Probabilistic models of uORF-mediated ATF4 translation control

Olivia N. J. M. Marasco[a,c], Marc R. Roussel[a,c,*], Nehal Thakor[b,c,d]

[a] *Alberta RNA Research and Training Institute, University of Lethbridge*
[b] *Southern Alberta Genome Sciences Centre and Department of Biological Sciences, University of Lethbridge*
[c] *Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada*
[d] *Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta, T2N 4Z6, Canada*

Abstract

ATF4 is a key transcription factor that activates transcription of genes needed to respond to cellular stress. Although the mRNA encoding ATF4 is present at constant levels in the cell during the initial response, translation of ATF4 increases under conditions of cellular stress while the global translation rate decreases. We study two models for the control system that regulates the translation of *ATF4*, both based on the Vattem-Wek hypothesis. This hypothesis is based on a race to reload, following the translation of a small upstream open reading frame (uORF), the ternary complex that brings the initiator tRNA to the ribosome as the 40S subunit scans along the mRNA, encountering first a start codon for an inhibitory uORF whose reading frame overlaps the start of the *ATF4* coding sequence. We develop a pair of simple, analytic, probabilistic models, one of which assumes all nucleotide triplets have identical kinetic properties, while the other recognizes the existence of triplets at which the ternary complex loads more efficiently. We also consider two different functions representing the dependence of the rate of initiation at uORF1 on the ternary complex concentration. In keeping with the theme of this Special Issue, we studied the properties of these models in a Maple document, which can easily be modified to consider different parameters, translation rate initiation functions, and so on.

Keywords: Translation control, ATF4, upstream open reading frame, probabilistic model, computer algebra, dynamic publication

*Corresponding author
Email addresses: o.marasco@uleth.ca (Olivia N. J. M. Marasco), roussel@uleth.ca (Marc R. Roussel)

1. Introduction

Cellular stress engages a multitude of regulatory responses that alter gene expression patterns through both global and specific changes in transcription and translation rates [1, 2], leading either to adaptation and recovery, or to cell death. This integrated stress response (ISR) relies on regulatory serine kinases—PKR (Protein Kinase R) which responds to viral infection; PERK (PKR-like Endoplasmic Reticulum Kinase) which responds to endoplasmic reticulum (ER) stress; GCN2 (General Control Non-derepressible protein 2) which responds to amino-acid deprivation; and HRI (Heme Regulated eIF2α kinase) which responds to heme deprivation—to phosphorylate the serine-51 residue on the α subunit of eukaryotic initiation factor 2 (eIF2) [1]. The initiator tRNA is normally conveyed to the 40S ribosomal subunit during translation initiation in a ternary complex (TC) consisting of eIF2, the initiator tRNA (Met-tRNA$_i^{Met}$) and GTP. Recycling of eIF2 following the delivery of the initiator tRNA requires the guanine nucleotide exchange factor (GEF) eIF2B, which facilitates the exchange of GDP in the "spent" complex for GTP. Phosphorylation of eIF2α greatly increases the affinity of eIF2B for eIF2, resulting in the sequestration of eIF2B in an inactive complex. Since eIF2B is less abundant in cells than eIF2 [3], this both ties up a fraction of the eIF2 complexes, and inhibits the necessary nucleotide exchange, thus globally repressing translation [4–6]. However, despite the repression of global protein synthesis, there are several mRNAs encoding effectors of the ISR whose translation levels increase under these conditions [1]. These ISR protein effectors go on to act as transcription factors for other targets that alleviate stress, promote apoptosis, or bring about the eventual derepression of global translation initiation [4].

One of the best characterized effectors of the ISR (and a master regulator of cell stress) is Activating Transcription Factor 4 (ATF4). Like many of the effectors of the ISR, the increase in ATF4 expression under conditions of global translation repression is attributed to the presence of upstream open reading frames (uORFs) within the 5' untranslated region (5'-UTR) of its mRNA [7]. As their name suggests, uORFs are mRNA sequences recognized by the translation initiation machinery located upstream (5') of the coding sequence of a gene [8]. 49% of human transcripts contain a uORF, and the positions of the start codons of these uORFs are conserved across species, indicating a functional role [9]. A uORF is translated to a short peptide, but it is the act of translating a uORF rather than the peptide that typically plays the regulatory role. The translation of a uORF can have various effects on the translation efficiency of a protein depending on the relative placement of the uORF and start codon of the protein-coding sequence (the main ORF, abbreviated mORF), although there is a marked tendency for uORFs to reduce protein expression levels [9].

So how do typically repressive uORFs result in increased translation of ATF4 under stress conditions? In a 2004 paper, Vattem and Wek proposed a mechanism resolving this conundrum [10]. The ATF4 transcript has two uORFs, the second of which overlaps the start codon of the protein-coding sequence. As noted above, under normal conditions, eIF2 forms a ternary complex with

GTP and the initiator tRNA. This ternary complex is required for translation initiation. Following translation of uORF1, the 40S ribosomal subunit resumes scanning [11]. Since the ternary complex is abundant, initiation at uORF2 is efficient, and the peptide encoded by uORF2 will be translated with high efficiency, preventing *ATF4* from being translated. However, under stress conditions, eIF2α is phosphorylated, which inhibits the exchange of GDP for GTP in spent ternary complexes, thus reducing the availability of active ternary complexes [1]. Accordingly, the rate of translation initiation decreases globally within the cell. However, the slow rate of reloading of ternary complexes means that some ribosomes that completed the translation of uORF1 will scan past the start codon of uORF2 before they acquire a new TC. If a TC loads between the start codons of uORF2 and the coding sequence, the latter can be translated. Thus, the Vattem-Wek mechanism describes a game of chance involving the probabilities of loading a TC before reaching uORF2, or between the start codons of uORF2 and of the protein-coding sequence. The concentration of TC in the cell biases this game towards one outcome or the other.

Our objective in this paper is to study the properties of this class of control systems in order to facilitate rational design of 5'-UTRs for synthetic biological applications. As an example of what we have in mind, consider that cancer cells are, for all intents and purposes, chronically stressed by their rapid growth, which causes the accumulation of proteins at a rate that challenges the endoplasmic reticulum's ability to fold them, and by the consequent issues with supply of nutrients and oxygen [12]. These stresses will cause eIF2α to be phosphorylated, which could be exploited to engineer an mRNA translated mostly in cancer cells [13]. Recent advances in mRNA therapies make such an approach at least plausible, albeit likely in a combination therapy [14]. A similar strategy could be used to target an antiviral peptide for expression in cells that have been infected by a virus: Viral double-stranded RNA activates PKR, which phosphorylates eIF2α [15], thus activating the translation of a suitably designed mRNA.

One of the two models developed below is similar to one studied by You et al. [16]. These authors developed a moderately detailed stochastic model of translation initiation for the yeast protein GCN4, the mRNA of which contains four uORFs. Unlike *ATF4*, none of these uORFs overlaps the start codon of the GCN4 main ORF. The control logic is thus slightly different, and is based on dissociation of the ribosome following translation of any of uORFs 2 to 4. Their model included a "factor X" originally hypothesized by Grant et al. [17] to explain discrepancies in rates of initiation between uORF4 and the GCN4 mORF. (To our knowledge, no such factor has been found.) Based on experimental data showing that uORFs 2 and 3 have very little effect on the translation of *GCN4*, they simplified their model to just two uORFs, which reduced their model to one with a strong family resemblance to the models we will be presenting. From their stochastic model, they obtained probabilistic expressions for the translation of *GCN4*. Their model was conditioned using experimental data from Grant et al. [17]. They then computed various properties of this control system, including notably the translation rate as a function of

initiation rate at uORF1.

In a joint experimental-theoretical study, Roy et al. used a very simple probabilistic model to interpret experimental data on the role of eIF3 in reinitiation following translation of a uORF [18]. This model assumed exponential decay in the competence for reinitiation during translation of a uORF and exponential gain in competence in the intercistronic region. The models presented below also result in exponential gain in competence following translation of uORF1.

Both of the models mentioned above, the only published models on the role of uORFs in translation initiation of which we are aware, dealt with different systems, and had different aims than the current study. It therefore seemed appropriate to develop and study a model of the paradigmatic ATF4 uORF-mediated stress-induced translation control system. Our first model, described in Section 2.1, is based on the nucleotide-by-nucleotide search of the mRNA for a start codon following termination of translation of uORF1. At each nucleotide, there is a stochastic race between loading a ternary complex and advancing to the next nucleotide. This competitive process is repeated again and again as the 40S subunit scans along the mRNA. Once a TC has been loaded, then the ORF whose start codon is most proximate downstream of the current position will be translated. Thus, if a TC is loaded before passing the start codon of uORF2, the latter is translated, but not ATF4. Alternatively, if a TC is loaded after passing uORF2's start codon but before passing ATF4's start codon, then the *ATF4* main ORF is translated. In this first model, which shares features with both the You et al. [16] and Roy et al. [18] models, we assume that every nucleotide position is the same. This reasoning leads to simple equations for the probabilities of translating uORF2 or ATF4, as well as the probability of translating neither if TC loading does not occur in a timely way. We also consider that the rate of initiation at uORF1 will depend on the TC concentration, considering two possible models for this dependence, and thus two variations on this first model. Considering the effect of TC concentration on the rate of translation initiation of uORF1 and its consequent effect on the overall ATF4 expression rate is a novel aspect of our models relative to earlier modeling efforts [16, 18].

The loading of the ternary complex involves an interaction between this complex, the 40S subunit and, critically, the mRNA. The rate constant for TC loading thus depends on the triplet of nucleotides presented by the 40S subunit [19]. (Note that we use the term "nucleotide triplet", or just "triplet", since these are not necessarily in-frame in any ORF, and therefore may not be codons in any reasonable sense of the word.) Our second model, developed in Section 2.2, considers two types of triplets, fast-loading triplets, which are a subset of the near-cognates of the start codon as well as the start codon itself, and slow-loading triplets. This leads us to a model for which we can estimate, at least roughly, all of the parameters, which we do in Section 2.3. One of the key parameters is later refined in Section 3.2.1 using experimental data available in the literature. This model also has two variations based on our assumptions about the dependence of the rate of translation initiation of uORF1 on the TC concentration.

We briefly study the properties of Model 1 in Section 3.1. A parameter esti-

mate derived from matching the model to experimental data strongly suggests that fast triplets are key to the behavior of this control system. We therefore move on to Model 2, whose properties are studied in Section 3.2. The model generally displays realistic behavior, although there are some interesting discrepancies with respect to mutants, discussed in Section 3.2.2. The response characteristics of the model with respect to the parameters that are available to the genetic engineer are discussed in detail in Section 3.2.3.

This paper is a contribution to a special issue on Dynamic Publication Media in Mathematical Biology. We have for the last few years often carried out at least parts of our computations in Maple documents and Matlab live scripts, typically using these media as organizational tools for our research. This particular project lent itself particularly well to study in a Maple document, which is provided as supplementary information to this paper. This gave us access to Maple's computer algebra engine, while also allowing the mixing of mathematics and narrative. Accordingly, we built a Maple document in which all of the computations are carried out and presented with explanatory notes and some additional calculations that would normally have ended up in a distinct supplementary information document. Some lessons learned along the way are discussed in the concluding section of this paper, along with the scientific conclusions of this study.

The calculations reported in this contribution were carried out using Maple version 2019.2, but should run correctly in any recent version of Maple, including the free Maple player available from MapleSoft at `https://www.maplesoft.com/products/maple/Mapleplayer`.

2. Model Descriptions

2.1. Model 1

Our first model is based on Vattem and Wek's description of the *ATF4* translation control system [10]. The structure of the mRNA is described in Fig. 1. A ribosome initially assembles to translate uORF1. Scanning reinitiates efficiently following the translation of short uORFs [11], but the ability to initiate translation depends on the arrival of a ternary complex to the scanning 40S subunit. Normally, when there is a plentiful supply of $\mathrm{eIF2}\cdot\mathrm{GTP}\cdot\mathrm{Met}-\mathrm{tRNA}_\mathrm{i}^{\mathrm{Met}}$ ternary complexes, a TC can be expected to be loaded into the scanning 40S subunit before it reaches the start codon of uORF2, forming an initiation-competent 43S complex which, once it reaches the start codon, will recruit first eIF5, forming a 48S pre-initiation complex, and then the 60S ribosomal subunit, allowing uORF2 to be translated [20]. Since uORF2 overlaps the start codon of ATF4, the latter will then not be translated. Under stress conditions, resulting in inactivation of eIF2 by phosphorylation of the α subunit, the probability that an active ternary complex arrives before the scanning 40S subunit reaches the start codon of uORF2 is depressed. This may allow the 40S subunit to scan through this start codon. There is then a chance that a ternary complex will bind to the 40S subunit before reaching the start codon of ATF4, where translation can be initiated.

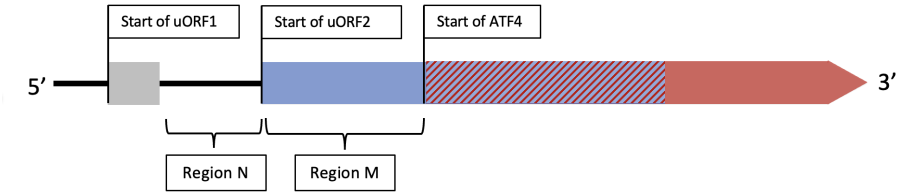

Figure 1: Qualitative illustration of the ATF4 mRNA showing the locations of uORF1 (grey), uORF2 (blue), and the ATF4 coding sequence (red) (GenBank accession number BC008090.1 [21]). Region $\mathcal{N}$ denotes the spacer between the end of uORF1 and the start of uORF2 while region $\mathcal{M}$ denotes the sequence from the start of the uORF2 reading frame to the start of the ATF4 reading frame.

In either case, translation of uORF1 initiates the process, i.e. translation of this uORF is a necessary precondition to the translation of downstream elements. Thus, our model focuses on the fate of the translation complex following termination of translation of uORF1 and resumption of scanning by the 40S subunit. This means that probabilities derived below are conditional on translation of uORF1 and on the resumption of scanning. Note that the rate of translation of uORF1 is itself dependent on the availability of ternary complex, a detail to which we will return shortly.

From the foregoing considerations, our model starts right after completion of the translation of uORF1. Scanning is a nucleotide-by-nucleotide process [20]. At any given nucleotide, we consider two competing processes: either the 40S subunit scans through to the next nucleotide at a specific rate k_s, or it acquires a ternary complex (TC), and is converted to a stable 43S complex capable of initiating translation. The latter process has a per-nucleotide rate of $k_a C$, where C is the concentration of ternary complex.

In order to reduce the model to these two processes, we are making the following assumptions:

- The 40S subunit only scans in the forward direction. Reverse scanning has been found to not be significant except when the upstream start codon is very close to the stop codon of a uORF [11, 22–24].
- Abortive scanning is negligible [22, 25]. We have looked at a model variation with abortive scanning, and other than depressing the expression probabilities, it has little effect on the results.
- Leaky scanning (bypassing an AUG codon without initiating translation) does not occur. Rates of leaky scanning are highly context-dependent, and we do not have sufficient information to estimate these rates for uORF2 and the mORF of the ATF4 mRNA.
- Translation initiates only at canonical AUG start codons.
- Formation of the 43S complex after binding of the TC is essentially irreversible [26].

Under these assumptions, there are three possible outcomes following the translation of uORF1:

1. The 40S ribosome regains competence in region $\mathcal{N}$ and translates uORF2 with probability P_1.

2. The 40S ribosome regains competence in region $\mathcal{M}$ after scanning through $\mathcal{N}$ and translates the ATF4 mORF with probability P_2.

3. The 40S ribosome does not regain competence before the start of the ATF4 gene and neither uORF2 nor the mORF is translated with probability P_3.

Under the assumptions above, the probability that the 40S subunit scans through to the next nucleotide *without* acquiring a ternary complex is

$$p_s = \frac{k_s}{k_a C + k_s} = \frac{1}{1 + C/K}, \tag{1}$$

where

$$K = k_s/k_a. \tag{2}$$

The 40S subunit scans through region $\mathcal{N}$ (Fig. 1), which we define as the region between the end of uORF1 up to and including the start codon of uORF2. If the 43S complex acquires a TC in this region, uORF2 will be translated. The probability that the 40S subunit *does not* acquire a TC in region $\mathcal{N}$ is p_s^N, where N is the number of nucleotides in region $\mathcal{N}$. The probability that uORF2 is translated is the complement of this probability, i.e.

$$P_1(C/K) = 1 - p_s^N = 1 - \left(\frac{1}{1 + C/K}\right)^N. \tag{3a}$$

In order for the mORF to be translated, the 40S subunit must scan through region $\mathcal{N}$ without acquiring a ternary complex, then must acquire a TC in region $\mathcal{M}$, the region starting right after the start codon of uORF2 up to and including the mORF start codon (of length M nucleotides, Fig. 1). The probability of acquiring a TC in region $\mathcal{M}$ *given that* the 40S subunit scanned through region $\mathcal{N}$ is, using identical reasoning to the above, $1 - p_s^M$. Thus, the probability that the mORF is translated is

$$\begin{aligned} P_2(C/K) &= p_s^N \left(1 - p_s^M\right) \\ &= \left(\frac{1}{1 + C/K}\right)^N \left[1 - \left(\frac{1}{1 + C/K}\right)^M\right]. \end{aligned} \tag{3b}$$

Finally, the probability that neither uORF2 nor the mORF is translated is simply the probability of scanning through both regions $\mathcal{N}$ and $\mathcal{M}$:

$$P_3(C/K) = p_s^{N+M} = \left(\frac{1}{1 + C/K}\right)^{N+M}. \tag{3c}$$

Somewhat similar equations were previously obtained by You and coworkers [16]. Note also the exponential dependence of P_2 on M, which also features in the work of Roy et al. [18].

Equation (3b) gives the conditional probability that ATF4 will be translated given that uORF1 has been translated and that scanning resumes following the translation of uORF1 under the assumptions mentioned above. The eventual production of ATF4 is however not solely dependent on this conditional probability. Suppose that v_{ATF4} is the rate of translation of ATF4 and k_d is the rate constant for degradation. Thus,

$$\frac{d[\mathrm{ATF4}]}{dt} = v_{\mathrm{ATF4}} - k_d[\mathrm{ATF4}]. \tag{4}$$

Note that v_{ATF4} depends on, among other things, the concentrations of ATF4 mRNA, which is constant, at least during the initial stages of the stress response [10]. A steady state is reached when $[\mathrm{ATF4}] = v_{\mathrm{ATF4}}/k_d$. This also represents the maximum concentration of ATF4 that can be reached for a given translation rate. Accordingly, the ATF4 concentration is bounded by a quantity proportional to the translation rate. This rate can be decomposed as the product of the rate of translation of uORF1 (v_{uORF1}), the probability that the 40S subunit resumes scanning (P_{res}), and the conditional probability that ATF4 is translated (P_2), i.e.

$$v_{\mathrm{ATF4}} = v_{\mathrm{uORF1}} P_{\mathrm{res}} P_2. \tag{5}$$

The dependence of v_{ATF4} on the mRNA concentration enters via v_{uORF1}, which should be proportional to [mRNA]. We assume that P_{res} is constant. The uORF1 translation rate on the other hand will depend on the state of the cell, and in particular on the availability of ternary complexes. We define the rate of translation of ATF4 relative to the rate at which scanning is resumed in an unstressed cell as

$$r(C) = \frac{v_{\mathrm{ATF4}}(C)}{v_{\mathrm{uORF1}}(\bar{C})\, P_{\mathrm{res}}} = \frac{v_{\mathrm{uORF1}}(C)}{v_{\mathrm{uORF1}}(\bar{C})} P_2(C/K), \tag{6}$$

where $\bar{C}$ is the ternary complex concentration in an unstressed cell. Note that $r(C)$ does not depend on the mRNA concentration, being a ratio of rates each proportional to [mRNA]. With the simple scale transformation

$$c = C/\bar{C}, \tag{7}$$

we can define

$$f(c) = \frac{v_{\mathrm{uORF1}}(c\bar{C})}{v_{\mathrm{uORF1}}(\bar{C})}, \tag{8}$$

such that

$$r(c) = f(c)\, P_2(\gamma c), \tag{9}$$

where

$$\gamma = \bar{C}/K = k_a \bar{C}/k_s. \tag{10}$$

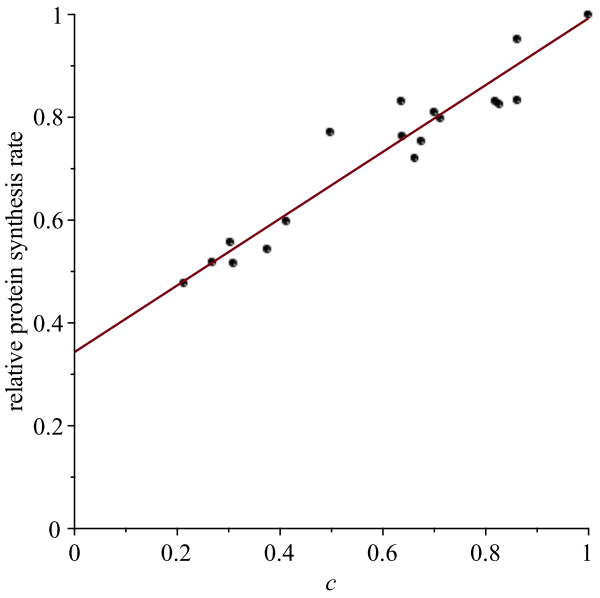

Figure 2: Global translation rate vs relative amount of eIF2α (c) for eIF2α abundances of 100% or less of wild-type. Data from from Firczcuk and coworkers [27, Fig. 1] extracted with Plot Digitizer (version 2.6.9).

Thus, only one kinetic parameter appears in the final equations describing the rate of expression of ATF4. Additionally note that in a study of eIF2α depletion, only values of $c \leq 1$ are of interest, with $c = 1$ indicating a cell in a normal (unstressed) condition.

In yeast, Firczuk and coworkers found that the global rate of translation is linearly dependent on the availability of eIF2α over a range of 20 to 100% of the wild-type abundance [27, Fig. 1]. (Interestingly, the rate of translation seems to saturate with respect to eIF2α above 100% of wild-type abundance, which suggests that evolution has tuned the synthesis of this translation initiation factor to provide a maximal rate of protein synthesis while making translation immediately responsive to falling levels of eIF2α.) We have replotted their data for $c < 1$ in Fig. 2 along with the line of best fit. The coordinate c of this plot assumes that the TC concentration is proportional to the amount of eIF2α available. The linearity of this plot indicates there is a component of the overall protein synthesis activity of the cell that is directly proportional to the eIF2α concentration and, presumably, to the concentration of ternary complexes containing unphosphorylated eIF2α. On the other hand, the non-zero intercept of this graph (assuming that linearity continues to hold) is likely due to translation by non-canonical mechanisms that do not require eIF2α [5], notably mechanisms that use either eIF2A [28] or eIF5B [29] to deliver initiator tRNA to the ribosome.

Since we do not know exactly how the rate of uORF1 translation depends on the TC concentration, we consider two alternative hypotheses:

A The rate of uORF1 translation follows the general trend of global translation, switching to a non-canonical mechanism as the TC concentration drops, i.e. the relative rate of uORF1 translation, $f(c)$, is given by the line of best fit in Fig. 2, viz.

$$f_A(c) \approx 0.34 + 0.65c. \tag{11}$$

The dependence of the rate of translation of ATF4 on the concentration of eIF5B is evidence in favor of switching to a non-canonical pathway for initiation on the ATF4 mRNA [30], thus supporting the use of the uORF1 initiation rate function (11) in this model. The resulting model, if correct, implies that different mechanisms may operate for initiation than for reinitiation, given that we assume strict dependence on the TC concentration in the latter case.

B We also consider the extreme that uORF1 translation initiation can only proceed by the eIF2α-dependent canonical translation initiation pathway. In this case, we would have

$$f_B(c) = c. \tag{12}$$

A complete model consists of Eqs. (3b), (9) and one of Eqs. (11) or (12). We call the corresponding models 1A and 1B.

2.2. Model 2

Model 1 assumes that the behavior is the same at every nucleotide. However, Kolitz and coworkers found that the rate constant for TC acquisition (k_a in our model) by a scanning 40S subunit varies greatly according to the nucleotide triplet presented during scanning, from 3.3×10^3 to $5.6 \times 10^6\,\mathrm{M^{-1}s^{-1}}$ [19]. The model introduced below takes a first step towards considering sequence effects by separating nucleotide positions into two sets: those that, taken with the following two bases, form a triplet with typical kinetic characteristics, and those that, with the following two bases, form a triplet with elevated TC loading rates.

To better understand the need for such a model, consider Fig. 3, in particular part B of the figure. There are significant differences in k_a values between the fast-loading triplets AUG, GUG and UUG, for which $k_a > 10^6\,\mathrm{M^{-1}s^{-1}}$, the intermediate CUG triplet, for which $k_a = 4.8 \times 10^4\,\mathrm{M^{-1}s^{-1}}$, and the slow-loading triplets whose k_a values are all below $10^4\,\mathrm{M^{-1}s^{-1}}$. Both region $\mathcal{N}$ (following the uORF1 stop codon up to and including the uORF2 start codon, colored blue in Fig. 3(a)) and region $\mathcal{M}$ (following the uORF2 start codon up to an including the mORF start codon, red in Fig. 3A) contain several triplets with large TC-loading rate constants. On the other hand, slower-loading near-cognates of AUG have k_a values similar to the rate constant for loading a TC

into the 40S subunit in the absence of mRNA [19]. This suggests that k_a is relatively constant except at triplets with good TC affinity. Moreover, the large gap between one set of rate constants and the rest suggests that we can, to a first approximation, consider only two types of nucleotide triplets.

We will use hats to distinguish probabilities in Model 2 from those in Model 1. We denote the probability of scanning through a nucleotide when the triplet positioned for recognition by the TC is an ordinary triplet by p_s, and by p_f the probability of scanning through a fast-loading triplet. The outcomes considered are the same as in the previous model. The first of these is that the ribosome regains competence in region $\mathcal{N}$ and translates uORF2. The probability of scanning through region $\mathcal{N}$ is therefore $p_s^{N-n}p_f^n$, where n is the number of fast-loading triplets in this region (including the uORF2 start codon). The probability of translating uORF2 is therefore

$$\hat{P}_1 = 1 - p_s^{N-n}p_f^n = 1 - \left(\frac{1}{1+C/K}\right)^{N-n}\left(\frac{1}{1+C/K_f}\right)^n, \tag{13a}$$

where $K_f = k_s/k_a^{(f)}$, with $k_a^{(f)}$ the mean "fast" value of k_a. Similarly, the probability of scanning through region $\mathcal{M}$ given that the 40S subunit had scanned through region $\mathcal{N}$ is $p_s^{M-m}p_f^m$, where m is the number of favorable triplets in region $\mathcal{M}$. Thus, the probability of translating the ATF4 mORF is

$$\begin{aligned}\hat{P}_2 &= p_s^{N-n}p_f^n\left(1 - p_s^{M-m}p_f^m\right)\\ &= \left(\frac{1}{1+C/K}\right)^{N-n}\left(\frac{1}{1+C/K_f}\right)^n\\ &\qquad\times\left[1-\left(\frac{1}{1+C/K}\right)^{M-m}\left(\frac{1}{1+C/K_f}\right)^m\right].\end{aligned} \tag{13b}$$

Finally, the probability of translating neither uORF2 nor the mORF is

$$\begin{aligned}\hat{P}_3 &= p_s^{N-n+M-m}p_f^{n+m}\\ &= \left(\frac{1}{1+C/K}\right)^{N-n+M-m}\left(\frac{1}{1+C/K_f}\right)^{n+m}.\end{aligned} \tag{13c}$$

Because of the appearance of a new parameter with units of concentration, an additional dimensionless parameter appears in this model. We can rewrite $\hat{P}_2$ as follows:

$$\begin{aligned}\hat{P}_2 = \left(\frac{1}{1+\gamma c}\right)^{N-n}&\left(\frac{1}{1+\gamma\kappa c}\right)^n\\ &\times\left[1-\left(\frac{1}{1+\gamma c}\right)^{M-m}\left(\frac{1}{1+\gamma\kappa c}\right)^m\right]\end{aligned} \tag{14}$$

with

$$\kappa = K/K_f = k_a^{(f)}/k_a, \tag{15}$$

A)

```
   1 cacagaugua guuuucucug cgcgugugcg uuuucccucc ucccccgccc ucagggucca
  61 cggccacc au gg cguau uag gggcagcagu gccugcggca gcauugggcu uugcagcggc
 121 ggcagcagca ccaggcucug cagcggcaac ccccagcggc uuaagcc aug gcgcuucuca
 181 cggcauucag cagcagcguu gcuguaaccg acaaagacac cuucgaauua agcacauucc
 241 ucgauuccag caaagcaccg caac aug acc gaaaugagcu uccugagcag cgagguguug
 301 gugggggacu ugaugucccc cuucgacccg ucggguuugg gggc uga aga aagccuaggu
 361 cucuuagaug auuaccugga gguggccaag cacuucaaac cucauggguu cuccagcgac
 421 aaggcuaagg cgggcuccuc cgaauggcug gcuguggaug gguuggucag ucccuccaac
 481 aacagcaagg aggaugccuu cuccgggaca gauuggaugu uggagaaaau ggauuugaag
 541 gaguucgacu uggaugcccu guuggguaua gaugaccugg aaaccaugcc agaugaccuu
 601 cugaccacgu uggaugacac uugugaucuc uuugcccccc uaguccagga gacuaauaag
 661 cagccccccc agacggugaa cccaauuggc caucucccag aaaguuuaac aaaacccgac
 721 caggguugccc ccuucaccuu cuuacaaccu cuucccccuuu ccccaggggu ccugucccuc
 781 acuccagauc auucccuuag uuuagagcug ggcagugaag uggauaucac ugaaggagau
 841 aggaagccag acuacacugc uuacguugcc augaucccuc agugcauaaa ggaggaagac
 901 acccccuucag auaaugauag uggcaucugu augagcccag aguccuaucu gggguccccu
 961 cagcacagcc ccucuaccag gggcucucca aauaggagcc ucccaucucc agguguucuc
1021 ugugggucug cccgucccaa accuuacgau ccuccuggag agaagauggu agcagcaaaa
1081 guaaagggug agaaacugga uaagaagcug aaaaaaaugg agcaaaacaa gacagcagcc
1141 acuagguacc gccagaagaa gagggcggag caggaggcuc uuacugguga gugcaaagag
1201 cuggaaaaga agaacgaggc ucuaaaagag agggcggauu cccuggccaa ggagauccag
1261 uaccugaaag auuugauaga agaggucccgc aaggcaaggg ggaagaaaag gguccccuag
1321 uugaggauag ucaggagcgu caaugugcuu guacauagag ugcuguagcu gugguucca
1381 auaaauuauu uuguagggaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa
```

Converted from ATF4 Sequence with GenBank Accession No. BC008090.1

B)

Fast-loading triplets: k_a values > 10^6 $M^{-1}s^{-1}$	AUG GUG UUG
Intermediate triplet: k_a = 4.8×10^4 $M^{-1}s^{-1}$	CUG
Slow-loading triplets: k_a values < 10^4 $M^{-1}s^{-1}$	AAG ACG AGG AUA AUC AUU

Region N mRNA (88 nt)

```
gggcagcagugccuugcggcagcauug
gccuuugcagcggcggcagcagcacc
aggcucugcagcggcaacccccagcg
gcuuaagccaug
```

Region M mRNA (97 nt)

```
uggcgcuucucacggcauucagcagc
agcguugcuguaaccgacaaagacac
cuucgaauuaagcacauuccucgauu
ccagcaaagcaccgcaacaugg
```

Figure 3: (A) Complete cDNA sequence BC008090.1 [21] modified to display the RNA sequence (uracil replacing thymine). The start and stop codons of uORF1 (69–80), uORF2 (168–358) and the ATF4 mORF (274–1378) are bolded and underlined. Region $\mathcal{N}$ is highlighted in blue and includes the start codon for uORF2; region $\mathcal{M}$ is highlighted in red and includes the start codon for ATF4. ORFs were identified using the online tool ORFfinder (`https://www.ncbi.nlm.nih.gov/orffinder`). (B) (Left) Table displaying near cognates of the start codon studied by Kolitz and coworkers grouped into fast-loading triplets ($k_a > 10^6\,\mathrm{M}^{-1}\mathrm{s}^{-1}$), the intermediate CUG triplet, and slow-loading triplets ($k_a < 10^4\,\mathrm{M}^{-1}\mathrm{s}^{-1}$) [19]. (Right) the sequences for regions $\mathcal{N}$ and $\mathcal{M}$. Fast-loading triplets are highlighted in green, and the intermediate CUG triplet is underlined. The final UG of AUG does not belong to the region the start codon terminates, hence the coloring. As explained in the text, it belongs to the next region due to nucleotide-wise scanning of the sequence, thus the UG at the beginning of region $\mathcal{M}$.

where k_a is now the basal (slow) TC loading rate constant.

Finally, we obtain two distinct models, 2A and 2B, using the two different assumptions for the rate of uORF1 translation initiation given above.

2.3. Parameters

Values for N and M were chosen based on the *Homo sapiens* activating transcription factor 4 (tax-responsive enhancer element B67) mRNA (cDNA clone MGC:9337 IMAGE:3454473), complete cds (Genbank accession number BC008090.1) [21]. This is the same sequence that was used as a base for the work of Vattem and Wek [10]. Region $\mathcal{N}$ begins right after the stop codon of uORF1, and runs up to the point where the 40S subunit is positioned at the start codon of uORF2. In nucleotides, this region has a length $N = 88$, since the last two nucleotides of the uORF2 start codon don't count: if the 40S subunit scans past the AUG and positions itself at the UGG triplet reached after a single additional nucleotide step, it has missed the uORF2 start codon. Similar reasoning gives us $M = 97$.

Model 1 has a single parameter, γ, defined by Eq. (10), which in turn depends on $\bar{C}$ and on K. To our knowledge, no measurements of active ternary complex concentrations have been reported in the literature. Firczuk and coworkers estimated the number of eIF2 complexes in yeast cells to be approximately 29 000 using their measured abundances of eIF2β and eIF2γ [27]. Using a cytoplasmic volume of 40 fL, they calculated a concentration of eIF2 of 1.2 μM. Their extensive global expression data were used to condition a model of translation. This model gives a steady-state TC concentration of approximately 0.17 μM.

Duncan and Hershey estimated that HeLa cells contain about 2.6×10^6 eIF2 complexes based on the quantitation of two-dimensional gel electrophoresis results [31]. Since the cytoplasmic volume of a HeLa cell is 0.94 pL [32], this gives a concentration of eIF2 of about 4.6 μM, of the same order of magnitude as the concentration from yeast. Of course, we don't know how much of this complex is in ternary complexes at any given time. In the simulations of Firczuk and coworkers, approximately 14% of the eIF2 complex appears in free ternary complex [27]. Assuming a similar ratio holds in human cells, we would have a concentration of ternary complex of approximately 0.66 μM. Since we are mostly interested in human cells, this is the value we use as our estimate of $\bar{C}$.

Model 1 assumes a single value for the parameter K. Reality is more complex, the rate constant for loading the TC depending on the triplet presented by the 40S subunit [19]. Later, we will estimate the effective value of γ from experimental data. This will allow us to estimate K using our value of $\bar{C}$. For now, we consider the bounds on K suggested by available experimental data. Berthelot and coworkers measured the average scanning speed as $k_s = 10\,\mathrm{s}^{-1}$ in an *in vitro* assay [22]. Kolitz and coworkers found that the TC loading rate constant, k_a, varied from 3.3×10^3 to $5.6 \times 10^6\,\mathrm{M}^{-1}\mathrm{s}^{-1}$ [19]. Thus, using Eq. (2), the effective value of K in Model 1 should lie somewhere between the extreme values 1.8×10^{-6} and 3.0×10^{-3} M. Correspondingly, using our estimate of $\bar{C}$ and Eq. (10), $\gamma \in (2.2 \times 10^{-4}, 0.37)$.

The situation is somewhat clearer in Model 2. The data of Kolitz and coworkers essentially breaks down into three sets: fast-loading triplets, CUG as a triplet of intermediate loading rate, and slow-loading triplets [19]. They did not look at every possible triplet, but only at near-cognates of the canonical start codon. One can assume however that the slow-loading triplets are representative of typical triplets, especially since the loading rate constant for these triplets is similar to that for a 40S subunit without an mRNA. We leave the exceptional CUG triplet out of this analysis since it has a much smaller k_a than the fast-loading triplets, and assuming that its somewhat elevated value won't have a significant influence on the average over all other triplets, i.e. that it is an outlier in the set of non-fast-loading triplets with a small influence on the overall rate of loading. The three fast-loading triplets have an average value of k_a of $4.7 \times 10^6\,\mathrm{M^{-1}s^{-1}}$, implying a value of $K_f = 2.1 \times 10^{-6}\,\mathrm{M}$. The slow-loading triplets (excluding the bare 40S subunit) average $k_a = 4.4 \times 10^3\,\mathrm{M^{-1}s^{-1}}$, implying $K = 2.3 \times 10^{-3}\,\mathrm{M}$. Thus, $\gamma = 2.9 \times 10^{-4}$ and $\kappa = 1.1 \times 10^3$. From the sequence analysis presented in Fig. 3, we see that there are $n = 4$ fast-loading triplets in region $\mathcal{N}$, and $m = 2$ fast-loading triplets in region $\mathcal{M}$.

3. Results

3.1. Model 1

Figure 4 shows the rate of expression [Eq. (9)] as a function of scaled TC concentration when $\gamma = 0.37$ for both models 1A and 1B. For this value of γ, there is essentially no ATF4 expression under normal conditions ($c = 1$) in either model. As c decreases, ATF4 expression increases, goes through a maximum, and then decreases again. The eventual decrease in expression to the left of the maximum is due both to the decreased TC reloading probability with falling concentration and, in Model 1B particularly, to the falling rate of translation initiation at uORF1. Interestingly, even when ATF4 is being maximally expressed, there is greater translation activity at uORF2, although the two asymptotically tend to similar limits as $c \to 0$ (details in Maple document).

Despite the similarities, there are some important quantitative differences. Notably, Model 1B is more sensitive to falling TC concentrations in the sense that its maximum response occurs at a higher value of c, but Model 1A results in a much higher rate of expression of ATF4 at the maximum. (Note the difference in vertical scales in the two insets.) Recalling that the rates of expression we calculate are rates relative to the rate of translation of uORF1 at the unstressed TC concentration, the very low value of the rate of expression of ATF4 at the maximum in Model 1B (relative rate of less than 1%) naturally leads to the question of whether this rate would lead to sufficient expression of ATF4 to allow a robust response to stress. Model 1B is amenable to some analysis allowing us to partially answer this question. If we introduce the rescaled variable $u = \gamma c$,

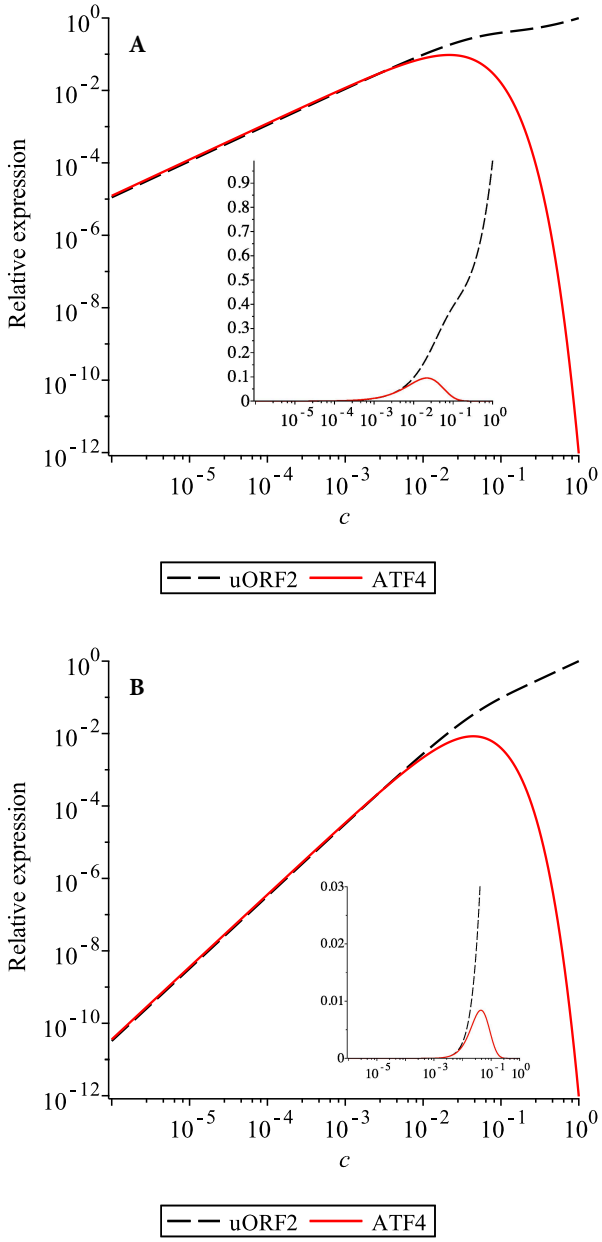

Figure 4: Relative translation rates of uORF2 and of ATF4 as a function of the scaled TC concentration at $\gamma = 0.37$, the maximum value from our parameter estimates. In both panels, the same data is plotted on log-log axes (main) and semilog axes (inset). (A) Model 1A. (B) Model 1B.

r for model 1B becomes

$$r_{1\mathrm{B}} = \frac{1}{\gamma} u \left[1 - \left(\frac{1}{1+u} \right)^M \right] \left(\frac{1}{1+u} \right)^N . \tag{16}$$

(Subscripts on r will be used henceforth to specify a particular model.) Let

$$A(u) = u \left[1 - \left(\frac{1}{1+u} \right)^M \right], \tag{17a}$$

$$B(u) = \left(\frac{1}{1+u} \right)^N . \tag{17b}$$

The following properties of these functions are trivial to prove:

1. Both functions are strictly positive for $u > 0$.
2. $A(u)$ is strictly increasing for $u \geq 0$.
3. $B(u)$ is strictly decreasing for $u \geq 0$.

Since $r_{1\mathrm{B}} = A(u)B(u)/\gamma$, and since $r'_{1\mathrm{B}} = \gamma^{-1}(A'B + AB')$, with primes indicating differentiation with respect to u, it follows that $r_{1\mathrm{B}}$ has a single maximum for $u \geq 0$. Given the form of Eq. (16), the value of u at the maximum, $u_{\max}$ can only depend on N and M. Substituting $u = u_{\max}(N, M)$ into Eq. (16), the maximum relative expression rate

$$r_{1\mathrm{B}}^{\max} = \frac{1}{\gamma} u_{\max}(N,M) \left[1 - \left(\frac{1}{1+u_{\max}(N,M)} \right)^M \right] \times \left(\frac{1}{1+u_{\max}(N,M)} \right)^N , \tag{18}$$

i.e. a function of N and M multiplied by γ^{-1}. It follows that small values of γ (smaller mean loading rate k_a) will favor higher expression levels. However, because $c_{\max}(\gamma, N, M) = u_{\max}(N, M)/\gamma$, decreasing γ will increase $c_{\max}$. In order for a falling TC concentration to result in a rising ATF4 expression, we must have $c_{\max} < 1$, i.e.

$$u_{\max}(N, M)/\gamma < 1. \tag{19}$$

There is therefore a trade-off between increasing ATF4 expression and obtaining the desired control function as c falls.

Figure 5 shows the effect of γ in both variations of Model 1. Qualitatively, the effect of γ is the same in both models, namely that decreasing γ shifts the expression maximum to higher concentrations and increases the rate of expression as predicted by the foregoing analysis. The effect on the maximum rate of expression is much more dramatic in Model 1B than in Model 1A given the former's strict proportionality between the rate of uORF1 translation initiation and c.

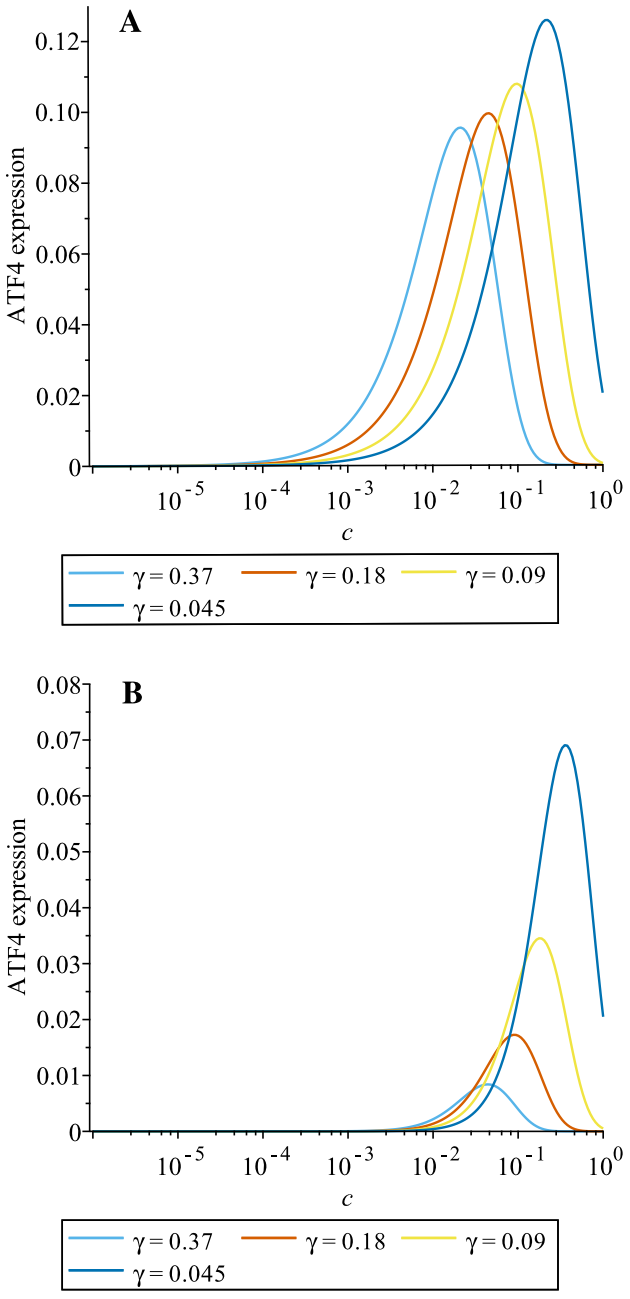

Figure 5: Relative expression of ATF4 in model 1, varying γ, with N and M corresponding to the *H. sapiens* ATF4 sequence. (A) Model 1A. (B) Model 1B.

The range of values of γ considered in Fig. 5 is roughly the range that could be considered physiologically plausible based on the location of the maximum and on the behavior near $c = 1$. If significantly smaller values of γ were considered, the maximum ATF4 expression would occur for $c > 1$, and the ATF4 expression would decrease with decreasing c for values of $c < 1$. This plausible range of values of γ is much closer to the upper limit than to the lower limit of possible γ values determined in Section 2.3. It follows that fast-loading triplets must play a preponderant role in determining the properties of this control system. This being the case, we turn immediately to Model 2, which explicitly takes the difference in loading rates between triplets into account.

3.2. Model 2

3.2.1. General behavior and a refined estimate for γ

Taking the parameters estimated in Section 2.3, we obtain the relative expression rates shown in Fig. 6. We see immediately that the parameters cannot be as we have estimated them, given that the ATF4 expression level decreases with decreasing c. We clearly need to shift the curve to the left. Model 2 has two parameters (not counting sequence-derived nucleotide counts), γ and κ. The estimate of κ derives from a single data set and should be a good estimate of the relative rate constants for TC loading. On the other hand, the estimate of γ derives from multiple sources of data, obtained under different conditions, some with substantial sources of uncertainty. Accordingly, the estimate of γ for Model 2 is at best an order-of-magnitude estimate.

As with Model 1, increasing γ shifts the ATF4 expression curves to the left (shown in the accompanying Maple document). Ideally, we would have data that gave relative levels ATF4 and TC in a specific experiment where cells had been exposed to a stress leading to phosphorylation of eIF2α. Unfortunately, to our knowledge, no such data set exists. In Vattem and Wek's paper, the authors used thapsigargin to induce endoplasmic reticulum (ER) stress [10]. Thapsigargin is used in these studies because of its high potency [33]. This suggests that the ATF4 expression in cells treated with thapsigargin might be close to maximal. In order to study the effect of eIF2α phosphorylation on the expression of ATF4, Vattem and Wek created a construct based on the ATF4 mRNA in which the coding sequence of ATF4 was replaced by the coding sequence of luciferase. This made it possible to measure expression by a simple luminescence assay. They found a low, but measurable, level of expression of luciferase even in unstressed cells [10, Fig. 2]. From their figure, we found that the level of expression of luciferase increased by a factor of approximately 5 following thapsigargin treatment. Similar expression ratios were observed in Vattem and Wek's Fig. 4, as well as in a study by Harding et al. [34]. Fixing all of the other parameters of the model, we can find the value of c at the maximum in ATF4 expression as a function of γ, $c_{\max}(\gamma)$ and then solve the equation

$$r(c_{\max}(\gamma))/r(1) = X \tag{20}$$

with $X = 5$ to obtain an improved estimate of γ. Doing this, we find $\gamma_{2A} = 1.66 \times 10^{-3}$ and $\gamma_{2B} = 2.53 \times 10^{-3}$, the subscript indicating the corresponding

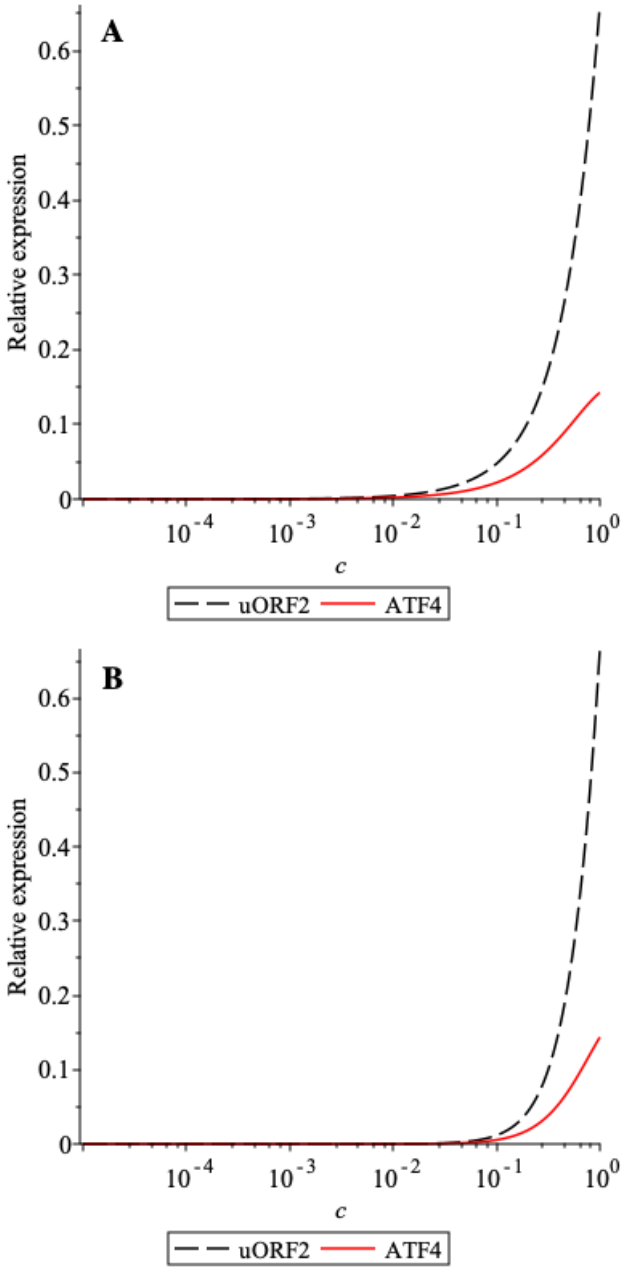

Figure 6: Relative expression of uORF2 and ATF4 in model 2 for the estimated parameters from Section 2.3. (A) Model 2A. (B) Model 2B. In both panels, dashed lines represent the relative expression of uORF2 while solid lines represent the relative expression of ATF4.

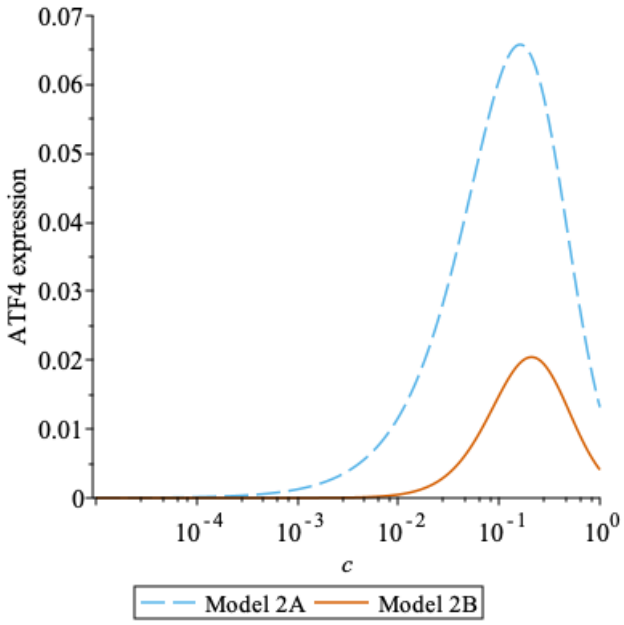

Figure 7: ATF4 expression vs TC concentration using the optimized values of γ computed as described in the text. For model 2A (dashed), $\gamma = 1.66 \times 10^{-3}$, and for model 2B (solid), $\gamma = 2.53 \times 10^{-3}$. All other parameters have the values estimated in Section 2.3.

model, which represent increases by factors of, respectively, 5.7 and 8.7 with respect to our original estimate. Given the very significant uncertainties in our original estimate of γ, this is reasonable agreement, supporting the idea that models of this class are capturing the dynamics of the ATF4 translation control system. The resulting ATF4 vs c response curves are shown in Fig. 7.

Since we do not know if thapsigargin maximally induces ATF4, the value of X could be larger than we have assumed here. Accordingly, it is useful to know how sensitive the estimate of γ is to X. We define the sensitivity coefficient of a quantity ϕ with respect to a parameter ξ by

$$S_\xi^\phi = \frac{\partial \log \phi}{\partial \log \xi} = \frac{\xi}{\phi} \frac{\partial \phi}{\partial \xi}. \tag{21}$$

Note that these logarithmic sensitivities give the fractional change in the quantity of interest for a given fractional change in the parameter [35]. By automatic differentiation in Maple (using the D operator), we find $S_X^{\gamma_{2A}} = 0.44$ and $S_X^{\gamma_{2B}} = 0.49$. These sensitivities indicate only a moderately strong dependence of γ on X. Thus, even a doubling of X would only change the value of γ by 40–50%. This suggests that our estimates of γ should be order-of-magnitude accurate, provided the true value of X is within an order of magnitude of the estimate presented above.

3.2.2. Mutants

Vattem and Wek carried out some genetic manipulations that have direct counterparts in our model. In this section, using the parameters estimated above, we verify the extent to which our model can reproduce these results.

In one set of experiments, Vattem and Wek compared the expression of ATF4 between the wild-type sequence and one in which the start codon of uORF2 had been mutated to AGG [10]. In our model, this corresponds to merging the $\mathcal{N}$ and $\mathcal{M}$ regions while losing one of the fast-loading triplets (the uORF2 start codon). Thus, this ΔuORF2 mutant would have $N = 0$, $n = 0$, $M = 88 + 97 = 185$ and $m = 4 - 1 + 2 = 5$. We immediately note an anomaly: for Model 2, when $N = n = 0$, Eq. (9) becomes, using Eq. (14),

$$r_{\Delta\text{uORF2}} = f(c)\left[1 - \left(\frac{1}{1+\gamma c}\right)^{M-m}\left(\frac{1}{1+\gamma\kappa c}\right)^{m}\right]. \tag{22}$$

Note that, since both versions of $f(c)$ considered are increasing functions of c, $r_{\Delta\text{uORF2}}$ is also clearly an increasing function of c, or conversely that it decreases with decreasing TC concentration. Thus, we would predict based on this model that thapsigargin treatment would cause a decrease in the rate of expression of ATF4 in this mutant. However, both in this mutant and in another one discussed below that also has a mutated uORF2 start codon, the experimental data show an increase in the expression of the reporter [10, Fig. 4]. It is difficult to reconcile these experimental observations with the model. We leave further discussion of this point to Section 4.

Despite this significant qualitative difference, the model and experiments agree that there is a large increase in ATF4 expression for the ΔuORF2 construct relative to the wild type. Assuming that thapsigargin treatment results in a similar change in c in either situation, we can calculate the fold-change in ATF4 expression between the wild-type and ΔuORF2 mutant, in both the untreated ($c = 1$) and thapsigargin-treated (stressed) conditions. For the latter, we evaluate the ratio of the mutant to the wild-type expression rates at $c = c_{\max}$. Figure 8 shows the results of these numerical experiments along with the experimental results. Model 2B tends to greatly overestimate the relative expression in the ΔuORF2 mutant under both conditions. Although agreement with Model 2A is not quantitative, it is much better.

In another set of experiments, a 120-nucleotide sequence lacking either "significant secondary structure" or start codons was inserted into region $\mathcal{N}$. This would increase N to $88 + 120 = 208$. While this seems straightforward on the surface, there is the issue of whether this insert included any fast-loading GUG or UUG triplets. Unfortunately, Vattem and Wek did not provide the sequence of this insert. Under unstressed conditions, this $(N + 120)$ mutant had 50% of the ATF4 expression of the wild-type [10, Fig. 4E]. If we take $n = 4$ (as in the wild-type) in Model 2A, the mutant expresses ATF4 at 82% of the wild-type rate. For $n = 5$, the ratio is 30%. The latter is a bit closer to the experimental result, so we retain this value. In Model 2B, the ratios at $n = 4$ and $n = 5$

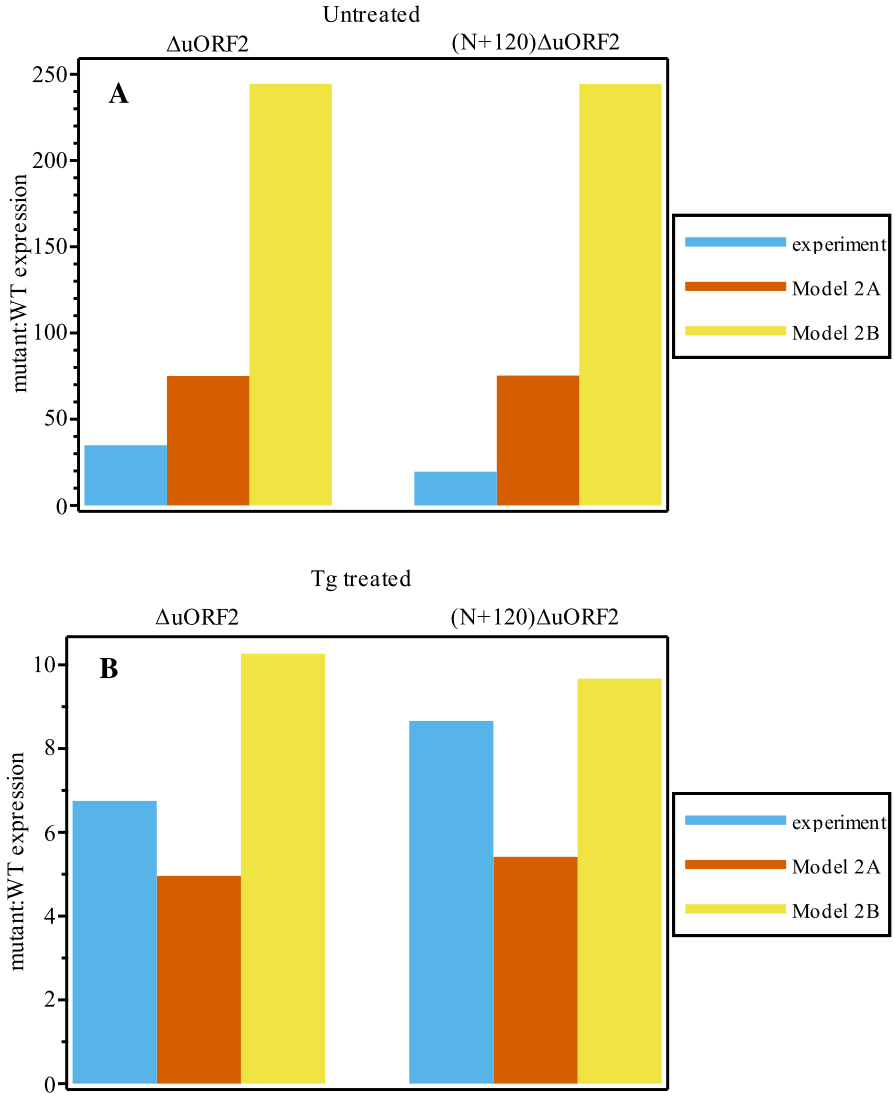

Figure 8: Expression of ATF4 in the ΔuORF2 and $(N + 120)$ΔuORF2 mutants relative to wild-type expression (A) in untreated (unstressed) conditions, and (B) with thapsigargin (Tg) treatment. The experimental results were digitized from Ref. 10, Fig. 4 using Plot Digitizer, version 2.6.9. With the exception of the parameters changed to create the mutants, model parameters are as in Fig. 7.

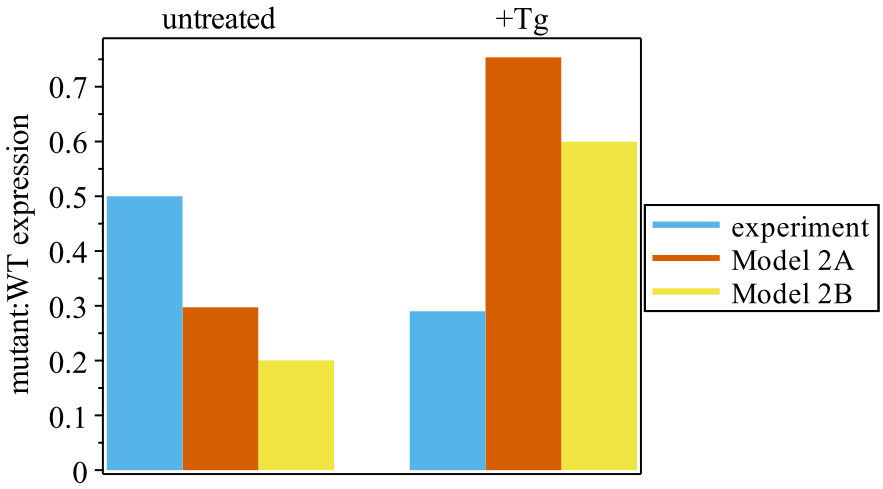

Figure 9: Expression of ATF4 in the $(N+120)$ mutants relative to wild-type expression. Data sources and treatment are as in Fig. 8.

slightly favor $n = 4$, i.e. an insert devoid of fast-loading triplets. For consistency, we take $n = 5$ in both cases.

The results for the $(N+120)$ mutants are shown in Fig. 9. Despite the value of n being chosen based on the untreated condition in Model 2A, Model 2B is somewhat better at predicting the effect of thapsigargin treatment, although both underestimate the effect of Tg by a large multiple.

In a final set of experiments, the 120 nt insert was combined with the mutated uORF2 start codon, yielding an $(N + 120)\Delta$uORF2 mutant with parameters $N = 0$, $n = 0$, $M = 208+97 = 305$ and $m = 5-1+2 = 6$. These model mutants display expression rates that are increasing functions of c for the same reason that the $(N + 120)$ mutants are. We again focused on the ratio of mutant to wild-type expression, reporting the results in Fig. 8. Model 2B somewhat overestimates this ratio in the untreated condition, but provides a much better prediction than model 2A in the thapsigargin-treated situation.

If we only compare the mutants to the wild-type at a fixed concentration, both versions of model 2 make qualitatively accurate predictions of the change in expression, appropriately predicting increases or decreases in expression. For mutants with a mutated uORF2 start codon however, the model predictions contradict the experimental data that thapsigargin increases ATF4 expression. Since a decreasing rate of ATF4 expression with decreasing TC concentration in mutants lacking a uORF overlapping with the start codon of the mORF appears to be an inevitable consequence of the control system as it is currently understood, we are at a loss to explain the experimental observations.

Having said that, we emphasize again that the two model variants qualitatively predict the effects of mutations at any given concentration and generally

show responses of the correct order of magnitude. Moreover, these results were obtained without attempting to optimize the parameters γ and κ. We used estimates of these parameters for the normally functioning control system, and modeled the mutants only by making the indicated adjustments to the sequence parameters N, M, n and m.

3.2.3. Characteristics of the control system subject to redesign

The parameters γ and κ are fixed by the interactions between the 40S subunit, the mRNA, and the ternary complex. Accordingly, we treat these as fixed quantities, retaining the value of κ estimated in Section 2.3 and the values of γ for the two models estimated above. The parameters that are subject to modification in the design of a control system based on the ATF4 5' UTR architecture are therefore N, M, n and m. There are other possibilities lying outside the scope of this model (engineering the 5' UTR in such a way as to affect the translation efficiency of uORF1, adding hairpin structures to the RNA, etc.), but we leave these for future work.

In most applications, it will likely be a desirable design feature that the mORF expression level remain low near $c = 1$ and that it rise to reasonably high concentrations under stress conditions. In Fig. 10, we present relative ATF4 (or, more generally, mORF) expression levels at $c = 1$ and at the maximum for Model 2A as a function of the engineerable parameters. The results for Model 2B are qualitatively identical, the main difference being the lower expression levels in the 2B model variation. The Maple document can be consulted for a full set of results.

It will first be noticed that N and M have modest effects on the expression of ATF4. This is of course not surprising given that these parameters designate the numbers of slow-loading triplets in regions $\mathcal{N}$ and $\mathcal{M}$, respectively. Clearly, the numbers of fast-loading triplets, n and m, matter a great deal more. Interestingly, the difference between the maximum and basal ATF4 expression levels is maximized near $n = 4$ (Fig. 10C), which is the value of n for the natural sequence. This suggests that evolution may have optimized n to maximize the amplitude of the ATF4 signal obtained in response to falling TC complex concentrations. Interestingly, the same selective pressure does not seem to be acting on m, otherwise the value of this parameter would be much larger than the $m = 2$ of the human ATF4 gene. It is possible that the amplitude of the response obtained for $m = 2$, given the value of n, is enough to activate the requisite genes, and that no selective pressure therefore exists that would sustain mutations in the sequence leading to larger values of m. Alternatively, it is possible that a higher level of expression of ATF4 would activate certain pathways too strongly or too early, notably pro-apoptotic pathways.

The amplitude of the response is not the only factor of importance. The change in the TC concentration required to reach the maximum is also a key characteristic. A designer of a synthetic gene will likely not want the maximum response to be reached at a value of c too close to a cell's normal operating conditions. Thus, Fig. 11 shows how $c_{\max}$, the value of c at which ATF4 is maximally expressed, depends on the sequence parameters for Model 2A. Results

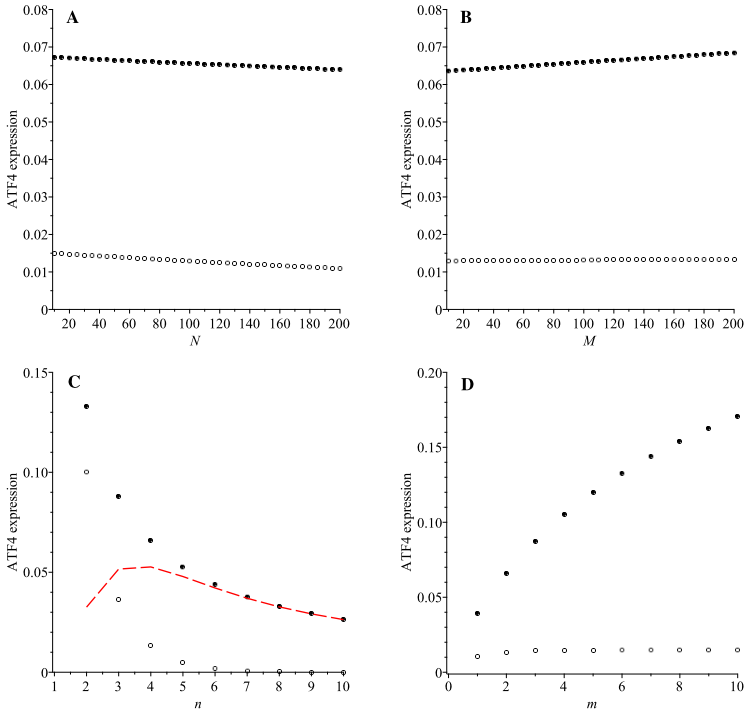

Figure 10: Relative ATF4 expression at $c = 1$ (open circles) and at the maximum (filled circles) in Model 2A plotted vs (A) N, (B) M, (C) n and (D) m. All parameters other than the one being scanned are fixed at the values for the *H. sapiens* ATF4 sequence, with γ values fixed as in Fig. 7. In panel (C), the dashed curve is the difference between the maximal and basal expression levels.

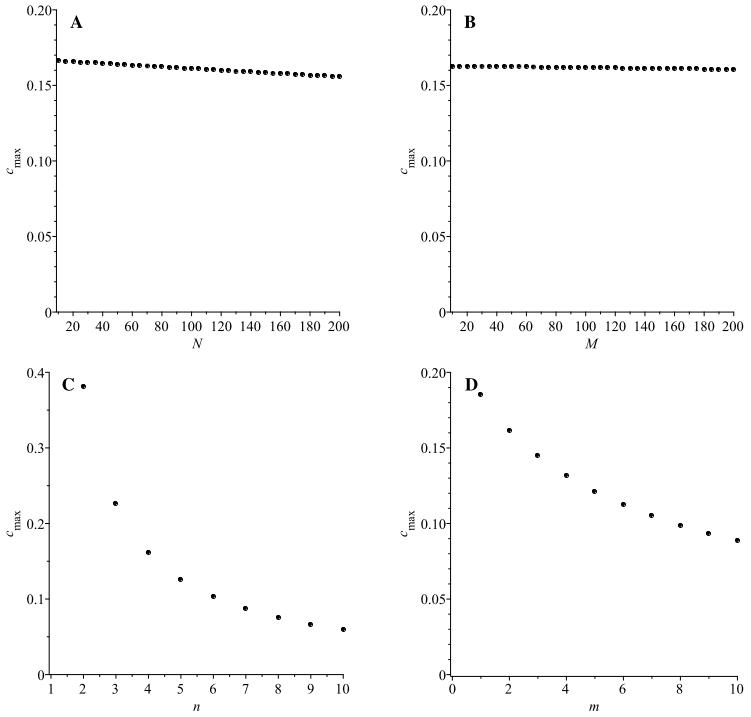

Figure 11: Relative TC concentration resulting in maximum ATF4 expression in Model 2A plotted vs (A) N, (B) M, (C) n and (D) m. All parameters other than the one being scanned are fixed at the values for the *H. sapiens* ATF4 sequence, with γ values fixed as in Fig. 7.

for Model 2B are qualitatively similar, and can be seen in the Maple document. Interestingly, increasing any of the sequence parameters tends to decrease $c_{\max}$, i.e. move it away from the unstressed value. N and M have, as expected, modest effects. Again, not surprisingly, it seems clear that a genetic engineer will want to focus on n and m.

An additional control property of interest is the initial sensitivity to a change in the TC concentration. We define

$$S_i = -\left.\frac{dr}{dc}\right|_{c=1} . \tag{23}$$

The negative sign makes S_i positive when the control system is functioning properly, i.e. when a decreasing TC concentration results in an increasing rate of expression. As might be guessed from our previous results, S_i depends very little on N or M (Maple document). It does depend more strongly on n and m, as seen in Fig. 12, which shows the dependence of the initial sensitivity on these two parameters for Model 2A. (Again, the results for Model 2B, available in our Maple document, are qualitatively similar.) The results in Fig. 12 show that a genetic engineer can make the initial sensitivity greater or lesser by varying n

or m. One might wish for a lesser value of S_i if the intention is to buffer the response against small fluctuations in c around $c = 1$. On the other hand, a value of S_i significantly different from zero would give this system the properties of a proportional controller near $c = 1$, with the value of S_i being the gain.

The full set of trade-offs involved in designing a stress-triggered control system based on the ATF4 architecture are now apparent. The behavior is relatively insensitive to N and M. The most significant design parameters are therefore n and m. As n increases, there is a rapid drop in the expression of the mORF both at the maximum and at $c = 1$ (Fig. 10C). The increase in n also causes an decrease in $c_{\max}$ (Fig. 11C), so the TC concentration has to fall farther to reach maximal mORF expression. As noted earlier, at least for the parameters considered here, the expression at $c = 1$ and at the maximum have their greatest difference near $n = 4$. If we want to protect cells with c values fluctuating near 1 from triggering of stress-induced translation, then we would favor larger values of n, which decreases S_i (Fig. 12A). However, we are limited in how much we can increase n by the drop in the level of mORF expression at the maximum.

The number of fast-loading triplets in region $\mathcal{M}$ may be a better variable to engineer: Note that the maximum mORF expression increases (Fig. 10D) and the value of $c_{\max}$ decreases as m increases. However, increasing m also increases the initial sensitivity, so the controller provides proportional control of increasing gain for larger values of m, which may or may not be desirable. It is interesting that in the natural sequence, m is just 2, even though there appear to be advantages to increasing m. The evolutionary constraint on m in the natural system may involve avoiding excessive ATF4 expression when its translation is fully induced by a stress condition. A genetic engineer need not be so constrained.

4. Discussion and conclusions

We have studied two models for the control of ATF4 translation in response to falling levels of TC, one that treats all nucleotide positions (i.e. all nucleotide triplets) as identical, while the other recognizes that some triplets facilitate binding of the TC to the 40S scanning complex. Moreover, we have considered two hypotheses for the dependence of the rate of translation initiation at uORF1, a factor that has not been considered in previous models. Estimating the single free parameter of the simple model immediately suggested that fast-loading triplets play a dominant role. We thus devoted most of our effort to a study of Model 2, which separates triplets into two sets, fast- and slow-loading. The model explains a number of properties of the natural system, although there is a discrepancy between its behavior and the observed effects of mutating the start codon of uORF2 [10]. These particular experimental results are difficult to understand. To our knowledge, these measurements have never been replicated. Replication of Vattem and Wek's experiments with the ΔuORF2 mutants would therefore be useful. If we take their results at face value, one possible explanation might be that higher TC concentrations favor termination rather than a

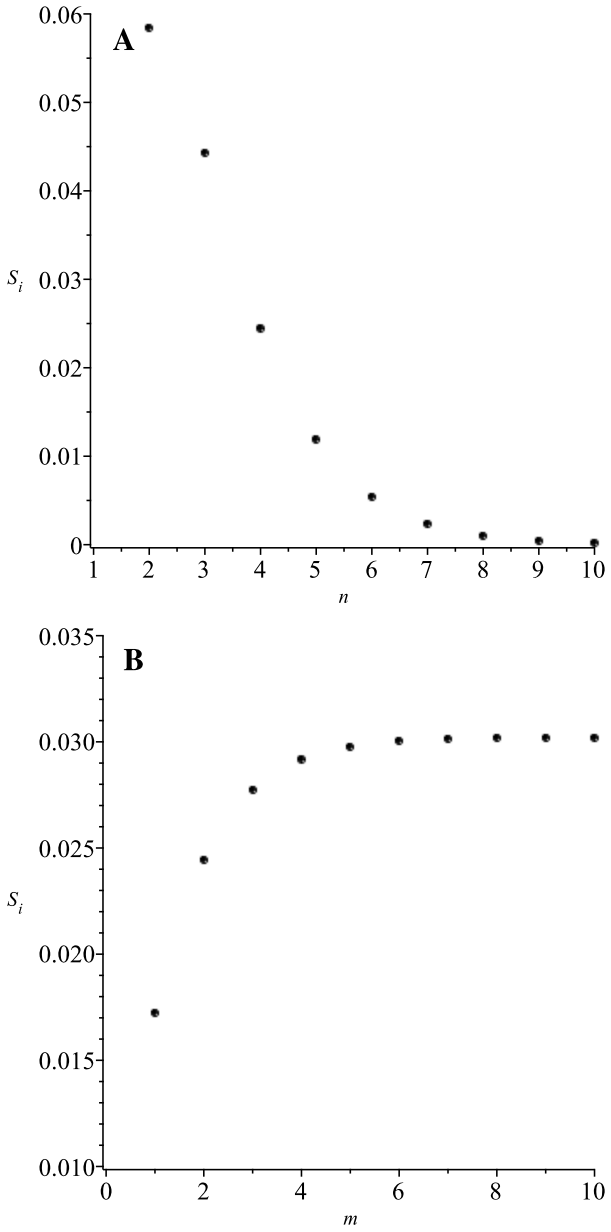

Figure 12: Initial sensitivity plotted vs (A) n and (B) m. All parameters other than the one being scanned are fixed at the values for the *H. sapiens* ATF4 sequence, with γ values fixed as in Fig. 7.

resumption of scanning following translation of uORF1, i.e. that P_{res} in Eq. (5) is not constant, but a decreasing function of c. To our knowledge, there is no direct experimental evidence for this hypothesis.

The assumed linear forms for the rate of initiation of translation at uORF1, one supported by experimental measurements of global translation rates and their dependence on the concentration of eIF2α [27], and the other based on a simple proportionality hypothesis, give qualitatively similar results. The simple proportionality hypothesis B, tends to produce lower levels of expression than hypothesis A, all other things being equal. Whether hypothesis B can produce ATF4 at a sufficient rate to account for its accumulation in cells subjected to stress depends very much on the rate at which uORF1 can maximally be expressed, as well as on P_{res}, as well as the *in vivo* half-life of ATF4. To our knowledge, these experimental measurements are not available. However, translation initiation rates in mammalian cells tend to be low, of the order of a few initiations per minute [36]. Thus, one might be tempted to think that mechanisms that favor efficient initiation at uORF1, even under conditions of low TC availability, would be favored.

Hypotheses A and B for the dependence of the rate of uORF1 translation initiation do not exhaust the possible forms of this dependence. It seems very likely that the true dependence is a non-decreasing function of c, a property shared by hypotheses A and B. However, both A and B are linear in the TC concentration, and one can imagine $f(c)$ being a nonlinear function, notably if the fall in eIF2α concentration somehow activates alternative translation initiation mechanisms. To our knowledge however, there is no evidence for such activation.

Hypotheses A and B do have one important difference that captures one important aspect of the range of behaviors that could be expected, namely that $f_A(c)$ approaches a non-zero limit as $c \to 0$, while $f_B(c) \to 0$. In the former case, the small-c behavior is mostly dependent (within a constant multiplicative factor) on the behavior of the conditional probability of translating the mORF, while in the latter case, the rate of initiation also contributes to the fall-off in the translation rate of the mORF. Clearly, time-resolved measurements of the rate of translation initiation at uORF1 and of the active TC concentration during the stress response would help clear up most of the issues raised above, and might yield new insights into the ATF4 control system, among other translation control systems involving uORFs [7, 8, 37, 38].

Analysis of Model 1B shows that there is a trade-off involving the value of γ: If we want to have a rising ATF4 concentration as the TC concentration falls, γ cannot be too small. On the other hand, smaller values of γ result in higher ATF4 expression levels. Although we have not done any analysis for these cases, we see the same trade-off at work in models 1B (Fig. 5A), 2A and 2B (Maple document). It is unlikely that any of the parameters that enter into the value of γ (Eq. (10)) can evolve strictly for the purpose of optimizing the ATF4 response as they are too central to too many other translation processes. Instead, it is the sequence parameters N, M, n and m that are subject to evolutionary optimization. Thus, in the context of Model 1B, the inequality (19) must be

read as a constraint on N and M given a value of γ. In general, the full set of sequence parameters is evolvable given the constraints imposed by the kinetic parameters of the translation machinery.

In all versions of the model studied, for realistic parameter values, ATF4 expression initially increases roughly linearly as c falls below 1 (e.g. Fig. 7), causing the system to behave like a proportional controller or, as Costa-Mattioli and Walter have described it, a rheostat [2]. Small amounts of stress cause a small increase in ATF4 expression, hopefully resulting in compensatory responses sufficient to move the cell back towards an unstressed state. It would be interesting to look at the binding affinities of ATF4 for the promoters of the genes it activates. In keeping with the rheostat model, it seems likely that the promoters with the highest affinity for ATF4 are responsible for initial, relatively mild restorative responses, while promoters with lesser affinity, which would only respond to higher concentrations of ATF4, would be responsible for more significant adaptations, as well as for, eventually, shifting the balance between anti- and pro-apoptotic proteins towards the latter.

Our analysis of Model 2 has shown that, perhaps not surprisingly, fast-loading triplets have the greatest influence on the characteristics of the ATF4 control system. Other than the start codon itself, for simplicity, we assumed that there were just two fast-loading triplets, GUG and UUG, based on the data of Kolitz et al. [19]. We neglected the effect of the intermediate-loading-rate triplet CUG, and of course we assumed that there were no other triplets that were significantly faster than the basal rate. Since Kolitz et al. only studied near-cognates of the start codon, there may well be other triplets that lead to disproportionately fast loading of the TC. Based on our results, it is possible that some of the observed effects of lengthening the region between stop and start codons, both in the Vattem and Wek study [10] and in others [18, 39] are due to the insertion of fast(er)-loading triplets, more so than to the lengthening per se. This is clearly an area in which further experimental studies would be helpful.

For the purpose of estimating κ, we took the *in vitro* data of Kolitz et al. [19] at face value. Lacking data to the contrary, we felt that this was the most sensible thing we could do. However, these assays were carried out in the absence of several key factors, notably eIF3, whose h subunit is important to reinitiation. It is possible that eIF3 facilitates loading of the TC at otherwise slow-loading triplets. If so, eIF3 might reduce the difference between fast- and slow-loading triplets, thus reducing the value of κ.

If we do assume that the data of Kolitz et al. are representative of the relative rates of TC loading *in vivo*, we should probably deal with CUG triplets separately, since they result in significantly slower loading than the fast-loading triplets, but significantly faster loading than the slow-loading codons. The ATF4 5' leader contains two CUG triplets in region $\mathcal{N}$, and one in region $\mathcal{M}$. Each detail we add increases the complexity of the model, and it may then be worth considering a model in which each triplet has a (potentially) distinct loading

probability. Thus, the probability of expressing the mORF becomes

$$\bar{P}_2 = \prod_{i \in \mathcal{N}} p_i \left(1 - \prod_{j \in \mathcal{M}} p_j \right), \tag{24}$$

where the bar distinguishes this model, Model 3, from the previous two, and p_k is the probability of scanning through nucleotide k without loading a TC. Each p_k is given by Eq. (1), but with K replaced by a local value of this parameter, K_k. The strong qualitative similarity between Models 1 and 2 suggest that this third model would at most display quantitative differences. However, these quantitative differences might still be of interest, particularly in the context of engineering gene sequences for precise responses.

One of the key issues in our models is the assumed linear forms of the dependence of the uORF1 translation initiation rate on the concentration of the TC. If, as the data of Firczuk et al. [27] suggest, the global rate of translation has a component that is directly proportional to the concentration of eIF2α, which their mass-spectrometry data also suggest is limiting in the formation of the eIF2 complex, then it is reasonable to assume that the eIF2 complex is limiting for initiation in the canonical pathway. Linearity with respect to the TC concentration is then not unreasonable. However, experiments usually measure the relative concentration of phosphorylated eIF2α, not the TC concentration. The two are nonlinearly related due to the role of the guanine nucleotide exchange factor (GEF) eIF2B [2]. eIF2B catalyzes GDP-GTP exchange in eIF2 following each cycle of translation initiation. It also forms a tight complex with phosphorylated eIF2. Thus, phosphorylated eIF2 is an inhibitor of the GEF required to recycle non-phosphorylated eIF2. Modeling these interactions, and coupling them to one of the models studied here, would allow us to make more direct use of some of the available experimental data on the expression of ATF4 and its relation to phosphorylated eIF2α levels.

A number of the comments above point to the possibility of testing our models further with additional experimental data. We also note that similar reasoning could be used to develop models for other uORF-mediated translation control systems.

The current Special Issue can be seen as encouraging reproducibility in mathematical and computational studies, on which much has been written in recent years [40–42]. Indeed, if a dynamic document containing all of the necessary mathematics, calculation methods and parameters is provided, it should be relatively easy to verify the work, and thus to reproduce it and extend it, potentially in a different environment. The latter raises the issue of interoperability, which is problematic at this time. There is, to our knowledge, no automated tool that will convert a Maple document to other symbolic algebra systems. Maple does have a tool for importing Mathematica documents. Given that Maple and Mathematica are commercial products, the lack of interoperability limits the extent to which code can be directly reused. Nevertheless, being able to at least see the code, which can be facilitated by providing a pdf of a document, but is

also possible using free "player" software provided by the vendors, increases the transparency of the research.

The premise of this Special Issue suggested by its title, Dynamic Publication Media in Mathematical Biology, is that one can write a manuscript in a computational environment and then generate a conventional manuscript from that. We are still far from this ideal. While most of these systems can generate LaTeX versions of a document, our experience has been that the LaTeX code generated is highly idiosyncratic. Moreover, the lack of any citation facilities within these environments means that, at best, an incomplete first draft would be generated. We eventually found that developing the worksheet and manuscript as two independent documents in parallel was more time-efficient than trying to use (in our case) a Maple document as a base for a manuscript. Having said that, starting our work in a Maple worksheet turned out to be an excellent brainstorming and outlining tool for the manuscript you now have in your hands.

CRediT authorship contribution statement

Olivia N. J. M. Marasco: Formal analysis, Investigation, Methodology, Writing - original draft. **Marc R. Roussel:** Conceptualization, Formal analysis, Methodology, Validation, Supervision, Funding acquisition, Writing – review & editing. **Nehal Thakor:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

MRR gratefully acknowledges Kamiko Bressler for bringing this problem to his attention. We also thank Elizabeth Trofimenkoff and Keiran Vanden Dungen for their assistance. This work was supported by the Natural Sciences and Engineering Research Council of Canada, by the University of Lethbridge Research Fund, and by an Institute Community Support Undergraduate Studentship to ONJMM from the Canadian Institutes of Health Research (Institute of Genetics).

References

[1] M. Holcik, N. Sonenberg, Translation control in stress and apoptosis, Nat. Rev. Mol. Cell Biol. 6 (2005) 318–327. doi:10.1038/nrm1618.

[2] M. Costa-Mattioli, P. Walter, The integrated stress response: From mechanism to disease, Science 368 (2020) eaat5314. doi:10.1126/science.aat5314.

[3] C. R. Singh, T. Udagawa, B. Lee, S. Wassink, H. He, Y. Yamamoto, J. T. Anderson, G. D. Pavitt, Change in nutritional status modulates the abundance of critical pre-initiation intermediate complexes during translation initiation *in vivo*, J. Mol. Biol. 370 (2007) 315–330. doi:10.1016/j.jmb.2007.04.034.

[4] K. Pakos-Zebrucka, I. Koryga, K. Mnich, M. Ljujic, A. Samali, A. M. Gorman, The integrated stress response, EMBO Rep. 17 (2016) 1374–1395. doi:10.15252/embr.201642195.

[5] M. Holcik, Could the eIF2α-independent translation be the Achilles heel of cancer?, Front. Oncol. 5 (2015) 264. doi:10.3389/fonc.2015.00264.

[6] D. K. Sharma, K. Bressler, H. Patel, N. Balasingam, N. Thakor, Role of eukaryotic initiation factors during cellular stress and cancer progression, J. Nucleic Acids 2016 (2016) 8235121. URL: https://www.hindawi.com/journals/jna/2016/8235121/. doi:10.1155/2016/8235121.

[7] A. G. Hinnebusch, I. P. Ivanov, N. Sonenberg, Translational control by 5'-untranslated regions of eukaryotic mRNAs, Science 352 (2016) 1413–1416. doi:10.1126/science.aad9868.

[8] D. R. Morris, A. P. Geballe, Upstream open reading frames as regulators of mRNA translation, Mol. Cell. Biol. 20 (2000) 8635–8642. doi:10.1128/MCB.20.23.8635-8642.2000.

[9] S. E. Calvo, D. J. Pagliarini, V. K. Mootha, Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 7507–7512. doi:10.1073/pnas.0810916106.

[10] K. M. Vattem, R. C. Wek, Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 11269–11274. doi:10.1073/pnas.0400541101.

[11] M. Kozak, Constraints on reinitiation of translation in mammals, Nucleic Acids Res. 29 (2001) 5226–5232. doi:10.1093/nar/29.24.5226.

[12] I. M. N. Wortel, L. T. van der Meer, M. S. Kilberg, F. N. van Leeuwen, Surviving stress: Modulation of ATF4-mediated stress responses in normal and malignant cells, Trends Endocrinol. Metab. 28 (2017) 794–806. doi:10.1016/j.tem.2017.07.003.

[13] J. P. Ferreira, W. L. Noderer, A. J. Diaz de Arce, C. L. Wang, Engineering ribosomal leaky scanning and upstream open reading frames for precise control of protein translation, Bioengineered 5 (2014) 186–192. doi:10.4161/bioe.27607.

[14] J. D. Beck, D. Reidenbach, N. Salomon, U. Sahin, Ö. Türeci, M. Vormehr, L. M. Kranz, mRNA therapeutics in cancer immunotherapy, Mol. Cancer 20 (2021) 69. doi:10.1186/s12943-021-01348-0.

[15] Y. Liu, M. Wang, A. Cheng, Q. Yang, Y. Wu, R. Jia, M. Liu, D. Zhu, S. Chen, S. Zhang, X.-X. Zhao, J. Huang, S. Mao, X. Ou, Q. Gao, Y. Wang, Z. Xu, Z. Chen, L. Zhu, Q. Luo, Y. Liu, Y. Yu, L. Zhang, B. Tian, L. Pan, M. U. Rehman, X. Chen, The role of host eIF2α in viral infection, Virol. J. 17 (2020) 112.

[16] T. You, I. Stansfield, M. C. Romano, A. J. P. Brown, G. M. Coghill, Analysing GCN4 translational control in yeast by stochastic chemical kinetics modelling and simulation, BMC Syst. Biol. 5 (2011) 131. doi:10.1186/1752-0509-5-131.

[17] C. M. Grant, P. F. Miller, A. G. Hinnebusch, Requirements for intercistronic distance and level of eukaryotic initiation factor 2 activity in reinitiation on *GCN4* mRNA vary with the downstream cistron, Mol. Cell. Biol. 14 (1994) 2616–2628.

[18] B. Roy, J. N. Vaughn, B.-H. Kim, F. Zhou, M. A. Gilchrist, A. G. von Arnim, The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames, RNA 16 (2010) 748–761. doi:10.1261/rna.2056010.

[19] S. E. Kolitz, J. E. Takacs, J. R. Lorsch, Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation, RNA 15 (2009) 138–152. doi:10.1261/rna.1318509.

[20] A. G. Hinnebusch, Molecular mechanism of scanning and start codon selection in eukaryotes, Microbiol. Mol. Biol. Rev. 75 (2011) 434–467. doi:10.1128/MMBR.00008-11.

[21] R. L. Strausberg, E. A. Feingold, L. H. Grouse, J. G. Derge, R. D. Klausner, F. S. Collins, L. Wagner, C. M. Shenmen, G. D. Schuler, S. F. Altschul, B. Zeeberg, K. H. Buetow, C. F. Schaefer, N. K. Bhat, R. F. Hopkins, H. Jordan, T. Moore, S. I. Max, J. Wang, F. Hsieh, L. Diatchenko, K. Marusina, A. A. Farmer, G. M. Rubin, L. Hong, M. Stapleton, M. B. Soares, M. F. Bonaldo, T. L. Casavant, T. E. Scheetz, M. J. Brownstein, T. B. Usdin, S. Toshiyuki, P. Carninci, C. Prange, S. S. Raha, N. A. Loquellano, G. J. Peters, R. D. Abramson, S. J. Mullahy, S. A. Bosak, P. J. McEwan, K. J. McKernan, J. A. Malek, P. H. Gunaratne, S. Richards, K. C. Worley, S. Hale, A. M. Garcia, L. J. Gay, S. W. Hulyk, D. K. Villalon, D. M. Muzny, E. J. Sodergren, X. Lu, R. A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. Madan, A. C. Young, Y. Shevchenko, G. G. Bouffard, R. W. Blakesley, J. W. Touchman, E. D. Green, M. C. Dickson, A. C. Rodriguez, J. Grimwood, J. Schmutz, R. M. Myers, Y. S. N. Butterfield, M. I. Krzywinski, U. Skalska, D. E.

Smailus, A. Schnerch, J. E. Schein, S. J. M. Jones, M. A. Marra, Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899–16903.

[22] K. Berthelot, M. Muldoon, L. Rajkowitsch, J. Hughes, J. E. G. McCarthy, Dynamics and processivity of 40S ribosome scanning on mRNA in yeast, Molecular Microbiology 51 (2004) 987–1001. doi:10.1046/j.1365-2958.2003.03898.x.

[23] M. Kozak, Regulation of translation via mRNA structure in prokaryotes and eukaryotes, Gene 361 (2005) 13–37. doi:10.1016/j.gene.2005.06.037.

[24] M. Sokabe, C. S. Fraser, J. W. B. Hershey, The human translation initiation multi-factor complex promotes methionyl-tRNA$_i$ binding to the 40S ribosomal subunit, Nucleic Acids Res. 40 (2012) 905–913. doi:10.1093/nar/gkr772.

[25] M. Kozak, Pushing the limits of the scanning mechanism for initiation of translation, Gene 299 (2002) 1–34. doi:10.1016/S0378-1119(02)01056-9.

[26] L. A. Passmore, T. M. Schmeing, D. Maag, D. J. Applefield, M. G. Acker, M. A. Algire, J. R. Lorsch, V. Ramakrishnan, The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome, Molecular Cell 26 (2007) 41–50. doi:10.1016/j.molcel.2007.03.018.

[27] H. Firczuk, S. Kannambath, J. Pahle, A. Claydon, R. Beynon, J. Duncan, H. Westerhoff, P. Mendes, J. E. G. McCarthy, An *in vivo* control map for the eukaryotic mRNA translation machinery, Mol. Syst. Biol. 9 (2013) 635. doi:10.1038/msb.2012.73.

[28] A. Sendoel, J. G. Dunn, E. H. Rodriguez, S. Naik, N. C. Gomez, B. Hurwitz, J. Levorse, B. D. Dill, D. Schramek, H. Molina, J. S. Weissman, E. Fuchs, Translation from unconventional 5' start sites drives tumour initiation, Nature 541 (2017) 494–499. doi:10.1038/nature21036.

[29] J. J. D. Ho, N. C. Balukoff, G. Cervantes, P. D. Malcolm, J. R. Krieger, S. Lee, Oxygen-sensitive remodeling of central carbon metabolism by archaic eIF5B, Cell Rep. 22 (2018) 17–26. doi:10.1016/j.celrep.2017.12.031.

[30] J. A. Ross, K. R. Bressler, N. Thakor, Eukaryotic initiation factor 5B (eIF5B) cooperates with eIF1A and eIF5 to facilitate uORF2-mediated repression of ATF4 translation, Int. J. Mol. Sci. 19 (2018) 4032. doi:10.3390/ijms19124032.

[31] R. Duncan, J. W. B. Hershey, Identification and quantitation of levels of protein synthesis initiation factors in crude HeLa cell lysates by two-dimensional polyacrylamide gel electrophoresis, J. Biol. Chem. 258 (1983) 7228–7235.

[32] A. Fujioka, K. Terai, R. E. Itoh, K. Aoki, T. Nakamura, S. Kuroda, E. Nishida, M. Matsuda, Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes, J. Biol. Chem. 281 (2006) 8917–8926. doi:10.1074/jbc.M509344200.

[33] M. Treiman, C. Caspersen, S. B. Christensen, A tool coming of age: Thapsigargin as an inhibitor of sarco-endoplasmic reticulum Ca^{2+}-ATPases, Trends Pharmacol. Sci. 19 (1998) 131–135. doi:10.1016/S0165-6147(98)01184-5.

[34] H. P. Harding, I. Novoa, Y. Zhang, H. Zeng, R. Wek, M. Schapira, D. Ron, Regulated translation initiation controls stress-induced gene expression in mammalian cells, Mol. Cell 6 (2000) 1099–1108.

[35] B. Ingalls, Sensitivity analysis: From model parameters to system behaviour, Essays Biochem. 45 (2008) 177–193. doi:10.1042/bse0450177.

[36] X. Yan, T. A. Hoek, R. D. Vale, M. E. Tanenbaum, Dynamics of translation of single mRNA molecules in vivo, Cell 165 (2016) 976–989. doi:10.1016/j.cell.2016.04.034.

[37] K. Wethmar, The regulatory potential of upstream open reading frames in eukaryotic gene expression, Wiley Interdiscip. Rev. RNA 5 (2014) 765–778. doi:10.1002/wrna.1245.

[38] S. Suresh, B. Chen, J. Zhu, R. J. Golden, C. Lu, B. M. Evers, N. Novaresi, B. Smith, X. Zhan, V. Schmid, S. Jun, C. M. Karacz, M. Peyton, L. Zhong, Z. Wen, A. A. Sathe, C. Xing, C. Behrens, I. I. Wistuba, G. Xiao, Y. Xie, Y.-X. Fu, J. D. Minna, J. T. Mendell, K. A. O'Donnell, eIF5B drives integrated stress response-dependent translation of PD-L1 in lung cancer, Nat. Cancer 1 (2020) 533–545.

[39] M. Kozak, Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes, Mol. Cell. Biol. 7 (1987) 3438–3445. doi:10.1128/mcb.7.10.3438.

[40] E. C. Hayden, Journal buoys code-review push, Nature 520 (2015) 276–277.

[41] L. A. Barba, The hard road to reproducibility, Science 354 (2016) 142.

[42] V. Porubsky, L. Smith, H. M. Sauro, Publishing reproducible dynamic kinetic models, Brief. Bioinform. 22 (2021) 1–13. doi:10.1093/bib/bbaa152.