

# Translational buffering by ribosome stalling in upstream open reading frames

Ty Bottorff<sup>1,2</sup>, Adam P. Geballe<sup>3,†</sup>, Arvind Rasi Subramaniam<sup>1,2,†</sup>

<sup>1</sup> Basic Sciences Division and Computational Biology Section of the Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

<sup>2</sup> Biological Physics, Structure and Design Graduate Program, University of Washington, Seattle, WA 98195, USA

<sup>3</sup> Human Biology and Clinical Research Divisions, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

† Corresponding authors: A.P.G: [ageballe@fredhutch.org](mailto:ageballe@fredhutch.org), A.R.S: [rasi@fredhutch.org](mailto:rasi@fredhutch.org)

## Abstract

Upstream open reading frames (uORFs) are present in over half of all human mRNAs. uORFs can potentially regulate the translation of downstream open reading frames by 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 performed systematic 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 translation 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 identify several human uORFs that repress main ORF translation via a similar terminal diproline motif. We propose that ribosome stalls in uORFs provide a general mechanism for buffering against reductions in main ORF translation during stress and developmental transitions.

## Introduction

About half of human mRNAs have at least one upstream open reading frame (uORF) in their 5' untranslated region<sup>1–3</sup>. Ribosome profiling studies estimate that at least twenty percent of these uORFs are actively translated<sup>4,5</sup>. 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<sup>6,7</sup>. Despite having poor initiation sequence contexts, many eukaryotic uORFs repress main ORF translation<sup>1,3,4,7–11</sup>. uORF mutations are implicated in several human diseases via changes to main ORF translation<sup>12,13</sup>. For example, uORF mutations in oncogenes and tumor suppressors can act as driver mutations in cancer<sup>14,15</sup>.

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 *GCN4* (*S. cerevisiae* homolog of human *ATF4*) mRNA render main ORF translation sensitive to cellular levels of the eIF2 $\alpha$ -GTP-tRNA<sub>Met</sub> ternary complex<sup>16,17</sup>. Although initiation rate usually limits translation,<sup>18–20</sup> inefficient elongation or termination on uORFs can regu-

late translation by barricading scanning ribosomes from reaching the main ORF<sup>21–26</sup>. Inefficient elongation can be driven by the nascent uORF peptide<sup>27,28</sup>, poorly translated codons in the uORF<sup>29,30</sup>, or small molecules such as amino acids or polyamines<sup>23,24</sup>. Further, interactions between scanning and elongating ribosomes on uORFs may cause dissociation of scanning ribosomes or enhanced initiation at start codons<sup>23,31,32</sup>.

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<sup>33</sup>, 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<sup>34</sup>. Quantitative modeling of uORF translation has been used to support the regulation re-initiation model for the *GCN4* mRNA<sup>35–37</sup>. A computational model predicted that elongating ribosomes can dislodge leading scanning ribosomes on uORFs and confer stress resistance to protein expression<sup>38</sup>. However, these models have not been compared against alternative models of uORF regulation that predict queuing or dissociation of scanning ribosomes that collide with paused elongating ribosomes<sup>21,23</sup>. A critical barrier for 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<sup>18,38</sup>, 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 cytomegalovirus *UL4* gene to test different kinetic models of uORF-mediated translational control<sup>21</sup>. The second uORF (uORF2 henceforth) in the *UL4* 5' UTR contains a terminal diproline motif that stalls 80S ribosomes by disrupting peptidyl transferase center activity<sup>27,28</sup>. 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<sup>39</sup>. Using this experimentally-integrated modeling approach, we find that the presence of 80S stalls in uORF2 of *UL4* 5' UTR confers resistance (called buffering henceforth) 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 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 translation against reduced ribosome loading across cellular and environmental transitions. Together, our results illustrate the value of experimentally-integrated kinetic modeling for comparisons of different uORF regulatory mechanisms and identification of novel experimental signatures from

complex molecular interactions.

## Results

### Models of uORF regulation of main ORF translation

We surveyed five previously proposed models of uORF regulation of main ORF translation (Fig. 1). We distinguished between these models using a combination of computational modeling and experimental reporter assays. In the constitutive repression model<sup>9</sup> (Fig. 1A), uORFs siphon away scanning ribosomes from the main ORF since re-initiation is usually infrequent<sup>40–43</sup>. In the 80S-hit dissociation model<sup>38</sup> (Fig. 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<sup>23</sup> (Fig. 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 rather than leaky scan past it. In the collision-mediated 40S dissociation model<sup>31,32</sup> (Fig. 1D), scanning ribosomes instead dissociate if they collide with a 3' stalled elongating ribosome.

Lastly, in the regulated re-initiation model<sup>16,44,45</sup> (Fig. 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 terminating at the first uORF, scanning ribosomes must reacquire a new eIF2 $\alpha$ -GTP-tRNA<sub>Met</sub> ternary complex (TC) before re-initiating. The time to reacquire a new TC correlates with the proportion of phosphorylated eIF2 $\alpha$ . Therefore, when cells are not stressed and the proportion of phosphorylated eIF2 $\alpha$  is lower, translation of the first uORF is followed by re-initiation at the second downstream uORF. Alternatively, when cells are stressed and the proportion of phosphorylated eIF2 $\alpha$  is higher, translation of the first uORF is instead followed by re-initiation at the main ORF.

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

To differentiate between proposed models of uORF regulation (Fig. 1), we used the well studied human cytomegaloviral *UL4* uORF2<sup>31</sup> as an experimental model. The 5' leader region preceding the *UL4* coding sequence contains three uORFs (Fig. 2A). uORF1 slightly reduces uORF2 repressiveness by siphoning scanning ribosomes away from uORF2, and uORF3 is irrelevant for repression<sup>31</sup>. uORF2 represses main ORF translation via a terminating ribosome stall that is dependent on the uORF2 peptide sequence<sup>21</sup> (Fig. 2A, irrelevant uORFs boxed in white, key uORF2 boxed in

green). The two C-terminal proline residues in uORF2 are necessary for the elongating ribosome stall. 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<sup>28</sup>. Reduced termination activity is further mediated through an interaction between the uORF2 nascent peptide and the GGQ motif within eRF1<sup>46</sup>. Even though the A-site of the uORF2-stalled ribosome is occupied by a stop codon, we will refer to it as an elongating ribosome stall since they are functionally equivalent for the purposes of this study. This terminology is also inclusive of elongation stalls within other uORFs<sup>22,23,47–51</sup>.

We inserted the uORF2-containing *UL4* leader sequence into a dual-luciferase reporter system (Fig. 2B) in which nanoluciferase (NLuc) signal provides a read-out of uORF2 repressiveness and firefly luciferase (FLuc) signal serves as normalization for transfection efficiency. We confirmed that uORF2 repressiveness depends on its translation and the terminal diproline-dependent elongating ribosome stall (Fig. 2C, top three rows compared to bottom wild-type). We used this *UL4*-based luciferase reporter to quantitatively dissect the kinetics of uORF-mediated translational regulation.

We complemented our experimental measurements with computational kinetic modeling of proposed models of uORF regulation (Fig. 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 rule-based framework for compact model specification<sup>52</sup>. We then expanded the model into the BioNetGen modeling language syntax<sup>53</sup> and inferred a reaction dependency graph for efficient simulation<sup>39</sup>. Next, we stochastically simulated the models using an agent-based Gillespie algorithm implemented in NFSim<sup>54</sup>. The molecules and reactions within the kinetic model are shown in Fig. S1A and Fig. S1B, respectively, and 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 constitutes our platform for differentiating between proposed uORF regulatory models.

While many rates of steps of translation have been estimated using single-molecule or ribosome profiling methods<sup>55–58</sup>, several critical parameters specific to *UL4* uORF2 have not been estimated. Therefore, we first calibrated our computational models to our reporter measurements on wild-type or mutant uORF2 (Fig. 2C) to derive estimates for these remaining, unknown pa-

rameters (marked with “This work” in Table 1). We used previously generated estimates for kinetic parameters not directly measured in our work (Table 1). We did not fit the constitutive repression and regulated re-initiation models (Fig. 1A,E) to our reporter measurements (Fig. 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 transcripts.

Simulations of the queuing-mediated enhanced repression (Fig. 1C) and collision-mediated 40S dissociation (Fig. 1D) models readily recapitulate measurements on NLuc protein output in wild-type and mutant *UL4* reporters (Fig. 2D, triangles and squares). The 80S-hit dissociation model (Fig. 1B), modified to include an elongating ribosome stall within the uORF, also recapitulates the reporter measurements (Fig. 2D, circles). However, this modified 80S-hit dissociation model requires the difference between the stronger Kozak and wild-type uORF initiation 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. 1B-D) are in line with literature estimates<sup>56–58</sup>. Although we do not measure how often terminating ribosomes at uORF2 re-initiate, the fractions derived here (50-70%, Table 1) are within the range of measured re-initiation fractions across mRNAs with different sequence features<sup>40–43</sup>. A complete description of the derivation of model parameters marked with a reference of “This work” can be found in the methods section. Given our computational recapitulation of experimental data, we then used our computational modeling platform to predict how translation would be perturbed following variations in other kinetic parameters.

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

While many kinetic parameters could be varied to help distinguish between proposed models of uORF regulation (Fig. 1), we honed in on the rate of ribosome loading onto the mRNA for two key reasons. Firstly, this rate is reduced 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 eIF2 $\alpha$  kinases (GCN2, PKR, PERK, and HRI) to reduce the concentration of eIF2 $\alpha$ -containing ternary complexes (TCs)<sup>59–61</sup>. A reduction in the concentration of eIF2 $\alpha$ -containing TCs reduces the rate of ribosome loading. Viral infection also leads to reduced ribosome loading via interferon-induced proteins with tetratricopeptide repeats (IFITs)<sup>62</sup>. Cellular



stress also reduces ribosome loading via inhibition of mTOR and sequestration of eIF4E by hypophosphorylated 4EBP<sup>63</sup>. Secondly, translated, repressive uORFs are enriched in transcripts buffered against reduced ribosome loading<sup>64–67</sup>. 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. 1), we investigated what uORF parameters combinations, if any, allow buffering against reduced ribosome loading rates.

We use the term ‘buffer’ to describe the observation of main ORF translation decreasing less than expected, or even increasing, with reduced ribosome loading in comparison to the constitutive repression model (Fig. 1A). Importantly, 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 main ORF translation reduced to a lower extent than the constitutive repression model when ribosome loading is reduced. Preferred translation refers to increased main ORF translation when ribosome loading is reduced. The constitutive repression model (Fig. 1A) has no buffering (Fig. 3A) since its repression is independent of the ribosome loading rate.

The 80S-hit dissociation model (Fig. 1B) displays buffering (Fig. 3B, left panel, yellow-green line) in agreement with previous work<sup>38</sup>. This behavior arises because the number of 5′ elongating ribosomes that collide with scanning ribosomes correlates with the ribosome loading rate. However, buffering depends on strongly initiating uORFs, minimal re-initiation, and longer uORFs (Fig. 3B, left panel, yellow-green line, Fig. S2A) as observed previously<sup>38</sup>. 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. Nevertheless, most eukaryotic uORFs only weakly initiate translation and are short<sup>1,3,4,8–11,31</sup>. *UL4* uORF2 is 22 codons long, and we estimate re-initiation to be frequent (Table 1). Accordingly, buffering is no longer observed (Fig. S2B) in this model when parameters derived from control *UL4* variants (Table 1) are used.

The queuing-mediated enhanced repression model<sup>23</sup> (Fig. 1C) displays buffering behavior (Fig. 3C, 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 uORF2 (Fig. S2C, left panel). Unlike the 80S-hit dissociation model (Fig. 1B), weakly initiating uORFs, such as *UL4* uORF2, still confer buffering in this model (Fig. 3C, 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. S2C and Fig. 3C, left vs. right panels). This sensitivity arises because  $d_{stall}$  determines if the P-site of a queued scanning ribosome is correctly positioned over the uORF start codon to productively increase uORF initiation (Fig. S2C, left panel). In the idealized case of homogeneously sized ribosomes (30 nt footprints<sup>55,68</sup>) 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<sup>41,69–71</sup> occurs with a high rate (Fig. S2D, middle panel). However, even a high rate of backward scanning is insufficient to compensate for misalignment between the uORF start codon and the P-site of the queued scanning ribosome (Fig. S2D, middle panel). To simplify our modeling interpretation, we considered *UL4* uORF2, that 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 on the start codon.

The collision-mediated 40S dissociation model (Fig. 1D) displays buffering (Fig. 3D, right panel, purple line) because the number of scanning ribosomes that collide with 3′ stalled elongating ribosomes is dependent upon the rate of ribosome loading. Buffering in this model requires the collision-induced 40S dissociation rate to be somewhat fast (Fig. 3D, right vs. left panels, Fig. S2E, teal and yellow-green lines). If this rate is too low (for example, 0 in Fig. 3D, left panel), this model reduces to the queuing-mediated enhanced repression model (Fig. 1C). However, unlike the queuing model, the collision-mediated 40S dissociation model is not sensitive to the distance between the stall and the start codon (Fig. S2F, purple lines). As in the queuing model (Fig. 1C), weakly initiating uORFs, such as *UL4* uORF2, can still confer buffering (Fig. 3D, right panel, purple line) in the collision-mediated 40S dissociation model (Fig. 1D). This effect arises because, unlike in the 80S-hit dissociation model (Fig. 1B), the elongation stall is now rate limiting for main ORF translation. Therefore, the mere presence of an elongation stall within the queuing-mediated

enhanced repression and collision-mediated 40S dissociation models imparts some degree of buffering.

In the regulated re-initiation model (Fig. 1E), buffering is observed (Fig. 3E, right panel, yellow-green line) because termination at the first uORF 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. 1E) depends upon the initiation efficiency and continued scanning fractions of the two uORFs (Fig. 3E). 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 concentration biases towards initiation at the second downstream uORF (Fig. S3A). Reductions from high ternary complex concentrations biases towards main ORF initiation; therefore main ORF translation can increase with decreased ribosome loading.

As such, our computational results provide the first systematic comparison of different mechanisms of uORF-mediated regulation (Fig. 1) and enable their comparison with experimental measurements below.

### ***UL4* uORF2 buffers against reductions in main ORF translation from reduced ribosome loading in an elongating ribosome stall-dependent manner**

We next tested whether the computational predictions of uORF-mediated buffering (Fig. 3) can be observed experimentally with *UL4* uORF2. To this end, we experimentally varied the rate of ribosome loading and measured effects on main ORF translation using our reporter system (Fig. 2B). Since no-stall uORF2 variants have similar protein expression to the no-start uORF2 variants (Fig. 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 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 that reduce the rate of 43S-cap binding without affecting mRNA stability (Fig. 4A)<sup>72,73</sup>. We varied the degree to which ribosome loading is reduced by altering the GC content of the stem loops. We observe that NLuc translation decreases less with reduced ribosome loading for the wild-type *UL4* reporter in comparison to the no-stall *UL4* variant (Fig. 4A, left panel, yellow vs. gray circles). Therefore, wild-type uORF2 imparts resistance to reductions in NLuc translation from stem loop-mediated reduced ribosome loading. When the wild-type data are normalized by the no-stall data, NLuc translation neg-

atively correlates with ribosome loading, indicative of buffering against reduced ribosome loading by wild-type uORF2 (Fig. 4A, right panel).

We also reduced ribosome loading with the drug 4E1RCat, which disrupts the interaction between the 5' cap-binding eIF4E and the scaffold eIF4G (Fig. 4B)<sup>74</sup>. We added a PEST tag<sup>75</sup> to increase the turnover of the NLuc protein to more accurately measure changes in main ORF translation during drug treatment. NLuc translation of wild-type *UL4* reporter decreases less in comparison to the no-stall control upon 4E1RCat treatment (Fig. 4B, 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 general buffering against reduced ribosome loading by wild-type uORF2 (Fig. 4B, right panel).

Finally, we added a short, synthetic uORF, 5' to the *UL4* uORF2, that siphons scanning ribosomes away from uORF2 (Fig. 4C). 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 in the wild-type *UL4* reporter as scanning ribosomes are increasingly siphoned off by improving the Kozak context of the synthetic uORF (Fig. 4C, left panel, yellow circles). While resistance is observed with the other strategies of reduced ribosome loading (Fig. 4A-B, left panels), preferred translation is observed here (Fig. 4C, left panel), perhaps because ribosome loading is reduced only within the transcript rather than at the 5' cap. However, 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 general buffering against reduced ribosome loading by wild-type uORF2 (Fig. 4C, right panel).

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

Given our experimental data of uORF2-mediated buffering of *UL4* reporters (Fig. 4), we narrowed our focus from the five surveyed models (Fig. 1) to the two (Fig. 1C-D) most relevant for *UL4* uORF2: the queuing-mediated enhanced repression (Fig. 1C) and collision-mediated 40S dissociation models (Fig. 1D). These two models are computationally predicted to confer buffering in an elongating ribosome stall-dependent manner without needing multiple uORFs (Fig. 3C-D). To differentiate between these models, we turned to our computational modeling prediction that, only in the queuing-mediated enhanced repression model (Fig. 1C), main

ORF protein output is sensitive to the distance between the uORF start codon and elongating ribosome stall (Fig. 3C). Our computational modeling of the queuing-mediated enhanced repression model (Fig. 5A, 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. 1D) predicts a much lower effect of  $d_{stall}$  on uORF repressiveness (Fig. 5A, left panel, purple line). The residual effect of  $d_{stall}$  on uORF repressiveness (Fig. 5A, left panel, purple line) in the collision-mediated 40S dissociation model (Fig. 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. 5A). However, backward scanning occurring as fast as forward scanning ( $\sim 5$  nt/s) is required to abolish the periodicity in the queuing model (Fig. 5A, right panel, yellow-green line). While there are estimates of how far ribosomes can backward scan<sup>69–71</sup>, 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 nt/s here) given the 5′-3′ directionality of scanning. Slower backward scanning ( $\sim 3.75$  nt/s) is sufficient to abolish periodicity in the collision-mediated 40S dissociation model (Fig. 5A, 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 stall of *UL4* uORF by adding codons to the 5′ end of uORF2. We observe less than 2-fold changes in translational regulation (Fig. 5B, 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. 5A, left panel). Our longest uORF mutant here (96 nt from start to stall in Fig. 5B) is less repressive, but this effect may be due to decreased elongating ribosome stall strength. In this case, the 32 amino acid-long nascent peptide can extend out of the exit tunnel and be bound by additional factors<sup>27,28</sup>. Thus, our experimental data does not match computational predictions of main ORF translation sensitivity to  $d_{stall}$  in the queuing-mediated enhanced repression model (Fig.

1C), and better supports the collision-mediated 40S dissociation model (Fig. 1D).

In the queuing-mediated enhanced repression model (Fig. 1C), buffering is uniquely predicted to be sensitive to the distance between the uORF start codon and elongating ribosome stall (Fig. 3C). 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. 4C), we observe that a 6 nt longer  $d_{stall}$  uORF still buffers against reduced ribosome loading (Fig. 5C, top two rows compared to bottom two rows). Since backward scanning of 15–17 nt has been observed<sup>69–71</sup>, 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. S2D, right panel). Thus, our experimental data does not match computational predictions of buffering sensitivity to  $d_{stall}$  in the queuing-mediated enhanced repression model (Fig. 1C), but is consistent with the collision-mediated 40S dissociation model (Fig. 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<sup>6</sup>, a database integrated from *de novo* transcriptome assembly and ribosome profiling with 3,577 uORFs<sup>76</sup>, and a database of proteins less than 100 residues in size derived from literature mining, ribosome profiling, and mass spectrometry with 1,080 uORFs<sup>77</sup>. We identified several human transcripts with uORFs with terminal diproline motifs: *C1orf43*, *C15orf59*, *TOR1AIP1*, and *ABCB9* (Fig. 6). We replaced *UL4* uORF2 in our reporter (Fig. 2B) with these human uORFs. We created terminal proline to alanine and no-start-codon mutants and measured the effects of these mutations on NLuc expression relative to the wild-type human uORFs. While many of the tested uORFs are repressive (Fig. 6, yellow vs. blue), unlike the human cytomegaloviral uORF2, these human uORFs still generally repress NLuc expression without their terminal diproline motif (Fig. 6, gray vs. blue), indicating additional contributions to translational repression from siphoning of scanning ribosomes at the start codon and other residues in the nascent peptide.



## 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 elongating ribosome stalls in *UL4* uORF2 buffer against reductions in main translation due to reduced ribosome loading (Fig. 4). 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. 1). Thus, we can predict which models of uORF regulation of translation allow buffering and which parameters are key for buffering in each model (Fig. 3). While uORFs are enriched in stress-resistant transcripts, not all uORFs provide buffering<sup>64</sup>. 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. 1A) is not predicted to provide buffering (Fig. 3C)<sup>64–67</sup>. Instead, we find that 5' UTRs containing one (or some combination) of the following enable buffering main ORF translation: scanning ribosome dissociation due to 80S hits from the 5' (Fig. 1B), a single uORF with an elongating ribosome stall (Fig. 1C-D), or multiple uORFs acting through the regulated re-initiation model (Fig. 1E).

Long, well-initiating uORFs that do not re-initiate well allow buffering (Fig. 3B, left panel, yellow-green line, Fig. S2A, yellow-green line) in the 80S-hit model (Fig. 1B), but these requirements are at odds with the typically short and poorly initiating nature of known uORFs<sup>1,3,4,8–11</sup>. Consequently, 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. S2B).

Our modeling of the regulated re-initiation model (Fig. 1E) agrees (Fig. 3E and Fig. S3A) with previous work<sup>35,36</sup> that buffering requires 1) two well-translated uORFs, and 2) frequent and rare continued scanning after termination at uORFs 1 and 2, respectively. The 30% of human transcripts that contain multiple uORFs can display buffering by the regulated re-initiation model; however, about 25% of human transcripts only have one uORF<sup>2</sup> and are unable to provide buffering via this model.

We narrowed our focus to the two models (Fig. 1C-D) that are most pertinent to *UL4* uORF2. Both the queuing-mediated enhanced repression (Fig. 1C) and collision-mediated 40S dissociation (Fig. 1D) models are predicted (Fig. 3C-D) to allow buffering with weakly initiating uORFs and elongating ribosome stalls. Both of these models include an elongating ribosome stall and only require a single uORF for buffering (Fig. 3C-D). Computational modeling not only predicted this buffering behavior but also allows us to differentiate between these two models. We predict that the queuing-mediated enhanced repression model (Fig. 1C) is uniquely sensitive to the distance between the uORF start codon and elongating ribosome stall (Fig. 5A, yellow-green line, Fig. 3C, purple lines, Fig. S2F, purple lines). We experimentally vary this distance and do not find any systematic changes to either main ORF protein output (Fig. 5B) or buffering (Fig. 5C). 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.

While we have no direct evidence of scanning ribosome dissociation, the requirement of crosslinking to retain scanning ribosomes during ribo-seq preparations<sup>78,79</sup> hints at their labile nature and potential dissociation *in vivo*. This dissociation could serve to maintain the free pool of 40S ribosomal subunits while still allowing regulation of main ORF translation. Collisions between scanning ribosomes and their subsequent dissociation have also recently been proposed in a model of initiation RQC<sup>80</sup>. 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. 1C)<sup>23</sup>, this model might be relevant in other scenarios. Translation from non-cognate start codons is resistant to cycloheximide, perhaps due to queuing-mediated enhanced initiation, but sensitive to reductions in ribosome loading<sup>81</sup>. Loss of eIF5A, which helps paused elongating ribosomes continue elongation, increases 5' UTR translation in 10% of studied genes in human cells, augmented by downstream in-frame pause sites within 67 codons, perhaps also through queuing-mediated enhanced initiation<sup>82</sup>. There is also evidence of queuing-enhanced uORF initiation in the 23 nt long *Neurospora crassa* arginine attenuator peptide<sup>83</sup> as well as in transcripts with secondary structure near and 3' to start codons<sup>84</sup>. Additional sequence elements in the mRNA might determine whether scanning ribosome collisions

result in queuing or dissociation.

*UL4* uORF2 is not unique in containing an elongating ribosome stall<sup>22,23,26,47–51</sup>. There are a variety of residues that may reduce the rate of elongation, either through changes in the activity of the peptidyl transferase, tRNA availabilities, or interactions between the nascent peptide and the ribosome<sup>85,86</sup>. uORFs are often short<sup>3</sup> and may therefore be better poised to stall ribosomes; the nascent peptides may be less likely to be released due to interactions with the ribosome. Thus, a key role for elongation ribosome stalls in uORFs might be to enable buffering. While very few uORFs have been mechanistically characterized<sup>1</sup>, other elongating ribosome stall-containing uORFs, such as the peptide sequence sensitive uORFs that regulate human methionine synthase<sup>47</sup> and antizyme inhibitor I<sup>23</sup>, might enable buffering. Conversely, uORFs in several single uORF transcripts known to buffer against stress, such as *SLC35A4*, *C19orf48*, and *IFRD1*<sup>64</sup>, 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<sup>87–89</sup>. It will be also interesting to consider the effect of cellular stress-reduced elongation rates<sup>90</sup> and increased re-initiation<sup>91</sup>, both of which might regulate uORF-mediated buffering, as well as elongating ribosome dissociation through known quality control pathways<sup>39,80,92–96</sup>.

## Author Contributions

T.A.B. designed research, performed computations and experiments, analyzed data, and wrote the manuscript. A.P.G. conceived the project, designed research, and wrote the manuscript. A.R.S. conceived the project, designed research, performed computations and experiments, analyzed data, wrote the manuscript, supervised the project, and acquired funding.

## Acknowledgements

We thank members of the Subramaniam lab, the Basic Sciences Division, and the Computational Biology Program at Fred Hutch for assistance with the project and discussions and feedback on the manuscript. The computations described here were performed on the Fred Hutch Cancer Research Center computing cluster. This

research was funded by NIH R21 AI156152 received by APG and NIH R35 GM119835, NSF MCB 1846521, and the Sidney Kimmel Scholarship received by ARS. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant Numbers NSF DGE-1762114 and DGE-2140004. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research was supported by the Genomics Shared Resource of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704) and Fred Hutch Scientific Computing (NIH grants S10-OD-020069 and S10-OD-028685). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Competing interests

None

## Data and Code Availability

All experimental data and programming code associated with this manuscript is available at [https://github.com/rasilab/bottorff\\_2022](https://github.com/rasilab/bottorff_2022).

## 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 nanoluciferase and firefly luciferase. NanoLuc expression is driven by a CMV promoter. Firefly luciferase 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 a construct gifted 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<sup>97</sup>. The stem loop<sup>73</sup> 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<sup>97</sup>. 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<sup>98</sup>. Table S1 lists the plasmids described in this study. Key plasmid maps are available at [https://github.com/rasilab/bottorff\\_2022](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, 24, or 12 well plates. If the plasmids were not transfected into a 96 well plate, then the cells were resuspended in 100  $\mu$ L media. Then, 20  $\mu$ L of these resuspended cells were added per well to a 96 well plate for the dual-luciferase assay. If the transfection was already in a 96 well plate, the ~110  $\mu$ L media was removed and replaced with 20  $\mu$ L media per well. The Promega dual-luciferase kit was used. Cells were lysed with 20  $\mu$ L ONE-Glo EX Luciferase Assay Reagent per well for three minutes to measure firefly (*Photinus pyralis*) luciferase activity. Then, 20  $\mu$ L NanoDLR Stop & Glo Reagent was added per well for 10 minutes to quench the firefly luciferase signal and provide the furimazine substrate needed to measure NanoLuc luciferase activity. Firefly luciferase activity serves as an internal control for transfection efficiency, and NanoLuc activity provides a readout of 5' UTR regulation of NanoLuc translation.

## Kinetic modeling

We specify our kinetic models using the PySB interface<sup>52</sup> to the BioNetGen modeling language<sup>53</sup> (Fig. S1). 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<sup>54</sup>.

## Molecules

Our kinetic models of eukaryotic translational control describe the interactions between 2 molecule types: mRNA and ribosome (composed of separate large and small subunits). Here, we describe these molecules' components, states, and binding partners (Fig. S1A).

mRNA molecules have the following components: 5' end and codon sites ( $c_i$ ). The mRNA 5' end can either be free of or occupied with a ribosome. The mRNA 5' end must be free 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 bonding sites for the ribosome A site. Ribosomes, both scanning and elongating, have the following components: A site ( $a$ ), 5' side ( $t$  for trailing), and 3' side ( $l$  for leading). These sites serve as bonding sites for either the mRNA (A site) or other ribosomes during collisions (5' or 3' side). Both scanning and elongating ribosomes have mRNA footprints of 10 codons in our simulations based on mammalian ribosome profiling data<sup>55,68</sup>.

## Reactions

We describe here each type of kinetic reaction in our models of eukaryotic translational control (Fig. S1B). We use a syntax similar to that of BioNetGen<sup>53</sup> to illustrate the kinetic reactions. We scale ternary complex (TC, 100) and ribosome subunit numbers (100 each) to the single mRNA present in the simulation. Simulation of a single mRNA is sufficient to infer 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<sup>55,68</sup>. Therefore, PIC loading can occur when the 5' most 30 nucleotides (nt) of the mRNA are not bound to any ribosomes. 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<sup>56</sup>. 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 nt/s following an estimate in a mammalian cell-free translation system<sup>99</sup>, and previous computational study<sup>38</sup>. 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<sup>55-58,100</sup>. 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 based on toeprinting assays over time with various inhibitory drugs<sup>101</sup>.

## 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<sup>55,102</sup>. 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 result in bonds 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 dis-

sociation 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.

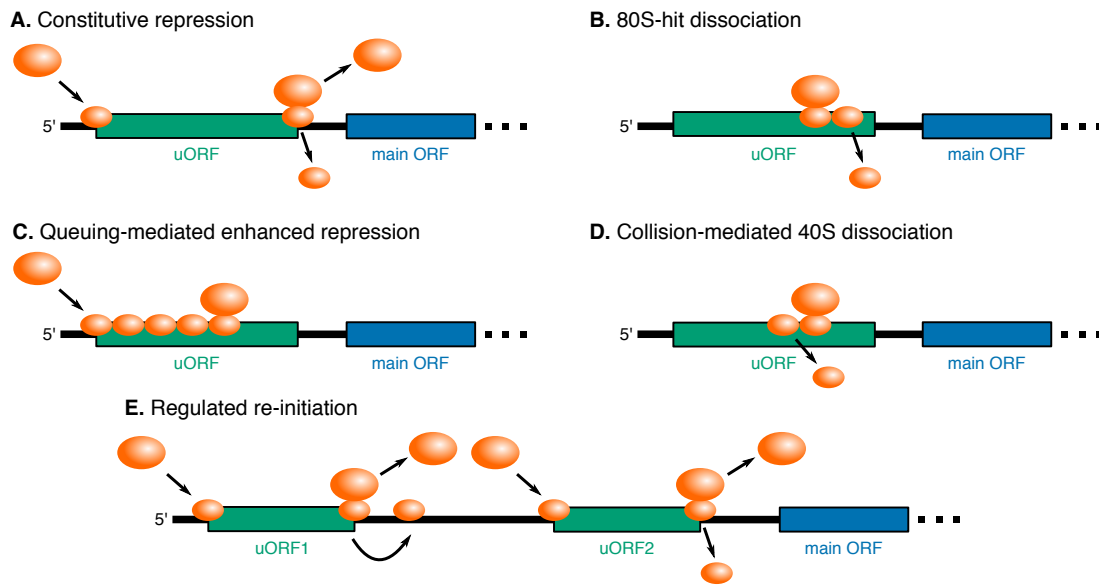
## Model calibration to reporter measurements

We derive the  $k_{cap\ bind}$  rates by spline interpolation of computationally modeled protein output fit to experimental data (Fig. 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<sup>6,76,77</sup>. The SmProt database<sup>77</sup> 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><sup>6</sup>. 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%2F541589-019-0425-0/MediaObjects/41589\\_2019\\_425\\_MOESM3\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1038%2F541589-019-0425-0/MediaObjects/41589_2019_425_MOESM3_ESM.xlsx)<sup>76</sup>. This database includes 3577 uORFs which we filter down to 3 that end in diproline motifs and that are less than 60 residues in length for ease of cloning, including *ABCB9*, *C15orf59*, and *PPP1R37*.

**Figure 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.

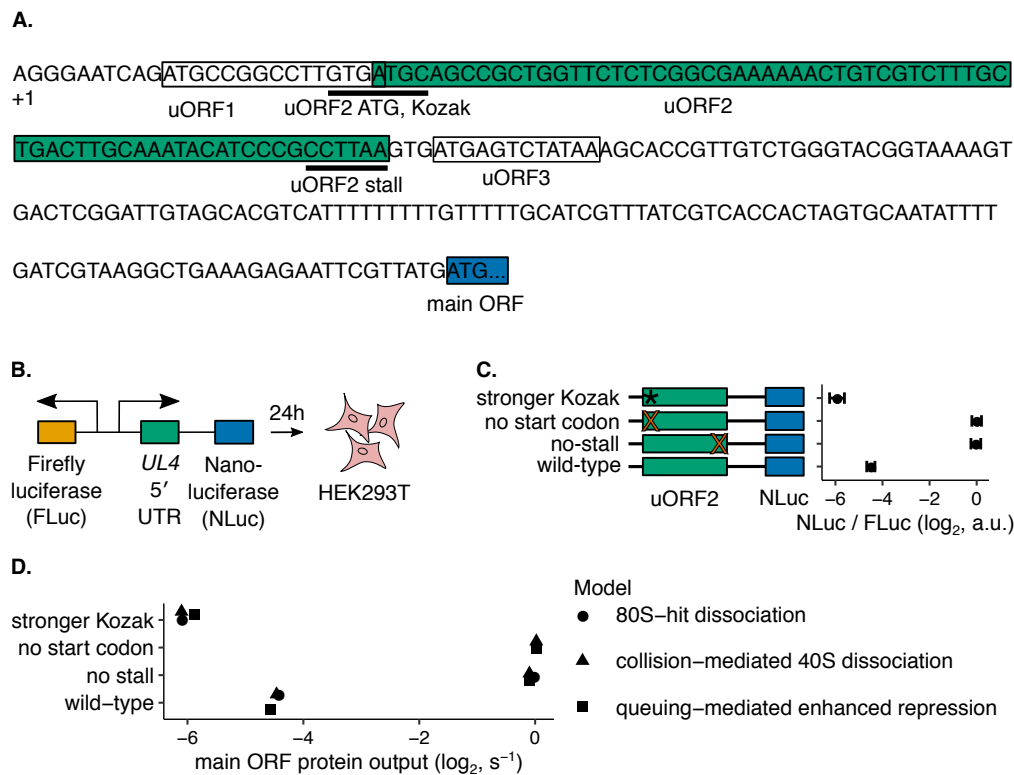
**(C) Queuing-mediated enhanced repression.** Scanning or elongating ribosomes form a queue behind a 3' stalled elongating ribosome. If the queue correctly positions a scanning ribosome at the uORF start codon, then the proportion of scanning ribosomes that initiate translation at the uORF increases.

**(D) Collision-mediated 40S dissociation.** Scanning ribosomes that collide with a 3' stalled elongating ribosome dissociate from the mRNA.

**(E) Regulated re-initiation.** Ribosomes initiate translation at the first uORF, and scanning continues after termination. Ribosomes re-initiate at the main ORF or the second downstream uORF when phosphorylated eIF2 $\alpha$  levels are high or low, respectively. The schematic is depicted in a low phosphorylated eIF2 $\alpha$  state.



**Figure 2**



**An experimental and computational platform for assessing uORF-mediated regulation of main ORF translation.**

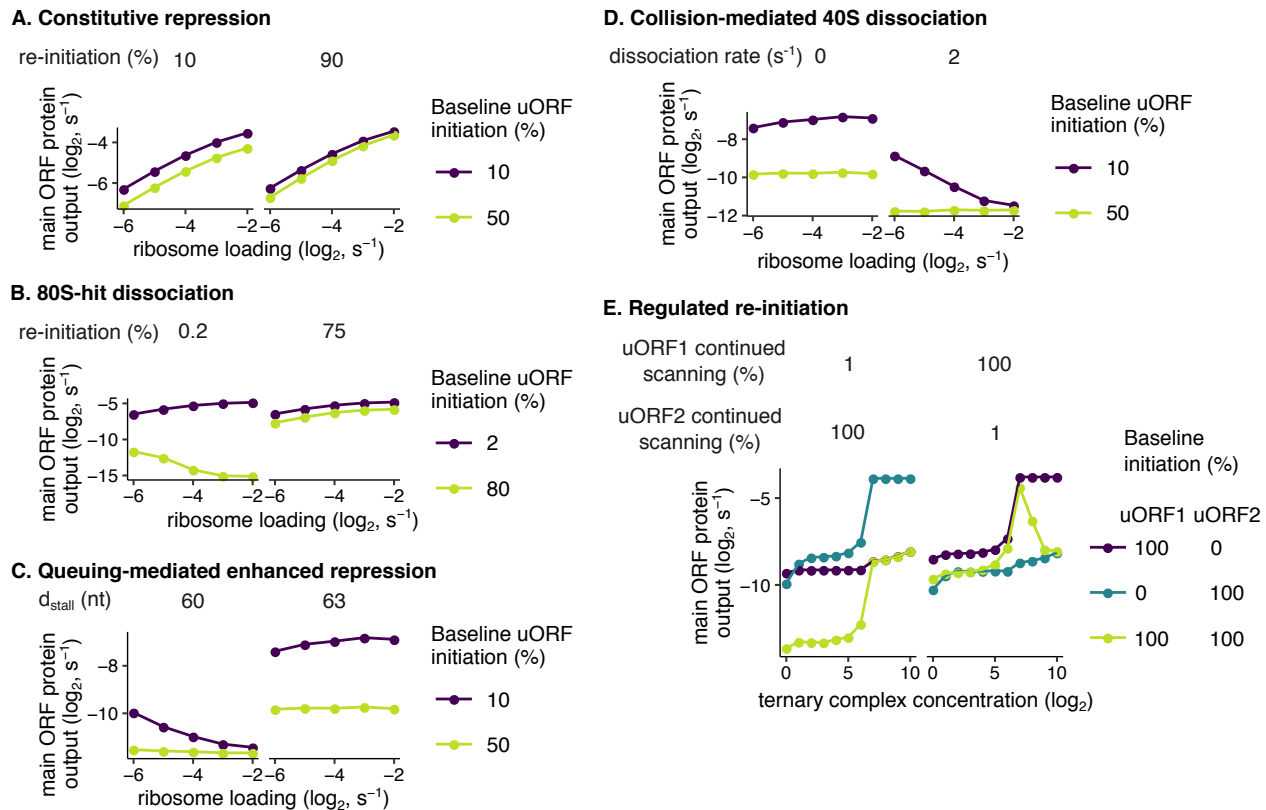
**(A)** The 236 nt 5' UTR of *UL4* mRNA from human cytomegalovirus contains 3 uORFs. The terminal proline and stop codons of uORF2 at which the P- and A-sites of the stalled ribosome are positioned is highlighted as uORF2 stall.

**(B)** A dual-luciferase reporter system measures 5' UTR repressiveness in HEK293T cells. Firefly luciferase signal serves as an internal control for transient transfection efficiency.

**(C)** The reporter system recapitulates the known elongating ribosome stall-dependent repression of translation by the *UL4* uORF<sup>21</sup>. 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 to alanine codon (GCT instead of CCT). Error bars show standard error of mean NLuc / FLuc ratios over 3 technical replicates. Data are normalized to a 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 a no-uORF start codon control. The parameter combination that best recapitulated the control behavior in Fig. 2C is displayed in Table 1. Error bars of simulated data are smaller than data points.

**Figure 3**



### Kinetic modeling predicts translational buffering by uORFs.

Buffering refers to a less than expected decrease (small positive slope), or even increase (negative slope), in main ORF translation with reduced ribosome loading.

**(A)** The constitutive repression model, without an elongating ribosome stall, has no buffering behavior. In this model, buffering does not occur at any uORF initiation or re-initiation rates. 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<sup>38</sup>. 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. The uORF is 100 residues long. Buffering occurs better for longer uORFs that have more time for elongating ribosomes to hit 3' scanning ribosomes (Fig. S2A, yellow-green line). 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 nt/s, and the elongation rate is 2 codons/s. There is no elongating ribosome stall in this model.

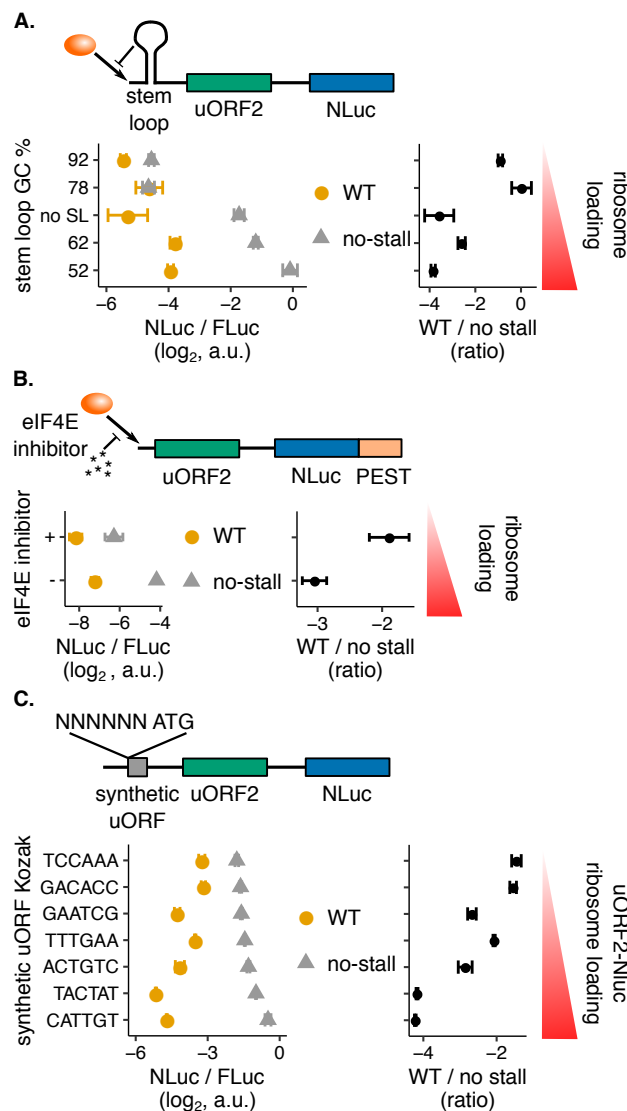
**(C)** Buffering in the queuing-mediated enhanced repression model depends on  $d_{stall}$ : the distance between the uORF start codon and elongating ribosome stall. In this model, uORF initiation can increase above baseline with increased ribosome loading when the distance between the uORF start codon and elongating ribosome 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. 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 residues 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. 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 eIF2 $\alpha$  phosphorylation mechanism; we instead vary the number of ternary complexes in this model. Error bars of simulated data are smaller than data points.

**Figure 4**



**The human cytomegaloviral uORF2 buffers against reductions in main ORF translation.**

The human cytomegaloviral *UL4* uORF2 is used in the dual-luciferase assay (Fig. 2B) in conjunction with various mechanisms to reduce ribosome loading.

**(A)** Ribosome loading is reduced using stem loops<sup>73</sup> with the indicated GC percentages. The -30 kcal/mol stem loops are positioned 8 bp from the 5' cap. The no-stem loop data has a 5' UTR 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.

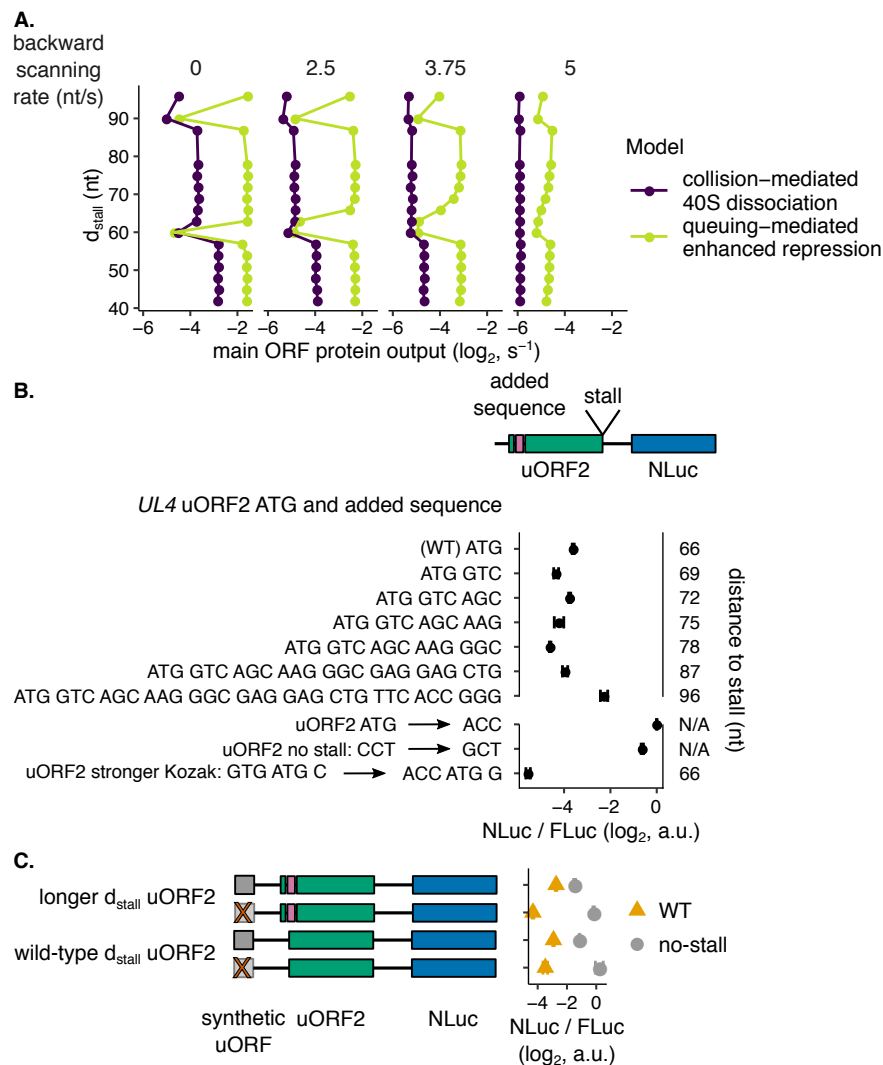
**(B)** Ribosome loading is reduced using the drug 4E1RCat (116  $\mu$ M)<sup>74</sup> that disrupts the interaction between eIF4E and eIF4G. NanoLuc has a C-terminal PEST tag to increase protein turnover<sup>75</sup> for the 3-hour drug treatment. The 5' UTR is 236 nt long. Data are normalized to a no-uORF start codon control without a PEST tag.

**(C)** Ribosome loading onto the uORF2-NanoLuc 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 data 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.

Right panels in A,B, 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). Error bars show standard error of mean NLuc / FLuc ratios over 3 technical replicates.



**Figure 5**



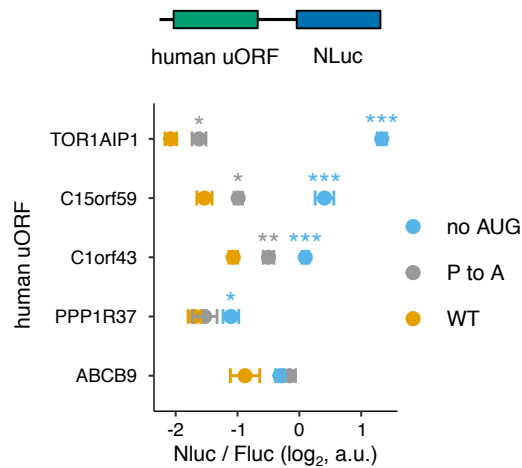
**Changes to the distance between the human cytomegaloviral uORF2 start codon and elongating ribosome stall do not change repressiveness or buffering as predicted by computational modeling of the collision-mediated 40S dissociation model.**

**(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 in main ORF translation across varied  $d_{stall}$ .  $d_{stall}$  refers to the distance between the start codon and 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. 2C, Table 1) are used here. The forward scanning rate is 5 nt/s. Data are normalized to a no-uORF start codon control. Error bars of simulated data are smaller than data points.

**(B)** Experimentally varying the distance between the human cytomegaloviral uORF2 start codon and elongating ribosome stall does not systematically affect its repression of main ORF translation. The human cytomegaloviral UL4 uORF2 is used in the dual-luciferase assay (Fig. 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 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 standard error of mean NLuc / FLuc ratios over 3 technical replicates. Data are normalized to a 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 translation. 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 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 6**



**Several human uORFs have repressive terminal diproline motifs.**

Terminal diproline motif-containing human uORFs are used in the dual-luciferase assay (Fig. 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 shown after running a two sample t-test: N.S. ( $P > 0.05$ ), \* ( $0.01 < P < 0.05$ ), \*\* ( $0.001 < P < 0.01$ ), \*\*\* ( $P < 0.001$ ). All comparisons with no shown significances are N.S. Data are normalized to a no-*UL4*-uORF2 start codon control.

## Figure S1

### A. Molecules in the kinetic model

Molecule	Component	State values or bond partners
Small ribosomal subunit ( $R_{40S}$ )	A-site ( $a$ )	bond to mRNA codon at position $i$ ( $c_i$ )
	5' side ( $\dagger$ )	bond to <i>trailing</i> ribosome
	3' side ( $\dagger$ )	bond to <i>leading</i> ribosome
	inter-subunit contact ( $isbi$ )	bond to large ribosomal subunit
Large ribosomal subunit ( $R_{60S}$ )	inter-subunit contact ( $isbi$ )	bond to small ribosomal subunit
Ternary complex ( $TC$ )	$R_{40S}$ contact	bond to $R_{40S}$
mRNA ( $M$ )	initiation footprint ( $start$ )	<i>clear</i> / <i>blocked</i>
	codon at position $i$ ( $c_i$ )	bond to ribosome A-site ( $a$ )

### B. Reactions in the kinetic model

$PIC_{43S}$  formation:  $R_{40S}(a) + TC \rightarrow PIC_{43S}(a)$   
 $PIC_{43S}$  formation (re-initiating ribosomes):  $R_{40S}(a!1) \circ M(c_i!1) + TC \rightarrow PIC_{43S}(a!1) \circ M(c_i!1)$   
 $PIC_{43S}$  loading:  $PIC_{43S}(a) + M(start \sim clear, c_i) \rightarrow PIC_{43S}(a!1) \circ M(start \sim blocked, c_i!1)$   
 Scanning:  $PIC_{43S}(a!1) \circ M(c_i!1) \rightarrow PIC_{43S}(a!1) \circ M(c_{i+1}!1)$   
 Backward scanning:  $PIC_{43S}(a!1) \circ M(c_i!1) \rightarrow PIC_{43S}(a!1) \circ M(c_{i-1}!1)$   
 Scanning (re-initiating ribosomes):  $R_{40S}(a!1) \circ M(c_i!1) \rightarrow R_{40S}(a!1) \circ M(c_{i+1}!1)$   
 Backward scanning (re-initiating ribosomes):  $R_{40S}(a!1) \circ M(c_i!1) \rightarrow R_{40S}(a!1) \circ M(c_{i-1}!1)$   
 Start codon selection:  $PIC_{43S}(a!1, isbi) \circ M(c_i!1) + R_{60S}(isbi) \rightarrow R_{40S}(a!1, isbi!2) \circ R_{60S}(isbi!2) \circ M(c_i!1) + TC$   
 Elongation:  $R_{40S}(a!1, isbi!2) \circ R_{60S}(isbi!2) \circ M(c_i!1) \rightarrow R_{40S}(a!1, isbi!2) \circ R_{60S}(isbi!2) \circ M(c_{i+1}!1)$   
 Termination:  $R_{40S}(a!1, isbi!2) \circ R_{60S}(isbi!2) \circ M(c_i!1) \rightarrow R_{40S}(a!1, isbi) \circ M(c_i!1) + R_{60S}(isbi)$   
 Recycling:  $R_{40S}(a!1) \circ M(c_i!1) \rightarrow R_{40S}(a) + M(c_i)$   
 $80S$ - $PIC_{43S}$  collision:  $R_{40S}(a!1, isbi!2, \dagger) \circ R_{60S}(isbi!2) \circ M(c_{i-10}!3, c_i!1) \circ PIC_{43S}(a!3, \dagger) \rightarrow$   
 $R_{40S}(a!1, isbi!2, \dagger!4) \circ R_{60S}(isbi!2) \circ M(c_{i-10}!3, c_i!1) \circ PIC_{43S}(a!3, \dagger!4)$   
 $80S$ -hit  $PIC_{43S}$  dissociation:  $R_{40S}(a!1, isbi!2, \dagger!4) \circ R_{60S}(isbi!2) \circ M(c_{i-10}!3, c_i!1) \circ PIC_{43S}(a!3, \dagger!4) \rightarrow$   
 $R_{40S}(a!1, isbi!2, \dagger) \circ R_{60S}(isbi!2) \circ M(c_{i-10}, c_i!1) + R_{40S}(a, \dagger) + TC$

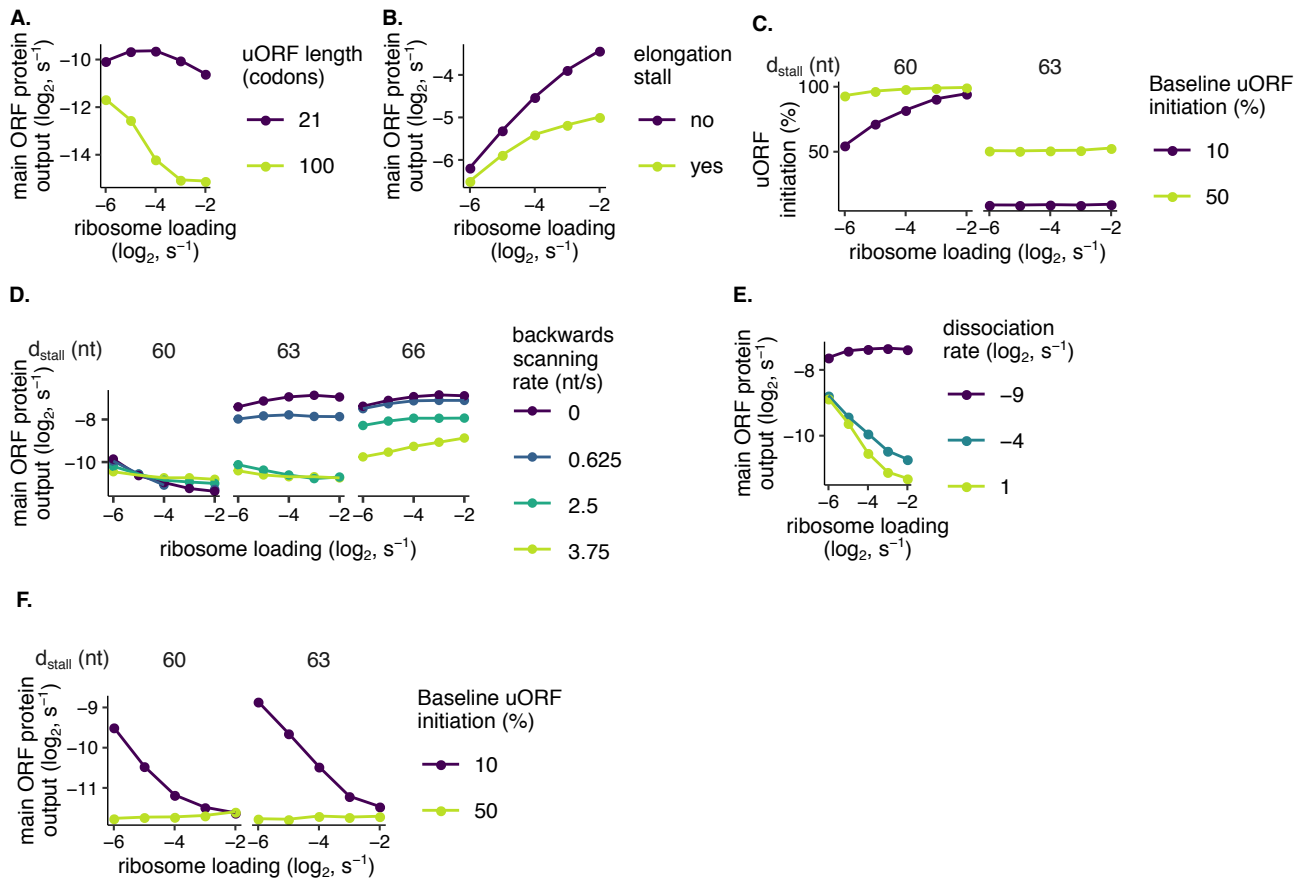
### Modeling workflow.

**(A)** Molecules in the kinetic model. Molecules have components each of which has state values or bond partners. For example, the mRNA ( $M$ ) initiation footprint ( $c_i$  to  $c_n$  where  $n$  is equal to the ribosome footprint size in nt) can either be *clear* of ribosomes, and therefore free to allow a  $PIC_{43S}$  loading reaction, or *blocked* by a ribosome and unable to allow this reaction.

**(B)** Reactions in the kinetic model. Exclamation points and the following numbers indicate component interactions following BioNetGen convention<sup>53</sup>. For example in the  $PIC_{43S}$  loading reaction, the A-site ( $a$ ) of the  $PIC_{43S}$  binds to the mRNA ( $M$ ) at codon 1 ( $c_1$ , the 5' cap). These interactions create the molecule bindings indicated by open circles. Plus signs between molecules indicate that they are not bound together. 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 forward and backward scan. Some reactions in the kinetic model, namely different types of collision and dissociation reactions, are not depicted here.

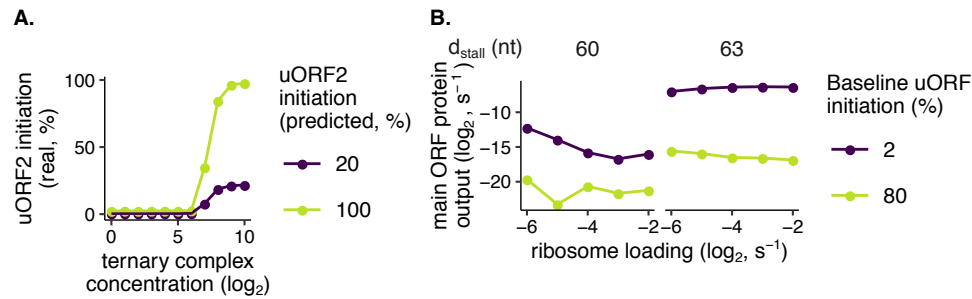


**Figure S2**



- (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. 3B, left panel, yellow-green line) and is stronger with longer uORFs (Fig. S2A, yellow-green line). However, we estimate re-initiation to be frequent (Table 1) following calibration of our modeling (Fig. 2D) to reporter measurements on wild-type or mutant uORF2 (Fig. 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. 1D) is insensitive to  $d_{stall}$ . All rates and labels are identical to the main modeling figure unless otherwise specified. Error bars of simulated data are smaller than data points.

**Figure S3**



**(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. 3C). Re-initiation is 0.2%.

All rates and labels are identical to the main modeling figure unless otherwise specified. Error bars of simulated data are smaller than data points.

**Table 1**

**Modeling parameter ranges and reporter measurement (Fig. 2C) fit values**

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 <sup>56-58</sup>
$k_{scan} (residues/s)$	1-10	5	5	5	<sup>38,99</sup>
$k_{start\ uORF2\ WT} (s^{-1})$	unknown	0.1	0.5	0.5	This work
WT uORF initiation (%)	unknown	2	10	10	This work
$k_{start\ uORF2\ strong\ Kozak} (s^{-1})$	unknown	20	5	5	This work
strong Kozak uORF initiation (%)	unknown	80	50	50	This work
$k_{elong} (residues/s)$	3-10	5	5	5	<sup>55-58</sup>
$k_{elong\ stall} (residues/s)$	0.001	0.001	0.001	0.001	<sup>101</sup>
$k_{terminate} (s^{-1})$	0.5-5	1	1	1	<sup>55</sup>
$k_{terminated\ ssu\ recycle\ uORF} (s^{-1})$	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 (residues)	21	21	21	21	<sup>32</sup>



**Table S1**

**List of plasmids used for this study**

Plasmid	Genotype	Figure	Source
pASHS64	pGL3-gp48UTR-NLuc-FLuc	Parent and 2	Promega pGL3 into which gp48 5' UTR-NLuc-FLuc was cloned (see methods)
pPNHS132	pcmv-gfpdropout-NLuc-pest-tsv40-ppgk1-FLuc-t2a-hyg-tbgh	Parent	YTK090
pTBHS1	pASHS64-uORF2-P22A	2	This work (exp1)
pTBHS2	pASHS64-no-AUG-uORF2	2	This work (exp1)
pTBHS3	pASHS64-cttaccatgg-uORF2	2	This work (exp1)
pTBHS12	pASHS64-5'-synthetic-uORF	4	This work (exp4)
pTBHS13	pASHS64-5'-synthetic-uORF-uORF2-P22A	4	This work (exp4)
pTBHS14	pASHS64-uORF2-1-codon-longer	5	This work (exp5)
pTBHS15	pASHS64-uORF2-2-codons-longer	5	This work (exp5)
pTBHS16	pASHS64-uORF2-3-codons-longer	5	This work (exp5)
pTBHS17	pASHS64-uORF2-4-codons-longer	5	This work (exp5)
pTBHS18	pASHS64-stem-loop-GC52	4	This work (exp6)
pTBHS19	pASHS64-stem-loop-GC62	4	This work (exp6)
pTBHS20	pASHS64-stem-loop-GC78	4	This work (exp6)
pTBHS21	pASHS64-stem-loop-GC92	4	This work (exp6)
pTBHS22	pASHS64-no-stem-loop-CAA-repeat	4	This work (exp6)
pTBHS23	pASHS64-5'-synthetic-uORF-no-AUG	4	This work (exp4)
pTBHS24	pASHS64-5'-synthetic-uORF-no-AUG-uORF2-P22A	4	This work (exp4)
pTBHS25	pASHS64-uORF2-7-codons-longer	5	This work (exp5)
pTBHS26	pASHS64-uORF2-10-codons-longer	5	This work (exp5)
pTBHS35	pASHS64-stem-loop-GC52-uORF2-P22A	4	This work (exp6)
pTBHS36	pASHS64-stem-loop-GC62-uORF2-P22A	4	This work (exp6)
pTBHS37	pASHS64-stem-loop-GC78-uORF2-P22A	4	This work (exp6)
pTBHS38	pASHS64-stem-loop-GC92-uORF2-P22A	4	This work (exp6)
pTBHS39	pASHS64-no-stem-loop-CAA-repeat-uORF2-P22A	4	This work (exp6)
pTBHS41	pASHS64-actgtc-5'-synthetic-uORF	4	This work (exp4)
pTBHS42	pASHS64-cattgt-5'-synthetic-uORF	4	This work (exp4)
pTBHS43	pASHS64-gaatcg-5'-synthetic-uORF	4	This work (exp4)
pTBHS44	pASHS64-tactat-5'-synthetic-uORF	4	This work (exp4)
pTBHS45	pASHS64-tccaaa-5'-synthetic-uORF	4	This work (exp4)
pTBHS46	pASHS64-ttgaa-5'-synthetic-uORF	4	This work (exp4)
pTBHS47	pASHS64-actgtc-5'-synthetic-uORF-uORF2-P22A	4	This work (exp4)
pTBHS48	pASHS64-cattgt-5'-synthetic-uORF-uORF2-P22A	4	This work (exp4)
pTBHS49	pASHS64-gaatcg-5'-synthetic-uORF-uORF2-P22A	4	This work (exp4)
pTBHS50	pASHS64-tactat-5'-synthetic-uORF-uORF2-P22A	4	This work (exp4)
pTBHS51	pASHS64-tccaaa-5'-synthetic-uORF-uORF2-P22A	4	This work (exp4)
pTBHS52	pASHS64-ttgaa-5'-synthetic-uORF-uORF2-P22A	4	This work (exp4)

Plasmid	Genotype	Figure	Source
pTBHS58	pASHS64-5'-synthetic-uORF-uORF2-2-codons-longer	S3	This work (exp4)
pTBHS59	pASHS64-5'-synthetic-uORF-uORF2-2-codons-longer-P22A	S3	This work (exp4)
pTBHS60	pASHS64-5'-synthetic-uORF-no-AUG-uORF2-2-codons-longer	S3	This work (exp4)
pTBHS61	pASHS64-5'-synthetic-uORF-no-AUG-uORF2-2-codons-longer-P22A	S3	This work (exp4)
pTBHS79	pASHS64-C1orf43-uORF	6	This work (exp12)
pTBHS83	pASHS64-TOR1AIP1-uORF	6	This work (exp12)
pTBHS84	pASHS64-HAUS4-uORF	6	This work (exp12)
pTBHS86	pASHS64-C15orf59-uORF	6	This work (exp12)
pTBHS87	pASHS64-PPP1R37-uORF	6	This work (exp12)
pTBHS89	pASHS64-ABCB9-uORF	6	This work (exp12)
pTBHS91	pASHS64-C1orf43-uORF-P12A	6	This work (exp12)
pTBHS95	pASHS64-TOR1AIP1-uORF-P25A	6	This work (exp12)
pTBHS98	pASHS64-C15orf59-uORF-P24A	6	This work (exp12)
pTBHS99	pASHS64-PPP1R37-uORF-P23A	6	This work (exp12)
pTBHS101	pASHS64-ABCB9-uORF-P40A	6	This work (exp12)
pTBHS127	pASHS64-C1orf43-no-AUG-uORF	6	This work (exp18)
pTBHS128	pASHS64-C15orf59-no-AUG-uORF	6	This work (exp18)
pTBHS131	pASHS64-ABCB9-no-AUG-uORF	6	This work (exp18)
pTBHS132	pASHS64-PPP1R37-no-AUG-uORF	6	This work (exp18)
pTBHS133	pASHS64-TOR1AIP1-no-AUG-uORF	6	This work (exp18)
pTBHS134	pPNHS132-uORF2-NLuc-PEST	4	This work (exp15)
pTBHS135	pPNHS132-uORF2-P22A-NLuc-PEST	4	This work (exp15)

# References

1. Wethmar, K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **5**, 765–778 (2014).
2. Ye, Y. *et al.* Analysis of human upstream open reading frames and impact on gene expression. *Hum Genet* **134**, 605–612 (2015).
3. 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).
4. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORFs are prevalent translational repressors in vertebrates. *The EMBO Journal* **35**, 706–723 (2016).
5. 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).
6. Chen, F. *et al.* Dissimilation of synonymous codon usage bias in virus–host coevolution due to translational selection. *Nature Ecology & Evolution* **4**, 589–600 (2020).
7. Zhang, H. *et al.* Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nature Communications* **12**, 1076 (2021).
8. Giess, A. *et al.* Deconstructing the individual steps of vertebrate translation initiation. *bioRxiv* 811810 (2019) doi:[10.1101/811810](https://doi.org/10.1101/811810).
9. 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).
10. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**, 8125–8148 (1987).
11. Dvir, S. *et al.* Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *PNAS* **110**, E2792–E2801 (2013).
12. Lee, D. S. M. *et al.* Disrupting upstream translation in mRNAs is associated with human disease. *Nat Commun* **12**, 1515 (2021).
13. Jürgens, L. *et al.* Somatic Functional Deletions of Upstream Open Reading Frame-Associated Initiation and Termination Codons in Human Cancer. *Biomedicines* **9**, 618 (2021).
14. Schulz, J. *et al.* Loss-of-function uORF mutations in human malignancies. *Sci Rep* **8**, 2395 (2018).
15. Smith, R. C. L. *et al.* Translation initiation in cancer at a glance. *J Cell Sci* **134**, (2021).
16. Vattam, 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).
17. 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).
18. Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601 (2013).
19. Gandin, V. *et al.* Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. *Nature* **455**, 684–688 (2008).
20. 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).
21. Cao, J. & Geballe, A. P. Coding sequence-dependent ribosomal arrest at termination of translation. *Molecular and Cellular Biology* **16**, 603–608 (1996).
22. 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).
23. 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).
24. 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).
25. 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).
26. Lin, Y. *et al.* Impacts of uORF codon identity and position on translation regulation. (2019) doi:[10.1184/R1/10110866.v1](https://doi.org/10.1184/R1/10110866.v1).
27. Bhushan, S. *et al.* Structural Basis for Translational Stalling by Human Cytomegalovirus and Fungal Arginine Attenuator Peptide. *Molecular Cell* **40**, 138–146 (2010).
28. Wilson, D. N., Arenz, S. & Beckmann, R. Translation regulation via nascent polypeptide-mediated ribosome stalling. *Current Opinion in Structural Biology* **37**, 123–133 (2016).
29. 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).

30. 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).
31. 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).
32. 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](https://doi.org/10.1128/JVI.67.9.5514-5521.1993).
33. Boersma, S. *et al.* Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. *Cell* **178**, 458–472.e19 (2019).
34. 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).
35. 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).
36. 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).
37. 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](https://doi.org/10.1016/j.mbs.2021.108762).
38. 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).
39. Park, H. & Subramaniam, A. R. Inverted translational control of eukaryotic gene expression by ribosome collisions. *PLoS Biol* **17**, e3000396 (2019).
40. 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).
41. Kozak, M. Constraints on reinitiation of translation in mammals. *Nucleic Acids Res* **29**, 5226–5232 (2001).
42. 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).
43. Kozak, M. Effects of Intercistronic Length on the Efficiency of Reinitiation by Eucaryotic Ribosomes. *MOL. CELL. BIOL.* **7**, 8 (1987).
44. 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).
45. 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).
46. 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).
47. Col, B., Oltean, S. & Banerjee, R. Translational Regulation of Human Methionine Synthase by Upstream Open Reading Frames. *Biochim Biophys Acta* **1769**, 532–540 (2007).
48. Yamashita, Y. *et al.* Sucrose sensing through nascent peptide-mediated ribosome stalling at the stop codon of Arabidopsis bZIP11 uORF2. *FEBS Letters* **591**, 1266–1277 (2017).
49. Tanaka, M. *et al.* The Minimum Open Reading Frame, AUG-Stop, Induces Boron-Dependent Ribosome Stalling and mRNA Degradation. *The Plant Cell* **28**, 2830–2849 (2016).
50. 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).
51. 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 and Cell Physiology* **55**, 1556–1567 (2014).
52. Lopez, C. F., Mühlich, J. L., Bachman, J. A. & Sorger, P. K. Programming biological models in Python using PySB. *Molecular Systems Biology* **9**, (2013).
53. Harris, L. A. *et al.* BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics* **32**, 3366–3368 (2016).
54. 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).
55. Ingolia, 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).
56. Yan, X., Hoek, Tim A., Vale, Ronald D. & Tanenbaum, Marvin E. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165**, 976–989 (2016).



57. Morisaki, T. *et al.* Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**, 1425–1429 (2016).
58. Wu, B., Eliscovich, C., Yoon, Y. J. & Singer, R. H. Translation dynamics of single mRNAs in live cells and neurons. *Science* **352**, 1430–1435 (2016).
59. Costa-Mattioli, M. & Walter, P. The integrated stress response: From mechanism to disease. *Science* **368**, (2020).
60. Pakos-Zebrucka, K. *et al.* The integrated stress response. *EMBO Rep* **17**, 1374–1395 (2016).
61. 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).
62. Vladimer, G. I., Górna, M. W. & Superti-Furga, G. IFITs: Emerging Roles as Key Anti-Viral Proteins. *Front Immunol* **5**, 94 (2014).
63. Spriggs, K. A., Bushell, M. & Willis, A. E. Translational Regulation of Gene Expression during Conditions of Cell Stress. *Molecular Cell* **40**, 228–237 (2010).
64. Andreev, D. E. *et al.* Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *eLife* **4**, e03971 (2015).
65. Sidrauski, C., McGeachy, A. M., Ingolia, N. T. & Walter, P. The small molecule ISRIB reverses the effects of eIF2 $\alpha$  phosphorylation on translation and stress granule assembly. *eLife* **4**, e05033 (2015).
66. Baird, T. D. *et al.* Selective mRNA translation during eIF2 phosphorylation induces expression of IBTK $\alpha$ . *Mol Biol Cell* **25**, 1686–1697 (2014).
67. Wei, S. *et al.* Ribosome profiling reveals translome remodeling in cancer cells in response to zinc oxide nanoparticles. *Aging* **13**, 23119–23132 (2021).
68. Jackson, R. & Standart, N. The awesome power of ribosome profiling. *RNA* **21**, 652–654 (2015).
69. Matsuda, D. & Dreher, T. W. Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. *RNA* **12**, 1338–1349 (2006).
70. Gu, 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).
71. Li, 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.
72. Kozak, M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* **9**, 5073–5080 (1989).
73. Babendure, 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).
74. Cencic, R. *et al.* Reversing chemoresistance by small molecule inhibition of the translation initiation complex eIF4F. *PNAS* **108**, 1046–1051 (2011).
75. Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences* **21**, 267–271 (1996).
76. Martinez, T. F. *et al.* Accurate annotation of human protein-coding small open reading frames. *Nature Chemical Biology* **16**, 458–468 (2020).
77. 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).
78. Archer, S. K., Shirokikh, N. E., Beilharz, T. H. & Preiss, T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* **535**, 570–574 (2016).
79. Giess, A. *et al.* Profiling of Small Ribosomal Subunits Reveals Modes and Regulation of Translation Initiation. *Cell Reports* **31**, 107534 (2020).
80. 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.
81. 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).
82. 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).
83. 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.
84. Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* **9**, 5134–5142 (1989).
85. Zhao, T. *et al.* Disome-seq reveals widespread ribosome collisions that promote cotranslational protein folding. *Genome Biol* **22**, 16 (2021).
86. Arpat, A. B. *et al.* Transcriptome-wide sites of collided ribosomes reveal principles of translational pausing. *Genome Res.* **30**, 985–999 (2020).
87. 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).

88. 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).
89. Mohammad, M. P., Munzarová Pondelícková, V., Zeman, J., Guňišová, 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).
90. Sanchez, M. *et al.* Cross Talk between eIF2 $\alpha$  and eEF2 Phosphorylation Pathways Optimizes Translational Arrest in Response to Oxidative Stress. *iScience* **20**, 466–480 (2019).
91. 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](https://doi.org/10.1038/s41589-021-00913-4).
92. 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](https://doi.org/10.1016/j.molcel.2020.11.033).
93. Wu, C. C.-C., Peterson, A., Zinshteyn, B., Regot, S. & Green, R. Ribosome Collisions Trigger General Stress Responses to Regulate Cell Fate. *Cell* (2020) doi:[10.1016/j.cell.2020.06.006](https://doi.org/10.1016/j.cell.2020.06.006).
94. Wallace, 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).
95. Aliouat, 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).
96. Jia, L. *et al.* Decoding mRNA translatability and stability from the 5' UTR. *Nat Struct Mol Biol* **27**, 814–821 (2020).
97. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).
98. Sambrook, J. *Molecular Cloning: A Laboratory Manual, Third Edition*. (Cold Spring Harbor Laboratory Press, 2001).
99. Vassilenko, 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).
100. Pichon, 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).
101. Cao, J. & Geballe, A. P. Ribosomal release without peptidyl tRNA hydrolysis at translation termination in a eukaryotic system. *RNA* **4**, 181–188 (1998).
102. Wu, 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).