Translational buffering by ribosome stalling in upstream open reading frames

Ty Alexander Bottorff

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctorate of Philosophy

University of Washington

2023

Reading Committee:

Rasi Subramaniam, Chair

Adam Geballe

Andrew Hsieh

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Translational buffering by ribosome stalling in upstream open reading frames

Ty Alexander Bottorff

Chair of the Supervisory Committee:

Rasi Subramaniam

Department of Biochemistry

Gene expression is the process by which information encoded in DNA is read to create functional proteins within a cell through DNA transcription and subsequent mRNA translation. Eukaryotic gene expression is highly complex and regulated and influenced by environmental signals, developmental cues, and feedback mechanisms at both the transcriptional and post-transcriptional levels. Upstream open reading frames (uORFs) are present in over half of all human mRNAs and can potently regulate the translation of downstream open reading frames. This regulation can operate through several mechanisms: siphoning away scanning ribosomes, regulating re-initiation, and allowing interactions between scanning and elongating ribosomes. However, the consequences of these different mechanisms for the regulation of protein expression remain incompletely understood. Here, we use computational modeling to help elucidate the effects of these different regulatory mechanisms on the regulation of protein expression. We performed systematic experimental measurements on the uORF-containing 5′ UTR of the cytomegaloviral *UL4* mRNA to test alternative models of uORF-mediated regulation in human cells. We find that a terminal diproline-dependent elongating ribosome stall in the *UL4* uORF prevents decreases in main ORF protein expression when ribosome loading onto the mRNA is reduced. This uORF-mediated buffering is insensitive to the location of the ribosome stall along the uORF. Computational kinetic modeling based on our measurements suggests that scanning ribosomes dissociate rather than queue when they collide with stalled elongating ribosomes within the *UL4* uORF. We propose that ribosome stalls in uORFs provide a general mechanism for buffering against reductions in main ORF translation during stress and develop-

mental transitions.

Table of Contents

List of Figures and Tables

Chapter 2

Chapter 3

Acknowledgements

Thank you to my parents, Carol and Steve, for their love and support in my education and life.

Thank you to the Subramaniam lab for fostering such a supportive and collaborative environment in which to perform research.

Finally, thank you to Rasi for his mentorship, flexibility, and guidance throughout my graduate career.

Chapter 1: Introduction

I. An overview of eukaryotic gene expression

Gene expression is the process by which information encoded in DNA is read to create functional proteins within a cell. Proteins are the 'workhorses of all living organisms'; they perform a wide range of functions including enzymatic activity, cell signaling, structural support, immune function, and regulation of gene expression[1]. Only established in the 1960s, the central dogma of molecular biology is a framework that explains genetic information encoded in DNA can be used to create proteins through the process of gene expression[2,3]. This dogma states that DNA serves as a template for the synthesis of a messenger RNA (mRNA) via the process of transcription[4].

Genes reside on chromosomes, 250 million to 2.5 billion base-pair long linear DNA molecules that package with histone proteins to compact into chromatin able to fit within nuclei[5]. Gene expression begins with the transcription of the DNA sequence of a gene into a pre-mRNA sequence by the enzyme RNA polymerase in the nucleus; genes are units of transcribed DNA[4]. Transcription machinery, transcription factors and RNA polymerase, must overcome the structural barrier of compacted chromatin to access the DNA sequence of a gene[6–8]. Once the structural barrier of compacted chromatin is surpassed, transcription factors recognize and bind the promoter region of a gene, a short DNA sequence that is recognized as the location to initiate transcription. Transcription factor binding regulates the recruitment of RNA polymerase; once bound, RNA polymerase synthesizes pre-mRNAs.

Eukaryotic pre-mRNAs are processed and modified before they leave the nucleus and enter the cytoplasm as mature mRNAs[9]. DNA and pre-processed mRNA sequences contain both exons, sequences intended to encode protein sequences, and introns, sequences not intended to encode protein sequences[10]. pre-mRNA processing involves the splicing of exons together, as introns as removed, and the addition of a 5' 7-methylguanosine cap and a poly(A) tail on its 3' end[4]. To generate poly(A) tails, RNAs are co-transcriptionally cleaved at poly(A) sites before being polyadenylated[11]. Deadenylated transcripts are not translated well and are substrates for decapping[12]. Processed, mature mRNAs are then exported into the cytoplasm through nuclear pores[13].

Eukaryotic gene expression is highly complex and regulated, allowing the integration of many different signaling and regulatory pathways[8]. Various factors can influence the expression of a gene, including environmental signals, developmental cues, and feedback mechanisms[14,15]. Gene expression is regulated at multiple levels: transcriptional, post-transcriptional, translational, and post-translational[16–18]. Transcriptional regulation includes the activity of transcriptional coactivators that can facilitate the recruitment of general transcription factors and RNA polymerase to the promoter region[19]. Additionally, chromatin is not a static barrier; transcription machinery assembly can also be regulated by changes to the local chromatin environment changes, such as by histone acetylation or methylation. Both the modification and the specific histone residue that is modified together determine if transcription is repressed or activated[19,20].

Post-transcriptional regulation involves the control of mRNA splicing, stability, and translation[21,22]. Alternative RNA splicing is a common regulational mechanism that allows different protein products to be produced from a single gene via the differential combination of exons, and sometimes introns[23,24]. In fact, 70% of human genes undergo alternative splicing. With respect to mRNA stability, mRNA half-lives vary over 100-fold across eukaryotic transcripts[25]. Variation in these mRNA half-lives changes the length of time an mRNA is present in the cytoplasm; longer mRNA half-lives leads to higher protein levels as the mRNA can be translated for a longer period of time before it is degraded[26,27]. Additionally, both proteins and RNAs can bind to RNAs to control their stability[28]. For example, microRNAs, short 21-24 nucleotide RNAs, bind the 3' UTR of RNAs to induce their decay[29–33]. RNA modifications, such as m6A, can also influence mRNA stability[34]. HuR and AUF1 are two exemplary RNA-binding proteins (RBPs) that regulate RNA decay through AU-rich elements[35,36].

II. A closer look at mRNA translation

mRNA translation is the process by which ribosomes use mRNA molecules as templates to synthesize proteins, polymers of amino acids[4,37]. The protein amino acid sequence, determined by the mRNA nucleotide sequence and, therefore also, the DNA nucleotide sequence, determines the structure and therefore the function of the protein[38]. This process occurs in three general stages: initiation, elongation, and termination. Initiation begins with the assembly of a 43S pre-initiation complex (PIC) of the small riboso-

mal subunit, a ternary complex (eIF2-GTP-Met-tRNAMeti), and various eIFs (1, 1A, 3, and 5)[39,40]. Then, canonical cap-dependent initiation proceeds with the small ribosomal subunit binding to the 5' end of the mRNA molecule with the support of eIF4A, eIF4B, and eIF4F, which together help 5' terminal secondary structure to facilitate 43S loading onto the mRNA[39]. The eIF4F cap-binding complex is comprised of the cap-binding eIF4E, RNA helicase eIF4A, and scaffold eIF4G, which also binds eIF3 and poly(A) binding protein (PABP)[39]. Once bound, the small ribosomal subunit scans along the 5' UTR of the mRNA molecule until it selects a start codon. The nucleotide sequence (Kozak) context around potential start codons affects the efficiency of start codon selection; start codons with weaker Kozak contexts are more often bypassed and not selected in a process termed leaky scanning[41–43].

Once the small ribosomal subunit selects a start codon, eIF5 and eIF5B promote eIF2-GTP hydrolysis, and a large ribosomal subunit (60S) freed from the anti-association factor eIF6 (that blocks premature small ribosomal subunit-large ribosomal subunit joining) joins to form the 80S elongating ribosome[39,44]. The ribosome has three sites of interest: the A (acceptor) site, where the next aminoacyl-tRNA (tRNA with an attached amino acid) binds; the P (peptidyl) site, where the peptidyl-tRNA is bound; and the E (exit) site, where the deacylated tRNA is bound[45,46]. Initiation ends with Met-tRNAMeti bound in the P site of the 80S ribosome with the tRNA's anticodon base-paired with the start codon of the mRNA and the second ORF codon in the A site[47].

Then, elongation begins with an aminoacyl-tRNA binding the A site depending on base-pairing interactions between its anticodon and the A site codon[48]; these base-pairing interactions trigger GTP hydrolysis by eEF1A[47]. Next, the nascent peptide is transferred from the peptidyl-tRNA in the P site to the amino group of the A site aminoacyl-tRNA to form a peptide bond and extend the peptidyl-tRNA[47]. After the peptide bond is formed, the uniquely dipthamide modified eEF2-GTP binds in the A site and promotes the translocation of the now deacylated tRNA into the E site of the ribosome[47,49]. Then, the ribosome moves along the mRNA molecule to position the next codon in the A site, and eEF2-GDP is released[47].

Ribosomes elongate at an average speed in the order of 10 amino acids per second, but ribosomes do not elongate at uniform speeds[50]. For instance, translation of sequential proline codons, a particularly

poor peptidyl donor and acceptor, can slow down ribosomes and necessitates the support of a specific elongation factor, eIF5A[50]. Damaged mRNA, higher-order mRNA structure, rare codons, and amino acid starvation can also induce ribosome stalls[51,52]. Sensibly, cells have evolved quality control mechanisms to respond to such ribosome stalls[53]. Ribosome stalls are typically recognized when a 5' ribosome collides with the 3' stalled ribosome; then, quality control mechanisms such as ribosome quality control (RQC) can respond by, for example, inducing co-translational degradation of the arrested peptide and recycling the stalled ribosome[51].

When a stop codon enters the A site of the ribosome, termination begins, mediated by a ternary complex of GTP and release factors eRF1 and eRF3[54]. eRF1 recognizes the stop codon, eRF3 hydrolyzes the GTP, and the nascent peptide is released from the peptidyl-tRNA in the P site[55]. ABCE1 splits the 80S ribosome to release the large ribosomal subunit[54]; the small ribosomal subunit may continue scanning for another downstream start codon in a process termed re-initiation[56]. The completed protein may undergo further modifications, such as completing folding or post-translational modifications, before it can assume its functional form[57].

Open reading frames in the 5' UTR, more accurately termed the 5' leader, are termed upstream open reading frames (uORFs). uORFs are present in about half of eukaryotic mRNAs[39] and are enriched in oncogene transcripts[42]. Following uORF translation and large ribosomal subunit dissociation, small ribosomal subunits can continue scanning for and re-initiate at downstream start codons given additional ternary complex binding[39]. uORFs are usually short stretches of codons; short uORF lengths help maintain high levels of re-initiation as eIFs are less likely to dissociate the less time it takes to translate a uORF[39]. uORF peptide sequences are not under strong selection to be maintained[58]. Accordingly, the synthesized small uORF peptides do not always have a biological function. However, the presence of uORFs does repress the expression of the main coding sequence of the mRNA in *cis* by reducing the number of ribosomes that make it to the main ORF start codon.

Post-translational gene regulation involves changes to protein activity, localization, and longevity via post-translational modifications: the addition or removal of small chemical groups such as methyl, phosphate,

acetyl, and ubiquitin groups[59]. For instance, ubiquitin marks proteins for degradation by the proteasome[60].

III. The role of computational modeling in understanding gene expression

Gene expression is a dynamic process that involves numerous molecular interactions, feedback mechanisms, and regulatory factors; this complexity makes it difficult to understand without the aid of experimental and computational techniques. Computational modeling has emerged as a powerful tool for understanding complex processes that can be difficult to directly probe experimentally[61–67]. Computational models aim to replicate a system's behavior by integrating the underlying mechanisms and interactions that govern the system. In doing so, modeling can serve a wide variety of roles for scientists, such as aiding in hypothesis testing, generating novel insights, and, importantly, demonstrating if proposed mechanisms explain an observed phenomenon[68–70]. For instance, modeling of mRNA turnover in yeast suggested that half-life measurements underestimated the true time that a mRNA persists in cells[71] and modeling of *Drosophila* germ granule development offered insights into how mRNA compositions arise[72]. Generally, computational modeling and visualization of gene expression can help elucidate the underlying mechanisms and provide a deeper understanding of how genes are expressed and regulated.

Most biologists are familiar with conceptual types of modeling maps, such as those depicting cellular regulatory networks of interacting proteins, such as common signaling cascades. These maps can be limited by their vague representation of the underlying mechanisms and interactions, and are often not intuitive to interpret given the complexities underlying these systems, including molecular promiscuity, compartmentalization, and binding competition, among others[73]. To standardize these efforts, previous studies have moved to computational models using a range of approaches, including statistical models, network models, and ordinary differential equation (ODE) models. While these approaches have been successful in generating insights into gene expression, they also have certain limitations.

Statistical models, for instance, are limited by the quality and quantity of the data used in the analysis, and may not capture the underlying mechanistic details. Statistical models grew and became known as Totally Asymmetric Exclusion Process (TASEP) models, which initially were used to model car traffic flow but have become widely used in translation modeling[74–77]. Network models, on the other hand, are limited by the

complexity of the interactions between the various components, and may not provide a complete picture of the underlying mechanisms. ODE models are difficult to use given their requirement of enumerating all potentially populated chemical species in a system[73], which can be numerous. For example, there are numerous potentially populated chemical species in an mRNA translation system as multiple ribosomes can exist on a single mRNA at the same time at different positions; this problem is referred to as combinatorial complexity[73]. Following this mRNA translation example, the number of these species grows exponentially with the number of ribosomes that can fit on the mRNA and becomes "essentially intractable" according to an early proponent of computer simulations, Richard Gordon[74].

Rule-based modeling is a promising approach for modeling gene expression given its ability to surpass the problem of combinatorial complexity by avoiding enumerating all potentially populated chemical species, as required by ODE models[73]. Rule-based modeling is a computational method that uses a set of rules to describe the interactions between the various components of a signaling pathway. These rules are based on the properties of the components, such as their binding affinities, and can be used to simulate the dynamics of the pathway over time.

Rule-based modeling represents biomolecular site dynamics very well given that rules include site information[73,78–82]. Given that rule-based models are based on assumptions of the modularity of interactions, model size correlates with the number of interactions rather than the number of tracked species as in equation-based models[73]. A specific variant of the Monte Carlo approach, agent-based modeling, can be used to simulate rule-based models. In this approach, modeled particles, agents such as ribosomes or mRNA molecules, are represented by individual variables, rather than pools of particles as in non-agent-based approaches. Agent-based modeling allows modification of variable states, for example a mRNA 5' cap being blocked or free[74]. While rule-based modeling has mostly been used in the context of cell signaling, here, we expand on the simulation of rule-based models to also simulate gene expression.

V. Thesis Objectives

Rationale

Here, we are interested in understanding the role of uORFs in regulating protein synthesis. uORFs have been of interest for decades given their prevalence, relevance in human disease, and the many possible repressive regulatory mechanisms they offer[58,83–90]. uORFs can siphon away scanning ribosomes from initiating at downstream main ORFs. Multiple uORFs can together regulate the re-initiation frequency at the main ORF[91]. Although the initiation rate usually limits translation[92–94], inefficient elongation or termination on uORFs can also repress protein expression by sterically blocking scanning ribosomes from reaching the main ORF[52,95–99]. Such inefficient elongation or termination can be driven by the nascent uORF peptide[100,101], poorly translated codons in the uORF[102,103], or small molecules such as amino acids or polyamines[97,98]. Scanning and elongating ribosomes can also interact together, necessitating additional eIFs and possibly causing dissociation of scanning ribosomes or enhanced initiation at start codons[97,104–107]. uORFs can also sometimes accelerate mRNA decay[108,109].

Thus, many uORF regulatory mechanisms have been proposed, but their implications for the regulation of protein expression are not clear. For example, it is unclear if different uORF regulatory mechanisms differentially regulate mRNA translation across different biochemical parameters. The goal of this work is to use computational modeling to explore the many implications of the regulation of protein expression by uORFs. We hypothesized that computational modeling could be used to gain insight into the interplay between these regulatory mechanisms and how they affect the regulation of mRNA translation by uORFs.

Approach

Computational modeling can synthesize the different steps of translation together to simulate how the different regulatory mechanisms influence the overall rates of uORF and main ORF translation. However, the rule-based computational models described here have historically mostly been used for simulating cell signaling; here, we adapted such modeling frameworks to also simulate gene expression.

Here, we use experimental measurements on the well-studied uORF-containing 5′ UTR of the human cytomegaloviral *UL4* mRNA to test different kinetic models of uORF-mediated translational control[95]. The

second uORF (henceforth uORF2) in the *UL4* 5′ UTR contains a terminal diproline motif that stalls 80S ribosomes by disrupting peptidyl transferase center activity[100,101]. We use an experimentally-integrated modeling approach and find that the presence of a terminal diproline-dependent 80S stall in uORF2 of *UL4* 5' UTR confers buffering of main ORF translation to reduced ribosome loading on the mRNA. Modeling suggests that collisions of scanning ribosomes with the stalled 80S ribosome confer this buffering behavior. Experimental variation of the distance between the uORF2 start codon and the main ORF start codon suggests that the buffering is insensitive to the location of the ribosome stall along the uORF and therefore supports a kinetic model of ribosome dissociation rather than queueing. We identify several human uORFs that repress main ORF translation via a similar terminal diproline motif to that of *UL4* uORF2.

Chapter 2: Translational buffering by ribosome stalling in upstream open reading frames

Portions published in: Bottorff T., Park H., Geballe AP, Subramaniam AR. Translational buffering by ribosome stalling in upstream open reading frames. *PLOS GENETICS* 2022, https://doi.org/10.1371/journal.pgen.1010460.

I. Introduction

About half of human mRNAs have at least one upstream open reading frame (uORF) in their 5′ untranslated region[110–112]. Ribosome profiling studies estimate that at least twenty percent of these uORFs are actively translated[87,113]. uORFs can regulate gene expression via the biological activity of the uORF peptide, but they also often *cis*-regulate translation of the downstream main ORF[114,115]. Despite having poor initiation sequence contexts, many eukaryotic uORFs repress main ORF translation[42,87,110,112,115–118]. uORF mutations are implicated in several human diseases via changes to main ORF translation[58,119]. For example, uORF mutations in oncogenes and tumor suppressors can act as driver mutations in cancer[120,121].

uORFs can regulate translation via a variety of molecular mechanisms. uORFs can constitutively repress translation by siphoning away scanning ribosomes from initiating at downstream main ORFs. Multiple uORFs can interact together to regulate the re-initiation frequency at the main ORF. For example, uORFs in the *S. cerevisiae GCN4* mRNA and the homologous human *ATF4* mRNA render main ORF translation sensitive to cellular levels of the eIF2α-GTP-$tRNA_{Met}$ ternary complex[91,122]. Although the initiation rate usually limits translation[92–94], inefficient elongation or termination on uORFs can also regulate protein expression by preventing scanning ribosomes from reaching the main ORF[52,95–99]. Inefficient elongation can be driven by the nascent uORF peptide[100,101], poorly translated codons in the uORF[102,103], or small molecules such as amino acids or polyamines[97,98]. Further, interactions between scanning and elongating ribosomes on uORFs may cause dissociation of scanning ribosomes or enhanced initiation at start codons[97,104,105].

Despite the plethora of proposed uORF regulatory mechanisms, their implications for the regulation of protein expression are not clear. For example, are some uORF regulatory mechanisms more effective than others at repressing protein expression across a wide range of biochemical parameters? How do uORFs

alter the response of main ORF translation to changes in cellular and environmental conditions? Answering these questions requires a joint accounting of how the different steps of translation, such as initiation, scanning, and elongation, together influence the overall rates of uORF and main ORF translation. Since it is not straightforward to monitor the rates of individual steps of translation[123], indirect measurements of protein expression are often necessary to infer the underlying mechanism of uORF-mediated regulation. Such inference requires rigorous kinetic models of uORF regulation that make testable experimental predictions for the effects of genetic mutations on protein expression.

Computational kinetic modeling has been widely used to study mechanisms of translational control[124]. Quantitative modeling of uORF translation has been used to support the regulated re-initiation model for the *GCN4* mRNA[125–127]. A computational model predicted that elongating ribosomes can dislodge leading scanning ribosomes on uORFs and confer stress resistance to protein expression[75]. However, these models have not been compared against alternative models of uORF regulation that predict queuing or dissociation of scanning ribosomes upon collision with paused elongating ribosomes[95,97]. A critical barrier to such comparison has been the lack of a computational framework for the specification and simulation of different kinetic models of uORF-mediated translational regulation. Such a computational framework is necessary for the identification of unique experimental signatures of each proposed model and their comparison with experimental measurements. Even though simulation code has been made available in many computational studies of mRNA translation[75,92], it is often highly tailored for specific models and cannot be easily modified to consider alternative regulatory mechanisms.

Here, we use experimental measurements on the well-studied uORF-containing 5′ UTR of the human cytomegaloviral *UL4* mRNA to test different kinetic models of uORF-mediated translational control[95]. The second uORF (henceforth uORF2) in the *UL4* 5′ UTR contains a terminal diproline motif that stalls 80S ribosomes by disrupting peptidyl transferase center activity[100,101]. For systematic model comparisons, we rely on a recent computational framework that allows easy specification and efficient simulation of arbitrary kinetic models of translational control[128]. Using this experimentally-integrated modeling approach, we find that the presence of 80S stalls in uORF2 of *UL4* 5′ UTR confers resistance (henceforth called buffering)

of main ORF translation to reduced ribosome loading on the mRNA. Modeling suggests that collisions of scanning ribosomes with the stalled 80S ribosome confer this buffering behavior. Experimental variation of the distance between the uORF2 start codon and the elongating ribosome stall supports a kinetic model in which scanning ribosomes dissociate rather than queue upon colliding with the 80S stall. We also identify several human uORFs that have repressive terminal diproline motifs similar to the *UL4* uORF2 80S stall. We propose that ribosome stalls in uORFs enable buffering of main ORF protein expression against reduced ribosome loading across cellular and environmental transitions. Together, our results illustrate the value of experimentally-integrated kinetic modeling for the comparison of different uORF regulatory mechanisms and the identification of novel experimental signatures from complex molecular interactions.

II. Results

Models of uORF regulation of main ORF translation

We surveyed five previously proposed models of uORF regulation of main ORF translation (Fig 2.1). We tested these models using a combination of computational modeling and experimental reporter assays. In the constitutive repression model[117] (Fig 2.1A), uORFs siphon away scanning ribosomes from the main ORF since re-initiation is usually infrequent[129–132]. In the 80S-hit dissociation model[75] (Fig 2.1B), elongating ribosomes that hit downstream scanning ribosomes cause the 3′ scanning ribosomes to dissociate from the mRNA. In the queuing-mediated enhanced repression model[97] (Fig 2.1C), a stalled elongating ribosome within the uORF allows upstream scanning ribosomes to queue in the 5′ region. This queuing can bias scanning ribosomes to initiate translation at the uORF start codon rather than leaky scan past it. In the collision-mediated 40S dissociation model[104,105] (Fig 2.1D), scanning ribosomes instead dissociate if they collide with a 3′ stalled elongating ribosome.

Lastly, in the regulated re-initiation model[91,133,134] (Fig 2.1E), for example in the *GCN4* (*S. cerevisiae* homolog of human *ATF4*) mRNA, translation of the first uORF is followed by re-initiation at either a second downstream uORF or the main ORF depending on the stress status of the cell. After termination at the first uORF stop codon, scanning ribosomes must reacquire a new eIF2α-GTP-$tRNA_{Met}$ ternary complex (TC) before re-initiating at a downstream start codon. The time to reacquire a new TC correlates with the proportion of phosphorylated eIF2α. Therefore, when cells are not stressed and the proportion of phosphorylated eIF2α is lower, translation of the first uORF is followed by re-initiation at the second downstream uORF start codon. Alternatively, when cells are stressed and the proportion of phosphorylated eIF2α is higher, translation of the first uORF is instead followed by re-initiation at the main ORF start codon.

Experimental system for testing different models of uORF-mediated translational regulation

To differentiate between proposed models of uORF regulation (Fig 2.1), we used the well-studied human cytomegaloviral *UL4* uORF2[104] as an experimental model (Fig 2.2A). uORF2 represses main ORF translation via an elongating ribosome stall that is dependent on the uORF2 peptide sequence[95] (Fig 2.2A, irrelevant uORFs boxed in white, key uORF2 boxed in green). The two C-terminal proline residues (un-

derlined and labeled as uORF2 stall in Fig 2.2A), regardless of codon usage, in uORF2 are necessary for the elongating ribosome stall[105]. These residues are poor substrates for nucleophilic attack to generate a peptide bond and also reorient the ribosomal peptidyl transferase center to reduce termination activity[101]. Termination activity is further reduced through interactions between the uORF2 nascent peptide and the GGQ motif within eRF1[135]. Even though the A-site of the uORF2-stalled ribosome is occupied by a stop codon, we refer to it as an elongating ribosome stall since they are functionally equivalent for this study. This terminology is also inclusive of elongation stalls within other uORFs[96,97,136–140]. The 5′ leader region preceding the *UL4* coding sequence contains two other uORFs besides uORF2. uORF1 slightly reduces uORF2 repressiveness by siphoning scanning ribosomes away from uORF2, and uORF3 is irrelevant for repression[104].

We inserted the uORF2-containing *UL4* leader sequence into a dual-luciferase reporter system (Fig 2.2B) in which nanoluciferase (NLuc) signal provides a readout of uORF2 repressiveness and firefly luciferase (FLuc) signal serves as normalization for transfection efficiency. This experimental platform can detect differences in luciferase activity over a 1000-fold range (Fig 2.3). We confirmed that uORF2 repressiveness depends on its translation and the terminal diproline-dependent elongating ribosome stall (Fig 2.2C). Near-cognate start codons within uORF2 do not contribute to the uORF2 repressiveness (Fig 2.3). We used this *UL4*-based luciferase reporter to quantitatively dissect the kinetics of uORF2-mediated translational regulation.

We complemented our experimental measurements with computational kinetic modeling of proposed models of uORF regulation (Fig 2.1). We aimed to find unique modeling predictions that would allow us to experimentally distinguish between the different models of uORF regulation. We specified the kinetics of each of the proposed models of uORF regulation using PySB, a framework for the compact specification of rule-based models[141]. We then expanded the model into the BioNetGen modeling language syntax[142] and inferred a reaction dependency graph for efficient simulation[128]. Next, we stochastically simulated the models using an agent-based Gillespie algorithm implemented in NFSim[143]. The molecules and reactions within the kinetic model are shown in Fig 2.4A and Fig 2.4B, respectively, and are described in detail in the

methods section. We experimentally tested predictions from this computational modeling and used the results to refine our model specifications. This iterative cycle of experimental testing and computational modeling constituted our platform for differentiating between proposed uORF regulatory models.

To derive estimates for unknown parameters ('This work' in Table 1), we first calibrated our computational models to our reporter measurements on wild-type or mutant uORF2 (Fig 2.2C). We did not fit the constitutive repression and regulated re-initiation models (Fig 2.1A,E) to our reporter measurements (Fig 2.2C) since these models cannot account for the critical role of the *UL4* uORF2 elongating ribosome stall in regulating main ORF translation in single uORF-containing mRNAs. We used previously generated estimates for kinetic parameters not directly inferred in our work (Table 1).

Simulations of the queuing-mediated enhanced repression (Fig 2.1C) and collision-mediated 40S dissociation (Fig 2.1D) models readily recapitulate measurements of NLuc protein output from wild-type and mutant *UL4* reporters (Fig 2.2D, triangles and squares). The 80S-hit dissociation model (Fig 2.1B), modified to include an elongating ribosome stall within the uORF, also recapitulates the reporter measurements (Fig 2.2D, circles). However, this modified 80S-hit dissociation model requires the difference between the stronger Kozak and wild-type uORF initiation fractions to be quite large (80% vs. 2% compared to 50% vs. 10% for 2 other models mentioned above, Table 1). The derived ribosome loading rates (~0.02/s for all three of these models (Fig 2.1B-D) are in line with literature estimates[144–146]. The re-initiation fractions derived here (50-70%, Table 1) are within the range of measured values across mRNAs with different sequence features[129–132]. A complete description of the derivation of model parameters can be found in the methods section.

Computational modeling predicts that different models of uORF regulation have unique parameters important for buffering

Following the calibration of our computational models to recapitulate experimental data, we used these models to predict how translation would be perturbed upon varying other kinetic parameters. While many kinetic parameters could be varied to help distinguish between proposed models of uORF regulation (Fig 2.1), we honed in on the rate of ribosome loading onto the mRNA for two key reasons. Firstly, this rate is re-

duced endogenously in response to a variety of cellular and environmental signals. Amino acid deprivation, ribosome collisions, dsRNA viral infection, unfolded proteins, and heme deprivation are sensed by one of the four eIF2α kinases (GCN2, PKR, PERK, and HRI), respectively, to reduce TC concentration[147–150]. A reduction in the concentration of eIF2α-containing TCs reduces the rate of ribosome loading. Viral infection also leads to reduced ribosome loading via interferon-induced proteins with tetratricopeptide repeats (IFITs)[151]. Cellular stress also reduces ribosome loading via inhibition of mTOR and sequestration of eIF4E by hypophosphorylated 4EBP[152]. Secondly, translated repressive uORFs are enriched in transcripts buffered against reduced ribosome loading[153–156]. Therefore, we were particularly interested in varying this ribosome loading rate to investigate if and how uORFs provide this buffering across various proposed models. For each of the five surveyed models of uORF regulation (Fig 2.1), we investigated what uORF parameter combinations, if any, allow buffering against reduced ribosome loading rates.

We use the term 'buffer' to describe the observation of main ORF protein output decreasing less than expected, or even increasing, with reduced ribosome loading in comparison to the constitutive repression model (Fig 2.1A). The constitutive repression model (Fig 2.1A) has no buffering (Fig 2.5A) since its repression is independent of the ribosome loading rate. Buffering requires an interaction between ribosome loading and the degree of translational repression. We use buffering as an overarching term that encompasses both resistance and preferred translation. Resistance refers to a decrease in main ORF protein output to a lower extent than in the constitutive repression model when ribosome loading is reduced. Preferred translation refers to increased main ORF protein output when ribosome loading is reduced.

The 80S-hit dissociation model (Fig 2.1B) displays buffering (Fig 2.5B, left panel, yellow-green line) in agreement with previous work[75]. This behavior arises because the number of 5′ elongating ribosomes that collide with scanning ribosomes correlates with the ribosome loading rate. However, buffering requires strong uORF initiation, minimal re-initiation, and a long uORF (Fig 2.5B, left panel, yellow-green line, Fig 2.6 panel A) as observed previously[75]. These observations can be rationalized as follows. Strong uORF initiation generates sufficient elongating ribosomes that hit and knock off 3′ scanning ribosomes. Minimal re-initiation prevents the many uORF-translating ribosomes from also translating the main ORF. Longer

uORFs offer more time for elongating ribosomes to catch up, hit, and knock off 3′ scanning ribosomes. However, most eukaryotic uORFs only weakly initiate translation and are short[42,87,104,110,112,116–118]. *UL4* uORF2 is 22 codons long, and we estimate re-initiation to be frequent (Table 1). Accordingly, buffering is no longer observed (Fig 2.6 panel B) in the 80S-hit dissociation model when parameters (Table 1) derived from control *UL4* variants (Fig 2.2C) are used.

The queuing-mediated enhanced repression model[97] (Fig 2.1C) displays buffering behavior (Fig 2.5C, left panel, purple line) since the number of scanning ribosomes that initiate translation at the uORF is dependent on the rate of ribosome loading. In this model, reduced ribosome loading decreases the average queue length of ribosomes behind the elongation stall and, thus, also the fraction of ribosomes that initiate at the uORF2 start codon (Fig 2.6 panel C, left). Unlike the 80S-hit dissociation model (Fig 2.1B), weakly initiating uORFs, such as *UL4* uORF2, still confer buffering in the queuing-mediated enhanced repression model (Fig 2.5C, left panel, purple line).

In the queuing-mediated enhanced repression model, enhanced uORF initiation and, therefore, buffering are sensitive to the distance between the uORF start codon and the elongating ribosome stall (d_{stall}) (Fig 2.6 panel C and Fig 2.5C, left vs. right panels). This sensitivity arises because d_{stall} determines if the P-site of a queued scanning ribosome is correctly positioned at the uORF start codon to productively increase uORF initiation (Fig 2.6 panel C, left). In the idealized case of homogeneously sized ribosomes (30 nt footprints[157,158]) and strict 5′-3′ scanning, d_{stall} must be an integer multiple of 30 nt for buffering to occur. This strong dependence of buffering on d_{stall} is relaxed when backward scanning[130,159–161] occurs with a high rate (Fig 2.6 panel D, middle). To simplify our modeling interpretations, we considered *UL4* uORF2, which is 22 codons long, to be 21 codons so that a queue behind the terminating ribosome stall positions a scanning ribosome's P-site exactly at the start codon.

The collision-mediated 40S dissociation model (Fig 2.1D) displays buffering (Fig 2.5D, right panel, purple line) because the number of scanning ribosomes that collide with 3′ stalled ribosomes depends on the rate of ribosome loading. Buffering in this model requires the collision-induced 40S dissociation rate to be somewhat fast (Fig 2.5D, right vs. left panels, Fig 2.6 panel E, teal and yellow-green lines). If this rate

is too low (for example, 0 in Fig 2.5D, left panel), this model reduces to the queuing-mediated enhanced repression model (Fig 2.1C). With an appreciable dissociation rate, the collision-mediated 40S dissociation model is not sensitive to the distance between the stall and the start codon (Fig 2.6 panel F, purple lines). As in the queuing model (Fig 2.1C), weakly initiating uORFs, such as *UL4* uORF2, can still confer buffering (Fig 2.5D, right panel, purple line) in the collision-mediated 40S dissociation model (Fig 2.1D). This effect arises because, unlike in the 80S-hit dissociation model (Fig 2.1B), the elongation stall is now rate-limiting for main ORF translation. Therefore, an elongation stall in the collision-mediated 40S dissociation model or the queuing-mediated enhanced repression model with permissive d_{stall} spacing imparts buffering.

In the regulated re-initiation model (Fig 2.1E), buffering is observed (Fig 2.5E, right panel, yellow-green line) because termination at the first uORF stop codon is followed by re-initiation at either the second downstream uORF or the main ORF depending on the ternary complex concentration. Buffering in the regulated re-initiation model (Fig 2.1E) depends on the initiation efficiency and continued scanning fraction (fraction of terminating but non-recycling ribosomes) of the two uORFs (Fig 2.5E). Continued scanning following termination at the first uORF must be frequent while continued scanning following termination at the second downstream uORF must be rare. Higher ternary complex concentrations bias towards initiation at the second downstream uORF (Fig 2.7 panel A). Lower ternary complex concentrations bias towards main ORF initiation; therefore main ORF translation can increase with decreased ribosome loading.

As such, our computational results provide the first systematic dissection of different mechanisms of uORF-mediated regulation (Fig 2.1) and enable their comparison with experimental measurements below.

UL4 uORF2 buffers against reductions in main ORF protein output from reduced ribosome loading in an elongating ribosome stall-dependent manner

We next tested whether the computational predictions of uORF-mediated buffering (Fig 2.5) can be observed experimentally with *UL4* uORF2. To this end, we experimentally varied the rate of ribosome loading and measured effects on main ORF protein output using our reporter system (Fig 2.2B). Since the no-stall uORF2 variants have similar protein expression to the no-start uORF2 variants (Fig 2.2C), luciferase signal from the no-stall uORF2 variants provides a readout of the ribosome loading rate. If buffering were

absent, then we would expect NLuc translation to be reduced equally between the no-stall and wild-type variants when ribosome loading is reduced.

We used three strategies to vary the rate of ribosome loading. We first used stem-loops near the 5′ cap to reduce the rate of 43S-cap binding without affecting mRNA stability (Fig 2.8A)[162,163]. We varied the degree to which ribosome loading is reduced by altering the GC content of the stem-loops; generally, higher GC content stem-loops are more stable and therefore cause greater reductions in ribosome loading. We observe that NLuc signal decreases less with reduced ribosome loading for the wild-type *UL4* reporter in comparison to the no-stall *UL4* variant (Fig 2.8A, left panel, yellow vs. gray circles). Therefore, NLuc protein output is resistant to stem-loop-mediated reduction in ribosome loading. When the wild-type data are normalized by the no-stall data, NLuc translation negatively correlates with ribosome loading, indicative of buffering against reduced ribosome loading by wild-type uORF2 (Fig 2.8A, right panel).

We also reduced ribosome loading with the drug thapsigargin, which induces the integrated stress response (ISR) by triggering ER stress (Fig 2.8B)[164]. We added a PEST tag[165] to increase the turnover of the NLuc protein to more accurately measure changes in main ORF translation during drug treatment. NLuc protein output from the wild-type *UL4* reporter decreases less in comparison to the no-stall control upon thapsigargin treatment (Fig 2.8B, left panel, yellow vs. gray circles), indicative of resistance. Again, when the wild-type data are normalized by the no-stall data, we observe that NLuc translation negatively correlates with ribosome loading, indicative of buffering against reduced ribosome loading by wild-type uORF2 (Fig 2.8B, right panel).

Finally, we added a short, synthetic uORF, 5′ to the *UL4* uORF2, to siphon scanning ribosomes away from uORF2 (Fig 2.8C). We varied the degree of ribosome siphoning by varying the Kozak context of the synthetic uORF, which in turn determines the rate of ribosome loading onto the uORF2-NLuc portion of the mRNA. Here, we observe that more NLuc is produced from the wild-type *UL4* reporter as scanning ribosomes are increasingly siphoned off by improving the Kozak context of the synthetic uORF (Fig 2.8C, left panel, yellow circles). While resistance is observed with the other strategies of reduced ribosome loading (Fig 2.8A-B, left panels), preferred translation is observed here (Fig 2.8C, left panel), perhaps

because ribosome loading is reduced at the scanning step instead of at the cap-binding step. Similar to the other two strategies, when the wild-type data are normalized by the no-stall data, NLuc translation negatively correlates with ribosome loading, indicative of buffering against reduced ribosome loading by wild-type uORF2 (Fig 2.8C, right panel).

Distance between the start codon and the stall does not systematically regulate uORF repressiveness or buffering

Given our experimental data demonstrating uORF2-mediated buffering of *UL4* reporters (Fig 2.8), we narrowed our focus from the five surveyed models (Fig 2.1) to the two (Fig 2.1C-D) most relevant for *UL4* uORF2: the queuing-mediated enhanced repression (Fig 2.1C) and collision-mediated 40S dissociation (Fig 2.1D) models. These two models are computationally predicted to confer buffering in an elongating ribosome stall-dependent manner without needing multiple uORFs (Fig 2.5C-D). To differentiate between these models, we turned to our computational modeling prediction that, only in the queuing-mediated enhanced repression model (Fig 2.1C), main ORF protein output is sensitive to the distance between the uORF start codon and the elongating ribosome stall (Fig 2.5C). Our computational modeling of the queuing-mediated enhanced repression model (Fig 2.10A, yellow-green line) predicts two broadly spaced clusters of main ORF protein output. Protein output from the main ORF is repressed when the start codon-stall distance is an integer multiple of the ribosome size. Protein output from the main ORF is high when the start codon-stall distance is not an integer multiple of the ribosome size. In contrast, the collision-mediated 40S dissociation model (Fig 2.1D) predicts a much lower effect of d_{stall} on uORF repressiveness (Fig 2.10A, left panel, purple line). The residual effect of d_{stall} on uORF repressiveness (Fig 2.10A, left panel, purple line) in the collision-mediated 40S dissociation model (Fig 2.1D) arises because the dissociation rate is low enough to allow rare queuing.

Backward scanning is predicted to diminish the periodicity in main ORF translation with varying d_{stall} lengths in both models (Fig 2.10A). However, backward scanning occurring as fast as forward scanning (~ 5 nucleotides/s) is required to abolish the periodicity in the queuing model (Fig 2.10A, right panel, yellow-green line). While there are estimates of how far ribosomes can backward scan[159–161], we are not aware

of any backward scanning rate estimates. It is unlikely that the rate of backward scanning approaches the rate of forward scanning (5 nucleotides/s here) given the 5′-3′ directionality of scanning. Slower backward scanning (~ 3.75 nucleotides/s) is sufficient to abolish periodicity in the collision-mediated 40S dissociation model (Fig 2.10A, middle panel, purple line). This effect is not surprising given that the presence of periodicity in the latter model arises from rare queuing behavior. Therefore, our computational predictions of greater periodicity in main ORF translation across varied d_{stall} in the queuing model hold even with backward scanning.

We then experimentally varied the distance between the start codon and the stall of the *UL4* uORF by adding codons to the 5′ end of uORF2. With *EYFP* donor sequences, we observe less than 2-fold changes in translational regulation (Fig 2.10B, top 7 rows) with no systematic trend with variations in uORF2 length, which is inconsistent with computational modeling predictions of the queuing-mediated enhanced repression model (Fig 2.10A, left panel). We observe similar results with a different donor sequence (Fig 2.9, top 7 rows), confirming the generality of the observed repression with changes in uORF2 length. With both donor sequences, the longest uORF mutants are less repressive, but this effect may be due to decreased elongating ribosome stall strength. In these cases, the longer nascent peptides can extend out of the exit tunnel and can be bound by additional factors[100,101] or cotranslationally fold to exert a pulling force[166] to relieve the stall. Thus, in summary, varying the length of the *UL4* uORF2 stall does not match computational predictions for the sensitivity of main ORF repression to d_{stall} in the queuing-mediated enhanced repression model (Fig 2.1C) and better supports the collision-mediated 40S dissociation model (Fig 2.1D).

In the queuing-mediated enhanced repression model (Fig 2.1C), buffering is uniquely predicted to be sensitive to the distance between the uORF start codon and the elongating ribosome stall (Fig 2.5C). We, therefore, asked whether or not buffering would still be experimentally observed with a disruption in this distance. Using our synthetic uORF method of reducing ribosome loading (Fig 2.10C), we observe that a 6 nt longer d_{stall} uORF still buffers against reduced ribosome loading (Fig 2.10C, top two rows compared to bottom two rows). Since backward scanning of 15-17 nt has been observed[159–161], one would expect that buffering would still be predicted in the queuing model even with an increase in d_{stall} of 6 nt. However,

our computational modeling predicts that even very fast backward scanning does not restore buffering when d_{stall} is disrupted by 6 nt (Fig 2.6 panel D, right). Thus, our experimental data does not match computational predictions of buffering sensitivity to d_{stall} in the queuing-mediated enhanced repression model (Fig 2.1C), but is consistent with the collision-mediated 40S dissociation model (Fig 2.1D).

Several human uORFs have repressive terminal diproline motifs

Given that the elongating ribosome stall in the human cytomegaloviral *UL4* uORF2 is dependent on a terminal diproline motif, we asked whether there are other human uORFs similarly ending in diproline motifs that are also repressive. We searched for such uORFs in three databases: a comprehensive database of ORFs in induced pluripotent stem cells and human foreskin fibroblasts with 1,517 uORFs[114], a database integrated from *de novo* transcriptome assembly and ribosome profiling with 3,577 uORFs[167], and a database of proteins less than 100 residues in size derived from literature mining, ribosome profiling, and mass spectrometry with 1,080 uORFs[168]. We identified several human transcripts with terminal diproline-containing uORFs: *C1orf43*, *C15orf59*, *TOR1AIP1*, *PPP1R37*, and *ABCB9*. We replaced *UL4* uORF2 in our reporter (Fig 2.2B) with these human uORFs. We mutated the terminal proline codon to an alanine codon as well as the start codon of these human uORFs and measured the effects of these mutations on NLuc protein output relative to the wild-type uORFs. While many of the tested uORFs are repressive (Fig 2.11, yellow vs. blue), unlike the human cytomegaloviral uORF2, these human uORFs still repress NLuc protein output without their terminal diproline motif (Fig 2.11, gray vs. blue), indicating additional contributions to translational repression from other residues in the nascent peptide and due to siphoning of scanning ribosomes at the start codon.

III. Discussion

In this study, we use a combination of computational modeling and experimental reporter measurements to dissect the kinetics of uORF-mediated translational regulation of the *UL4* mRNA of human cytomegalovirus. We find that the elongating ribosome stall in *UL4* uORF2 buffers against reductions in main ORF protein output due to reduced ribosome loading (Fig 2.5). Using an experimentally-integrated modeling approach, we differentiate between models of regulation that can explain this observation. Our computational framework allows easy specification and efficient simulation of several previously proposed kinetic models of uORF regulation (Fig 2.1). While uORFs are enriched in stress-resistant transcripts, not all uORFs provide buffering[153]. We can predict which models of uORF regulation allow buffering and which parameters are key for buffering in each model (Fig 2.5). To our knowledge, our work is the first systematic investigation of what uORF metrics impart buffering in each kinetic model of uORF regulation.

uORFs are generally thought to simply siphon away scanning ribosomes from main ORFs, but this simple behavior in the constitutive repression model (Fig 2.1A) is not predicted to provide buffering (Fig 2.5C)[153–156]. Instead, we find that 5′ UTRs containing one (or some combination) of the following enable buffering of main ORF translation: scanning ribosome dissociation due to 80S hits from the 5′ end (Fig 2.1B), a single uORF with an elongating ribosome stall (Fig 2.1C-D), or multiple uORFs acting through the regulated re-initiation model (Fig 2.1E).

Long, well-initiating uORFs that do not re-initiate well allow buffering (Fig 2.5B, left panel, yellow-green line, Fig 2.6 panel A, yellow-green line) in the 80S-hit model (Fig 2.1B), but these requirements are at odds with the typically short and poorly initiating nature of known uORFs[42,87,110,112,116–118]. Indeed, when we use parameters specific to *UL4* uORF2 for the 80S-hit model (Table 1), namely that uORF2 initiates poorly, re-initiates well, and is not very long, buffering is no longer predicted (Fig 2.6 panel B).

Computational predictions from the regulated re-initiation model (Fig 2.1E) agree (Fig 2.5E and Fig 2.7 panel A) with previous work[125,126] showing that buffering requires: 1) two well-translated uORFs and 2) frequent and rare continued scanning after termination at uORFs 1 and 2, respectively. Since 30% of human transcripts contain multiple uORFs, some of these might enable buffering by the regulated re-

initiation model. However, about 25% of human transcripts only have one uORF[111] and cannot provide buffering under this model.

We narrowed our focus to the two models (Fig 2.1C-D) that are most pertinent to *UL4* uORF2. Both the queuing-mediated enhanced repression (Fig 2.1C) and collision-mediated 40S dissociation (Fig 2.1D) models are predicted to allow buffering (Fig 2.5C-D) with weakly initiating uORFs and elongating ribosome stalls. Both of these models require only a single uORF for buffering (Fig 2.5C-D). Computational modeling not only predicts this buffering behavior but also allows us to differentiate between these two models. We predict that the queuing-mediated enhanced repression model (Fig 2.1C) is uniquely sensitive to the distance between the uORF start codon and elongating ribosome stall (Fig 2.10A, yellow-green line, Fig 2.5C, purple lines, Fig 2.6 panel F, purple lines). We experimentally vary this distance and do not find any systematic changes in either main ORF protein output (Fig 2.10B) or buffering (Fig 2.10C). Based on our results, we propose that scanning ribosomes dissociate rather than queue when encountering a 3′ stalled elongating ribosome on uORF2 of *UL4* mRNA.

Scanning ribosomes have been predicted to dissociate upon encountering stable secondary structures on the mRNA[169]. Collisions between scanning ribosomes and their subsequent dissociation have also been proposed in a model of initiation quality control[170]. This dissociation could serve to maintain the free pool of 40S ribosomal subunits while still allowing regulation of main ORF translation. Collisions between scanning and elongating ribosomes and subsequent quality control are not well understood; what we describe as scanning ribosome dissociation here may be rescue by a quality control pathway.

Although our data from *UL4* uORF2 does not support the queuing-mediated enhanced repression model (Fig 2.1C)[97], this model might describe translation kinetics on other mRNAs. Translation from near-cognate start codons is resistant to cycloheximide, perhaps due to queuing-mediated enhanced initiation, but sensitive to reductions in ribosome loading[171]. Loss of eIF5A, which helps paused ribosomes continue elongation, increases 5′ UTR translation on human mRNAs with pause sites proximal to the start codon, perhaps also through queuing-mediated enhanced initiation[172]. There is also evidence of queuing-enhanced uORF initiation in the 23 nt long *Neurospora crassa* arginine attenuator peptide[173] as well as in transcripts with

secondary structure near and 3′ to start codons[174]. Additional sequence elements in the mRNA might determine whether scanning ribosome collisions result in queuing or dissociation. Small subunit profiling data[175] from human uORFs that have conserved amino acid-dependent elongating ribosome stalls do not show evidence of scanning ribosome queues (Fig 2.12 panel A), consistent with the collision-mediated 40S-dissociation model. However, subtle queues might not be observed given low read counts arising from insufficient capture of small ribosomal subunits in these experiments.

In our modeling, we assume homogenous footprint lengths of 30 nt for both scanning and elongating ribosomes. Even though heterogeneously sized footprints have been observed for small ribosomal subunits[175–177] and elongating ribosomes[178,179], our modeling of homogenous footprint length is appropriate for the following reasons. Firstly, concerning the small ribosomal subunit footprints, crosslinking of associated eIFs is thought to be the main source of length heterogeneity[175,176], and homogenous 30 nt footprints are observed in the absence of crosslinking[176]. Secondly, in the context of the strong, minutes-long *UL4* uORF2 elongating ribosome stall[180], collided ribosomes, if they do not dissociate, will wait for long periods in a queue relative to normal scanning or elongating ribosomes, during which associated eIFs likely dissociate[176]. Thirdly, a sizable fraction of mRNAs exhibit cap-tethered translation in which eIFs must dissociate from ribosomes before new cap-binding events, and therefore collisions, can occur[176]. Elongating ribosome footprint heterogeneity is much less drastic than that observed for scanning ribosomes and likely arises from different conformational states such as empty or occupied A sites[178,179]. While different elongating ribosome footprints arise from differences in mRNA accessibility to nucleases, it is unclear whether the distance between two collided ribosomes changes across different ribosome conformations.

In addition to the *UL4* viral uORF studied here, several human uORFs are known to contain an elongating ribosome stall[52,96,97,136]. Apart from terminal diprolines, other motifs such as Arg-X-Lys at E-P-A sites[181] or specific dipeptides such as Gly-Ile, Asp-Ile, Gly-Asp[182] can also cause elongation stalls. There are a variety of other mechanisms that may reduce the rate of elongation, such as mRNA stem-loops and G-quadruplexes[183], low tRNA availability[184,185], or interactions between the nascent peptide and the

ribosome[186,187]. uORFs are often short[112] and may therefore be better poised to stall ribosomes since the nascent peptides might not be accessible to co-translational factors that pull the nascent peptide out of the ribosome[188,189]. Thus, a key role for elongation ribosome stalls in uORFs might be to enable buffering. While few uORF stalls have been mechanistically characterized[110], other elongating ribosome stall-containing uORFs, such as the ones in *MTR*[136] and *AMD1*[190] mRNAs, might enable buffering; the elongating ribosome stall-containing uORFs in *AZIN1*[97], *PPP1R15A* (*GADD34*)[191], and *DDIT3* (*CHOP*)[96] have already been shown to enable buffering. Conversely, uORFs in several single uORF transcripts known to buffer against stress, such as *SLC35A4*, *C19orf48*, and *IFRD1*[153], might act through elongating ribosome stalls.

The computational models considered here can be readily extended to incorporate more complex mechanisms of translational control. For example, in our models, initiation proceeds via a cap-severed mechanism in which multiple scanning ribosomes can be present in the 5′ UTR at the same time. If we were to model cap-tethered initiation, strong uORF elongating ribosome stalls would eventually sever this connection, similar to how the cap-eIF-ribosome connection is severed during the usually longer translation of main ORFs[176,192,193]. It will also be interesting to consider the effect of cellular stress-reduced elongation rates[194] and increased re-initiation[195], both of which might regulate uORF-mediated buffering, as well as elongating ribosome dissociation through known quality control pathways[128,170,196–201]. Translation heterogeneity among isogenic mRNAs has been observed in several single-molecule translation studies[123,144–146,202]. This heterogeneity may arise from variability in intrasite RNA modifications[203], RNA binding protein occupancy, or RNA localization. We do not capture these sources of heterogeneity in our modeling since the observables in our simulations are averaged over long simulated time scales and used to predict only bulk experimental measurements. However, the models studied here can readily be extended through compartmentalized and state-dependent reaction rates[142] to account for the different sources of heterogeneity observed in single-molecule studies.

IV. Materials and Methods

Plasmid construction

The parent cloning vector was created as follows. A commercial vector (Promega pGL3) with ampicillin resistance was used to clone NLuc and FLuc. NLuc expression is driven by a CMV promoter. FLuc expression is driven in the opposite direction within the plasmid and serves as an internal transfection control. The human cytomegaloviral *UL4* 5′ UTR was PCR amplified from HCMV genomic DNA. To create mutant 5′ UTR versions of the parent pGL3-FLuc-NLuc vector, the vector was digested with KpnI/EcoRI unless otherwise noted. 1 or 2 PCR-amplified fragments with 20-30 bp homology arms were then cloned using isothermal assembly[204]. The stem-loop[163] 5′ UTR mutants were cloned as follows. The stem-loops were ordered as oligonucleotides with overhangs for ligation into ClaI and NotI sites. The oligonucleotides were annealed and used in PCR reactions to add CMV homology arms. An AAVS1 parent vector was digested with ClaI and NotI. These stem-loops were then inserted into the ClaI/NotI restriction digested parent vector by isothermal assembly[204]. The stem-loops were then PCR amplified off of this plasmid and inserted into the pGL3-Fluc-*UL4*-5′-UTR-NLuc parent vector described above. The several tested human uORFs were PCR amplified from human genomic DNA and inserted into a PstI/EcoRI digested parent. The inserted sequences were confirmed by Sanger sequencing. Kozak context and stall codon mutations were introduced in the PCR primers used for amplifying inserts before isothermal assembly. Standard molecular biology procedures were used for all other plasmid cloning steps[205]. Key plasmid maps are available at https://github.com/rasilab/bottorff_2022 as SnapGene `.dna` files. Plasmids will be sent upon request.

Cell culture

HEK293T cells were cultured in Dulbecco's modified Eagle medium (DMEM 1X, with 4.5 g/L D-glucose, + L-glutamine, - sodium pyruvate, Gibco 11965-092) and passaged using 0.25% trypsin in EDTA (Gibco 25200-056).

Dual-luciferase reporter assay

Plasmid constructs were PEI or Lipofectamine 3000 (Invitrogen, L3000-008) transiently transfected into HEK293T cells for 12-16h in 96 well plates. After the 12-16h transfection, the ~110 μL media was removed and replaced with 20 μL media per well. The Promega dual-luciferase kit was used. Cells were lysed with 20 μL ONE-Glo EX Luciferase Assay Reagent per well for three minutes to measure firefly (*Photinus pyralis*) luciferase activity. Then, 20 μL NanoDLR Stop & Glo Reagent was added per well for 10 minutes to quench the FLuc signal and provide the furimazine substrate needed to measure NLuc luciferase activity. FLuc activity serves as an internal control for transfection efficiency, and NLuc activity provides a readout of 5′ UTR regulation of NLuc translation.

Kinetic modeling

We specify our kinetic models using the PySB interface[141] to the BioNetGen modeling language[142] (Fig 2.4). The Python script is parsed by BioNetGen into a `.bngl` file and converted into an `xml` file for use as input to the agent-based stochastic simulator NFsim[143].

Molecules

Our kinetic models of eukaryotic translational control describe the interactions between 3 molecule types: mRNA, ribosome (composed of separate large and small subunits), and ternary complex. Here, we describe these molecules' components, states, and binding sites (Fig 2.4A). mRNA molecules have the following components: 5′ end and codon sites (c_i). The mRNA 5′ end can either be free of (*clear*) or occupied with a ribosome (*blocked*). The mRNA 5′ end must be clear for a 43S to bind, which leaves the 5′ end blocked until the ribosome scans (or elongates) sufficiently 3′ downstream. The mRNA codon sites serve as binding sites for the ribosome A site. Small ribosomal subunits have the following components: inter-subunit binding interface (*isbi*), ternary complex contact (*tc*), 5′ side (*t* for trailing), 3′ side (*l* for leading), and A site (*a*). The inter-subunit binding interface site allows interactions between small and large ribosomal subunits; large ribosomal subunits also have the inter-subunit binding interface (*isbi*) components. The 5′ and 3′ side sites serve as binding sites for other ribosomes during collisions (5′ or 3′ side). The A site serves as a binding site for the mRNA. Both scanning and elongating ribosomes have mRNA

footprints of 10 codons in our simulations based on mammalian ribosome profiling data[157,158]. Ternary complex molecules have a single component *ssusite* that serves as a binding site for the small ribosomal subunit.

Reactions

We describe here each type of kinetic reaction in our models of eukaryotic translational control (Fig 2.4B). We use a syntax similar to that of BioNetGen[142] to illustrate the kinetic reactions. We scale TC and ribosome subunit numbers (100 each) to the single mRNA present in the simulation. Simulation of a single mRNA over several rounds of translation is sufficient to infer steady-state translation dynamics.

Initiation: PIC (43S) formation

Small ribosomal subunits must bind TCs to form pre-initiation complexes (PICs, 43Ss) before loading onto mRNAs. We assume that PIC formation is irreversible. PIC formation is not rate-limiting in our simulations; we set the rate of 43S-cap binding ($k_{cap\ bind}$) to be rate-limiting and to a total rate (independent of [43S]) to match the overall initiation rate to that of cellular estimates. Therefore, we arbitrarily set the second-order PIC formation rate (40S-TC binding rate, $k_{ssu\ tc\ bind}$) to 0.01 * TC^{-1} * SSU^{-1} such that 100 40S-TC binding events occur per second, which is much higher than the 43S-cap binding rate.

Initiation: PIC (43S) loading onto mRNA

We model ribosome footprints at 30 nt following mammalian ribosome profiling data[157,158]. Therefore, PIC loading can occur when the 5′ most 30 nucleotides (nt) of the mRNA are not bound to any ribosome. The rate at which PICs load onto the 5′ end of the mRNA, $k_{cap\ bind}$, is varied over a 100-fold range from the maximum ribosome loading rate, 0.125/s, based on single-molecule estimations in human cells[144]. PICs can load onto the mRNA when a ribosome footprint-sized region at the 5′ mRNA end is free of ribosomes. PIC loading results in the 5′ end being blocked until this ribosome scans or elongates past a ribosome footprint from the 5′ cap. We assume that PIC loading is irreversible.

Initiation: Scanning and start codon selection

The scanning rate is 5 nucleotides/s following estimates in a mammalian cell-free translation system[206]

and a previous computational study[75]. Small ribosomal subunit A sites must be positioned exactly over start codons to initiate translation. The uORF start codon is 25 nt from the 5′ cap. We vary the rate at which this start codon selection occurs at the uORF in our modeling. Start codon selection releases the TC bound to the small ribosomal subunit. We assume that TC is regenerated instantaneously. The start codon selection rate divided by the sum of this start codon selection rate, the scanning rate, and the backward scanning rate equals the baseline initiating fraction. This calculation of the baseline initiating fraction will underestimate the initiating fraction in the case of correctly positioned 3′ ribosome queues (as in the queuing-mediated enhanced repression model). We assume that start codon selection is irreversible.

Elongation

Elongation results in the ribosome A site moving from codon c_i to codon c_{i+1}. The rate of elongation is set to 5 codons/s following single-molecule method and ribosome profiling estimates in mammalian cells of 3-18 codons/s[144–146,157,207]. Elongation may only proceed if there is no occluding 3′ ribosome; in other words, elongation may only proceed from codon c_i to codon c_{i+1} if the next 3′ ribosome's A site is bound to a codon no more 5′ than c_{i+11}. The elongation rate at the stall within the uORF is set to 0.001/s[180].

Termination, continued scanning, and re-initiation

Termination results in the dissociation of the large ribosomal subunit, but the small ribosomal subunit may continue scanning and subsequently re-initiate if a new TC is acquired before the next start codon is encountered. The termination rate is set to 1/s given that ribosome density tends to be higher at stop codons than within ORFs[157,178]. The recycling rate of terminated small ribosomal subunits after uORF translation is varied to model the effect of varied continued scanning after uORFs on the regulation of main ORF translation. The scanning rate divided by the sum of the scanning rate and this recycling rate equals the continued scanning fraction.

Collisions and dissociations

A collision between two ribosomes requires them to be separated by exactly one ribosome footprint in distance on the mRNA and results in binding between the 5′ side of the leading (3′ most) ribosome and the 3′ side of the trailing (5′ most) ribosome. Abortive (premature) termination of ribosomes results in their

dissociation from the mRNA and any collided ribosomes they are bound to. Different models have different non-zero dissociation rates. For instance in the 80S-hit model, the following rates are equal and non-zero: $k_{scan\ term\ 5\ hit\ 80s}$, $k_{scan\ term\ both\ hit\ 80s\ 80s}$, $k_{scan\ term\ both\ hit\ 80s\ 40s}$. These rates relate to the dissociation of scanning ribosomes upon collisions with a 5′ elongating ribosome. Both hit refers to collisions with ribosomes on both sides. In the collision-mediated 40S dissociation model, the following rates are equal and non-zero: $k_{scan\ term\ 3\ hit\ 40s}$, $k_{scan\ term\ 3\ hit\ 80s}$, $k_{scan\ term\ both\ hit\ 40s\ 40s}$, $k_{scan\ term\ both\ hit\ 40s\ 80s}$, $k_{scan\ term\ both\ hit\ 80s\ 40s}$, $k_{scan\ term\ both\ hit\ 80s\ 80s}$. These rates relate to the dissociation of scanning ribosomes upon collisions with a 3′ scanning or elongating ribosome. The *in vivo* abortive termination rates of scanning ribosomes are not known. Small ribosomal subunits that make it to the 3′ end of the mRNA through leaky scanning of all (u)ORFs always dissociate.

Model calibration to reporter measurements

We derive the $k_{cap\ bind}$ rates by spline interpolation of computationally modeled protein output fit to experimental data (Fig 2.2C). We minimized the root mean square error between modeled protein output across variations in these parameters and the experimental data.

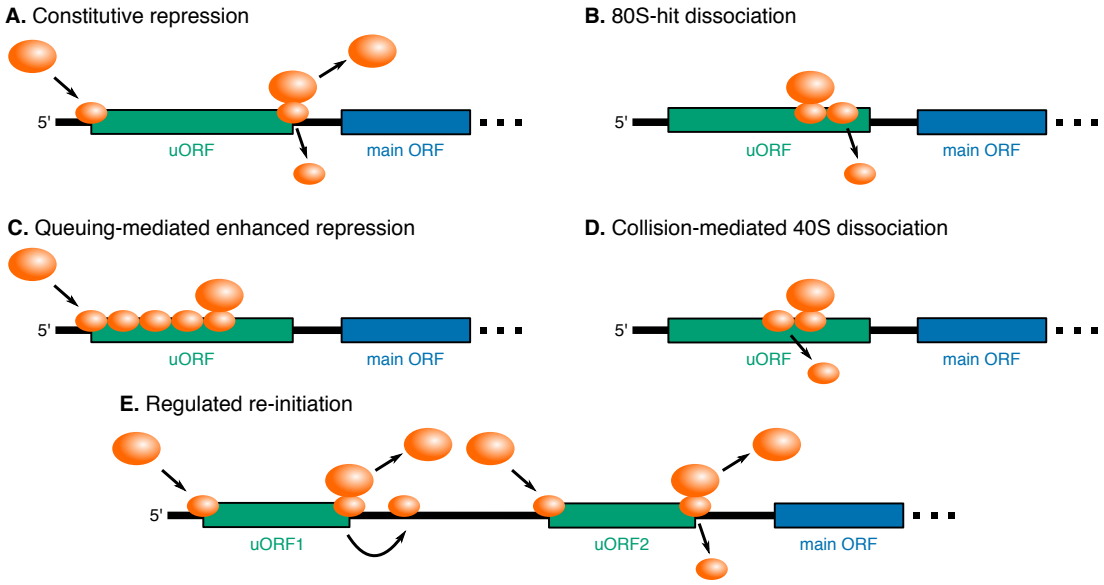
Human uORF search

We import uORF lists from several databases[114,167,168]. The SmProt database[168] includes 3162 uORFs from ribosome profiling data, which we filter down first to 1080 uORFs after filtering for aligned matches, available Kozak context, near-cognate start codons, and non-duplicates. Two of these uORFs end in diproline motifs, including *C1orf43*. Another database is a set of high-confidence ORFs derived from ribosome profiling of human-induced pluripotent stem cells (iPSCs) or foreskin fibroblast cells (HFFs) and was downloaded from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4720255/bin/NIHMS741295-supplement-3.csv[114]. This database includes 1517 high-confidence (ORF-RATER score > 0.8) uORFs from either iPSCs or HFFs, which we filter down to 3 that end in diproline motifs, including *ABCB9*, *C1orf43*, and *TOR1AIP1*. The third database derives from HEK293T, HeLa, and K562 cells using ribosome profiling and was downloaded from https://static-content.springer.com/esm/art%3A10.1038%2Fs41589-019-0425-0/MediaObjects/41589_2019_425_MOESM3_ESM.xlsx[167]. This database includes 3577 uORFs

which we filter down to 3 that end in diproline motifs and that are less than 60 codons in length for ease of cloning, including *ABCB9*,*C15orf59*, and *PPP1R37*.

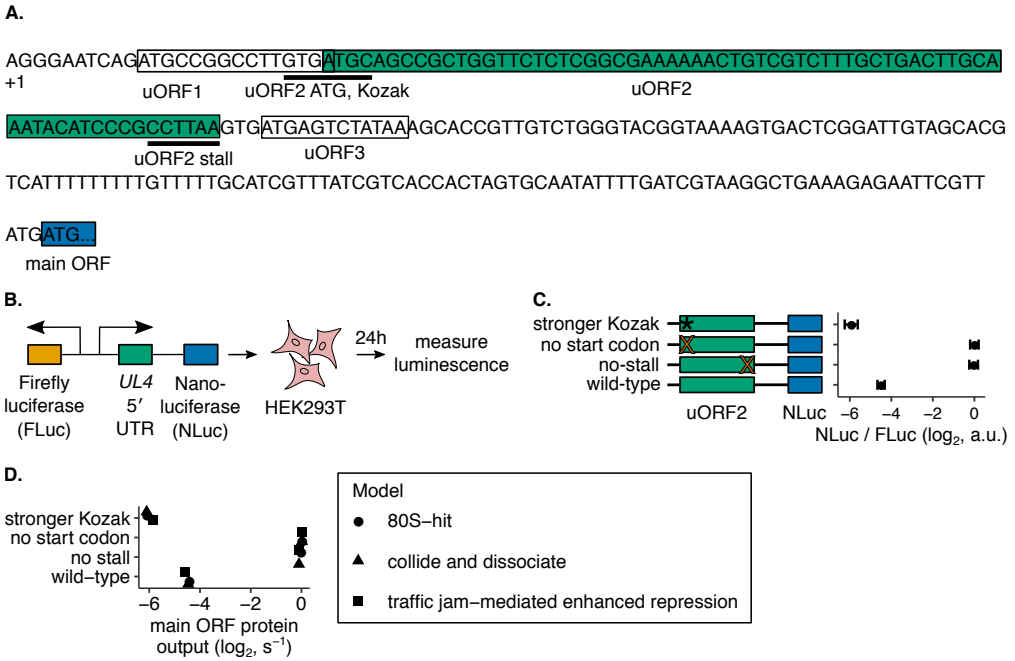
V. Figures

Figure 2.1



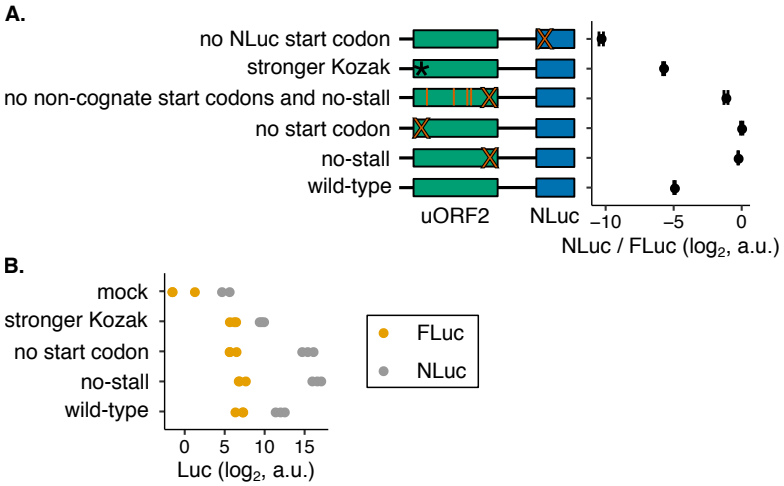
Models of uORF regulation considered in this study.
(A) *Constitutive repression.* The uORF constitutively siphons away a proportion of scanning ribosomes from the main ORF.
(B) *80S-hit dissociation.* Elongating ribosomes that collide with 3′ scanning ribosomes cause the leading scanning ribosome to dissociate from the mRNA.
(C) *Queuing-mediated enhanced repression.* Scanning or elongating ribosomes form a queue behind a 3′ stalled elongating ribosome. If the queue correctly positions a scanning ribosome at the uORF start codon, then the proportion of scanning ribosomes that initiate translation at the uORF start codon increases.
(D) *Collision-mediated 40S dissociation.* Scanning ribosomes that collide with a 3′ stalled ribosome dissociate from the mRNA.
(E) *Regulated re-initiation.* Ribosomes initiate translation at the first uORF start codon, and scanning continues after termination at the stop codon of the first uORF. Ribosomes re-initiate at the main ORF start codon or the second downstream uORF start codon when phosphorylated eIF2α levels are high or low, respectively. The schematic is depicted in a low phosphorylated eIF2α state.

Figure 2.2



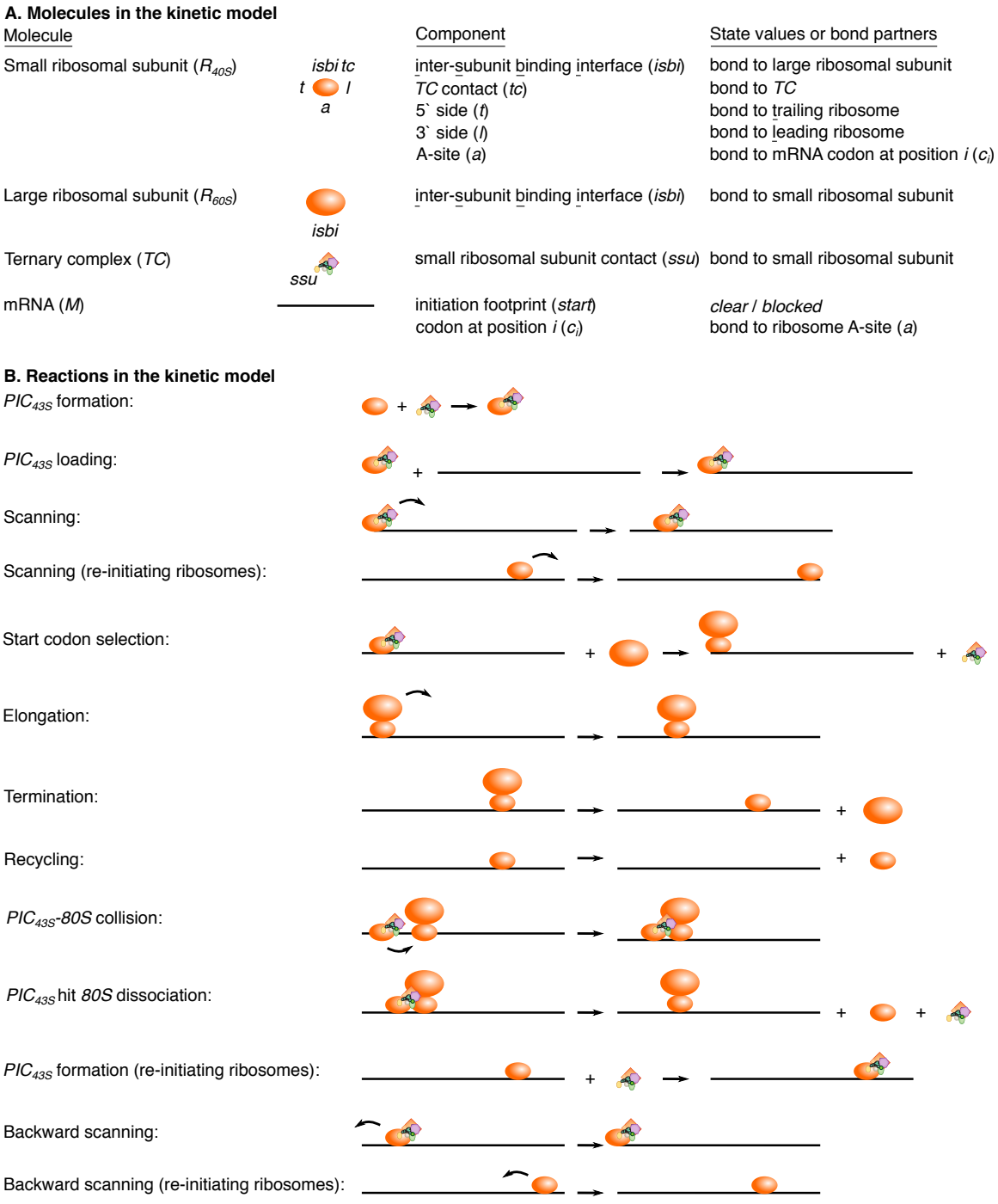
An experimental and computational platform for assessing uORF-mediated regulation of main ORF translation.
(A) The 236 nt 5′ UTR of *UL4* mRNA from human cytomegalovirus contains 3 uORFs. The terminal proline and stop codons of uORF2 at which the P- and A-sites of the stalled ribosome are positioned are highlighted as uORF2 stall.
(B) A dual-luciferase reporter system for measuring 5′ UTR repressiveness in HEK293T cells. FLuc signal serves as an internal control for transfection efficiency.
(C) The reporter system recapitulates the known elongating ribosome stall-dependent repression of protein expression by the *UL4* uORF2[95]. The indicated mutations improve the uORF2 Kozak context (ACCATGG instead of GTGATGC), remove the start codon (ACC instead of ATG), or remove the elongating ribosome stall by mutating the terminal proline codon to an alanine codon (GCT instead of CCT). Error bars show the standard error of mean NLuc / FLuc ratios over 3 biological replicates. Data are normalized to the no-uORF start codon control.
(D) Computationally predicted uORF regulation in the 80S-hit dissociation, queuing-mediated enhanced repression, and collision-mediated 40S dissociation models. Data are normalized to the no-uORF start codon control. The parameter combination that best recapitulated the control behavior in Fig 2.2C is displayed in Table 1. Error bars for simulated data are smaller than data markers.

Figure 2.3



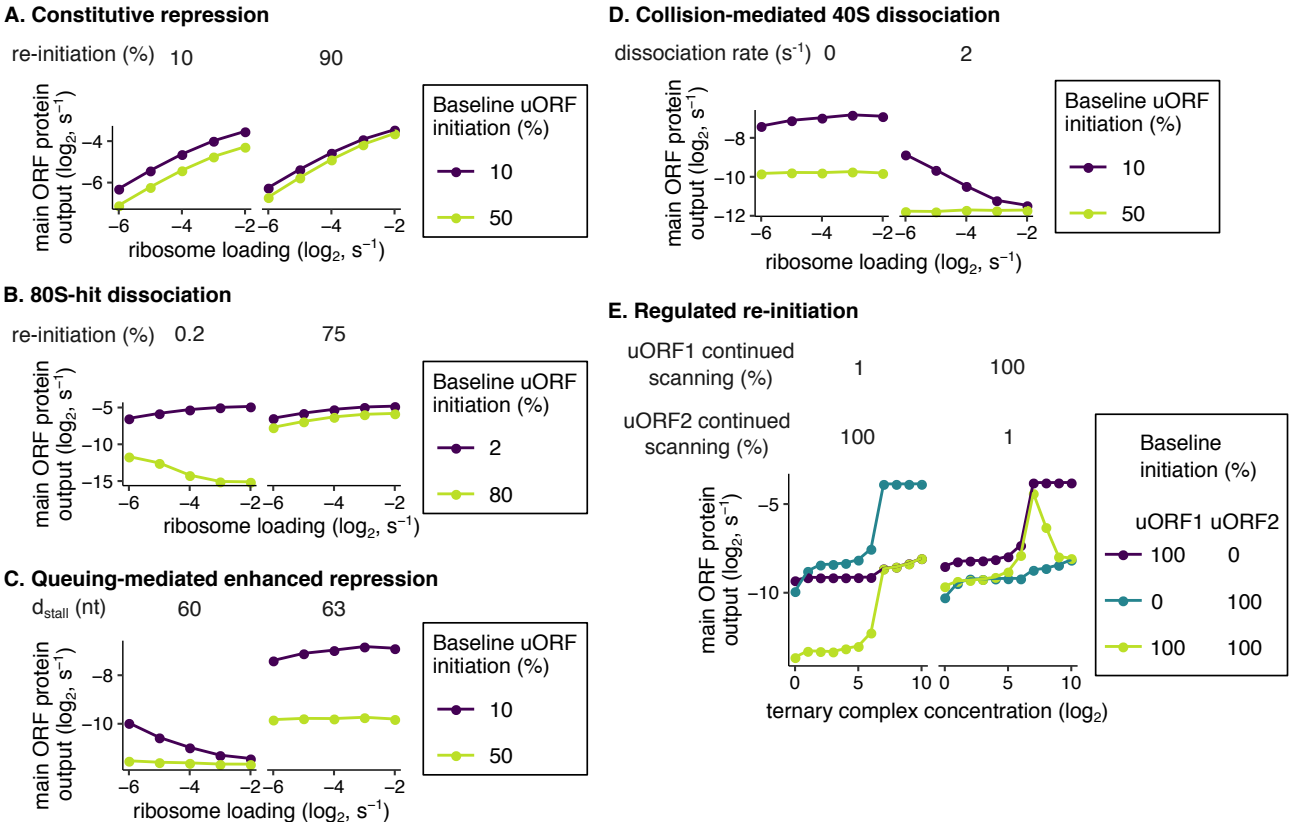
The dynamic range of the dual-luciferase reporter assay is sufficiently large
(A) The indicated mutations not present in Fig. 2.2C (the no NLUc start codon and no uORF2 near-cognate start codons) remove two adjacent NLuc ATG codons (ATGATG to ACCACC) and remove 4 uORF2 near-cognate start codons (CTG to CTA or TTG to TTA, red bars), respectively. The no NLuc start codon mutant abolishes NLuc signal, and the no uORF2 near-cognate start codons mutant does not greatly affect uORF2 repressiveness. Error bars show the standard error of mean NLuc / FLuc ratios over 3 biological replicates. Data are normalized to the no-uORF start codon control.
(B) Raw FLuc and NLuc signals for indicated mutations from Fig. 2.2C. Mock refers to transfection of no plasmid. Multiple data points indicate biological replicates.

Figure 2.4



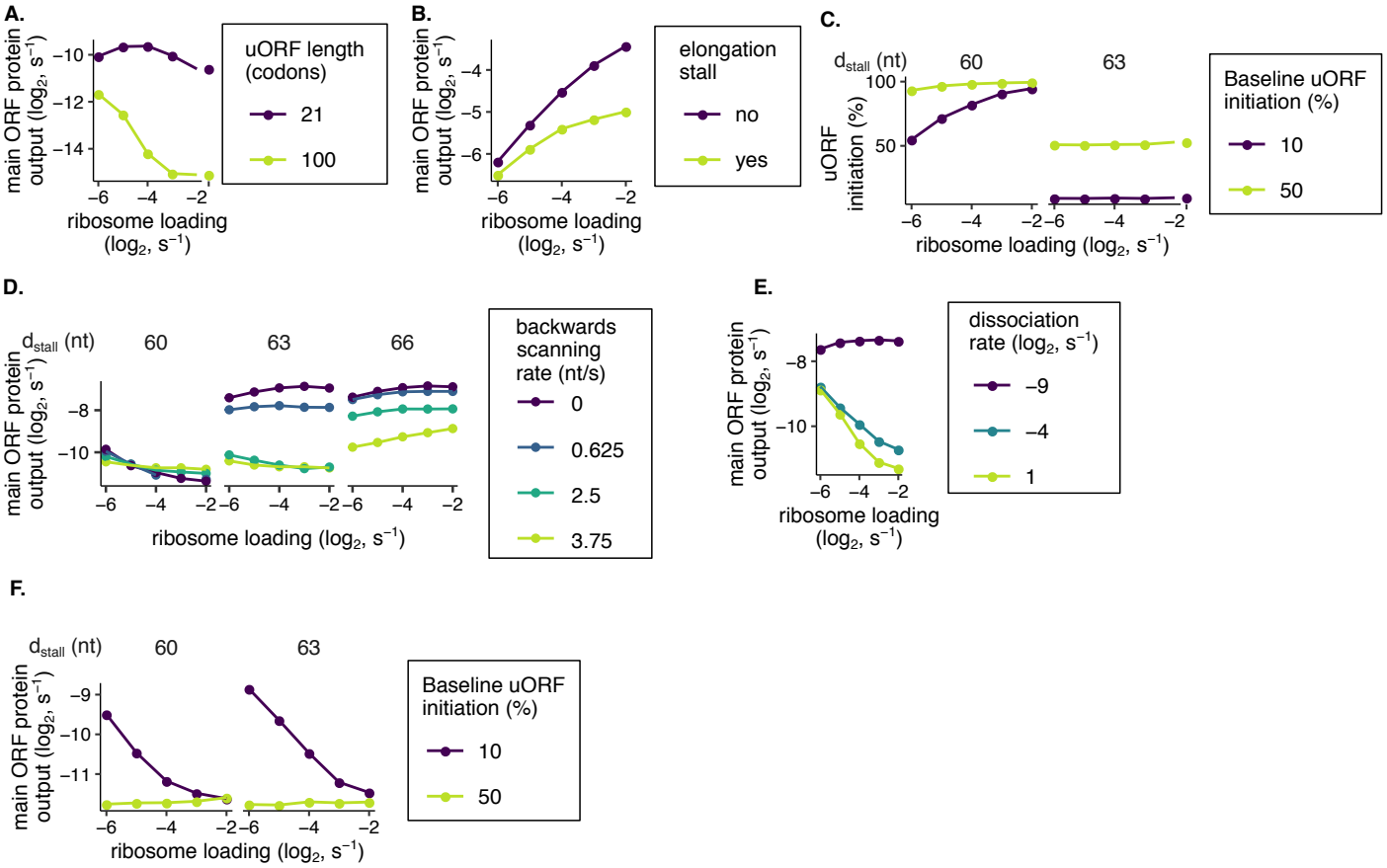
Modeling workflow.
(A) Molecules in the kinetic model. Molecules have components each of which has state values or binding sites for other molecules (called *bond* in BioNetGen). Components are abbreviated in parentheses by how they are referenced in the model specification. For example, the mRNA (*M*) initiation footprint (c_1 to c_n where n is equal to the ribosome footprint size in nt) can either be *clear* of ribosomes, and therefore free for a PIC_{43S} loading reaction to occur, or *blocked* by a ribosome, preventing this reaction.
(B) Visual representations of the reactions in the kinetic model. Re-initiation necessitates several additional reactions. PIC_{43S} formation (R_{40S} binding *TC*) can occur if the R_{40S} is bound to the mRNA; this *TC* re-binding is required for start codon selection competence. R_{40S} molecules can scan forward or backward. Some reactions in the kinetic model, such as different types of collision and dissociation reactions, are not depicted here.

Figure 2.5



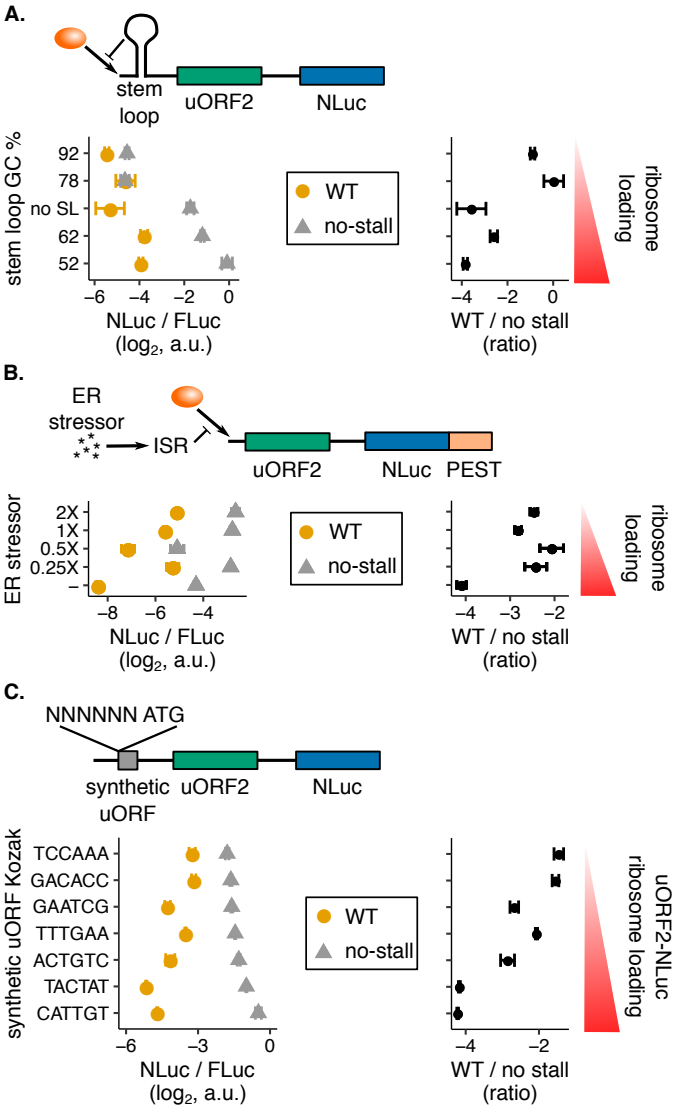
Kinetic modeling predicts translational buffering by uORFs.
Buffering refers to a smaller than expected decrease (small positive slope), or even increase (negative slope), in main ORF protein output with reduced ribosome loading.
(A) The constitutive repression model, without an elongating ribosome stall, has no buffering behavior. uORFs simply siphon away scanning ribosomes from the main ORF. **(B)** Buffering in the 80S-hit dissociation model depends on uORF initiation and re-initiation frequencies[75]. For buffering to occur in this model, uORFs must initiate well enough to have elongating ribosomes hit 3′ scanning ribosomes (yellow-green line). uORFs must also not continue scanning at high frequencies following termination (left panel); frequent continuation of scanning coupled with high uORF initiation allows many scanning ribosomes to make it to the main ORF. Buffering occurs better for longer uORFs that have more time for elongating ribosomes to hit 3′ scanning ribosomes (Fig 2.6 panel A, yellow-green line). Here, the uORF is 100 codons long. The dissociation rate is $200s^{-1}$, so 99% of scanning ribosomes hit by 5′ elongating ribosomes dissociate rather than continue scanning. The scanning rate is 2 nucleotides/s, and the elongation rate is 2 codons/s. There is no elongating ribosome stall in this model.
(C) Buffering in the queuing-mediated enhanced repression model depends on d_{stall}: the distance between the uORF start codon and elongating ribosome stall. In this model, uORF initiation can increase above baseline with increased ribosome loading when d_{stall} is an integer multiple of the ribosome footprint (30 nt, left panel). When this condition is met, buffering occurs. For d_{stall} values of 60 and 63 nt, the uORF length is 21 and 22 codons, respectively.
(D) Buffering in the collision-mediated 40S dissociation model depends on the dissociation rate. Here, d_{stall} is 63 nt; with a low dissociation rate, this model reduces to the queuing-mediated enhanced repression model.
(E) Buffering in the regulated re-initiation model depends on uORF initiation and continued scanning frequencies. For buffering to occur, several conditions must be met. At least 2 uORFs are required, both of which must be well-translated (yellow-green line). Continued scanning following termination at the first uORF must be frequent, and continued scanning following termination at the second downstream uORF must be rare (right panel). The second downstream uORF is 3 codons long. There is no elongating ribosome stall in this model.
uORFs are located 25 nt from the 5′ cap. 99% of scanning ribosomes that make it to the main ORF will initiate translation; 1% will leaky scan. Unless otherwise stated, parameters (Table 1) obtained from calibrating models to reporter measurements on wild-type or mutant uORF2 (Fig 2.2C) are used here. Ribosome loading is the $k_{cap\ bind}$ rate for non-regulated re-initiation models. We model changes in ribosome loading via changes in $k_{cap\ bind}$ as that rate is easier to match to *in vivo* estimates of ribosome loading. However, buffering in the regulated re-initiation model is dependent on an eIF2α phosphorylation mechanism; we instead vary the number of ternary complexes in this model. Error bars of simulated data are smaller than data markers.

Figure 2.6



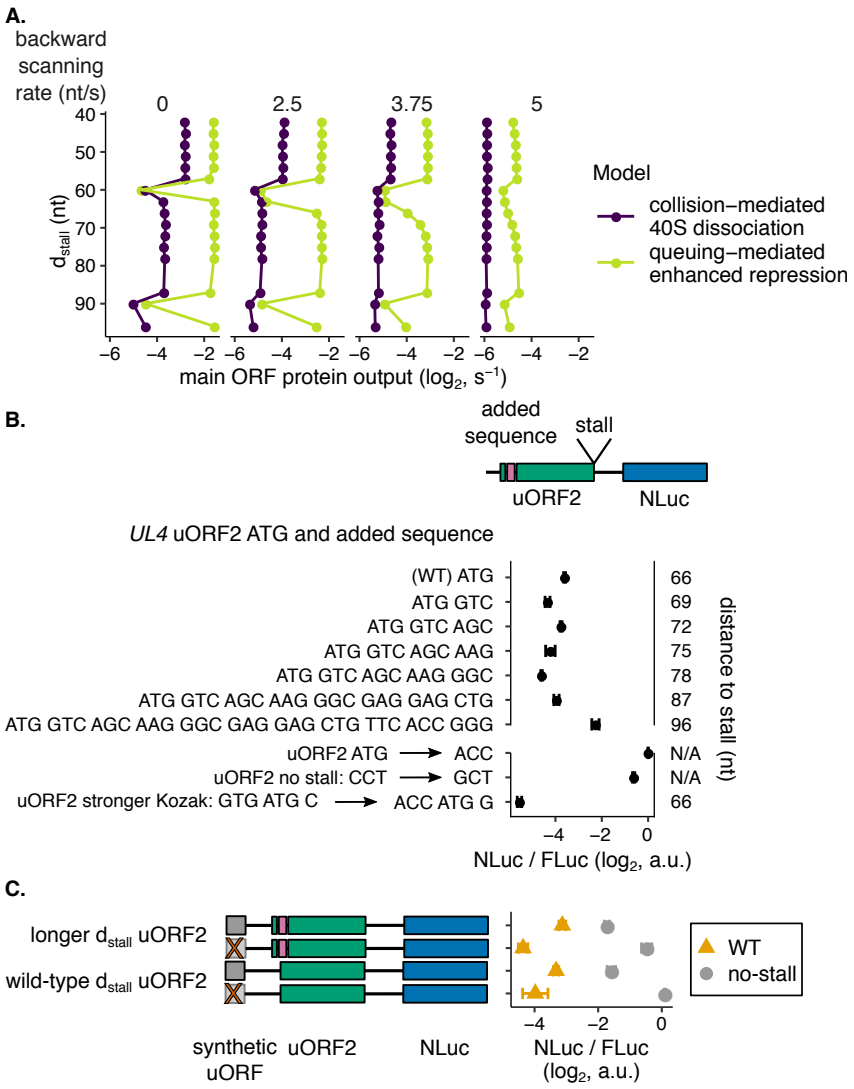
Additional modeling buffering predictions (A) Buffering in the 80S-hit dissociation model is affected by uORF length. Re-initiation is 0.2%. uORF initiation is 80%.
(B) Buffering in the 80S-hit dissociation model is lost with control-matched parameters. Buffering in the 80S-hit dissociation model requires strong uORF initiation and rare re-initiation (Fig. 2.6B, left panel, yellow-green line) and is stronger with longer uORFs (Fig. 2.5A, yellow-green line). However, we estimate re-initiation to be frequent (Table 1) following calibration of our modeling (Fig. 2.2D) to reporter measurements on wild-type or mutant uORF2 (Fig. 2.2C). uORF initiation is 2%. When the elongating ribosome stall is present, d_{stall} is 63 nt to prevent reduction to the queuing-mediated enhanced repression model.
(C) Queuing-mediated enhanced uORF initiation is sensitive to d_{stall}. As the rate of ribosome loading increases, the average queue size increases and allows enhanced uORF initiation only when d_{stall} equals an integer multiple of the ribosome footprint (30 nt).
(D) Backward scanning only relaxes the depending of buffering on d_{stall} in the queuing-mediated enhanced repression model when d_{stall} is close to an integer multiple of the ribosome footprint (30 nt). The forward scanning rate is 5 nt/s. For d_{stall} values of 60, 63, 66 nt, the uORF length is 21, 22, 23 codons, respectively.
(E) Buffering in the collision-mediated 40S dissociation model occurs even with a rather low dissociation rate. Here, d_{stall} is 63 nt.
(F) Buffering in the collision-mediated enhanced repression model (Fig. 2.1D) is insensitive to d_{stall}. All rates and labels are identical to Fig. 2.6 unless otherwise specified. Error bars of simulated data are smaller than data points.

Figure 2.7



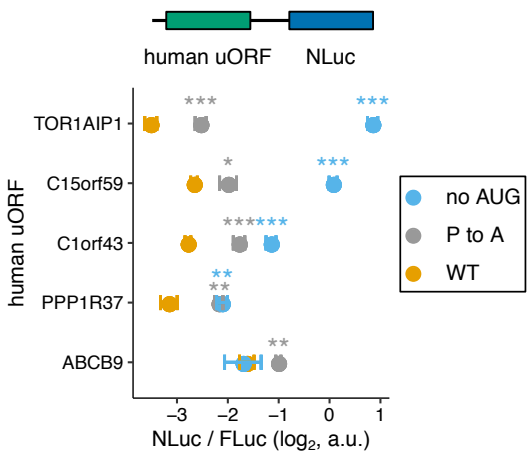
The human cytomegaloviral uORF2 buffers against reductions in main ORF protein output.
The human cytomegaloviral *UL4* uORF2 is used in the dual-luciferase assay (Fig 2.2B) in conjunction with three experimental strategies to reduce ribosome loading.
(A) Ribosome loading is reduced using stem-loops[163] with the indicated GC percentages. All stem-loops are positioned 8 nt from the 5′ cap and have the same predicted stability of -30 kcal/mol. The no-stem-loop construct has a CAA repeat instead of a stem-loop. The 5′ UTR is 287 nt long. Data are normalized to a no-uORF start codon control without a stem-loop.
(B) Ribosome loading is reduced using the drug thapsigargin (1X = 1 μM)[208], which induces the integrated stress response (ISR) by triggering ER stress. NLuc has a C-terminal PEST tag to turnover[165] of protein produced before the 6-hour drug treatment. The 5′ UTR is 236 nt long. Data are normalized to a no-uORF start codon control without a PEST tag. Error bars show the standard error of mean NLuc / FLuc ratios over 4 biological replicates.
(C) Ribosome loading onto the uORF2-NLuc portion of the transcript is reduced using a 5′ synthetic uORF: ATG GGG TAG. The synthetic uORF Kozak is varied to alter ribosome loading. The variants are vertically ordered by the no-stall means. The 5′ UTR is 262 nt long. Data are normalized to a no-uORF start codon control without a synthetic uORF.
The right panels in A, B, and C show wild-type (WT) mean values normalized by the corresponding no-stall values. The no-stall uORF2 mutants lack their terminal diproline motifs (P22A mutation). Unless stated otherwise, error bars show the standard error of mean NLuc / FLuc ratios over 3 biological replicates.

Figure 2.8



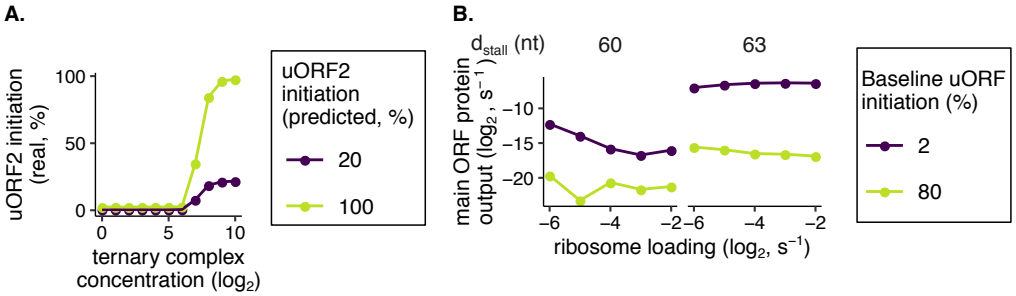
Changes to the distance between the human cytomegaloviral uORF2 start codon and the elongating ribosome stall do not change repressiveness or buffering, consistent with the collision-mediated 40S dissociation model.
(A) Computational modeling predicts greater changes in uORF repressiveness with changes in d_{stall} in the queuing-mediated enhanced repression model. Fast backward scanning abolishes this periodicity. d_{stall} refers to the distance between the start codon and the elongating ribosome stall. As backward scanning increases in rate (moving right along panels), the collision-mediated enhanced repression model loses periodicity (middle panel, purple line) before the queuing-mediated enhanced repression model (right panel, yellow-green line). Parameters that best recapitulated reporter measurements on wild-type or mutant uORF2 (Fig 2.2C, Table 1) are used here. The forward scanning rate is 5 nucleotides/s. Data are normalized to a no-uORF start codon control. Error bars of simulated data are smaller than data markers.
(B) Experimentally varying the distance between the human cytomegaloviral uORF2 start codon and the elongating ribosome stall does not systematically affect its repression of main ORF protein output. The human cytomegaloviral *UL4* uORF2 is used in the dual-luciferase assay (Fig 2.2B) in conjunction with various length inserts from the N-terminus of the *EYFP* main ORF. The *EYFP* main ORF sequence is inserted directly 3′ to the uORF2 start codon. The added sequence increases the distance between the uORF2 start codon and the elongating ribosome stall. The bottom three controls improve the uORF2 Kozak context, remove the start codon, and remove the elongating ribosome stall. Error bars show the standard error of mean NLuc / FLuc ratios over 3 biological replicates. Data are normalized to the no-uORF start codon control.
(C) Experimentally varying the human cytomegaloviral uORF2 d_{stall} does not strongly regulate the capacity of buffering against reductions in main ORF protein output. Ribosome loading is reduced with a 5′ synthetic uORF: ATG GGG TAG. The no-stall uORF2 mutants lack their terminal diproline motifs (P22A mutation). No-synthetic uORF mutants (ATG to AAG) are depicted by transparent, gray bars with red Xs and have a higher relative ribosome loading rate onto the uORF2-NLuc portion of the transcript. The distance between the uORF2 start codon and the elongating ribosome stall is varied as indicated by adding 6 nt, GTC AGC, from the N-terminus of the *EYFP* main ORF. Data are normalized to a no-uORF start codon control without a synthetic uORF.

Figure 2.9



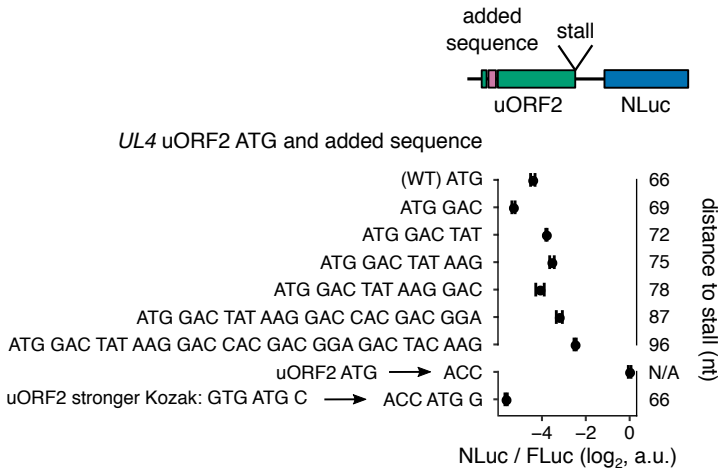
Several human uORFs have repressive terminal diproline motifs.
Terminal diproline motif-containing human uORFs are used in the dual-luciferase assay (Fig 2.2B). The terminal proline codon in each uORF is mutated to an alanine codon in the P to A mutant. Start codons are mutated to ACC for the no-AUG mutants. P values comparing the indicated mutants to the wild-type are from a two-sample t-test: * ($0.01 < P < 0.05$), ** ($0.001 < P < 0.01$), *** ($P < 0.001$). Error bars show the standard error of mean NLuc / FLuc ratios over 5 biological replicates. Data are normalized to a no-*UL4*-uORF2 start codon control.

Figure 2.10



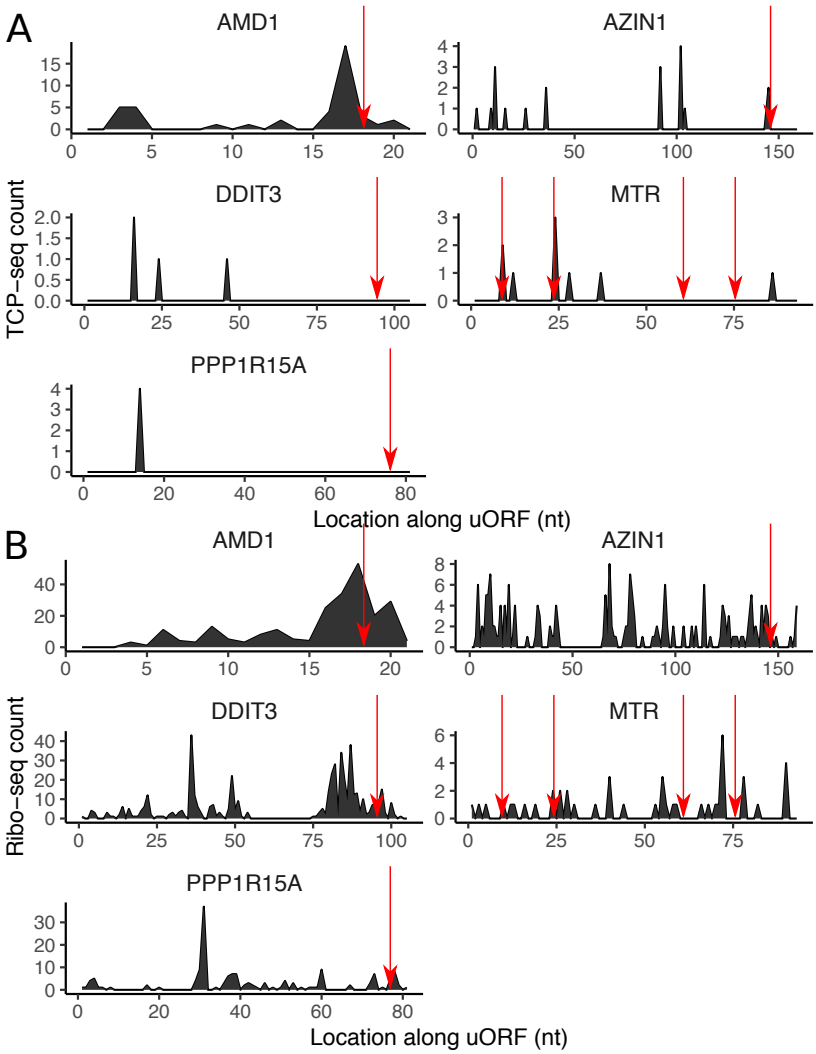
Additional modeling buffering predictions
(A) Initiation at the second downstream uORF is dependent on high ternary complex concentration. Initiation at the first uORF is 100%. Continued scanning fractions at both uORFs are 100%. Following termination at the first uORF, initiation at the second downstream uORF depends on if a new ternary complex has been acquired since termination at the first uORF. Only when ternary complex concentration is high does this real uORF2 initiation fraction approach the predicted fraction.
(B) With an elongating ribosome stall, the 80S-hit dissociation model acquires d_{stall}-dependent buffering similar to that in the queuing-mediated enhanced repression model (Fig. 2.6C). Re-initiation is 0.2%.
All rates and labels are identical to Fig. 2.6 unless otherwise specified. Error bars of simulated data are smaller than data points.

Figure 2.11



Experimentally increasing the distance between the human cytomegaloviral uORF2 start codon and elongating ribosome stall using FLAG donor sequence.
The human cytomegaloviral *UL4* uORF2 is used in the dual-luciferase assay (Fig. 2.2B) in conjunction with various length inserts from the N-terminus of the *FLAG* main ORF. The *FLAG* main ORF sequence is inserted directly 3′ to the uORF2 start codon. The added sequence increases the distance between the uORF2 start codon and elongating ribosome stall. The bottom two controls improve the uORF2 Kozak context and remove the start codon. Error bars show the standard error of mean NLuc / FLuc ratios over 3 biological replicates. Data are normalized to a no-uORF start codon control.

Figure 2.12



Ribosome density within elongation stall-containing human uORFs.
(A) Small ribosomal subunit (TCP-seq[175]) coverage data.
(B) Elongating ribosome (Ribo-seq, A site global aggregate) coverage data.
AMD1, *AZIN1*, *DDIT3* (*CHOP*), *MTR* and *PPP1R15A* (*GADD34*) uORF amino acid sequences are MAGDIS, IPPKKRRRFTRLFGPLSHGELSDQVYNYPEGLGEVLYREQFDFNAEPPWEPS, MLKMSGWQRQSQNQSWNLRRECSRRKCIFIHHHT, MSRRPPLPVFSWVLFRAVPRLRLWPRVSGC, and MNALASLTVRTCDRFWQTEPALLPPG, respectively. Elongation stall locations are marked with red arrows. Coverage data were downloaded from GWIPS[209].

Table 1

Parameter ranges and fit values for modeling

Parameter	Value range	Fit value (80S-hit dissociation)	Fit value (queuing-mediated enhanced repression)	Fit value (collision-mediated 40S dissociation)	Reference
$k_{cap\ bind}$ (s^{-1})	0.02-0.06	0.016	0.023	0.025	This work[144–146]
k_{scan} (nt/s)	1-10	5	5	5	[75,206]
$k_{start\ uORF2\ WT}$ (s^{-1})	unknown	0.1	0.5	0.5	This work
WT uORF2 initiation (%)	unknown	2	10	10	This work
$k_{start\ uORF2\ strong\ Kozak}$ (s^{-1})	unknown	20	5	5	This work
strong Kozak uORF2 initiation (%)	unknown	80	50	50	This work
k_{elong} (codons/s)	3-10	5	5	5	[144–146,157]
$k_{elong\ stall}$ (codons/s)	0.001	0.001	0.001	0.001	[180]
$k_{terminate}$ (s^{-1})	0.5-5	1	1	1	[157]
$k_{terminated\ ssu\ recycle\ uORF}$ (s^{-1})	unknown	2	5	5	This work
Re-initiation (%)	unknown	75	50	50	This work
$k_{dissociate}$ (s^{-1})	unknown	2	0	2	This work
uORF length (codons)	21	21	21	21	[105]

Chapter 3: GEsim enables efficient simulation and visualization of kinetic models for gene expression processes

I. Introduction

Motivation for kinetic modeling

Many biological "omics" methodologies create large amounts of data[210] which can lead to many interesting biological conclusions but complicate extractions of biological conclusions. Much of the data generated by these methods is not directly interpretable and requires some form of modeling to understand. In the case of gene expression data, this complexity results from gene expression processes involving several molecular complexes acting simultaneously on the same gene or mRNA or several temporally choreographed processes affecting an outcome such as the production of mRNA or protein. There is a clear need for quantitative kinetic models for evaluating alternate hypotheses and inferring rates and parameters from experimental observables.

Limitations of previous modeling studies

Previous studies have used modeling to investigate gene expression processes, but prior models, while highly complex, have not been modular or easily extensible. There has not been any framework for visualizing models or simulation results. Parameters are often presented as tables, but models themselves are not transparent. As modeling and simulation are intertwined, it is difficult to glean modeling assumptions from computer code.

The rule-based modeling framework

Rule-based modeling is a powerful approach to understanding biological data, primarily in cell signaling. It has been used to model many biological processes. These models are compact and can efficiently generate large reaction networks. The number of reactions and chemical species can exponentially grow with only a linear increase in the number of rules and parameters. PySB, a Python-based modeling framework[141], is used to specify rule-based model kinetics and interface with the BioNetGen language[142,211]. The agent-based Gillespie algorithm in NFsim, a network-free simulation framework, is

used to stochastically simulate models. This approach treats molecules as objects and performs simulations directly on them. The simulation is repeated until the desired number of time steps have been reached.

Major advances in this work

Extending the applicability of modeling to gene expression processes

Previously, NFsim searched for rate changes across all reactions at every time step, which is inefficient for large models. Here, we create a connectivity graph once at the beginning of the simulation. This connectivity graph identifies only the reactions that need to be updated (reactions that can now occur based on the last step update) at each time step. Connected reactions share at least 1 reactant or product component; reactant and product molecules in connected reactions don't share incompatible component states or bonds. Rate changes are only searched for connected reactions. The connectivity changes made here are a major cornerstone of the GEsim (gene expression simulation) framework.

Generalized visualization

The second major advancement to create GEsim is an automatic generation of a simulation visualization for any model based on user inputs for the appearance of molecules, the location of molecule sites, the appearance of molecule site states, and how molecules and site states should transform for different reactions. For example, the user can specify how ribosomes bind to mRNA and whether the mRNA remains static or moves to the ribosome. The molecules and site states are represented by SVG, scalable vector graphics, classes, with each molecule represented as an object that includes a list of associated sites and their respective representations. The site states are also associated with their respective representations. Visualization of simulation results allows users to gain insight into biological processes and understand the behavior of complex systems. The ability to input data on the appearance and behavior of molecules and site states allows for a highly customizable and accurate representation of the simulation output.

II. Results and Discussion

Analysis of ribosome collisions

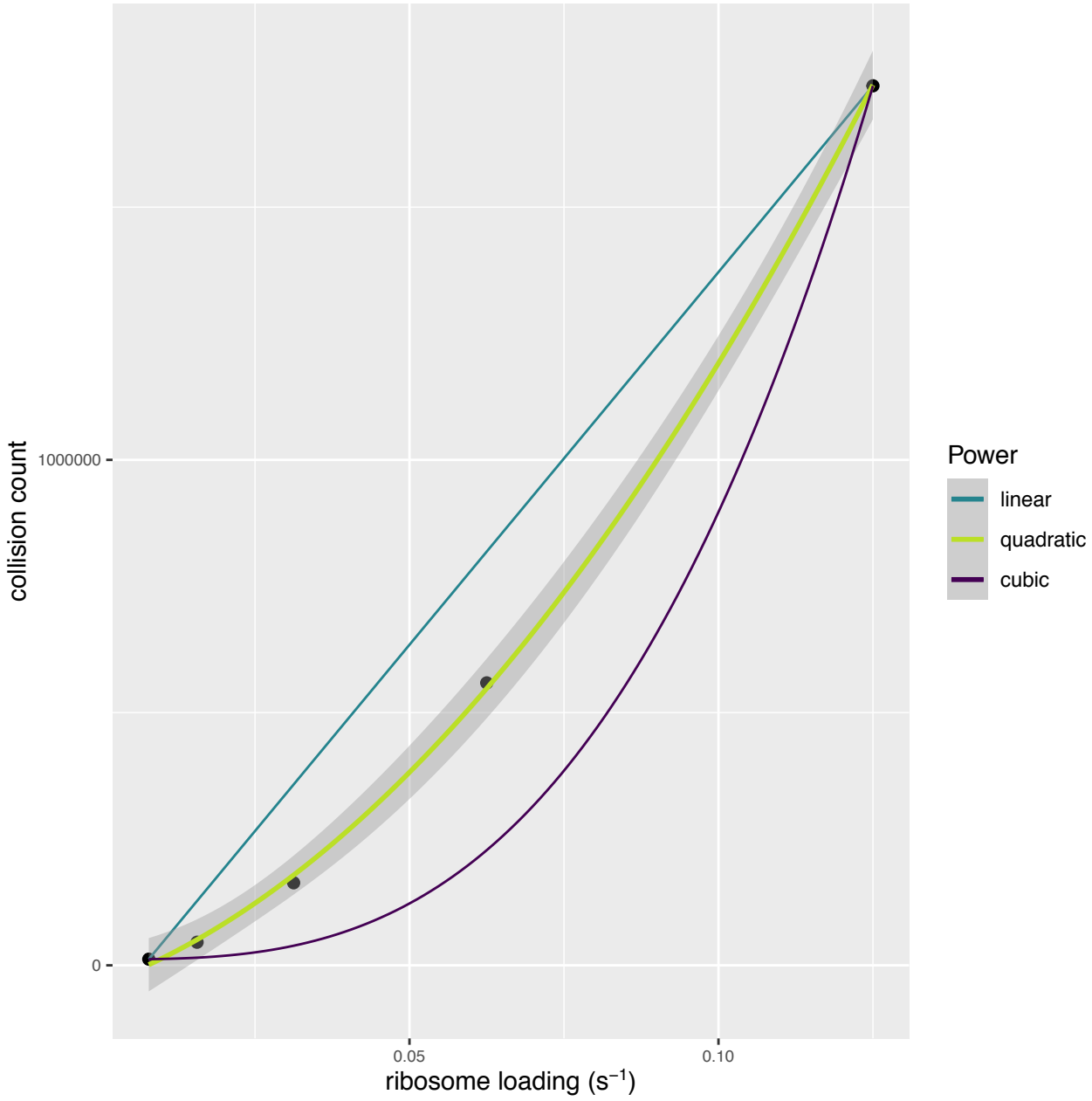
To test the modeling framework, we simulated a simple model of ribosome collisions. The frequency of ribosome collisions is expected to scale quadratically with the initiation rate because the number of possible pairs of ribosomes that could collide increases quadratically with the number of ribosomes in the system. We simulated various initiation rates, near the estimated physiological ribosome loading rate of 0.05/s[144–146], and measured the frequency of ribosome collisions. Linear, quadratic, and cubic lines (legend colors) as well as a 95% confidence interval for the quadratic fit (grey area) are shown in Fig 3.1 We observe that ribosome collisions do, in fact, scale quadratically with the initiation rate. This investigation of the relationship between ribosome collisions and initiation rate yields confidence in the modeling framework as this quadratic relationship is mathematically expected.

Future directions

Next, we will verify simulations and analyses of a TASEP phase diagram and simulate more complex models that capture additional biological features. One specific, complex model that would integrate multiple gene expression processes and display the power of computational modeling would be the SARS-CoV-2 viral life cycle, with the Nsp1 protein being of special interest. This viral protein represses translation of both viral and host mRNAs, induces mRNA degradation, and blocks nuclear mRNA export; Nsp1 preferentially represses expression of host mRNAs[212–214].

III. Figures

Figure 3.1



Ribosome collisions scale quadratically with the ribosome loading rate.
A quadratic regression model is fit for the ribosome collision frequency data. Linear, quadratic, and cubic lines are displayed (see legend colors). A 95% confidence interval for the quadratic fit (grey area) is also shown.

Chapter 4: Conclusions

In this dissertation, I report the use of an experimentally-integrated computational modeling approach to investigate mechanisms of uORF regulation of main ORF translation in eukaryotes. While individual kinetic models have been separately investigated before, to our knowledge, this work is the first systematic investigation of what uORF metrics impart buffering in each kinetic model of uORF regulation. We discover that ribosomes likely dissociate, or are rescued, rather than queue when they collide with a 3' stalled elongating ribosome. This discovery transforms the prevalent model of queuing ribosomes in the translation field. We aspire for others to also incorporate such experimentally-integrated computational modeling into their research to further our collective understanding of gene expression. We hope that the framework used here and its visualization tools will be both straightforward to use and offer novel insights.

The experimentally-integrated computational modeling approach used in this dissertation provides a comprehensive understanding of uORF regulation of main ORF translation. Additionally, the visualization tools developed in this study enable researchers to easily interpret and analyze the complex data. However, there are also some potential pitfalls to consider. For example, the accuracy of the computational models is dependent on the quality and availability of experimental data, and there may be limitations to the extent to which experimental results can be extrapolated to other systems. In terms of assessing the generalizability of the modeling approach, it will be important to test the models in other eukaryotic systems and compare the results with experimental data. To push the field forward, it will be important to validate the predictions of the computational models through further experimental studies. Additionally, it will be important to incorporate other factors that can influence translation regulation, such as cap-tethered initiation, elongation time-derived re-initiation frequencies, and cellular stress-reduced elongation rates, into the modeling framework.

The findings in this dissertation suggest that ribosomes likely dissociate or are rescued when they collide with a stalled elongating ribosome. However, the specific proteins responsible for this rescue mechanism are currently unknown. We are interested in discovering the proteins that rescue small ribosomal subunits from mRNAs. Identifying these proteins would provide new insights into the regulation of gene expression.

One potential approach for discovering these proteins is to perform a whole-genome CRIPSR screen to determine factors that mediate stall-induced repression on the main ORF. Such a potential strategy involves infecting millions of cells, sorting cells based on fluorescent ratios, and testing changes in reporter activity by knocking out genes, such as eIFs, known to modulate uORF-mediated repression. Once candidate proteins have been identified, their functions can be further validated through biochemical and biophysical assays.

References

1. Morris, R., Black, K. A. & Stollar, E. J. Uncovering protein function: From classification to complexes. *Essays in Biochemistry* **66**, 255–285 (2022).

2. Tsugita, A. & Fraenkel-Conrat, H. THE AMINO ACID COMPOSITION AND c-TERMINAL SEQUENCE OF a CHEMICALLY EVOKED MUTANT OF TMV*. *Proc Natl Acad Sci U S A* **46**, 636–642 (1960).

3. Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961).

4. Mercadante, A. A., Dimri, M. & Mohiuddin, S. S. Biochemistry, replication and transcription. in *StatPearls* (StatPearls Publishing, 2022).

5. Kornberg, R. D. Chromatin structure: A repeating unit of histones and DNA. *Science* **184**, 868–871 (1974).

6. Mizuguchi, G., Tsukiyama, T., Wisniewski, J. & Wu, C. Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol Cell* **1**, 141–150 (1997).

7. Czarnota, G. J., Bazett-Jones, D. P., Mendez, E., Allfrey, V. G. & Ottensmeyer, F. P. High resolution microanalysis and three-dimensional nucleosome structure associated with transcribing chromatin. *Micron* **28**, 419–431 (1997).

8. Reményi, A., Schöler, H. R. & Wilmanns, M. Combinatorial control of gene expression. *Nat Struct Mol Biol* **11**, 812–815 (2004).

9. Jurado, A. R., Tan, D., Jiao, X., Kiledjian, M. & Tong, L. Structure and function of pre-mRNA 5′-end capping quality control and 3′-end processing. *Biochemistry* **53**, 1882–1898 (2014).

10. Solov'ev, V. V. & Kolchanov, N. A. [Exon-intron structure of eukaryotic genes can be due to the nucleosome organization of chromatin and to its related characteristics of gene expression regulation]. *Dokl Akad Nauk SSSR* **284**, 232–237 (1985).

11. Nicholson, A. L. & Pasquinelli, A. E. Tales of detailed poly(a) tails. *Trends Cell Biol* **29**, 191–200 (2019).

12. Passmore, L. A. & Coller, J. Roles of mRNA poly(a) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol* **23**, 93–106 (2022).

13. Maquat, L. E. Nuclear mRNA export. *Curr Opin Cell Biol* **3**, 1004–1012 (1991).

14. Takamatsu, N. *et al.* Heat shock factor 1 induces a short burst of transcription of the clock gene Per2 during interbout arousal in mammalian hibernation. *J Biol Chem* 104576 (2023) doi:10.1016/j.jbc.2023.104576.

15. Yang, X. *et al.* GATA4 forms a positive feedback loop with CDX2 to transactivate MUC2 in bile acids-induced gastric intestinal metaplasia. *Gut Liver* (2023) doi:10.5009/gnl220394.

16. Jones, B. Gene regulation: Transcription factor clutch control. *Nat Rev Genet* **13**, 380 (2012).

17. Zhang, C. & Peng, G. Non-coding RNAs: An emerging player in DNA damage response. *Mutat Res Rev Mutat Res* **763**, 202–211 (2015).

18. Picard, F., Loubière, P., Girbal, L. & Cocaign-Bousquet, M. The significance of translation regulation in the stress response. *BMC Genomics* **14**, 588 (2013).

19. Venters, B. J. & Pugh, B. F. How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol* **44**, 117–141 (2009).

20. Sproul, D., Gilbert, N. & Bickmore, W. A. The role of chromatin structure in regulating the expression of clustered genes. *Nat Rev Genet* **6**, 775–781 (2005).

21. Tomecki, R. & Drazkowska, K. An integrative approach uncovers transcriptome-wide determinants of mRNA stability regulation in saccharomyces cerevisiae. *FEBS J* **288**, 3418–3423 (2021).

22. Petasny, M. *et al.* Splicing to keep cycling: The importance of pre-mRNA splicing during the cell cycle. *Trends Genet* **37**, 266–278 (2021).

23. Kornblihtt, A. R. *et al.* Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**, 153–165 (2013).

24. Gehring, N. H. & Roignant, J.-Y. Anything but ordinary – emerging splicing mechanisms in eukaryotic gene regulation. *Trends in Genetics* **37**, 355–372 (2021).

25. Hasan, A., Cotobal, C., Duncan, C. D. S. & Mata, J. Systematic analysis of the role of RNA-binding proteins in the regulation of RNA stability. *PLoS Genet* **10**, e1004684 (2014).

26. Mitchell, P. & Tollervey, D. mRNA stability in eukaryotes. *Curr Opin Genet Dev* **10**, 193–198 (2000).

27. Tatosyan, K. A., Ustyantsev, I. G. & Kramerov, D. A. RNA degradation in eukaryotic cells. *Mol Biol* **54**, 485–502 (2020).

28. Ross, J. Control of messenger RNA stability in higher eukaryotes. *Trends Genet* **12**, 171–175 (1996).

29. Song, C.-R. *et al.* mRNA 3' -UTR-mediate translational control through PAS and CPE in sheep oocyte. *Theriogenology* **201**, 30–40 (2022).

30. Farberov, L. *et al.* Multiple copies of microRNA binding sites in long 3'UTR variants regulate axonal translation. *Cells* **12**, 233 (2023).

31. Valinezhad Orang, A., Safaralizadeh, R. & Kazemzadeh-Bavili, M. Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. *Int J Genomics* **2014**, 970607 (2014).

32. Godfrey, T. C. *et al.* The microRNA-23a cluster regulates the developmental HoxA cluster function during osteoblast differentiation. *J Biol Chem* **293**, 17646–17660 (2018).

33. Tsan, Y.-C., Morell, M. H. & O'Shea, K. S. miR-410 controls adult SVZ neurogenesis by targeting neurogenic genes. *Stem Cell Res* **17**, 238–247 (2016).

34. Boo, S. H. & Kim, Y. K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp Mol Med* **52**, 400–408 (2020).

35. Grammatikakis, I., Abdelmohsen, K. & Gorospe, M. Posttranslational control of HuR function. *Wiley Interdiscip Rev RNA* **8**, 10.1002/wrna.1372 (2017).

36. White, E. J. F., Matsangos, A. E. & Wilson, G. M. AUF1 regulation of coding and noncoding RNA. *Wiley Interdiscip Rev RNA* **8**, 10.1002/wrna.1393 (2017).

37. Holschuh, K. & Gassen, H. G. Mechanism of translocation. Binding equilibria between the ribosome, mRNA analogues, and cognate tRNAs. *J Biol Chem* **257**, 1987–1992 (1982).

38. Strader, C. D. Mutagenesis approaches for elucidation of protein structure-function relationships. *Curr Protoc Neurosci* **Chapter 4**, Unit 4.9 (2001).

39. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113–127 (2010).

40. Akirtava, C. & McManus, C. J. Control of translation by eukaryotic mRNA transcript leaders-insights from high-throughput assays and computational modeling. *Wiley Interdiscip Rev RNA* **12**, e1623 (2021).

41. Kozak, M. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature* **308**, 241–246 (1984).

42. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**, 8125–8148 (1987).

43. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).

44. Brina, D., Miluzio, A., Ricciardi, S. & Biffo, S. eIF6 anti-association activity is required for ribosome biogenesis, translational control and tumor progression. *Biochim Biophys Acta* **1849**, 830–835 (2015).

45. Lührmann, R. Dinucleotide codon-anticodon interaction as a minimum requirement for ribosomal aa-tRNA binding: Stabilisation by viomycin of aa-tRNA in the a site. *Nucleic Acids Res* **8**, 5813–5824 (1980).

46. Rheinberger, H. J. The function of the translating ribosome: Allosteric three-site model of elongation. *Biochimie* **73**, 1067–1088 (1991).

47. Dever, T. E., Dinman, J. D. & Green, R. Translation elongation and recoding in eukaryotes. *Cold Spring Harb Perspect Biol* **10**, a032649 (2018).

48. Grunberger, D., Weinstein, I. B. & Jacobson, K. B. Codon recognition by enzymatically mischarged valine transfer ribonucleic acid. *Science* **166**, 1635–1637 (1969).

49. Liu, S. *et al.* Diphthamide modification on eukaryotic elongation factor 2 is needed to assure fidelity of mRNA translation and mouse development. *Proceedings of the National Academy of Sciences* **109**, 13817–13822 (2012).

50. Joazeiro, C. A. P. Ribosomal stalling during translation: Providing substrates for ribosome-associated protein quality control. *Annual Review of Cell and Developmental Biology* **33**, 343–368 (2017).

51. Ikeuchi, K., Izawa, T. & Inada, T. Recent progress on the molecular mechanism of quality controls induced by ribosome stalling. *Frontiers in Genetics* **9**, (2019).

52. Lin, Y. *et al.* Impacts of uORF codon identity and position on translation regulation. (2019) doi:10.1184/R1/10110866.v1.

53. Filbeck, S., Cerullo, F., Pfeffer, S. & Joazeiro, C. A. P. Ribosome-associated quality-control mechanisms from bacteria to humans. *Mol Cell* **82**, 1451–1466 (2022).

54. Hellen, C. U. T. Translation termination and ribosome recycling in eukaryotes. *Cold Spring Harb Perspect Biol* **10**, a032656 (2018).

55. Kisselev, L. L. & Frolova LYu, null. Termination of translation in eukaryotes. *Biochem Cell Biol* **73**, 1079–1086 (1995).

56. Lu, P. D., Harding, H. P. & Ron, D. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *J Cell Biol* **167**, 27–33 (2004).

57. Dimaline, R. Post-translational modification of peptide messengers in the gut. *Q J Exp Physiol* **73**, 873–902 (1988).

58. Lee, D. S. M. *et al.* Disrupting upstream translation in mRNAs is associated with human disease. *Nat Commun* **12**, 1515 (2021).

59. Wang, Y.-C., Peterson, S. E. & Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res* **24**, 143–160 (2014).

60. Ciechanover, A. & Schwartz, A. L. The ubiquitin-proteasome pathway: The complexity and myriad functions of proteins death. *Proc Natl Acad Sci U S A* **95**, 2727–2730 (1998).

61. Walker, D. C., Hill, G., Wood, S. M., Smallwood, R. H. & Southgate, J. Agent-based computational modeling of wounded epithelial cell monolayers. *IEEE Trans Nanobioscience* **3**, 153–163 (2004).

62. Vodovotz, Y. & An, G. Agent-based models of inflammation in translational systems biology: A decade later. *Wiley Interdiscip Rev Syst Biol Med* **11**, e1460 (2019).

63. Wilm, A., Kühnl, J. & Kirchmair, J. Computational approaches for skin sensitization prediction. *Crit Rev Toxicol* **48**, 738–760 (2018).

64. Goldberg, A. P. *et al.* Emerging whole-cell modeling principles and methods. *Curr Opin Biotechnol* **51**, 97–102 (2018).

65. Kühn, C. & Gennemark, P. Modeling yeast osmoadaptation at different levels of resolution. *J Bioinform Comput Biol* **11**, 1330001 (2013).

66. Scotti, M., Stella, L., Shearer, E. J. & Stover, P. J. Modeling cellular compartmentation in one-carbon metabolism. *Wiley Interdiscip Rev Syst Biol Med* **5**, 343–365 (2013).

67. An, G. & Christley, S. Addressing the translational dilemma: Dynamic knowledge representation of inflammation using agent-based modeling. *Crit Rev Biomed Eng* **40**, 323–340 (2012).

68. Brodland, G. W. How computational models can help unlock biological systems. *Seminars in Cell & Developmental Biology* **47-48**, 62–73 (2015).

69. Loewe, L. & Hillston, J. Computational models in systems biology. *Genome Biology* **9**, 328 (2008).

70. Atwell, K., Dunn, S.-J., Osborne, J. M., Kugler, H. & Albert Hubbard, E. J. How computational models contribute to our understanding of the germ line. *Mol Reprod Dev* **83**, 944–957 (2016).

71. Cao, D. & Parker, R. Computational modeling of eukaryotic mRNA turnover. *RNA* **7**, 1192–1212 (2001).

72. Valentino, M. *et al.* Computational modeling offers new insight into drosophila germ granule development. *Biophys J* **121**, 1465–1482 (2022).

73. Chylek, L. A., Harris, L. A., Faeder, J. R. & Hlavacek, W. S. Modeling for (physical) biologists: An introduction to the rule-based approach. *Phys Biol* **12**, 045007 (2015).

74. Haar, T. von der. Mathematical and computational modelling of ribosomal movement and protein synthesis: An overview. *Comput Struct Biotechnol J* **1**, e201204002 (2012).

75. Andreev, D. E. *et al.* TASEP modelling provides a parsimonious explanation for the ability of a single uORF to derepress translation during the integrated stress response. *eLife* **7**, e32563 (2018).

76. Szavits-Nossan, J. & Evans, M. R. Dynamics of ribosomes in mRNA translation under steady- and nonsteady-state conditions. *Phys. Rev. E* **101**, 062404 (2020).

77. Erdmann-Pham, D. D., Dao Duc, K. & Song, Y. S. The key parameters that govern translation efficiency. *Cell Syst* **10**, 183–192.e6 (2020).

78. Chylek, L. A., Wilson, B. S. & Hlavacek, W. S. Modeling biomolecular site dynamics in immunoreceptor signaling systems. *Adv Exp Med Biol* **844**, 245–262 (2014).

79. Prasad, A. Computational modeling of signal transduction networks: A pedagogical exposition. *Methods Mol Biol* **880**, 219–241 (2012).

80. Sekar, J. A. P. & Faeder, J. R. Rule-based modeling of signal transduction: A primer. *Methods Mol Biol* **880**, 139–218 (2012).

81. Terfve, C. & Saez-Rodriguez, J. Modeling signaling networks using high-throughput phosphoproteomics. *Adv Exp Med Biol* **736**, 19–57 (2012).

82. Hlavacek, W. S. *et al.* Rules for modeling signal-transduction systems. *Sci STKE* **2006**, re6 (2006).

83. Morris, D. R. & Geballe, A. P. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**, 8635–8642 (2000).

84. Chen, H.-H. & Tarn, W.-Y. uORF-mediated translational control: Recently elucidated mechanisms and implications in cancer. *RNA Biol* **16**, 1327–1338 (2019).

85. Alghoul, F., Laure, S., Eriani, G. & Martin, F. Translation inhibitory elements from Hoxa3 and Hoxa11 mRNAs use uORFs for translation inhibition. *eLife* **10**, e66369 (2021).

86. Barbosa, C., Peixeiro, I. & Romão, L. Gene expression regulation by upstream open reading frames and human disease. *PLOS Genetics* **9**, e1003529 (2013).

87. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORFs are prevalent translational repressors in vertebrates. *The EMBO Journal* **35**, 706–723 (2016).

88. Zhang, H., Wang, Y. & Lu, J. Function and evolution of upstream ORFs in eukaryotes. *Trends in Biochemical Sciences* **44**, 782–794 (2019).

89. Silva, J., Fernandes, R. & Romão, L. Gene expression regulation by upstream open reading frames in rare diseases. *Journal of Rare Diseases Research & Treatment* **2**, (2017).

90. Nelde, A. *et al.* Upstream open reading frames regulate translation of cancer-associated transcripts and encode HLA-presented immunogenic tumor antigens. *Cell. Mol. Life Sci.* **79**, 171 (2022).

91. Vattem, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences* **101**, 11269–11274 (2004).

92. Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601 (2013).

93. Gandin, V. *et al.* Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. *Nature* **455**, 684–688 (2008).

94. Sharma, A. K. *et al.* A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLoS Comput Biol* **15**, (2019).

95. Cao, J. & Geballe, A. P. Coding sequence-dependent ribosomal arrest at termination of translation. *Molecular and Cellular Biology* **16**, 603–608 (1996).

96. Young, S. K., Baird, T. D. & Wek, R. C. Translation regulation of the glutamyl-prolyl-tRNA synthetase gene EPRS through bypass of upstream open reading frames with noncanonical initiation codons. *Journal of Biological Chemistry* **291**, 10824–10835 (2016).

97. Ivanov, I. P. *et al.* Polyamine Control of Translation Elongation Regulates Start Site Selection on the Antizyme Inhibitor mRNA via Ribosome Queuing. *Mol Cell* **70**, 254–264.e6 (2018).

98. Wei, J., Wu, C. & Sachs, M. S. The arginine attenuator peptide interferes with the ribosome peptidyl transferase center. *Mol Cell Biol* **32**, 2396–2406 (2012).

99. Wang, Z., Gaba, A. & Sachs, M. S. A highly conserved mechanism of regulated ribosome stalling mediated by fungal arginine attenuator peptides that appears independent of the charging status of arginyl-tRNAs *. *Journal of Biological Chemistry* **274**, 37565–37574 (1999).

100. Bhushan, S. *et al.* Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell* **40**, 138–146 (2010).

101. Wilson, D. N., Arenz, S. & Beckmann, R. Translation regulation via nascent polypeptide-mediated ribosome stalling. *Current Opinion in Structural Biology* **37**, 123–133 (2016).

102 Meijer, H. A. & Thomas, A. A. M. Ribosomes stalling on uORF1 in the xenopus Cx41 5' UTR inhibit downstream translation initiation. *Nucleic Acids Res* **31**, 3174–3184 (2003).

103 Young, S. K., Palam, L. R., Wu, C., Sachs, M. S. & Wek, R. C. Ribosome elongation stall directs gene-specific translation in the integrated stress response. *Journal of Biological Chemistry* **291**, 6546–6558 (2016).

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

105 Degnin, C., Schleiss, M., Cao, J. & Geballe, A. Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (Gp48) transcript. *Journal of virology* (1993) doi:10.1128/JVI.67.9.5514-5521.1993.

106 Smirnova, V. V. *et al.* Ribosomal leaky scanning through a translated uORF requires eIF4G2. *Nucleic Acids Research* **50**, 1111–1127 (2022).

107 David, M. *et al.* DAP5 drives translation of specific mRNA targets with upstream ORFs in human embryonic stem cells. *RNA* **28**, 1325–1336 (2022).

108 Wu, H.-Y. L. & Hsu, P. Y. Actively translated uORFs reduce translation and mRNA stability independent of NMD in arabidopsis. (2021) doi:10.1101/2021.09.16.460672.

109 Oliveira, C. C. & McCarthy, J. E. G. The relationship between eukaryotic translation and mRNA stability: A SHORT UPSTREAM OPEN READING FRAME STRONGLY INHIBITS TRANSLATIONAL INITIATION AND GREATLY ACCELERATES mRNA DEGRADATION IN THE YEAST SACCHAROMYCES CEREVISIAE*. *Journal of Biological Chemistry* **270**, 8936–8943 (1995).

110 Wethmar, K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **5**, 765–778 (2014 Nov-Dec).

111 Ye, Y. *et al.* Analysis of human upstream open reading frames and impact on gene expression. *Hum Genet* **134**, 605–612 (2015).

112. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. 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 (2009).

113. Rodriguez, C. M., Chun, S. Y., Mills, R. E. & Todd, P. K. Translation of upstream open reading frames in a model of neuronal differentiation. *BMC Genomics* **20**, 391 (2019).

114. Chen, J. *et al.* Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146 (2020).

115. Zhang, H. *et al.* Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nature Communications* **12**, 1076 (2021).

116. Giess, A. *et al.* Deconstructing the individual steps of vertebrate translation initiation. *bioRxiv* 811810 (2019) doi:10.1101/811810.

117. Chew, G.-L., Pauli, A. & Schier, A. F. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nature Communications* **7**, 11663 (2016).

118. Dvir, S. *et al.* Deciphering the rules by which 5′-UTR sequences affect protein expression in yeast. *PNAS* **110**, E2792–E2801 (2013).

119. Jürgens, L. *et al.* Somatic functional deletions of upstream open reading frame-associated initiation and termination codons in human cancer. *Biomedicines* **9**, 618 (2021).

120. Schulz, J. *et al.* Loss-of-function uORF mutations in human malignancies. *Sci Rep* **8**, 2395 (2018).

121. Smith, R. C. L. *et al.* Translation initiation in cancer at a glance. *J Cell Sci* **134**, (2021).

122. Dever, T. E. *et al.* Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell* **68**, 585–596 (1992).

123 Boersma, S. *et al.* Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* **178**, 458–472.e19 (2019).

124 Zur, H. & Tuller, T. Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution. *Nucleic Acids Res* **44**, 9031–9049 (2016).

125 Abastado, J. P., Miller, P. F. & Hinnebusch, A. G. A quantitative model for translational control of the GCN4 gene of saccharomyces cerevisiae. *New Biol* **3**, 511–524 (1991).

126 You, T., Stansfield, I., Romano, M. C., Brown, A. J. & Coghill, G. M. Analysing GCN4 translational control in yeast by stochastic chemical kinetics modelling and simulation. *BMC Syst Biol* **5**, 131 (2011).

127 Marasco, O. N. J. M., Roussel, M. R. & Thakor, N. Probabilistic models of uORF-mediated ATF4 translation control. *Mathematical Biosciences* 108762 (2021) doi:10.1016/j.mbs.2021.108762.

128 Park, H. & Subramaniam, A. R. Inverted translational control of eukaryotic gene expression by ribosome collisions. *PLoS Biol* **17**, e3000396 (2019).

129 Poyry, T. A. A. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes & Development* **18**, 62–75 (2004).

130 Kozak, M. Constraints on reinitiation of translation in mammals. *Nucleic Acids Res* **29**, 5226–5232 (2001).

131 Luukkonen, B. G., Tan, W. & Schwartz, S. 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 (1995).

132 Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**, 8125–8148 (1987).

133 Grant, C. M. & Hinnebusch, A. G. Effect of sequence context at stop codons on efficiency of reinitiation in GCN4 translational control. *Mol Cell Biol* **14**, 606–618 (1994).

134 Hinnebusch, A. G. Translational regulation of yeast GCN4: A WINDOW ON FACTORS THAT CONTROL INITIATOR-tRNA BINDING TO THE RIBOSOME *. *Journal of Biological Chemistry* **272**, 21661–21664 (1997).

135 Janzen, D. M., Frolova, L. & Geballe, A. P. Inhibition of translation termination mediated by an interaction of eukaryotic release factor 1 with a nascent peptidyl-tRNA. *Mol Cell Biol* **22**, 8562–8570 (2002).

136 Col, B., Oltean, S. & Banerjee, R. Translational Regulation of Human Methionine Synthase by Upstream Open Reading Frames. *Biochim Biophys Acta* **1769**, 532–540 (2007).

137 Yamashita, Y. *et al.* Sucrose sensing through nascent peptide-meditated ribosome stalling at the stop codon of arabidopsis bZIP11 uORF2. *FEBS Letters* **591**, 1266–1277 (2017).

138 Tanaka, M. *et al.* The minimum open reading frame, AUG-stop, induces boron-dependent ribosome stalling and mRNA degradation. *Plant Cell* **28**, 2830–2849 (2016).

139 Hou, C. *et al.* Global analysis of truncated RNA ends reveals new insights into ribosome stalling in plants. *The Plant Cell* **28**, 2398–2416 (2016).

140 Uchiyama-Kadokura, N. *et al.* Polyamine-responsive ribosomal arrest at the stop codon of an upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA decay in arabidopsis thaliana. *Plant Cell Physiol.* **55**, 1556–1567 (2014).

141 Lopez, C. F., Muhlich, J. L., Bachman, J. A. & Sorger, P. K. Programming biological models in Python using PySB. *Molecular Systems Biology* **9**, (2013).

142 Harris, L. A. *et al.* BioNetGen 2.2: Advances in rule-based modeling. *Bioinformatics* **32**, 3366–3368 (2016).

143 Sneddon, M. W., Faeder, J. R. & Emonet, T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Methods* **8**, 177–183 (2011).

144 Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165**, 976–989 (2016).

145Morisaki, T. *et al.* Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**, 1425–1429 (2016).

146Wu, B., Eliscovich, C., Yoon, Y. J. & Singer, R. H. Translation dynamics of single mRNAs in live cells and neurons. *Science* **352**, 1430–1435 (2016).

147Costa-Mattioli, M. & Walter, P. The integrated stress response: From mechanism to disease. *Science* **368**, (2020).

148Pakos-Zebrucka, K. *et al.* The integrated stress response. *EMBO Rep* **17**, 1374–1395 (2016).

149Wu, C. C.-C., Peterson, A., Zinshteyn, B., Regot, S. & Green, R. Ribosome collisions trigger general stress responses to regulate cell fate. *Cell* **182**, 404–416.e14 (2020).

150Darnell, A. M., Subramaniam, A. R. & O'Shea, E. K. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Molecular Cell* **71**, 229–243.e11 (2018).

151Vladimer, G. I., Górna, M. W. & Superti-Furga, G. IFITs: Emerging roles as key anti-viral proteins. *Front Immunol* **5**, 94 (2014).

152Spriggs, K. A., Bushell, M. & Willis, A. E. Translational regulation of gene expression during conditions of cell stress. *Molecular Cell* **40**, 228–237 (2010).

153Andreev, D. E. *et al.* Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. *eLife* **4**, e03971 (2015).

154Sidrauski, C., McGeachy, A. M., Ingolia, N. T. & Walter, P. The small molecule ISRIB reverses the effects of eIF2α phosphorylation on translation and stress granule assembly. *eLife* **4**, e05033 (2015).

155Baird, T. D. *et al.* Selective mRNA translation during eIF2 phosphorylation induces expression of IBTKα. *Mol Biol Cell* **25**, 1686–1697 (2014).

156Wei, S. *et al.* Ribosome profiling reveals translatome remodeling in cancer cells in response to zinc oxide nanoparticles. *Aging* **13**, 23119–23132 (2021).

157Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**, 789–802 (2011).

158Jackson, R. & Standart, N. The awesome power of ribosome profiling. *RNA* **21**, 652–654 (2015).

159Matsuda, D. & Dreher, T. W. Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. *RNA* **12**, 1338–1349 (2006).

160Gu, Y., Mao, Y., Jia, L., Dong, L. & Qian, S.-B. Bi-directional ribosome scanning controls the stringency of start codon selection. *Nat Commun* **12**, 6604 (2021).

161Li, K., Kong, J., Zhang, S., Zhao, T. & Qian, W. *Distance-dependent inhibition of translation initiation by downstream out-of-frame AUGs reveals that ribosome scans in a brownian ratchet process*. http://biorxiv.org/lookup/doi/10.1101/2021.11.01.466764 (2021) doi:10.1101/2021.11.01.466764.

162Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* **9**, 5134–5142 (1989).

163Babendure, J. R., Babendure, J. L., Ding, J.-H. & Tsien, R. Y. Control of mammalian translation by mRNA structure near caps. *RNA* **12**, 851–861 (2006).

164Guan, B.-J. *et al.* A unique ISR program determines cellular responses to chronic stress. *Mol Cell* **68**, 885–900.e6 (2017).

165Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences* **21**, 267–271 (1996).

166Nilsson, O. B. *et al.* Cotranslational protein folding inside the ribosome exit tunnel. *Cell Reports* **12**, 1533–1540 (2015).

167Martinez, T. F. *et al.* Accurate annotation of human protein-coding small open reading frames. *Nature Chemical Biology* **16**, 458–468 (2020).

168 Hao, Y. *et al.* SmProt: A database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief Bioinform* **19**, 636–643 (2018).

169 Neumann, T. & Tuller, T. Modeling the ribosomal small subunit dynamic in saccharomyces cerevisiae based on TCP-seq data. *Nucleic Acids Research* gkac021 (2022) doi:10.1093/nar/gkac021.

170 Garshott, D. M. *et al.* iRQC, a surveillance pathway for 40S ribosomal quality control during mRNA translation initiation. *bioRxiv* 2021.04.20.440649 (2021) doi:10.1101/2021.04.20.440649.

171 Kearse, M. G. *et al.* Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev* **33**, 871–885 (2019).

172 Manjunath, H. *et al.* Suppression of Ribosomal Pausing by eIF5A Is Necessary to Maintain the Fidelity of Start Codon Selection. *Cell Reports* **29**, 3134–3146.e6 (2019).

173 Gaba, A., Wang, H., Fortune, T. & Qu, X. Smart-ORF: A single-molecule method for accessing ribosome dynamics in both upstream and main open reading frames. *Nucleic Acids Research* (2020) doi:10.1093/nar/gkaa1185.

174 Kozak, M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* **9**, 5073–5080 (1989).

175 Wagner, S. *et al.* Selective translation complex profiling reveals staged initiation and co-translational assembly of initiation factor complexes. *Mol Cell* **79**, 546–560.e7 (2020).

176 Bohlen, J., Fenzl, K., Kramer, G., Bukau, B. & Teleman, A. A. Selective 40S Footprinting Reveals Cap-Tethered Ribosome Scanning in Human Cells. *Molecular Cell* **79**, 561–574.e5 (2020).

177 Young, D. J., Meydan, S. & Guydosh, N. R. 40S ribosome profiling reveals distinct roles for Tma20/Tma22 (MCT-1/DENR) and Tma64 (eIF2D) in 40S subunit recycling. *Nat Commun* **12**, 2976 (2021).

178Wu, C. C.-C., Zinshteyn, B., Wehner, K. A. & Green, R. High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Molecular Cell* **73**, 959–970.e5 (2019).

179Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* **3**, e01257 (2014).

180Cao, J. & Geballe, A. P. Ribosomal release without peptidyl tRNA hydrolysis at translation termination in a eukaryotic system. *RNA* **4**, 181–188 (1998).

181Han, P. *et al.* Genome-wide survey of ribosome collision. *Cell Reports* **31**, 107610 (2020).

182Arpat, A. B. *et al.* Transcriptome-wide sites of collided ribosomes reveal principles of translational pausing. *Genome Res.* **30**, 985–999 (2020).

183Tuck, A. C. *et al.* Mammalian RNA decay pathways are highly specialized and widely linked to translation. *Molecular Cell* **77**, 1222–1236.e13 (2020).

184Rubio, A., Ghosh, S., Mülleder, M., Ralser, M. & Mata, J. Ribosome profiling reveals ribosome stalling on tryptophan codons and ribosome queuing upon oxidative stress in fission yeast. *Nucleic Acids Research* **49**, 383–399 (2021).

185Rooijers, K., Loayza-Puch, F., Nijtmans, L. G. & Agami, R. Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat Commun* **4**, 2886 (2013).

186Ramu, H. *et al.* Nascent peptide in the ribosome exit tunnel affects functional properties of the a-site of the peptidyl transferase center. *Mol Cell* **41**, 321–330 (2011).

187Dimitrova, L. N., Kuroha, K., Tatematsu, T. & Inada, T. Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. *J Biol Chem* **284**, 10343–10352 (2009).

188 Gamerdinger, M. *et al.* Early scanning of nascent polypeptides inside the ribosomal tunnel by NAC. *Molecular Cell* **75**, 996–1006.e8 (2019).

189 Weber, R. *et al.* 4EHP and GIGYF1/2 Mediate Translation-Coupled Messenger RNA Decay. *Cell Reports* **33**, 108262 (2020).

190 Law, G. L., Raney, A., Heusner, C. & Morris, D. R. Polyamine regulation of ribosome pausing at the upstream open reading frame of s-adenosylmethionine decarboxylase *. *Journal of Biological Chemistry* **276**, 38036–38043 (2001).

191 Young, S. K., Willy, J. A., Wu, C., Sachs, M. S. & Wek, R. C. Ribosome reinitiation directs gene-specific translation and regulates the integrated stress response. *J Biol Chem* **290**, 28257–28271 (2015).

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

193 Mohammad, M. P., Munzarová Pondělíčková, V., Zeman, J., Gunišová, S. & Valášek, L. S. In vivo evidence that eIF3 stays bound to ribosomes elongating and terminating on short upstream ORFs to promote reinitiation. *Nucleic Acids Res* **45**, 2658–2674 (2017).

194 Sanchez, M. *et al.* Cross talk between eIF2α and eEF2 phosphorylation pathways optimizes translational arrest in response to oxidative stress. *iScience* **20**, 466–480 (2019).

195 Shu, X. E., Mao, Y., Jia, L. & Qian, S.-B. Dynamic eIF3a o-GlcNAcylation controls translation reinitiation during nutrient stress. *Nat Chem Biol* (2021) doi:10.1038/s41589-021-00913-4.

196 Yan, L. L. & Zaher, H. S. Ribosome quality control antagonizes the activation of the integrated stress response on colliding ribosomes. *Molecular Cell* (2020) doi:10.1016/j.molcel.2020.11.033.

197 Wu, C. C.-C., Peterson, A., Zinshteyn, B., Regot, S. & Green, R. Ribosome Collisions Trigger General Stress Responses to Regulate Cell Fate. *Cell* **182**, 404–416.e14 (2020).

198 Ferrin, M. A. & Subramaniam, A. R. Kinetic modeling predicts a stimulatory role for ribosome collisions at elongation stall sites in bacteria. *eLife* **6**, e23629.

199Wallace, E. W. J. *et al.* Quantitative global studies reveal differential translational control by start codon context across the fungal kingdom. *Nucleic Acids Research* **48**, 2312–2331 (2020).

200Aliouat, A. *et al.* Divergent effects of translation termination factor eRF3A and nonsense-mediated mRNA decay factor UPF1 on the expression of uORF carrying mRNAs and ribosome protein genes. *RNA Biol* **17**, 227–239 (2019).

201Jia, L. *et al.* Decoding mRNA translatability and stability from the 5′ UTR. *Nat Struct Mol Biol* **27**, 814–821 (2020).

202Wang, H., Sun, L., Gaba, A. & Qu, X. An in vitro single-molecule assay for eukaryotic cap-dependent translation initiation kinetics. *Nucleic Acids Res* **48**, e6 (2020).

203Yu, J. *et al.* Dynamic m6A modification regulates local translation of mRNA in axons. *Nucleic Acids Research* **46**, 1412–1423 (2018).

204Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).

205Sambrook, J. *Molecular Cloning: A Laboratory Manual, Third Edition*. (Cold Spring Harbor Laboratory Press, 2001).

206Vassilenko, K. S., Alekhina, O. M., Dmitriev, S. E., Shatsky, I. N. & Spirin, A. S. Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. *Nucleic Acids Res* **39**, 5555–5567 (2011).

207Pichon, X. *et al.* Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *Journal of Cell Biology* **214**, 769–781 (2016).

208Cencic, R. *et al.* Reversing chemoresistance by small molecule inhibition of the translation initiation complex eIF4F. *PNAS* **108**, 1046–1051 (2011).

209Michel, A. M. *et al.* GWIPS-viz: Development of a ribo-seq genome browser. *Nucleic Acids Res* **42**, D859–864 (2014).

210Stephens, Z. D. *et al.* Big data: Astronomical or genomical? *PLoS Biol* **13**, e1002195 (2015).

211Faeder, J. R., Blinov, M. L. & Hlavacek, W. S. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol Biol* **500**, 113–167 (2009).

212Fisher, T. *et al.* Parsing the role of NSP1 in SARS-CoV-2 infection. *bioRxiv* 2022.03.14.484208 (2022) doi:10.1101/2022.03.14.484208.

213Mendez, A. S. *et al.* The n-terminal domain of SARS-CoV-2 nsp1 plays key roles in suppression of cellular gene expression and preservation of viral gene expression. *Cell Rep* **37**, 109841 (2021).

214Bujanic, L. *et al.* The key features of SARS-CoV-2 leader and NSP1 required for viral escape of NSP1-mediated repression. *RNA* **28**, 766–779 (2022).