIN1* start codon. The *AZIN1* uCC- *arg-2* AAP fusion has the *AZIN1* uCC start site (AUU) but the polyamine sensing region is replaced by the Arg sensing *arg-2* AAP. Mouse

AZINI specific regions are depicted in green and *N. crassa arg-2* specific regions are depicted in purple. *N. crassa arg-2* AAP D12 (shown in red) is essential for Arg sensing and a D12N mutation abolishes regulation of AAP translation in the presence of excess Arg.

To generate the phRL reporters containing the OF uCC control (Fig. 2.1C), two fragments were used to generate a DNA insert. The first fragment contained the OF uCC with either an ACG, AUG or AAA start codon, and the second contained the *GGP* 5'UTR between the uCC and the *GGP* start codon. The first fragments were generated by PCR, as described above, using plasmid pC5867 as the template. For the fragment with an ACG start site, primers NheIStopACGuCC_F and FSuCC_R2 were used. For the fragment with an AUG start site, primers NheIStopATGuCC_F and FSuCC_R2 were used. The 5'UTR fragment was generated by PCR using plasmid pC5570 as a template and FSuCC_F2 and phRL_hifi_AvaI_R. After generating the separate fragments, they were fused by PCR after combining 1:1 ratios of the first fragment with the 5'UTR fragment and using the forward primer from fragment one with primer phRL_hifi_AvaI_R. The fused PCR product was digested with NheI and AvaI and cloned into phRL as described above.

Table 2.5. The RLuc reporters used in *in vivo* assays to study uCC regulation

<u>GGP uCC reporters</u>	<u>Stock ID</u>	<u>Source</u>
phRL	pC5906	Ivanov et al., 2008.
ACG uCC	pC5914	This study
AUG uCC	pC5915	This study
AAA uCC	pC5916	This study
ACG OF	pC5956	This study
AUG OF	pC5957	This study

2.1.8. Generation of the uCC-FLuc fusions for mammalian transfection experiments (Table 2.6 and Table 2.7)

The p2luc backbone was used for construction of mammalian GGP-uCC FLuc fusion reporters (Fig. 2.4B) (Grentzmann et al., 1998). The p2luc plasmid contains the SV40 early enhancer/promoter upstream of a HindIII restriction site, *rluc*, and a polylinker with the following sites: Sall, BamHI, and a SacI site. Following the polylinker SacI site, the FLuc coding sequence lacks its own start codon. The SV40 late polyadenylation signal is downstream of FLuc.

p2Luc was digested with *HindIII* and *BamHI* to and then gel purified to remove *rluc*. To generate DNA fragments for HiFi Assembly into p2luc, fragments encoding the GGP uCC plus an extra 12 nucleotides of the GGP 5'UTR upstream of the uCC start codon were generated by PCR (Fig. B). Because they were designed for HiFi Assembly, appropriate homology to p2luc upstream of *HindIII* and downstream of *BamHI* was added

using the primer sets. The existing *in vitro* transcription plasmids were used as templates for PCR for each uCC except the S11A substitution (Table 2.2). To generate uCC fragments with an AUG start site primers p2lucATG_F and p2lucBamHI_uCC_R were used. To generate uCC fragments with an ACG start, primer p2lucACG_F and p2lucBamHI_uCC_R were used. For the OF uCC reporters (Fig. 2.1C), primer p2lucBamHI_FS_R was used to add a T nucleotide and restore the reading frame 5' of FLuc. After PCR amplification the fragments and digested p2luc were assembled using HiFi assembly as described above. Correct incorporation of the DNA fragments was confirmed by sequencing using primers occ1096 and uccSeq_R.

To generate the reporters with the S11A substitution, a 300-bp DNA fragment encoding the uCC with the Ser11 codon (AGC) mutated to Ala (GCC) was synthesized (G-block #11). The synthesized block was directly assembled into *HindIII* and *BamHI* digested p2luc to generate plasmid pC5980. To generate the AUG initiated S11A reporter pC6045, the G-block fragment was used as a template for PCR using the primers p2lucATG_F and p2lucBamHI_uCC_R. The fragment was cloned into p2luc as described above.

Table 2.6. The p2Luc based uCC-FLuc fusions used in *in vivo* assays

<u>uCC-FLuc fusions</u>	<u>Stock ID</u>	<u>Source</u>
p2luc	pC5150	Grentzmann et al., 1998
ACG uCC	pC5946	This study
AUG uCC	pC5947	This study
ACG OF	pC5949	This study
AUG OF	pC5950	This study

Table 2.7. The p2Luc based uCC substitution-FLuc fusions used in *in vivo* assays

<u>uCC-FLuc fusion substitutions</u>	<u>Stock ID</u>	<u>Source</u>
GGG/3A AUG	pC5958	This study
PSD/3A AUG	pC5960	This study
GG/2A AUG	pC5961	This study
PS/2A AUG	pC5963	This study
GGRG/4A AUG	pC5964	This study
R34A AUG	pC5965	This study
9A AUG	pC5967	This study
S11A AUG	pC6045	This study
GGG/3A ACG	pC5969	This study
PSD/3A ACG	pC5971	This study
GG/2A ACG	pC5972	This study
PS/2A ACG	pC5974	This study
GGRG/4A ACG	pC5975	This study
R34A ACG	pC5976	This study
9A ACG	pC5978	This study
S11A ACG	pC5980	This study

2.1.9. Generation of the phRL based reporters containing alanine substitutions within the uCC (Table 2.8)

To generate the phRL plasmids containing alanine substitutions within the uCC, two fragments were used to generate a DNA insert. The first fragment contained the ACG initiated uCC with the alanine substitution based on the template used. The second fragment contained the *GGP* 5'UTR between the uCC and the *GGP* start codon. Three different fragments were generated for the uCC: for plasmid pC6119 template plasmid pC5958 was used with primers ACGSubsAmpHiFi_F and uCCSubs_R, for plasmid pC6121 template plasmid pC5964 was used with the same primer pair, and for plasmid pC6123 template plasmid pC6045 was used. The 5'UTR fragment was generated by PCR using plasmid pC5914 as a template and uCC5UTR_F phRL_hifi_AvaI_R primers. After generating the separate fragments, they were fused by PCR after combining 1:1 ratios of the first fragment with the 5'UTR fragment and using the forward primer ACGSubsAmpHiFi_F with primer phRL_hifi_AvaI_R. The fragments were used in HiFi assembly with *NheI* and *AvaI* digested phRL plasmid as described above. Correct assembly was verified by sequencing using primers occ1098 and phRLuCC_R.

Table 2.8. The phRL reporters containing uCC substitutions

<u>uCC substitutions</u>	<u>Stock ID</u>	<u>Source</u>
GGG/3A	pC6119	This study
GGRG/4A	pC6121	This study
S11A	pC6123	This study

Table 2.8. All uCCs initiate at the native ACG.

2.1.10. Generation of the alanine insertions within the uCC for the uCC-FLuc fusion reporters used in mammalian transfection experiments (Table 2.9)

To generate the reporters with the alanine insertions within the uCC, four G-blocks were synthesized (Blocks #12-15) (LifeSct). All four blocks were synthesized for insertion in p2luc between HindIII and BamHI using HiFi Assembly. The synthesized blocks had regions of homology to the p2luc vector upstream of HindIII and downstream of BamHI. All four blocks were synthesized with an ACG uCC start site. p2luc was digested with HindIII and BamHI and gel purified. To generate the reporters with an ACG start site Blocks #12-15 were amplified by PCR using p2lucACG_F and p2lucBamHI_uCC_R. The DNA fragments were cloned in digested p2luc by HiFi Assembly as described above. To change the start site to an AUG, Blocks #12-15 were amplified by PCR using p2lucATG_F and p2lucBamHI_uCC_R. The DNA fragments were cloned in digested p2luc by HiFi Assembly as described above. Correct assembly was confirmed by sequencing using occ109 and uCCseqR.

Table 2.9. The p2Luc based FLuc reporters containing alanine insertions within the uCC

<u>uCC-FLuc alanine insertions</u>	<u>Stock ID</u>	<u>Source</u>
+1A@P28 AUG	pC6198	This study
+1A@P28 ACG	pC6194	This study
+3A@P28 AUG	pC6199	This study
+3A@P28 ACG	pC6195	This study
+1A@P44 AUG	pC6200	This study
+1A@P44 ACG	pC6196	This study
+3A@P44 AUG	pC6201	This study
+3A@P44 ACG	pC6197	This study

2.1.11. Generation of the stem-loop uCC FLuc reporters for mammalian transfection experiments (Table 2.10)

To generate the p2Luc reporters containing the GGP leader sequence between the uCC and FLuc (Fig. 2.4C), DNA fragments containing uCC sequence followed by the sequence 5' of the GGP coding region were generated by PCR as described above. To generate a reporter with an ACG initiated uCC followed by the 5'UTR of *GGP*, plasmid pC6240, reporter pC5915 was used as a template for primers HindIIIled_ACG_uCC_F and BamHI_AUG_ledR. To generate plasmid pC6241, an ACG initiated OF uCC, primers HindIIIled_ACG_uCC_F and BamHI_AUG_ledR were used with plasmid pC5956 as the template. To generate the AAA initiated OF reporter pC6242, primer HindIII_led_NS_uCC_F and primer BamHI_AUG_ledR were used with plasmid pC5956

as the template. The p2Luc backbone was digested with HindIII and BamHI and gel purified, as described above. The PCR products were also digested with HindIII and BamHI but purified using the High Pure PCR Product Purification Kit (Roche) as described above. The vectors and inserts were ligated together using Quick Ligase (NEB) as described previously. The colonies were screened by PCR using occ1096 and uCCseqR and the LiTaq 2X PCR mix (LifeSct) and positive colonies were further screened by confirmation by sequencing using primers occ1096 and uccSeq_R.

To generate the reporters containing a stem-loop structure 17-bases from the start of transcription and 29-bases upstream of the uCC start site, two primers containing an 89 base stem loop with a $\Delta G = -41.3$ kcal/mol were designed to generate HindIII compatible sticky ends. To anneal the primers into a double stranded insert, the primers were diluted to 100 μ M in duplex buffer (100 mM potassium acetate, 30 mM HEPES pH 7.5). Equal volumes of each primer were added to a microcentrifuge tube, and the primers were heated to 94°C for 2 minutes. The block was turned off and the solution was allowed to slowly cool to RT. The annealed primers were stored at -20°C.

The reporters, pC6240, pC6241, and pC6242 were digested with HindIII and rSAP and then gel purified. The annealed primers fragment was ligated into the HindIII digested pC6240, pC6241, and pC6242 vectors using Quick Ligase as described above. Colonies were screened by PCR using occ1096 and trunc3_R and the LiTaq 2X PCR mix (LifeSct). Positive colonies further screened by sequencing using occ1096 or SLseqF and uCCseqR or SLseqR.

Table 2.10. The p2Luc based FLuc reporters containing the GGP uCC regulating FLuc and a stem-loop to reduce ribosome loading

<u>uCC FLuc reporters</u>	<u>Stock ID</u>	<u>Source</u>
ACG uCC	pC6240	This study
ACG OF	pC6241	This study
AAA OF	pC6242	This study
SL-ACG uCC	pC6243	This study
SL-ACG OF	pC6244	This study
SL-AAA OF	pC6245	This study

2.1.12. Generation of the phRL based stem-loop uCC RLuc reporters for mammalian transfection experiments (Table 2.11)

To generate the reporters containing a stem-loop structure 49-bases from the start of transcription and 325-bases upstream of the uCC start site, two primers containing an 89 base stem loop with a $\Delta G = -41.3$ kcal/mol were designed to generate HindIII compatible sticky ends. The primers were annealed as described above for the p2luc plasmid. For assembly, plasmids pC5914 and pC5956 were digested with HindIII and rSAP and then gel purified. The annealed primers fragment was ligated into the HindIII digested pC5914 and pC5956 vectors using Quick Ligase as described above. Colonies were screened by PCR using occ1096 and trunc3_R and the LiTaq 2X PCR mix (LifeSct). Positive colonies further screened by sequencing using occ1096 or SLseqF and uCCseqR or SLseqR.

Table 2.11. The phRL based RLuc reporters containing the GGP uCC regulating RLuc and a stem-loop to reduce ribosome loading

<u>uCC RLuc reporters with SL</u>	<u>Stock ID</u>	<u>Source</u>
SL- ACG uCC	pC6215	This study
SL- ACG OF	pC6216	This study

2.1.13. Generation of the phRL based Trip12-extended uCC reporters for mammalian transfection experiments (Table 2.12)

To generate a 65-codon extension of the uCC, two fragments were generated for fusion by PCR. The first fragment was amplified using plasmid pC5802 (Ivanov et al., 2018) as the Trip12 template. Three different combinations of primers were used to generate the Trip12 fragments with the appropriate combination of start sites for Trip12 and for the uCC. To generate the Trip12 fragment with no start site for Trip12 and no start site for the uCC (AAA AAA, AAA AAA) primer Trip12AAAAAA_F and primer modAAAAAAuCC_R were used. To generate the Trip12 fragment with no start site for Trip12 but an ACG start site for the uCC (AAA AAA, ATC ACG) primer Trip12AAAAAA_F and primer modATCACGuCC_R were used. To generate the extended uCC fragment with the Trip12 ACG start site and no start site in the uCC (ATC ACG, AAA AAA), primer Trip12ATCACG_F and primer modAAAAAAuCC_R were used.

For the second fragment, containing the majority of the uCC, two different fragments were generated using plasmid pC5914 as the template. To keep the ACG start site intact, a fragment was amplified using primers modATCACGuCC_F and

phRL_hifi_AvaI_R. Primers modAAAAAAuCC_F and phRL_hifi_AvaI_R were used to remove the start site.

To generate the Trip12-uCC extensions combinations of fragment 1 and fragment 2 were combined 1:1 with the appropriate primers and fused by PCR. To generate the uCC extension without start sites, the fragments amplified with Trip12AAAAAA_F and modAAAAAAuCC_R for the Trip12 extension, and modAAAAAAuCC_F and phRL_hifi_AvaI_R for the uCC were combined and fused using primers HF_NheI_Trip12_F and phRL_hifi_AvaI_R. To generate the a Trip12-uCC fusion where Trip12 lacked a start site and the uCC start site was left intact, the fragments amplified with Trip12AAAAAA_F and modATCACGuCC_R for the extension, and modATCACGuCC_F and phRL_hifi_AvaI_R for the uCC were combined and fused by PCR using primers HF_NheI_Trip12_F and phRL_hifi_AvaI_R. To generate the Trip12 extended uCC, the fragments amplified with Trip12ATCACG_F and modAAAAAAuCC_R for the extension, and modAAAAAAuCC_F and phRL_hifi_AvaI_R for the uCC were combined and fused by PCR using primers HF_NheI_Trip12_F and phRL_hifi_AvaI_R.

After generating the three inserts, phRL was digested with NheI and AvaI and gel purified as described above. The three inserts were cloned into the digested vector using HiFi Assembly as described above. Correct insertion was confirmed by sequencing with primers occ1098 and phRLuCC_R.

Table 2.12. The phRL based Trip12-uCC extensions used in *in vivo* assays

<u>Mammalian phRL Trip12-uCC extensions</u>	<u>Stock ID</u>	<u>Source</u>
AAA AAA	pC5982	This study
AAA ACG	pC5983	This study
ACG AAA	pC5984	This study

2.1.14. Generation of the Trip12-extended uCC FLuc fusion reporter for mammalian transfection experiments (Table 2.13)

To generate the Trip12-uCC-FLuc fusion, p2luc was digested with HindIII and BamHI as described above. Plasmid pC5982 was used as a template for PCR with primers HindIIIp2lucTrip12_F and p2lucBamHI-uCC_R to generate a 441 bp product with a HindIII restriction site, an AUG start site at Trip12, no start site at the beginning of the uCC, and a direct fusion to FLuc at the BamHI site. The fragment was digested with HindIII and BamHI as described above and cloned into p2luc using Quick Ligase. Correct insertion into p2luc was confirmed by sequencing using occ1096 and uCCseqR.

Table 2.13. The p2Luc-based Trip12-uCC-FLuc fusion control

<u>Mammalian p2Luc Trip12-uCC FLuc fusion</u>	<u>Stock ID</u>	<u>Source</u>
Trip12-uCC	pC6202	This study

Table 2.13. The fusion initiates at an AUG codon.

2.1.15. Other reporters for mammalian transfection experiments

Table 2.14. The phRL based AZIN-AAP controls used in *in vivo* assays (Fig. 2.4D)

<u>Previously published phRL reporters</u>	<u>Stock ID</u>	<u>Source</u>
AAP	IV2475	Ivanov, 2018
AAP D12N	IV2478	Ivanov, 2018

2.1.16. Peptidyl-tRNA release assay templates (Table 2.15)

The templates for the peptidyl-tRNA release assay were generated as G-blocks #16, #17, and #18. Each block contains the T7 promoter for *in vitro* transcription and a transcription unit encoding a FLAG-epitope tag fused in-frame with a 390-bp linker fragment encoding residues 21-81 from the human chorionic gonadotropin beta subunit with other amino acids, followed by the *GGP* uCC or hCMV uORF2 gp48/UL4 peptide coding region (Fig. 2.2B,D). G-block #16 encodes the FLAG-linker-GGP uCC peptide, G-block #17 encodes the FLAG-linker-GGP [OF] peptide, and G-block #18 encodes the FLAG-linker-gp48/UL4 control peptide.

Plasmids pC6301 and pC6302 were made by cloning into pUC19. G-block #16 was resuspended in water to a concentration of 5 ng/ μ l. For PCR, the resuspended G-block was diluted 1:100 in water and amplified using primers EcoRI_uCC_F and HindIII_pramod_R as described above. G-block #18 was diluted as described above but amplified with primers EcoRI_uCC_F and gp48_R. The G-block #16 fragment was digested with EcoRI and HindIII, as described above, and the G-block #18 fragment was digested with EcoRI and BamHI. pUC19 was digested with EcoRI and BamHI with rSAP, or EcoRI and HindIII

with rSAP, and gel purified as described above. The fragments were cloned into pUC19 using Quick Ligase as described above.

To generate the FLAG tagged uCC FLuc fusion plasmid pC6217 (Fig. 2.2C), G-block #15 was amplified by PCR with primers Pramod_FLUC_F and uCC_SalI_FLUC_R as described above. To generate the FLAG tagged OF FLuc fusion pC6218, G-block #16 was amplified with primers Pramod_FLUC_F and OF_SalI_FLUC_R. Plasmid pC5568 was digested with NheI and SalI and gel purified. The fragments were cloned into the purified pC5568 vector using HiFi DNA Assembly as described above. Correct insertion was confirmed by sequencing using uCCseq_F and uCCseq_R.

Table 2.15. The FLAG tagged peptidyl-tRNA stalling assay templates

<u><i>In vitro</i> translation stalling templates</u>	<u>Stock ID</u>	<u>Source</u>
uCC	pC6301	This study
OF	None	Not in vector-block #17
uCC-FLuc	pC6217	This study
OF-FLuc	pC6218	This study
UL4/gp48	pC6302	This study

Table 2.15. All of the fusions contain FLAG-linker fused directly upstream of the indicated region.

2.2. Mammalian cell transfection procedures

Human osteosarcoma U2OS (Nancy Kedersha, Harvard University) cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g/L glucose and L-glutamine without sodium pyruvate (Corning) supplemented with 10 % fetal bovine serum (FBS) (Gibco), 1 mM sodium pyruvate (Corning), and penicillin/streptomycin (Quality Biological). Cells were maintained in 100 mm plates at 37°C and 5% CO₂ and split at least once per week using 0.25% trypsin- ethylenediaminetetraacetic acid (EDTA) (Gibco).

L-ascorbate for cell culture was prepared by dissolving L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (Sigma-Aldrich) in supplemented DMEM to a concentration of 200 mM. The stock solution was filter sterilized and stored at 4°C.

Plasmids for transfection were prepared using the QIAprep Spin Miniprep Kit (Qiagen). For a miniprep, 3 mls of *E. coli* containing the plasmid of interest were grown overnight in LB-Amp (100 mg/ml). The cells were pelleted in a 1.5 ml tube by centrifugation at 6800 xg for 3 minutes at RT. The pellet was resuspended in 250 µl Buffer P1 and then 250 µl of Buffer P2 was added and the solution was mixed by inverting 4-6 times. After mixing, 350 µl Buffer N3 was added and the solution was mixed by inverting 4-6 times. The sample was spun for 10 minutes at 17,900 xg. After clarifying the cell lysate, 800 µl of supernatant was added to the QIAprep 2.0 spin column. The column was spun for 30 s and the flow-through was discarded. The column was washed with 0.5 ml Buffer PB and centrifuged for 30 s and the flow-through was discarded. The column was washed a second time with 0.75 ml Buffer PE and centrifuged for 30 s and the flow-through was discarded. The column was transferred to a new collection tube and spun for 1 minute to remove residual buffer. The column was then placed in a clean 1.5 ml tube and 50 µl of

Buffer EB was added to the column. The column was incubated for 1 minute and then spun for 1 minute to elute the DNA. The plasmid concentration was determined using a NanoDrop ND-1000 Spectrophotometer and diluted to 10 ng/ μ l before use in transfection experiments.

To introduce the reporter plasmids into the U2OS cells, reverse transfection was used. For transfections utilizing FLuc (p2Luc plasmids), each transfection contained 5 ng of reporter plasmid and 1 ng of RLuc plasmid as a transfection control. For transfections utilizing the RLuc (phRL plasmids), each transfection contained 2 ng of reporter plasmid and 5 ng of FLuc plasmid as a transfection control. For each transfection, a master mix was made containing 0.2 μ l of Lipofectamine 2000 Reagent (Invitrogen) per well and 25 μ l per well of Opti-MEM I Reduced-Serum Medium (Gibco). The master mix was mixed gently by inversion and incubated for five minutes while the reporters were aliquoted into individual 1.5 ml microcentrifuge tubes. After five minutes, the control plasmid was added to the master mix and gently mixed. The master mix was then added to the aliquoted reporters and gently mixed. Before the transfection, 25 μ l of the reporter transfection mix was added to each well of a 96-well half area white tissue culture treated plate (Corning).

After transferring the transfection reagent, the cells were prepared. To achieve a final concentration of 50 mM L-ascorbate in a final volume of 50 μ l per well, 12.5 μ l of 200 mM L-ascorbate was added for each well to a 1.5 ml tube. As a control, 1.5 ml tubes containing the same volume of DMEM were used. 70-80% confluent cells were removed from the plate using 0.25% trypsin-EDTA (Gibco) and resuspended in supplemented DMEM. The cells were counted, and the volume adjusted to achieve a cell/ μ l count of 800 cells/ μ l so 12.5 μ l would contain approximately 10,000 cells. The cells were added to either

the 200 mM L-ascorbate in the 1.5 ml tube or the DMEM control at a volume of 12.5 μ l per well. This ensured that 10,000 cells were seeded in each well. The cell master mix was gently mixed by inverting and 25 μ l of the mix was added to each well achieving a cell count of 10,000 per well in 50 mM L-ascorbate or DMEM. The cells were incubated for 20-22 hours.

2.3. Dual luciferase assay

Buffers used for assays of FLuc (Table 2.16) and RLuc (Table 2.17) activity were made in the lab. All stock solutions were prepared by dissolving the indicated reagent in water except for coelenterazine, used for the RLuc assay buffer, which was dissolved in methanol.

After incubation, the Centro XS³ LB 960 luminometer (Berthold Technologies) was prepared by soaking the buffer pump lines in 80% ethanol for 10-15 minutes. After soaking, the lines were washed for 40 wash cycles 80% ethanol and then for 40 wash cycles with deionized (DI) water. After washing, the pumps were primed with the assay buffers and the appropriate assay and wells were selected. While the luminometer lines were soaking, 1X passive lysis buffer was prepared (Promega). The tissue culture plates were removed from the incubator one at a time and the medium was removed using a multichannel pipet and 25 µl of 1X passive lysis buffer was added to each well. The plate was placed on a shaker for 10 minutes. The plates were analyzed on the luminometer using the dual luciferase settings: firefly dispense 50 µl middle speed, delay duration 2.0, firefly counting time 10.00, Renilla dispense 50 µl middle speed, delay duration 2.0, Renilla counting time 10.00.

Table 2.16. Components of the FLuc Assay Buffer

FLuc Assay Buffer			
Reagent	Stock Concentration	Final Concentration	Volume (500 mls)
Glycylglycine	250 mM	25 mM	50 ml
K ₂ PO ₄ pH 8.0	1 M	15 mM	7.5 ml
EGTA	200 mM	4 mM	10 ml
MgSO ₄	500 mM	15 mM	15 ml
Coenzyme A (CoA)	10 mM	0.1 mM	5 ml
Luciferin	7.5 mM	75 μM	5 ml
Dithiothreitol (DTT)	500 mM	1 mM	1 ml
ATP	200 mM	2 mM	5 ml
Commercial Firefly reagent (Promega)		5 %	25 ml
DI water (add after adjusting pH of assay buffer to 8.0 using NaOH)			

Table 2.17. Components of the RLuc Assay Buffer

RLuc Assay Buffer			
Reagent	Stock Concentration	Final Concentration	Volume (500 mls)
NaCl	2.5 M	1.1 M	220 ml
EDTA	500 mM	2.2 mM	2200 μ l
K ₂ PO ₄ pH 5.1	1 M	220 mM	111.1 ml
Bovine serum albumin (BSA)	100 mg/ml	0.44 mg/ml	2200 μ l
NaN ₃	1 M	1.3 mM	650 μ l
Coelentrerazine	1.43 mM	1.43 μ M	500 μ l
Commercial Renilla reagent (Promega)		5 %	25 ml
DI water (add after adjusting pH of assay buffer to 5.0 using HCl)			

2.4. Detection of eIF2 α phosphorylation

To monitor eIF2 α phosphorylation, U2OS cells were transfected with plasmid pC5914 and 600,000 cells per well were plated in 6-well plates. Cells were left untreated (negative control), or 23 hours after transfection cells were treated with 50 mM L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (+ L-AA) or with 100 mM sodium arsenite (positive control) and then incubated for an additional 1 hour before harvesting.

To begin harvesting, the plates were placed on ice. The cells were washed one time with chilled phosphate buffered saline (PBS) and then harvested by adding 30 μ l of chilled radioimmunoprecipitation assay (RIPA) buffer (150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 25 mM Tris pH 7.4, DI water to volume) with Complete Protease Inhibitor (Roche). The cells were scraped off the plate and the resuspended cells were transferred to 1.5 ml microcentrifuge tubes on ice. The cells were sonicated 10 times with one second pulses and then spun at 17,949 xg for 10 minutes at 4°C. The cleared cell lysates were transferred to clean microcentrifuge tubes and Bradford assays were used to quantify the protein concentration for each sample. The samples were diluted to 10 and 20 μ g per 15 μ l in RIPA buffer and 4X NuPAGE lithium dodecyl sulfate (LDS) Sample Buffer (Invitrogen) was added along with 1 mM DTT to a final volume of 33 μ l. The samples were heated to 70°C for 10 minutes and then 15 μ l was loaded into each well of a 4-20% Criterion TGX Precast Gel (BIO-RAD).

After running the gel (200 V for 1 hour), the proteins were transferred using the BIO-RAD TGX system (50 V for 30 minutes at 4°C) to a nitrocellulose membrane. The membrane was stained with Ponceau-S stain to visualize proteins and then trimmed. The Ponceau S stain was removed by washing the membrane with water and TBS-T (Tris-

buffered saline, 0.1% Tween 20) and the membrane was then incubated with blocking solution (5% non-fat milk in TBS-T) for one hour on a nutator at RT. The membrane was probed overnight with rabbit anti-eIF2 α -(Phospho-Ser51) (Abcam ab32157) primary antibody (diluted 1:1,000 in blocking solution) at 4°C on a nutator. The membrane was then washed three times for five minutes with X-ml TBS-T and then incubated for 1 hour with a 1:10,000 dilution of enhanced chemiluminescence (ECL) Anti-Rabbit IgG HRP conjugated antibody (GE Healthcare) in blocking solution on a nutator at RT. After probing with this secondary antibody, the membrane was washed six times for five minutes with TBS-T. Immunoconjugates were detected by incubating the membrane for 5 minutes in Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare). The membrane was then exposed to film, which was developed using an SRX-101A medical film processor (Konica Minolta).

After developing the Western blot, the membrane was stripped using the harsh stripping method. The stripping buffer (For 100 ml: 20 ml 10% SDS, 12.5 ml Tris-HCl pH 6.8 0.5 M, 67.5 ml DI water, 0.8 ml 2-mercaptoethanol (BME)) for 30 minutes at 50 °C with agitation. After stripping the membrane was rinsed with tap water for 1 minute, then washed 5 times for 1 minute in TBS-T. The membrane was re-blocked in blocking solution, as above, for one hour and then probed overnight with a 1:1,000 dilution of rabbit anti-eIF2 α total (Bethyl) primary antibody in 5% non-fat milk. The membrane was then washed with TBS-T three times for five minutes and then incubated for 1 hour with 1:10,000 dilution of ECL Anti-Rabbit IgG HRP conjugated antibody (GE Healthcare). After probing with the secondary antibody, the membrane was washed six times for five minutes each with TBS-T and then incubated for 5 minutes with Amersham ECL Prime Western Blotting

Detection Reagent (GE Healthcare). The membrane was then exposed to film, which was developed as described above.

2.5. qPCR quantification of luciferase mRNA levels

To validate the luciferase assays, RLuc and FLuc mRNA levels were quantified using quantitative reverse transcription polymerase chain reaction (RT-qPCR). U2OS cells were transfected as described but volumes were increased 40X for 6-well plates. Transfected cells were grown for approximately 24 hours and then the RNA was extracted. The 6-well plates were placed on ice, the growth medium was removed, and the wells were washed with 2 mls of PBS. The cells were resuspended in 1 ml of Trizol (ThermoFisher) using a cell scraper and pipetting. The resuspended cells were stored at -80 °C or processed immediately. For processing, the sample was incubated for 5 minutes at RT then 0.2 mL of chloroform was added to the sample and it was mixed by hand for 15 seconds. The sample was incubated for 2 minutes at RT and then the sample was spun at 12,000 xg for 15 minutes at 4 °C. The upper aqueous phase was removed and transferred to a new microcentrifuge tube and 1.5 µl of glycoblue (ThermoFisher) was added along with 0.5 mL of 100% isopropanol. The sample was incubated at RT for 10 minutes and then the RNA was pelleted by centrifugation at 12,000 xg for 10 minutes at 4 °C. The supernatant was removed, and the pellet was washed with 1 mL of 75% ethanol by vortexing. The pellet was preserved by centrifugation at 7500 xg for 5 minutes at 4 °C. The supernatant was removed, and the pellet was dried for 5-10 minutes. The pellet was resuspended in 44 µl of water and the RNA concentration was quantified using a nanodrop.

To remove plasmid contamination up to 10 µg of nucleotide was treated with TURBO DNase in a 50 µl reaction. For each reaction, the extracted RNA was added at a maximum volume of 43.5 µl or 10 µg with water added to adjust the volume to 43.5 µl, 5 µl of 10X TURBO DNase Buffer was added and then 2.5 µl of TURBO DNase was added

and the sample was gently mixed. The sample was incubated at 37 °C for 30 minutes and then 10 µl of resuspended DNase Inactivation Reagent was added. The sample was incubated at RT for 5 minutes with occasional mixing. The DNase was removed by centrifugation at 10,000 xg for 1.5 min and the RNA was transferred to a fresh tube. RNA was used immediately for RT-qPCR or stored at -80 °C.

Multiplexed TaqMan one-step RT-qPCR was performed using TaqMan assays with HEX, FAM, and ROX as a reference dye. For each reaction, 5 µl of 4X TaqMan Fast Virus 1-Step Master Mix (ThermoFisher) was combined with 0.25 µl of ActB probe mix, 0.25 µl of either FLuc assay or RLuc assay, 4.5 µl of water, and 5 µl of purified RNA. The assays were run on a Stratagene Mx3000P (Agilent) using the fast cycling mode. Reverse transcription at 50 °C for 5 minutes followed by RT inactivation/denaturation at 95 °C for 20 seconds. This was followed by 40 cycles of denaturation at 95 °C for 3 seconds and annealing/extension at 60 °C for 30 seconds.

For the TaqMan assays PrimeTime Std qPCR Assays were designed and purchased from IDT. Each assay contained two primers (5.0 nmoles) and a probe (2.5 nmoles) and they were diluted in 500 µl IDTE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) to 20X.

FLuc assay set:

Primer 1: 5' -GTGTTGGGCGCGTTATTTATC-3'

Primer 2: 5' -TAGGCTGCGAAATGTTTCATACT-3'

Probe: /56-FAM/TTGCGCCCG/ZEN/CGAACGACATTTATA/3IABkFQ/

RLuc assay set:

Primer 1: 5' - CTGATCTGATCGGAATGGGTAAG -3'

Primer 2: 5' - CAAGCGGTGAGGTACTTGTAG -3'

Probe: /56-FAM/AAGAGCGGG/ZEN/AATGGCTCATATCGC/3IABkFQ/

N001101.1.pt.ACTB set:

Primer 1: 5'- ACCTTCTACAATGAGCTGCG-3'

Primer 2: 5'- CCTGGATAGCAACGTACATGG-3'

Probe: /5HEX/ATCTGGGTC/ZEN/ATCTTCTCGCGGTTG/3IABkFQ/

For analysis, ΔC_t was calculated as the ActB C_t minus the luciferase C_t . Outliers were identified and removed using 1.5 times the interquartile range (IQR) as the cut-off for each replicate. After removing the outliers, $\Delta\Delta C_t$ was calculated by calculating the average ΔC_t of the control (untreated 0 mM) for each reporter and subtracting it from the ΔC_t for each value. Fold gene expression was calculated for each well using $2^{-\Delta\Delta C_t}$.

2.6. *In vitro* transcription of mRNAs

To transcribe the mRNAs for *in vitro* translation assays, large amounts of concentrated plasmid were needed so a ZymoPure II Plasmid Maxiprep Kit (Zymo) was used. For each plasmid preparation, a 3 ml preculture of *E. coli* containing the plasmid of interest were grown overnight in LB + Amp (100 mg/ml) at 37°C with shaking at 250 rpm. One ml of the preculture was used to inoculate 100 ml of LB + Amp in a 500 ml flask and the culture was grown for 3 hours at 37°C with shaking at 250 rpm. The bacterial culture was transferred to a 500 ml tube and spun at 3,400 xg for 10 minutes to pellet the cells. The supernatant was discarded, and the pellet was resuspended in 14 ml of ZymoPure P1. After resuspending the pellet, 14 ml of ZymoPure P2 was added and mixed by inversion 6 times. After mixing, 14 ml of ZymoPure P3 was added and the solution was mixed by inversion. The ZymoPure Syringe Filter was placed in a tube rack and the Luer Lock was plugged. The cell lysate was added to the filter and allowed to sit for 5 minutes. The Luer Lock plug was removed from the syringe and the syringe was placed in a 50 ml conical tube. The plunger was placed into the syringe and used to push the solution through the filter until 33-35 ml of the cleared lysate was in the conical tube. Then 14 ml of ZymoPure Binding Buffer was added to the cleared lysate and mixed by inversion at least 10 times. The 50 ml reservoir was moved from the top of the Zymo-Spin V-P Column Assembly and the assembly was placed in a 50 ml conical tube, and then 14 ml of the binding buffer-lysate was added to the assembly. The column was spun at 500 xg for 2 minutes and the flow-through was discarded. This was repeated until all the lysate had passed through the column. After binding the plasmid to the filter, the filter was washed with 5 ml of ZymoPure Wash 1 and centrifuging the assembly at 500 xg for 2 minutes and the flow-

through was discarded. Then 5 ml of ZymoPure Wash 2 and centrifuging the assembly at 500 xg for 2 minutes and the flow-through was discarded. The wash step was repeated once and then 15 ml conical reservoir was discarded and the Zymo-Spin V-P column was placed in a collection tube and spun at 10,000 xg for 1 minute. The column was transferred to a clean 1.5 ml tube and 400 μ l of ZymoPure Elution Buffer was added. The column was incubated for 2 minutes and then spun at 10,000 xg for 1 minute. The plasmid concentration was determined using a NanoDrop ND-1000 Spectrophotometer.

For each transcription reaction, 20 μ g of purified plasmid was digested in a 50 μ l digestion with 5 μ l of DraI (NEB) overnight at 37°C. After digestion, mRNAs were transcribed in a 500 μ l reaction with T7 RNA polymerase overnight at 37°C (Table 2.18 and Table 2.19).

Table 2.18. Composition of the *in vitro* transcription reaction

<i>In vitro</i> Transcription Reaction			
Reagent	Stock Concentration	Final Concentration	Volume for 500 μ l
Ribomax buffer	5X	1X	80 μ l
DTT	1 M (5X)	1X	20 μ l
NTPs	25 mM each	5 mM each	100 μ l
Ribolock	40U/ μ l	0.4U/ μ l	5 μ l
T7 RNAP (His6)	100X	1X	5 μ l
YIPP (NEB)	100X	1X	5 μ l
Digested DNA	10X	1X	50 μ l
Water			235 μ l

Table 2.18. For the ribomax buffer see Table 2.19. The NTPs (ThermoFisher) were purchased as a four nucleotide kits at 100 mM each and mixed in equal parts to achieve a final concentration of 25 mM of each nucleotide. Ribolock RNase Inhibitor (ThermoFisher) works best in the reaction mix. The T7 RNA polymerase was purified in the lab. Yeast Pyrophosphatase, Inorganic (YIPP) was purchased from NEB. All reagents and plastics from the plasmid preparation on were RNase Free.

Table 2.19. Composition of the ribomax buffer

5X Ribomax Buffer			
Reagent	Stock Concentration	Final Concentration	Volume 5 ml
HEPES-KOH pH 7.4	1 M	400 mM	2000 μ l
MgCl ₂	1 M	120 mM	600 μ l
Spermidine	1 M	10 mM	50 μ l
DTT	1 M	200 mM	1000 μ l
Water			1350 μ l

Table 2.19. The 5X ribomax buffer was aliquoted into 500 μ l aliquots and stored at -20°C. All reagents and plastics from the plasmid preparation on were RNase Free.

After transcribing the mRNAs overnight, 133 μ l (the maximum volume) of the transcription product was purified with the RNA Clean & Concentrator Kit-25 (Zymo Research). After transferring 133 μ l of the transcription reaction to a clean 1.5 ml tube, 2

volumes of RNA Binding Buffer (267 μ l) were added and then an equal volume of 100% ethanol was added (400 μ l). The solution was mixed and transferred to the Zymo-Spin IIC Column in a collection tube and spun for 30 s at 14,000 xg. The flow-through was discarded and 400 μ l of RNA Prep Buffer was added to the column and spun for 30 s. The flow-through was discarded and 700 μ l of RNA Wash Buffer was added to the column and spun for 30 s. The flow-through was discarded and 400 μ l of RNA Wash Buffer was added to the column and the column was spun for 2 minutes. The column was transferred to a clean 1.5 ml tube and 80 μ l of RNase-Free Water was added to the column matrix and incubated for 1 minute before spinning for 30 s.

The mRNAs were capped using the Vaccinia Capping System (NEB) scaled up four times (58 μ l of purified mRNA in a total volume of 80 μ l). The mRNA was added to a clean 1.5 ml tube and heated at 65°C for 5 minutes then placed on ice for 5 minutes. After denaturing the mRNA, 2 μ l of RNasin (Promega) added followed by 8 μ l 10X Capping Buffer, 4 μ l GTP (10 mM), 4 μ l S-adenosyl methionine (SAM) (2 mM, diluted from 32 mM stock), and 4 μ l Vaccinia Capping Enzyme. The capping reaction was incubated at 37°C for 30 minutes. After capping, the entire capping reaction was purified again using the RNA Clean & Concentrator Kit-25 (Zymo Research), with the RNA Binding Buffer adjusted to the correct volume (160 μ l) were added and then an equal volume of 100% ethanol was added (240 μ l). The purified capped mRNAs were eluted in 50 μ l. The mRNAs were quantified using a Bioanalyzer and an RNA Pico Chip (Agilent Technologies).

2.7. *In vitro* translation of mRNAs

Purified mRNAs were translated using Promega's Flexi Rabbit Reticulocyte Lysate System (RRL) (Promega) (Table 2.20). The total concentration of mRNA added varied to keep translation in the linear range but unless noted was 25 ng per reaction. Magnesium concentration of the Flexi Rabbit Reticulocyte Lysate varies by lot, so the amount of magnesium acetate added varies by lot but unless noted the magnesium concentration for the assay was 1.88 mM. L-ascorbic acid was buffered in 750 mM HEPES buffer at a 5X concentration was added to the translation reaction immediately before mRNA was added. Reactions were incubated at 30°C for 30 minutes and then flash frozen in liquid nitrogen. The reactions were thawed and Luciferase Assay System (Promega) was used to measure FLuc activity with a luminometer as described above but for a single luciferase.

Table 2.20. Composition of a Promega Flexi RRL reaction

Standard Translation Reaction (25 µl reaction volume)	
Component	Volume (µl)
Flexi Rabbit Reticulocyte Lysate	16.5
Amino Acid Mixture, Minus Leucine, 1 mM	0.25
Amino Acid Mixture, Minus Methionine, 1mM	0.25
Magnesium Acetate, 25 mM	To 1.88 mM
Potassium Chloride, 2.5 M	0.7
RNasin Plus (Promega)	0.5
mRNA	0.5
HEPES Buffered Ascorbate or HEPES	5
RNase Free Water	To 25

2.8. Western blot detection of translation products

In vitro transcribed capped and purified mRNA from plasmids pC5568 (uCC-FLuc) and pC5867 (OF-FLuc) was translated as described above with or without 5 mM ascorbate but the reactions were allowed to proceed for 1 hour. The reactions were flash frozen in liquid nitrogen and then thawed. The *in vitro* translation reactions were mixed so that 15 μ l of loaded reaction would contain either 5 μ g or 10 μ g of total protein by diluting the translation reaction with water and adding 4X NuPAGE LDS Sample Buffer (Invitrogen), 10X NuPAGE Sample Reducing Agent (Invitrogen), and 10X BME. Previous Bradford assays on RRL reactions show that the protein concentration is approximately 181 μ g protein/ μ l. The samples were incubated at 70°C for 10 minutes, centrifuged at max RT for 1 minute, and then placed back at 70°C.

After preparing the samples, 15 μ l of each was loaded onto a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and run with NuPAGE MOPS buffer plus NuPAGE Antioxidant (Invitrogen) at 180 V for about 1 hour. The proteins were transferred to a 0.45-micron nitrocellulose membrane (Invitrogen) with NuPAGE transfer buffer and 20% methanol at 30 V for 1 hour 10 minutes. The membrane was rinsed with DI water and then Ponceau stained for several minutes.

The blot was trimmed and cut at the 39 kDa marker and then washed once with TBS-T. The blots were blocked in 5% BSA in TBS-T for the lower part of the blot that contained GAPDH and the upper blot that contained uCC-FLuc was blocked with 3% non-fat milk in TBS-T for 1 hour at RT on a nutator. The lower part of the membrane was incubated with anti-GAPDH (Abcam) at 1:5,000 in 5% BSA in TBS-T at 4°C overnight and the upper part of the blot was incubated with anti-Luci17 (Abcam) at 1:500 in 3% non-

fat milk in TBS-T at 4°C overnight on a nutator. After incubation, the blots were washed 3 times for 5 minutes in TBS-T then ECL-anti-mouse IgG, Horseradish Peroxidase linked whole antibody (GE Healthcare) at 1:10,000 was added. For the lower blot the secondary antibody was diluted in 5% BSA in TBS-T, and for the upper blot the secondary antibody was diluted in 3% non-fat milk in TBS-T. The blots were incubated with secondary antibody for 1 hour at RT on a nutator and then washed 5 times for 5 minutes with TBS-T. After washing, SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) was added to the blots for 5 minutes. Amersham Hyperfilm MP (GE Healthcare) was used for 15-minute exposures to the FLuc blot. GAPDH required a shorter exposure time of 15 seconds. The films were developed as described above.

2.9. Peptidyl-tRNA release assay

The templates for the peptidyl-tRNA release assay were generated as G-blocks. Each block contains the T7 promoter for *in vitro* transcription, FLAG tag for peptide detection, a linker fragment from human chorionic gonadotropin beta subunit, and the peptide of interest is fused in-frame to the tag and linker (Fig. 2.B-D). FLAG-tagged uCC template was generated by PCR using 2X LiSpark Max SuFi PCR Master Mix (LifeSct), G-block #16, and the primers gp48_F and uCCStall_R. The OF FLAG-tagged uCC was generated using the same primers but with G-block #17 as a template. The stalling control gp48 template was generated using G-block #18 and the primers gp48_F and gp48_R. The PCR reactions were scaled up 4X for a final total volume of 100 μ l. The PCR products were pooled and purified using the High Pure PCR Purification Kit (Roche) and eluted in 50 μ l elution buffer, as described above. DNA concentration was determined using a NanoDrop ND-1000 Spectrophotometer.

The purified PCR product was used downstream in the T7 RibomAX Express Large Scale RNA Production System (Promega). The reactions were scaled up 2X: 40 μ l RibomAX Express T7 2X Buffer, 32 μ l purified PCR product (~4 μ g), 8 μ l T7 Express Enzyme Mix. The reactions were incubated at 37°C for 30 minutes. After *in vitro* transcription, the mRNAs were purified with the RNA Clean & Concentrator Kit (Zymo), as described after capping, and quantified using a NanoDrop ND-1000 Spectrophotometer.

In vitro translation was done as described above except for two modifications. The mRNAs were denatured at 65°C for 5 minutes then placed in an ice bath and translation of the mRNAs was done at 25°C for 90 minutes for all five mRNAs at 0 mM or 5 mM L-ascorbate. The reactions were stopped using liquid nitrogen. After the samples thawed, 20

μl was aliquoted into two tubes of 10 μl each. One tube was treated with 2 μl of RNase free water, the other tube was treated with 2 μl of RNase A at 0.1 mg/ml (Qiagen). The samples were mixed and incubated at RT for 20 minutes.

After incubation, the samples were diluted 1:10 in water and prepared for Western blot so that 15 μl of prepared sample would contain 1 μg of total protein, which was calculated to be 3 μl of diluted sample in 30 μl of sample buffer mix. Each preparation contained treated sample, 2X Laemmli buffer (BIO-RAD), 10X NuPAGE Sample Reducing Agent (Invitrogen), 10X BME, and water to adjust the volume to 33 μl per sample. The samples were heated to 70°C for 10 minutes, centrifuged at max RT for 1 minute, placed back at 70l, and 15 μl of each sample was loaded onto a 10% NuPAGE Bis-Tris gel (Invitrogen). The gel was run with NuPAGE MOPS buffer with antioxidant (Invitrogen) at 140 V until the dye front reached the bottom of the gel. The proteins were transferred to a 0.45-micron nitrocellulose membrane (Invitrogen) with NuPAGE transfer buffer and 20% methanol at 30 V for 1 hour 15 minutes. The membrane was rinsed with DI water and then Ponceau stained for several minutes.

After Ponceau staining, the membrane was trimmed, washed with TBS-T, and blocked at RT for 1 hour in 3% non-fat milk in TBS-T on a nutator. The blocked blot was incubated at 4°C overnight on a nutator with mouse monoclonal anti-FLAG M2 antibody (Sigma F3165) at 1:2,000 in 3% non-fat milk in TBS-T. After incubating overnight, the blot was washed 3 times for 5 minutes with TBS-T. ECL-anti-mouse IgG, Horseradish Peroxidase linked whole antibody (GE Healthcare) was used for incubation at 1:10,000 in 3% non-fat milk in TBS-T for 1 hour at RT on a nutator. After incubation, the blot was washed 5 times for 5 minutes with TBS-T. After washing, Amersham ECL Western

Blotting Detection Reagent (GE Healthcare) was added to the blot for 5 minutes. Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) was used for the blots if a 15-minute exposure was not sufficient. Amersham Hyperfilm MP (GE Healthcare) was used for exposures and the film was developed as described above.

3. Results

3.1. A noncanonical uORF in the plant GGP mRNA contains highly conserved amino acid motifs

3.1.1. Significance and rationale for study

We propose that the *GGP* uORF, also referred to as a uCC for simplicity in the text, regulates GGP expression through translational control in response to L-ascorbate because it has the three hallmark features of a uCC: conservation of a near-cognate initiation codon upstream of the conserved region; absolute conservation of amino acid motifs within the uORF-encoded peptide; and evidence of high ribosome density within the uORF including a peak indicating a translation pause site.

A key feature of the AZIN uCC is the sequence conservation at the amino acid level rather than at the nucleotide level, indicating that the encoded peptide has a function rather than the RNA. As first described by Laing et al., alignment of sequences of the uORF-like element in the 5'-UTR of the GGP mRNA from multiple plant species revealed strong amino acid sequence conservation near the C-terminus in dicotyledonous plants. Importantly, the codons encoding the conserved amino acid residues varied at the wobble position in the aligned sequences, indicating that the amino acid sequence was conserved, not the nucleotide sequence (Laing et al., 2015). Of note, the sequence alignments used by Laing et al. did not include many single-celled algae species. The inclusion of additional sequences from more genetically diverse species is expected to enhance the usefulness of the alignment as an aid to identify which amino acid residues are critical for regulation of GGP mRNA translation.

To broaden the diversity of species included in the sequence alignment, I compiled the sequences of the 5'UTRs from a highly diverse set of GGP mRNAs. The NCBI transcriptome shotgun assembly (TSA) and expressed sequence tag (EST) databases were searched using *A. thaliana* GGP as a reference sequence and I specifically searched for algae species. The ancestors of land plants diverged from green algae between 450 and 500 million years ago (MYA) (Ulsvkov et al., 2013), maximizing the diversity in the sequence alignment. The sequences were translated in all six reading frames, and the text was searched for amino acid motifs highlighted by Laing et al., specifically the GGRG motif and the C-terminal glycine repeat (Laing et al., 2015).

3.1.2. The GGP uORF contains highly conserved motifs

Newly identified sequences were compiled with the sequences previously identified by Laing et al. to generate a 21 species alignment using CLUSTALW (3.1A). The alignment showed that there are five perfectly conserved motifs, but that the length of the uORF varies. The shortest GGP uORF is only 32 codons and the longest could be as long as 103 codons; however, the longest uORFs contain multiple potential in-frame near-cognate start codons, so the translated uORF might be shorter than 103 codons. In ribosome profiling studies of *A. thaliana*, ribosome footprints extended to a potential near-cognate ACG start codon suggesting that the GGP uORF could be at least 65 codons long (Laing et al., 2015). As discussed below, the greatest amino acid sequence conservation in the uORF includes a nearly 30 amino acid span near the C-terminus. There is no conservation in the N-terminal region of the uORF, which ranges in size from at least six codons to around 30 codons before the first conserved motif. There is also no conservation in the length of the C-terminal region between the last conserved polyglycine motif and the stop

codon with the distance ranging from one to 16 codons between the last glycine residue and the stop codon. In addition to the lack of conservation within the C-terminal region of the uORF, there is also no conservation in the distance between the uORF stop codon and the AUG start codon of the *GGP* coding sequence (CDS) (data not shown). In some species the *GGP* coding sequence start codon was over 100 nucleotides downstream of the uCC stop codon, while in other cases the 3' end of the uORF overlapped the beginning of the *GGP* coding sequence. In these latter cases, the *GGP* start codon was always located 3' of the last conserved motif in the uORF.

The most striking feature of the *GGP* uCC is the near perfect conservation of five motifs or single residues (Fig 3.1A, B) within a fixed span of 23 amino acids. Starting at the N-terminus of this conserved region, the first motif, GGRG, shows 100% conservation. Five codons after the GGRG motif, the sequence GGxPx_D (x = any residue) is also 100% conserved. A hydrophobic residue often follows the aspartate and then after a four-codon gap, there is a C-terminal glycine repeat. In most cases, the repeat consists of three consecutive glycine residues; however, in at least two organisms, only the C-terminal two glycines are conserved. As noted above, another striking feature is the perfect conservation of the length (23 amino acids) of the conserved region from the start of the first conserved motif to the last glycine residue in the C-terminal glycine repeat.

A

```

Volvox_carteri_f_na      -----MIFESHPIK-----DLAELRFRVPTGGGRCLRFSGGRKQDIAQAGGLRA-----56
Chlamydomonas_debary    -----MHTA-----LASEDLPVRAAAVCSGRVRFSCGGRKDAQAEAGGGKRV-----54
Chlamydomonas_sphaer    -----MHQQLRKT-----RVTPVPRAAASVCSGRVRFSCGGRKDALSEAGGGVA-----58
Gonium_pectoralis     -----TKTACQKTPVCPALP-----CARSRPAVAVGGGRCCRFAGGRKQGLCAAGGGRI-----53
Coleochaete_obthacula  -----LHSLHQSPKASNLPOCH-P-----VRKCK-VIKLGTVAPEASVWRDSDRYGSIAIHLCSLVLSRPAVGGGRSEGSVDLCLAGGGQWFLGFSQAQIFLC*103
Spirogyra_praeensis    -----NSREQ-----NFRLDYLXCELSFVGGGRGANSSEGGSDLFLAGGGTFFNLYF-----54
Tetraselmis_subcordi   -----TEVVNCTLD-----TVHQISTNFFTTAN-VFTRPAGFAAASPRRSGGRANFSAAGCPSDFLRLAGGTV-----81
Klebsormidium_flaeci   -----LVLCLGTTA-----CPVPOGSA-----QVLSKLLSSPALHGRGANFSEGGPSDFLRLAGGGSAV*-----64
Coccomyxa_subellipso   -----IEMLVRLSP-----NFMHLRPEVKAFASEA-----AVCRAL-----DCAHQPSKQGLHKGGRCLRFSEGGPSDSLHLAAGGSL-----53
Chlamydomonas_reinhardtii -----TIKPSVAGL-----IQ-----NVTGR-----OSMASOPAVTANRSGGRANFSCGGRKDALSKAGGGQ*-----69
Trebouxia              -----TIKPSVAGL-----IQ-----NVTGR-----LRAVTFPVALHSGGRVRFSLGGCPSDLHLAGGGQ-----60
Chlamydomonas_acidophilus -----CRQPAPEARRPRKTRK-----LPAPC-----SVALRHMGLRRAGGRVFPFSLGGAPSDAVRKSGGGRRA*-----61
Auxenochlorella_pyrenoidosa -----LQADSAGRAN-----PAPP-----SHSMFSEVLIHRSGGRVFPFSLGGGRPQAVRKSGGGRRA*-----57
Chlorella_vulgaris     -----YSLAKTLO-----AKAFPRCLLFGKRLASAFVVRK-----TKRNP-----MSQOTHRSGGRVFPFSLGGGRPQAVRKSGGGRCA-----33
Chlamydomonas_moewusii -----VWFEHPTPRGCRAGRVVSRHEEGSAAQTVPDLSFGLAHMQAHYINQRPFMPVWV-ATE-----HCFGP-----MEVALVRRRSGGRCCPSLGGGRPQAVRKSGGGRVRC-----38
Auxenochlorella_pretiosa -----SFF-----PAFDLKKSGEQVIGTITHGSRGV-----SS-HVHIVRQKCLIEINPLPHGGRGALPSEGGSDLLFLLAGGGSSFNFFSFRF*-----83
Plectonochaete_thermophila -----SFF-----PAFDLKKSGEQVIGTITHGSRGV-----SS-HVHIVRQKCLIEINPLPHGGRGALPSEGGSDLLFLLAGGGSSFNFFSFRF*-----83
Arabidopsis_thaliana

```

B

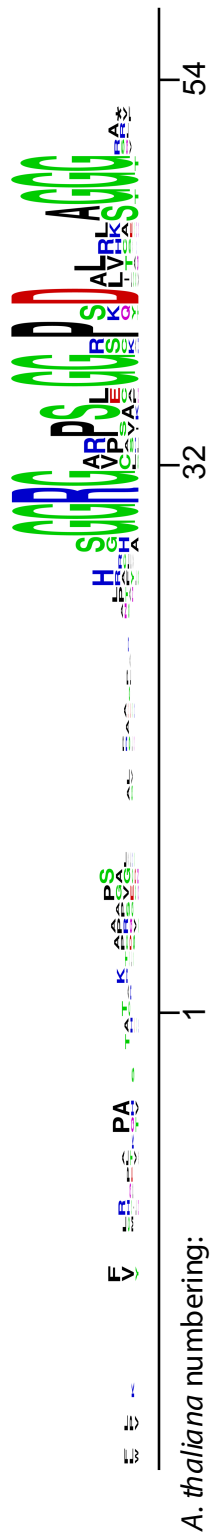


Figure 3.1. The GGP uCC is highly conserved across plants

Figure 3.1. The GGP uCC is highly conserved across plants

(A) The predicted amino acid sequences of the GGP uCC from 20 algae species and *A. thaliana* (red), representing approximately 450 million years of divergence, were aligned using Clustal Omega. Absolutely conserved amino acids are shown in orange, while residues conserved in >90% of the sequences are shown in blue. (B) WebLogo logogram of the alignment from panel A.

3.2. GDP-L galactose phosphorylase (GGP) mRNA translation is regulated by ascorbic acid in a uORF-dependent manner in mammalian cells

3.2.1. Significance and rationale for study

We propose that the GGP uORF, referred to as a uCC for simplicity in the text, regulates GGP expression through translational control in response to L-ascorbate. Ivanov et al. used reporters containing the *AZINI* uCC upstream of the luciferase coding sequence to examine polyamine and *AZINI* uCC regulation of *AZINI* expression in mammalian cell culture (Ivanov et al., 2018). Increased translation of the upstream uCC inhibited luciferase expression and polyamines regulated downstream luciferase reporter expression in an *AZIN* uCC-specific manner. If the hypothesized GGP uCC regulates GGP expression in response to L-ascorbate, replacing GGP with luciferase should lead to regulation of luciferase expression by the GGP uCC in response to high levels of L-ascorbate.

Laing et al. (2015) used plants for their studies but had to deal with the challenge of manipulating ascorbate levels indirectly by knocking down or overexpressing GGP, which is an essential gene (Laing et al., 2015). GGP and the GGP uORF are specific to plants, so it was conceivable that regulation would not work in a mammalian cell culture system; however, as human cells do not make ascorbate and it is not present in cell culture medium or FBS, ascorbate concentrations can be controlled in the mammalian cell culture system by simply adding ascorbate to the culture medium (Hata and Senoo, 1989; Takamizawa et al., 2004).

As addition of ascorbate to cell culture medium can change the pH of the medium and cause cell death, alternative methods have been developed to increase ascorbate concentrations in mammalian cells. Addition of ascorbate sesquimagnesium phosphate to

culture medium does not alter the pH or cause cell death (data not shown), and upon uptake into cells ascorbate sesquimagnesium phosphate is dephosphorylated by intracellular phosphatases and converted to biologically active dephosphorylated L-ascorbate.

3.2.2. GDP-L galactose phosphorylase (GGP) mRNA translation is regulated by ascorbic acid in a uORF-dependent manner

To monitor translational control conferred by the GGP uCC in the presence of ascorbate, reporters were generated containing the *A. thaliana* GGP uCC upstream of RLuc in place of GGP (Fig 2.4A, Fig. 3.2A). A CMV promoter was used to ensure high mRNA levels and boost the potential luciferase signal. The *A. thaliana* GGP uORF contains a non-conserved internal OF uORF that was disrupted by changing the AUG start codon to AAG (Fig 2.1A, B). Disrupting the internal uORF changes the uCC amino acid sequence His13 (CAT) to Gln13 (CAA), but the amino acid change is upstream of the conserved region. In order to keep the putative uCC near-cognate initiation site (ACG) in its native context, 12 nucleotides of the *A. thaliana* 5'UTR upstream of the uCC was included (Fig 2.4A). An OF control was generated where one nucleotide was removed to shift the frame directly upstream of the first conserved GGRG motif (Fig 2.1C). The stop codon was put back in-frame with the insertion of one nucleotide upstream of the stop codon. Downstream of the uCC stop codon, the 111 nucleotides between the uCC stop codon and the GGP start codon were left intact before the RLuc AUG initiation site.

To examine whether the GGP uCC regulates GGP expression in mammalian cells in response to 50 mM ascorbate, GGP-RLuc reporters were transfected into U2OS cells and the cells were incubated in the presence or absence of 50 mM ascorbate (Fig 3.2A). To

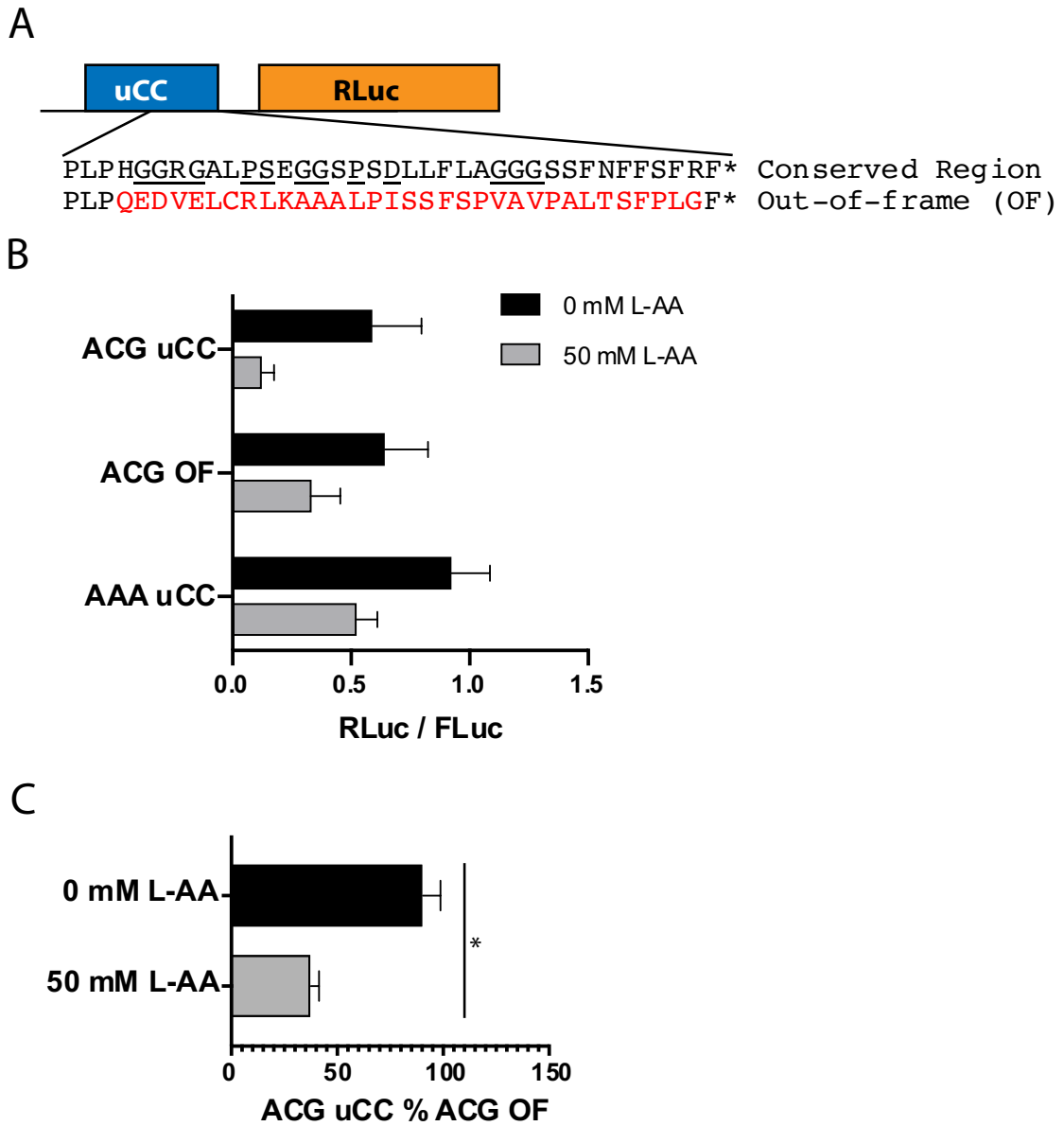


Figure 3.2. GGP mRNA translation is regulated by ascorbate in a uCC-specific manner *in vivo*

Figure 3.2. GGP mRNA translation is regulated by ascorbate in a uCC-specific manner *in vivo*

(A) GGP-Rluc reporters containing the *A. thaliana* GGP 5' leader with either the wild-type or OF control uCC. (ACG-initiated), OF, or uCC. (B) U2OS cells were transfected with the GGP-Rluc reporters from panel A containing the wild-type uCC, the OF control uCC, or the wild-type uCC lacking a start codon (ACG mutated to AAA). Cells were incubated 24 h in DMEM in the absence or presence of 50 mM ascorbate-P (L-AA). Rluc activity was normalized to the activity of an FLuc reporter transfection control. (C) wild-type reporter (ACG uCC) activity was calculated as a percentage of the OF reporter (ACG uCC % ACG OF) was calculated at 0 mM and 50 mM ascorbate-P. Error bars represent standard deviation (SD), * $p < 0.005$ (Student's two-tailed t-test; n=6, assayed in duplicate).

control for differences in transfection efficiency, an FLuc reporter was used for normalization. The FLuc reporter was plasmid p2luc which contains the SV40 promoter and terminator and no uCC upstream of FLuc. GGP-RLuc expression without translation of the uCC was monitored using a reporter with the uCC initiation site disrupted (AAA codon).

In the absence of ascorbate, highest luciferase expression was obtained with the AAA reporter lacking a uCC start codon. Both the wild-type and OF reporters yielded lower levels of luciferase, indicating the translation of the uCC (even at low levels from the near cognate start codon) impairs downstream translation (Fig. 3.2B). There is no uCC-

Table 3.1. Relative reporter mRNA levels – related to Fig. 3.2.

Reporter	Description	Treatment	Relative mRNA level (ACG uCC 0 mM set at 1)		Fold induction (0 mM L-AA / 50 mM L-AA)		
			RLuc mRNA	SD	RLuc mRNA	SD	RLuc activity
pC5914	ACG uCC	0 mM L-AA	1.00	0.51			
pC5914	ACG uCC	50 mM L-AA	0.19	0.30	0.19	0.30	0.21
pC5956	ACG OF	0 mM L-AA	0.98	0.41			
pC5956	ACG OF	50 mM L-AA	0.05	0.02	0.05	0.02	0.52

Table 3.1. Relative RLuc mRNA levels for experiments in Fig. 3.2 determined by qPCR. Standard deviation based on three biological replicates.

specific translational control in the absence of ascorbate as GGP-RLuc levels are the same for the wild-type and OF reporters. Upon addition of 50 mM ascorbate, GGP-RLuc expression was repressed to a greater extent by the wild-type, but not the OF.

In order to quantify the uCC-specific repression of GGP-RLuc expression in response to ascorbate, the data were normalized to the OF (Fig. 3.2C). At 0 mM ascorbate, no uCC-specific repression was observed and comparing the expression of the ACG uCC reporter as a percentage of the ACG OF reporter (ACG uCC % ACG OF) was nearly 100%. However, at 50 mM ascorbate, a dramatic uCC-specific decrease in GGP-RLuc expression was observed. Addition of ascorbate caused a 60% decrease ($p < 0.005$) in GGP-RLuc expression with the reporter containing an intact uORF (ACG initiation codon). These results allow two important conclusions: first, ascorbate can regulate GGP mRNA translation in mammalian cells; and second, ascorbate regulation of GGP mRNA translation is dependent on the amino acid sequence of the uCC (regulation is lost with the OF mutant).

To quantify how treatment with 50 mM ascorbate affects RLuc mRNA levels, RT-qPCR was performed. Human U2OS cells were transfected with the reporters used in Fig. 3.2 in the manner used for luciferase assays but scaled up 40X and using 6-well cell culture plates. After incubation in the presence or absence of 50 mM ascorbate, total RNA was isolated using Trizol Reagent and treated with TURBO DNase to remove plasmid contamination. Multiplexed TaqMan one-step RT-qPCR assays were performed. In short, cDNA was generated using random priming in the same tube as the qPCR assay. After cDNA generation, RLuc and *ActB* levels were quantified simultaneously using TaqMan assays targeting RLuc or *ActB*. In a TaqMan assay, two primers amplify a portion of the target cDNA and a probe that is specific to the amplified region anneals to the ssDNA. The probe contains a dye at one end and a quencher at the other. During the amplification process the dye is displaced by polymerase generating a fluorescent signal specific to the dye. RLuc was detected using a probe with FAM dye while *ActB* was detected using a probe with HEX dye. Relative RLuc mRNA levels were calculated using $2^{-\Delta\Delta Ct}$ utilizing *ActB* as a control housekeeping gene.

For the experiment in Fig. 3.2, relative GGP-RLuc mRNA levels were normalized to the relative GGP-RLuc mRNA level for the untreated ACG-initiated uCC (Table 3.1). The fold-induction in GGP-RLuc mRNA levels in response to 50 mM ascorbate was then calculated for each reporter. The fold-induction in RLuc activity in response to 50 mM ascorbate from Fig. 3.2 is also shown in Table 3.1. Treatment with 50 mM ascorbate reduces GGP-RLuc mRNA levels by about 5-fold for the reporter containing the wild-type *GGP* uCC and luciferase activity decreased to the same extent. In contrast, treatment with 50 mM ascorbate reduced GGP-RLuc mRNA levels by about 20-fold for the reporter

containing the OF, but luciferase activity only decreased 2-fold. The lack of correlation in relative mRNA levels and RLuc activity indicates that the changes in RLuc activity upon addition of ascorbate are not simply due to changes in mRNA levels.

3.2.3. Conservation of the motifs is essential for regulation

In order to examine whether the GGRG motif and the C-terminal glycine repeat (GGG) were necessary for ascorbate regulation of GGP, the conserved residues were replaced by alanine residues in reporters where GGP was replaced by RLuc as described above (Fig. 3.3A). As an additional control to test whether any alanine substitution in the uORF impaired regulation, the non-conserved residue Ser11 was also substituted by Ala. The S11A reporter was expected to show wild-type regulation of RLuc.

To determine whether alanine substitutions of the GGRG motif and C-terminal glycine repeat would derepress GGP-RLuc expression in response to 50 mM ascorbate, the alanine substitution reporters along with reporters containing the wild-type uCC or the OF control were transfected into U2OS cells. As shown in Fig. 3.3B, after normalization to the FLuc transfection control, expression of the GGP-RLuc reporters containing a wild-type uCC or the S11A mutant uCC showed substantial repression in the presence of ascorbate. In order to quantify the impact of the motif mutations on the ascorbate-specific regulation of GGP-RLuc expression, the normalized RLuc/FLuc levels for each reporter in the presence of ascorbate were normalized to values obtained in the absence of ascorbate (Fig. 3.3C). As expected, the reporters containing the wild-type uCC and S11A mutant uCC showed a greater than 60% reduction in GGP-RLuc expression upon addition of ascorbate.

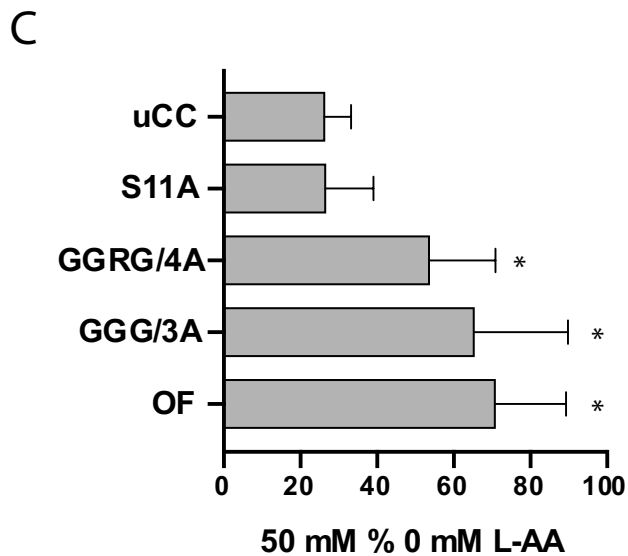
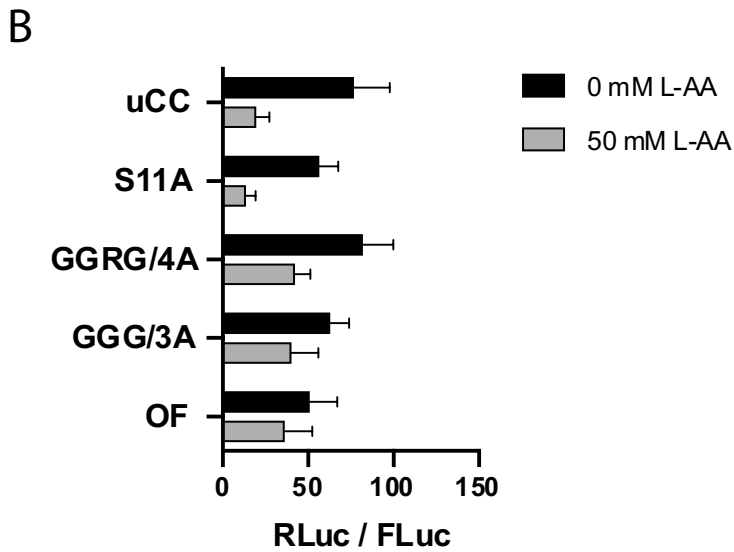
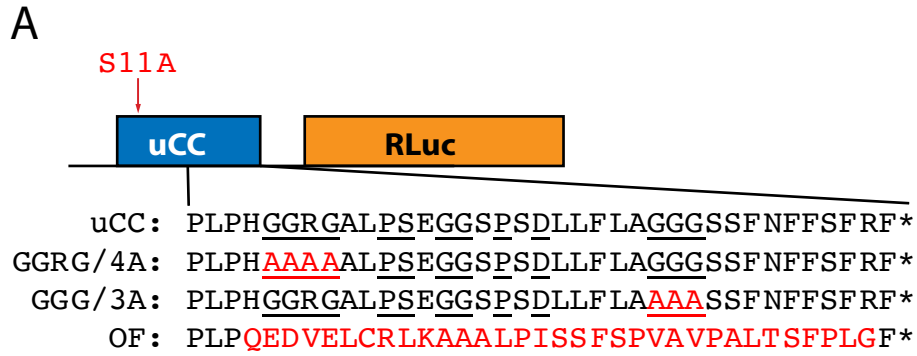


Figure 3.3. Conservation of the uCC amino acid motifs is essential for ascorbate regulation

Figure 3.3. Conservation of the uCC amino acid motifs is essential for ascorbate regulation

(A) GGP-RLuc reporters containing the *A. thaliana* GGP 5' leader with the wild-type or the indicated mutant uCCs (mutated residues in red). (B) U2OS cells were transfected with the indicated GGP-RLuc reporters and incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P (L-AA). RLuc activity was normalized to a co-transfected FLuc reporter (RLuc/FLuc). (C) The response of each reporter to ascorbate-P was calculated by dividing reporter activity in the presence versus absence of ascorbate (50 mM % 0 mM L-AA). Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; n=6, assayed in duplicate).

In contrast, the alanine substitutions for both the GGRG and GGG motifs, like the OF mutation in the uCC, severely impaired ascorbate-dependent repression of GGP-RLuc expression with at least a two-fold decrease in the repression ratio compared to a reporter with a wild-type uCC ($p < 0.005$). These results indicate that both the GGRG and GGG motifs are necessary for uCC regulation of GGP in response to ascorbate.

To quantify how treatment with 50 mM ascorbate affects GGP-RLuc mRNA levels, RT-qPCR was performed. Human U2OS cells were transfected with the reporters used in Fig. 3.3 in the manner used for luciferase assays but scaled up 40X and using 6-well cell culture plates. After incubation in the presence or absence of 50 mM ascorbate, total RNA was isolated using Trizol Reagent and treated with TURBO DNase to remove plasmid contamination. Multiplexed TaqMan one-step RT-qPCR assays were performed to

Table 3.2. Relative reporter mRNA levels – related to Fig. 3.3.

Reporter	Description	Treatment	Relative mRNA level		Fold induction		
			(ACG uCC 0 mM set at 1)		(0 mM L-AA / 50 mM L-AA)		
			RLuc mRNA	SD	RLuc mRNA	SD	RLuc activity
pC5914	ACG uCC	0 mM L-AA	1.00	0.51			
pC5914	ACG uCC	50 mM L-AA	0.19	0.30	0.19	0.30	0.27
pC6123	S11A	0 mM L-AA	0.93	0.18			
pC6123	S11A	50 mM L-AA	0.43	0.09	0.46	0.10	0.26
pC6121	GGRG/4A	0 mM L-AA	3.53	7.71			
pC6121	GGRG/4A	50 mM L-AA	0.16	0.05	0.05	0.02	0.52
pC6119	GGG/3A	0 mM L-AA	2.53	4.75			
pC6119	GGG/3A	50 mM L-AA	0.17	0.08	0.07	0.03	0.65
pC5956	OF	0 mM L-AA	0.98	0.41			
pC5956	OF	50 mM L-AA	0.05	0.02	0.05	0.02	0.72

Table 3.2. Relative RLuc mRNA levels for experiments in Fig. 3.3 determined by qPCR. Standard deviation based on three biological replicates.

quantify RLuc and *ActB* mRNA levels. Relative RLuc mRNA levels were calculated using $2^{-\Delta\Delta Ct}$ utilizing *ActB* as a control housekeeping gene.

For the experiment, relative RLuc mRNA levels for the mutants were normalized to the relative RLuc mRNA level for the untreated wild-type (ACG-initiated uCC) control (Table 3.2). The fold induction in RLuc mRNA levels in response to 50 mM ascorbate was then calculated for each reporter and displayed along with the fold induction in RLuc activity (from Fig. 3.3).

Treatment with 50 mM ascorbate reduced GGP-RLuc mRNA levels by about 5-fold for the wild-type reporter (ACG uCC) and luciferase activity decreased by about the same amount; however, whereas treatment with 50 mM ascorbate reduced GGP-RLuc

mRNA levels by about 20-fold for the OF, GGRG/4A, and GGG/3A substitutions, the luciferase activity decreased only 2-fold or less. Thus, the reduced repression of these three mutants by ascorbate is not due to higher mRNA levels. For the S11A mutant, treatment with ascorbate reduced GGP-RLuc mRNA levels by only about 2-fold while luciferase activity decreased almost 5-fold, consistent with the conclusion that the S11A mutation did not impair regulation.

3.2.4. The GGP uORF is specifically sensitive to ascorbic acid *in vivo*

To examine the specificity of the GGP uCC to ascorbate and determine whether ascorbate can affect regulation of a different uCC, the GGP wild-type uCC and GGP OF were used alongside the AAP reporter and its control, AAP D12N (Fig. 3.4A). The AAP is a well-described uORF that regulates production of CPAase A, an arginine biosynthetic enzyme subunit, in response to arginine (Luo et al., 1995; Werner et al., 1987). Mutating a single highly conserved aspartate residue, D12, in the AAP abolishes arginine-specific repression of CPAase A production (Freitag et al., 1996; Werner et al., 1987). Reporter constructs containing the AAP and the D12N control were developed by Ivaylo Ivanov and shown to respond to the addition of 25 mM Arg in mammalian cell culture (Ivanov et al., 2018). The reporters contain the mouse *azin1* uCC with the last 10 residues replaced by the C-terminal two-thirds of the *N. crassa* AAP sequence (Fig. 2.4D). The essential D12 residue is included in the wild-type AAP sequence but substituted for Asn (D12N) in the AAP control. Similar to the GGP-RLuc reporters, the AAP and AAP-D12N sequences are inserted in the 5' leader of an RLuc reporter. Addition of 25 mM Arg represses RLuc synthesis from the AAP, but not the AAP-D12N reporter (Ivanov, data not published).

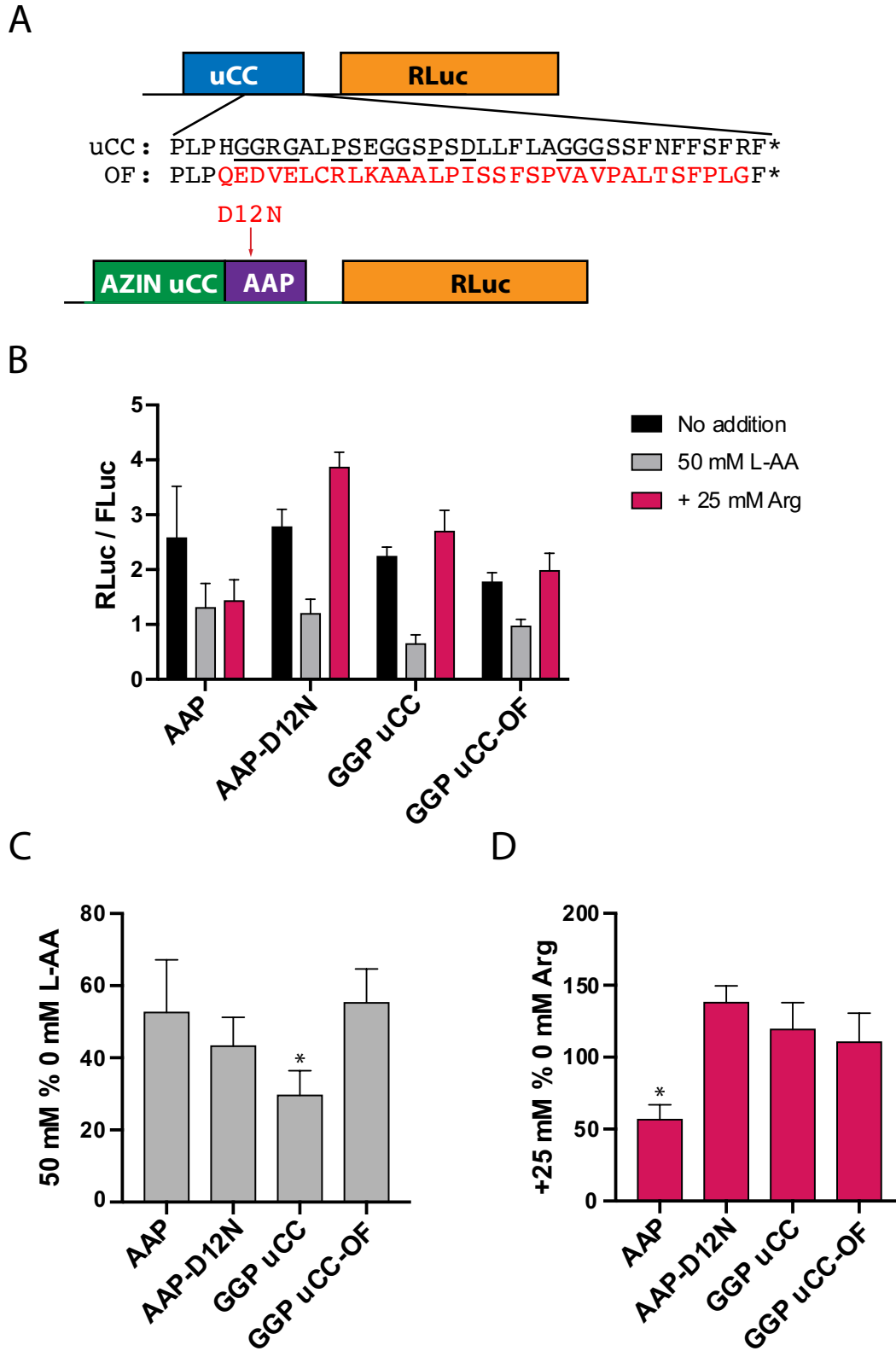


Figure 3.4. Regulation of GGP is specific to ascorbate and ascorbate does not regulate AAP

Figure 3.4. Regulation of GGP is specific to ascorbate and ascorbate does not regulate AAP

(A) GGP-RLuc reporters containing the *A. thaliana* GGP 5' leader or an AZIN-AAP fusion (mutated residues in red). (B) U2OS cells were transfected with the indicated GGP-RLuc reporters and incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P (L-AA) or in medium supplemented with an additional 25 mM arginine. RLuc activity was normalized to a co-transfected FLuc reporter (RLuc/FLuc). (C) The response of each reporter to ascorbate-P was calculated by dividing reporter activity in the presence versus absence of ascorbate (50 mM % 0 mM L-AA) and the significance of the response was calculated against the GGP OF reporter. (D) The response of each reporter to the addition of 25 mM arginine was calculated by dividing reporter activity in the presence versus absence of arginine (+25 mM % 0 mM Arg) and the significance of the response was calculated against the AAP-D12N reporter. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=6$, assayed in duplicate).

U2OS cells were transfected with the wild-type and OF GGP-RLuc reporters, as well as the AAP and AAP-D12N RLuc reporters, and the cells were grown in DMEM medium containing no additions, 50 mM ascorbate, or 25 mM Arg. In DMEM, RLuc expression was high for all four reporters (Fig. 3.4B). The addition of 50 mM ascorbate repressed expression from the wild-type GGP-RLuc reporter more substantially than any of the other three reporters (Fig. 3.4B). As expected, addition of 25 mM Arg repressed RLuc expression for the reporter containing the AAP, but not the AAP-D12N

Table 3.3. Relative reporter mRNA levels – related to Fig. 3.4.

Reporter	Description	Treatment	Relative mRNA level (ACG uCC 0 mM set at 1)		Fold induction (0 mM L-AA / 50 mM L-AA)			Fold induction (0 mM L-AA / +25 mM Arg)		
			RLuc mRNA	SD	RLuc mRNA	SD	RLuc activity	RLuc mRNA	SD	RLuc activity
IV2475	AAP	No addition	1.00	0.15						
IV2475	AAP	50 mM L-AA	0.05	0.02	0.05	0.02	0.53			
IV2475	AAP	25 mM Arg	0.59	0.03				0.59	0.03	0.58
IV2478	AAP-D12N	No addition	1.05	0.36						
IV2478	AAP-D12N	50 mM L-AA	0.17	0.06	0.16	0.05	0.44			
IV2478	AAP-D12N	25 mM Arg	1.33	0.93				1.26	0.88	1.39
pC5914	GGP uCC	No addition	1.00	0.51						
pC5914	GGP uCC	50 mM L-AA	0.19	0.30	0.18	0.28	0.30			
pC5914	GGP uCC	25 mM Arg	1.77	1.54				1.62	1.41	1.21
pC5956	GGP uCC-OF	No addition	0.98	0.41						
pC5956	GGP uCC-OF	50 mM L-AA	0.05	0.02	0.05	0.02	0.56			
pC5956	GGP uCC-OF	25 mM Arg	12.02	22.22				12.25	22.63	1.12

Table 3.3. Relative RLuc mRNA levels for experiments in Fig. 3.4 determined by qPCR. Standard deviation based on three biological replicates.

mutant (Fig. 3.4B). The addition of 25 mM Arg did not repress GGP-RLuc repression for either the wild-type uCC or OF control reporters.

To examine the magnitude of each reporters' response to ascorbate and Arg, the normalized RLuc levels were further normalized to the levels obtained in the absence of additions (Fig. 3.4C). After normalization, only the wild-type GGP uCC showed a significant inhibition in response to ascorbate ($p < 0.005$) with RLuc levels dropping nearly 50% compared to the other reporters. As expected, the AAP, known to respond to Arg, did not respond to ascorbate. Addition of Arg reduced expression from the AAP reporter by 3-fold ($p < 0.005$), whereas expression from the AAP-D12N, wild-type uCC and OF was not affected by the addition of Arg (Fig. 3.4D). Thus, the different uCC elements exhibit specificity for their effector, with the AAP specifically responding to Arg and the GGP

uCC specifically responding to ascorbate. Presumably, this metabolite specificity is dictated by the encoded amino acid sequence of the respective uCCs.

To quantify how treatment with 50 mM ascorbate or 25 mM Arg affects RLuc mRNA levels, RT-qPCR was performed as described previously. Relative RLuc (to *ActB* control) mRNA levels for each reporter in presence of ascorbate or Arg were normalized to the relative RLuc mRNA levels in the untreated samples. As shown in Table 3.3, the fold induction in RLuc mRNA levels in response to 50 mM ascorbate or 25 mM arginine was calculated and displayed along with the fold induction in RLuc activity. Treatment with 50 mM ascorbate lead to a greater than 10-fold reduction in RLuc mRNA levels for the AAP and AAP-D12N reporters, but RLuc activity was reduced only 2-fold. RLuc mRNA levels were also reduced for the GGP uCC and GGP OF reporters; however, the GGP uCC reporter showed the least response to 50 mM ascorbate at the mRNA level (only 5-fold reduction) but the greatest decrease in RLuc activity (over 3-fold). The lack of correlation in relative mRNA levels and RLuc activity in samples treated with 50 mM ascorbate indicates that the changes in RLuc activity are not simply due to changes in mRNA levels. Analysis of arginine regulation in cells transfected with AAP and AAP-D12N reporters showed that luciferase activity correlated with mRNA levels, so we cannot draw a strong conclusion regarding arginine regulation of the AAP and AAP-D12N reporters in these experiments. Of note, the reduced mRNA levels of the AAP reporter in the presence of Arg is consistent with previous reports that addition of Arg triggers nonsense-mediated decay (Gaba et al., 2005). The GGP reporters showed a similar response to Arg as observed for the AAP-D12N reporter with elevated mRNA levels

compared to the wild-type AAP reporter, and for unknown reasons the OF GGP-RLuc reporter showed a massive increase mRNA levels in the presence of Arg.

3.2.5. Treatment of mammalian cells with ascorbic acid does not trigger phosphorylation of eIF2 α

Cellular stress responses are known to downregulate global translation and upregulate translation of specific mRNAs through phosphorylation of the translation factor eIF2 α (Abastado et al., 1991; Dever et al., 1995; Hinnebusch, 2005; Lee et al., 2009; Palam et al., 2011; Vattem and Wek, 2004; Young and Wek, 2016; Young et al., 2015b). In mammalian cells, four distinct protein kinases transduce different stress signals resulting in phosphorylation of eIF2 α (Taniuchi et al., 2016). This phosphorylation impairs the function of eIF2 and thereby inhibits general translation. However, the translation of certain uORF-containing mRNAs can be upregulated under conditions of eIF2 α phosphorylation (Dever et al., 1995). While this latter mechanism does not appear relevant to the GGP mRNA, as its translation is repressed by ascorbate, we wondered if translation of the GGP mRNA was hypersensitive to induction of a general stress response. To test whether ascorbate treatment induces a general stress response, we used immunoblot analyses to examine eIF2 α phosphorylation in U2OS cells expressing the wild-type GGP-RLuc reporter and grown in the presence or absence of 50 mM ascorbate or of 100 μ M arsenite, an inducer of oxidative stress that activates the eIF2 α kinase HRI (McEwen et al., 2005). Whole cell extracts were subjected to immunoblot analysis using antibodies that specifically detected phosphorylation of eIF2 α on Ser51 or that detect eIF2 α independent of Ser51 phosphorylation (Fig. 3.5A). Normalizing the levels of eIF2 α -P to total eIF2 α (Fig. 3.5B) revealed that addition of ascorbate does not lead to increased levels of eIF2 α -

P, indicating that ascorbate regulation of GGP mRNA translation is likely not mediated by activation of a general cellular stress response.

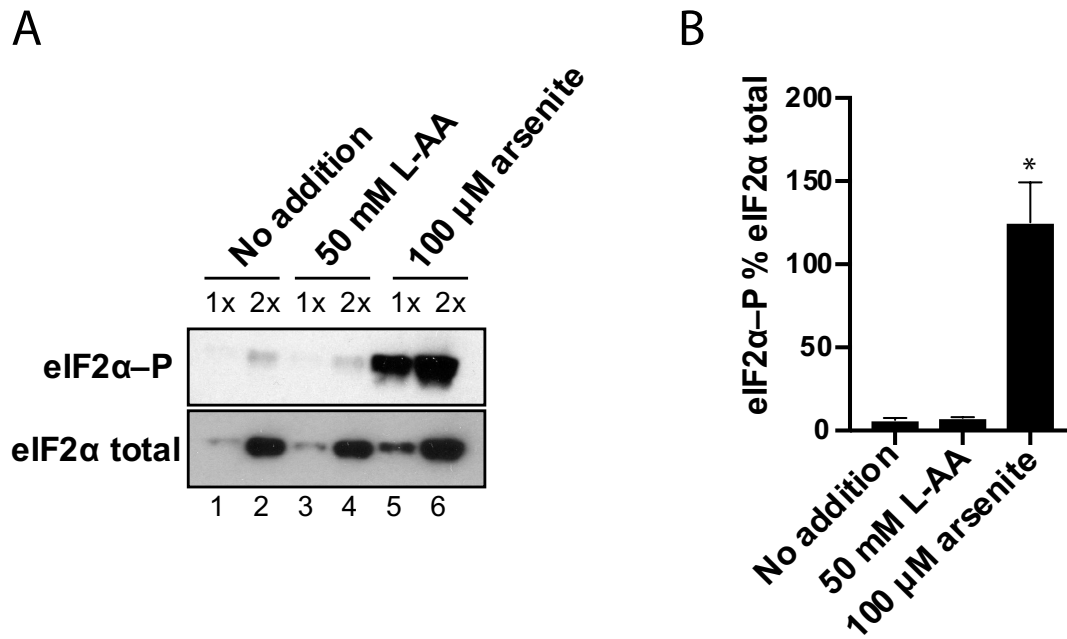


Figure 3.5. Ascorbate-P does not regulate GGP mRNA translation through eIF2 α phosphorylation

Figure 3.5. Ascorbate-P does not regulate GGP mRNA translation through eIF2 α phosphorylation

(A) U2OS cells were grown in medium containing 0 mM or 50 mM ascorbate-P (L-AA) or with the addition of 100 μ M arsenite and harvested for Western blotting. The membrane was probed for eIF2 α -P then stripped and probed for total eIF2 α . The blot shown is representative of three experimental replicates. (B) eIF2 α -P % eIF2 α total was calculated for each of the three experiments. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; n=3).

3.3. L-ascorbic acid enhances translation of the ACG-initiated GGP uORF in mammalian cell culture (*in vivo*)

3.3.1. Significance and rationale for study

The studies using the GGP-Luc reporters revealed that ascorbate regulation of GGP mRNA translation is dependent on the conserved uCC that initiates at a near cognate ACG codon. In previous studies to assess whether polyamines affected *AZINI* uCC translation, pairs of AZIN-Luc fusion reporters with the native AUU start codon or with an AUG start codon to monitor maximal initiation were utilized by Ivanov et al. (2018). With the native near-cognate start codon, enhanced initiation can be more easily detected after normalizing to the AUG-initiated reporter. Expression from the AUG-initiated uCC-luciferase reporters was inhibited upon addition of polyamines, due to the elongation defect in the uCC. Expression of the reporter with the near cognate start codon will suffer the same elongation defect as the AUG-initiated reporter but might also show enhanced initiation. Accordingly, normalizing the expression of the near-cognate initiated uCC-luciferase reporter to the AUG-initiated reporter will normalize the elongation defect and reveal the enhancement of near-cognate initiation. For the *AZINI* uCC, addition of spermidine triggered a 2-fold increase in AUU-initiated reporter activity following normalization to the corresponding AUG-initiated reporter (Ivanov et al., 2018).

To examine whether ascorbate controls translation of the GGP uCC, uCC-Luc reporters were generated by fusing an eight-codon linker containing a BamHI site between the last sense codon of the uCC and a Firefly luciferase reporter. The reporters were under control of the SV40 promoter and also contained an SV40 terminator (Fig. 2.4B). Like the GGP-RLuc reporter described above, the uCC start site context was maintained by

including 12 nucleotides of the *A. thaliana* 5'UTR upstream of the uCC start codon (Fig 2.4B). An OF control was generated by deleting one nucleotide directly upstream of the first conserved GGRG motif (Fig 2.1C) and then inserting one nucleotide immediately upstream of the linker to put FLuc back in-frame. As described above, the reporters were generated in pairs with either the native start site (ACG) or an AUG start site.

3.3.2. Ascorbic acid controls GGP uCC translation *in vivo*

In order to determine whether ascorbate affects GGP uCC translation, sets of FLuc reporters were transfected into U2OS cells and the cells were incubated in the presence or absence of 50 mM ascorbate (Fig 3.6A). To control for differences in transfection efficiency, an RLuc reporter was used for normalization. As expected, in the absence of 50 mM ascorbate, the ACG-initiated fusions show much lower uCC-FLuc expression than AUG-initiated fusions (Fig. 3.6B). The addition of 50 mM ascorbate lead to enhanced expression from the ACG-initiated uCC-FLuc fusion that was so great that it was visible even before normalization (Fig. 3.6B). In contrast, 50 mM ascorbate decreased expression from the AUG-initiated uCC-FLuc fusion, likely due to impaired translation elongation on the uCC. Both of these effects were lost with the OF-FLuc reporter. Likewise, expression of FLuc reporters lacking the uCC sequences and initiating with either an ACG or AUG codon did not respond to 50 mM ascorbate, indicating that ascorbate does not simply (or directly) affect the translation start codon selection (Fig. 3.6B).

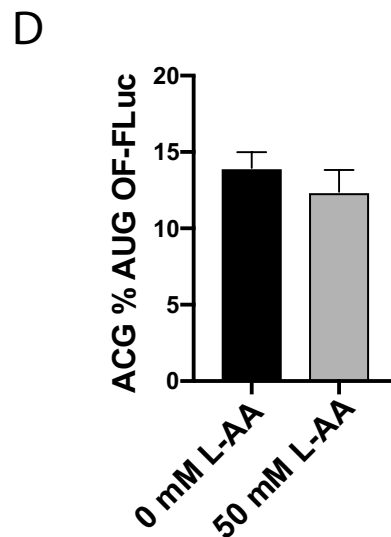
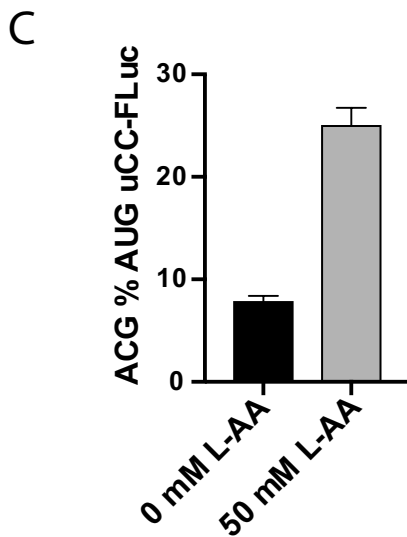
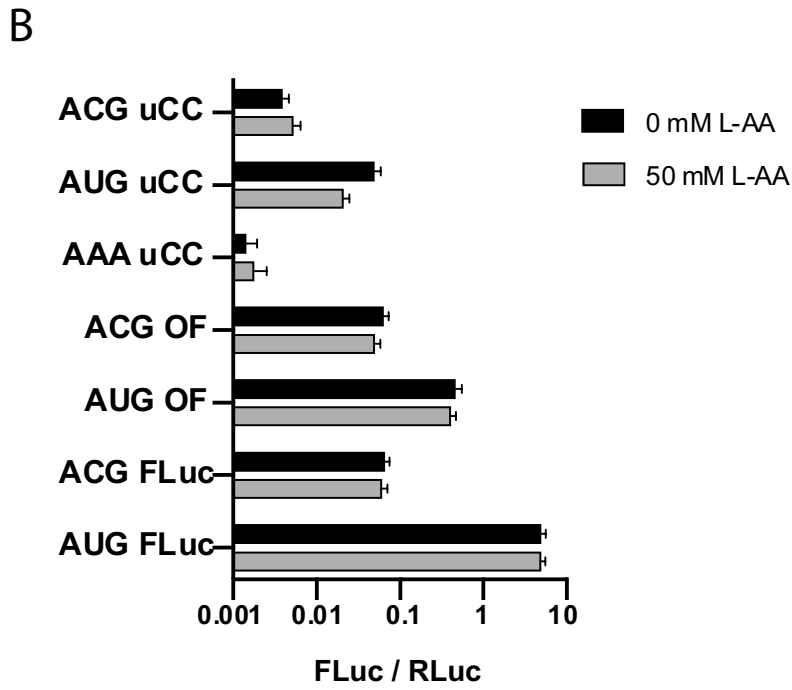
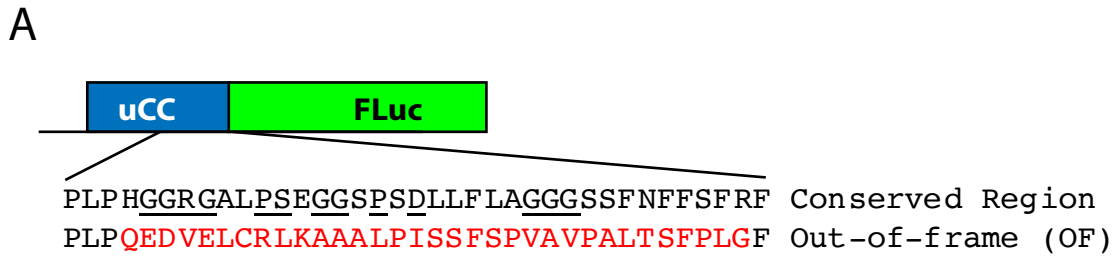


Figure 3.6. Ascorbate controls GGP uCC translation *in vivo*

Figure 3.6. Ascorbate controls GGP uCC translation *in vivo*

(A) GGP uCC-FLuc fusion reporters containing the *A. thaliana* GGP uCC fused in-frame to FLuc (mutated residues in red). (B) U2OS cells were transfected with the indicated FLuc reporters and incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P. FLuc activity was normalized to a co-transfected RLuc reporter (FLuc/RLuc). (C-D) Enhancement of initiation for the wild-type or OF-FLuc fusion was calculated by dividing the expression of the ACG-initiated reporter by the expression of the AUG-initiated reporter (ACG % AUG uCC-FLuc or ACG % AUG OF-FLuc). Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=6$, assayed in duplicate).

The ability of ascorbate to enhance uCC-FLuc expression was calculated by comparing the expression from the ACG- and AUG-initiated wild-type and OF-FLuc reporters (Fig 3.6C, D). After normalization to the expression of the AUG-initiated reporter, the addition of 50 mM ascorbate was found to increase synthesis of the ACG-initiated reporter by 3-fold ($p < 0.005$). This enhancement of uCC-FLuc expression is dependent on the wild-type uCC sequence as no enhancement was seen with the OF-FLuc fusion (Fig. 3.6D). Taken together, these results show that ascorbate enhances initiation at the ACG near-cognate start codon of the GGP uCC and that this enhancement is dependent on the encoded amino acid sequence of the uCC.

To quantify how treatment with 50 mM ascorbate affects FLuc mRNA levels, RT-qPCR was performed as described above except that FLuc specific primers were used in place of RLuc primers for the RT-PCR reactions.

Table 3.4. Relative reporter mRNA levels – related to Fig. 3.6.

Reporter	Description	Treatment	Relative mRNA level		Fold induction		
			(ACG uCC 0 mM set at 1)		(0 mM L-AA / 50 mM L-AA)		
			FLuc mRNA	SD	FLuc mRNA	SD	FLuc activity
pC5946	ACG uCC	0 mM L-AA	1.00	0.26			
pC5946	ACG uCC	50 mM L-AA	0.67	0.26	0.67	0.26	1.35
pC5947	AUG uCC	0 mM L-AA	1.22	1.01			
pC5947	AUG uCC	50 mM L-AA	0.42	0.14	0.34	0.11	0.43
pC5949	ACG OF	0 mM L-AA	1.02	0.37			
pC5949	ACG OF	50 mM L-AA	2.08	1.26	2.04	1.23	0.78
pC5950	AUG OF	0 mM L-AA	1.00	0.25			
pC5950	AUG OF	50 mM L-AA	1.21	0.92	1.21	0.92	0.88

Table 3.4. Relative FLuc mRNA levels for experiments in Fig. 3.6 determined by qPCR. Standard deviation based on three biological replicates.

The relative FLuc mRNA levels were all normalized to the levels of the wild-type untreated ACG-initiated uCC-FLuc fusion. As shown in Table 3.4, the fold induction of FLuc mRNA levels in response to 50 mM ascorbate was then calculated for each reporter and displayed along with the fold induction in FLuc activity from Fig. 3.6. Treatment with 50 mM ascorbate lead to a reduction in FLuc mRNA levels for the in-frame uCC-FLuc fusions, but the reduction in mRNA levels was not correlated with FLuc activity. In fact, the ACG-initiated uCC-FLuc fusion had a nearly 2-fold decrease in mRNA levels but there was an increase in FLuc activity. Therefore, the ability of ascorbate to increase expression from the ACG-initiated uCC-FLuc reporter is not due to an increase in FLuc mRNA levels.

3.3.3. Conservation of the GGP uCC motifs is essential for initiation regulation

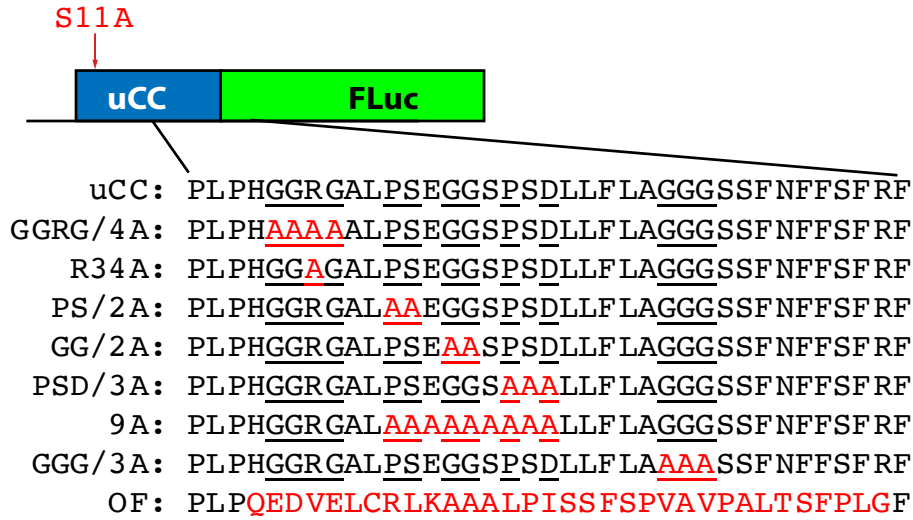
The GGP uCC contains five perfectly conserved amino acid sequence motifs in addition to a well, though not perfectly, conserved PS motif (Fig 3.1A, B). In order to

determine which motifs are essential for ascorbate control of uCC translation, the conserved motifs were individually substituted with alanine residues in the native ACG-initiated and the mutant AUG-initiated uCC-Luc fusion constructs (Fig. 3.7A). As a control, the non-conserved residue Ser11, upstream of the highly conserved region, was also mutated to alanine.

Following transfection into U2OS cells, and growth in the absence or presence of 50 mM ascorbate, reporter expression was measured and normalized to a co-transfected RLuc transfection control. As expected, the S11A substitution did not affect the response of the ACG-initiated or AUG-initiated fusions to 50 mM ascorbate, and the uCC-FLuc / RLuc levels responded like the wild-type uCC-FLuc fusion (Fig. 3.7B). For both the wild-type and S11A mutant, expression from the ACG-initiated fusion remained the same or increased in presence of ascorbate while the expression of the AUG-initiated fusion decreased. In contrast, alanine substitutions of any of the seven conserved motifs impaired the ability of ascorbate to enhance initiation of the ACG reporter and to repress expression from the AUG-initiated reporter, comparable to what was observed with the OF-FLuc fusion.

Normalization of the ACG-initiated uCC-FLuc reporter activities to the activity observed with the respective AUG-initiated reporter, revealed that ascorbate stimulated initiation for the wild-type and S11A mutant uCCs, but not for the OF reporter or any of the reporters in which a conserved uCC motif was substituted by alanine residues (Fig. 3.7C). This impact of the motif mutations blunting the effect of ascorbate to stimulate uCC translation is most easily observed by normalizing the ACG % AUG uCC-FLuc levels at 50 mM ascorbate to the same ratio in the absence of ascorbate (Fig. 3.7D). Interestingly,

A



B

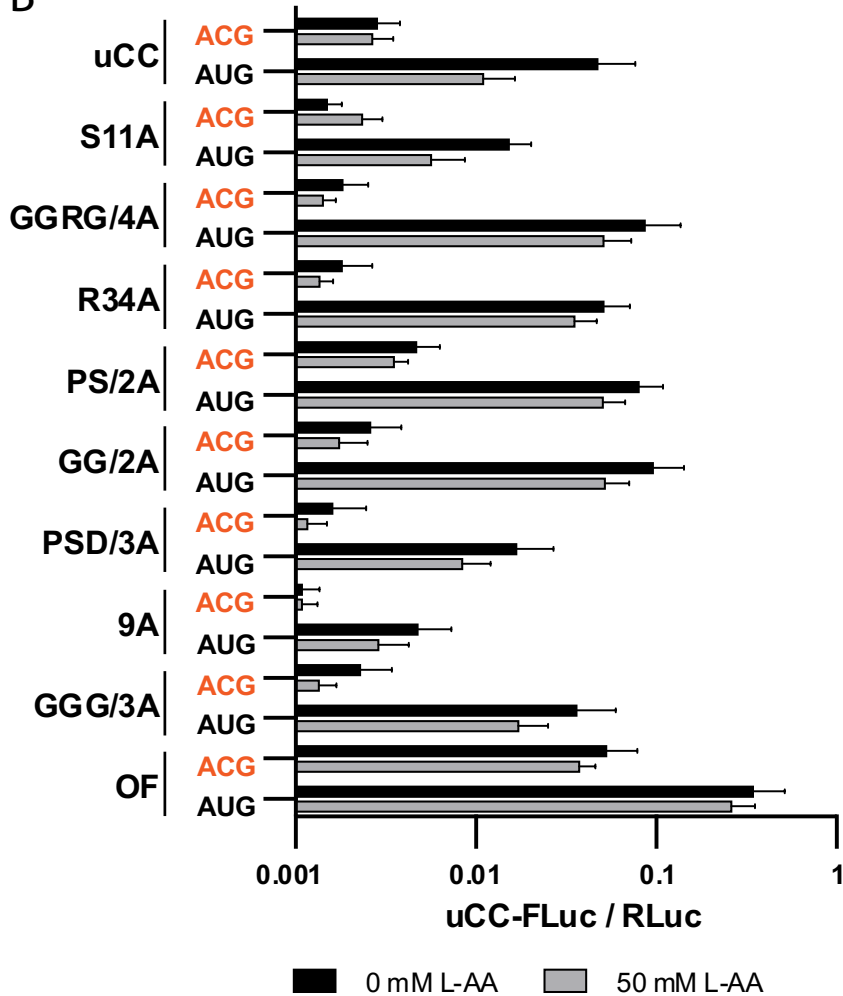


Figure 3.7. Conservation of the motifs is essential for ascorbate control of GGP uCC translation *in vivo*

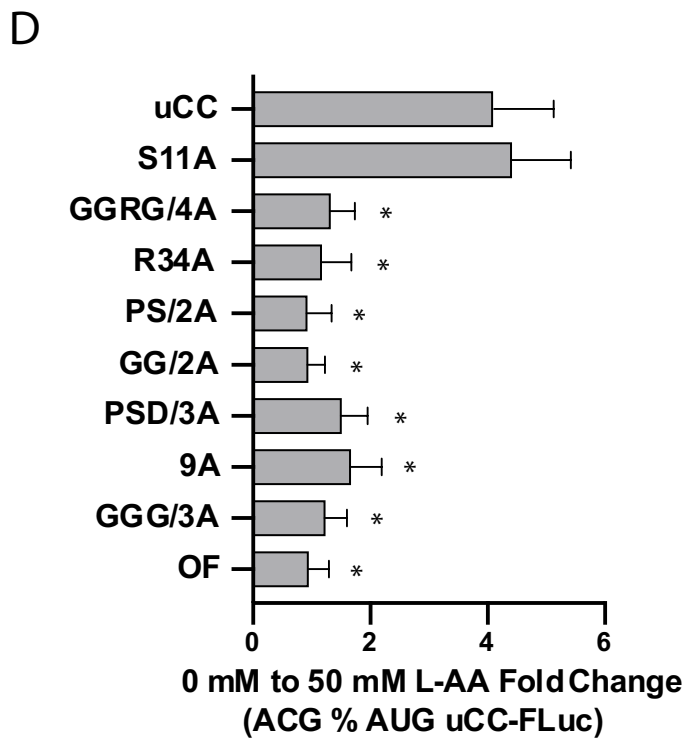
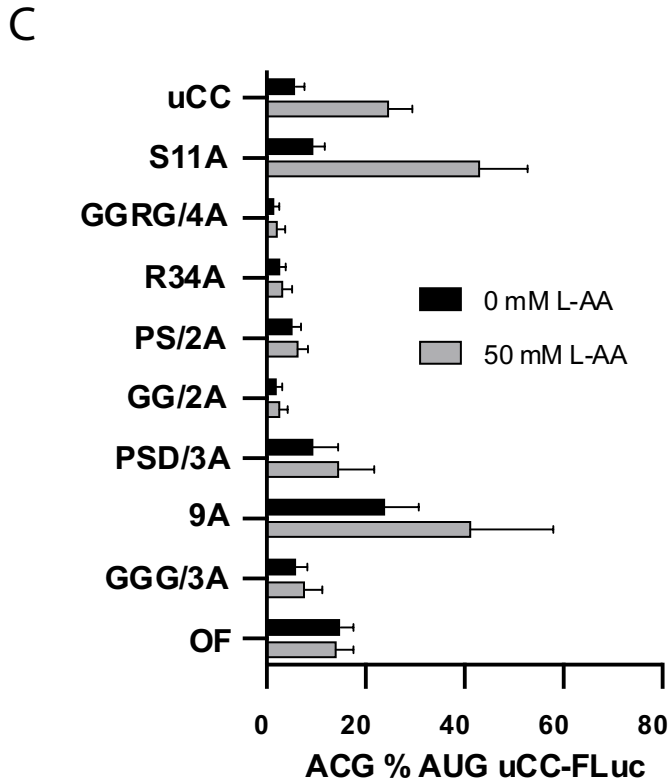


Figure 3.7. Conservation of the motifs is essential for ascorbate control of GGP uCC translation *in vivo*

Figure 3.7. Conservation of the motifs is essential for ascorbate control of GGP uCC translation *in vivo*

(A) GGP uCC-FLuc fusion reporters containing the *A. thaliana* GGP uCC fused in-frame to FLuc (mutated residues in red). (B) U2OS cells were transfected with the indicated FLuc reporters and incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P (L-AA). FLuc activity was normalized to a co-transfected RLuc reporter (FLuc/RLuc). (C) Enhancement of initiation for the uCC-FLuc fusion was calculated by dividing the expression from the ACG-initiated reporter by the expression from the corresponding AUG-initiated reporter (ACG % AUG uCC-FLuc). (D) The induction by 50 mM L-AA was calculated by normalizing the ACG % AUG uCC-FLuc levels at 50 mM L-AA by the values at 0 mM L-AA for each mutant (0 mM to 50 mM L-AA fold change). The significance of induction was calculated compared to the wild-type uCC. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=6$, assayed in duplicate).

the single R34A substitution had the same effect as mutating the entire GGRG motif. In both cases the fold change upon addition of ascorbate was reduced from over 4-fold to less than 2-fold. Though the PS motif is not perfectly conserved (Fig. 3.1), alanine substitution for this motif abolished the ascorbate-specific effect on uCC-FLuc translation (Fig. 3.7D). Substitutions of the GG and PSD motifs in the middle of the uCC had the same effect as substituting the whole nine-amino acid region with alanines or disrupting the entire conserved region with the OF mutation. Finally, substituting the C-terminal glycine residues with three alanine residues also abolished the ascorbate-specific induction of uCC-FLuc translation (Fig. 3.7D). Thus, all of the conserved motifs contribute to the ascorbate-

induced regulation of initiation at the uCC near-cognate (ACG) start site and translational control of GGP synthesis.

To quantify how treatment with 50 mM ascorbate affects uCC-FLuc mRNA levels, RT-qPCR was performed as described previously. The FLuc mRNA levels were first normalized to the levels of *ActB* mRNA and then these relative uCC-FLuc mRNA levels were all normalized to the level of the wild-type uCC-FLuc reporter mRNA from the untreated sample (no ascorbate) (Table 3.5). Finally, as shown in Table 3.5, the fold induction of each uCC-FLuc mRNA in response to 50 mM ascorbate was then calculated and displayed along with fold induction in FLuc activity obtained from Fig. 3.7. Treatment with 50 mM ascorbate did not lead to a consistent increase or decrease in uCC-FLuc mRNA levels across all the reporters. For the wild-type uCC and S11A substitution, the uCC-FLuc mRNA levels did not correlate with FLuc activity, consistent with the conclusion that ascorbate was controlling translation of these mRNAs. Likewise, for the constructs containing alanine substitutions at the conserved motifs, the diminished impact of ascorbate on FLuc activity cannot be accounted for by systematic changes in FLuc mRNA levels. This lack of correlation in relative mRNA levels and uCC-RLuc activity is consistent with the conclusion that ascorbate is affecting the translation of the uCC

Table 3.5. Relative reporter mRNA levels – related to Fig. 3.7.

Reporter	Description	Treatment	Relative mRNA level		Fold induction		
			(ACG uCC 0 mM set at 1)		(0 mM L-AA / 50 mM L-AA)		
			FLuc mRNA	SD	FLuc mRNA	SD	FLuc activity
pC5946	uCC ACG	0 mM L-AA	1.00	0.26			
pC5946	uCC ACG	50 mM L-AA	0.67	0.26	0.67	0.26	0.95
pC5947	UCC AUG	0 mM L-AA	1.22	1.01			
pC5947	UCC AUG	50 mM L-AA	0.42	0.14	0.34	0.11	0.23
pC5980	S11A ACG	0 mM L-AA	1.04	0.40			
pC5980	S11A ACG	50 mM L-AA	1.15	0.38	1.11	0.37	1.56
pC6045	S11A AUG	0 mM L-AA	1.26	1.12			
pC6045	S11A AUG	50 mM L-AA	0.92	0.25	0.74	0.20	0.37
pC5975	GGRG/4A ACG	0 mM L-AA	1.13	0.70			
pC5975	GGRG/4A ACG	50 mM L-AA	0.51	0.21	0.46	0.19	0.78
pC5964	GGRG/4A AUG	0 mM L-AA	0.98	0.18			
pC5964	GGRG/4A AUG	50 mM L-AA	0.54	0.17	0.55	0.18	0.59
pC5976	R344 ACG	0 mM L-AA	1.45	1.56			
pC5976	R344 ACG	50 mM L-AA	1.12	0.71	0.78	0.49	0.76
pC5965	R34A AUG	0 mM L-AA	1.33	1.14			
pC5965	R34A AUG	50 mM L-AA	1.74	2.05	1.30	1.53	0.69
pC5974	PS/2A ACG	0 mM L-AA	1.09	0.57			
pC5974	PS/2A ACG	50 mM L-AA	0.99	0.84	0.90	0.77	0.75
pC5963	PS/2A AUG	0 mM L-AA	1.04	0.43			
pC5963	PS/2A AUG	50 mM L-AA	0.81	0.63	0.77	0.60	0.63
pC5972	GG/2A ACG	0 mM L-AA	1.25	1.10			
pC5972	GG/2A ACG	50 mM L-AA	0.37	0.25	0.29	0.20	0.67
pC5961	GG/2A AUG	0 mM L-AA	1.02	0.33			
pC5961	GG/2A AUG	50 mM L-AA	0.66	0.32	0.64	0.31	0.55
pC5971	PSD/3A ACG	0 mM L-AA	1.09	0.59			
pC5971	PSD/3A ACG	50 mM L-AA	0.69	0.32	0.63	0.30	0.73
pC5960	PSD/3A AUG	0 mM L-AA	1.17	0.83			
pC5960	PSD/3A AUG	50 mM L-AA	0.79	0.29	0.68	0.25	0.50
pC5978	9A ACG	0 mM L-AA	1.05	0.48			
pC5978	9A ACG	50 mM L-AA	1.96	1.30	1.87	1.24	1.00
pC5967	9A AUG	0 mM L-AA	1.02	0.35			
pC5967	9A AUG	50 mM L-AA	1.01	0.90	0.99	0.88	0.60
pC5969	GGG/3A ACG	0 mM L-AA	1.03	0.35			
pC5969	GGG/3A ACG	50 mM L-AA	0.65	0.19	0.63	0.18	0.59
pC5958	GGG/3A AUG	0 mM L-AA	1.07	0.46			
pC5958	GGG/3A AUG	50 mM L-AA	1.14	0.67	1.07	0.63	0.48
pC5949	OF ACG	0 mM L-AA	1.02	0.37			
pC5949	OF ACG	50 mM L-AA	2.08	1.26	2.04	1.23	0.71
pC5950	OF AUG	0 mM L-AA	1.00	0.25			
pC5950	OF AUG	50 mM L-AA	1.21	0.92	1.21	0.92	0.76

Table 3.5. Relative FLuc mRNA levels for experiments in Fig. 3.7 determined by qPCR. Standard deviation based on three biological replicates.

3.3.4. Conservation of the motif spacing is essential for regulation of uCC

translation

The multiple sequence alignments of plant GGP uCC elements revealed that the spacing between the conserved motifs in the uCCs is perfectly maintained across all the plant species (Fig. 3.1A). To assess whether the spacing between the conserved motifs is important, we examined the effect of inserting residues between conserved motifs within the uCC (Fig. 3.8A). One or three alanine residues were inserted upstream of the first conserved motif or between the conserved proline and aspartic acid residues in the middle of the conserved uCC motifs. As described for the previous uCC-FLuc reporters, the alanine insertion reporters were generated in pairs with the native (ACG) or AUG start sites. The reporters, along with the RLuc transfection control plasmid, were transfected into U2OS cells grown in the presence or absence of 50 mM ascorbate.

FLuc activity was normalized to RLuc activity from the co-transfected control plasmid (Fig. 3.8B). The reporters with one or three alanine residues inserted after Pro28 behaved the same as the wild-type uCC-FLuc fusion: the ACG-initiated fusion expression increased and AUG-initiated fusion expression decreased with the addition of 50 mM ascorbate. Thus, inserting alanine residues upstream of the conserved region did not affect translation of the ACG- or AUG-initiated uCC-RLuc fusions. In contrast, inserting one or three alanine residues after Pro44, thus changing the spacing between the conserved Pro44

and Asp46 residues, caused the fusions to behave like the OF control with modest reductions in uCC-FLuc expression upon addition of 50 mM ascorbate.

To quantify the enhancement of initiation in the presence of ascorbate, luciferase activity obtained with the ACG-initiated uCC-FLuc reporter was normalized to the levels observed with the AUG-initiated control (Fig. 3.8C). After normalization, the reporters with the alanine residues inserted upstream of the conserved region showed a noticeable increase in ACG-initiated uCC-FLuc expression at 50 mM compared to 0 mM ascorbate. This was not true for the insertions within the conserved region.

To more clearly reveal the impacts of the insertions on uCC translation, the data were normalized again comparing the levels observed in the presence versus the absence of ascorbate. As shown in Fig. 3.8D), the native uCC, as well as the mutants containing one or three alanine residues upstream of the conserved motifs, showed a 2–4-fold increase in uCC translation in the presence of ascorbate. In contrast, adding even a single alanine residue in the middle of the conserved region, like the OF mutation, abolished the ability of ascorbate to induce initiation at the ACG start site of the native uCC (Fig. 3.8D) and to inhibit translation elongation on the AUG-initiated uCC (Fig. 3.7D). As a single amino acid insertion within the middle of the conserved region abolished regulation, these results provide a rationale for the perfect conservation of spacing between the conserved motifs in the amino acid sequence alignment (Fig. 3.1A, B).

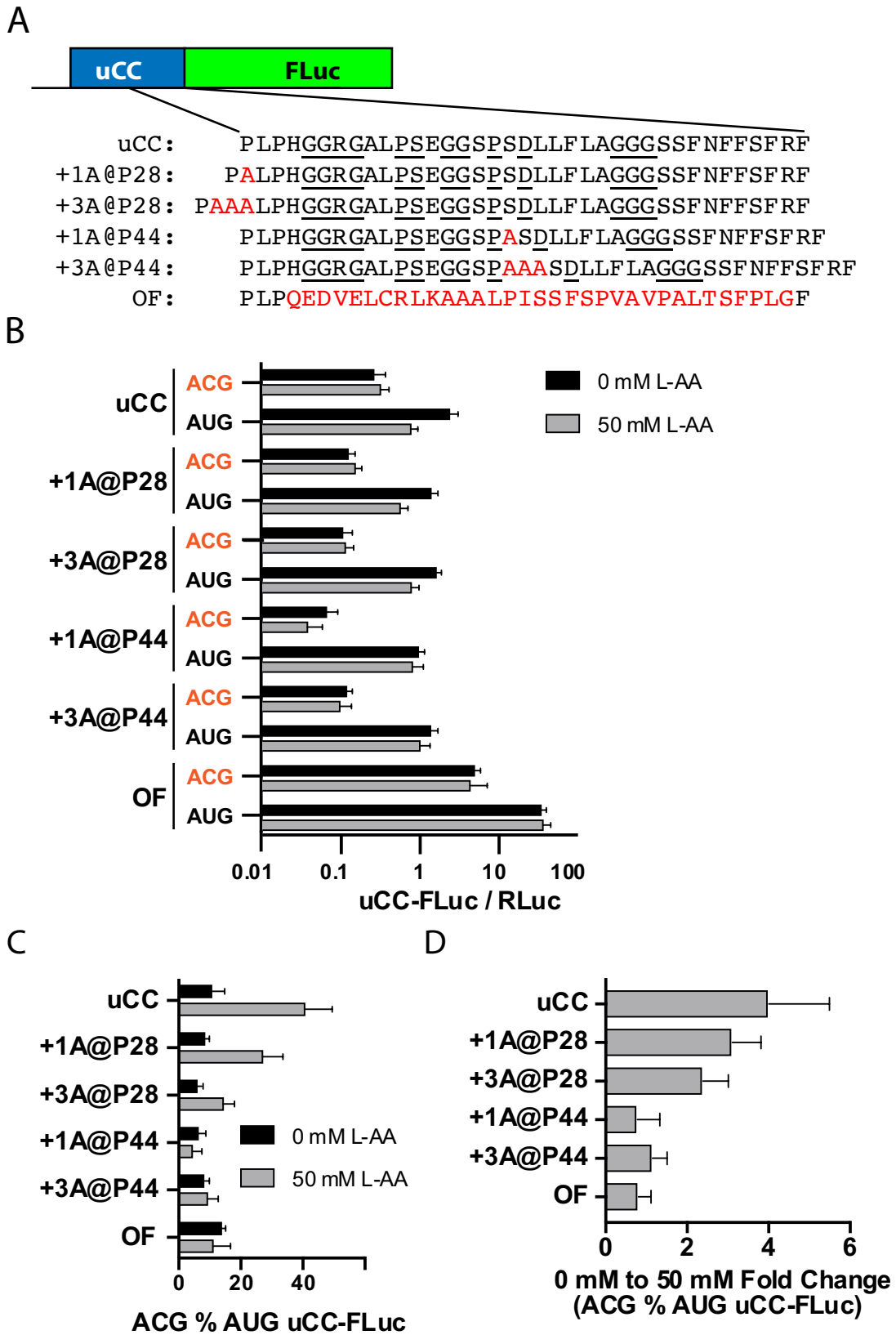


Figure 3.8. Conservation of motif spacing is essential for regulation by ascorbate

Figure 3.8. Conservation of motif spacing is essential for regulation by ascorbate

(A) GGP uCC-FLuc fusion reporters containing the *A. thaliana* GGP uCC fused in-frame to FLuc (inserted and mutated residues in red). (B) U2OS cells were transfected with the indicated FLuc reporters and incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P (L-AA). FLuc activity was normalized to a co-transfected RLuc reporter (FLuc/RLuc). (C) Enhancement of initiation for the uCC-FLuc fusion was calculated by normalizing the activity from the ACG-initiated reporter to the activity from the corresponding AUG-initiated reporter (ACG % AUG uCC-FLuc). (D) The induction of uCC translation by 50 mM L-AA was calculated by dividing the normalized ACG % AUG uCC-FLuc levels at 50 mM L-AA by the ACG % AUG uCC-FLuc at 0 mM L-AA for each mutant (0 mM to 50 mM L-AA fold change). The significance of induction was calculated in comparison to the wild-type uCC. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=6$, assayed in duplicate).

To quantify how treatment with 50 mM ascorbate affects FLuc mRNA levels, RT-qPCR was performed. Human U2OS cells were transfected with the reporters used in Fig. 3.8 in the manner used for luciferase assays but scaled up 40X and using 6-well cell culture plates. After incubation in the presence or absence of 50 mM ascorbate, total RNA was isolated and relative FLuc mRNA levels were determined as described previously.

The relative FLuc mRNA levels were determined in comparison to the endogenous *ActB* mRNA and then these ratios were normalized to the relative FLuc mRNA level obtained for the untreated ACG-initiated uCC-FLuc reporter. As shown in

Table 3.6. Relative reporter mRNA levels – related to Fig. 3.8.

Reporter	Description	Treatment	Relative mRNA level (ACG uCC 0 mM set at 1)		Fold induction (0 mM L-AA / 50 mM L-AA)		
			FLuc mRNA	SD	FLuc mRNA	SD	FLuc activity
pC5946	uCC ACG	0 mM L-AA	1.00	0.26			
pC5946	uCC ACG	50 mM L-AA	0.67	0.26	0.67	0.26	1.21
pC5947	UCC AUG	0 mM L-AA	1.22	1.01			
pC5947	UCC AUG	50 mM L-AA	0.42	0.14	0.34	0.11	0.33
pC6194	+1A@P28 ACG	0 mM L-AA	0.98	0.19			
pC6194	+1A@P28 ACG	50 mM L-AA	0.80	0.71	0.82	0.73	1.23
pC6198	+1A@P28 AUG	0 mM L-AA	1.18	0.73			
pC6198	+1A@P28 AUG	50 mM L-AA	0.24	0.09	0.20	0.08	0.41
pC6195	+3A@P28 ACG	0 mM L-AA	1.08	0.53			
pC6195	+3A@P28 ACG	50 mM L-AA	17.12	43.16	15.86	39.99	1.09
pC6199	+3A@P28 AUG	0 mM L-AA	1.01	0.32			
pC6199	+3A@P28 AUG	50 mM L-AA	1.06	0.28	1.05	0.27	0.48
pC6196	+1A@P44 ACG	0 mM L-AA	1.12	0.66			
pC6196	+1A@P44 ACG	50 mM L-AA	0.31	0.10	0.28	0.09	0.58
pC6200	+1A@P44 AUG	0 mM L-AA	1.11	0.62			
pC6200	+1A@P44 AUG	50 mM L-AA	0.46	0.08	0.41	0.07	0.85
pC6197	+3A@P44 ACG	0 mM L-AA	1.00	0.29			
pC6197	+3A@P44 ACG	50 mM L-AA	1.33	0.95	1.33	0.95	0.82
pC6201	+3A@P44 AUG	0 mM L-AA	1.46	1.56			
pC6201	+3A@P44 AUG	50 mM L-AA	3.18	4.58	2.17	3.13	0.73
pC5949	OF ACG	0 mM L-AA	1.02	0.37			
pC5949	OF ACG	50 mM L-AA	2.08	1.26	2.04	1.23	0.88
pC5950	OF AUG	0 mM L-AA	1.00	0.25			
pC5950	OF AUG	50 mM L-AA	1.21	0.92	1.21	0.92	1.07

Table 3.6. Relative FLuc mRNA levels for experiments in Fig. 3.8 determined by qPCR; standard deviation based on three biological replicates.

Table 3.6, the fold induction in FLuc mRNA levels in response to 50 mM ascorbate was then calculated for each reporter and displayed along with the fold induction in FLuc activity obtained from Fig. 3.8. Treatment with 50 mM ascorbate did not lead to a

consistent increase or decrease in FLuc mRNA levels across all the sets of reporters. In cells transfected with the uCC ACG reporter, FLuc mRNA levels decreased while FLuc activity increased. This was also true for the +1A@P28 ACG reporter, confirming that ascorbate induced translation of these mRNAs. While the +1A@P44 reporter showed decreases in both mRNA levels and Luc activity, which might indicate that the reduced mRNA levels contributed to the reduced uCC-Luc activity, the +3A@P44 reporter with native ACG start codon showed elevated mRNA levels but not FLuc activity. Thus, the apparent insensitivity associated with the +3A@P44 mutation cannot be attributed to masking mRNA changes. Taken together, are consistent with the conclusion that disrupting the spacing of the conserved motifs in the uCC interferes with ascorbate control of uCC translation.

3.4. Ascorbate regulates uCC translation in rabbit reticulocyte lysate (*in vitro*)

3.4.1. Significance and rationale for study

While the experiments in U2OS cells established that ascorbate acts through the uCC to regulate GGP synthesis, the cell-based assays limit the ability to test the ascorbate specificity of regulation and to monitor ribosome stalling during uCC translation. Both of these tests can be more easily performed using *in vitro* translation extracts. In their studies examining the AAP and arginine-mediated regulation of translation, Wang and Sachs (1997, 1999) used several different *in vitro* translation systems including commercial wheat germ extracts, and lab-prepared *N. crassa* and *S. cerevisiae* extracts. As the GGP mRNA is native to plants including wheat, we first considering using wheat germ extracts to monitor ascorbate regulation of GGP mRNA translation. However, in preliminary experiments, we found that addition of ascorbate dramatically inhibited translational activity in the wheat germ extracts (data not shown), rendering them unsuitable for studying ascorbate regulation of GGP mRNA translation. In contrast to the wheat germ extracts, we found that general translational activity in commercial rabbit reticulocyte lysates (RRLs) was maintained upon addition of ascorbate, and therefore chose to study GGP mRNA translation in RRLs. There is precedent for studying nascent chain dependent stalling in non-native translation systems. AAP, a fungal peptide, has been studied in wheat germ extracts as well as rabbit reticulocyte lysates (Wang et al., 1999; Wang and Sachs, 1997; Wu et al., 2012). In addition, AdoMet-induced translational stalling on the *CGSI* mRNA from *A. thaliana* has been studied using rabbit reticulocyte lysates (Onouchi et al., 2008). Despite the fact that mammals do not express cystathionine gamma-synthase,

encoded by *CGSI*, AdoMet induced a sequence-dependent translation elongation arrest in rabbit reticulocyte lysate (Onouchi et al., 2008).

The mRNAs for *in vitro* translation assays were generated by *in vitro* transcription. The GGP 5'-UTR from the transcription start site up to the stop codon of the uCC was cloned downstream of a T7 phage promoter (Fig. 2.2A). Like the *in vivo* reporters, the uCC was modified to remove the internal OF uORF. As preliminary experiments, see below, revealed that start codon selection stringency was relaxed in the RRLs, this system would not be effective for studying ascorbate stimulation of uCC translation; however, as ascorbate impaired AUG-initiated uCC-luciferase expression in cells (Fig. 3.6), we hypothesized that a similar response could be examined *in vitro*. Therefore, the translation start site of the uCC was changed to an AUG codon, but kept in its native context. An OF control was also constructed by deleting one nucleotide prior to the highly conserved motifs and inserting one nucleotide after the last sense codon of the uCC (Fig. 2.1A-C). The uCC was fused directly to FLuc, followed by a 22-nt 3'-UTR from the LucPolyA50 vector, a Poly(A)₅₅ tail, and a *Dra*I restriction site for linearization of the plasmid (Fig. 2.2A). In brief, plasmids were linearized and then T7 RNA polymerase was used to transcribe the mRNAs. The mRNAs were capped by incubation with purified recombinant vaccinia virus capping enzyme, and the mRNAs were translated in RRLs and translational activity was monitored by measuring luciferase activity.

3.4.2. Optimization of RRLs for *in vitro* translation

In order to ensure reproducibility, account for batch-to-batch variation in RRLs, and to faithfully monitor regulation of translation, it is critical to ensure that mRNA concentrations and incubation times are in the linear range of response. To determine the

linear range for time and mRNA concentration, a simple FLuc reporter was used (Fig. 2.3). The reporter, under the control of the phage T7 promoter, contained the GGP 5'-UTR from the transcription start site up to the start codon of the uCC fused to the AUG start codon of FLuc. To ensure that FLuc initiation could only occur at the AUG start codon, nucleotide substitutions were used to create stop codons in all three reading frames immediately upstream of the start site (Fig. 2.3). As described above for the uCC-FLuc fusions, the FLuc stop codon was left intact, followed by a 22-nt 3'-UTR from the LucPolyA50 vector, a Poly(A)₅₅ tail, and a DraI restriction site for linearization of the plasmid.

The FLuc mRNA was translated in RRL and aliquots were removed every 15 minutes up to 2 hours and then analyzed for luciferase activity (Fig. 3.9A). FLuc activity was near background levels at 15 minutes but reached a plateau by 1 hour, so 30 min was selected as the optimal incubation time for the *in vitro* translation assays. To determine the linear range for mRNA concentrations in the assay, 25 ng to 100 ng of the same luciferase mRNA preparation was added to RRL and incubated for 30 min before the FLuc activity was measured (Fig. 3.9B). Luciferase activity nearly doubled when the mRNA amount was increased from 25 ng to 50 ng; however, no further increases in luciferase activity were observed when the mRNA level exceeded 50 ng. Therefore, 25 ng was selected as the ideal amount of mRNA for the RRL assays.

While ascorbate enhanced initiation at the ACG start codon of the uCC in the cell-based assays (Fig. 3.6), it was unclear whether this could be observed *in vitro*. Ionic strength, particularly the magnesium concentration, is known to dramatically affect start codon selection stringency in RRL (Kearse and Wilusz, 2017; Kozak, 1989). To optimize the magnesium concentration in the RRL, we generated a set of FLuc reporters with three

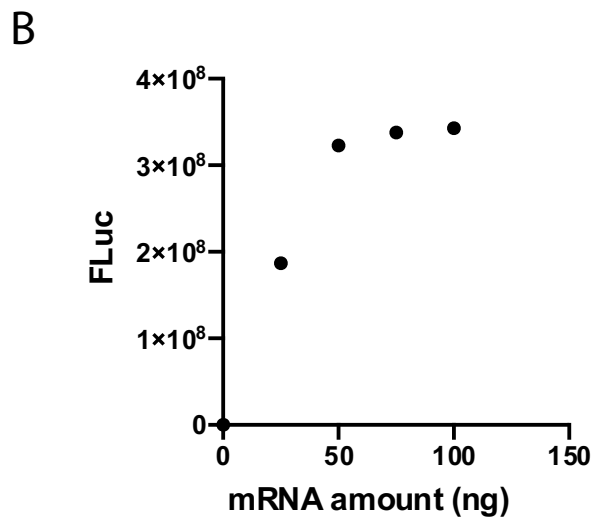
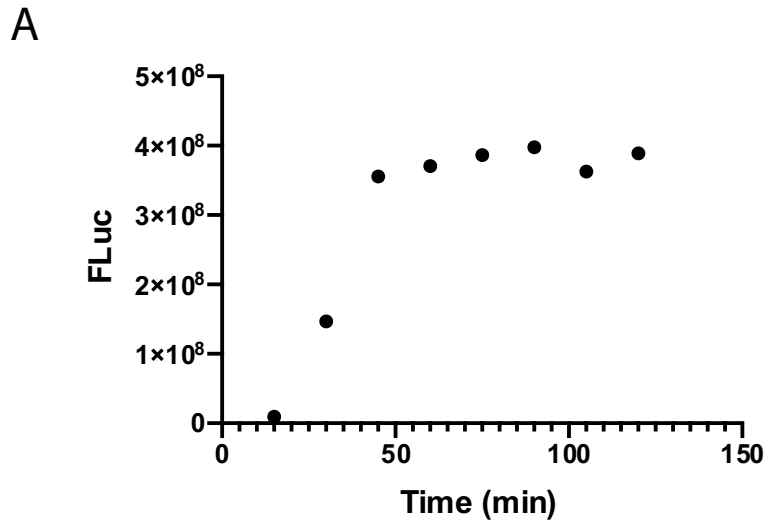


Figure 3.9. Determination of the linear range for incubation time and amount of FLuc mRNA translated in RRL

Figure 3.9. Determination of the linear range for incubation time and amount of FLuc mRNA translated in RRL

(A) 25 ng of FLuc mRNA was translated in RRL, and luciferase activity was measured at increasing time points from 15 min to 120 min. (B) FLuc mRNA amounts from 0 ng to 100 ng were translated in RRL, and luciferase activity was measured at 30 min.

different start codons: AUG, ACG, and AAA (Fig. 2.3). To prevent any luciferase activity due to initiation at potential upstream start sites, stop codons were added in all three frames within the first 14 nucleotides upstream of the start codon (Fig. 2.3). The FLuc reporter mRNAs were transcribed in vitro, capped, and then translated in RRL in the presence of increasing concentrations of magnesium. The magnesium concentration in Promega's Flexi RRL in vitro translation assay system varies from lot to lot. The basal magnesium concentration in the RRLs used for our studies was 1.8 mM. The ACG-, AAA- and AUG-initiated luciferase reporter mRNAs were incubated in the RRLs in the presence of 1.8 mM to 2.2 mM magnesium (increasing by increments of 0.1 mM). Luciferase activities for all three reporters were measured and normalized to the values obtained with the AUG-initiated luciferase reporter (Fig. 3.10). At 1.8 mM magnesium, translation was impaired and the AUG-initiated FLuc activity was low (data not shown); the ACG-initiated luciferase activity was even lower at ~10% of the level obtained with the AUG-initiated reporter (Fig. 3.10). At 2.0 mM magnesium, the yield of AUG-initiated luciferase reached its maximum; however, start codon selection stringency was relaxed and the luciferase activity from the ACG-initiated reporter was now ~15% of the level obtained with the

AUG-initiated reporter. At magnesium concentrations above 2.0 mM, the luciferase activity from the ACG-initiated reporter was over 20% of the value obtained with the wild-type reporter. According to previous studies, an ACG codon is used at ~12% the efficiency of an AUG start codon in mammalian cells (Ivanov et al., 2010b; Kearse and Wilusz, 2017). Based on this information and the results of magnesium titration assay, we determined that 1.9 mM magnesium (Fig. 3.10) was the optimal concentration to obtain acceptable levels of translation with normal start site selection stringency.

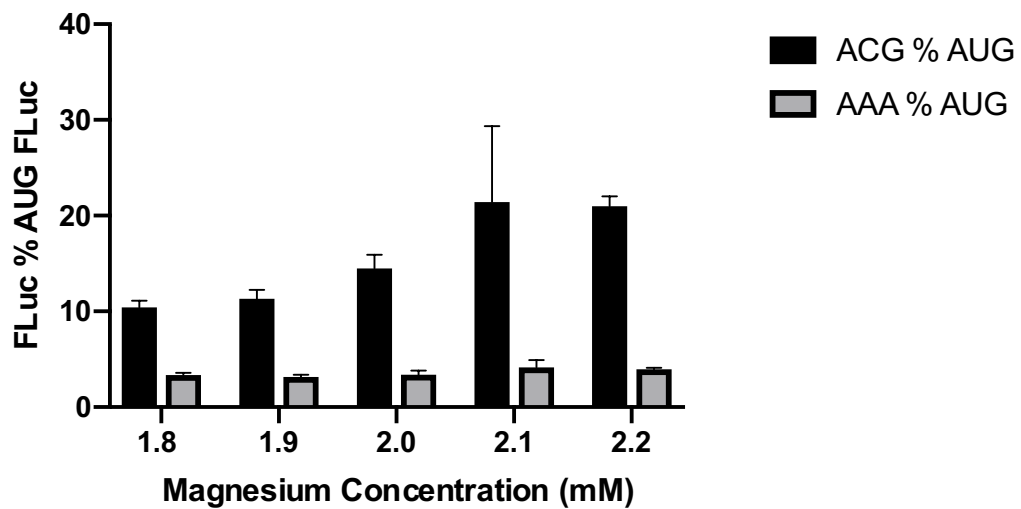


Figure 3.10. Optimization of RRL magnesium concentration

Figure. 3.10. Optimization of RRL magnesium concentration

(A) FLuc transcripts with three start sites (AAA, ACG, or AUG) were translated in RRL with increasing concentrations of magnesium acetate. The luciferase activities from the ACG- and AAA-initiated mRNAs were normalized to the activity from the AUG-initiated transcript (ACG % AUG and AAA % AUG) and then plotted to define the magnesium concentration that optimized translational efficiency and start codon selection stringency (n=3).

3.4.3. Ascorbate regulates uCC translation in rabbit reticulocyte lysate (*in vitro*)

To test whether ascorbate affects uCC translation *in vitro*, mRNAs encoding AUG-initiated wild-type and OF-FLuc fusions (Fig. 3.11A) were translated in RRL in the presence and absence of ascorbate. Adding unbuffered ascorbate to RRL disrupted the pH of the extract and impaired the translational activity. To avoid this complication, the ascorbate was buffered by dissolving ascorbate in 750 mM HEPES. To monitor the impact of ascorbate on GGP uCC translation, increasing concentrations of ascorbate were added to RRLs and then the wild-type or OF-FLuc mRNA was added to the translation mix. Translational activity was determined by measuring FLuc activity (Fig 3.11B). Because the translational efficiency of the *in vitro* transcribed mRNAs varied among different preparations and in different translation reactions, the FLuc activity for each reaction was normalized to the activity observed with the same mRNA in a reaction lacking ascorbate (Fig. 3.11C). With the addition of 5 mM ascorbate, there was a 4-fold decrease in uCC-

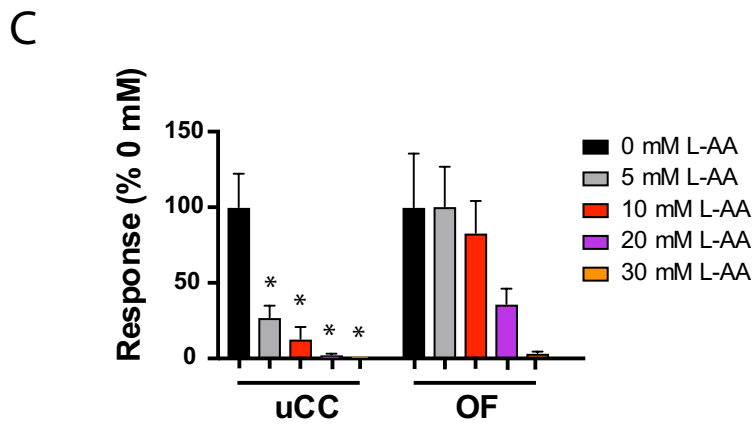
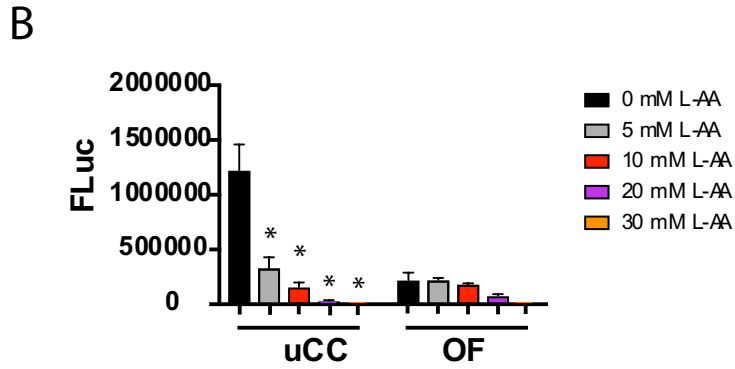
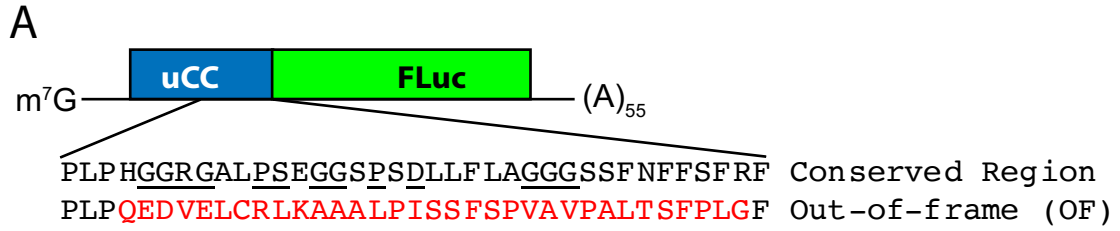


Figure 3.11. Ascorbate regulates uCC elongation in RRL

Figure 3.11. Ascorbate regulates uCC elongation in RRL

(A) AUG-initiated *in vitro* transcribed GGP uCC-FLuc mRNA fusions, or an OF control, were (B) translated in RRL in the presence of increasing concentrations of HEPES-buffered L-ascorbic acid (L-AA). (C) FLuc activity at each L-AA concentration was normalized to the level obtained in the absence of L-AA for each mRNA. Error bars represent propagation of error, * $p < 0.005$ (Student's two-tailed t-test; $n=3$).

FLuc activity (to 27%; $p < 0.005$), but no decrease with the OF-FLuc control. At 10 mM ascorbate, the inhibition of the uCC-FLuc reporter activity was even greater (to 13%); however, translation of the OF-FLuc control was also slightly affected (to 83%). At 20 mM ascorbate, translation of the uCC-FLuc mRNA was barely detectable, while the OF control was reduced by only about 60%. When the ascorbate concentration was increased to 30 mM, the FLuc activity for both the wild-type and OF fusions was severely inhibited. Presumably, at this high concentration of ascorbate, the translational activity of the RRL, the stability of the mRNA reporter, or the luciferase enzymatic activity was severely impaired. Based on these data, we conclude that 5 mM ascorbate is the optimal concentration to monitor uCC-specific inhibition of translation in RRLs.

To confirm that the reduction in FLuc activity upon addition of 5 mM ascorbate is due to a reduction in translation of the uCC-FLuc mRNA and not due to ascorbate affecting FLuc enzymatic activity, we used Western analyses to assess FLuc protein levels (Fig. 3.12). The wild-type and OF-FLuc mRNAs (Fig. 3.12A) were translated in RRL in the presence or absence of 5 mM ascorbate. The reaction products were separated by

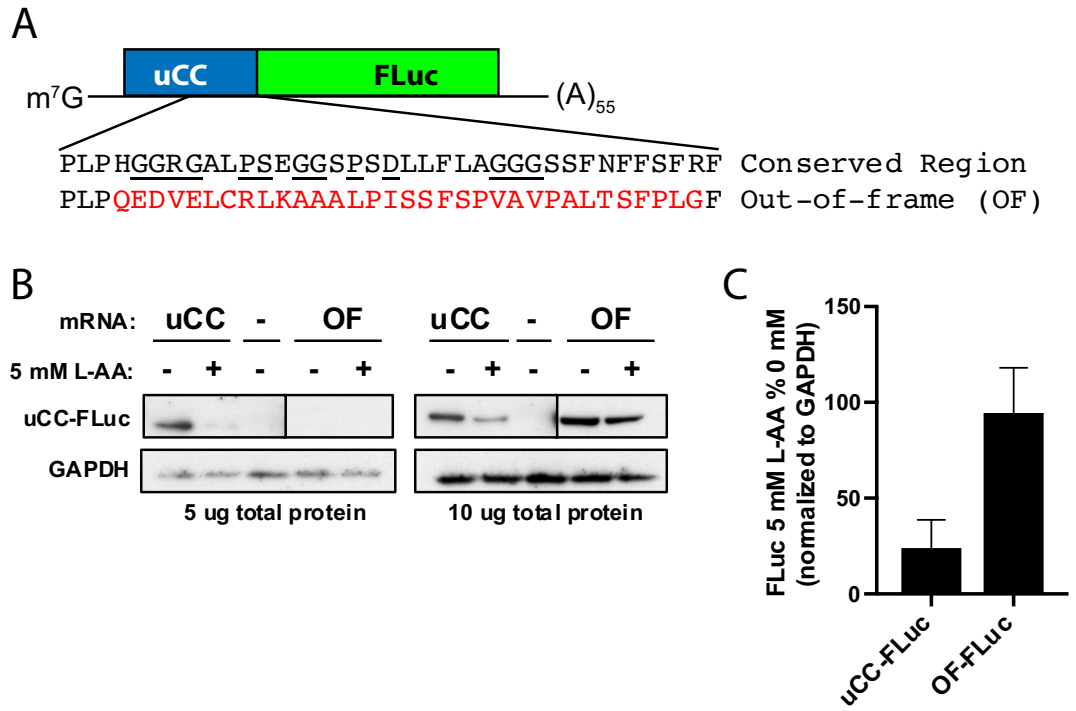


Figure 3.12. The reduction in FLuc levels in the presence of ascorbate is due to a uCC-specific reduction in uCC-FLuc mRNA translation *in vitro*

Figure 3.12. The reduction in FLuc levels in the presence of ascorbate is due to a uCC-specific reduction in uCC-FLuc mRNA translation *in vitro*

In vitro transcribed and AUG-initiated wild-type or OF GGP uCC-FLuc mRNAs (A) were translated in RRL in the absence or presence of 5 mM HEPES buffered L-ascorbic acid (L-AA). (B) Translation products (5 or 10 μ g) were separated on a 4-12% Bis-Tris NuPAGE gel, and then transferred to a nitrocellulose membrane. The membrane was then probed with anti-FLuc and anti-GAPDH antibodies. The blot is representative of three independent experiments. (C) Band intensity on the Western blot was quantified using ImageJ software and the reduction in total FLuc protein in the presence of L-AA was calculated by first normalizing to GAPDH and then normalizing again to FLuc levels obtained in the absence of L-AA. Error bars represent SD, $p=0.01$ (Student's two-tailed t-test; $n=3$).

electrophoresis on a NuPAGE gel, transferred to a nitrocellulose membrane, and the membrane was then probed with antibodies that detect FLuc and GAPDH (Fig. 3.12B).

Addition of ascorbate dramatically decreased the yield of FLuc protein from the wild-type, but not the OF reporter (Fig. 3.12B). To quantify the changes in FLuc protein levels, the Western signals for FLuc were normalized to the GAPDH signals to correct for loading differences between lanes. Next, the corrected FLuc levels from the reactions containing 5 mM ascorbate were normalized to the FLuc levels in the reactions lacking ascorbate. As shown in Fig. 3.12C, addition of ascorbate significantly lowered uCC-FLuc synthesis to around 25% of the level obtained with the OF-FLuc reporter ($p=0.01$). These results are consistent with the luciferase reporter assays and indicate that the reduction in

FLuc activity in the in vitro translation assays upon addition of ascorbate is due to a uCC-specific inhibition of translation.

3.4.4. The uCC motifs are essential for ascorbate inhibition of uCC-FLuc translation

In order to examine the role of the conserved uCC motifs in a potential ascorbate induced elongation pause, each of the five motifs was disrupted using alanine substitutions (Fig. 3.13A). As described above for the mammalian cell reporter assays, the GGRG motif was modified to four alanines and the perfectly conserved Arg34 residue was also individually mutated to alanine. The conserved PS motif was replaced by two alanine residues, as was the GG motif, and three alanine residues were substituted for the PSD motif in the middle of the uCC conserved region. The block of the uCC beginning with the PS motif and ending with the PSD motif was also mutated to a block of nine alanine residues. Finally, the C-terminal glycine repeat was replaced by three alanine residues. Like the set of alanine substitutions used for the mammalian cell reporter assays, an S11A substitution was included as a control. As Ser11 is upstream of the GGP uCC conserved region, we did not expect an alanine substitution at this position to affect ascorbate regulation of uCC-FLuc translation.

The mRNAs containing the alanine substitution mutations, along with mRNAs containing the wild-type uCC and the OF control, were translated in RRL in the presence or absence of 5 mM ascorbate. Total FLuc activity varied for each reporter, but only the wild-type and S11A mutant uCC-FLuc fusions showed a specific reduction in FLuc activity in the presence of 5 mM ascorbate (Fig. 3.13B). Like the OF control, every substitution targeting a conserved motif resulted in a loss of sensitivity to ascorbate. To

account for differences in translation efficiency between the reporters even in the absence of ascorbate, the FLuc activity from each reporter obtained from reactions performed in the presence of 5 mM ascorbate was normalized to the activity obtained from the same reporter in the absence of ascorbate (Fig. 3.13C). Mirroring the *in vivo* results, only the wild-type and S11A mutant uCC-FLuc fusions showed a significant reduction in FLuc activity in the presence of ascorbate ($p < 0.005$). Alanine substitutions at any of the conserved motifs abolished sensitivity to ascorbate to the same extent as the OF control. These results are nearly identical to the results obtained when the same substitutions were made in the reporters for mammalian cell experiments. Taken together, these results further support our hypothesis that all of the conserved motifs play a role in reducing uCC-FLuc translation in the presence of ascorbate. We hypothesize that the reduction in uCC-FLuc activity is due to an elongation pause during translation of the conserved region in the presence of ascorbate.

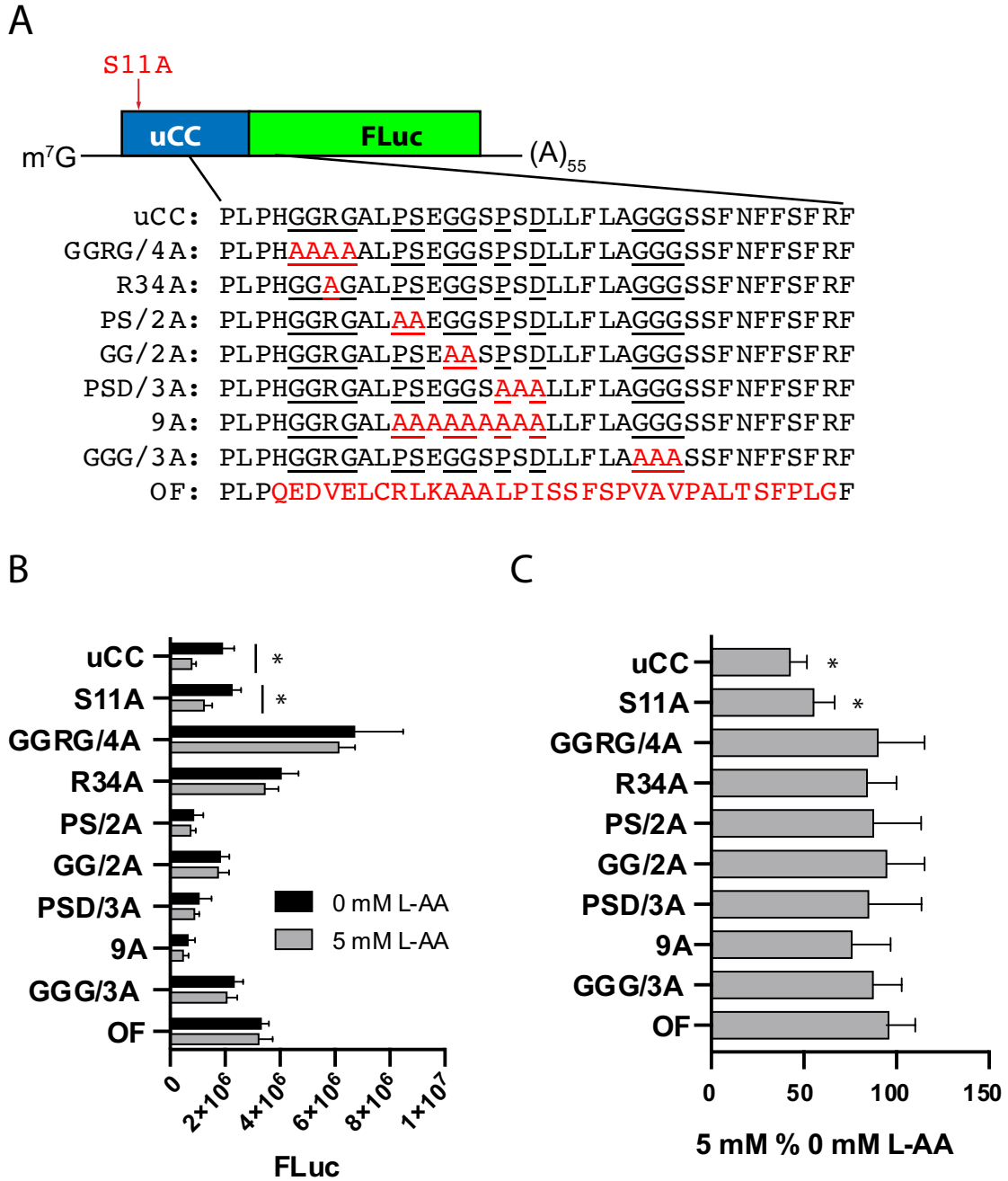


Figure 3.13. Conservation of the uCC motifs is essential for ascorbate control of GGP uCC elongation *in vitro*

Figure 3.13. Conservation of the uCC motifs is essential for ascorbate control of GGP uCC elongation *in vitro*

AUG-initiated *in vitro* transcribed mRNAs with alanine substitutions of conserved motifs (A) were translated in RRL in the absence or presence of 5 mM HEPES-buffered L-ascorbic acid (L-AA). (B) FLuc activity was measured for each sample. (C) Because FLuc activity varied by mRNA, the response to L-AA was calculated for each mRNA by normalizing the activity obtained in the presence of L-AA to the activity obtained in the absence of L-AA (5 mM % 0 mM L-AA). Significance [* $p < 0.005$ (Student's two-tailed t-test; $n=3$)] depicts the difference in FLuc activity at 5 mM compared to 0 mM L-AA; error bars represent propagation of error.

3.4.5. uCC-FLuc mRNA translation is inhibited specifically by L-ascorbate

In order to examine the specificity of L-ascorbate for translational control of the GGP uCC, we tested whether four chemicals related to L-ascorbate could also regulate uCC translation (Fig. 3.14A-E). The chemicals tested included the D-isoform of L-ascorbic acid, D- isoascorbic acid (D-iso), differing in structure from L-ascorbic acid at one chiral center. We also tested dehydroascorbic acid (DHAA), the oxidized form of L-ascorbic acid, which is present in cells and likely in the translation reaction mix because L-ascorbic acid spontaneously oxidizes to DHAA. In addition, we tested L-Galactono-1,1-lactone (Gal.), the precursor one step upstream of L-ascorbic acid in the plant ascorbate biosynthetic pathway, and L-Gulono-1,4-lactone (Gul.), the precursor one step upstream of L-ascorbic

acid in the animal ascorbate biosynthetic pathway. L-ascorbate is the oxidized product of these two precursors that differ from each other at one chiral center.

To test whether these structurally similar chemicals could also regulate uCC translation, the wild-type and OF uCC-FLuc mRNAs (Fig. 3.15A) were translated in RRL in the presence or absence of 5 mM ascorbate or the related chemicals. As shown in Fig. 3.15B, although FLuc activity was lower for the wild-type versus the OF uCC-FLuc reporter even in the presence of just the HEPES buffer, both chemical- and reporter-specific effects were visible before normalization of the data. Addition of ascorbate reduced uCC-FLuc activity from the wild-type reporter by about 3-fold, while uCC-FLuc activity from the OF reporter was reduced only about 1.5-fold. D-isoascorbic acid had a non-specific effect and reduced both wild-type and OF uCC-FLuc translation about 1.5-fold. The addition of DHAA severely impaired translation in the RRL and almost no measurable FLuc activity was observed with either mRNA upon addition of 5 mM DHAA. Unlike DHAA, the addition of Gal. or Gul. had little effect on translation of either reporter. For both the wild-type and OF uCC-FLuc reporters, the FLuc activity obtained in reactions containing each chemical was normalized to the FLuc activity obtained in reactions containing an equal volume of HEPES buffer. As shown in Fig. 3.15C, the wild-type uCC fusion was twice as sensitive to 5 mM ascorbate as the OF control ($p < 0.005$). In contrast, no uCC-specific decrease in FLuc activity was observed in response to 5 mM D-iso, Gal., or Gul. These data show that regulation of GGP uCC translation is highly specific to ascorbic acid, and that the sensing mechanism can distinguish between the L and D isoforms as well as the oxidative state of the molecule.

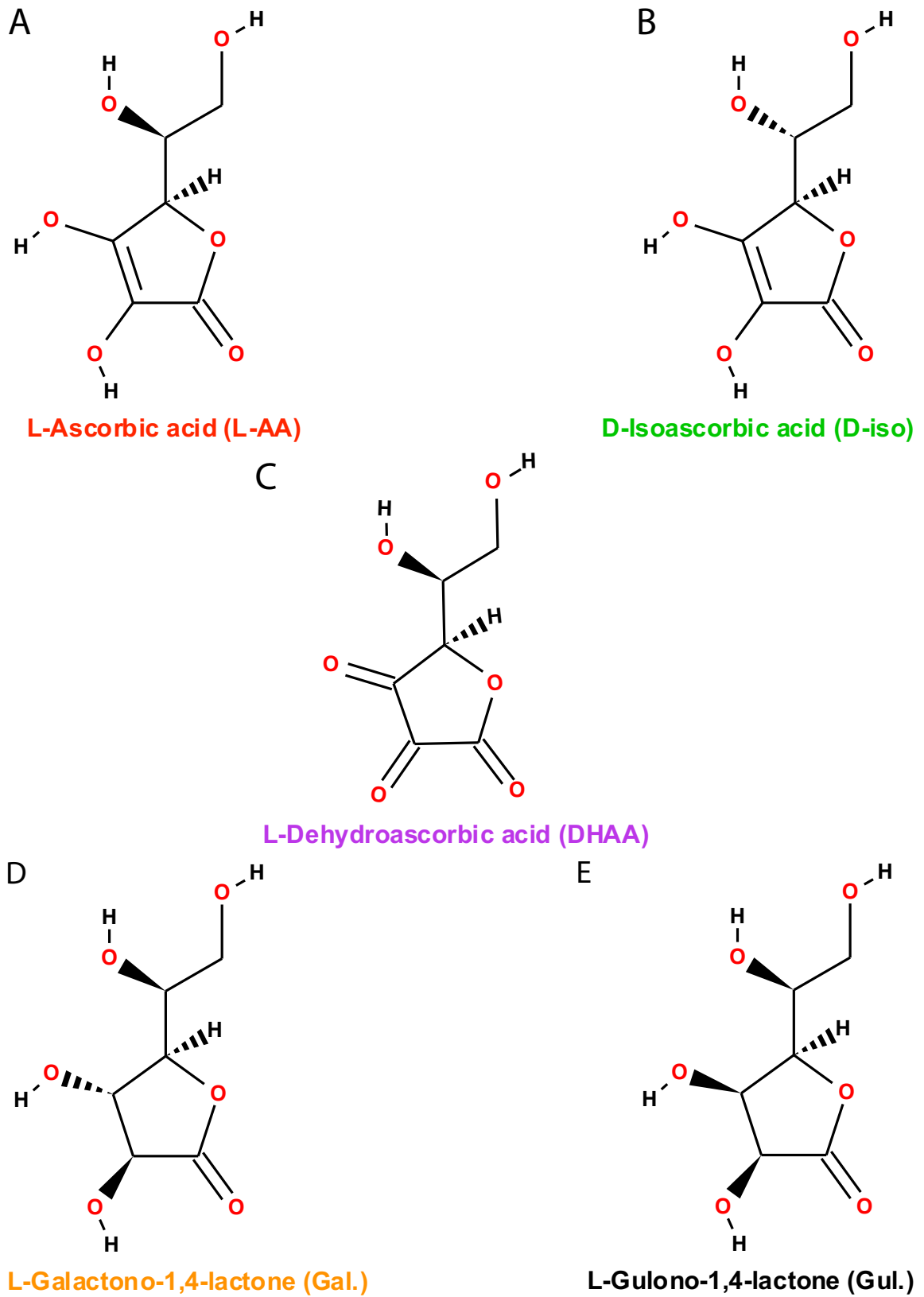


Figure 3.14. Chemical structures of L-ascorbic acid and chemical analogs

Figure 3.14. Chemical structures of L-ascorbic acid and chemical analogs

Structures of (A) L-ascorbic acid (L-AA, red label), (B) D-isoascorbic acid (D-iso, green label), (C) L-Dehydroascorbic acid (DHAA, purple label), (D) L-Galactono-1,4-lactone (Gal., orange label), and (E) L-Gulono-1,4-lactone (Gul., black label).

FLuc activity obtained in reactions containing each chemical was normalized to the FLuc activity obtained in reactions containing an equal volume of HEPES buffer for each reporter (Fig. 3.15C). Although ascorbate reduced OF-FLuc activity as well as uCC-FLuc activity, after normalization the uCC fusion was twice as sensitive to 5 mM ascorbate as the OF control ($p < 0.005$). There was no uCC specific decrease in FLuc activity in response to 5 mM D-iso, Gal., or Gul, but D-iso appeared to reduce translation to about the same extent as ascorbate (note similar decreases in FLuc yield for both the wild-type and OF-FLuc reporter mRNAs). Gal. and Gul. only modestly impaired translation of both the wild-type and OF reporters. These data show that regulation of GGP uCC translation is highly specific to ascorbic acid and can distinguish between the L and D isoforms of the molecule. The regulatory mechanism can also distinguish between the oxidative states of the chemicals.

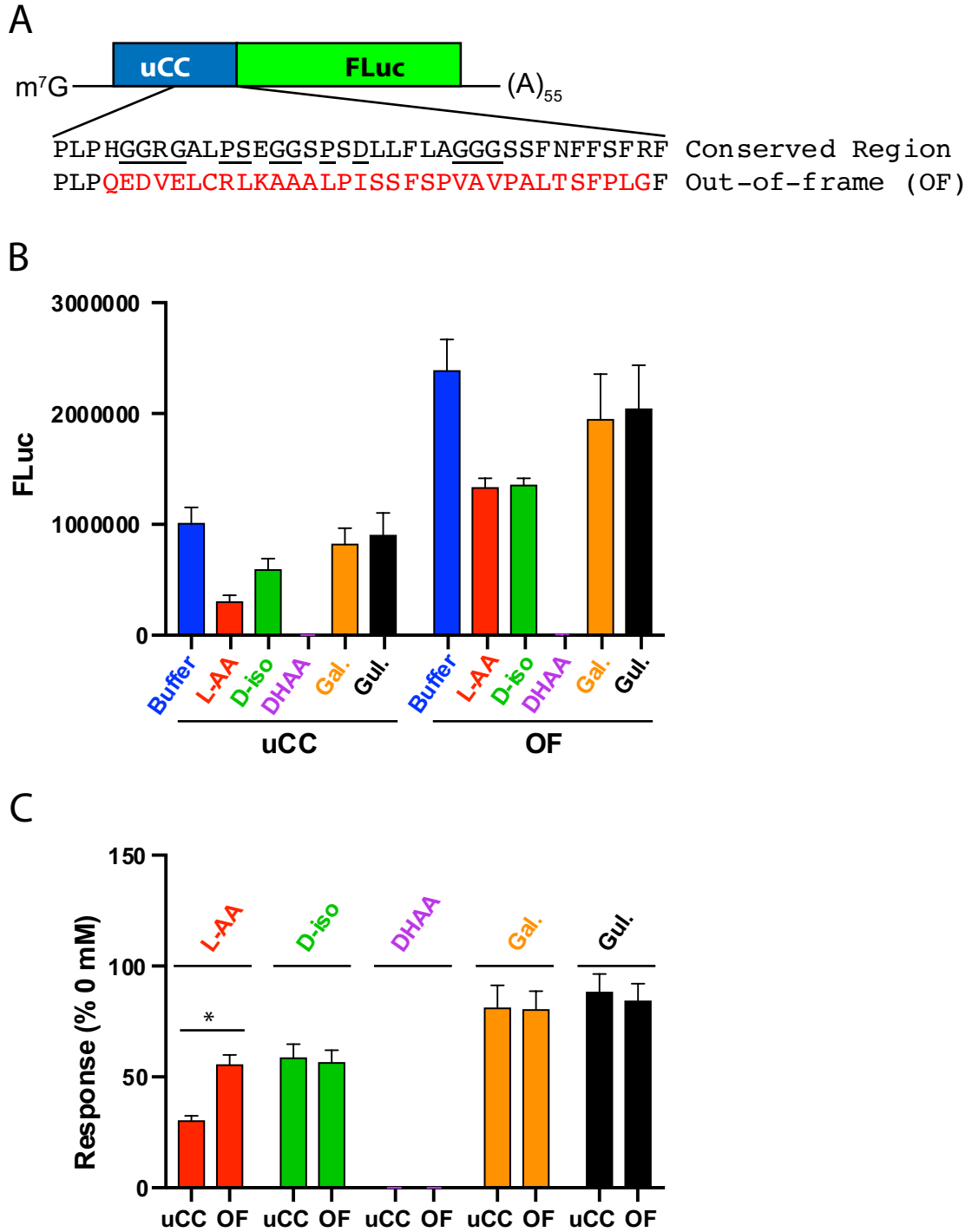


Figure 3.15. The uCC-specific reduction in uCC-FLuc translation is unique to L-ascorbic acid

Figure 3.15. The uCC-specific reduction in uCC-FLuc translation is unique to L-ascorbic acid

AUG-initiated *in vitro* transcribed wild-type or OF GGP uCC-FLuc mRNA fusions (A) were translated in RRL with the following additions: 5 mM HEPES (buffer, blue) or 5 mM HEPES-buffered L-AA (red), D-iso (green), DHAA (purple), Gal. (orange), or Gul. (black). (B) FLuc activity levels obtained following translation of each mRNA following addition of 5 mM of each chemical. (C) The response to each chemical was calculated by normalizing the FLuc activity obtained in the presence of each chemical versus the buffer control (Response (% 0 mM)). Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; n=3).

3.5. Ascorbate regulation of the uCC is mediated by ribosome pausing during translation elongation

3.5.1. Significance and rationale for study

In the uCC model developed from the study of *AZINI* mRNA translational control by polyamines (Ivanov et al., 2018) and *arg-2* mRNA translational control by arginine (Freitag et al., 1996; Wang et al., 1998; Wei et al., 2012), inhibition of translation elongation on the uCC is mediated in part by conserved residues in the uCC and by components of the elongating ribosome, either through inhibition of an elongation factor or the PTC by the metabolite. The results of the mammalian cell and *in vitro* RRL assays reported here suggest that ascorbate impairs translation elongation in a GGP uCC-specific manner. We performed *in vitro* and *in vivo* assays to test the hypothesis that translating ribosomes stall while translating the GGP uCC in the presence of ascorbate.

3.5.2. uCC regulation is due to ribosome pausing during translation

During translation elongation, the nascent chain is attached to the peptidyl-tRNA until a stop codon enters the A site and termination factors release the completed polypeptide. If translation stalls during elongation, the nascent chain will remain attached to the peptidyl-tRNA. *In vitro* assays examining the production of completed polypeptides and peptidyl-tRNA can be used to assess stalling during translation elongation. Translation reaction products can be resolved by SDS-PAGE and then subjected to Western analysis to detect the translated peptide. The presence of a tRNA adds approximately 25 kDa to the mass of the nascent chain. The presence of peptidyl-tRNA products in the reaction can be assessed by RNase treatment. The tRNA is susceptible to nuclease digestion and will be removed from the nascent chain with the addition of RNaseA. The molecular weight of a

peptidyl-tRNA product decrease by about 25 kDa following treatment with RNase. If the elongation stall occurs close to the stop codon, the nascent chain released by RNase digestion will co-migrate with the protein produced by authentic translation termination. Comparison of the levels of peptidyl-tRNA and free peptide in assays not treated with RNase can provide an estimate of the severity of the elongation defect. Because the aminoacyl bond linking the nascent chain to the tRNA is susceptible to high pH, it is essential that the sample buffer and SDS-PAGE be at neutral or slightly acidic pH.

Analysis of peptidyl-tRNA levels has previously been used to demonstrate translational stalling during termination at a uORF on the human cytomegalovirus (hCMV) UL4 gene encoding the glycopeptide gp48. Before adapting this assay for analysis of GGP uCC translation, we examined stalling on the UL4 uORF as a proof of concept. The 5' leader of the hCMV UL4 mRNA contains three AUG-initiated uORFs (Schleiss et al., 1991). The uORF2, consisting of 22 codons, represses gp48 synthesis during the early stages of hCMV infection (Janzen et al., 2002; Schleiss et al., 1991). The repressive nature of uORF2 is due to a C-terminal PP-stop motif that impairs translation termination, leading to persistence of the uORF2 peptidyl-tRNA linkage with the peptidyl-tRNA^{Pro} in the P site and the termination codon in the A site (Bhushan et al., 2010; Degnin et al., 1993; Janzen et al., 2002). In order to monitor uORF2 translation in vitro, a FLAG epitope was inserted at the N-terminus of the uORF (immediately downstream of the AUG start codon). It is critical to tag the uCC at the N- versus the C-terminus to ensure that the epitope would be synthesized regardless of any elongation defect (Fig. 3.16A). An 130-residue linker peptide derived from the human chorionic gonadotropin beta protein was added between the FLAG epitope and the uORF to increase the molecular weight of the peptide and to ensure that

the FLAG-linker-uORF product would be well-resolved on SDS-PAGE from the abundant globin proteins present in the RRL. To conduct the assay, the mRNA encoding the uORF2 fusion protein was translated in RRL in the presence or absence of 5 mM ascorbate. Following translation, the reaction was split into two fractions: one was treated with RNaseA to degrade the tRNA and release the nascent chain; while the second was treated with water to leave the peptidyl-tRNA intact. Following PAGE, Western analyses using anti-FLAG antibodies were used to monitor the levels of the peptidyl-tRNA and the released nascent chain peptide.

As shown in Fig. 3.16B, two prominent products were detected when the FLAG-linker-uORF2 construct was translated in RRL: an ~39 kDa product and a 25 kDa product corresponding to the predicted mass of the FLAG-linker-uORF2 peptide. Treatment of the sample with RNase prior to PAGE results in loss of the 39 kDa product and increased amounts of the 25 kDa protein. This behavior is consistent with the notion that the 39 kDa product is peptidyl-tRNA stalled at the stop codon and that treatment with RNase releases the full-length peptide. As expected, addition of 5 mM ascorbate to the translation reactions did not affect the levels of the peptidyl-tRNA or released peptide (Fig. 3.16B).

To make a detectable peptidyl-tRNA reporter for the GGP uCC, the same FLAG epitope and human chorionic gonadotropin beta peptide linker used in the gp48 uORF2 experiments was added at the N-terminus of the uCC (immediately downstream of the start codon) (Fig. 3.16C). As a control, corresponding constructs were generated using the uCC with the same OF mutation as described for the mammalian and *in vitro* translation reporters. In addition to these FLAG-linker-uCC constructs, two additional constructs were



B

L-AA:	0 mM	5 mM		
mRNA:	uORF2		—	—
RNaseA:	—	+	—	+

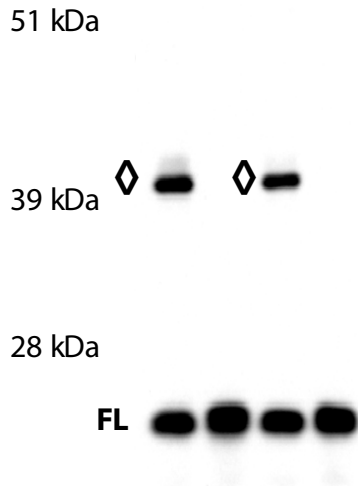


Figure 3.16. uCC regulation is due to ribosome pausing during translation

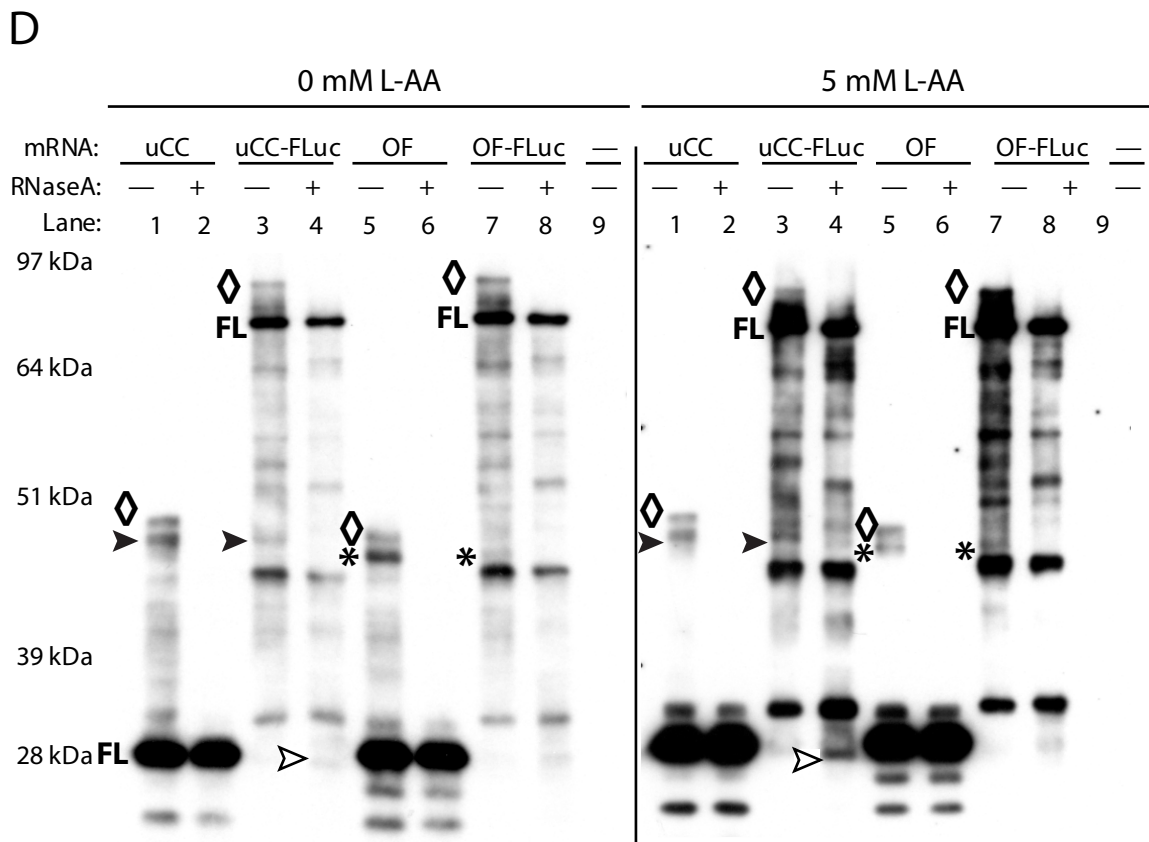
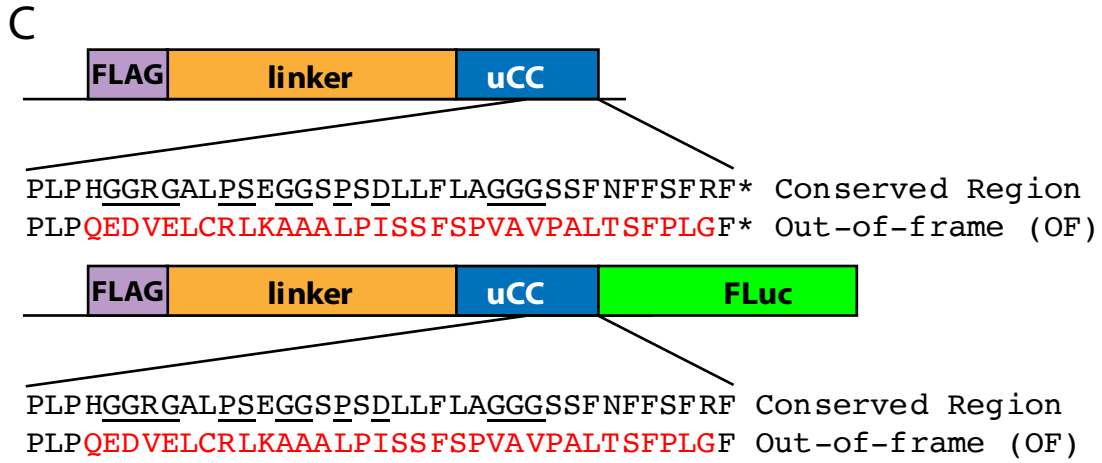


Figure 3.16. uCC regulation is due to ribosome pausing during translation

Figure 3.16. uCC regulation is due to ribosome pausing during translation

mRNAs encoding AUG-initiated FLAG-linker-gp48 uORF2 fusions (A) were transcribed *in vitro* and then (B) translated in RRL in the absence or presence of 5 mM HEPES-buffered L-AA. Following translation, the reactions were split into two equal halves, and one half was treated with RNaseA to digest the tRNA. The samples were prepared with Laemmli buffer, run on a 10% Bis-Tris NuPAGE gel, and then immunoblotted using anti-FLAG antibodies. FL = Full-length released peptide, diamond = full-length peptidyl-tRNA (due to stall at the stop codon). (C) Schematics of wild-type and OF AUG-initiated FLAG-linker-uCC and FLAG-linker-uCC-FLuc fusions. (D) The *in vitro* transcribed mRNAs were translated in RRL in the absence or presence of 5 mM HEPES-buffered L-AA. Following translation, the reactions were split, and one half was treated with RNaseA to digest the tRNA and release the peptide. The samples were prepared with Laemmli buffer, run on a 10% Bis-Tris NuPAGE gel, and then immunoblotted using anti-FLAG antibodies. FL= Full-length released peptide; closed arrow = peptidyl-tRNA from stalled elongation complex; open arrow = released peptide after RNase digestion of stalled translation products; diamond = peptidyl-tRNA from stalled termination complex; star = unidentified product due to elongation stall in the OF control.

made by inserting the FLAG-linker element immediately after the AUG start codon of the wild-type and OF uCC-FLuc fusions. These latter constructs were made because, in our preliminary experiments, we found that translation termination was inefficient in the RRL, and it was difficult to resolve FLAG-linker-uCC products stalled during elongation from

those stalled at termination. With the FLAG-linker-uCC-FLuc fusions, the products stalled at termination are approximately 50 kDa larger than the products stalled during elongation. The translation assays with the FLAG-linker-uCC constructs were performed as described above (Fig. 3.16B) for FLAG-linker-uORF2 experiments (Fig. 3.16B).

Translation of the FLAG-linker-uCC and FLAG-linker-uCC-FLuc mRNAs generated multiple products. As shown in Fig. 3.16D, translation of the FLAG-linker-uCC mRNA (lane 1) yielded three prominent products. The highest molecular weight (MW) product (marked with a diamond) corresponds to peptidyl-tRNA from a complex paused at the uCC stop codon. This product was not observed during synthesis of the FLAG-linker-uCC-FLuc fusion (lane 3); however, it was observed with the FLAG-linker-OF-uCC fusion (lane 5). These results are consistent with the notion that termination is slow in the RRL and that the largest product in lanes 1 and 5 is due to slow termination when translating the FLAG-linker-uCC (wild-type and OF) constructs. Because the termination site is moved far downstream in the FLAG-linker-uCC-FLuc construct, the corresponding termination product is much larger in lane 3 (diamond). As RNase treatment eliminates these largest products for both the uCC and uCC-FLuc fusions (lanes 2 and 4), we propose that slow termination results in accumulation of peptidyl-tRNA containing the full-length nascent chain.

The second largest product observed upon translation of the FLAG-linker-uCC fusion is likely the result of an elongation stall during translation of the uCC (Fig. 3.16D, lane 1, marked with filled arrow). A product of the same size was observed when translating the FLAG-linker-uCC-FLuc fusion (lane 3), but not when translating the OF reporters (lanes 5 and 7). Consistent with the notion that this product is peptidyl-tRNA

from a stalled elongation complex, treatment with RNaseA eliminated this product (lanes 2 and 4, black arrow). Notably, RNase treatment of reactions translating the FLAG-linker-uCC-FLuc fusion not only eliminated the peptidyl-tRNA stall product (lane 4, black arrow), but generated a novel peptide (lane 4, open arrow) that migrated slightly faster than the full-length Flag-linker-uCC peptide. We propose that this product is the nascent chain released from the stalled elongation complex, suggesting that the elongation stall is near the C-terminus of the uCC. As RNase treatment of the reactions translating the wild-type and OF FLAG-linker-uCC-FLuc fusions did not eliminate all of the products between the size of the FL FLAG-linker-uCC and FLAG-linker-uCC-FLuc peptides (lanes 4 and 8), we propose that ribosomes stall at other non-specific sites when synthesizing the FLAG-linker-FLuc fusions or that the proteins are subject to proteolysis in the lysates.

Translation of the OF constructs (lanes 5 and 7) produced a peptidyl-tRNA product (marked by asterisk) that migrated close to the size of the stalled FLAG-linker-uCC products (black arrow in lanes 1 and 3). As these products do not perfectly co-migrate on PAGE, we propose that it may reflect pausing during elongation, perhaps on difficult to translate sequences like proline residues. Importantly, the proportion of this unknown product (asterisk) in the reactions translating the OF control mRNAs decreases upon addition of ascorbate, indicating that the pause is not induced by ascorbate. As the yield of all products was higher in the reactions containing 5 mM ascorbate, and the ratio of the yields of the termination stall (diamond) and elongation stall (black arrow) products for the FLAG-linker-uCC fusions (lane 1) were similar in the reactions containing and lacking ascorbate, we cannot confirm that ascorbate triggered more stalling during translation of the uCC.

3.5.3. Ribosome profiling in plants reveals ribosome pausing on the conserved C-terminal glycine repeat of the uCC

Ribosome profiling provides a snapshot of all the translating ribosomes in a cell at a particular moment and provides details about the relative levels of translation of particular mRNAs and the utilization of specific initiation sites. Ribosome profiling is also used to infer elongation rates and identify translation pause sites within the transcriptome. Previously published ribosome profiling data from *Arabidopsis* clearly shows a high ribosome density on the GGP uCC (Laing et al., 2015). Using previously published ribosome protection data (Juntawong et al., 2014; Liu et al., 2013), Laing et al. (2015) analyzed the distribution of ribosomes along the *GGP* sequence. Their analysis revealed that the uCC has a higher ribosome occupancy and density than the mORF.

Our collaborator Ying Wen Betty Chung at Cambridge University has been using ribosome profiling to study translation in plants. Her work has focused on the single-celled green algae *Chlamydomonas reinhardtii* and on the tomato *Solanum lycopersicum*. The *C. reinhardtii* data was previously published (Chung et al., 2017) using cells grown in Tris-acetate-phosphate medium under constant illumination with white light as previously described (Chung et al., 2015). For *S. lycopersicum*, the data were generated from seeds vernalized at 4 °C in the dark for two weeks. Seedlings were germinated and grown under constant light at 23 °C. The two-week old plants were snap frozen with liquid nitrogen prior to cryogenic grinding with lysis buffer (20 mM Tris–Cl pH7.5, 140 mM KCl, 5 mM MgCl₂, 100 µg/mL cycloheximide, 100 µg/mL chloramphenicol, 0.05 mM DTT, 0.1% NP40 and 5% sucrose).

Although the ascorbate concentration in the plant tissues used for generation of the ribosome profiling libraries is unknown, ribosome footprints were readily detected on both the uCC and mORFs of the GGP mRNA from *C. reinhardtii* and *S. lycopersicum* (Fig. 3.17A, B). The ribosome footprint reads were higher in the uCC than in the mORF for both species, but especially for *S. lycopersicum*. A very clear stall was observed with the final glycine codon of the C-terminal tri-glycine repeat in the P site for both uCCs (lower panels in Figs. 3.17A-B). In addition, in the *C. reinhardtii* profile a second peak is observed 30-nt or 10 codons upstream of the glycine residue (Fig. 3.17A). As a single ribosome protects 30-nts, the distance between the two peaks is indicative of 80S ribosome queuing with the progress of the upstream ribosome blocked upon collision with the ribosome paused on the Gly residue. These data support a model where translating ribosomes stall during translation of the final conserved motif leading to a queue of subsequent scanning and translating ribosomes.

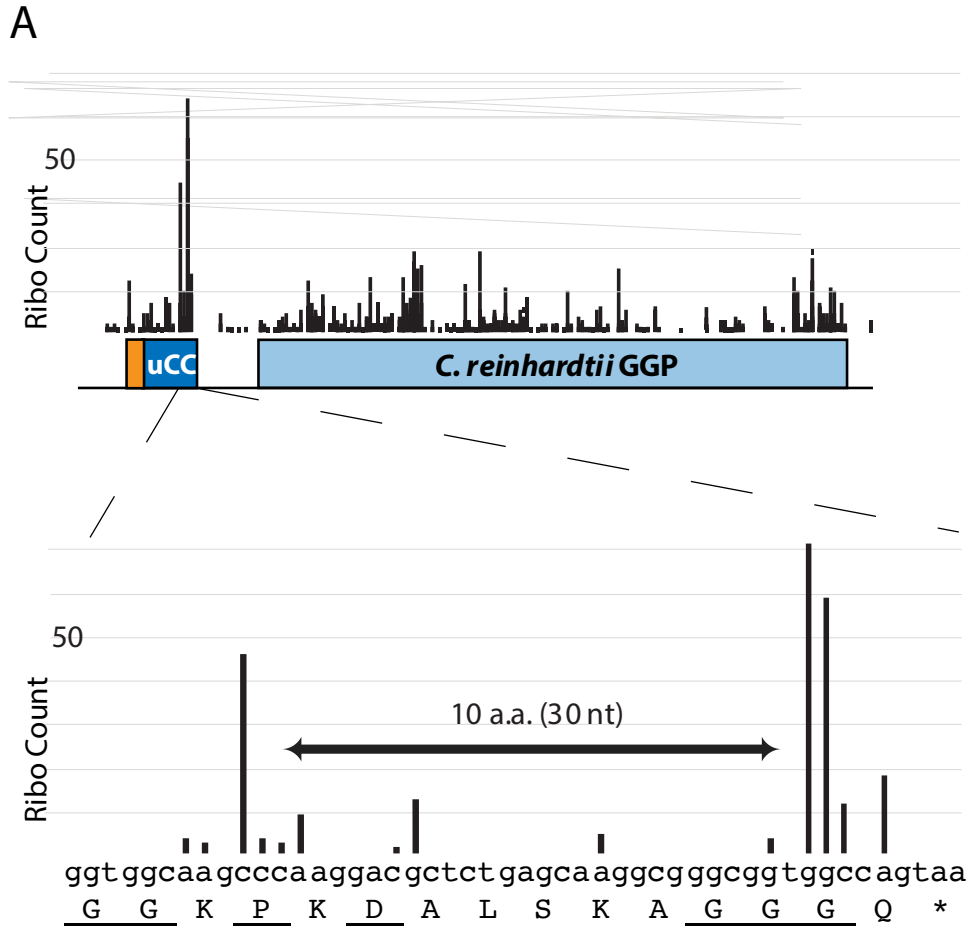


Figure 3.17. Ribosome profiling reveals a ribosome stall at the C-terminal glycine of the last conserved motif in the GGP uCC

B

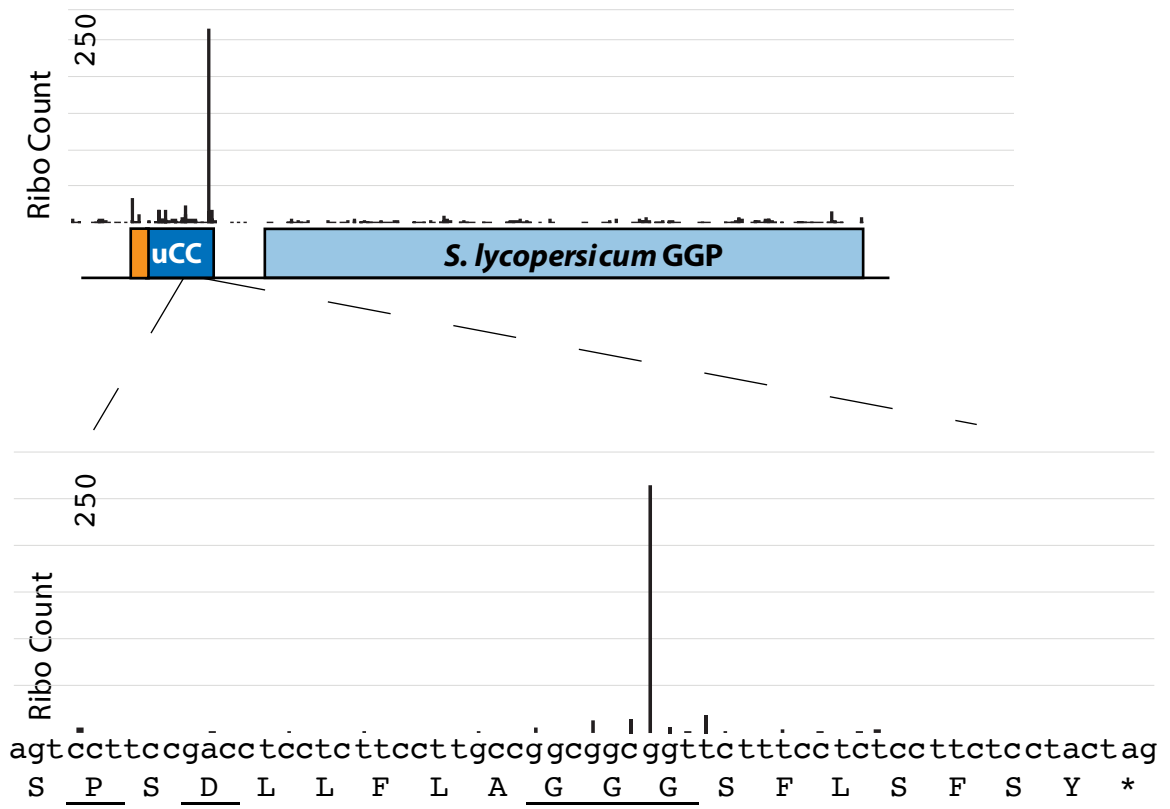


Figure 3.17. Ribosome profiling reveals a ribosome stall at the C-terminal glycine of the last conserved motif in the GGP uCC

Figure 3.17. Ribosome profiling reveals a ribosome stall at the C-terminal glycine residue of the last conserved motif in the GGP uCC

(A) Ribosome-protected fragments from *C. reinhardtii* and (B) *S. lycopersicum* germinated and grown under constant light for 2 weeks were mapped to the *GGP* mRNA, with the uCC (dark blue) and mORF (light blue) depicted as rectangles in the schematic. The data were provided by Ying Wen Betty Chung at Cambridge University. The zoomed images depict the ribosome footprints over the C-terminal conserved motifs (underlined) in the GGP uCC.

3.6. Ribosome queuing on the uCC is essential for enhancement of initiation at the uCC start site

3.6.1. Significance and rationale for study

The third feature of the uCC model, and the hardest to show, is that ribosome queuing is responsible for enhanced initiation at the weak start site of the uCC. In the uCC model, under the proper conditions (+metabolite) conserved amino acid motifs within the uCC trigger an elongation pause. The elongation pause leads to a queue of scanning and translating ribosomes. The presence of a ribosome queue causes scanning ribosomes to spend more time in the vicinity of the weak start site, leading to increased initiation and translation of the uCC and, in turn, decreased translation of the mORF. Ribosome queuing cannot reliably be seen using techniques such as ribosome profiling because the queue is likely a randomized mix of scanning 40S and translating 80S ribosomes, with only the latter detected by standard ribosome profiling approaches. Instead of trying to observe queuing directly, a simpler approach is to disrupt queuing or its impact using multiple methods and observing whether this leads to a loss of uCC-specific regulation.

There are multiple ways to interfere with queuing *in vivo*. One method is to reduce ribosome loading on the mRNA. The drug 4EGI-1 blocks the association of the mRNA cap-binding factor eIF4E with the scaffolding protein eIF4G within the eIF4F complex (Moerke et al., 2007; Sekiyama et al., 2015). As eIF4F is responsible for binding the ribosome near the cap of an mRNA, disruption of the eIF4F complex by 4EGI-1 impairs ribosome loading onto mRNAs and thereby inhibits general translation. As proof of principle, in their studies on the *AZINI* uCC, Ivanov et al. (2018) found that 4EGI-1, at concentrations that reduce general translation by 80-95%, impaired uCC translation and

derepressed *AZINI* expression in the presence of polyamines. To examine whether 4EGI-1 has the same effect on the GGP uCC, we examined the impact of 4EGI-1 on ascorbate induction of GGP uCC translation. Unfortunately, inconsistent results in these assays, including with the transfection control reporter, prevented the use of 4EGI-1 to test whether ribosome queuing was important for translational control of the GGP uCC (data not shown).

In an alternate approach to test the importance of ribosome queuing to enhance uCC translation, Ivanov et al. (2018) extended the length of the *AZINI* uCC. Based on the assumption that the length of the uCC has been optimized to the duration of the elongation pause to allow ribosome queuing back to the uCC start codon, extending the uCC would prevent the ribosome queue from reaching the weak initiation site and thus would impair uCC translation. In their studies on the *AZINI* uCC, Ivanov et al. (2018) used a segment from the mouse *Trip12* gene that lacks any in-frame near-cognate start codons over a span of 65 codons. Inserting the *Trip12* sequence immediately downstream of the AUU start codon of the *AZINI* uCC resulted in complete loss of regulation in response to spermidine (Ivanov et al., 2018). A potential complication of this approach is if the *Trip12* extension interferes with metabolite sensing by the uCC rather than simply affecting the length of the uCC. In light of this possible complication, a third test of the queuing model will be to reduce ribosome loading by inserting a stem-loop structure near the 5' end of the mRNA. Secondary structure near the cap of an mRNA impairs ribosome binding and inhibits translation (Kozak, 1986a), and stem-loop (SL) structures formed by palindromic sequences with a high free energy have been shown to impair ribosome loading and scanning (Babendure et al., 2006; Pelletier and Sonenberg, 1985).

3.6.2. Lengthening the GGP uCC impairs ascorbate regulation in mammalian cell culture (*in vivo*); however, the extension may interfere with ascorbate sensing

To test whether lengthening the GGP uCC interferes with regulation in response to ascorbate, I doubled the uCC length by adding a 65-codon extension from the mouse *Trip12* mRNA immediately upstream of the uCC start codon. This natural sequence lacks any in-frame near-cognate start codons (Fig. 3.18A). The start codon for the extension was either AAA (no initiation) or ACG (near cognate), and the start site for the uCC was either the natural near-cognate ACG codon or a non-cognate AAA codon. Three reporters in which the uCC as a uORF controls RLuc expression were generated: one with no start codon for the extension and the native ACG codon at the uCC start site (AAA ACG), this reporter contains the normal uCC; a second with an ACG start codon before the *Trip12* sequences and no uCC start site (ACG AAA), this reporter has the lengthened uCC; and a third normalization control with no cognate or near-cognate start codons at either the *Trip12* or the uCC start sites (AAA AAA) (Fig. 3.18A).

The three reporters were transfected into U2OS cells, which were then grown in the presence or absence of 50 mM ascorbate. The RLuc expression from the reporters was normalized to the expression of the FLuc transfection control (Fig. 3.18B). At 0 mM ascorbate, RLuc expression was high for all three reporters, but addition of 50 mM ascorbate repressed RLuc expression from the normal uCC reporter (AAA ACG). In contrast, RLuc expression was not repressed from the reporter with the lengthened uCC (ACG AAA).

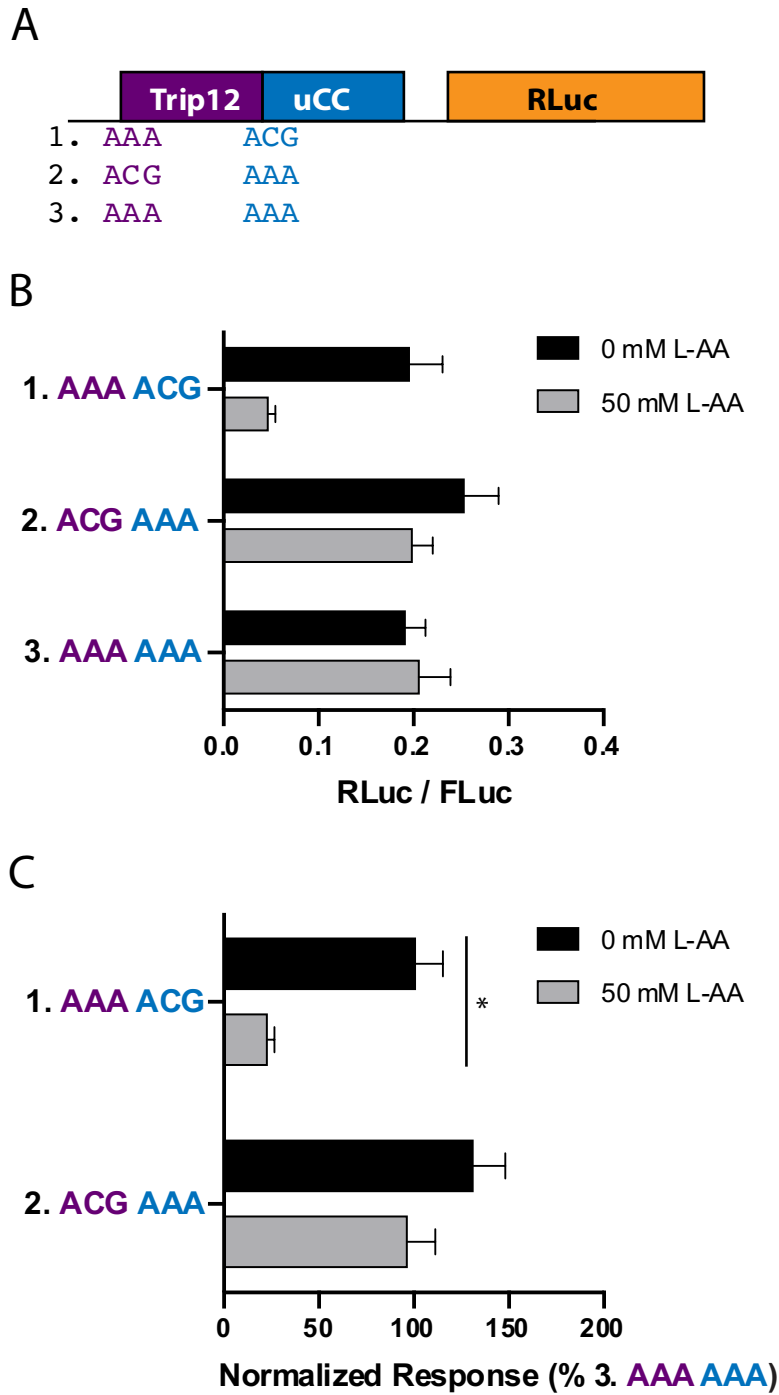


Figure 3.18. Lengthening the GGP uCC leads to a loss of regulation *in vivo*

Figure 3.18. Lengthening the GGP uCC leads to a loss of regulation *in vivo*

(A) The uCC length was doubled by inserting a 65-codon extension from the mouse *Trip12* gene directly upstream of the uCC start codon. The mouse *Trip12* sequence was chosen because of its lack of any in-frame near cognate start codons. The start codon for the extension was either AAA (no initiation) or ACG (near cognate), and the start site for the uCC was either the natural near-cognate ACG codon or a non-cognate AAA codon. This generated three reporters 1. AAA ACG, 2. ACG AAA, or 3. AAA AAA. (B) The indicated reporters were transfected into U2OS cells and the cells were incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P. RLuc activity was normalized to a co-transfected FLuc reporter (RLuc/FLuc). (C) L-AA-induced repression of RLuc expression was calculated by normalizing the RLuc/FLuc ratios for the normal and extended uCC constructs to the ratio for the no uCC control (% AAA AAA). Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; n=6, assayed in duplicate).

The expression of the reporters containing the intact or lengthened uCC were normalized to the expression of the reporter lacking start codons (AAA AAA) in order to calculate the relative regulation conferred by the native and extended uCC (Fig. 3.18C). As expected, the AAA ACG reporter containing the native uCC showed a significant, 4-fold, inhibition in response to added ascorbate. In contrast, addition of ascorbate did not significantly affect expression from the ACG AAA reporter with the extended uCC. These results support the notion that the ascorbate-induced ribosome queue on the GGP uCC is

sufficient to promote initiation at the native start site, but not when the start site is moved further upstream.

To ensure that the loss of regulation observed upon extending the length of the uCC by inserting the *Trip12* sequences (ACG AAA reporter) was not due to the *Trip12* sequences interfering with ascorbate sensing, I generated a Trip12-uCC-FLuc fusion in which a canonical AUG start codon was inserted upstream of the *Trip12* sequences and the uCC start codon was replaced by an AAA codon (Fig. 3.19A). Related constructs were generated in which the native uCC or the OF mutant, both with an AUG start codon, were fused to FLuc.

Following transfection of the reporters into U2OS cells, the cells were incubated with or without 50 mM ascorbate, harvested, and then FLuc expression was normalized to the expression of the RLuc transfection control (Fig. 3.19B). Both the uCC-FLuc fusion and Trip12-uCC-FLuc fusion showed very low expression compared to the control OF uCC-FLuc fusion even in the absence of added ascorbate. Consistent with previous results (Fig. 3.6B), addition of 50 mM ascorbate repressed expression of the uCC-FLuc fusion (Fig. 3.19B).

The response to ascorbate was calculated by normalizing the expression of each reporter in the presence of 50 mM ascorbate to the expression in the absence of ascorbate (Fig. 3.19C). As expected, the expression of the wild-type uCC-FLuc fusion was inhibited in the presence of ascorbate; however, unexpectedly, the expression of the Trip12-uCC-FLuc fusion showed no reduction in the presence of ascorbate. In fact, Trip12-uCC-FLuc expression was modestly increased in the presence of ascorbate. These results suggest that the *Trip12* extension interferes with the ability of the uCC to respond normally in the

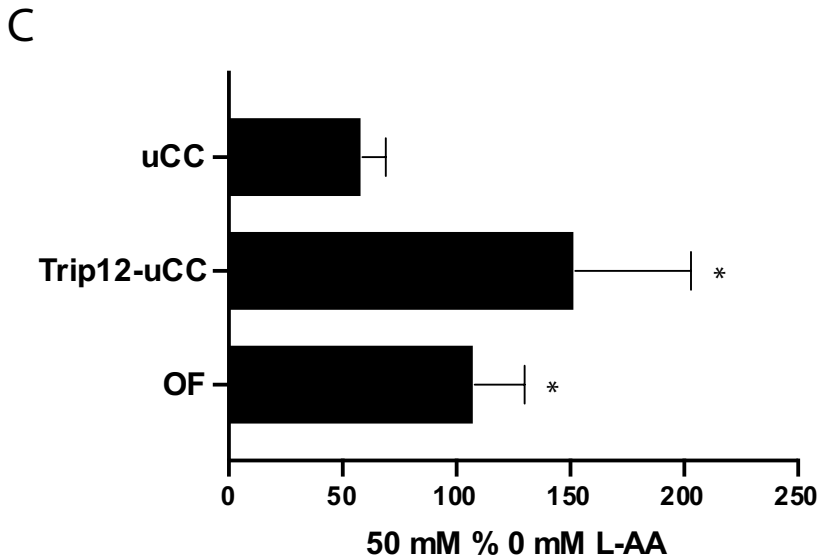
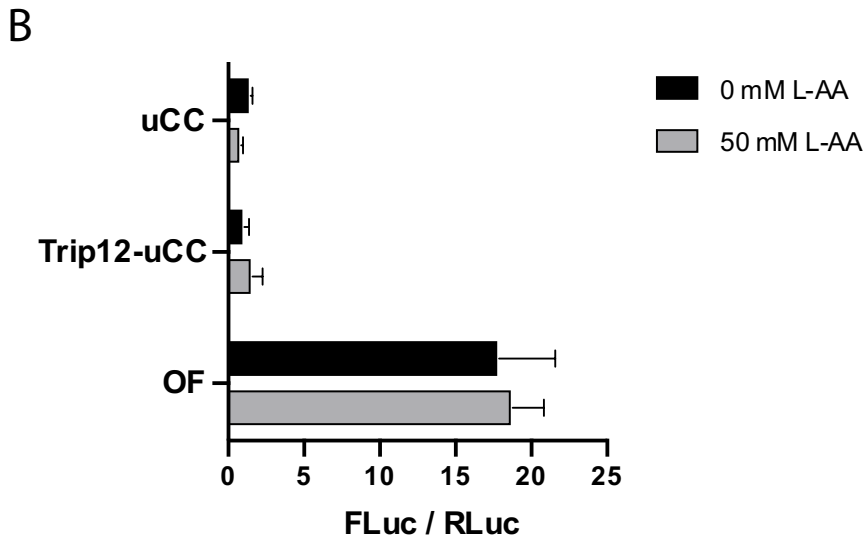
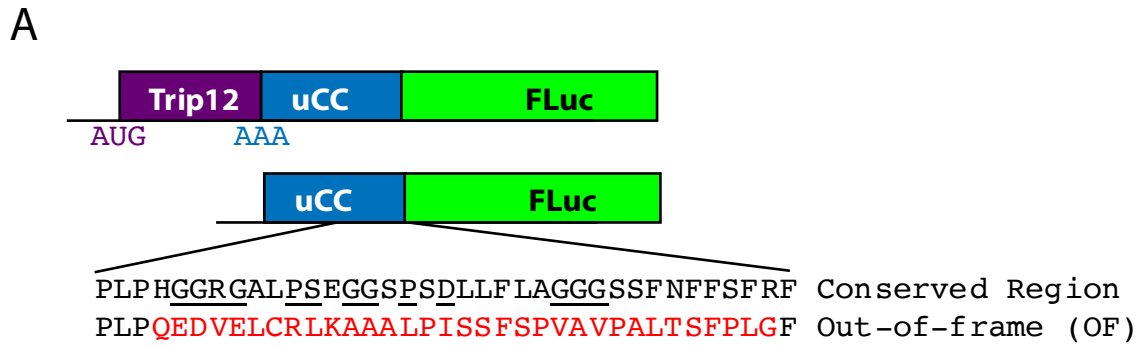


Figure 3.19. Inserting *Trip12* sequences to lengthen the GGP uCC impairs ascorbate sensing

Figure 3.19. Inserting *Trip12* sequences to lengthen the GGP uCC impairs ascorbate sensing

(A) The length of the uCC was doubled by adding a 65-codon extension from the mouse *Trip12* gene directly upstream of the uCC start codon. The mouse *Trip12* gene was chosen because it lacks in-frame near-cognate start codons. The *Trip12* extension with an AUG start codon was fused in-frame to the uCC and FLuc sequences and the start codon for the uCC was changed to a non-cognate AAA codon to generate the Trip12-uCC-FLuc fusion. The wild-type and OF uCC-FLuc fusions were used as controls. (B) The indicated reporters were transfected into U2OS cells and the cells were incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P. FLuc activity was normalized to a co-transfected RLuc reporter (FLuc/RLuc). (C) To determine how the extension effected L-AA sensing, repression of FLuc activity was calculated by normalizing the FLuc/RLuc ratio obtained in the presence of L-AA to the same ratio obtained in the absence of L-AA (50 mM % 0 mM L-AA) for each reporter. The significance of the response was calculated in comparison to the wild-type uCC-FLuc reporter. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; n=6, assayed in duplicate).

presence of ascorbate; thus, we cannot rely on these experiments to assess the importance of ribosome queuing for translational control by the GGP uCC.

3.6.3. Reducing ribosome loading with stem loop structures impairs GGP regulation in mammalian cell culture (*in vivo*)

As an alternate approach to test the queuing model, stem-loop (SL) structures were inserted near the 5' end of the GGP reporter mRNAs (Fig. 3.20A). The post-translational modification was generated from a highly palindromic sequence with a high free energy ($\Delta G = -41.3$ kcal/mol) and was previously used to reduce ribosome scanning (Baird et al., 2014; Vattem and Wek, 2004). The reporters were generated using the GGP-RLuc reporters with the post-translational modification placed 49-nt downstream from the 5' cap, impeding scanning (Babendure et al., 2006), and 325-nt upstream of the uCC start codon (Fig. 3.20B).

Following transfection of the reporters into U2OS cells, the cells were incubated with or without 50 mM ascorbate, harvested, and then RLuc expression was normalized to the expression of the FLuc transfection control (Fig. 3.20C). As expected, the addition of a post-translational modification near the 5' end of the GGP reporter mRNAs reduced RLuc expression. To more easily assess the impact of the post-translational modification on regulation of GGP mRNA translation, the RLuc activity for the wild-type uCC reporters was normalized to the activity of the corresponding OF reporter. These calculations were performed separately for the reporters with and without the post-translational modification (Fig. 3.20D). After normalizing to the OF control, addition of ascorbate was found to lower RLuc expression from the reporters both lacking and containing the post-translational modification.

Finally, the fold regulation in response to 50 mM ascorbate was calculated for the reporters containing or lacking the post-translational modification. As shown in Fig. 3.20E, the presence of the post-translational modification did not significantly affect ascorbate regulation of GGP mRNA translation. This result was surprising; however, the RLuc

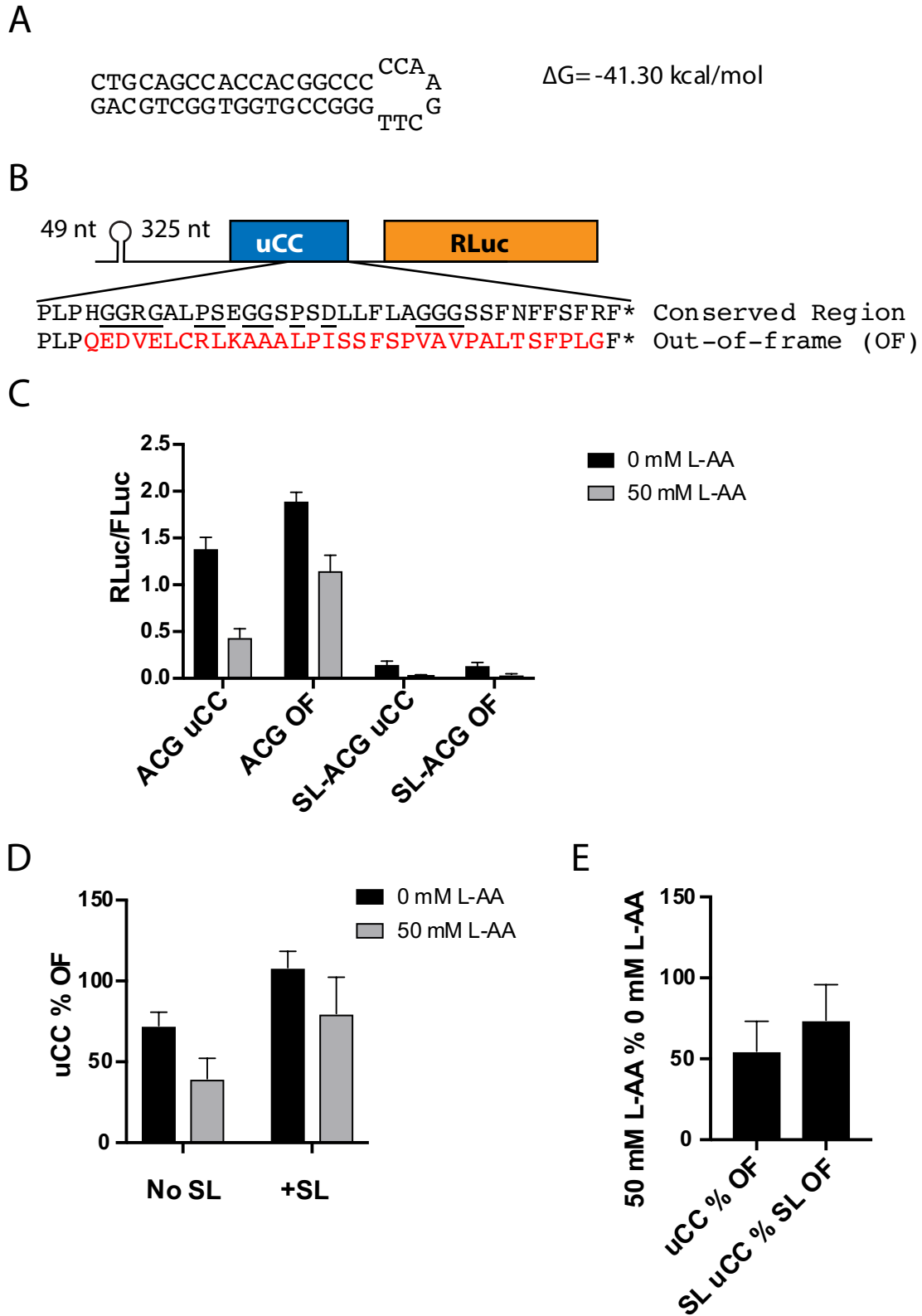


Figure 3.20. Insertion of a SL into the 5' leader of a GGP-RLuc reporter in a pHRL vector (CMV promoter) did not significantly impact ascorbate regulation

Figure 3.20. Insertion of a SL into the 5' leader of a GGP-RLuc reporter in a phRL vector (CMV promoter) did not significantly impact ascorbate regulation

(A) A SL was generated from a highly palindromic sequence with a high free energy ($\Delta G = -41.3$ kcal/mol) and (B) inserted 49 nt from the transcription start site and 325 nt upstream of the wild-type or OF uCC in a GGP-RLuc reporter. (C) The indicated reporters were transfected into U2OS cells and the cells were incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P. RLuc activity was normalized to a co-transfected FLuc reporter (RLuc/FLuc). (D) The uCC-specific response to 50 mM ascorbate-P was calculated by normalizing the response of the wild-type reporters to the OF control (uCC % OF) for reporters without (No SL) or with (+SL) the SL. (E) The impact of the SL on the response to ascorbate-P was calculated by normalizing the results from panel D in the presence to the absence of L-AA (50 mM L-AA % 0 mM L-AA). Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=6$, assayed in duplicate).

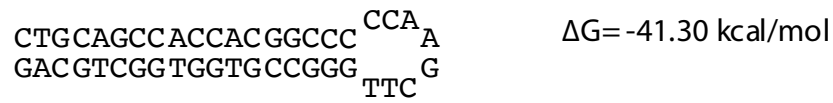
reporter is based on the vector phRL and expressed under the control of a CMV promoter that is known to use alternate transcription start sites. Perhaps the post-translational modification imposed even greater heterogeneity in transcription start sites, and the apparent lack of impact of the post-translational modification on uCC control might be due to the artifacts associated with the reporters.

Because of the potential issues regarding transcription start site heterogeneity with the CMV promoter in the RLuc reporters, the experiment was repeated using FLuc reporters which use an SV40-based reporter that has much greater fidelity in using a single

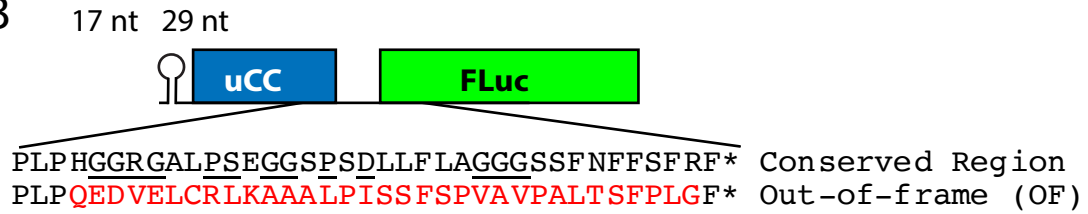
transcription start site (Fig 2.4C, Fig. 3.21A,B). In these reporters, the same post-translational modification described above was placed only 17 nt from the 5' cap, which is expected to strongly impede ribosome binding (Kozak, 1986a), and 29 nt upstream of the uCC start codon. Downstream of the uCC, the 105 nucleotides between the uCC stop codon and the GGP start codon were left intact and immediately followed by the AUG start codon of FLuc. The reporters were transfected into U2OS cells and following 22 hr incubation in the presence or absence of 50 mM ascorbate cells were harvested. The FLuc activity in the extracts was normalized to the RLuc activity from the transfection control. As shown in Fig. 3.21C, the presence of the post-translational modification substantially reduced FLuc expression in the reporters containing the wild-type or OF uCC as well as reporters in which the uCC start codon was mutated to AAA.

As observed above in the assays with the RLuc reporters, the corrected FLuc levels were normalized to the OF control for the reporters either containing (+SL) or lacking (No SL) the post-translational modification. As observed previously, the reporters lacking the post-translational modification showed a significant response to ascorbate, with an approximately 2-fold reduction in normalized FLuc activity upon addition of ascorbate ($p < 0.005$) (Fig. 3.21D). In contrast, insertion of the post-translational modification, which will reduce ribosome loading, blocked ascorbate regulation via the uCC (Fig. 3.21D). In the absence of the post-translational modification, we propose that ribosome loading allows a queue to develop on the uCC in the presence of ascorbate, which enhances initiation at the ACG start codon of the uCC. The increased uCC translation, in turn, blocks GGP (FLuc translation). In contrast, insertion of the post-translational modification reduces ribosome loading and prevents the ribosome queue from forming despite the translation elongation

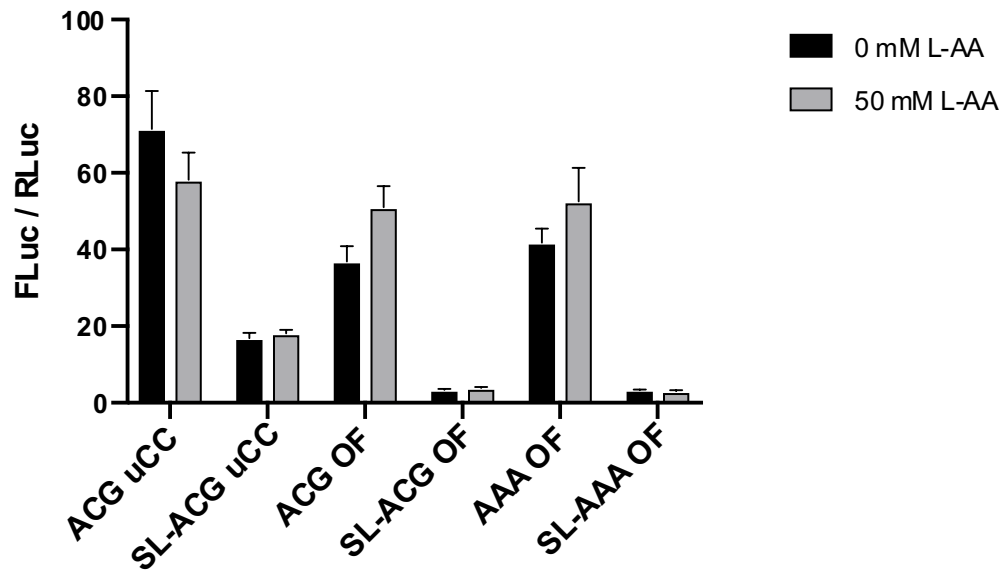
A



B



C



D

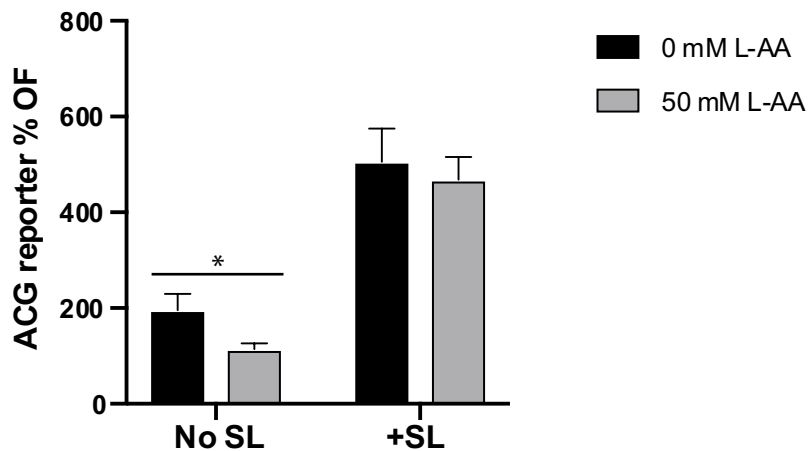


Figure 3.21. Insertion of a SL into the 5' leader of a GGP-FLuc reporter under the control of an SV40 promoter impairs ascorbate regulation *in vivo*

Figure 3.21. Insertion of a SL into the 5' leader of a GGP-FLuc reporter under the control of an SV40 promoter impairs ascorbate regulation *in vivo*

(A) A SL was generated from a highly palindromic sequence with a high free energy ($\Delta G = -41.3$ kcal/mol) and (B) inserted 17 nt from the transcription start site and 29 nt upstream of the wild-type, OF, or defective (AAA initiation codon) uCC in a GGP-FLuc reporter. (C) The indicated reporters were transfected into U2OS cells and the cells were incubated for 24 hr in DMEM in the absence or presence of 50 mM ascorbate-P. FLuc activity was normalized to a co-transfected RLuc reporter (FLuc/RLuc). (D) The effect of the SL was calculated by normalizing the response of the wild-type reporters to the OF control (ACG reporter % OF) for the reporters without (No SL) and with (+SL) the SL. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=6$, assayed in duplicate).

pause on the uCC in the presence of ascorbate. In the absence of ribosome queuing, scanning ribosomes regularly bypass the ACG start codon of the uCC (leaky scan) and initiate translation at the GGP start site, resulting in nearly equivalent expression of the FLuc reporter in the presence and absence of ascorbate.

To quantify how treatment with 50 mM ascorbate affects FLuc mRNA levels, RT-qPCR was performed as described previously. The FLuc mRNA levels were first normalized to the levels of *ActB* mRNA and then these relative FLuc mRNA levels were all normalized to the level of the wild-type (ACG-initiated uCC) reporter. As shown in Table 3.7, the fold induction in FLuc mRNA levels in response to 50 mM ascorbate was determined for each reporter and displayed along with the fold induction in FLuc activity

Table 3.7. Relative reporter mRNA levels – related to Fig. 3.21.

Reporter	Description	Treatment	Relative mRNA level		Fold induction		
			(ACG uCC 0 mM set at 1)		(0 mM L-AA / 50 mM L-AA)		
			FLuc mRNA	SD	FLuc mRNA	SD	FLuc activity
pC6240	ACG uCC	0 mM	1.00	0.80			
pC6240	ACG uCC	50 mM	0.82	0.38	0.82	0.38	0.81
pC6243	SL-ACG uCC	0 mM	1.23	1.26			
pC6243	SL-ACG uCC	50 mM	0.20	0.11	0.17	0.09	1.07
pC6241	ACG OF	0 mM	1.15	0.27			
pC6241	ACG OF	50 mM	1.30	1.26	1.13	1.09	1.38
pC6244	SL-ACG OF	0 mM	0.90	0.59			
pC6244	SL-ACG OF	50 mM	1.48	1.38	1.65	1.54	1.16

Table 3.7. Relative FLuc mRNA levels for experiments in Fig. 3.21 determined by qPCR. Standard deviation based on three biological replicates.

(from Fig. 3.21). No clear correlation was observed between FLuc mRNA levels and FLuc activity. In cells expressing the wild-type uCC reporter, a modest decrease in mRNA levels and FLuc activity was observed after treatment with 50 mM ascorbate; however, this correlation was not observed with the OF control or in the reporters containing the post-translational modification. In fact, in cells expressing the wild-type uCC reporter with the post-translational modification, a 5-fold reduction in FLuc mRNA levels was observed with no reduction in FLuc activity. The lack of correlation in relative mRNA levels and FLuc activity indicates that the changes in FLuc activity are not simply due to alterations in mRNA levels. Taken together, these results support the conclusion that impairing ribosome loading to dampen queue formation impedes ascorbate and uCC-mediated regulation of GGP mRNA translation.

4. Discussion

4.1. The first described uCC: control of *AZINI* through a polyamine regulated uCC

The 5' leader is a common target for controlling mRNA translation through initiation (Hinnebusch, 2014; Jackson et al., 2010). In addition to the m⁷G cap, mRNA secondary structure, the length of the leader, upstream AUG codons and uORFs affect the efficiency of translation (Kozak, 1991b). Furthermore, selection of the translation start site by scanning ribosomes is sensitive to the context of the start codon as well as to the use of near-cognate start codons. Ivaylo et al. (2018) revealed a mechanism of translational control by polyamines that regulates expression of *AZINI* through a uORF initiated with a near-cognate start codon. Because of the noncanonical start site, this type of uORF was defined as a uCC. A uCC has three defining characteristics: initiation at a weak start site, typically a near-cognate start codon; patterns of conservation consistent with coding regions, including synonymous rather than non-synonymous substitutions, that are essential for regulation through a translational pause; and finally, the translational pause leads to ribosome queuing that both blocks downstream scanning, and enhances initiation at the weak start site (Ivanov et al., 2018).

As discussed above, a noncanonical uORF within the *AZINI* mRNA was the first described example of a uCC and was used to define the features of a uCC. *AZINI* encodes a regulator of cellular polyamine synthesis. Polyamines are essential for a variety of basic cellular functions; however, excess polyamines have detrimental effects. Accordingly, the key enzymes in the polyamine biosynthetic pathway are tightly regulated, including at the translational level, by changes in intracellular polyamine concentrations (Ivanov et al., 2010a). Elevated polyamine concentrations repress *AZINI* synthesis through translational

control and repression is dependent on the uCC encoded in the *AZINI* mRNA (Ivanov et al., 2008).

The *AZINI* uCC initiates at a near-cognate AUU codon that is essential for leaky scanning. Introduction of a cognate AUG codon in place of the AUU codon results in constitutive repression of *AZINI* synthesis (Ivanov et al., 2008), which is consistent with the notion that the 52-codon-long uCC prevents efficient reinitiation at downstream start codons. Replacing the near-cognate AUU start codon with a non-cognate UUU codon results in constitutive derepression of *AZINI* synthesis (Ivanov et al., 2008). Thus, *AZINI* expression is inversely correlated with uCC translation, such that ribosomes that translate the uCC fail to synthesize AZIN. The coding sequence of the *AZINI* uCC is required for polyamine stimulation of uCC translation from the near-cognate AUU start codon (Ivanov et al., 2018). In a comparison of *AZINI* orthologs from vertebrates and metazoans, and uCC sequences from *ODC* homologs in invertebrates, Ivanov et al. (2018) identified a nearly perfectly conserved PPW motif near the C-terminus of the uCC. Mutations that alter the PPW motif abolish polyamine induction of uCC translation. Ribosome profiling of HEK293T cells pretreated with DFMO (to deplete polyamines) or with DFMO + 2 mM spermidine (high polyamines) revealed that polyamines inversely control ribosome occupancy on the uCC and mORF of the *AZINI* mRNA (Ivanov et al., 2018). Under low polyamine conditions, ribosome occupancy was high in the mORF and low in the uCC, correlating with high *AZINI* expression; however, under high polyamine concentrations, the relative ribosome occupancy in the uCC versus the mORF increased 22-fold. The most prominent ribosome footprints clustered in three distinct regions of the uCC. The middle cluster corresponds to ribosomes pausing during elongation with the Trp codon of the PPW

motif in the decoding A site of the ribosome and the two proline residues at the end of the nascent chain and linked directly to the P-site tRNA (Ivanov et al., 2018). The other two peaks correspond to ribosomes at the termination codon and 10 codons upstream of the termination codon. The latter peak is one ribosome width upstream of the termination codon and could represent an 80S ribosomes queued behind a ribosome paused during termination at the conserved PS-stop motif. The results of the mutational analyses and ribosome profiling assays are consistent with a model where the PPW and PS-stop motifs confer polyamine-dependent translation elongation and termination pauses, respectively, that impede the flow of scanning ribosomes that leaky scan past the near cognate start codon of the uCC. Interestingly, PPW motifs were the most prominent pause sites detected in bacteria lacking elongation factor (EF) EF-P, an elongation factor known to be important for translation of polyproline sequences (Woolstenhulme et al., 2015). The homologous eukaryotic translation factor eIF5A (Dever et al., 2014; Saini et al., 2009) has also been shown to promote translation of polyproline sequences (Gutierrez et al., 2013; Saini et al., 2009; Schuller et al., 2017), so Ivanov et al. (2018) hypothesized that eIF5A is required for elongation at the PPW motif in the uCC and that polyamines inhibit elongation at the PPW motif by inhibiting eIF5A. In a reconstituted yeast *in vitro* translation system, spermidine concentrations of 2 mM or greater inhibited synthesis of a peptide fragment containing a PPW motif (Ivanov et al., 2018). Interestingly, the polyamine spermidine was found to crosslink to helices H74 and H93 of the 23S rRNA near the PTC of bacterial ribosomes (Xaplanteri et al., 2005) and eIF5A interacts with the same region of eukaryotic ribosomes (Gutierrez et al., 2013; Ivanov et al., 2018), raising the possibility that polyamines compete with eIF5A for a common binding site on the ribosome. In support of this competition

model, synthesis of the PPW-containing peptide was maintained at high levels even under high polyamine conditions when the eIF5A concentration was increased 100-fold (Ivanov et al., 2018).

In order to test the importance of queuing for translational control of *AZINI* synthesis, Ivanov et al. (2018) used two methods to impair queuing. The first was to reduce the rate of ribosome loading on mRNAs with the drug 4EGI-1, which blocks the function of the cap-binding protein that is required for ribosome binding to the mRNA. The addition of 4EGI-1 inhibited uCC translation and depressed mORF expression on an mRNA containing the 5' leader from *AZINI* (Ivanov et al., 2018). The second strategy was to lengthen the uCC, which was hypothesized to be of optimal length to enable precise control of uCC translation in response to varying polyamine levels. Doubling the length of the uCC derepressed expression of the mORF of an *AZINI* reporter in the presence of polyamines (Ivanov et al., 2018). Presumably, the ribosome queue was of insufficient length to position a ribosome in the vicinity of the repositioned uCC start site.

Having elucidated this novel mechanism of translational control involving ribosome queuing on a non-conventional uORF, Ivanov et al. (2018) speculated that the translational control of other mRNAs might rely on distinct uCC-like elements that respond to different metabolites.

4.2. GDP-L galactose phosphorylase (GGP) mRNA was an ideal candidate for regulation by a uCC that responds to L-ascorbic acid

The uCC rules that were established during the study of the *AZINI* uCC apply to the *GGP* uCC making it an ideal candidate for a uCC that responds to ascorbate (Fig. 4.1). First, bioinformatics indicated that the *GGP* noncanonical uORF, or uCC, initiated at a weak start site. The *GGP* mRNA *Arabidopsis* does not have an in-frame AUG codon upstream of the uCC and the uCC is predicted to initiate at an ACG near-cognate codon. Here, I showed that mutating the predicted near-cognate uCC start codon (ACG) of the *Arabidopsis* *GGP* mRNA to a noncognate (AAA) codon reduced expression of a uCC-FLuc fusion to near-background levels in mammalian cells (Fig. 3.6A,B). This indicates that the near-cognate ACG codon is most likely required for initiation of translation on *GGP* uCC. Like the *AZINI* uCC, the *GGP* uCC has highly conserved motifs that are conserved at the amino acid level (Fig. 3.1). I showed that the conserved GGRG and GGG motifs are essential for repression of *GGP* expression in the presence of ascorbate (Fig. 3.3). In addition, mutation of any of the conserved motifs or alteration of their spacing impaired the ability of ascorbate to promote initiation at the ACG start codon of a uCC-FLuc fusion (Figs. 3.7, 3.8). Finally, like the *AZINI* uCC, under repressing conditions ribosomes pause during translation of the *GGP* uCC and this pause regulates uCC translation via a ribosome queuing mechanism. Ribosome profiling of two plant *GGP* mRNAs revealed that ribosomes stall at the final glycine residue of the C-terminal tri-glycine repeat in both uCCs (lower panels in Figs. 3.17A-B). I proposed that ribosome pausing on the uCC leads to queuing of subsequent scanning and translating ribosomes. Consistent with this hypothesis, disrupting queuing, by impairing ribosome loading with a

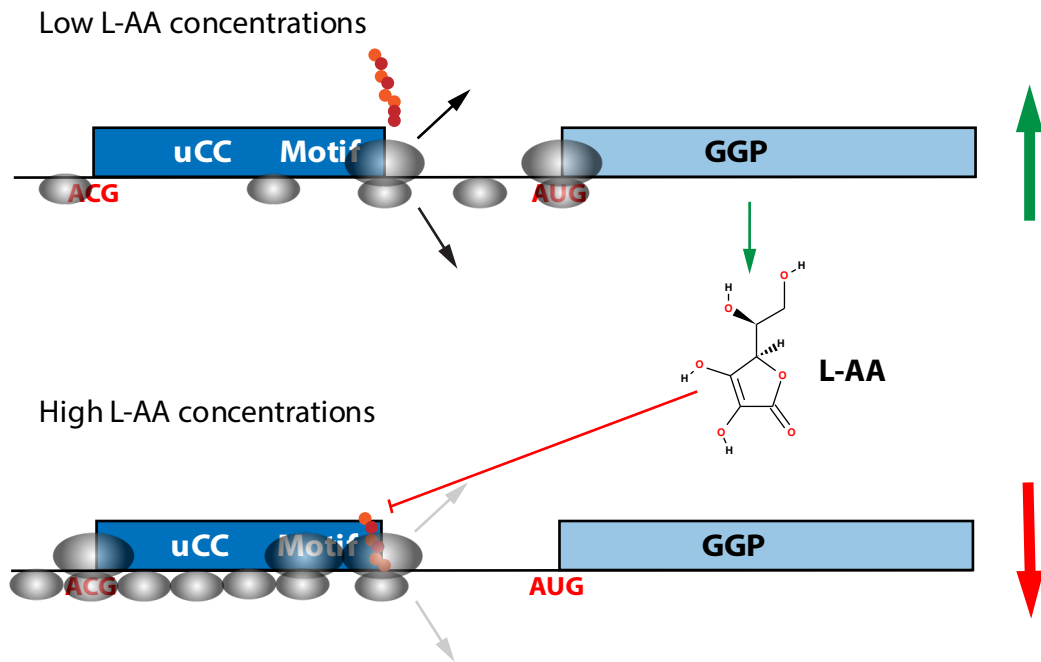


Figure 4.1. Proposed model of *GGP* regulation by the uCC and L-ascorbic acid

Figure 4.1. Proposed model of *GGP* regulation by the uCC and L-ascorbic acid

(Upper panel) Under low ascorbate (L-AA) conditions, most scanning ribosomes leaky scan over the uCC (dark blue) start codon (red ACG) without initiating and then initiate translation at the downstream *GGP* mORF (light blue). The few ribosomes that initiate translation at the uCC ACG start codon, synthesize the uCC peptide, and then disengage from the mRNA. As *GGP* expression increases, L-AA concentrations increase. (Lower panel) Under conditions of elevated L-AA concentrations, the occasional ribosome that initiates at the ACG start codon of the uCC will translate the conserved motifs at the 3' end of the uCC. L-AA will enter the peptide exit tunnel of the translating ribosome and interact with nascent chain and exit tunnel residues causing the ribosome to stall. Subsequent scanning ribosomes, which mostly leaky scan over the uCC start codon, and the occasional elongating ribosome form a queue behind the paused ribosome. As the queue develops, a scanning ribosome will spend an extended time traversing in the vicinity of the uCC start codon, providing more opportunities for initiation. The enhanced rate of initiation on the uCC reinforces the elongation stall, prevents ribosomes from leaky scanning to the *GGP* mORF, and thus represses *GGP* synthesis.

post-translational modification inserted near the 5' end of the mRNA, depressed *GGP* expression in the presence of ascorbate (Fig. 3.21). Based on these and other findings in this thesis, I conclude that the *GGP* noncanonical uORF is in fact a uCC that regulates *GGP* expression in response to ascorbate.

We propose that the uCC mechanism of gene regulation is a homeostatic response that allows cells to fine tune metabolite concentrations rapidly. As such, the system might be optimized to narrowly control GGP expression through a range of ascorbate concentrations. In their study, Laing et al. (2015) utilized transient transfection of *N. benthamiana* leaves to study ascorbate regulation of GGP expression. They generated reporters containing the kiwifruit (*Actinidia eriantha*) GGP promoter including the 5'-UTR (with the uCC) upstream of FLuc and expressed the gene in *N. benthamiana* leaves (Laing et al., 2015). To increase leaf ascorbate levels, they co-expressed the coding sequence of kiwifruit GGP without the 5'-UTR (or the uCC) under a strong constitutive promoter. Overexpression of GGP increased ascorbate levels approximately 2-fold from 27.3 mg/100 gFW to 49.3 mg/100 gFW. Laing et al. (2015) observed an approximately 5-fold reduction in FLuc activity in leaves with higher ascorbate levels. This regulation is very similar to the approximately 4-fold reduction we observed in the mammalian cell culture system when the *GGP* uCC was placed upstream of RLuc and the cells were treated with 50 mM ascorbate (Fig. 3.2B). After normalization to the OF control (Fig. 3.2C), uCC specific repression is only 2-fold for the GGP reporter, but this level of regulation is not too dissimilar from the 7-fold repression of the *AZINI* uCC by spermidine observed by Ivanov et al. (2018). In addition, the 3-fold repression of the GGP uCC-FLuc activity that we observe in RRL (Fig. 3.11C, and 3.13C) is similar to the 1.5-4.0-fold repression of the AAP-FLuc fusions by arginine in translation extracts reported by Wang et al. (1999). Taken together, these results indicate that the uCC system of translational control likely exerts only moderate levels of regulation under physiological conditions; however, the similar levels of regulation seen across different

uCCs and translation systems indicates that the uCC system is robust and possibly universal in eukaryotes.

4.3. How does the *GGP* uCC detect changes in L-ascorbic acid concentration?

Although we established that the *GGP* uCC represses *GGP* expression in response to elevated ascorbate concentrations, we were unable to determine how ascorbate is sensed by the translating ribosome or how it induces an elongation pause. The *AZINI* uCC contains one highly conserved motif, a PPW sequence near the C-terminus, that requires eIF5A for its translation (Ivanov et al., 2018). High polyamine concentrations impair eIF5A function (Ivanov et al., 2018), which is especially important for translation of particular motifs, including polyproline stretches and PPW (Gutierrez et al., 2013; Ivanov et al., 2018; Manjunath et al., 2019). Unlike the *AZINI* uCC, the *GGP* uCC does not have polyproline motifs or motifs that have been identified as requiring eIF5A (Fig. 3.1) (Manjunath et al., 2019); therefore, there it is unlikely that eIF5A plays a role in regulating *GGP* expression. In addition, the conserved motifs in the *GGP* uCC, with five perfectly conserved motifs spanning exactly 23 codons, are more extensive than the simple PPW motif in the *AZINI* uCC. Each of the conserved motifs in the *GGP* uCC, as well as their spacing, plays a role in regulation of *GGP* expression. Substitutions within any of the motifs or inserting even a single amino acid within the conserved region, abolished ascorbate-specific induction of uCC translation (Figs. 3.7, 3.8). Based on the ribosome profiling data, we were able to conclude that the C-terminal tri-glycine repeat serves as a stall site; but how pausing is regulated by ascorbate is unknown (lower panels Fig. 3.17A,B).

We propose that the *GGP* uCC detects ascorbate through an interaction between ascorbate, the uCC nascent chain, and the ribosome exit tunnel in a mechanism similar to the well-characterized co-factor mediated translation arrest on the TnaC mRNA in bacteria (Fig. 4.2). The TnaC nascent peptide stalls the ribosome in the presence of L-tryptophan to

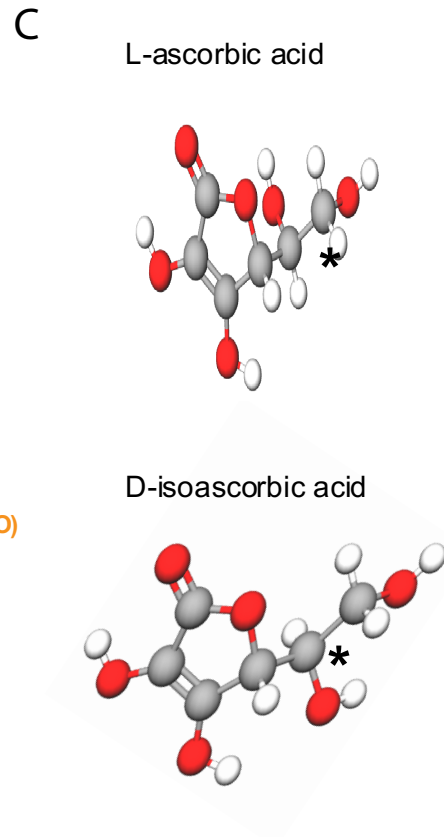
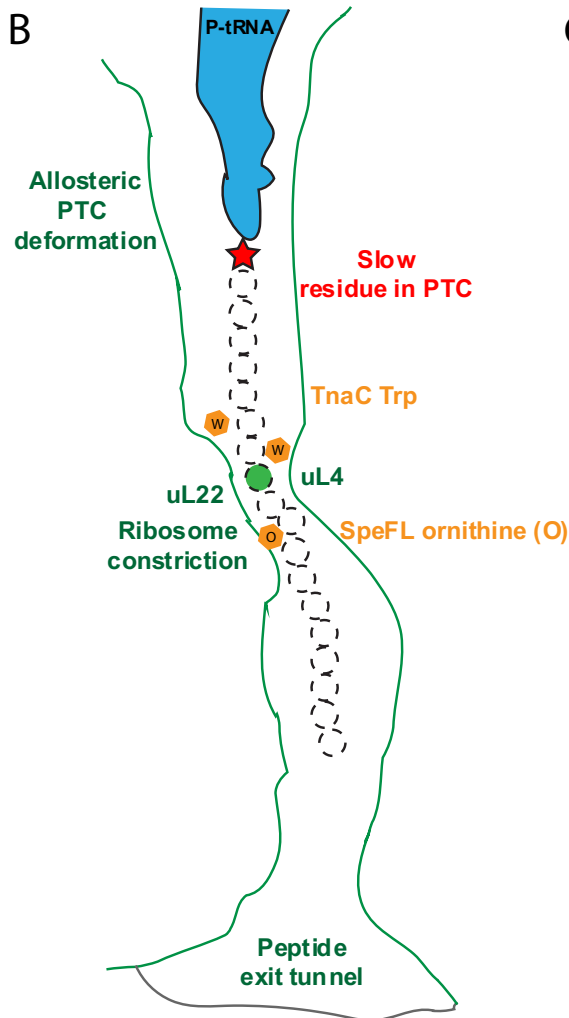
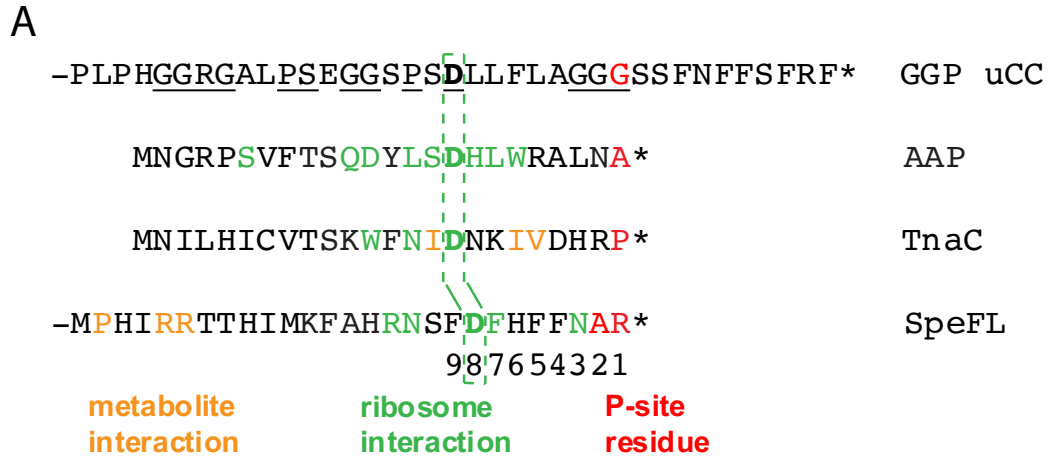


Figure 4.2. *GGP* uCC and other metabolite-sensing nascent peptides

Figure 4.2. GGP uCC and other metabolite-sensing nascent peptides

(A) Amino acid sequence of the conserved C-terminal portion of the *A. thaliana* GGP uCC nascent peptide aligned from the stalled P-site residue (red) with three other described stalling peptides: *N. crassa* AAP, *E. coli* TnaC, and the C-terminal portion of *E. coli* SpeFL. Residues reported to interact with a metabolite (Trp for TnaC or ornithine for SpeFL) are shown in orange; residues reported to interact with rRNA or ribosomal proteins in the ribosome exit tunnel are shown in green. The conserved Asp residues located 8 or 9 residues before the stall (Asp16 in TnaC and AAP, shown to interact with rRNA A2058 in cryo-EM structures, Asp46 in the GGP uCC and Asp27 in SpeFL) are in bold. (B) Schematic illustrating the common features of the ribosome stalling peptides shown in A. Interactions of the nascent peptide with ribosomal components including the proteins uL4 and uL22 that form the tunnel constriction trigger allosteric conformational rearrangements of PTC nucleotides that, combined with the poorly reactive substrates in the A and P sites, results in inhibition of the PTC. The schematic was adapted from Figure 4 in Wilson et al. (2016). (C) Ball and stick model of L-ascorbic acid and D-isoascorbic acid with the chiral center marked with an asterisk (carbon, gray; oxygen, red; hydrogen, white).

upregulate expression of specific enzymes related to L-tryptophan uptake and catabolism (Cruz-Vera et al., 2005; Cruz-Vera and Yanofsky, 2008; Gong and Yanofsky, 2002b). TnaC is a 24-amino acid peptide that stalls with the Pro24 codon in the P site and the UGA stop codon in the A site. Stalling requires mid-tunnel residues Trp12 (Gollnick and Yanofsky, 1990; Gong and Yanofsky, 2002b) and Asp16 (Cruz-Vera and Yanofsky, 2008)

as well as the PTC-proximal Pro24 at an exact distance from Trp12/Asp16 (Fig. 4.2A) (Gong and Yanofsky, 2002b) for termination arrest. In structural studies, binding of two L-tryptophan molecules in pockets formed by TnaC and the ribosome exit tunnel caused the PTC to change conformation (Fig. 4.2B) (Bischoff et al., 2014). The pockets were formed by TnaC residues between Asn17 and Asp21. The invariant Trp12 and Asp16 residues interact with the ribosomal proteins uL22 and uL4 that form the central constriction, a narrow passage in the ribosome peptide exit tunnel (Bischoff et al., 2014). The *GGP* uCC has a minimum length similar to that of TnaC and my studies showed that the conservation of the spacing between motifs is essential for ribosome stalling on the *GGP* uCC in the presence of ascorbate.

It is possible that the conserved *GGP* uCC C-terminal region acts together with residues in the ribosome exit tunnel to recognize ascorbate. In a recent paper, del Valle et al. (2019) elucidated a mechanism whereby polyamine synthesis in bacteria is regulated through interactions between ornithine, a specific peptide nascent chain, and residues in the ribosome exit tunnel. In addition to being essential in eukaryotes, polyamines play an essential role in bacterial cell growth, stress adaptation, and virulence. Pathogenic bacteria adjust the expression and activity of polyamine biosynthetic enzymes in response to environmental stresses and metabolic cues. In *E. coli*, the inducible ornithine decarboxylase gene *speF* is present in an operon with *potE*, an ornithine-putrescine antiporter. A regulatory uORF, *speFL*, is located upstream of *speF* and controls synthesis of the *speF-potE* mRNA. The 34-amino acid SpeFL peptide acts as an ornithine sensor. In a cryo-EM structure of an *E. coli* ribosome stalled during translation of *speFL* in the presence of ornithine, the ribosome and SpeFL sensor domain form a selective binding

pocket that accommodates a single ornithine molecule (Fig. 4.2B) (del Valle et al., 2019). Ornithine pre-associates with the ribosome and is then held in place by the SpeFL sensor domain, which leads to compaction of the C-terminal SpeFL effector domain, blocking the action of release factor RF1. Ribosome stalling on *speFL* masks a predicted Rho utilization site, allowing transcription to proceed and *speF* and *potE* to be expressed. Under low ornithine conditions, the leading ribosome on *speFL* terminates efficiently and then dissociates from the mRNA, exposing part of the predicted Rho utilization site and leading to attenuation of transcription prior to the *speF* coding region.

By analogy to SpeFL, we propose that the C-terminal region of the *GGP* uCC folds in ribosome exit tunnel enabling the conserved amino acids to interact with ascorbate and/or residues in the ribosome exit tunnel causing the ribosome to stall during elongation at the terminal glycine residue of the tri-glycine repeat. If ascorbate binds low in the exit tunnel (near the PTC), the terminal glycine residues could contribute to ascorbate recognition. Based on the sequence of the uCC motifs, we hypothesize that multiple motifs might interact to sense ascorbate and to induce the elongation pause (Fig. 4.2A). The N-terminal motifs, the GGRG motif, and possibly the GG motif could serve as ascorbate sensors within the ribosome exit tunnel. They could form a pocket for ascorbate along with rRNA and ribosomal protein residues in the ribosome exit tunnel. In a further analogy to SpeFL, where conserved rare codons in the middle of the peptide are thought to slow elongation and allow an N-terminal part of the nascent chain to fold prior to synthesis of the C-terminal effector domain, the PxD motif in the *GGP* uCC might slow elongation to allow N-terminal motifs to adopt a conformation necessary for sensing ascorbate. Notably, proline, glycine and aspartic acid motifs are known to stall elongation in eIF5A-depleted

cells and are thought to represent slowly translated residues (Pelechano and Alepuz, 2017; Schuller and Green, 2018; Schuller et al., 2017). As translational control of the GGP uCC functions in both plants and mammalian cells, the ribosome peptide exit tunnel residues contributing to uCC and ascorbate recognition are likely conserved through evolution. In contrast to TnaC and SpeFL, which arrest at termination (Bischoff et al., 2014; del Valle et al., 2019), interactions between the uCC nascent chain, ascorbate, and residues in the ribosome exit tunnel cause the elongating ribosome to stall at the C-terminal tri-glycine repeat. Consistent with this notion of an elongation rather than termination stall on the GGP uCC, the distance between the conserved motifs and the stop codon, which is conserved in TnaC and SpeFL, is not conserved among different *GGP* uCCs (Fig. 3.1).

Though we do not know the ascorbate concentration in the tissues used for the ribosome profiling study, it is conceivable that ribosomes always pause during translation of the C-terminal tri-glycine repeat of the *GGP* uCC. However, we propose that *GGP* regulation is dependent on the duration of the pause and how frequently it occurs. The pause would need to last long enough to allow a queue of scanning and translating ribosomes to form that positions a scanning ribosome in the vicinity of the weak initiation site of the uCC. In theory, interactions between the duration of the pause, the distance between the pause site and the uCC start codon (queuing distance), and the ribosome loading rate could fine tune *GGP* expression, and therefore ascorbate concentrations, in different plant species or tissues.

4.4. Are there other uCCs and what is the advantage of a uCC?

As discussed in earlier sections, translational regulation via nascent chain-mediated ribosome stalling is not rare. In bacteria, nascent chain-mediated ribosome stalling can regulate gene expression in multiple ways including transcription antitermination via termination stalling (TnaC and SpeFL) (Bischoff et al., 2014; del Valle et al., 2019) and translation induction via translation elongation stalling (SecM and ErmCL among others) (Arenz et al., 2014; Ito and Chiba, 2013; Sarker and Oliver, 2002; Wilson et al., 2016; Zhang et al., 2015). In the case of the bacterial examples described above, gene expression is induced rather than repressed by ribosome stalling. In addition to the regulatory mechanisms discussed above, riboswitches are a common means of genetic regulation at the mRNA level in bacteria (Garst et al., 2011). Riboswitches are most commonly found in the 5'-leader of bacterial mRNAs, and through direct interactions with metabolites regulate expression of the coding region via a secondary structural switch. A typical riboswitch contains two distinct functional domains. An aptamer domain recognizes the effector molecule by adopting a compact three-dimensional fold to scaffold the ligand binding pocket (Winkler and Breaker, 2003). The RNA receptor is able to discriminate between chemically related metabolites with high selectivity. The second domain, the expression platform, contains a secondary structural switch that interfaces with the transcriptional or translational machinery (Garst et al., 2011). Regulation is achieved by a region of overlap between the two domains called the switching sequence, whose pairing directs folding of the RNA into one of two mutually exclusive structures in the expression platform that represents the on and off states of the mRNA (Garst et al., 2011).

One highly conserved group of riboswitches is the thiamine pyrophosphate (TPP)-binding riboswitches, which have been identified in all kingdoms (Sudarsan et al., 2003). In eukaryotes, the TPP riboswitches are in introns or the 5'- or 3'-UTR of their target mRNAs. In the case of *A. thaliana*, the TPP riboswitch AtRs is located in the 3' region of the *thiC* gene. ThiC catalyzes the conversion of 5-aminoimidazole ribotide to hydroxymethyl pyrimidine phosphate in the thiamine-biosynthetic pathway. The TPP binding switch is proposed to regulate mRNA processing and stability of *thiC* in plants (Sudarsan et al., 2003). A 2.9 angstrom-resolution crystal structure revealed that the *A. thaliana* TPP riboswitch recognizes TPP via conserved residues located within two parallel sensor helices (Thore et al., 2006).

We believe that the uCC could be an additional mechanism that eukaryotes use to directly sense and respond to changes in metabolite concentrations. Currently, only two uCCs, the *AZINI* uCC and the *GGP* uCC, have been formally identified as uCCs, but we believe that several well-described uORF-encoded arrest sequences are actually uCCs. Although not formally defined as a uCC, we believe that the *N. crassa* AAP, the well-described stalling uORF-encoded peptide, is, in fact, a uCC. The *N. crassa arg-2* mRNA, and the related *S. cerevisiae CPAI* mRNA, encode a 23-residue and 24-residue peptide, respectively, upstream of the arginine biosynthetic enzyme subunit CPSase A (Luo et al., 1995; Pierard and Schroter, 1978; Werner et al., 1987; Werner et al., 1985). Acting in *cis*, the AAP represses expression of *CPAI* and *arg-2* in response to arginine (Delbecq et al., 1994; Luo et al., 1995; Werner et al., 1987). Unlike the mouse *AZINI* uCC, or the *A. thaliana GGP* uCC, which initiate at near-cognate AUU or ACG start codons, respectively (Ivanov et al., 2018; Laing et al., 2015), the *CPAI* and *arg-2* uORFs (AAPs) initiate at an

AUG codon in poor context (Wang et al., 1999; Werner et al., 1987). Of note, some *GGP* uCCs also initiate at AUG codons in poor context (Fig. 3.1). Similar to a near-cognate start codon, AUG start codons in poor context allow frequent leaky scanning (Kozak, 1986b). The AAP also shows conservation at the amino acid level; four conserved residues in the middle of the AAP are necessary for arrest as are two conserved residues near the C-terminus that are proximal to the PTC in the arrested complex (Delbecq et al., 2000; Hood et al., 2007; Ito and Chiba, 2013; Spevak et al., 2010). Finally, there is evidence that AAP is regulatory via an elongation pause and ribosome queuing. To test the metabolite specificity of uCCs, Ivanov et al. (2018) and I used a reporter that replaced the *AZIN1* uCC regulatory elements with those from the *N. crassa* AAP. This fusion uCC contained the near-cognate AUU start site and 5' portion of the mouse *AZIN1* uCC linked to the *N. crassa* AAP. Both Ivanov et al. (2018) and I (Fig. 3.4) showed that the chimeric uCC responded to Arg, but not to polyamines or ascorbate. These results indicate first, the different uCCs specifically respond to distinct metabolites, and second, that the Arg-induced translational stall on the AAP, like the polyamine- and ascorbate-induced stalls on the *AZIN1* and *GGP* uCCs, respectively, leads to ribosome queuing and enhanced initiation at the weak start site of the chimeric uCC and represses translation of the mORF.

In the case of the AAP, it seems likely that arginine binds in the ribosome exit tunnel and interacts with the nascent chain to affect PTC activity, similar to the proposed model for regulation of the *GGP* uCC and bacterial TnaC by ascorbate and Trp, respectively (Fig. 4.2B). While Arg induces stalling at the AAP stop codon, the stop codon is not a prerequisite for stalling (Wang et al., 1998); therefore, the AAP possesses both anti-termination and anti-elongation qualities. Mutation of the AAP residues that lie in the

middle of the exit tunnel in the stalled complex abolishes Arg-induced stalling, indicating that the peptide interacts with the ribosome exit tunnel (Freitag et al., 1996; Ito and Chiba, 2013; Wang and Sachs, 1997). Consistent with this idea, in a cryo-EM structure of the stalled AAP, the highly conserved residues of the AAP were seen interacting the ribosome tunnel constriction (Fig. 4.2A,B) (Bhushan et al., 2010), and mutation of these residues abolished translational stalling (Freitag et al., 1996). Similar to TnaC, conserved amino acid residues in the AAP arrest sequences interact with conserved rRNA nucleotides Ala24/Asn23-U2585/A2062, Asp16-A2058, and Asp12/Gln11-A751 as well as with extensions of ribosomal proteins L4 and L17 (Bhushan et al., 2010; Bischoff et al., 2014). The rRNA residue A2062, in the vicinity of the PTC, is close to the site of AAP compaction in the ribosome exit tunnel, and in the structure of the stalled AAP complex A2062 adopts a distinct conformation that is similar to the change in conformation of the corresponding bacterial rRNA nucleotide during TnaC stalling that is incompatible with release factor action (Bhushan et al., 2010; Bischoff et al., 2014). Intriguingly, the perfectly conserved *GGP* uCC residue Asp46 is located eight residues upstream from the stalling site, the same distance as AAP residue Asp16 and the TnaC residue Asp16 that are both described as interacting with rRNA in the exit tunnel (Fig. 4.2A,B) (Bhushan et al., 2010; Bischoff et al., 2014), raising the possibility that interactions of Asp46 (part of the PxD motif) with the exit tunnel contributes to ascorbate-induced stalling during translation of the *GGP* uCC. In addition, the specificity of regulation to L-ascorbate rather than D-isoascorbate also points to L-ascorbate binding in the ribosome exit tunnel (Fig. 4.2C).

In addition to the well-described AAP, we believe that the mRNAs encoding growth arrest and DNA damage-inducible protein (GADD34 or PPP1R15A) and DNA

damage inducible transcript 3 (also known as C/EBP homologous protein (CHOP)) both contain uCCs. *GADD34* and *CHOP* are members of a group of genes whose transcript levels are increased by ATF4 in response to the integrated stress response and unfolded protein response (Harding et al., 2000). The integrated stress response and unfolded protein response lead to increased phosphorylation of eIF2 α and inhibition of eIF2B, reducing global translation while increasing translation of select transcripts including ATF4. GADD34, a protein phosphatase 1 (PP1c) regulatory subunit, binds to PP1 and directs it to dephosphorylate eIF2 α -P and restore protein synthesis (Brush et al., 2003).

The 5' leaders of the *Drosophila melanogaster* and mouse *GADD34* transcripts each contain 2 uORFs, with the first being dispensable for regulation (Lee et al., 2009; Malzer et al., 2013; Young et al., 2015b). uORF2 in the *D. melanogaster* *GADD34* mRNA overlaps out-of-frame with the *GADD34* coding sequence and translation of this uORF blocks GADD34 production because ribosomes that translated the uORF terminate 3' of the *GADD34* coding sequence start codon (Malzer et al., 2013). Mouse uORF2 terminates 23 nt upstream of the *GADD34* coding sequence start codon and is thought to inhibit translation through inefficient translation termination (Young et al., 2015b). In their study of *GADD34*, Young et al. (2015) utilized mouse embryonic fibroblast (MEF) cells expressing a wild-type version of eIF2 α or eIF2 α -S51A, a mutant that cannot be phosphorylated. To study regulation of GADD34 under ER stress, the cells were treated with thapsigargin. Unlike the previously described uCCs, uORF2 initiates at an AUG in good, but imperfect, context (GGCGACAAUGU) (Palam et al., 2011). The current hypothesis for eIF2 α -P control of *GADD34* expression is that phosphorylation of eIF2 causes scanning ribosomes to leaky scan over the uORF2 imperfect start codon without

initiating. Consistent with this proposal, placing the uORF2 AUG start codon in perfect context greatly reduced expression of a *GADD34* reporter (Young et al., 2015b). This context mutation also decreased stress-induced induction of *GADD34* reporter expression from 3.3-fold in the reporter with the wild-type context to only 2.7-fold after optimization of the start codon context (Young et al., 2015b). Like the other described uCCs, the uORF2 peptide is necessary for regulation. Mutations that alter the reading frame of uORF2 lead to loss of regulation in response to ER stress and increased basal expression 9-fold (Young et al., 2015b). The C-terminal PPG* motif appears to be essential for regulation via inefficient termination, and replacing this motif with alanine residues, or inserting an alanine residue between the PPG motif and the stop codon, increased reporter expression although some regulation in response to ER stress was still observed (Young et al., 2015b). Consistent with this idea of defective termination, toe-printing experiments demonstrated that ribosomes stall at the stop codon (Young et al., 2015b). Intriguingly, uORF2 contains a perfectly conserved Asp13 residue that would be in a similar position as the TnaC Asp16 residue, AAP Asp12 residue, and *A. thaliana* *GGP* uCC Asp46 residue and located upstream from the stall site. The conservation of residues that would be in the ribosome exit tunnel (Ito and Chiba, 2013; Wilson et al., 2016), especially Asp13, and the evidence of a termination defect, suggest that *GADD34* uORF2 could function in a manner similar to TnaC, where an interaction between the nascent chain, the ribosome exit tunnel, and perhaps some unknown regulatory molecule, results in a change in the PTC geometry that is unfavorable for translation termination. In addition to the imperfect start codon context and translation stalling in uORF2 supporting the uCC hypothesis, addition of a post-translational modification upstream of uORF2 led to a loss of ER stress-dependent

regulation of *GADD34* reporter expression (Young et al., 2015b), supporting the notion that a ribosome queue promotes uORF2 translation and represses *GADD34* synthesis.

CHOP is an ER stress-regulated transcription factor that regulates metabolic genes during ER stress (Chikka et al., 2013). Like *GADD34*, *CHOP* is regulated by ribosomal leaky scanning of a uORF with two in-frame AUG start codons in good, but imperfect, context, with the second AUG codon acting as the primary start site (Palam et al., 2011). In their study of the *CHOP* uORF, Young et al. (2016) found that improving the context around both AUG codons reduced basal *CHOP* reporter expression by about 2-fold and decreased induction upon ER stress from 2.7-fold to 2.1-fold. Like the *GGP* uCC and *AAP*, the *CHOP* uORF contains perfectly conserved residues that would be in the ribosome exit tunnel when ribosomes stall near the 3' end of the uORF. Mutating the three perfectly conserved RRK motif, located 9-11 residues upstream of the stop codon, to alanines increased *CHOP* reporter expression 2.4-fold, while mutating the conserved IFI motif, located 5-7 residues from the C-terminus, resulted in a 4.9-fold increase in *CHOP* reporter expression, similar to the increase seen with the OF control (Young et al., 2016). Using a toe-printing assay, Young et al (2016) observed an elongation stall at the IFI motif; however, there are an additional four perfectly conserved residues, HHHT, immediately C-terminal to the IFI motif. Mutating the histidine residues in this HHHT motif to alanines did not significantly affect *CHOP* reporter expression and the terminal Thr residue was not mutated in the study (Young et al., 2016). As it is difficult to understand why residues C-terminal to the stall would be conserved, it seems likely that the stall could actually occur at termination rather than during elongation; however, additional studies will be needed to clarify this point.

Both the *GADD34* uORF2 and the *CHOP* uORF possess the hallmarks of uCC elements: a conserved peptide sequence, evidence of an elongation or termination stall (though in this case the stall appears to be constitutive rather than regulated), and a weak start codon, though the AUG codon in good but imperfect context is much better than the near-cognate start codons of the *AZINI* and *GGP* uCCs. While increased eIF2 α phosphorylation under ER stress conditions was proposed to increase leaky scanning and enable ribosomes to bypass the inhibitory uORFs due to the imperfect context (Young et al., 2016; Young and Wek, 2016; Young et al., 2015b), an alternative possibility is that eIF2 α phosphorylation impairs ribosome loading. The decreased ribosome loading would reduce queue formation on the uORFs and enable more ribosomes to leaky scan over the uORFs, thus increasing *GADD34* and *CHOP* expression.

In addition to the *GGP* uCC, over 100 different plant genes encode conserved peptide uORFs in their 5' leaders (van der Horst et al., 2020). Besides regulation of *GGP* expression by ascorbate, two other groups of plant uORFs have been described as being regulated by metabolites. The first is the *AdoMetDC* uORF regulated by polyamines that was described in the Introduction. The second is the *Arabidopsis* S1-group bZIP transcription factors regulated by sucrose (Kang et al., 2010). The bZIP transcription factors are regulators of metabolism and control amino acid and sugar metabolism as well as resource allocation (Kang et al., 2010). Translation of the bZIP mORFs is regulated by a uORF, uORF2b, in a sucrose-dependent manner (Rahmani et al., 2009). Like the AAP, uORF2b initiates translation at an AUG codon in poor context, allowing for leaky scanning (Rahmani et al., 2009). In addition, the peptide encoded by uORF2b is highly conserved and a frame shift mutation completely abolished sucrose-induced repression of translation

(Rahmani et al., 2009). Surprisingly, substituting three Arg residues at the beginning of the conserved region with alanines did not abolish regulation (Rahmani et al., 2009; Yamashita et al., 2017); however, a different study found that the alanine substitutions also led to stalling through changes in the PTC (Yamashita et al., 2017). The C-terminal 10 amino acids of uORF2b are critical for sucrose-induced ribosome stalling (Yamashita et al., 2017). Changing the position of the stop codon also abolished sucrose-induced repression of translation, indicating that the stall might occur during termination (Rahmani et al., 2009; Yamashita et al., 2017). Taken together, the weak start site, conservation of the peptide, and regulated stall, indicate that bZIP uORF2b might be uCC regulated by sucrose.

We believe that with further investigation many of the conserved uORFs identified through bioinformatics approaches will be classified as uCCs. Regulated translational stalls are used in bacteria to induce downstream gene expression in response to metabolites and antibiotics (Ito and Chiba, 2013; Wilson et al., 2016), and the uCC could be the repressive eukaryotic equivalent of the bacterial systems. The uCC allows cells to rapidly respond to changes in metabolites at the level of translation, rather than synthesizing or degrading mRNAs, or degrading excess enzymes. In our uCC model, the interplay between the ribosome stall time and ribosome loading determines the development of the ribosome queue. The distance from the uCC start codon to the stall site is critical because the ribosome stall time, and ribosome loading, have to be sufficient for a queue to develop from the stall site to the start codon in order to enhance initiation at the weak start site. As evidenced by the alignment of *GGP* uCC sequences (Fig. 3.1), the distance from the start codon to the stall site can vary greatly across species for the same uCC. The variation in distance from the stall to the start site could be in response to different needs for the particular metabolite or

differences in the rate of ribosome loading between species and mRNAs. A species that requires higher levels of the metabolite could have a longer distance from the stall site to the initiation site to make the uCC less repressive. In addition, stress responses can play a role in queuing potential as evidenced by *GADD34* and *CHOP* regulation. Phosphorylation of eIF2 α derepressed *GADD34* and *CHOP* expression (Young et al., 2016; Young et al., 2015b), and we propose that initiation at the imperfect uORF start sites is reduced through decreased ribosome loading, and a therefore a decrease in queuing. It will be of interest to determine if phosphorylation of eIF2 α can override the metabolite-induced regulation of the *AZIN1* and *GGP* uCCs, perhaps enabling even greater control of metabolite levels under certain stress conditions.

4.5. What can we learn from the uCC model?

In addition to being a novel form of gene regulation in eukaryotes, the uCC model poses interesting questions about ribosome queuing, translation initiation dynamics, and the potential use of uCCs in biotechnology. How does a ribosome queue form and is it a mixture of scanning and translating ribosomes? Does the queue affect mRNA turnover and further reduce expression of the mORF? How does the queue affect initiation at weak start codons?

Currently, there is no way to monitor ribosome queuing that is not purely 80S translating ribosomes. Ribosome profiling of the *C. reinhardtii* *GGP* mRNA showed evidence of a second peak 30-nt or 10 codons upstream of the stall peak, indicating a queued 80S ribosome (Fig. 3.17); however, no further peaks consistent with further queuing could be resolved. The crosslinking-dependent 40S ribosome profiling technique (Archer et al., 2016) might provide insights into how the ribosome queue forms. In our study, we measured reporter mRNA levels in cells after treatment with ascorbate, but we did not measure the native mRNA under conditions that lead to ribosome queue formation. Kinetic traps such as hairpins or truncations that prevent further translation are known to trigger the no-go mRNA decay pathway as a quality control mechanism (Doma and Parker, 2006). Finally, the enhancement of initiation at weak start codons by ribosome queuing has been described for uCCs and for drug-induced ribosome queuing (Ivanov et al., 2018; Kearse et al., 2019). Aberrant translation initiation at non-AUG start codons is associated with multiple cancers and neurodegenerative diseases, so how it is regulated and induced is of interest to biomedical researchers. In a study of initiation at non-AUG start codons, Kearse et al (2019) used reporters with distinct start codons and ribosome profiling to

examine how translation from non-AUG start codons responds to protein synthesis inhibitors in human cells. Expression from multiple reporters with non-AUG start codons was resistant to cycloheximide (CHX), a protein synthesis inhibitor that severely slows but does not completely abrogate elongation (Kearse et al., 2019). The resistance of these reporters initiated at non-AUG start codons to CHX suggests that slowly elongating ribosomes can trigger queuing or stacking of subsequent scanning PICs and/or translating ribosomes, enabling preferential enhancement of recognition of weak non-AUG start codons. Similar to findings with the *AZINI* (Ivanov et al., 2018) and *GGP* (Fig. 3.21) uCCs, Kearse et al. (2019) also found that limiting PIC formation or scanning sensitizes non-AUG-initiated translation to CHX, probably by reducing queuing potential. The Kearse study supports the uCC queuing model where a stalled, or slowed elongating ribosome, leads to queuing of subsequent scanning and translating ribosomes to enhance initiation at the weak start codon (Fig. 4.1) (Ivanov et al., 2018; Kearse et al., 2019).

The uCC represents an interesting target for modifying gene expression. In plant biology, control of *GGP* expression is widely studied as a mechanism to increase vitamin C production in agricultural products such as grains, fruits, and vegetables (Broad et al., 2019; Zhang et al., 2018; Zhang, 2013). In a study of four different plant mRNAs containing uORFs: *AtBR11*, *A. thaliana* phytohormone brassinosteroid receptor; *AtVCT2*, *A. thaliana* GGP; *LsGGP1*, lettuce GGP; and *LsGGP2*, a second lettuce GGP, deleting the 5' portion of the uORF/uCC resulted in dramatically increased production of the mORF-encoded protein, either phytohormone brassinosteroid receptor or GGP (Zhang et al., 2018), presumably by impeding the uCC mechanism of regulation. Notably, editing the *AtVCT2* or *LsGGP1* uCC to prevent its translation increased ascorbate concentrations by

almost 2-fold in plant leaves(Zhang et al., 2018). Removing the uCC from other mRNAs could be used to increase metabolite concentrations; however, in some cases the loss of regulation might impair fitness or survival.

In addition to removing uCCs to derepress enzyme expression in the presence of the regulatory metabolite, uCCs could be used to repress target gene expression. The *GGP* uCC and *AZINI* uCC repress the native downstream mORF and also function in heterologous mRNAs to control mORF translation. A uCC could be inserted upstream of different target genes to regulate expression in the presence of the metabolite sensed by the uCC. In this scenario, the gene of interest would be expressed at some high level by default but could be turned off by increasing the concentration of a metabolite. Increasing the concentration of a metabolite could be complicated in a whole organism if the metabolite is essential, so finding complementary pairs of novel metabolites (drugs) and uCCs translation would be useful. In bacteria, macrolide antibiotics are known to cause peptide-specific stalls (Ito and Chiba, 2013; Wilson et al., 2016). One example is the *ErmCL* leader peptide where translating ribosomes stall in the presence of erythromycin; however, it is unclear whether the macrolide antibiotic-dependent stalling of *ErmCL* leader peptide translation would function with eukaryotic ribosomes. However, the ribosome PTC and exit tunnel are highly conserved, and eukaryotic arrest peptides have been shown to function in multiple systems. For example, the ascorbate regulation of the *GGP* uCC functions in mammalian cells and the RRL *in vitro* translation system even though the *GGP* uCC is only naturally present in plants (Figs 3.2, 3.4 and 3.11). Likewise, the fungal and mammalian *AZINI* uCCs (Ivanov et al., 2018), the fungal AAP (Fig. 3.4) (Fang et al., 2004; Ivanov et al., 2018; Spevak et al., 2010), and the human viral gp48/UL4 uORF2 (Fig

3.16A) (Bhushan et al., 2010) function in various mammalian and fungal systems. Many of the eukaryotic uCC/uORF-encoded peptides had similar stalling motifs to their bacterial counterparts and interacted with homologous rRNA and ribosomal proteins in the peptide exit tunnel, so it is possible that bacterial stalling peptides could work with eukaryotic ribosomes (Bhushan et al., 2010; Bischoff et al., 2014; Ito and Chiba, 2013; Wilson et al., 2016). Finally, small molecules have been found to induce stalling during translation of specific peptides (Li et al., 2019; Liaud et al., 2019). The drug-like molecule PF846 blocks production of the proprotein convertase subtilisin/kexin type 9 (PCSK9), an important target for regulating plasma low-density lipoprotein cholesterol levels, by interfering with translation elongation (Li et al., 2019; Lintner et al., 2017; Petersen et al., 2016). Early studies of PF846 showed that the drug-like molecule selectively stalled ribosomes translating specific peptide sequences (Lintner et al., 2017). A cryo-EM structure of a PF846-stalled human ribosome nascent chain complex showed that PF846 binds in the ribosome exit tunnel in a eukaryotic-specific pocket formed by the 28S rRNA and alters the path of the nascent chain (Li et al., 2019). PF846 can also block termination when proline-rich sequences occur immediately adjacent to the stop codon (Li et al., 2019). Unfortunately, both PF846 and its similar sister compound 8503 have toxic effects due to off-target effects (Liaud et al., 2019); however, this type of drug-like molecule could be optimized to be more specific. The structural studies on these metabolite-induced stalling events leads to questions about how ascorbate and L-Arg trigger translational stalling on the *GGP* uCC and the AAP, and how the sequence of these uCCs confers metabolite specificity.

In summary, my studies on *GGP* translation control, combined with the previous studies on *AZIN1* and *arg-2* regulation, establishes a new paradigm for translational control involving an elongation pause and ribosome queuing to stimulate initiation at an upstream weak translation start site. The uCC has common features with the well-described bacterial leader peptides where ribosome stalling in response to specific metabolites or cellular conditions controls transcript levels or downstream mORF translation. Similar amino acid motifs in the arrest peptides and similar interactions of the nascent chains with the ribosome exit tunnel suggest that they share a common mechanism of translational stalling. Finally, the uCC system poses interesting questions for future studies on how the metabolites are sensed, how ribosome queuing occurs, how ribosome queuing affects mRNA stability, how the ribosome queue enhances initiation at the upstream weak start site, and whether this system can be exploited to manipulate gene expression for biotechnology applications.

5. eIF2 γ mutations and MEHMO syndrome

5.1. Introduction to the role of eIF2 γ in translation

eIF2 is a heterotrimeric eukaryotic translation initiation factor, which is formed by the subunits: eIF2 α , eIF2 β , and eIF2 γ . eIF2 forms a stable ternary complex with GTP and the initiator methionyl-tRNA (Met-tRNA_i^{Met}). The ternary complex along with eIFs 1, 1A, 3, and 5 bind to the 40S ribosomal subunit to form the 43S preinitiation complex (PIC) (Fig. 5.1A)(Hussain et al., 2014; Llacer et al., 2015). In the scanning model of translation initiation, the 43S PIC attaches to an mRNA near the 5' cap and then linearly scans the mRNA leader in a 3' direction in search of an AUG start codon in a suitable context (Hinnebusch, 2011). Base pairing interactions between the anticodon loop of tRNA_i^{Met} in the 43S complex and an AUG codon signals the end of scanning and triggers the eIF5-dependent irreversible hydrolysis of the GTP bound to eIF2 (Hinnebusch, 2011). Following GTP hydrolysis, eIF2-GDP and many of the other initiation factors are released from the 43S complex. Subsequent binding of the GTPase eIF5B promotes joining of the large 60S ribosomal subunit to form the 80S initiation complex, where Met-tRNA_i^{Met} is base paired to the AUG start codon in the P site. After GTP hydrolysis by eIF5B, this terminal 80S complex is ready to begin the elongation phase of protein synthesis (Hinnebusch, 2014).

Active eIF2-GTP is regenerated from eIF2-GDP by the guanine exchange nucleotide exchange factor eIF2B, and this recycling reaction is a key regulatory point for eukaryotic translation. Phosphorylation of the eIF2 α subunit converts eIF2 into a competitive inhibitor of eIF2B, limiting the availability of eIF2-GTP for ternary complex formation (Hinnebusch, 2005). This inhibition of eIF2B results in impaired general translation but enhanced translation of a specific subset of mRNAs in a process known as

the integrated stress response (Baird and Wek, 2012; Harding et al., 2000). In mammalian cells four distinct protein kinases share a common eIF2 α kinase domain linked to different regulatory domains that trigger activation of the kinases in response to different stress signals. PERK, one of these four eIF2 α kinases, feeds into the integrated stress response by sensing ER stress and triggering the translational branch of the unfolded protein response to reduce the burden of misfolded protein accumulation. The decrease in ternary complex levels that accompanies eIF2 α phosphorylation impairs global translation, providing time for the cell to cope with the stressor, while increasing the translation of select mRNAs containing regulatory uORFs including the mRNA encoding the transcription factor ATF4. The yeast *S. cerevisiae* contains only a single eIF2 α kinase, GCN2, and phosphorylation of eIF2 α derepresses translation of the mRNA encoding the transcription factor GCN4 (Hinnebusch, 2014). The integrated stress response and translational control of *GCN4* are useful *in vivo* reporters of eIF2 function in mammalian and yeast cells, respectively (Alone et al., 2008; Baird and Wek, 2012; Borck et al., 2012; Dever et al., 1995; Erickson et al., 1997; Harding et al., 2003; Hinnebusch, 1985; Shin et al., 2011).

As there are no protein structures available for human eIF2 or the isolated eIF2 γ subunit, cryo-EM images of the yeast 48S PIC give the most insight into the structure of eIF2 and interactions with other PIC components (Hussain et al., 2014; Llacer et al., 2015). As shown in Figure 5.1A and B, eIF2 ternary complex interacts with the intersubunit face of the 40S subunit. The γ subunit consists of three domains: an N-terminal GTP binding domain (G) and two β -barrel domains DII and DIII (Fig. 5.1B and C). The γ subunit is the center of the eIF2 complex and ternary complex, and eIF2 γ has binding sites for eIF2 α ,

eIF2 β , Met-tRNA^{Met}, and GTP. The eIF2 β subunit binds to the G domain on the face opposite the side that contacts GTP (shown in dark blue in Fig. 5.1B). Domain DII of eIF2 γ provides the binding site for

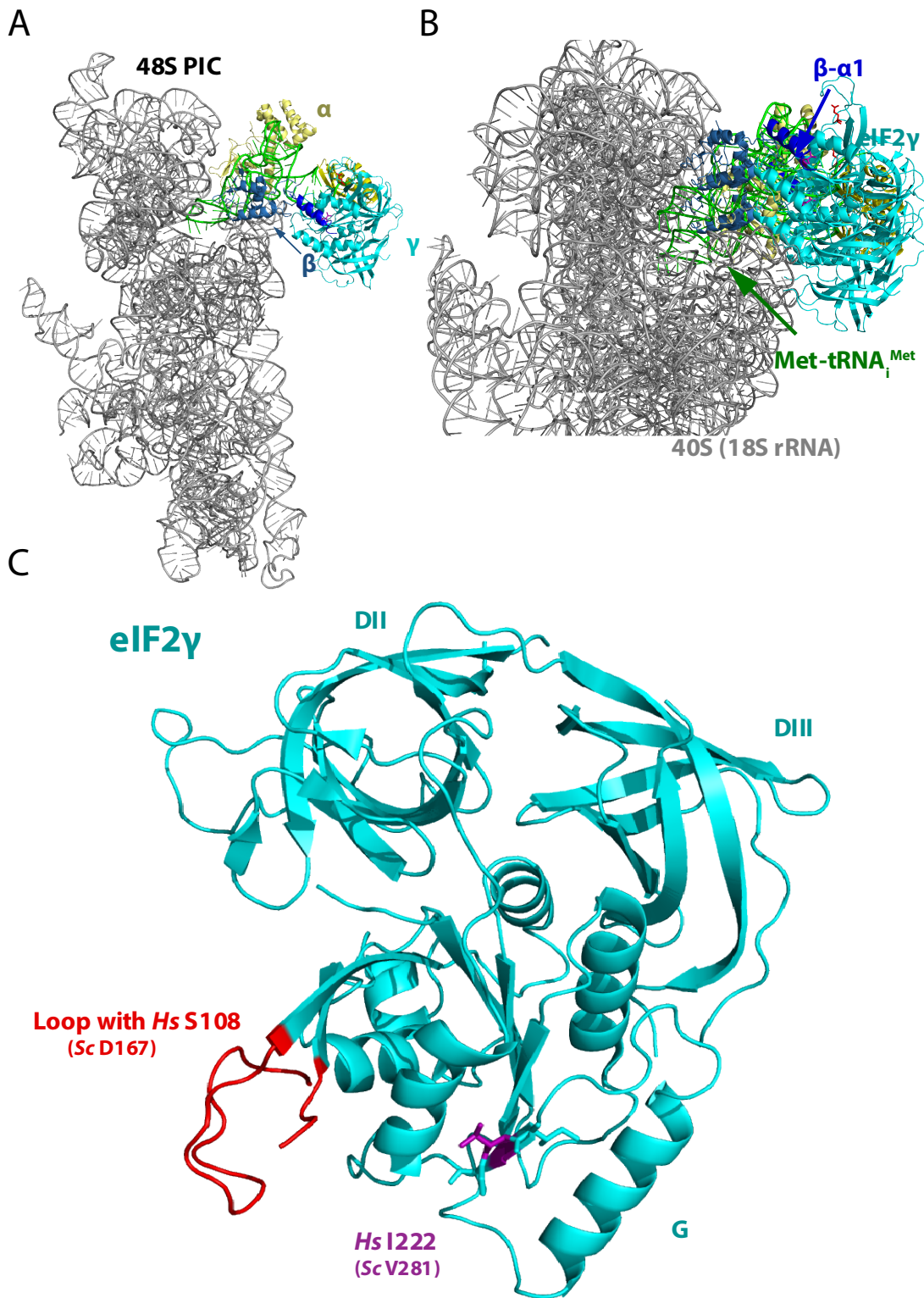


Figure 5.1. Putative MEHMO mutation S108R maps to an unstructured loop in eIF2 γ

Figure 5.1. Putative MEHMO mutation S108R maps to an unstructured loop in eIF2 γ

(A) Ribbons representation of the yeast 48S preinitiation complex (PIC; PDB 3JAP). Only the 18S rRNA of the 40S subunit (grey), Met-tRNA_i^{Met} (green), eIF2 α (N-terminal domains, light yellow; C-terminal domain, yellow), eIF2 β (helix α 1, blue; C-terminal domain, slate), and eIF2 γ (cyan) are shown. (B) The image from panel A rotated and zoomed to highlight the eIF2 γ subunit. The sites of two mutations identified in patients are highlighted: *Homo sapiens* (*Hs*) S108 (corresponding to *S. cerevisiae* *Sc* D167; red) and *Hs* I222 (*Sc* V281; purple). (C) Zoom and rotation of eIF2 γ from panel B. The other eIF2 subunits and components of the 48S complex have been removed for clarity. The sites of the eIF2 γ mutations are colored as in panel B and labeled.

eIF2 α (light yellow), and the acceptor arm of Met-tRNA_i^{Met} binds in a groove of eIF2 γ between the G domain and domains DII and DIII.

In 2012, Borck et al. reported a case where three male family members (two brothers and a maternal uncle) were affected with an X-linked intellectual disability (XLID) (Borck et al., 2012). The patients exhibited a variety of clinical features including moderate to severe intellectual disability, microcephaly, short stature, and facial dysmorphic features. In addition, one brother had seizures and the uncle had microgenitalism and obesity. A single variant, c.665T>C, was detected in the X-chromosome gene *EIF2S3* that encodes humans eIF2 γ . The gene variant was reported to result in an isoleucine-to-threonine missense substitution, p.Ile222Thr (eIF2 γ -I222T).

Yeast structural data (Fig. 5.1 B and C, I222 shown in purple) and biochemical studies of human cells overexpressing eIF2 γ -I222T and of yeast with the analogous mutation showed a disruption in eIF2 β binding to eIF2 γ (Borck et al., 2012). The yeast eIF2 γ mutation impaired translation and reduced the stringency of translation start site selection *in vivo* in a manner that was suppressed by overexpressing eIF2 β . These results thus link intellectual disability to impaired translation initiation through the partial loss of eIF2 function.

The clinical manifestation of the family studied by Borck et al. is very similar to MEHMO syndrome (Borck et al., 2012; Steinmuller et al., 1998). MEHMO syndrome is a rare severe X-linked neurodevelopmental disorder (MIM#300148), characterized by intellectual disability and clinical symptoms including epileptic seizures, hypogonadism, hypogonitalism, microcephaly, and obesity. Steinmuller et al. (1998) coined the term MEHMO in a study of a large family with five affected boys and identified the disease locus as the short arm of chromosome X (Xp11.3-22.13). Though Steinmuller et al. (1998) were not able to determine the genetic cause of the condition, based on the X linkage and clinical similarities with the family described by Borck et al. (2012), we suspected that MEHMO syndrome is caused by mutations in eIF2 γ .

Rare diseases can be hard to study in humans, so model organisms are commonly used. Yeast are an advantageous model organism for rare disease studies because they are simple to genetically manipulate (Smith and Snyder, 2006; Yang et al., 2017). In yeast, eIF2 γ is essential for viability and is encoded by *GCD11* (Hannig et al., 1993). Mutations in *GCD11* alter translational regulation of *GCN4* and the selection of the start site for protein synthesis (Dorris et al., 1995). In this study, I utilized yeast functional assays to assess a novel maternally inherited missense *EIF2S3* variant (c.324T>A). The substitution

changes a conserved serine codon to arginine (p.Ser108Arg or eIF2 γ -S108R). The serine residue is conserved in human, rat, mouse, chicken, frog, and *D. melanogaster*, but is not conserved in budding yeast (Skopkova et al., 2017). The analogous residue to S108 in human eIF2 γ is D167 in the yeast factor. This residue is located in a flexible loop that is part of a zinc-binding ribbon in the G domain (Fig. 5.1 C, shown in red) (Hussain et al., 2014; Llacer et al., 2015; Roll-Mecak et al., 2004).

5.2. Materials and methods

5.2.1. *GCD11* mutant plasmids

The plasmid pC2860, a derivative of the single copy *LEU2* vector YCplac111 and encoding a His8-tagged derivative of yeast eIF2 γ (*GCD11*) under control of the native promoter (Alone et al., 2008) served as the starting point for all mutant constructs. The His8 tag was inserted between codons 2 and 3 of the *GCD11* open reading frame. Site directed mutagenesis, conducted by Joo-Ran Kim, was used to generate various mutant constructs (Table 5.1). A series of four mutations were made at *GCD11* residue D167 to generate: a *GCD11* mutant with the wild-type human amino acid residue eIF2 γ -D167S; the human mutation eIF2 γ -D167R; and eIF2 γ -D167K and eIF2 γ -D167W, two mutations that were predicted to generate a more severe phenotype based on the size and charge of their side-chains. The residues flanking D167 in yeast eIF2 γ are not homologous to human eIF2 γ , so a series of mutants were made where *GCD11* residues 164-173 were swapped for their corresponding human eIF2 γ residues, generating His₈-*GCD11*-(human 164-173), referred to as ‘swap’ for simplicity. The plasmid pC5346 (swap) was used to generate three mutations; the eIF2 γ -human₁₆₄₋₁₇₃-S167R (swap-S167R) mutation that corresponds to the patient mutation; and the eIF2 γ -human₁₆₄₋₁₇₃-S167K (swap-S167K) and eIF2 γ -human₁₆₄₋₁₇₃-S167W (swap-S167W) mutations.

Table 5.1. The yeast *GCD11* plasmids used in this study

Description	Plasmid ID	Source
<i>His₈-GCD11, LEU2</i>	pC2860	Alone et al., 2008
<i>His₈-GCD11-D167S</i>	pC5342	This study
<i>His₈-GCD11-D167R</i>	pC5343	This study
<i>His₈-GCD11-D167K</i>	pC5344	This study
<i>His₈-GCD11-D167W</i>	pC5345	This study
<i>His₈-GCD11-human₁₆₄₋₁₇₃</i> (swap)**	pC5346	This study
<i>His₈-GCD11-human₁₆₄₋₁₇₃** -</i> <i>S167W</i>	pC5347	This study
<i>His₈-GCD11-human₁₆₄₋₁₇₃</i> <i>** -S167K</i>	pC5348	This study
<i>His₈-GCD11-human₁₆₄₋₁₇₃</i> <i>** -S167R</i>	pC5349	This study
<i>His₈-GCD11-V281K</i>	pC4187	Borck et al., 2012
<i>His₈-GCD11-V281T</i>	pC4188	Borck et al., 2012

Table 5.1. ** For simplicity the yeast to human 164-173 substitutions are referred to as ‘swap’ in the figures.

5.2.2. Yeast transformation

S. cerevisiae cells were transformed using the lithium acetate (LiOAc)/single-stranded carrier DNA/PEG method (Gietz and Schiestl, 2007). Briefly, stock solutions of LiOAc (Table 5.2) and polyethylene glycol (PEG) (Table 5.3) were prepared as follows.

Table 5.2. Stock 10X LiOAc solution

1 M LiOAc/TE (10X)	
Lithium acetate	10.2 g
1X TE pH 7.5	100 ml

Table 5.2. Filter sterilize the 10X solution before storage. To make the 1X (0.1M) working solution, dilute the 10X stock solution in 1X TE pH 7.5 and filter sterilize.

Table 5.3. The 44% PEG stock solution

44% PEG	
PEG 3350	220 g
1X TE pH 7.5	310 ml

Table 5.3. To sterilize, autoclave the 44% PEG stock solution. To make the 40% PEG working stock, mix 45 ml of 44% PEG with 5 ml of 10X lithium acetate in TE.

The day prior to transformation, inoculate the yeast strain(s) in three ml of yeast extract peptone dextrose (YPD) medium and incubate overnight at 30°C. The next day, dilute the overnight culture 1:50 in fresh YPD medium (1 ml culture in 49 ml YPD in a 250 ml flask). Incubate at 30°C with shaking for three to five hr until OD₆₀₀ is 0.5-0.8. Pellet the cells in a 50 ml Falcon tube by centrifugation for 5 min at 1462 x g, and then discard the supernatant. Wash the cell pellet with five to ten ml of 1X LiOAc/TE, re-pellet the cells by centrifugation for five min at 1462 x g, and then discard the supernatant. Resuspend the cells in one ml 1X LiOAc/TE, then transfer the cells to a microcentrifuge tube and incubate at 30°C for 30 min. While the cells are incubating, prepare the DNA for

transformation by mixing 5-10 μ l of calf thymus DNA, 25-50 ng of plasmid DNA, and 350 μ l of 40% PEG in a microcentrifuge tube. Following the incubation, pellet the cells at 13,000 x g for 30 s in a microcentrifuge and resuspend the pellet in 1/100 of the original yeast culture volume in 1X LiOAc/TE (e.g. use 500 μ l 1X LiOAc/TE for the pellet from a 50 ml culture). Add 50 μ l of the competent cells to each of the tubes containing DNA and then mix by vortexing. Incubate the DNA and competent cell mixture at RT for 30 min. Then, mix the samples on a vortex, heat shock by incubating at 42°C for 15 min, and then spin down the cells and resuspend the pellet in 200 μ l synthetic minimal medium. Finally, spread the cells on YPD plates to recover overnight at 30°C. The next day, replica plate onto restrictive plates to select for transformants.

5.2.3. Generation of *GCD11* mutant yeast strains

As *GCD11* is essential for yeast viability, plasmid shuffling (Boeke et al., 1987) was used to introduce mutant forms of eIF2 γ into a yeast strain lacking chromosomal *GCD11*. The *gcd11 Δ* strain J292 (*MAT α* , *leu2-3*, *leu2-112*, *ura3-52*, *his3*, *gcn2 Δ ::loxP*, *gcd11 Δ ::KanMX*, p[*GCD11*, *URA3*]) was used for all studies. J292 contains *GCD11* on a plasmid with a *URA3* marker that allows for negative selection after the introduction of a second copy of *GCD11* on a different (*LEU2*) plasmid. Following introduction of a *GCD11* mutant on a *LEU2* plasmid, plasmid shuffling was used to select for cells that lost the wild-type *GCD11*, *URA3* plasmid (Boeke et al., 1987). In short, yeast strain J292 was transformed with each of the 13 *gcd11* mutant plasmids. Clonally purified transformants were patched on synthetic minimal plates supplemented with histidine, and then replica printed to medium containing 5-fluororotic acid (5-FOA). Cells that have lost the *GCD11*, *URA3* plasmid will grow on the 5-FOA medium. The 5-FOA-resistant colonies were

clonally purified by streaking on synthetic minimal plates supplemented with uracil and histidine. The purified strains carrying a mutant *GCD11*, *LEU2* plasmid and lacking the wild-type *GCD11*, *URA3* plasmid were used for subsequent studies (Table 5.4). The resulting strains were colony purified and then patched on plates. For growth assays, the strains were streaked on synthetic minimal plates supplemented with uracil and histidine and grown at 30°C for three days.

Table 5.4. The yeast strains generated in this study

Yeast genotypes	Strain ID	Source
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[GCD11, URA3]</i>	J292	(Borck et al., 2012)
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11, LEU2]</i>	J1440	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-D167S, LEU2]</i>	J1441	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-D167R, LEU2]</i>	J1442	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-D167K,LEU2]</i>	J1443	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-D167W,LEU2]</i>	J1444	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-(human₁₆₄₋₁₇₃), LEU2]</i>	J1445	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-human₁₆₄₋₁₇₃-S167W, LEU2]</i>	J1446	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-human₁₆₄₋₁₇₃-S167K,LEU2]</i>	J1447	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-human₁₆₄₋₁₇₃-S167R,LEU2]</i>	J1448	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-V281K,LEU2]</i>	J1449	This study
<i>Mata leu2-3, leu2-112, ura3-52, his3, gcn2Δ, gcd11Δ::KanMax p[His8-GCD11-V281T,LEU2]</i>	J1450	This study

Table 5.4. For simplicity, the yeast to human 164-173 substitutions are referred to as ‘swap’ in the figures.

5.2.4. β -galactosidase assay

For the measurement of *GCN4-lacZ* expression, derivatives of yeast strain J292 expressing wild-type or mutated forms of *GCD11* were transformed with the plasmid p180 (Table 5.5). Transformants were grown in synthetic minimal medium supplemented with histidine to OD \sim 0.8, and β -galactosidase activities were determined as described below. To measure translation start site selection stringency, the derivatives of yeast strain J292 were transformed with the plasmids p3989 (*His4(AUG)-lacZ*) or p3990 (*His4(UUG)-lacZ*) (Table 5.5). Transformants were grown to OD \sim 0.8, and β -galactosidase activities were determined as described below.

Table 5.5. β -galactosidase assay lacZ reporters

Reporter	Plasmid ID	Source
<i>GCN4-lacZ, URA3</i>	p180	(Hinnebusch, 1985)
<i>His4(AUG)-lacZ, URA3</i>	p3989	(Cigan et al., 1988b)
<i>His4(UUG)-lacZ, URA3</i>	p3990	(Cigan et al., 1988b)

To initiate β -galactosidase assays, triplicate five ml pre-cultures in synthetic minimal medium supplemented with histidine were incubated overnight at 30°C with shaking. The next day, the pre-cultures were used to inoculate 5 ml cultures of synthetic minimal plus histidine medium at an OD of 0.5 in 14 ml round bottom plastic Falcon tubes. The cultures were incubated for six hr at 30°C with shaking at 250 rpm, then the cells were pelleted at 1462 x g, transferred to a 1.5 ml microcentrifuge tube, and the cell pellets stored at -80 °C overnight.

The breaking buffer (Table 5.6) was chilled on ice, and chilled Z-buffer (Table 5.7) was prepared by adding 0.75 g of ortho-Nitrophenyl- β -galactoside (ONPG) per liter of Z-buffer. The frozen cells were resuspended in 200 μ l breaking buffer (Table 5.6) and lysed using two-thirds volume glass beads. The cell suspension was mixed on a vortex for 10 min at 4°C. The supernatant was transferred to a clean chilled microcentrifuge tube and the solution was clarified by centrifugation at maximum speed for 15 min at 4°C. The supernatant was transferred to a cold 1.5 ml microcentrifuge tube.

Total protein concentration was measured using a Bradford assay. Bradford Reagent (BIO-RAD) was prepared by mixing one-part 5X reagent with four parts water and filtering. Each sample was then prepared by adding one ml of 1x Bradford reagent and 1 μ l of sample and then mixed on a vortex. The samples were incubated for five min at RT and then the absorbance was measured at 595 nm. To calculate the protein concentration, a set of protein standards containing 0.2 mg/ml, 0.4 mg/ml, 0.6 mg/ml, 0.8 mg/ml and 1 mg/ml BSA were used to create a standard curve.

Two hours before beginning the β -galactosidase assays, an ortho-Nitrophenyl- β -galactoside (ONPG) solution was prepared by mixing 0.2 g of ONPG in 50 ml Z-buffer (Table 5.7). For the β -galactosidase assays, 950 μ l of Z-buffer was added to a 5 ml round bottom falcon tube and then 50 μ l of sample was added. The samples were incubated for five minutes at 28 °C and then 200 μ l ONPG solution was added to each sample and the start time for each sample was recorded. The samples were incubated at 28 °C until pale yellow, the time was recorded, and the reaction was terminated by adding 500 μ l Na₂CO₃ stock solution (5.3 g Na₂CO₃ in 50 ml water). The absorbance was measured at 420 nm. The β -galactosidase concentration was calculated using the following formula:

$(\text{absorbance at } 420 \text{ nm} * 1.7) / (0.0045 * \text{protein concentration} * \text{volume}(\text{ml}) * \text{time}(\text{min}))$

protein concentration = absorbance at 595 nm / 0.069, where 0.069 is the slope generated by the BSA standard curve.

Table 5.6. Composition of the β -galactosidase assay breaking buffer

Breaking Buffer	
Reagent	Volume (40 mls)
ddH ₂ O	28 ml
1 M Tris pH 8.0	4 ml
20% Glycerol	8 ml
BME	2.88 μ l
40 mM AEBSF**	400 μ l

Table 5.6. ** (2-aminoethyl)benzenesulfonyl fluoride hydrochloride serine protease inhibitor

Table 5.7. Composition of the β -galactosidase assay Z-buffer

Z-Buffer	
Reagent	Volume (250 mls)
Na ₂ HPO ₄	4.03 g
NaH ₂ PO ₄	1.38 g
2 M KCl	1.25 ml
MgSO ₄	0.062 g
water	To 250 ml

5.3. Results

A male patient displaying microcephaly, hypospadias and cryptorchidism, developmental delay and obesity was given a probable diagnosis of MEHMO syndrome by Dr. Clesson Turner, M.D. at Walter Reed National Military Medical Center (Skopkova et al., 2017). While the patient did not have all the clinical features of MEHMO syndrome, having intellectual disability, hypogonadism, microcephaly, and obesity, but not epilepsy, whole exome sequencing revealed that he had a novel maternally inherited missense *EIF2S3* variant (c.324T>A). The substitution changes a conserved serine codon to arginine (p.Ser108Arg). The serine residue is conserved in human, rat, mouse, chicken, frog, and *D. melanogaster* but is not conserved in budding yeast (Skopkova et al., 2017). The analogous residue to S108 in human eIF2 γ is D167 in the yeast factor. This residue is located in a flexible loop that is part of a zinc-binding ribbon in the G domain (Fig. 5.1 C, shown in red) (Hussain et al., 2014; Llacer et al., 2015; Roll-Mecak et al., 2004). This region is not fully resolved in the yeast cryo-EM structure, but it does not appear to interact with other components of the 48S PIC (Fig. 5.1 B), and the function of the Zn²⁺-binding element in eIF2 γ is unknown.

To test the impact of the eIF2 γ -S108R substitution on eIF2 function, a yeast/human (*Sc/Hs*) chimeric form of eIF2 γ was generated by replacing yeast eIF2 γ amino acid residues 164-173 with the corresponding residues from human eIF2 γ (Fig. 5.2). In this chimeric protein, S167 corresponds to the native human eIF2 γ residue S108. The S167 residue was mutated to Arg to correspond with the patient S108R mutation or to Trp or Lys. The latter mutations, which are anticipated to be more severe than the Arg substitution, were generated in case the eIF2 γ -S167R mutation does not confer a phenotype in yeast. In

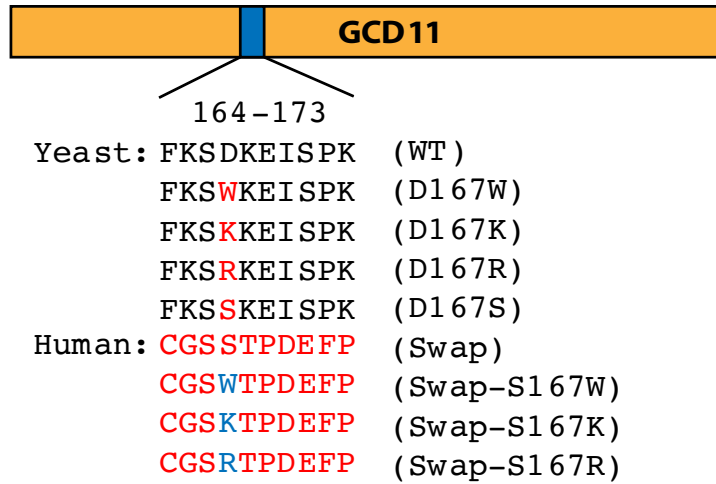


Figure 5.2. Schematic of the eIF2 γ (yeast GCD11) mutations generated for this study

Figure 5.2. Schematic of the eIF2 γ (yeast GCD11) mutations generated for this study

To study the impact of S108 mutations in human eIF2 γ , mutations were made at the corresponding D167 residue in yeast eIF2 γ (GCD11) and in a human-yeast chimera. (Upper) GCD11 is displayed as a rectangle with the targeted region highlighted in blue. The wild-type yeast sequence for residues 164-173 is depicted in black and amino acid substitutions (red) were made at residue D167 to generate the normal human residue S108 (D167S), and the mutations D167R, D167K, and D167W. A chimeric (swap) mutant was made by substituting the human residues 105-114 in place of the corresponding yeast residues 164-173 (Swap). The same S167 amino acid substitutions (blue) were made in this chimeric allele.

addition to mutating the yeast-human chimeric allele, the D167 residue in the native yeast *GCD11* gene, which corresponds to S108 in the human protein, was mutated to Ser, the native human residue, as well as to Arg, the patient's mutation, and to Trp and Lys. The yeast strain J292 was transformed with the plasmids containing wild-type or mutant forms of *GCD11*, followed by loss of the original plasmid-borne GCD11 plasmid, so that only the introduced eIF2 γ was expressed (Tables 5.1 and 5.4). In addition to the yeast mutants corresponding to the human eIF2 γ -S108R mutant, mutant strains expressing two other previously described eIF2 γ mutants were also generated, eIF2 γ -V281K and eIF2 γ -V281T, which correspond to mutations at human eIF2 γ residue I222 (Borck et al., 2012). The eIF2 γ -V281K and eIF2 γ -V281T mutants were included as controls because the mutations

were previously shown to affect growth, *GCN4-lacZ* reporter expression, and translation start site selection stringency (Borck et al., 2012).

Yeast expressing the eIF2 γ -D167S or the eIF2 γ -human₁₆₄₋₁₇₃ (swap) as the sole form of eIF2 γ grew like yeast expressing wild-type yeast eIF2 γ with no discernable growth defect at 30°C (Fig. 5.3). Similarly, yeast expressing the eIF2 γ -D167R and the eIF2 γ -human₁₆₄₋₁₇₃-S167R mutants, corresponding to the patient mutation, had no discernable growth defect. Yeast expressing the eIF2 γ -D167K and eIF2 γ -D167W mutations also had no discernable growth defect. Finally, like the other mutations at residue 167, the eIF2 γ -human₁₆₄₋₁₇₃-S167W and eIF2 γ -human₁₆₄₋₁₇₃-S167K mutants also conferred no discernable growth defect when expressed as the sole form of eIF2 γ in yeast. As a positive control in this growth assay, I tested the yeast eIF2 γ -V281T and eIF2 γ -V281K mutants, corresponding to the human eIF2 γ -I222T mutation identified in the previously described patients with XLID (Borck et al., 2012). As described previously, the eIF2 γ -V281K showed a striking slow growth phenotype at 30°C (Fig. 5.3).

To further assess the impact of the human eIF2 γ -S108R mutation on eIF2 γ function, a *GCN4-lacZ* reporter was introduced into isogenic yeast strains expressing: wild-type-eIF2 γ ; the yeast eIF2 γ -D167S, eIF2 γ -D167R, eIF2 γ -D167K, and eIF2 γ -D167W mutations; the eIF2 γ -human₁₆₄₋₁₇₃(swap), eIF2 γ -human₁₆₄₋₁₇₃-S167R, eIF2 γ -human₁₆₄₋₁₇₃-S167K, and eIF2 γ -human₁₆₄₋₁₇₃-S167W; or as a positive control, yeast eIF2 γ -V281T and eIF2 γ -V281K, which were previously shown to impair eIF2 function and cause derepression of *GCN4* expression (Borck et al., 2012). The *GCN4-lacZ* reporter system is a useful tool to measure eIF2 function. In yeast, phosphorylation of eIF2 α by the protein

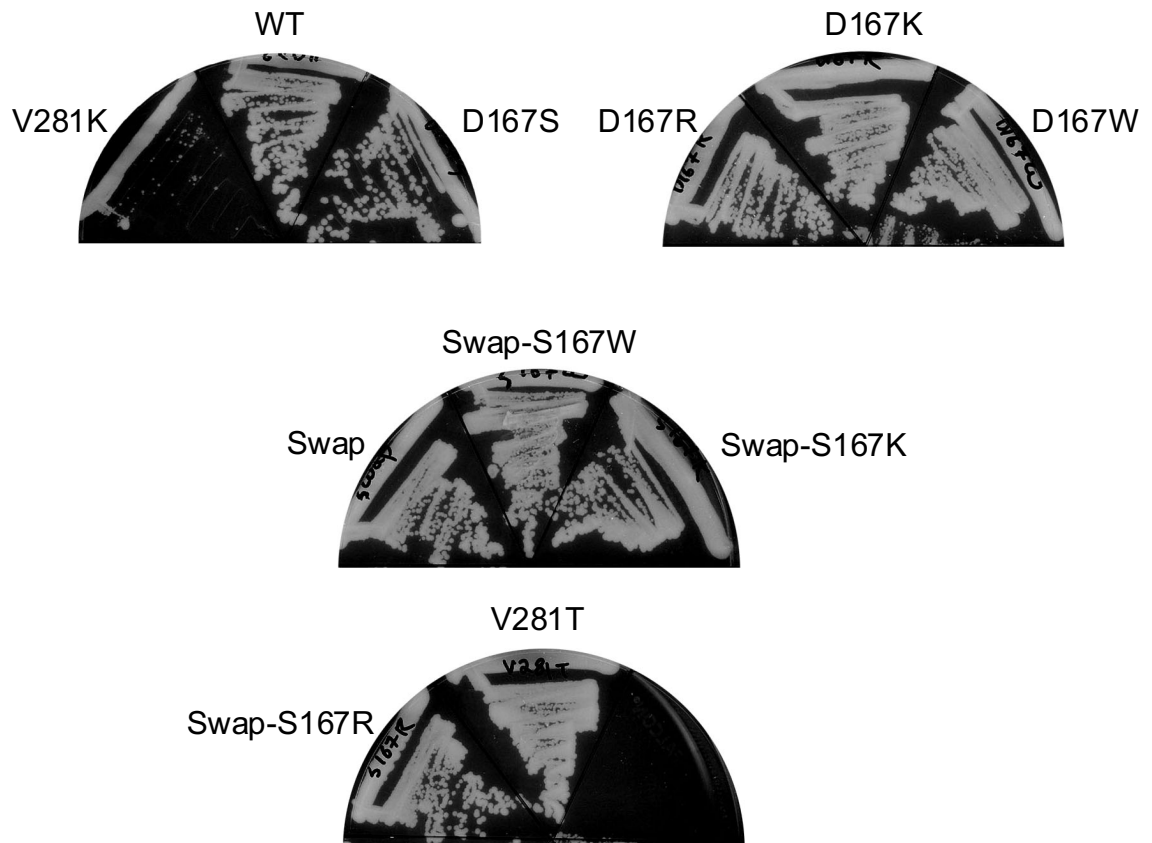


Figure 5.3. Mutations at residue D167 of yeast eIF2 γ do not impair cell growth

Figure 5.3. Mutations at residue D167 of yeast eIF2 γ do not impair cell growth

Yeast expressing the indicated eIF2 γ mutants, as the sole source of eIF2 γ , were streaked on synthetic minimal medium and incubated at 30 °C for 3 days.

kinase GCN2 mediates translational control of the transcriptional activator *GCN4* (Dever et al., 1992). The *GCN4* mRNA contains four short uORFs prior to the *GCN4* mORF. Regulated reinitiation on the uORFs, which is sensitive to eIF2 function, controls synthesis of GCN4. Ribosomes bind near the 5' cap of the mRNA, scan down and translate uORF1 (Fig. 5.4A). Following translation of uORF1, the post-termination 40S subunit remains attached to the mRNA and resumes scanning (Hinnebusch, 2011). If ternary complex levels are high, indicating normal eIF2 function, nearly all of the scanning 40S subunits will re-acquire a ternary complex, which enables the scanning ribosome to recognize subsequent start codons and re-initiate translation at uORF2, uORF3, or uORF4. Following translation of these latter uORFs, especially uORF3 and uORF4, the ribosome dissociates from the mRNA and fails to translate the *GCN4* mORF. However, if ternary complex levels are low or ternary complex binding to the ribosome is impaired, the 40S subunits scanning down the mRNA after translating uORF1 are more likely to scan past uORFs 2-4 without reinitiating translation. If these scanning 40S subunits acquire a ternary complex after scanning past uORF4, the ribosomes can then re-initiate translation at the *GCN4* start codon. Thus, mutations that lower ternary complex levels or impair eIF2 function will induce the expression of a *GCN4-lacZ* reporter.

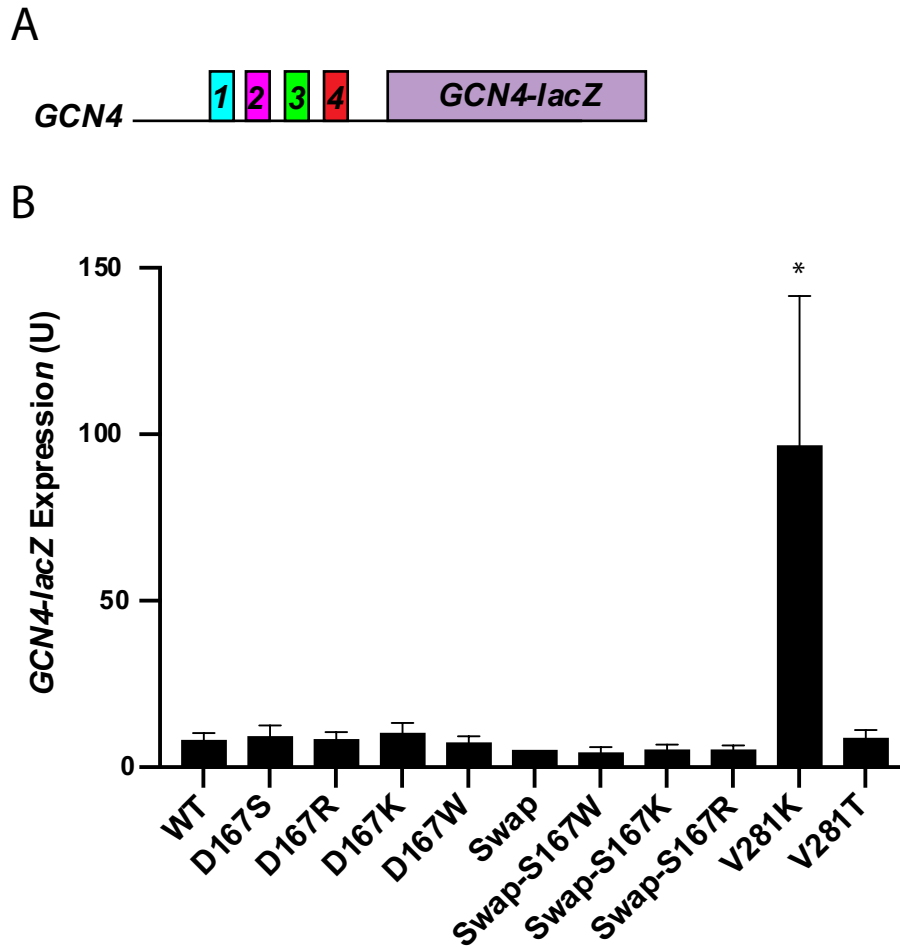


Figure 5.4. Mutations at residue D167 of yeast eIF2 γ do not affect *GCN4* expression

Figure 5.4. Mutations at residue D167 of yeast eIF2 γ do not affect *GCN4* expression

(A) Schematic of *GCN4-lacZ* reporter construct containing the four uORFs that control *GCN4* mRNA translation. (B) Yeast expressing the indicated eIF2 γ mutants, as the sole source of eIF2 γ , were transformed with the *GCN4-lacZ* reporter. β -Galactosidase activities and standard deviations were determined from three independent transformants. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=3$, assayed in duplicate, assessed against wild-type).

As expected, *GCN4-lacZ* expression was low in yeast expressing wild-type eIF2 γ or the eIF2 γ -D167S mutant in which the wild-type human residue was substituted for the native yeast residue (Fig. 5.4B). Likewise, the eIF2 γ -D167R mutant, which mimics the human mutation, did not increase *GCN4-lacZ* expression. Notably, this lack of impact of the patient mutation on *GCN4-lacZ* expression was previously observed for the yeast eIF2 γ -V281T mutation that mimics the human eIF2 γ -I222T mutation that causes MEHMO syndrome. Like the eIF2 γ -D167R mutation, the eIF2 γ -V281T mutation conferred only a mild growth defect and did not increase *GCN4-lacZ* expression (Fig. 5.3 and Fig. 5.4B). As neither of these mutations is lethal in humans, where alterations in translation are expected to be more impactful than in single-celled yeast, perhaps it is not surprising that these patient mutations do not confer a phenotype in yeast. Unexpectedly, the eIF2 γ -D167K and eIF2 γ -D167W mutations, which were chosen with the notion that they might more substantially impair eIF2 function, did not impact *GCN4-lacZ* expression. Similar results were observed when the mutations were introduced into the yeast-human eIF2 γ

chimera. As expected, the wild-type yeast-human chimera (referred to as the Swap mutant) showed wild-type levels of *GCN4-lacZ* expression. Moreover, consistent with the results obtained when mutating native yeast eIF2 γ , the mutations Swap-S167W, Swap-S167K, and Swap-S167R did not impact *GCN4-lacZ* expression. Taken together, these results indicate that mutations at residue D167 in yeast eIF2 γ have minimal or no impact on eIF2 function.

Mutations that impair eIF2 function can also lower translation initiation fidelity and allow ribosomes to initiate translation at non-AUG codons (Hinnebusch, 2011). Translation is normally stringently restricted to initiate at an AUG codon; however, mutations that weaken Met-tRNA_i^{Met} binding to eIF2, enhance eIF2 GTPase activity, or disrupt eIF2 integrity, enhance initiation at non-AUG codons and confer a suppressor of initiation (*Sui*⁻) phenotype in yeast (Huang et al., 1997). Yeast *Sui*⁻ mutations were first identified through the restoration of the expression of *his4* alleles with a non-AUG start codon. *Sui*⁻ mutations enable translation to initiate more efficiently at the third, UUG codon of the *his4* open reading frame, reestablishing growth in histidine-free medium (Yoon and Donahue, 1992). To assess stringency of translation start codon selection, *His4-lacZ* reporter constructs with either an AUG or UUG start codon (Fig. 5.5A) were transformed into yeast strains expressing: wild-type-eIF2 γ ; the yeast eIF2 γ -D167S, eIF2 γ -D167R, eIF2 γ -D167K, and eIF2 γ -D167W mutants; the eIF2 γ -human₁₆₄₋₁₇₃ (swap), eIF2 γ -human₁₆₄₋₁₇₃-S167R, eIF2 γ -human₁₆₄₋₁₇₃-S167K, and eIF2 γ -human₁₆₄₋₁₇₃-S167W mutants; or as positive controls, the yeast eIF2 γ -V281T and eIF2 γ -V281K mutants, which were previously shown to impair eIF2 function and decrease the stringency of translation start codon selection (Borck et al., 2012).

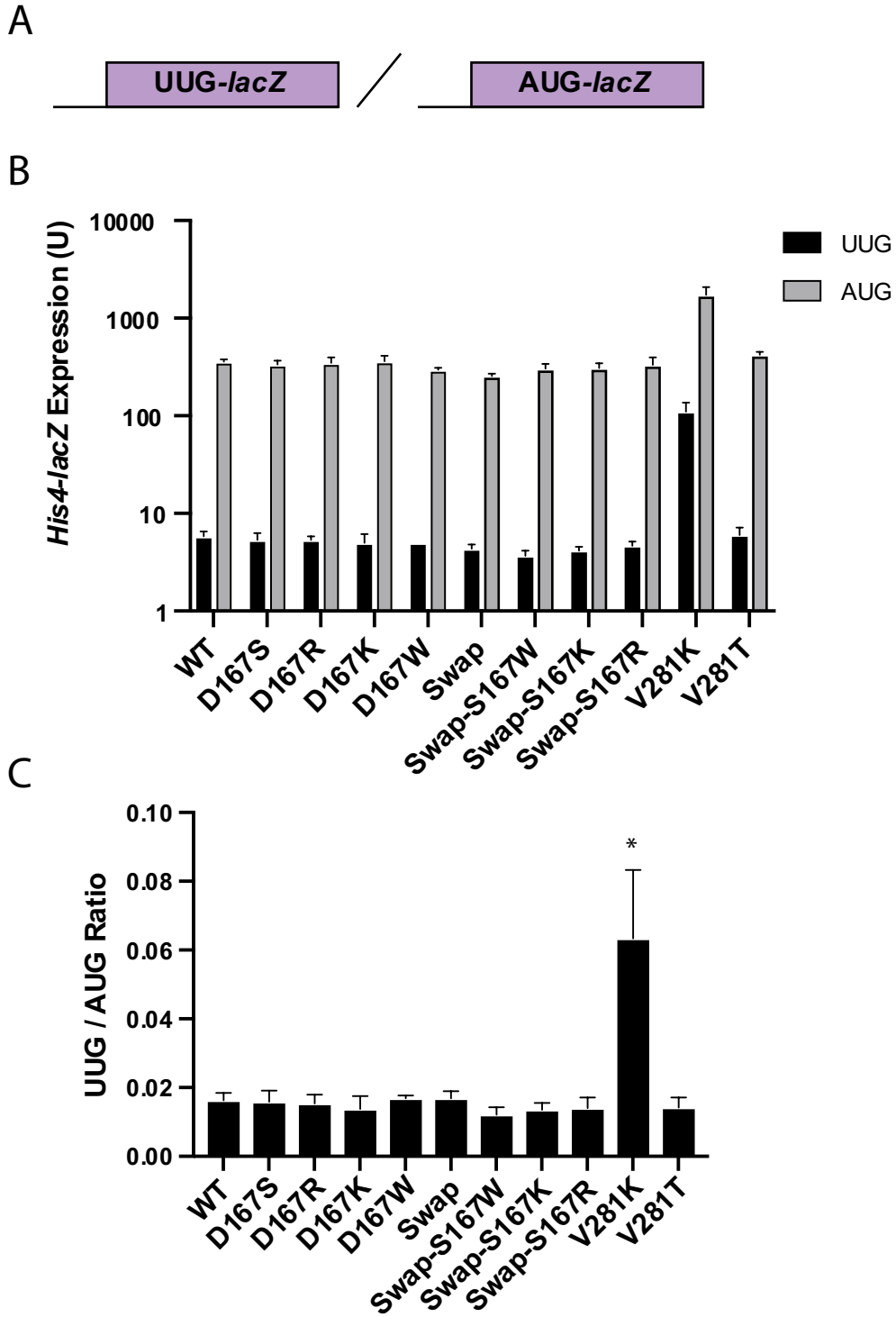


Figure 5.5. Mutations at residue D167 of yeast eIF2 γ do not alter translation start site fidelity

Figure 5.5. Mutations at residue D167 of yeast eIF2 γ do not alter translation start site fidelity

(A) Schematics of *HIS4(AUG)*- and *HIS4(UUG)*-*lacZ* reporter constructs. (B) The *HIS4(AUG)*- and *HIS4(UUG)*-*lacZ* reporter constructs were introduced into yeast expressing the indicated eIF2 γ mutants, as the sole-source of eIF2 γ . β -galactosidase activities (B) and mean ratios of the activities (C) were determined. Error bars represent SD, * $p < 0.005$ (Student's two-tailed t-test; $n=3$, assayed in duplicate, assessed against wild-type (*Sc*)).

Consistent with previous reports, expression of the *His4(UUG)*-*lacZ* reporter was substantially poorer than expression of the *His4(AUG)*-*lacZ* reporter in cells expressing wild-type eIF2 γ , as well as in cells expressing yeast eIF2 γ -D167S and eIF2 γ -human₁₆₄₋₁₇₃ (swap) (Donahue and Cigan, 1988), leading to a low UUG/AUG initiation ratio of less than 0.02 (Fig. 5.5C). Similarly, expression of the *His4(UUG)*-*lacZ* was substantially poorer than expression of the *His4(AUG)*-*lacZ* reporter in cells expressing the eIF2 γ -D167R, eIF2 γ -D167K, and eIF2 γ -D167W mutants, also leading to a low UUG/AUG ratio of less than 0.02. Poor expression of the *His4(UUG)*-*lacZ* compared to the *His4(AUG)*-*lacZ* reporter was also seen in cells expressing eIF2 γ -human₁₆₄₋₁₇₃-S167R (swap-S167R), eIF2 γ -human₁₆₄₋₁₇₃-S167K (swap-S167K), and eIF2 γ -human₁₆₄₋₁₇₃-S167W (swap-S167W). In contrast, expression of the *His4(UUG)*-*lacZ* reporter was substantially increased along with expression of the *His4(AUG)*-*lacZ* reporter in cells expressing the previously described eIF2 γ -V281K mutant. *His4(UUG)*-*lacZ* expression was approximately 10-times

higher in yeast expressing eIF2 γ -V281K compared to yeast expressing wild-type eIF2 γ , leading to an increase in the UUG/AUG initiation ratio from 0.02 in the wild-type to 0.06 in the cells expressing eIF2 γ -V281K, a greater than 3-fold increase. The 3-fold increase in the UUG/AUG ratio in cells expressing eIF2 γ -V281K is consistent with previous findings (Borck et al., 2012); however, unlike as in the previous study, *His4(UUG)-lacZ* expression was not elevated in cells expressing eIF2 γ -V281T. The UUG/AUG ratio remained below 0.02, at a similar level to cells expressing wild-type eIF2 γ . In contrast, the eIF2 γ -V281T mutation was previously shown to increase the UUG/AUG ratio approximately 2-fold (Borck et al., 2012), perhaps reflecting a difference in growth conditions or assay sensitivity in this assay compared to the previous study.

5.4. Discussion and conclusions

Our lab, together with collaborators, recently demonstrated that mutations in eIF2 cause MEHMO syndrome (Skopkova et al., 2017). While the original MEHMO patient exhibited intellectual disability, epilepsy, hypogonadism, microcephaly, and obesity (Steinmuller et al., 1998), the male patient with the c.324T>A (p.Ser108Arg) mutation in *EIF2S3*, the mutation examined in this thesis, has not experienced seizures and has not been diagnosed with epilepsy. The eIF2 γ -S108R mutation described is one of five different mutations in *EIF2S3* that has been linked to MEHMO syndrome to date. In 2017, our collaborators described a 4 bp deletion in the *EIF2S3* gene (c.1394_1397delTCAA) in two families with sons exhibiting MEHMO syndrome phenotypes (Skopkova et al., 2017). Moreover, a separate family in which four males were afflicted with an XLID that was associated with the same 4 bp deletion in *EIF2S3* was reported (Moortgat et al., 2016). Definitive linkage of mutations in *EIF2S3* to MEHMO syndrome was provided when our collaborators obtained archival DNA samples from the index patient. Sequencing of the *EIF2S3* gene from this patient (Steinmuller et al., 1998) revealed the same 4 bp deletion that is predicted to cause a frameshift and premature stop codon (p.Ile465Serfs*4) that impacts the C-terminal eight residues of eIF2 γ (Skopkova et al., 2017). The eIF2 γ -fs mutation decreases eIF2 heterotrimer formation by impairing binding of eIF2 α to eIF2 γ (Young-Baird et al., 2020). The impaired interaction is not due to an alteration in the site on eIF2 γ where eIF2 α binds, because the mutation affects the C-terminal residues of eIF2 γ which are far from the loop in Domain II that was previously shown to interact with eIF2 α (Hussain et al., 2014; Roll-Mecak et al., 2004; Yatime et al., 2007; Young-Baird et al., 2019a){Young-Baird, 2020 #405}. Instead, Young-Baird et al. (2019) propose that the

eIF2 γ -fs mutation alters the interaction of eIF2 γ with the chaperone CDC123, consistent with structural studies that map the CDC123 interaction site to the C-terminus of eIF2 γ (Bieganowski et al., 2004; Perzlmaier et al., 2013). As CDC123 was previously shown to facilitate eIF2 α binding to eIF2 γ (Perzlmaier et al., 2013), the altered interaction with CDC123 likely accounts for defective eIF2 α binding to the eIF2 γ -fs mutant.

In their 2012 paper, Borck et al. described three patients with XLID who carried a c.665T>C mutation in *EIF2S3* that caused the missense mutation eIF2 γ -I222T (Borck et al., 2012). The I222T mutation lies in a hydrophobic cleft on the backside of the GTP binding domain (Fig. 5.1B and C, purple) that forms the binding site for helix α 1 of the eIF2 β subunit and disrupts eIF2 β binding to eIF2 γ (Borck et al., 2012). The disruption in eIF2 β binding is consistent with the homologous archaeal aIF2 complex crystal structures available at the time (Stolboushkina et al., 2008; Yatime et al., 2007), and the subsequent yeast high resolution cryo-EM structures (Hussain et al., 2014; Llacer et al., 2015).

In addition to the family in which four males carried the eIF2 γ -fs mutation, Moortgat et al. (2016) also described a family where two male siblings exhibited severe microcephaly, growth retardation, seizures, and intellectual disability, some of the features of MEHMO syndrome. The family had a c.777T>G mutation causing a p.Ile259Met missense mutation in the domain DII Met-tRNA_i^{Met} binding pocket of eIF2 γ (Moortgat et al., 2016; Young-Baird et al., 2019b). In a study to examine how the eIF2 γ -I259M mutation impairs eIF2 function, Young-Baird et al. (2019b) generated the corresponding mutation in yeast, eIF2 γ -I318M. In yeast, the eIF2 γ -I318M mutation caused derepression of *GCN4* expression and relaxed the stringency of translation start site selection. The eIF2 γ -I318M mutation impairs Met-tRNA_i^{Met} binding to eIF2 (Young-Baird

et al., 2019b). The impairment in Met-tRNA_i^{Met} binding is likely due to an indirect effect because I318 does not directly contact the Met-tRNA_i^{Met}. However, the I318 residue is adjacent to R319, which is part of the Met-tRNA_i^{Met} binding pocket. The eIF2 γ -R319D mutation is lethal in yeast (Roll-Mecak et al., 2004) and Young-Baird et al. (2019) propose that the eIF2 γ -I318M mutation impacts the position of R319 in the binding pocket.

The most recently described mutation in *EIF2S3* was reported in a family with three affected male children, a cousin and twin brothers (Gregory et al., 2019). The children had hypoglycemia, short stature with growth hormone and thyroid stimulating hormone deficiencies, and a pancreatic phenotype. Their phenotype was linked to a c.1294C>T *EIF2S3* mutation (p.Pro432Ser) that alters a highly conserved residue in domain III of eIF2 γ (Gregory et al., 2019). To date, the impact of this mutation on eIF2 function has not been determined, but the corresponding mutation in yeast, eIF2 γ -P490S, resulted in a modest 1.7-fold increase in *GCN4-LacZ* expression and a modestly relaxed stringency of translation start site selection (Gregory et al., 2019).

All five *EIF2S3* mutations that have been linked to ID or MEHMO syndrome have been introduced into the yeast gene to examine the impact on cell growth, *GCN4* expression and the stringency of translation start site. Of the five mutations, only the eIF2 γ -S108R (yeast eIF2 γ -D167R) had no phenotype in any of the yeast assays. The eIF2 γ -I222T (yeast eIF2 γ -V281T) mutation conferred a mild growth defect in yeast, but substituting lysine for the valine residue (V281K) resulted in a severe growth defect (Borck et al., 2012). The C-terminal I465Sfs*4 mutation, eIF2 γ -fs, introduced into a yeast-human chimeric allele in which the C-terminal thirteen residues of yeast eIF2 γ were replaced by the corresponding sixteen residues of human eIF2 γ , caused no growth phenotype at 30 °C but conferred a

mild slow-growth phenotype at 37 °C (Skopkova et al., 2017). The eIF2 γ -I259M mutation (yeast eIF2 γ -I318M mutation) caused a pronounced slow growth phenotype at 30 °C that was nearly as severe as the eIF2 γ -V281K mutation (Young-Baird et al., 2019b). Finally, like the eIF2 γ -D167R mutation, the P432S (yeast eIF2 γ -P490S) mutation did not impact yeast cell growth (Gregory et al., 2019). Thus, despite all five of the *EIF2S3* mutations conferring a phenotype in the patients, the less severe human *EIF2S3* mutations do not have a corresponding slow growth phenotype in yeast.

The *GCN4-lacZ* and UUG/AUG *His4-lacZ* reporter assays are much more sensitive to disturbances in eIF2 function than yeast cell growth. The eIF2 γ -V281T (human eIF2 γ -I222T) mutation lead to over a 2-fold increase in *GCN4-lacZ* expression while the lysine substitution led to a nearly 40-fold increase in *GCN4-lacZ* expression (Borck et al., 2012). The eIF2 γ -V281T mutation also affected translation start site selection stringency with a two-fold increase in the UUG to AUG initiation ratio, while the lysine mutation caused a three-fold increase in the UUG to AUG initiation ratio. Like the eIF2 γ -V281K mutation, the eIF2 γ -fs mutation in the yeast-human chimeric allele caused an approximately two-fold increase in *GCN4-lacZ* expression and also increased the UUG to AUG initiation ratio by about 2-fold (Skopkova et al., 2017). The eIF2 γ -I318M mutation that impairs Met-tRNA binding led to a dramatic 16-fold increase in *GCN4-lacZ* expression that mirrored the effect seen with the V281K mutation (Young-Baird et al., 2019b). The eIF2 γ -I318M mutation also substantially increased the UUG to AUG initiation ratio, indicating that weakened Met-tRNA^{Met} binding to eIF2 both decreases ternary complex levels and enables initiation at non-AUG codons. While neither the D167R mutation nor the P490S mutation in eIF2 γ affected yeast cell growth, the P490S mutation caused modest increases in both *GCN4-*

lacZ expression and the UUG to AUG initiation ratio (Gregory et al., 2019). Taken together, these results reveal that of all of the yeast models of human eIF2 γ mutations, the D167R mutation has the least effect on eIF2 function. However, I make this conclusion cautiously because in my assays the eIF2 γ -V281T mutation, which previously was shown to modestly affect the yeast reporter assays (Borck et al., 2012; Skopkova et al., 2017), also failed to impact *GCN4-lacZ* expression or UUG to AUG initiation ratios (data not shown). This lack of impact of the V281T mutation in my assays might reflect subtle differences in experimental technique that mask the effects of the mild V281K and D167R mutations in eIF2 γ . At odds with this simple explanation, the fact that more extreme amino acid changes at D167, substituting with oppositely charged lysine or bulky tryptophan, likewise had no impact indicates that the residue D167 of yeast eIF2 γ is not as sensitive to amino acid changes as the residues altered by the other eIF2 γ mutations. Due to the fact that the yeast eIF2 γ -D167 mutations do not confer a phenotype, we have no evidence that the human eIF2 γ -S108R mutation is pathogenic. Therefore, we cannot, at this time, definitively conclude that the eIF2 γ -S108R mutation is the cause of the patient's disease.

Mutations in *EIF2S3* affect eIF2 function and may cause pathology through the integrated stress response (Skopkova et al., 2017; Young-Baird et al., 2019a). Induced pluripotent stem cells (iPSCs) derived from a patient with eIF2 γ -fs mutation showed impaired general translation and cell growth (Young-Baird et al., 2019a). Western blot analyses revealed elevated levels of ATF4, CHOP, and GADD34, key components of the integrated stress response pathway, but no increase in eIF2 α -P, indicating that the integrated stress response is induced independently of eIF2 α phosphorylation, likely due to decreased eIF2 activity (Young-Baird et al., 2019a). As described previously,

translational regulation of *ATF4*, *CHOP* and *GADD34* is mediated, in part, by uORFs in the 5' leaders of the transcripts that couple expression to the level of eIF2 α -P and eIF2 activity (Wek, 2018; Young and Wek, 2016). Translational control of *ATF4* is mediated by two uORFs (Harding et al., 2000; Lu et al., 2004; Vattam and Wek, 2004). As in the *GCN4* mRNA, the *ATF4* uORF1 acts as a positive element by enabling downstream translational re-initiation. When eIF2 ternary complex levels are high, ribosomes that resume scanning after translation of uORF1 quickly reacquire a new eIF2 ternary complex and reinitiate translation at uORF2; however, uORF2 overlaps out-of-frame with the *ATF4* coding sequence, and termination at the uORF2 stop codon is 3' of the *ATF4* coding sequence start codon. Translation of uORF2 precludes the ribosome from initiating translation at the *ATF4* coding sequence, resulting in low *ATF4* expression when eIF2 ternary complex levels are high. Under conditions that lower eIF2 ternary complex levels, ribosomes that resume scanning after translating uORF1 take a longer time to acquire a new eIF2 ternary complex. This additional time allows the ribosome to scan past the inhibitory uORF2 start codon and initiate translation at the *ATF4* coding sequence start codon, increasing *ATF4* expression when eIF2 ternary complex concentrations are low. Similarly, as described above, decreased eIF2 ternary complex concentrations increase expression of *CHOP* and *GADD34* by increasing the bypass of the inhibitory uORF (uCC) (Young et al., 2016; Young and Wek, 2016; Young et al., 2015b).

The eIF2 γ -fs mutation impairs eIF2 function, mimicking constitutive eIF2 α phosphorylation and leading to chronic activation of the integrated stress response. The mutation also leads to hyperactivation of the integrated stress response and apoptosis under stress conditions (Young-Baird et al., 2019a), which could contribute to the

symptoms of MEHMO syndrome. *ATF4* induction via eIF2 α phosphorylation has also been linked to regulation of learning and memory in mice , which could explain the ID phenotype seen in MEHMO patients (Young-Baird et al., 2019a). ATF4 is a transcription factor that triggers a program of gene expression to alleviate stress (Walter and Ron, 2011); however, prolonged or pronounced expression of *ATF4* prompts maladaptive responses including induction of apoptosis (Matsumoto et al., 2013). Among the ATF4 transcriptional targets is the pro-apoptotic transcription factor CHOP (Zinszner et al., 1998), whose expression was shown to be elevated in eIF2 γ -fs iPSCs (Young-Baird et al., 2019a).

The S108R mutation exists in a flexible loop of eIF2 γ (Fig 5.1) (Hussain et al., 2014; Llacer et al., 2015). The flexible loop is part of a zinc-binding ribbon within the G domain, but it is unclear whether the human zinc-binding ribbon actually binds zinc. The eIF2 γ proteins from diverse species including the archaeon *Methanococcus jannaschii*, *S. cerevisiae*, and *D. melanogaster* have four cysteine residues within the loop that could be involved in coordinating a zinc ion (Roll-Mecak et al., 2004). Mutation of the single invariant cysteine residue in yeast eIF2 γ results in a severe slow growth phenotype (Erickson et al., 1997; Roll-Mecak et al., 2004); however, it is not known if the mutation affects zinc binding. Of the four cysteine residues conserved among *M. jannaschii* , *S. cerevisiae*, and *D. melanogaster*, mammals have lost three, replacing them with leucine or threonine. Loss of zinc binding during evolution was observed in zinc-binding ribosomal proteins, such as ribosomal protein L33 (Krishna et al., 2003). In the case of these ribosomal proteins, the integrity of the structure is preserved through main chain hydrogen bonds and van der Waals interactions among the hydrophobic amino acids replacing the

zinc-binding cysteines (Krishna et al., 2003; Roll-Mecak et al., 2004). The human eIF2 γ -S108R mutation is in the middle of a region of the zinc ribbon that is not highly conserved outside animals. The loop is absent in archaea and the serine is replaced by an aspartate in *S. cerevisiae* and by an asparagine in *S. pombe*. It is possible that the human eIF2 γ -S108R mutation, which introduces a positively charged amino acid side chain, slightly destabilizes the region and leads to lower levels of eIF2 γ . The fact that mutating the residue to tryptophan or lysine also did not interfere with eIF2 function in yeast indicates that the residue is not extremely sensitive to changes in charge or structure.

Rare human diseases such as MEHMO syndrome affect fewer than 1 person in 100,000, but over 8,000 rare diseases have been identified to date; therefore, although each disease only affects a small number of people, in total, rare diseases affect a significant percentage of the population (Hmeljak and Justice, 2019). Rare diseases can be hard to study in humans, so model organisms are commonly used. For MEHMO syndrome, patient-derived cell lines and budding yeast (*S. cerevisiae*) have been used to understand how *EIF2S3* mutations affect eIF2 activity. Yeast are advantageous for these studies because they are simple to genetically manipulate (Smith and Snyder, 2006; Yang et al., 2017). Although the *GCD11* gene encoding eIF2 γ is essential for yeast viability, we were easily able to introduce mutant forms of eIF2 γ using plasmid shuffling. The majority of the *EIF2S3* mutations examined to date conferred a phenotype in yeast, indicating that the corresponding human mutation is likely pathogenic due to eIF2 γ dysfunction. The limitations of the yeast model system were illustrated by the yeast eIF2 γ -D167 mutations. Whereas the human eIF2 γ -S108R mutation is likely pathogenic in the patient, the corresponding yeast mutation, eIF2 γ -D167R, did not confer any phenotypes. This lack of

phenotype could be because the residues flanking eIF2 γ -D167 are not highly conserved from humans to yeast and, thus, are permissive to substitutions. Alternatively, the putative zinc binding ribbon might have distinct functions in humans and yeast, and the mutation specifically impairs the mammalian function of this element. While we cannot rule out the possibility that the lack of phenotype in yeast is because the mutation is not pathogenic and not the cause of the patients' symptoms, we favor the notion that eIF2 function is slightly compromised in yeast but our assays lack the sensitivity to detect the defects and that modest impairments in eIF2 function will have greater impacts in developing humans than in single-celled yeast.

Although the X-chromosome comprises only about five percent of the human genome, it accounts for about 15 percent of the genes currently associated with intellectual disability (Neri et al., 2018). As of 2017, 141 genes have been linked to XLID and geneticists expect that as many as 80 other X-chromosomally encoded genes could also contribute to XLIDs (Neri et al., 2018). With the introduction of cheap genome sequencing, it is likely that more XLID syndromes will be attributed to mutations in *EIF2S3*. Although, to date, only the eIF2 γ -I465Sfs*4 (eIF2 γ -fs) mutation has been linked to activation of the integrated stress response, it seems likely that other eIF2 γ mutations will also induce the integrated stress response (Skopkova et al., 2017; Young-Baird et al., 2019a). Out of the five eIF2 γ mutations that have been studied in yeast, four were found to derepress expression of *GCN4*, the yeast homolog of *ATF4* (Borck et al., 2012; Gregory et al., 2019; Skopkova et al., 2017; Young-Baird et al., 2019b). Patient derived iPSCs with eIF2 γ -fs were shown to have elevated integrated stress response associated factors including ATF4, CHOP, and GADD34 (Young-Baird et al., 2019a). The small molecule drug ISRIB was

identified as an inhibitor of the integrated stress response (Sidrauski et al., 2013; Sidrauski et al., 2015a) and activator of the eIF2 guanine nucleotide exchange factor eIF2B (Sekine et al., 2015; Sidrauski et al., 2015b; Zyryanova et al., 2018). In light of these properties of ISRIB, Young-Baird et al. (2019a) tested whether ISRIB could suppress the effects of the eIF2 γ -fs mutation in patient-derived iPSCs. Treatment with ISRIB rescued the cell growth defect and restored global translation and blocked the basal depression of ATF4, CHOP, and GADD34 expression in the mutant cells (Young-Baird et al., 2019a). The current model is that treatment with ISRIB enhances eIF2B activity, increasing the fraction of GTP-bound eIF2 complexes, and thereby restoring translational activity in the mutant iPSCs (Young-Baird et al., 2019a). While studies in mouse models of MEHMO syndrome are needed to further assess the efficacy of ISRIB, these studies raise the possibility that the integrated stress response might represent a druggable target for reducing MEHMO symptoms in these patients.

6. References

6.1. Appendix: Glossary of abbreviations

3'-UTR: 3' untranslated region

4EGI-1: eIF4E/eIF4G interaction inhibitor 1

5-FOA: 5-fluoroorotic acid

5'-UTR: 5' untranslated region

A site: aminoacyl site

AEBSF: 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride

Amp: ampicillin

AREs: AU-rich elements

ASL: anticodon stem loop

ATP: adenosine triphosphate

BME: 2-mercaptoethanol

BSA: bovine serum albumin

C-terminal: carboxy terminal

CDS: coding sequence

CHX: cycloheximide

CoA: coenzyme A

Cryo-EM: cryogenic electron microscopy

Ct: threshold cycle

D-iso: D-isoascorbic acid

DFMO: difluoromethylornithine

DHAA: L-Dehydroascorbic acid

DI: deionized

DMEM: Dulbecco's Modified Eagle Medium

DNA: deoxyribonucleic acid

dNTPs: deoxyribonucleotide triphosphates

dsRNA: double stranded RNA

DTT: Dithiothreitol

E site: exit site

E1: ubiquitin activating enzyme

E2: ubiquitin conjugating enzyme

E3: ubiquitin ligase

ECL: enhanced chemiluminescence

EDTA: Ethylenediaminetetraacetic acid

eEF: eukaryotic elongation factor

EF: bacterial elongation factor

EGTA: ethylene glycol-bis(β -aminoethyl ether)-N,N,N',N'-tetraacetic acid

eIF: eukaryotic initiation factor

ER: endoplasmic reticulum

eRF: eukaryotic release factor

EST: expressed sequence tag

FBS: fetal bovine serum

FL: full-length

Gal.: L-Galactono-1,4-lactone

GDP: guanosine-5'-diphosphate

GEF: guanine nucleotide exchange factor

GTP: guanosine -5'-triphosphate

Gul.: L-Gulono-1,4-lactone

HEK293T cells: human embryonic kidney 293 cells with T antigen

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

ISR: integrated stress response

L-AA: L-ascorbic acid

LB: Luria broth

LDS: lithium dodecyl sulfate

MEF: mouse embryonic fibroblast cells

Met-tRNA^{Met}_i: initiator tRNA^{Met}

miRNA: micro RNA

MOPS: 3-(N-morpholino)propanesulfonic acid buffer

mORF: main open reading frame

mRNA: messenger RNA

mRNP: messenger ribonucleoprotein

MW: molecular weight

MYA: million years ago

N-terminal: amino terminal

NC: nascent chain

NCBI: National Center for Biotechnology Information

NEB: New England Biolab

NMD: nonsense-mediated decay

nt: nucleotide

NTPs: ribonucleotide triphosphates

OD: optical density

OF: out-of-frame

ONPG: ortho-Nitrophenyl- β -galactoside

ORF: open reading frame

P Site: peptidyl site

PBS: phosphate-buffered saline

PCR: polymerase chain reaction

PEG: polyethylene glycol

P_i: inorganic phosphate

PIC: translation preinitiation complex

Pol II: RNA polymerase II

poly(A): polyadenylation

Post-TC: post-termination ribosomal complex

PTC: peptidyl transferase center

PTM: post-translational modification

RIPA: radioimmunoprecipitation assay

RNA: ribonucleic acid

RRL: rabbit reticulocyte lysate

rRNA: ribosomal RNA

RT-qPCR: Quantitative reverse transcription polymerase chain reaction

RT: room temperature

SAM: S-adenosyl methionine

SD: standard deviation

SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SDS: sodium dodecyl sulfate

siRNA: small interfering RNA

stem-loop: stem-loop

SOC: super optimal broth

TBS-T: tris-buffered saline and polysorbate 20 (Tween 20)

TC: ternary complex

TE: tris EDTA buffer

tRNA: transfer RNA

TSA: transcriptome shotgun assembly

TSS: transcription start site

U2OS cells: human bone osteosarcoma epithelial cells

uAUG: upstream AUG

uCC: upstream conserved coding region

uORF: upstream open reading frame

UPR: unfolded protein response

XLID: X-linked intellectual disability

YIPP: pyrophosphatase, inorganic (yeast)

YPD: yeast extract peptone dextrose medium

6.2. Bibliography

Abastado, J.P., Miller, P.F., Jackson, B.M., and Hinnebusch, A.G. (1991). Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. *Mol Cell Biol* *11*, 486-496.

Agius, F., Gonzalez-Lamothe, R., Caballero, J.L., Munoz-Blanco, J., Botella, M.A., and Valpuesta, V. (2003). Engineering increased vitamin C levels in plants by overexpression of a D-galacturonic acid reductase. *Nat Biotechnol* *21*, 177-181.

Algire, M.A., Maag, D., and Lorsch, J.R. (2005). Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol Cell* *20*, 251-262.

Alone, P.V., Cao, C., and Dever, T.E. (2008). Translation initiation factor 2gamma mutant alters start codon selection independent of Met-tRNA binding. *Mol Cell Biol* *28*, 6877-6888.

Alves, R., and Savageau, M.A. (2000). Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys J* *79*, 2290-2304.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* *432*, 112-118.

Archer, S.K., Shirokikh, N.E., Beilharz, T.H., and Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* *535*, 570-574.

Arenz, S., Meydan, S., Starosta, A.L., Berninghausen, O., Beckmann, R., Vazquez-Laslop, N., and Wilson, D.N. (2014). Drug sensing by the ribosome induces translational arrest via active site perturbation. *Mol Cell* 56, 446-452.

Babendure, J.R., Babendure, J.L., Ding, J.H., and Tsien, R.Y. (2006). Control of mammalian translation by mRNA structure near caps. *RNA* 12, 851-861.

Baird, T.D., Palam, L.R., Fusakio, M.E., Willy, J.A., Davis, C.M., McClintick, J.N., Anthony, T.G., and Wek, R.C. (2014). Selective mRNA translation during eIF2 phosphorylation induces expression of IBTKalpha. *Mol Biol Cell* 25, 1686-1697.

Baird, T.D., and Wek, R.C. (2012). Eukaryotic initiation factor 2 phosphorylation and translational control in metabolism. *Adv Nutr* 3, 307-321.

Bartoli, C.G., Yu, J., Gomez, F., Fernandez, L., McIntosh, L., and Foyer, C.H. (2006). Interrelationships between light and respiration in the control of ascorbic acid synthesis and accumulation in *Arabidopsis thaliana* leaves. *J Exp Bot* 57, 1621-1631.

Behrmann, E., Loerke, J., Budkevich, T.V., Yamamoto, K., Schmidt, A., Penczek, P.A., Vos, M.R., Burger, J., Mielke, T., Scheerer, P., *et al.* (2015). Structural snapshots of actively translating human ribosomes. *Cell* 161, 845-857.

Benne, R., and Hershey, J.W. (1978). The mechanism of action of protein synthesis initiation factors from rabbit reticulocytes. *The Journal of biological chemistry* 253, 3078-3087.

Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J., and McCarthy, J.E. (2004). Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol Microbiol* *51*, 987-1001.

Bertolotti, A., Zhang, Y., Hendershot, L.M., Harding, H.P., and Ron, D. (2000). Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat Cell Biol* *2*, 326-332.

Bhushan, S., Meyer, H., Starosta, A.L., Becker, T., Mielke, T., Berninghausen, O., Sattler, M., Wilson, D.N., and Beckmann, R. (2010). Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell* *40*, 138-146.

Bieganowski, P., Shilinski, K., Tschlis, P.N., and Brenner, C. (2004). Cdc123 and checkpoint forkhead associated with RING proteins control the cell cycle by controlling eIF2gamma abundance. *The Journal of biological chemistry* *279*, 44656-44666.

Bischoff, L., Berninghausen, O., and Beckmann, R. (2014). Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell Rep* *9*, 469-475.

Boeke, J.D., Trueheart, J., Natsoulis, G., and Fink, G.R. (1987). 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol* *154*, 164-175.

Bohlen, J., Fenzi, K., Kramer, G., Bukau, B., and Teleman, A. (2019). Selective 40S footprinting reveals that scanning ribosomes remain cap-tethered in human cells. *bioRxiv Preprint*.

Borck, G., Shin, B.S., Stiller, B., Mimouni-Bloch, A., Thiele, H., Kim, J.R., Thakur, M., Skinner, C., Aschenbach, L., Smirin-Yosef, P., *et al.* (2012). eIF2gamma mutation that disrupts eIF2 complex integrity links intellectual disability to impaired translation initiation. *Mol Cell* *48*, 641-646.

Broad, R.C., Bonneau, J.P., Beasley, J.T., Roden, S., Philips, J.G., Baumann, U., Hellens, R.P., and Johnson, A.A.T. (2019). Genome-wide identification and characterization of the GDP-L-galactose phosphorylase gene family in bread wheat. *BMC Plant Biol* *19*, 515.

Brush, M.H., Weiser, D.C., and Shenolikar, S. (2003). Growth arrest and DNA damage-inducible protein GADD34 targets protein phosphatase 1 alpha to the endoplasmic reticulum and promotes dephosphorylation of the alpha subunit of eukaryotic translation initiation factor 2. *Mol Cell Biol* *23*, 1292-1303.

Budkevich, T., Giesebrecht, J., Altman, R.B., Munro, J.B., Mielke, T., Nierhaus, K.H., Blanchard, S.C., and Spahn, C.M. (2011). Structure and dynamics of the mammalian ribosomal pretranslocation complex. *Mol Cell* *44*, 214-224.

Bulley, S., Wright, M., Rommens, C., Yan, H., Rassam, M., Lin-Wang, K., Andre, C., Brewster, D., Karunairetnam, S., Allan, A.C., *et al.* (2012). Enhancing ascorbate in fruits and tubers through over-expression of the L-galactose pathway gene GDP-L-galactose phosphorylase. *Plant Biotechnol J* *10*, 390-397.

Bulley, S.M., Rassam, M., Hoser, D., Otto, W., Schunemann, N., Wright, M., MacRae, E., Gleave, A., and Laing, W. (2009). Gene expression studies in kiwifruit and gene over-

expression in Arabidopsis indicates that GDP-L-galactose guanyltransferase is a major control point of vitamin C biosynthesis. *J Exp Bot* *60*, 765-778.

Burns, J.J. (1957). Missing Step in Man, Monkey and Guinea Pig required for the Biosynthesis of L-Ascorbic Acid. *Nature* *180*, 553-553.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* *106*, 7507-7512.

Cao, J., and Geballe, A.P. (1995). Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon. *J Virol* *69*, 1030-1036.

Carmody, S.R., and Wentz, S.R. (2009). mRNA nuclear export at a glance. *J Cell Sci* *122*, 1933-1937.

Cavener, D.R., and Ray, S.C. (1991). Eukaryotic start and stop translation sites. *Nucleic Acids Res* *19*, 3185-3192.

Chang, C.P., Vesole, D.H., Nelson, J., Oldstone, M.B., and Stinski, M.F. (1989). Identification and expression of a human cytomegalovirus early glycoprotein. *J Virol* *63*, 3330-3337.

Chen, J.J., and London, I.M. (1995). Regulation of protein synthesis by heme-regulated eIF-2 alpha kinase. *Trends Biochem Sci* *20*, 105-108.

Chen, J.J., Throop, M.S., Gehrke, L., Kuo, I., Pal, J.K., Brodsky, M., and London, I.M. (1991). Cloning of the cDNA of the heme-regulated eukaryotic initiation factor 2 alpha (eIF-2 alpha) kinase of rabbit reticulocytes: homology to yeast GCN2 protein kinase and human double-stranded-RNA-dependent eIF-2 alpha kinase. *Proc Natl Acad Sci U S A* *88*, 7729-7733.

Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* *127*, 1389-1400.

Chiabudini, M., Tais, A., Zhang, Y., Hayashi, S., Wolfle, T., Fitzke, E., and Rospert, S. (2014). Release factor eRF3 mediates premature translation termination on polylysine-stalled ribosomes in *Saccharomyces cerevisiae*. *Mol Cell Biol* *34*, 4062-4076.

Chikka, M.R., McCabe, D.D., Tyra, H.M., and Rutkowski, D.T. (2013). C/EBP homologous protein (CHOP) contributes to suppression of metabolic genes during endoplasmic reticulum stress in the liver. *The Journal of biological chemistry* *288*, 4405-4415.

Chu, D., Barnes, D.J., and von der Haar, T. (2011). The role of tRNA and ribosome competition in coupling the expression of different mRNAs in *Saccharomyces cerevisiae*. *Nucleic Acids Res* *39*, 6705-6714.

Chung, B.Y., Deery, M.J., Groen, A.J., Howard, J., and Baulcombe, D.C. (2017). Endogenous miRNA in the green alga *Chlamydomonas* regulates gene expression through CDS-targeting. *Nat Plants* *3*, 787-794.

Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C., and Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA* 21, 1731-1745.

Cigan, A.M., Feng, L., and Donahue, T.F. (1988a). tRNAⁱ(met) functions in directing the scanning ribosome to the start site of translation. *Science* 242, 93-97.

Cigan, A.M., Pabich, E.K., and Donahue, T.F. (1988b). Mutational analysis of the HIS4 translational initiator region in *Saccharomyces cerevisiae*. *Mol Cell Biol* 8, 2964-2975.

Clements, J.M., Laz, T.M., and Sherman, F. (1988). Efficiency of translation initiation by non-AUG codons in *Saccharomyces cerevisiae*. *Mol Cell Biol* 8, 4533-4536.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11, 2755-2766.

Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875-886.

Cruz-Rus, E., Amaya, I., Sanchez-Sevilla, J.F., Botella, M.A., and Valpuesta, V. (2011). Regulation of L-ascorbic acid content in strawberry fruits. *J Exp Bot* 62, 4191-4201.

Cruz-Rus, E., Amaya, I., and Valpuesta, V. (2012). The challenge of increasing vitamin C content in plant foods. *Biotechnol J* 7, 1110-1121.

Cruz-Vera, L.R., Rajagopal, S., Squires, C., and Yanofsky, C. (2005). Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. *Mol Cell* 19, 333-343.

Cruz-Vera, L.R., and Yanofsky, C. (2008). Conserved residues Asp16 and Pro24 of TnaC-tRNA^{Pro} participate in tryptophan induction of Tna operon expression. *J Bacteriol* 190, 4791-4797.

Degnin, C.R., Schleiss, M.R., Cao, J., and Geballe, A.P. (1993). Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J Virol* 67, 5514-5521.

del Valle, A.H., Seip, B., Cervera-Marzal, I., Sacheau, G., Seefeldt, A.C., and Innis, C.A. (2019). Ornithine capture by a translating ribosome controls bacterial polyamine synthesis. *bioRxiv*, 604074.

Delbecq, P., Calvo, O., Filipkowski, R.K., Pierard, A., and Messenguy, F. (2000). Functional analysis of the leader peptide of the yeast gene CPA1 and heterologous regulation by other fungal peptides. *Curr Genet* 38, 105-112.

Delbecq, P., Werner, M., Feller, A., Filipkowski, R.K., Messenguy, F., and Pierard, A. (1994). A segment of mRNA encoding the leader peptide of the CPA1 gene confers repression by arginine on a heterologous yeast gene transcript. *Mol Cell Biol* 14, 2378-2390.

Dever, T.E., Dinman, J.D., and Green, R. (2018). Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harb Perspect Biol* 10.

Dever, T.E., Feng, L., Wek, R.C., Cigan, A.M., Donahue, T.F., and Hinnebusch, A.G. (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell* 68, 585-596.

Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol* 4, a013706.

Dever, T.E., Gutierrez, E., and Shin, B.S. (2014). The hypusine-containing translation factor eIF5A. *Crit Rev Biochem Mol Biol* 49, 413-425.

Dever, T.E., Yang, W., Astrom, S., Bystrom, A.S., and Hinnebusch, A.G. (1995). Modulation of tRNA(iMet), eIF-2, and eIF-2B expression shows that GCN4 translation is inversely coupled to the level of eIF-2.GTP.Met-tRNA(iMet) ternary complexes. *Mol Cell Biol* 15, 6351-6363.

Dittmar, K.A., Sorensen, M.A., Elf, J., Ehrenberg, M., and Pan, T. (2005). Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep* 6, 151-157.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* 440, 561-564.

Donahue, T.F., and Cigan, A.M. (1988). Genetic selection for mutations that reduce or abolish ribosomal recognition of the HIS4 translational initiator region. *Mol Cell Biol* 8, 2955-2963.

Dong, J., Lai, R., Nielsen, K., Fekete, C.A., Qiu, H., and Hinnebusch, A.G. (2004). The essential ATP-binding cassette protein RLI1 functions in translation by promoting preinitiation complex assembly. *The Journal of biological chemistry* 279, 42157-42168.

Dorris, D.R., Erickson, F.L., and Hannig, E.M. (1995). Mutations in GCD11, the structural gene for eIF-2 gamma in yeast, alter translational regulation of GCN4 and the selection of the start site for protein synthesis. *EMBO J* 14, 2239-2249.

Dowdle, J., Ishikawa, T., Gatzek, S., Rolinski, S., and Smirnoff, N. (2007). Two genes in *Arabidopsis thaliana* encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. *Plant J* 52, 673-689.

Duan, G., and Walther, D. (2015). The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol* 11, e1004049.

Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature structural & molecular biology* 14, 15-22.

Elf, J., Nilsson, D., Tenson, T., and Ehrenberg, M. (2003). Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* 300, 1718-1722.

Erickson, F.L., Harding, L.D., Dorris, D.R., and Hannig, E.M. (1997). Functional analysis of homologs of translation initiation factor 2gamma in yeast. *Mol Gen Genet* 253, 711-719.

Fang, P., Spevak, C.C., Wu, C., and Sachs, M.S. (2004). A nascent polypeptide domain that can regulate translation elongation. *Proc Natl Acad Sci U S A* 101, 4059-4064.

Fang, P., Wang, Z., and Sachs, M.S. (2000). Evolutionarily conserved features of the arginine attenuator peptide provide the necessary requirements for its function in translational regulation. *The Journal of biological chemistry* 275, 26710-26719.

Flaherty, S.M., Fortes, P., Izaurrealde, E., Mattaj, I.W., and Gilmartin, G.M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc Natl Acad Sci U S A* 94, 11893-11898.

Franks, T.M., and Lykke-Andersen, J. (2008). The control of mRNA decapping and P-body formation. *Mol Cell* 32, 605-615.

Freitag, M., Dighde, N., and Sachs, M.S. (1996). A UV-induced mutation in neurospora that affects translational regulation in response to arginine. *Genetics* 142, 117-127.

Fresco, L.D., and Buratowski, S. (1996). Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *RNA* 2, 584-596.

Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* 295, 2258-2261.

Gaba, A., Jacobson, A., and Sachs, M.S. (2005). Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol Cell* 20, 449-460.

Gao, Y., Badejo, A.A., Shibata, H., Sawa, Y., Maruta, T., Shigeoka, S., Page, M., Smirnoff, N., and Ishikawa, T. (2011). Expression analysis of the VTC2 and VTC5 genes encoding GDP-L-galactose phosphorylase, an enzyme involved in ascorbate biosynthesis, in *Arabidopsis thaliana*. *Biosci Biotechnol Biochem* 75, 1783-1788.

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

Gietz, R.D., and Schiestl, R.H. (2007). Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2, 35-37.

Gillman, E.C., Slusher, L.B., Martin, N.C., and Hopper, A.K. (1991). MOD5 translation initiation sites determine N6-isopentenyladenosine modification of mitochondrial and cytoplasmic tRNA. *Mol Cell Biol* 11, 2382-2390.

Gollnick, P., and Yanofsky, C. (1990). tRNA(Trp) translation of leader peptide codon 12 and other factors that regulate expression of the tryptophanase operon. *J Bacteriol* 172, 3100-3107.

Gong, F., and Yanofsky, C. (2002a). Analysis of tryptophanase operon expression in vitro: accumulation of TnaC-peptidyl-tRNA in a release factor 2-depleted S-30 extract prevents Rho factor action, simulating induction. *The Journal of biological chemistry* 277, 17095-17100.

Gong, F., and Yanofsky, C. (2002b). Instruction of translating ribosome by nascent peptide. *Science* 297, 1864-1867.

Gregory, L.C., Ferreira, C.B., Young-Baird, S.K., Williams, H.J., Harakalova, M., van Haaften, G., Rahman, S.A., Gaston-Massuet, C., Kelberman, D., Gosgene, *et al.* (2019). Impaired EIF2S3 function associated with a novel phenotype of X-linked hypopituitarism with glucose dysregulation. *EBioMedicine* 42, 470-480.

Grentzmann, G., Ingram, J.A., Kelly, P.J., Gesteland, R.F., and Atkins, J.F. (1998). A dual-luciferase reporter system for studying recoding signals. *RNA* 4, 479-486.

Guhaniyogi, J., and Brewer, G. (2001). Regulation of mRNA stability in mammalian cells. *Gene* 265, 11-23.

Gutierrez, E., Shin, B.S., Woolstenhulme, C.J., Kim, J.R., Saini, P., Buskirk, A.R., and Dever, T.E. (2013). eIF5A promotes translation of polyproline motifs. *Mol Cell* 51, 35-45.

Hancock, R.D., Galpin, J.R., and Viola, R. (2000). Biosynthesis of L-ascorbic acid (vitamin C) by *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* 186, 245-250.

Hannig, E.M., Cigan, A.M., Freeman, B.A., and Kinzy, T.G. (1993). GCD11, a negative regulator of GCN4 expression, encodes the gamma subunit of eIF-2 in *Saccharomyces cerevisiae*. *Mol Cell Biol* *13*, 506-520.

Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* *19*, 20-30.

Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., and Ron, D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell* *6*, 1099-1108.

Harding, H.P., Zhang, Y., and Ron, D. (1999). Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* *397*, 271-274.

Harding, H.P., Zhang, Y., Zeng, H., Novoa, I., Lu, P.D., Calton, M., Sadri, N., Yun, C., Popko, B., Paules, R., *et al.* (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell* *11*, 619-633.

Hata, R., and Senoo, H. (1989). L-ascorbic acid 2-phosphate stimulates collagen accumulation, cell proliferation, and formation of a three-dimensional tissuelike substance by skin fibroblasts. *J Cell Physiol* *138*, 8-16.

Hayden, C.A., and Jorgensen, R.A. (2007). 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* *5*, 32.

Herold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U., Carmo-Fonseca, M., Bork, P., and Izaurralde, E. (2000). TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. *Mol Cell Biol* 20, 8996-9008.

Hershey, J.W., Sonenberg, N., and Mathews, M.B. (2012). Principles of translational control: an overview. *Cold Spring Harb Perspect Biol* 4.

Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *The Journal of biological chemistry* 258, 8206-8214.

Hinnebusch, A.G. (1984). Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc Natl Acad Sci U S A* 81, 6442-6446.

Hinnebusch, A.G. (1985). A hierarchy of trans-acting factors modulates translation of an activator of amino acid biosynthetic genes in *Saccharomyces cerevisiae*. *Mol Cell Biol* 5, 2349-2360.

Hinnebusch, A.G. (1997). Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. *The Journal of biological chemistry* 272, 21661-21664.

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* 59, 407-450.

Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol Mol Biol Rev* 75, 434-467, first page of table of contents.

Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. *Annu Rev Biochem* 83, 779-812.

Hinton, T.M., Coldwell, M.J., Carpenter, G.A., Morley, S.J., and Pain, V.M. (2007). Functional analysis of individual binding activities of the scaffold protein eIF4G. *The Journal of biological chemistry* 282, 1695-1708.

Hmeljak, J., and Justice, M.J. (2019). From gene to treatment: supporting rare disease translational research through model systems. *Dis Model Mech* 12.

Hood, H.M., Spevak, C.C., and Sachs, M.S. (2007). Evolutionary changes in the fungal carbamoyl-phosphate synthetase small subunit gene and its associated upstream open reading frame. *Fungal Genet Biol* 44, 93-104.

Horikami, S.M., De Ferra, F., and Moyer, S.A. (1984). Characterization of the infections of permissive and nonpermissive cells by host range mutants of vesicular stomatitis virus defective in RNA methylation. *Virology* 138, 1-15.

Horinouchi, S., and Weisblum, B. (1980). Posttranscriptional modification of mRNA conformation: mechanism that regulates erythromycin-induced resistance. *Proc Natl Acad Sci U S A* 77, 7079-7083.

Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. *Cell* *136*, 763-776.

Huang, H.K., Yoon, H., Hannig, E.M., and Donahue, T.F. (1997). GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev* *11*, 2396-2413.

Hussain, T., Llacer, J.L., Fernandez, I.S., Munoz, A., Martin-Marcos, P., Savva, C.G., Lorsch, J.R., Hinnebusch, A.G., and Ramakrishnan, V. (2014). Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* *159*, 597-607.

Ibba, M., and Soll, D. (2004). Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev* *18*, 731-738.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* *324*, 218-223.

Inoue, K., Ohno, M., Sakamoto, H., and Shimura, Y. (1989). Effect of the cap structure on pre-mRNA splicing in *Xenopus* oocyte nuclei. *Genes Dev* *3*, 1472-1479.

Ishii, E., Chiba, S., Hashimoto, N., Kojima, S., Homma, M., Ito, K., Akiyama, Y., and Mori, H. (2015). Nascent chain-monitored remodeling of the Sec machinery for salinity adaptation of marine bacteria. *Proc Natl Acad Sci U S A* *112*, E5513-5522.

Ito, K., and Chiba, S. (2013). Arrest peptides: cis-acting modulators of translation. *Annu Rev Biochem* *82*, 171-202.

Ivanov, I.P., Atkins, J.F., and Michael, A.J. (2010a). A profusion of upstream open reading frame mechanisms in polyamine-responsive translational regulation. *Nucleic Acids Res* *38*, 353-359.

Ivanov, I.P., Loughran, G., and Atkins, J.F. (2008). uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs. *Proc Natl Acad Sci U S A* *105*, 10079-10084.

Ivanov, I.P., Loughran, G., Sachs, M.S., and Atkins, J.F. (2010b). Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proc Natl Acad Sci U S A* *107*, 18056-18060.

Ivanov, I.P., Shin, B.S., Loughran, G., Tzani, I., Young-Baird, S.K., Cao, C., Atkins, J.F., and Dever, T.E. (2018). Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. *Mol Cell* *70*, 254-264 e256.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* *11*, 113-127.

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

Janzen, D.M., Frolova, L., and Geballe, A.P. (2002). Inhibition of translation termination mediated by an interaction of eukaryotic release factor 1 with a nascent peptidyl-tRNA. *Mol Cell Biol* 22, 8562-8570.

Jiang, H.Y., Wek, S.A., McGrath, B.C., Lu, D., Hai, T., Harding, H.P., Wang, X., Ron, D., Cavener, D.R., and Wek, R.C. (2004). Activating transcription factor 3 is integral to the eukaryotic initiation factor 2 kinase stress response. *Mol Cell Biol* 24, 1365-1377.

Jiao, X., Chang, J.H., Kilic, T., Tong, L., and Kiledjian, M. (2013). A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Mol Cell* 50, 104-115.

Johansson, M., leong, K.W., Trobro, S., Strazewski, P., Aqvist, J., Pavlov, M.Y., and Ehrenberg, M. (2011). pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. *Proc Natl Acad Sci U S A* 108, 79-84.

Joshi, C.P., Zhou, H., Huang, X., and Chiang, V.L. (1997). Context sequences of translation initiation codon in plants. *Plant Mol Biol* 35, 993-1001.

Juntawong, P., Girke, T., Bazin, J., and Bailey-Serres, J. (2014). Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc Natl Acad Sci U S A* 111, E203-212.

Kang, S.G., Price, J., Lin, P.C., Hong, J.C., and Jang, J.C. (2010). The arabidopsis bZIP1 transcription factor is involved in sugar signaling, protein networking, and DNA binding. *Mol Plant* 3, 361-373.

Karim, M.M., Svitkin, Y.V., Kahvejian, A., De Crescenzo, G., Costa-Mattioli, M., and Sonenberg, N. (2006). A mechanism of translational repression by competition of Paip2 with eIF4G for poly(A) binding protein (PABP) binding. *Proc Natl Acad Sci U S A* 103, 9494-9499.

Kearse, M.G., Goldman, D.H., Choi, J., Nwaezeapu, C., Liang, D., Green, K.M., Goldstrohm, A.C., Todd, P.K., Green, R., and Wilusz, J.E. (2019). Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev* 33, 871-885.

Kearse, M.G., and Wilusz, J.E. (2017). Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev* 31, 1717-1731.

Kemper, W.M., Berry, K.W., and Merrick, W.C. (1976). Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2Balpha and M2Bbeta. *The Journal of biological chemistry* 251, 5551-5557.

King, C.G., and Waugh, W.A. (1932). The Chemical Nature of Vitamin C. *Science* 75, 357-358.

Kiser, K.B., and Schmidt, M.G. (1999). Regulation of the Escherichia coli secA gene is mediated by two distinct RNA structural conformations. *Curr Microbiol* 38, 113-121.

Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* *8*, 761-773.

Kolitz, S.E., Takacs, J.E., and Lorsch, J.R. (2009). Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA* *15*, 138-152.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* *184*, 868-871.

Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., Trakhanov, S., Scott, W.G., and Noller, H.F. (2008). Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A* *105*, 19684-19689.

Koutmou, K.S., Radhakrishnan, A., and Green, R. (2015). Synthesis at the Speed of Codons. *Trends Biochem Sci* *40*, 717-718.

Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* *15*, 1109-1123.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res* *12*, 857-872.

Kozak, M. (1986a). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci U S A* *83*, 2850-2854.

Kozak, M. (1986b). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283-292.

Kozak, M. (1987). Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol Cell Biol* 7, 3438-3445.

Kozak, M. (1989). Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* 9, 5073-5080.

Kozak, M. (1991a). A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr* 1, 111-115.

Kozak, M. (1991b). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *The Journal of biological chemistry* 266, 19867-19870.

Kozak, M. (1992). Regulation of translation in eukaryotic systems. *Annu Rev Cell Biol* 8, 197-225.

Kozak, M. (2001). Constraints on reinitiation of translation in mammals. *Nucleic Acids Res* 29, 5226-5232.

Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003). Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* 31, 532-550.

Laing, W.A., Martinez-Sanchez, M., Wright, M.A., Bulley, S.M., Brewster, D., Dare, A.P., Rassam, M., Wang, D., Storey, R., Macknight, R.C., *et al.* (2015). An upstream open

reading frame is essential for feedback regulation of ascorbate biosynthesis in Arabidopsis. *Plant Cell* 27, 772-786.

Lanker, S., Bushman, J.L., Hinnebusch, A.G., Trachsel, H., and Mueller, P.P. (1992). Autoregulation of the yeast lysyl-tRNA synthetase gene GCD5/KRS1 by translational and transcriptional control mechanisms. *Cell* 70, 647-657.

Laurberg, M., Asahara, H., Korostelev, A., Zhu, J., Trakhanov, S., and Noller, H.F. (2008). Structural basis for translation termination on the 70S ribosome. *Nature* 454, 852-857.

Law, G.L., Raney, A., Heusner, C., and Morris, D.R. (2001). Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. *The Journal of biological chemistry* 276, 38036-38043.

Lawless, C., Pearson, R.D., Selley, J.N., Smirnova, J.B., Grant, C.M., Ashe, M.P., Pavitt, G.D., and Hubbard, S.J. (2009). Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. *BMC Genomics* 10, 7.

Le Quesne, J.P., Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Dysregulation of protein synthesis and disease. *J Pathol* 220, 140-151.

Lebreton, A., Tomecki, R., Dziembowski, A., and Seraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* 456, 993-996.

Lee, Y.Y., Cevallos, R.C., and Jan, E. (2009). An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2alpha phosphorylation. *The Journal of biological chemistry* 284, 6661-6673.

Levin, D.H., Petryshyn, R., and London, I.M. (1980). Characterization of double-stranded-RNA-activated kinase that phosphorylates alpha subunit of eukaryotic initiation factor 2 (eIF-2 alpha) in reticulocyte lysates. *Proc Natl Acad Sci U S A* 77, 832-836.

Li, W., Ward, F.R., McClure, K.F., Chang, S.T., Montabana, E., Liras, S., Dullea, R.G., and Cate, J.H.D. (2019). Structural basis for selective stalling of human ribosome nascent chain complexes by a drug-like molecule. *Nature structural & molecular biology* 26, 501-509.

Li, X., and Coffino, P. (1992). Regulated degradation of ornithine decarboxylase requires interaction with the polyamine-inducible protein antizyme. *Mol Cell Biol* 12, 3556-3562.

Liaud, N., Horlbeck, M.A., Gilbert, L.A., Gjoni, K., Weissman, J.S., and Cate, J.H.D. (2019). Cellular response to small molecules that selectively stall protein synthesis by the ribosome. *PLoS Genet* 15, e1008057.

Lind, J. (1953). Lind's treatise on scurvy. In *The diagnostics, or signs*, G. Stewart CP, eds., ed. (Edinburgh: Edinburgh University Press), pp. 113–132.

Linster, C.L., and Van Schaftingen, E. (2007). Vitamin C. Biosynthesis, recycling and degradation in mammals. *FEBS J* 274, 1-22.

Lintner, N.G., McClure, K.F., Petersen, D., Londregan, A.T., Piotrowski, D.W., Wei, L., Xiao, J., Bolt, M., Loria, P.M., Maguire, B., *et al.* (2017). Selective stalling of human translation through small-molecule engagement of the ribosome nascent chain. *PLoS Biol* *15*, e2001882.

Liu, M.J., Wu, S.H., Wu, J.F., Lin, W.D., Wu, Y.C., Tsai, T.Y., Tsai, H.L., and Wu, S.H. (2013). Translational landscape of photomorphogenic Arabidopsis. *Plant Cell* *25*, 3699-3710.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* *127*, 1223-1237.

Llacer, J.L., Hussain, T., Marler, L., Aitken, C.E., Thakur, A., Lorsch, J.R., Hinnebusch, A.G., and Ramakrishnan, V. (2015). Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol Cell* *59*, 399-412.

Lopatin, A.N., Makhina, E.N., and Nichols, C.G. (1995). The mechanism of inward rectification of potassium channels: "long-pore plugging" by cytoplasmic polyamines. *J Gen Physiol* *106*, 923-955.

Lorence, A., Chevone, B.I., Mendes, P., and Nessler, C.L. (2004). myo-inositol oxygenase offers a possible entry point into plant ascorbate biosynthesis. *Plant Physiol* *134*, 1200-1205.

Lu, J., and Deutsch, C. (2008). Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol* *384*, 73-86.

Lu, P.D., Harding, H.P., and Ron, D. (2004). Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *J Cell Biol* 167, 27-33.

Luo, Z., Freitag, M., and Sachs, M.S. (1995). Translational regulation in response to changes in amino acid availability in *Neurospora crassa*. *Mol Cell Biol* 15, 5235-5245.

Luukkonen, B.G., Tan, W., and Schwartz, S. (1995). Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J Virol* 69, 4086-4094.

Machida, K., Shigeta, T., Yamamoto, Y., Ito, T., Svitkin, Y., Sonenberg, N., and Imataka, H. (2018). Dynamic interaction of poly(A)-binding protein with the ribosome. *Sci Rep* 8, 17435.

Malkin, L.I., and Rich, A. (1967). Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J Mol Biol* 26, 329-346.

Malzer, E., Szajewska-Skuta, M., Dalton, L.E., Thomas, S.E., Hu, N., Skaer, H., Lomas, D.A., Crowther, D.C., and Marciniak, S.J. (2013). Coordinate regulation of eIF2alpha phosphorylation by PPP1R15 and GCN2 is required during *Drosophila* development. *J Cell Sci* 126, 1406-1415.

Manjunath, H., Zhang, H., Rehfeld, F., Han, J., Chang, T.C., and Mendell, J.T. (2019). Suppression of Ribosomal Pausing by eIF5A Is Necessary to Maintain the Fidelity of Start Codon Selection. *Cell Rep* 29, 3134-3146 e3136.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* 19, 1512-1517.

Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J.F., Gesteland, R.F., and Hayashi, S. (1995). Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* 80, 51-60.

Matsufuji, S., Miyazaki, Y., Kanamoto, R., Kameji, T., Murakami, Y., Baby, T.G., Fujita, K., Ohno, T., and Hayashi, S. (1990). Analyses of ornithine decarboxylase antizyme mRNA with a cDNA cloned from rat liver. *J Biochem* 108, 365-371.

Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *The Journal of biological chemistry* 255, 11992-11996.

Matsumoto, H., Miyazaki, S., Matsuyama, S., Takeda, M., Kawano, M., Nakagawa, H., Nishimura, K., and Matsuo, S. (2013). Selection of autophagy or apoptosis in cells exposed to ER-stress depends on ATF4 expression pattern with or without CHOP expression. *Biol Open* 2, 1084-1090.

McEwen, E., Kedersha, N., Song, B., Scheuner, D., Gilks, N., Han, A., Chen, J.J., Anderson, P., and Kaufman, R.J. (2005). Heme-regulated inhibitor kinase-mediated phosphorylation of eukaryotic translation initiation factor 2 inhibits translation, induces stress granule formation, and mediates survival upon arsenite exposure. *The Journal of biological chemistry* 280, 16925-16933.

McNicholas, P., Salavati, R., and Oliver, D. (1997). Dual regulation of *Escherichia coli* *secA* translation by distinct upstream elements. *J Mol Biol* 265, 128-141.

Meurs, E., Chong, K., Galabru, J., Thomas, N.S., Kerr, I.M., Williams, B.R., and Hovanessian, A.G. (1990). Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell* 62, 379-390.

Minois, N., Carmona-Gutierrez, D., and Madeo, F. (2011). Polyamines in aging and disease. *Aging (Albany NY)* 3, 716-732.

Mittal, N., Guimaraes, J.C., Gross, T., Schmidt, A., Vina-Vilaseca, A., Nedialkova, D.D., Aeschmann, F., Leidel, S.A., Spang, A., and Zavolan, M. (2017). The Gcn4 transcription factor reduces protein synthesis capacity and extends yeast lifespan. *Nat Commun* 8, 457.

Moazed, D., and Noller, H.F. (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature* 342, 142-148.

Moerke, N.J., Aktas, H., Chen, H., Cantel, S., Reibarkh, M.Y., Fahmy, A., Gross, J.D., Degtarev, A., Yuan, J., Chorev, M., *et al.* (2007). Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. *Cell* 128, 257-267.

Moortgat, S., Desir, J., Benoit, V., Boulanger, S., Pendeville, H., Nassogne, M.C., Lederer, D., and Maystadt, I. (2016). Two novel EIF2S3 mutations associated with syndromic intellectual disability with severe microcephaly, growth retardation, and epilepsy. *Am J Med Genet A* 170, 2927-2933.

Mueller, P.P., and Hinnebusch, A.G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. *Cell* 45, 201-207.

Muller-Moule, P. (2008). An expression analysis of the ascorbate biosynthesis enzyme VTC2. *Plant Mol Biol* 68, 31-41.

Murakami, Y., Ichiba, T., Matsufuji, S., and Hayashi, S. (1996). Cloning of antizyme inhibitor, a highly homologous protein to ornithine decarboxylase. *The Journal of biological chemistry* 271, 3340-3342.

Murakami, Y., Matsufuji, S., Kameji, T., Hayashi, S., Igarashi, K., Tamura, T., Tanaka, K., and Ichihara, A. (1992). Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. *Nature* 360, 597-599.

Naidu, K.A. (2003). Vitamin C in human health and disease is still a mystery? An overview. *Nutr J* 2, 7.

Najarian, D., Dihanich, M.E., Martin, N.C., and Hopper, A.K. (1987). DNA sequence and transcript mapping of MOD5: features of the 5' region which suggest two translational starts. *Mol Cell Biol* 7, 185-191.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., and Miura, K. (2008). Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Res* 36, 861-871.

Neri, G., Schwartz, C.E., Lubs, H.A., and Stevenson, R.E. (2018). X-linked intellectual disability update 2017. *Am J Med Genet A* 176, 1375-1388.

Niederberger, P., Miozzari, G., and Hutter, R. (1981). Biological role of the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1, 584-593.

Nika, J., Rippel, S., and Hannig, E.M. (2001). Biochemical analysis of the eIF2beta gamma complex reveals a structural function for eIF2alpha in catalyzed nucleotide exchange. *The Journal of biological chemistry* 276, 1051-1056.

Nishikimi, M., and Yagi, K. (1991). Molecular basis for the deficiency in humans of gulonolactone oxidase, a key enzyme for ascorbic acid biosynthesis. *Am J Clin Nutr* 54, 1203S-1208S.

Nojima, T., Hirose, T., Kimura, H., and Hagiwara, M. (2007). The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *The Journal of biological chemistry* 282, 15645-15651.

Nurenberg, E., and Tampe, R. (2013). Tying up loose ends: ribosome recycling in eukaryotes and archaea. *Trends Biochem Sci* 38, 64-74.

Onouchi, H., Haraguchi, Y., Nakamoto, M., Kawasaki, D., Nagami-Yamashita, Y., Murota, K., Kezuka-Hosomi, A., Chiba, Y., and Naito, S. (2008). Nascent peptide-mediated translation elongation arrest of *Arabidopsis thaliana* CGS1 mRNA occurs autonomously. *Plant Cell Physiol* 49, 549-556.

Padayatty, S.J., and Levine, M. (2016). Vitamin C: the known and the unknown and Goldilocks. *Oral Dis* 22, 463-493.

Pain, V.M. (1986). Initiation of protein synthesis in mammalian cells. *Biochem J* 235, 625-637.

Palam, L.R., Baird, T.D., and Wek, R.C. (2011). Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. *The Journal of biological chemistry* 286, 10939-10949.

Park, M.H., Cooper, H.L., and Folk, J.E. (1981). Identification of hypusine, an unusual amino acid, in a protein from human lymphocytes and of spermidine as its biosynthetic precursor. *Proc Natl Acad Sci U S A* 78, 2869-2873.

Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R., and Ramakrishnan, V. (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* 26, 41-50.

Pavlov, M.Y., Watts, R.E., Tan, Z., Cornish, V.W., Ehrenberg, M., and Forster, A.C. (2009). Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc Natl Acad Sci U S A* 106, 50-54.

Peabody, D.S. (1987). Translation initiation at an ACG triplet in mammalian cells. *The Journal of biological chemistry* 262, 11847-11851.

Peabody, D.S. (1989). Translation initiation at non-AUG triplets in mammalian cells. *The Journal of biological chemistry* 264, 5031-5035.

Pegg, A.E. (2009). Mammalian polyamine metabolism and function. *IUBMB Life* 61, 880-894.

Pelechano, V., and Alepuz, P. (2017). eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res* 45, 7326-7338.

Pelletier, J., and Sonenberg, N. (1985). Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* 40, 515-526.

Perzmaier, A.F., Richter, F., and Seufert, W. (2013). Translation initiation requires cell division cycle 123 (Cdc123) to facilitate biogenesis of the eukaryotic initiation factor 2 (eIF2). *The Journal of biological chemistry* 288, 21537-21546.

Pesole, G., Gissi, C., Grillo, G., Licciulli, F., Liuni, S., and Saccone, C. (2000). Analysis of oligonucleotide AUG start codon context in eukaryotic mRNAs. *Gene* 261, 85-91.

Pestova, T.V., and Kolupaeva, V.G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev* 16, 2906-2922.

Petersen, D.N., Hawkins, J., Ruangsiriluk, W., Stevens, K.A., Maguire, B.A., O'Connell, T.N., Rocke, B.N., Boehm, M., Ruggeri, R.B., Rolph, T., *et al.* (2016). A Small-Molecule Anti-secretagogue of PCSK9 Targets the 80S Ribosome to Inhibit PCSK9 Protein Translation. *Cell Chem Biol* 23, 1362-1371.

Pierard, A., and Schroter, B. (1978). Structure-function relationships in the arginine pathway carbamoylphosphate synthase of *Saccharomyces cerevisiae*. *J Bacteriol* 134, 167-176.

Pisarev, A.V., Hellen, C.U., and Pestova, T.V. (2007). Recycling of eukaryotic posttermination ribosomal complexes. *Cell* 131, 286-299.

Pisarev, A.V., Skabkin, M.A., Pisareva, V.P., Skabkina, O.V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U., and Pestova, T.V. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol Cell* 37, 196-210.

Porrúa, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. *Trends Genet* 32, 508-522.

Poyry, T.A., Kaminski, A., and Jackson, R.J. (2004). What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev* 18, 62-75.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. *Cell* 108, 501-512.

Purvis, I.J., Bettany, A.J., Santiago, T.C., Coggins, J.R., Duncan, K., Eason, R., and Brown, A.J. (1987). The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J Mol Biol* 193, 413-417.

Radhakrishnan, A., Chen, Y.H., Martin, S., Alhusaini, N., Green, R., and Collier, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell* 167, 122-132 e129.

Radhakrishnan, A., and Green, R. (2016). Connections Underlying Translation and mRNA Stability. *J Mol Biol* 428, 3558-3564.

Raghavan, A., Ogilvie, R.L., Reilly, C., Abelson, M.L., Raghavan, S., Vasdewani, J., Krathwohl, M., and Bohjanen, P.R. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Res* 30, 5529-5538.

Rahmani, F., Hummel, M., Schuurmans, J., Wiese-Klinkenberg, A., Smeekens, S., and Hanson, J. (2009). Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol* 150, 1356-1367.

Ramanathan, A., Robb, G.B., and Chan, S.H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Res* 44, 7511-7526.

Rasanu, N.M., V.; Matei, Nicoleta; Soceanu, Alina. (2005). Determination of Vitamin C in different stages of fruits growing. *Analele Universităţii din Bucureşti I-II*, 167-172.

Ravid, T., and Hochstrasser, M. (2008). Diversity of degradation signals in the ubiquitin-proteasome system. *Nat Rev Mol Cell Biol* 9, 679-690.

Resch, A.M., Ogurtsov, A.Y., Rogozin, I.B., Shabalina, S.A., and Koonin, E.V. (2009). Evolution of alternative and constitutive regions of mammalian 5'UTRs. *BMC Genomics* 10, 162.

Richter, J.D., and Lasko, P. (2011). Translational control in oocyte development. *Cold Spring Harb Perspect Biol* 3, a002758.

Rodnina, M.V. (2013). The ribosome as a versatile catalyst: reactions at the peptidyl transferase center. *Curr Opin Struct Biol* 23, 595-602.

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 21, 327-335.

Roll-Mecak, A., Alone, P., Cao, C., Dever, T.E., and Burley, S.K. (2004). X-ray structure of translation initiation factor eIF2gamma: implications for tRNA and eIF2alpha binding. *The Journal of biological chemistry* 279, 10634-10642.

Roth, A.C. (2012). Decoding properties of tRNA leave a detectable signal in codon usage bias. *Bioinformatics* 28, i340-i348.

Rowlands, A.G., Panniers, R., and Henshaw, E.C. (1988). The catalytic mechanism of guanine nucleotide exchange factor action and competitive inhibition by

phosphorylated eukaryotic initiation factor 2. *The Journal of biological chemistry* 263, 5526-5533.

Ruan, H., Shantz, L.M., Pegg, A.E., and Morris, D.R. (1996). The upstream open reading frame of the mRNA encoding S-adenosylmethionine decarboxylase is a polyamine-responsive translational control element. *The Journal of biological chemistry* 271, 29576-29582.

Sabatini, D.D., and Blobel, G. (1970). Controlled proteolysis of nascent polypeptides in rat liver cell fractions. II. Location of the polypeptides in rough microsomes. *J Cell Biol* 45, 146-157.

Sabi, R., and Tuller, T. (2014). Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res* 21, 511-526.

Safaei, N., Kozlov, G., Noronha, A.M., Xie, J., Wilds, C.J., and Gehring, K. (2012). Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol Cell* 48, 375-386.

Saikia, M., Wang, X., Mao, Y., Wan, J., Pan, T., and Qian, S.B. (2016). Codon optimality controls differential mRNA translation during amino acid starvation. *RNA* 22, 1719-1727.

Saini, A.K., Nanda, J.S., Lorsch, J.R., and Hinnebusch, A.G. (2010). Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA(i)(Met) binding to the ribosome. *Genes Dev* 24, 97-110.

Saini, P., Eyler, D.E., Green, R., and Dever, T.E. (2009). Hypusine-containing protein eIF5A promotes translation elongation. *Nature* *459*, 118-121.

Salzberg, S.L. (2018). Open questions: How many genes do we have? *BMC Biol* *16*, 94.

Sarker, S., and Oliver, D. (2002). Critical regions of secM that control its translation and secretion and promote secretion-specific secA regulation. *J Bacteriol* *184*, 2360-2369.

Schleiss, M.R., Degnin, C.R., and Geballe, A.P. (1991). Translational control of human cytomegalovirus gp48 expression. *J Virol* *65*, 6782-6789.

Schuller, A.P., and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. *Nat Rev Mol Cell Biol* *19*, 526-541.

Schuller, A.P., Wu, C.C., Dever, T.E., Buskirk, A.R., and Green, R. (2017). eIF5A Functions Globally in Translation Elongation and Termination. *Mol Cell* *66*, 194-205 e195.

Schwartz, S., Felber, B.K., Benko, D.M., Fenyo, E.M., and Pavlakis, G.N. (1990). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* *64*, 2519-2529.

Schwartz, S., Felber, B.K., and Pavlakis, G.N. (1992). Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. *Mol Cell Biol* *12*, 207-219.

Sekine, Y., Zyryanova, A., Crespillo-Casado, A., Fischer, P.M., Harding, H.P., and Ron, D. (2015). Stress responses. Mutations in a translation initiation factor identify the target of a memory-enhancing compound. *Science* 348, 1027-1030.

Sekiyama, N., Arthanari, H., Papadopoulos, E., Rodriguez-Mias, R.A., Wagner, G., and Leger-Abraham, M. (2015). Molecular mechanism of the dual activity of 4EGI-1: Dissociating eIF4G from eIF4E but stabilizing the binding of unphosphorylated 4E-BP1. *Proc Natl Acad Sci U S A* 112, E4036-4045.

Shin, B.S., Kim, J.R., Walker, S.E., Dong, J., Lorsch, J.R., and Dever, T.E. (2011). Initiation factor eIF2gamma promotes eIF2-GTP-Met-tRNAⁱ(Met) ternary complex binding to the 40S ribosome. *Nature structural & molecular biology* 18, 1227-1234.

Shoemaker, C.J., and Green, R. (2011). Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc Natl Acad Sci U S A* 108, E1392-1398.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. *Nature structural & molecular biology* 19, 594-601.

Sidrauski, C., Acosta-Alvear, D., Khoutorsky, A., Vedantham, P., Hearn, B.R., Li, H., Gamache, K., Gallagher, C.M., Ang, K.K., Wilson, C., *et al.* (2013). Pharmacological brake-release of mRNA translation enhances cognitive memory. *Elife* 2, e00498.

Sidrauski, C., McGeachy, A.M., Ingolia, N.T., and Walter, P. (2015a). The small molecule ISRIB reverses the effects of eIF2 α phosphorylation on translation and stress granule assembly. *Elife* 4.

Sidrauski, C., Tsai, J.C., Kampmann, M., Hearn, B.R., Vedantham, P., Jaishankar, P., Sokabe, M., Mendez, A.S., Newton, B.W., Tang, E.L., *et al.* (2015b). Pharmacological dimerization and activation of the exchange factor eIF2B antagonizes the integrated stress response. *Elife* 4, e07314.

Simonovic, M., and Steitz, T.A. (2009). A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. *Biochim Biophys Acta* 1789, 612-623.

Skabkin, M.A., Skabkina, O.V., Dhote, V., Komar, A.A., Hellen, C.U., and Pestova, T.V. (2010). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev* 24, 1787-1801.

Skopkova, M., Hennig, F., Shin, B.-S., Turner, C.E., Stanikova, D., Brennerova, K., Stanik, J., Fischer, U., Henden, L., Müller, U., *et al.* (2017). EIF2S3 Mutations Associated with Severe X-Linked Intellectual Disability Syndrome MEHMO. *38*, 409-425.

Slusher, L.B., Gillman, E.C., Martin, N.C., and Hopper, A.K. (1991). mRNA leader length and initiation codon context determine alternative AUG selection for the yeast gene MOD5. *Proc Natl Acad Sci U S A* 88, 9789-9793.

Smith, M.G., and Snyder, M. (2006). Yeast as a model for human disease. *Curr Protoc Hum Genet Chapter 15*, Unit 15 16.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731-745.

Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A., and Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1-- mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* 100, 311-321.

Sorensen, M.A., and Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J Mol Biol* 222, 265-280.

Sothiselvam, S., Liu, B., Han, W., Ramu, H., Klepacki, D., Atkinson, G.C., Brauer, A., Remm, M., Tenson, T., Schulten, K., *et al.* (2014). Macrolide antibiotics allosterically predispose the ribosome for translation arrest. *Proc Natl Acad Sci U S A* 111, 9804-9809.

Spevak, C.C., Ivanov, I.P., and Sachs, M.S. (2010). Sequence requirements for ribosome stalling by the arginine attenuator peptide. *The Journal of biological chemistry* 285, 40933-40942.

Steinmuller, R., Steinberger, D., and Muller, U. (1998). MEHMO (mental retardation, epileptic seizures, hypogonadism and -genitalism, microcephaly, obesity), a novel syndrome: assignment of disease locus to xp21.1-p22.13. *Eur J Hum Genet* 6, 201-206.

Stewart, M. (2019). Polyadenylation and nuclear export of mRNAs. *The Journal of biological chemistry* 294, 2977-2987.

Stoecklin, G., Mayo, T., and Anderson, P. (2006). ARE-mRNA degradation requires the 5'-3' decay pathway. *EMBO Rep* 7, 72-77.

Stoecklin, G., and Muhlemann, O. (2013). RNA decay mechanisms: specificity through diversity. *Biochim Biophys Acta* 1829, 487-490.

Stolboushkina, E., Nikonov, S., Nikulin, A., Blasi, U., Manstein, D.J., Fedorov, R., Garber, M., and Nikonov, O. (2008). Crystal structure of the intact archaeal translation initiation factor 2 demonstrates very high conformational flexibility in the alpha- and beta-subunits. *J Mol Biol* 382, 680-691.

Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* 9, 644-647.

Svirbely, J.L., and Szent-Gyorgyi, A. (1932). The chemical nature of vitamin C. *Biochem J* 26, 865-870.

Svirbely, J.L., and Szent-Gyorgyi, A. (1933). The chemical nature of vitamin C. *Biochem J* 27, 279-285.

Takamizawa, S., Maehata, Y., Imai, K., Senoo, H., Sato, S., and Hata, R. (2004). Effects of ascorbic acid and ascorbic acid 2-phosphate, a long-acting vitamin C derivative, on the proliferation and differentiation of human osteoblast-like cells. *Cell Biol Int* 28, 255-265.

- Taniuchi, S., Miyake, M., Tsugawa, K., Oyadomari, M., and Oyadomari, S. (2016). Integrated stress response of vertebrates is regulated by four eIF2alpha kinases. *Sci Rep* 6, 32886.
- Thanaraj, T.A., and Argos, P. (1996). Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci* 5, 1973-1983.
- Thore, S., Leibundgut, M., and Ban, N. (2006). Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* 312, 1208-1211.
- Ulvskov, P., Paiva, D.S., Domozych, D., and Harholt, J. (2013). Classification, naming and evolutionary history of glycosyltransferases from sequenced green and red algal genomes. *PLoS One* 8, e76511.
- Unbehaun, A., Borukhov, S.I., Hellen, C.U., and Pestova, T.V. (2004). Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev* 18, 3078-3093.
- van den Heuvel, J.J., Bergkamp, R.J., Planta, R.J., and Raue, H.A. (1989). Effect of deletions in the 5'-noncoding region on the translational efficiency of phosphoglycerate kinase mRNA in yeast. *Gene* 79, 83-95.
- van der Horst, S., Filipovska, T., Hanson, J., and Smeekens, S. (2020). Metabolite Control of Translation by Conserved Peptide uORFs: The Ribosome as a Metabolite Multisensor. *Plant Physiol* 182, 110-122.

Varshavsky, A. (2005). Regulated protein degradation. *Trends Biochem Sci* 30, 283-286.

Vassilenko, K.S., Alekhina, O.M., Dmitriev, S.E., Shatsky, I.N., and Spirin, A.S. (2011).

Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. *Nucleic Acids Res* 39, 5555-5567.

Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates

ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A* 101, 11269-11274.

Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. *Crit Rev*

Biochem Mol Biol 44, 117-141.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: deadenylation by the Ccr4-not

and Pan2-Pan3 complexes. *Biochim Biophys Acta* 1829, 561-570.

Walsh, D., Mathews, M.B., and Mohr, I. (2013). Tinkering with translation: protein

synthesis in virus-infected cells. *Cold Spring Harb Perspect Biol* 5, a012351.

Walter, P., and Ron, D. (2011). The unfolded protein response: from stress pathway to

homeostatic regulation. *Science* 334, 1081-1086.

Wang, Y., and Casero, R.A., Jr. (2006). Mammalian polyamine catabolism: a therapeutic

target, a pathological problem, or both? *J Biochem* 139, 17-25.

Wang, Z., Fang, P., and Sachs, M.S. (1998). The evolutionarily conserved eukaryotic

arginine attenuator peptide regulates the movement of ribosomes that have translated

it. *Mol Cell Biol* 18, 7528-7536.

Wang, Z., Gaba, A., and Sachs, M.S. (1999). A highly conserved mechanism of regulated ribosome stalling mediated by fungal arginine attenuator peptides that appears independent of the charging status of arginyl-tRNAs. *The Journal of biological chemistry* 274, 37565-37574.

Wang, Z., and Sachs, M.S. (1997). Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*. *Mol Cell Biol* 17, 4904-4913.

Wei, J., Wu, C., and Sachs, M.S. (2012). The arginine attenuator peptide interferes with the ribosome peptidyl transferase center. *Mol Cell Biol* 32, 2396-2406.

Wei, J., Zhang, Y., Ivanov, I.P., and Sachs, M.S. (2013). The stringency of start codon selection in the filamentous fungus *Neurospora crassa*. *The Journal of biological chemistry* 288, 9549-9562.

Wek, R.C. (2018). Role of eIF2alpha Kinases in Translational Control and Adaptation to Cellular Stress. *Cold Spring Harb Perspect Biol* 10.

Wek, S.A., Zhu, S., and Wek, R.C. (1995). The histidyl-tRNA synthetase-related sequence in the eIF-2 alpha protein kinase GCN2 interacts with tRNA and is required for activation in response to starvation for different amino acids. *Mol Cell Biol* 15, 4497-4506.

Werner, M., Feller, A., Messenguy, F., and Pierard, A. (1987). The leader peptide of yeast gene CPA1 is essential for the translational repression of its expression. *Cell* 49, 805-813.

Werner, M., Feller, A., and Pierard, A. (1985). Nucleotide sequence of yeast gene CP A1 encoding the small subunit of arginine-pathway carbamoyl-phosphate synthetase. Homology of the deduced amino acid sequence to other glutamine amidotransferases. *Eur J Biochem* *146*, 371-381.

West, S., Proudfoot, N.J., and Dye, M.J. (2008). Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol Cell* *29*, 600-610.

Wheeler, G.L., Jones, M.A., and Smirnoff, N. (1998). The biosynthetic pathway of vitamin C in higher plants. *Nature* *393*, 365-369.

Wilson, D.N., Arenz, S., and Beckmann, R. (2016). Translation regulation via nascent polypeptide-mediated ribosome stalling. *Curr Opin Struct Biol* *37*, 123-133.

Wilson, D.N., and Beckmann, R. (2011). The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr Opin Struct Biol* *21*, 274-282.

Winkler, W.C., and Breaker, R.R. (2003). Genetic control by metabolite-binding riboswitches. *ChemBiochem* *4*, 1024-1032.

Wohlgemuth, I., Brenner, S., Beringer, M., and Rodnina, M.V. (2008). Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *The Journal of biological chemistry* *283*, 32229-32235.

Wolucka, B.A., and Van Montagu, M. (2003). GDP-mannose 3',5'-epimerase forms GDP-L-gulose, a putative intermediate for the de novo biosynthesis of vitamin C in plants. *The Journal of biological chemistry* 278, 47483-47490.

Woolstenhulme, C.J., Guydosh, N.R., Green, R., and Buskirk, A.R. (2015). High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep* 11, 13-21.

Wu, C., Wei, J., Lin, P.J., Tu, L., Deutsch, C., Johnson, A.E., and Sachs, M.S. (2012). Arginine changes the conformation of the arginine attenuator peptide relative to the ribosome tunnel. *J Mol Biol* 416, 518-533.

Xaplanteri, M.A., Petropoulos, A.D., Dinos, G.P., and Kalpaxis, D.L. (2005). Localization of spermine binding sites in 23S rRNA by photoaffinity labeling: parsing the spermine contribution to ribosomal 50S subunit functions. *Nucleic Acids Res* 33, 2792-2805.

Yabuta, Y., Mieda, T., Rapolu, M., Nakamura, A., Motoki, T., Maruta, T., Yoshimura, K., Ishikawa, T., and Shigeoka, S. (2007). Light regulation of ascorbate biosynthesis is dependent on the photosynthetic electron transport chain but independent of sugars in *Arabidopsis*. *J Exp Bot* 58, 2661-2671.

Yamashita, Y., Takamatsu, S., Glasbrenner, M., Becker, T., Naito, S., and Beckmann, R. (2017). Sucrose sensing through nascent peptide-mediated ribosome stalling at the stop codon of *Arabidopsis* bZIP11 uORF2. *FEBS Lett* 591, 1266-1277.

Yang, F., Sun, S., Tan, G., Costanzo, M., Hill, D.E., Vidal, M., Andrews, B.J., Boone, C., and Roth, F.P. (2017). Identifying pathogenicity of human variants via paralog-based yeast complementation. *PLoS Genet* *13*, e1006779.

Yatime, L., Mechulam, Y., Blanquet, S., and Schmitt, E. (2007). Structure of an archaeal heterotrimeric initiation factor 2 reveals a nucleotide state between the GTP and the GDP states. *Proc Natl Acad Sci U S A* *104*, 18445-18450.

Yoon, H.J., and Donahue, T.F. (1992). The suil suppressor locus in *Saccharomyces cerevisiae* encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. *Mol Cell Biol* *12*, 248-260.

Young, D.J., Guydosh, N.R., Zhang, F., Hinnebusch, A.G., and Green, R. (2015a). Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3'UTRs In Vivo. *Cell* *162*, 872-884.

Young, S.K., Palam, L.R., Wu, C., Sachs, M.S., and Wek, R.C. (2016). Ribosome Elongation Stall Directs Gene-specific Translation in the Integrated Stress Response. *The Journal of biological chemistry* *291*, 6546-6558.

Young, S.K., and Wek, R.C. (2016). Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. *The Journal of biological chemistry* *291*, 16927-16935.

Young, S.K., Willy, J.A., Wu, C., Sachs, M.S., and Wek, R.C. (2015b). Ribosome Reinitiation Directs Gene-specific Translation and Regulates the Integrated Stress Response. *The Journal of biological chemistry* 290, 28257-28271.

Young-Baird, S.K., Lourenco, M.B., Elder, M.K., Klann, E., Liebau, S., and Dever, T.E. (2019a). Suppression of MEHMO Syndrome Mutation in eIF2 by Small Molecule ISRIB. *Mol Cell*.

Young-Baird, S.K., Lourenco, M.B., Elder, M.K., Klann, E., Liebau, S., and Dever, T.E. (2020). Suppression of MEHMO Syndrome Mutation in eIF2 by Small Molecule ISRIB. *Mol Cell* 77, 875-886 e877.

Young-Baird, S.K., Shin, B.S., and Dever, T.E. (2019b). MEHMO syndrome mutation EIF2S3-I259M impairs initiator Met-tRNA^{iMet} binding to eukaryotic translation initiation factor eIF2. *Nucleic Acids Res* 47, 855-867.

Yu, C.H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M.S., and Liu, Y. (2015). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Mol Cell* 59, 744-754.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature structural & molecular biology* 16, 274-280.

Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C. (2018). Genome editing of upstream open reading frames enables translational control in plants. *Nat Biotechnol* 36, 894-898.

Zhang, J., Pan, X., Yan, K., Sun, S., Gao, N., and Sui, S.F. (2015). Mechanisms of ribosome stalling by SecM at multiple elongation steps. *Elife* 4.

Zhang, P., McGrath, B.C., Reinert, J., Olsen, D.S., Lei, L., Gill, S., Wek, S.A., Vattam, K.M., Wek, R.C., Kimball, S.R., *et al.* (2002). The GCN2 eIF2alpha kinase is required for adaptation to amino acid deprivation in mice. *Mol Cell Biol* 22, 6681-6688.

Zhang, Y. (2013). *Ascorbic Acid in Plants* (Springer).

Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* 407, 401-405.

Zinder, J.C., and Lima, C.D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* 31, 88-100.

Zinszner, H., Kuroda, M., Wang, X., Batchvarova, N., Lightfoot, R.T., Remotti, H., Stevens, J.L., and Ron, D. (1998). CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. *Genes Dev* 12, 982-995.

Zitomer, R.S., Walthall, D.A., Rymond, B.C., and Hollenberg, C.P. (1984). *Saccharomyces cerevisiae* ribosomes recognize non-AUG initiation codons. *Mol Cell Biol* 4, 1191-1197.

Zyryanova, A.F., Weis, F., Faille, A., Alard, A.A., Crespillo-Casado, A., Sekine, Y., Harding, H.P., Allen, F., Parts, L., Fromont, C., *et al.* (2018). Binding of ISRIB reveals a regulatory site in the nucleotide exchange factor eIF2B. *Science* 359, 1533-1536.

7. Curriculum vitae

Leda Eleanor Lotspeich-Cole

Education

- Johns Hopkins University-National Institutes of Health Graduate** 2020
Partnership Program, Baltimore, MD and Bethesda, MD, *Ph.D.*,
Biology
- Johns Hopkins University Bloomberg School of Public Health,** 2011
Baltimore, MD, *Sc.M.*, *Molecular Microbiology and Immunology*
Vaccine Science and Policy Certificate 2010
- University of Alaska Fairbanks**, Fairbanks, AK, *B.S. magna cum* 2009
laude, B.S., *Biological Sciences*

Experience

- National Institutes of Health, Eunice Kennedy Shriver National* 2014-present
Institute of Child Health and Human Development, Bethesda, MD
Predoctoral Fellow, Principle Investigator: Thomas Dever, Ph.D.
- National Institutes of Health, National Institutes of Allergy and* 2013-2014
Infectious Diseases, Bethesda, MD
Predoctoral Fellow, Principle Investigator: Stefan Muljo, Ph.D.

United States Food and Drug Administration, Center for Biologics **2011-2012**
Evaluation and Research, Bethesda, MD

ORISE Fellow, Principle Investigator: Sheldon Morris, Ph.D.

The Johns Hopkins University Bloomberg School of Public Health, **2009-2011**
Baltimore, MD

Graduate Student, Principle Investigator: Egbert Hoicyk, Ph.D.

University of Alaska Fairbanks, Institute of Arctic Biology, Fairbanks, **2008-2009**
AK

Undergraduate Research Assistant, Principle Investigator: Karsten
Hueffer, D.V.M, Ph.D.

Publications

1. *EIF2S3* mutations associated with severe X-linked intellectual disability syndrome MEHMO. Skopkova M, Hennig F, Shin BS, Turner CE, Stanikova D, Brennerova K, Stanik J, Fischer U, Henden L, Müller U, Steinberger D, Kurdiova T, Ukropec J, Nyitrayova O, Kolnikova K, Klimes I, Borck G, Bahlo M, Haas SA, Kim JR, **Lotspeich-Cole LE**, Gasperikova D, Dever TE, Kalscheuer VM. *Hum Mutat*, 2017.
2. T-bet modulates the antibody response and immune protection during murine malaria. Oakley MS, Sahu B, **Lotspeich-Cole L**, Majam V, Pham PT,

Sengupta Banerjee A, Kozakai Y, Morris SL, and Kumar S. *Eur J Immunol*, 2014.

3. The transcription factor T-bet regulated parasite and promotes pathogenesis during *Plasmodium berghei* ANKA murine malaria. Oakley MS, Sahu B, **Lotspeich-Cole L**, Solanki N, Majam V, Pham PT, Banerjee R, Kozaki Y, Derrick SC, Kumar S, and Morris SL. *Journal Immunol.*, 2013.
4. Oakley MS, Gerald N, Anantharaman V, Gao Y, Majam V, Mahajan B, Pham PT, **Lotspeich-Cole L**, Meyers TG, McCutchan TF, Morris SL, Aravind L, Kumar S. Radiation Induced Cellular and Molecular Alterations in Asexual Intraerythrocytic *Plasmodium falciparum* Parasites. *J Infect Dis.*, 2013.
5. Detection of *Francisella tularensis* in Alaskan mosquitoes (Diptera: Culicidae) and assessment of a laboratory model for transmission. Triebenbach AN, Vogl SJ, **Lotspeich-Cole L**, Sikes DS, Happ GM, Hueffer K. *J Med Entomol*, 2010.