

Research Proposal

**Noise in translation initiation from an mRNA with a secondary structure:
computational design, statistical characteristics, and control mechanisms**

Hossein Hosseini

hosseini@uleth.ca

Alberta RNA Research and Training Institute (ARRTI)

University of Lethbridge

Abstract

In bottom-up design approaches in synthetic biology, predictable expression of each component of a synthetic circuit or network is highly important. Toward achieving this goal, synthetic biologists have developed *cis* and *trans* regulatory motifs in different steps of gene expression process. Nevertheless, designing highly reliable and robust regulatory motifs have stayed a challenge and under intense studies. The main contributor to this challenge is the stochasticity of the gene expression. Although natural circuits/networks in cells evolutionarily adapted to work in this noisy environment, synthetic circuits/networks are suffering due to our insufficient knowledge about the effect of mRNA/gene features on noise. In this work, my goal is to characterize noise generation in translation initiation phase, in particular, the mechanisms in this phase that causes bursty expression, on the molecular level. To this end, we design a library of RNA *cis*-motifs in the 5' UTR region of both bacterial and eukaryotic cells to achieve gene expression with a predictable noise level. Characterizing noise at the molecular level will help us to not only reduce noise in gene expression but also to generate noisy expression in desired levels. Despite the fact that noise is destructive and an unwanted by-product especially in synthetic systems, e.g., when trying to maintain homeostasis in a protein level, it owns functional roles in several cellular situations. Therefore having a library with predictable noise is of high interest to synthetic biologists.

1 Introduction

The gene expression, the central dogma in molecular biology, is intuitively stochastic [1–3]. This stochasticity, termed ‘gene expression noise’ is reflected in the variations of the synthesized messenger RNAs and proteins from a gene within a single cell over time and among genetically identical cells at a time point. Looking over studies on gene expression noise, one can divide them into three separate but codependent categories: (i) origin or source of noise (ii) consequences of noise (iii) controlling noise [4]. The origin or source of noise in gene expression is categorized into two classes: intrinsic and extrinsic [5, 6]. Intrinsic noise is associated with the inherent random fluctuations arising directly from the biochemical processes in the gene expression (transcription, translation, mRNA and protein degradation). The extrinsic noise results from the variations in the concentration, state, and the location of involved biomolecules such as RNA polymerases, ribosomes and other transcription/translation factors in the intercellular environment [7]. In general, noise is regarded to be an unwanted entailment and cells are evolutionarily adapted to mitigate its effect. The well-studied strategies that cells use are negative autoregulation and negative feedback loops. However, Noise can also have functionalities such as noise-regulated genetic circuits [8, 9]. For instance, it can drive phenotypic variabilities and bistabilities by which an increasing number of pathogenic organisms can survive, and function in the host cells [10, 11].

While the extrinsic noise sources usually affect all genes or most of the gene within a cell, intrinsic noise is gene-specific thus each gene has its noise characteristics [12]. To understand the total intrinsic noise generated in a single gene’s expression, one should specify the contribution of each step involved in gene expression, from the transcription of a gene to the translation of the corresponding RNA to the degradation of the RNAs and proteins. All steps of gene expression contribute to the overall generated intrinsic noise [13]. The contribution of transcription into overall noise, especially the initiation phase of transcription, is the most-studied one since the general assumption was that the variation in RNA level could explain the disparity in protein level. The stochastic kinetics of the mRNA production has been shown to follow either Poissonian mode or bursty mode [14]: The former one refers to the synthesis of RNAs with random independent transcription initiation events with a uniform distribution over time. The latter one denotes to an on-off switch form situation where a gene intermittently switches between two states: high transcription initiation activity (on state) and inactive state (off state). The bursty mode has been shown to be noisier than the Poissonian mode [14]. Promoter architecture in eukaryotes and positive supercoiling formation in bacteria is attested to be the source of this bursty mode [15, 16]. However, recent genome-wide studies revealed that the contribution of transcription to overall variation in protein level is less than 50% [17–19]. Aforementioned

indicates the importance of post-transcription events such as protein biosynthesis and degradation [20]. Translational bursting is one of these events that can generate a significant variation in protein level [21]

In this work, we will study the source of intrinsic stochasticity in protein synthesis, particularly the translation initiation phase. This phase is triggered by the assembly of subunits of a highly elaborate machine, ribosome, on a target RNA. RNA and its intrinsic properties play the pivotal role here to affect the protein synthesis. RNAs are versatile linear macromolecules that can adopt secondary structures due to the hydrogen bond formation between its complementary nucleotides. Each RNA has its own distinctive structural personality [22]. Here we aim to understand how the RNA secondary structure can contribute to noise in gene expression. Although the recent works have provided us a good understanding of the mechanistic framework, the molecular aspects of noise generation in the process of translation initiation from a structured RNA have remained less well-understood. We are specifically interested in determining the underlying molecular mechanisms of bursting in translation initiation from a structured RNA. We want to see if we can generate bursty translation initiation events with predictable burst size and frequency by only using RNA structural features.

Why do we care about bursting? Interestingly bursting is dynamic (changes in time) so it can generate a dynamic pulsatile expression at the protein level. This dynamic expression play a significant functional role in genetic circuits [23]. For example, p53 which is a master regulator has been shown to express in a pulsatile way to keep its functionality. Otherwise, it ends up with a change in the cell fate [24]. Another example is clock genes where their expression is also dynamic and rhythmic [25]. Translational bursting provides a good platform to generate a controllable dynamics in the expression level of the gene of interest. If you can control the expression dynamics of a single gene by bursting events in translation initiation phase, it provides a vast potential to design circuits and networks with interesting outputs. For example, if you put a bursty motif in a gene of a repressilator system, how will system's outcome change? Thus, bursting can open up possible intriguing questions in the domain of synthetic biology to be asked and work on. Synthetic biology demands new devices/methods on its way to design predictable or new synthetic circuits and networks to exploit the uncertainties imposed by intrinsic stochasticity. To achieve this goal, noise-oriented designs are of high priority. It indicates the vital importance of gaining knowledge about the noise characteristics of every involved gene's stochastic expression in a circuit or network.

In the following subsections, I briefly review some relevant essential topics and then I will discuss the proposed objectives.

1.1 Translation

Translation is a fundamental part of gene expression in which the protein synthesis occurs. Traditionally, it has been divided into four phases: initiation, elongation, termination and ribosome recycling. Each one of these phases is regulated by translational factors to assure an accurate synthesis of the protein. The translation initiation is the most complex and highly-regulated phase of translation where a high number of initiation factors are incorporated to recruit a mRNA molecule. In eukaryotes, it consists of five essential steps: eIF4F cap-binding complex formation on 5'-end of mRNA, 43S pre-initiation complex (PIC) formation, 5' to 3' scanning, start codon recognition and 60S ribosome subunit joining and formation of 80S ribosome ready to commence the elongation phase [26]. In prokaryotes, 30S PIC formation is cap-independent and directly bind to the translation initiation region (TIR), where the Shine-Dalgarno sequences and start codon are accessible to quickly trigger an initiation event by joining the 50S subunit and formation of the 70S ribosome. During elongation, the 80S/70S ribosome translocates along the mRNA, codon by codon, while forming a polypeptide chain. For each elongation round, the correct aa-tRNA binds to the A-site of the ribosome. Next, a peptide bond is formed, and at the end, mRNA and tRNA are translocated by one codon [27]. Upon recognition of stop codon at the end of the mRNA, the release factors are incorporated to terminate elongation so that it ends up releasing a newly formed protein. During the recycling step, the mRNA is discharged, and the ribosome is separated into its components, ready to be used for another round of translation.

1.1.1 Translation initiation mechanisms in eukaryotic cells

The 5' untranslated region of the eukaryotic mRNA plays a crucial role in translation initiation mechanism [28]. Figure 1 shows the sequential steps from ribosome recycling (after a round of translation) to 80S complex formation. Ribosomal recycling refers to dissociation of the ribosome subunits, several translation factors and the tRNA from the mRNA after recognition of the stop codon (UGA) by release factors. Translation initiation starts by the assembly of the 43S pre-initiation complex (PIC). This assembly, triggered by incorporating several stimulatory factors such as eIF3, eIF1A, eIF1, contains eIF2 tertiary complex (met-tRNA_i·GTP·eIF2). After the 43S complex formation, it binds to the capped 5' end of the mRNA. The formation of the eIF4F complex facilitates this binding process (see figure 1) on capped 5' end and also the poly(A) binding protein (PABP). After ribosome attachment to the mRNA, it starts scanning from 5' end to the 3' end along the 5' UTR until its complimentary anti-codon recognizes the start codon (AUG) on the met-tRNA_i. This event is followed by hydrolysis of the eIF2-bound GTP and release of the phosphate (P_i) leading to the formation of the 48S initiation complex. The next step is the 60S

ribosome subunit joining which is catalyzed by another initiation factor (eIF5B), followed by the release of all the initiation factors and formation of 80S ribosomal complex ready to start the elongation then termination phases. [28].

Translation initiation is a highly regulated phase of translation, and different regulatory mechanisms can control it [29, 30]. Two sets of cis-regulatory elements have been recognized on the 5' untranslated region of RNAs: (1) sequence-specific and (2) structural motifs [31]. Sequence-specific regulatory elements such as alternative start codons upstream of the main open reading frame (mORF) can provide an alternative regulatory mechanism leading to expression of protein isoforms with different functionality. Also, multiple upstream open reading frames can promote or inhibit translation of the main ORF. Regulatory structures such as the general secondary structures and G-quadruplexes reduce the translation efficiency and can be unfolded/ linearized by stimulating eIF4A helicase activity [32–34]. Another well-studied structural element which is present in the 5' UTR and the intergenic region of some cellular RNAs and most of the viral RNAs is called the internal ribosome entry site (IRES). This highly structured element can recruit 40S ribosomal subunit directly on 5' UTR, mostly by the aid of canonical initiation factors and the IRES trans-acting factors (ITAFs). However, there are cases that no initiation factor or ITAF is required to recruit ribosome.

1.1.2 Translation initiation mechanisms in bacteria

Although the translation process is similar in both eukaryotes and bacteria, there are key differences between them:

- (i) Translation in bacteria is coupled with transcription. In fact, the translation of the mRNA by ribosome occurs while it is getting transcribed by an RNA polymerase [35]. In contrast, in eukaryotes, the transcribed mRNA is first exported from nucleus to cytoplasm, and it is translated in the cytoplasm.
- (ii) The initiating amino acids are different (N-formylmethionine in bacteria and methionine in eukaryotes).
- (iii) The regulatory processes in eukaryotes are more complex than in bacteria [36].

The final goal during translation initiation phase in bacteria is to position the initiator tRNA ($fMet-tRNA^{fMet}$) over the start codon of the mRNA [37]. Aforementioned requires the incorporation of ribosomal subunits, three initiation factors, and the initiator tRNA. To this end, translation initiation starts by the assembly of 30S initiation complex (30S IC) on mRNA 5' end. To form this complex, first, 30S ribosome subunit binds to the IF3 forming 30S-IF3 complex. Then, this complex plus IF1 and IF2 (a GTPase) binds to the mRNA and form the 30S IC. The IF2 promotes the joining of large ribosome subunit (the 50S) to the 30S IC so forming the 70S IC, occurring with

the IF3 release. The GTP hydrolysis by IF2 and release of phosphate causes movement of the initiator tRNA into the peptidyl-transferase center of the 70S ribosome [37].

Before the codon-anti-codon matching, the 30S IC goes through several interactions with mRNAs 5'UTR sequences [38]. Initially, 30S IC is assembled on the single-stranded region upstream of the start codon termed standby sites [39]. That occurs because of the electrostatic interaction between positively-charged domains of the ribosome and negatively-charged mRNA's phosphate backbone. Next, the 30S IC slides through the mRNA by strand displacement mechanism until it recognizes the Shine-Dalgarno (SD) sequence (known as ribosome binding sites (RBS)) which binds with a high affinity to 9 nt-length anti-SD sequences at the 3' end of the 16S rRNA within the 30S IC [40]. Concurrently, the fMet-tRNA performs the base-pairing with the start codon downstream the SD sequence. The region between the SD sequence and the start codon which is crucial for an efficient translation is termed "spacing" region. Note that during sliding process, the 30S IC unfolds the structured RNA regions because of its helicase activity [41]. The structured RNAs play an inhibitory role here and sometimes a decisive regulatory role when it occurs along the RNA, which is addressed in the following subsection.

1.1.3 The role of RNA secondary structures in translation process

While RNAs carry information on their primary linear sequences of nucleotides, they can form secondary and tertiary structures by pairing with local and distant nucleotides. Examples are hairpins (stem loops), pseudoknots, G-quadruplexes and more complex structures such as riboswitches and Internal ribosome entry sites (IRESes). These structures play a significant role on about every step of gene expression [42] and highly dynamic influenced by environmental signals and molecular effectors such as metabolites (small molecules), specific protein binders, chaperones, helicases, RNA cleavage, and thermally and mechanically induced changes [43]. Equilibrium fluctuations over the RNA structures implemented by various effectors can overcome the energy barriers between different conformations of the RNA and leads to transitions to a favorable conformation. One example of interest is the translation-induced unfolding of hairpin structures in 5' UTR of the RNA by eIF4A among others.

The formation of stable structures within sequences of an RNA can mediate the translation process. Other than IRESes, discussed in next subsection, which can initiate translation independent of 5' cap, the other structural elements such as step-loops and G-quadruplexes can affect the elongating ribosomal machinery [28]. The regulatory effects of these structures depend on the thermodynamic stability and their position concerning the 5' cap, the ribosomal binding site (RBS) or the start codon [44–47].

In case of stem-loop structures, Kozak showed that placing a stem-loop with $\Delta G = -30$ kcal/mol at position +12 downstream the cap can inhibit the 43S PIC binding but if one place it at position +52, the PIC, once bound, can penetrate the stem-loop [47]. She also showed that a stem-loop with $\Delta G = -72$ kcal/mol could not be unwound by the PIC in the 5' UTR while 80S ribosome can do it at elongation phase. Therefore, it is clear that if you disrupt these structures, the translation efficiency will increase [48]. However, in a recent work by Jagodnik *et al.*, they showed that placing a stem-loop structure within coding sequences near the start codon of the *E. coli* *fepA* mRNA can activate initiation events [49]. These structures impede the movement of the PIC along with the mRNA since they act as a thermodynamic barrier. For stem-loops with low stabilities, the DEAD-box RNA helicase eIF4A in conjunction with eIF4B and eIF4H can unwind the secondary structures so the PIC can bind and scan the mRNA [50, 51]. Other RNA helicases can perturb the secondary structures and increase the level of translation in different situations such as DHX29, DHX9, DHX3 [52–54]. In contrast, there are situations where binding of regulatory proteins to a stem-loop can inhibit the translation process. The stem-loop, termed iron response element (IRE), is one of the well-studied cases of this type where proteins can bind to this element blocking the loading of the PIC on the cap [55, 56].

The G-quadruplexes refers to the formation of four-stranded secondary structures from the guanine-rich sequences of an RNA [57]. These structures are thermodynamically stable and highly conserved [58, 59]. Within the structure, each of four guanine bases makes a cyclic planar arrangement by eight hydrogen bonds called a G-quartet. The presence of a cation stabilizes the G-quartets, mostly K^+ in its center [60] and stacking these planner G-quartets makes the G-quadruplexes. The presence of G-quadruplex structures in 5' UTR of a mRNA can inhibit the translation initiation process [61].

Another highly structured RNA motif identified in the 5' UTR region of some viruses and eukaryotic RNAs, known as IRESes (Internal Ribosome Entry Sites) IRESs, are found in some viral RNAs [62] and eukaryotic RNAs [63]. They can initiate translation in a cap-independent way bypassing the need for the transcription initiation factor, eIF4E, that binds to the cap structure at the 5' end. Viral IRESs have been broadly studied and characterized, and that is because they share structural and sequence-based similarities. In contrast, cellular IRESs share no detectable conserved sequences. Based on the number of factors required for translation initiation by IRESs, they have been classified into four groups based on their mechanisms and the needed factors to recruit the ribosome machinery. One of well studied IRES RNAs is the RNA of the cricket paralysis virus. The RNA of this virus contains two IRES

elements: One at 5'-end of the RNA ($\text{IRES}_{5'\text{end}}$) and the other one in the intergenic region of the RNA (IRES_{IGR}). A recent study showed that the $\text{IRES}_{5'\text{end}}$ is a type III IRES so in need of the eukaryotic translation initiation factors (eIFs) to initiate translation [64]. However, the second IRES element in the intergenic region of this is not in need of any eIFs, and the initiation occurs in a non-AUG codon without methionine initiator tRNA which is essential in canonical initiation [65–67]. The recent structural and biochemical research have provided valuable insight on mechanism, function, and structure of IRES_{IGR} . IRES_{IGR} contains 190 nucleotides and forms a compact structure upstream the start codon GCU (Alanine) [68].

1.2 Bursting: the main source of large variations in gene expression

The number of synthesized RNAs from a gene and produced proteins from a single mRNA are subject to variations during time (expression noise) and among isogenic cells. These variations are originated from different sources. Among them, bursting is considered to be one of the main sources of variation/noise. Bursting refers to the production of both RNAs and proteins in short periods of time in a pulsatile form (see figure 5 b). Essentially, three parameters are used to characterize this phenomenon: burst size, duration, and frequency. These parameters have been shown to be regulated by different mechanisms. For example, the cis-regulatory elements such as TATA box strength in transcription can tune the burst size and nucleosome occupancy in transcription or ribosomal binding site (RBS) availability in translation can tune the burst frequency [69].

Bursting the in transcription level has been observed in both prokaryotes and eukaryotes. In Bacteria, studies show that bursting arises from the formation of supercoiling structures ahead of elongating RNA polymerase [15, 70] while in yeast and higher eukaryotes, it is mainly originated from alterations in nucleosome positioning and occupancy (switching between active and repressed promotor) [71]. In the translational level, bursting is also reported in both prokaryote and eukaryote systems [12, 72].

1.2.1 Stochastic modeling of bursting in gene expression

Modeling approaches have always provided valuable insights to understanding underlying dynamics and kinetics of gene expression. From early works by McAdams and Arkin where they theoretically analysed stochasticity in prokaryotic gene expression and predicted that proteins are produced in short bursts in random time intervals [3] till now, people have used different modelling approaches to understand stochasticity (noise) in gene expression from the scale of a single gene with its cis- and trans-regulatory effectors to genetic circuits to complex regulatory networks. The

interplay between different cellular events such as physiological changes (e.g., cell growth rate and cell size) [73], signaling stimuli and noise in gene expression has been of interest of modeling because of the massive importance of gene expression noise. Depending on the question one can ask and the level of complexity of the system, people have used either coarse-grained or fine-grained approaches. In case of gene expression, as it is comprised of roughly thousands of chemical reaction, people try to break the whole system into a few critical components as shown in figure 5 a.

2 Research objectives and approach

The overall objective of this study is to characterize the molecular origin of intrinsic noise in the translation of RNAs whose secondary structures affect the rate of translation initiation, and after that to engineer the translation initiation phase to control noise dynamics in gene expression. To be specific, we want to understand the molecular mechanisms of how translational bursting can rise from structured RNAs and how bursting can affect the gene expression dynamics. Moreover, we want to analyze if we can control gene expression dynamics by modulating burst characteristics (amplitude and frequency) in the translation initiation level. To address these objectives, I propose to use a combination of *in vitro* analysis and stochastic modeling approaches.

2.1 Molecular mechanisms of translational bursting from structured RNAs in bacteria

Recently it has been shown that structured RNA folding kinetics and ribosome binding rate collectively can control translation initiation via a mechanism called ‘ribosomal drafting’ [38]. This mechanism occurs when RNA refolding rate after a round of initiation is less than ribosome’s binding rate so that successive ribosomes can bind to already unfolded mRNA and initiate. So they showed that slow-folding RNAs could result in higher rates of gene expression than fast-folding RNAs. I propose that the interplay between ribosome binding rate and RNA structure refolding rate can be used as a controllable and predictable strategy to trigger a predictable bursty behavior in the translation initiation phase. Therefore I aim to look into this specific triggering event in translation initiation and see how the underlying molecular mechanisms proceed. Also, I want to understand how this bursting dynamics at this level can affect the dynamics of protein level. Since they are occurring at different time-scales, it might be interesting to see the consequences of this delay in a single RNA level.

To investigate this objective, I propose that as a first step I will identify mRNAs with sequence-structure features that have the potential to generate bursty initiation events. Stem-loops and G-quadruplex structures in the 5’ end

of the mRNA will be the focus of this identification process. Natural and synthetic RNAs will be analyzed by kinetic programs such as Kinfold [74] and RBS calculators [75] to simulate RNA structural dynamics and translation initiation rate. *In vitro* experiments will be used to test and validate the RNAs. The smFRET microscopy and flow cytometry will be used to quantify molecular mechanisms of the bursting dynamics in translation initiation and protein population level respectively. Finally, stochastic modeling and simulation will be used in parallel to corroborate or support the experimental results.

a. Structured RNA library for prokaryotic gene expression

In bacteria, the Shine-Dalgarno (SD) sequence in the 5' UTR of the RNA is a key element that can determine the translation initiation efficiency. Therefore, to tune/adjust the initiation efficiency, you can either mutate the nucleotides of SD sequence or conceal it on a secondary structure like a stem-loop to restrict its accessibility to ribosomes. In recent years, translation initiation rate calculators have been developed by using thermodynamic models based on the free energy of the interaction between the ribosome 30S and the RNA's 5' UTR to predict the translation initiation rate [76–78]. To incorporate the RNA secondary structure, they measure the minimum folding energy in its equilibrium state. However, studies showed that there are situations where the RNA can get unfolded and while it is out of equilibrium and unfolded can recruit ribosomes [38] leading to higher initiation rate than predicted by equilibrium-based models.

I propose that by utilizing RNA structure-sequence features, I build up a library of 5' UTR RNA motifs capable of generating translation bursts with predictable burst size and frequency. These standard units can be used as a functional component of a synthetic device or a system. The hypothesis to generate such events is based on two quantities: (i) the RNA refolding rate and (ii) the ribosomal binding rate. Keeping these two rates at similar levels will generate bursty events while increasing or decreasing these rates will be used as a tool to obtain different burst sizes and frequencies. Here are the criteria for designing such a library:

- RNA sequences will be used where the ribosome binding rate is of the same order as the stem-loop refolding rate. Thus, we need to choose RNAs with narrow distributions of these rates. The reason for having similar rates is the point that having imbalanced rate will end up either a non-stop succeeding initiation events or the opposite where single initiation events will occur randomly.
- Both the Shine-Dalgarno sequence and AUG start codon should be on the stem part sequestered by base-pairing. Figure 3, a shows the position of SD sequences and AUG and the variable region that will be under

mutation to get the desirable stem-loop sequences.

- The first coding sequences downstream the start codon should be fast codons to increase the elongation speed.
- The free energy of the stem-loop should be in the range of $-10\text{kcal/mol} < \Delta G < -30\text{kcal/mol}$ to provide enough inhibitory effect on ribosomes translocation along the RNA. Note that the higher ΔG does not imply slower folding/unfolding rate. As shown previously [38], two RNAs with the same ΔG can have different folding/unfolding kinetics.
- I will select the ones with shorter in length since the short-length stem-loops minimize the risk of degradation by RNases.

b. Predicting the translational bursting and noise gene expression using a stochastic kinetic model

Before starting the wet-lab work, I put forward carrying out the stochastic Gillespie simulations [79] using two stochastic kinetic models: a single-molecule model to look at translation initiation bursting from a single mRNA, and a population-level model to look at the effects on the distribution of proteins in a cell and/or the detectability of bursting in a cellular population of proteins. This strategy will help to predict the burst characteristics and also gene expression level. Figure 3,b shows the pathway from the initiation phase to termination of translation. On state is defined as when RNA stays unfolded after first initiation event so that the subsequent ribosomes can bind sequentially without spending energy to re-unfold the stem-loop structure. This cycle of initiation events (green arrows in figure 3,b) will be continued (generating a burst) until the RNA gets enough time to trigger nucleation and zipping up so getting back to its folded state. The ribosome will need to spend energy to unfold the stem-loop again (causing a delay) to start the next burst. This delay defines the burst frequency.

The Gillespie algorithm [79] is a perfect tool that will be used to not only predict the kinetics of translation initiation but predict the expression level and noise in expression level as well.

c. RNA library cloning

The library of stem-loops with their specific bursting characteristic predicted by stochastic simulation will be incorporated into the expression cassette. Figure 3,c shows a schematic view of this construct. (expression part of the plasmid). The plasmids expressing two reporter genes (CFP and RFP) will be designed in this way: One of the reporter genes will contain a stem-loop in the 5' UTR region, and another gene will be a control mRNA without a stem-loop in its 5' UTR. The stem-loop structure will be isolated by upstream standby sites and downstream

insulator sequences to prevent any unwanted secondary structures around the stem-loop. RNA sequences containing standby-stem-loop-insulator will be prepared using PCR assembly of several oligonucleotides. Then they will be replaced within 5' end of the mRNA on the plasmid by using appropriate restriction enzymes.

I will use already designed two reporter expressing plasmids in our lab. However, I need to insert aforementioned RNA sequences containing 'standby-stem-loop-insulator' between the reporter gene and the constitutive promoter sequences. The plasmids will be transformed into *E. coli* BL21-Gold(DE3) for the *in vivo* flow cytometry analysis and RNA quantification by q-PCR.

d. Flow cytometry and RT-qPCR to quantify noise in protein level

To characterize noise in protein level, the expression of the reporter genes *in vivo* in the cells containing the cloned plasmids is quantified in the steady-state condition (mid-exponential phase of growth ($OD_{600} = 0.2$)). The flow cytometry machine in our lab will be utilized by following a standard protocol [80]. This method is capable of rapidly analyzing/ quantifying the fluorescence intensity of the signal output of each cell across the population of cells and gives the fluorescence distribution. For each cloned stem-loop RNA system, which is a two-reporter construct, the fluorescence distribution of the reporters will be obtained. Then from each fluorescence distribution, the mean expression and the variance will be exploited. The expression noise for each reporter will be quantified by measuring the squared coefficient of variation (CV^2):

$$CV^2 = \frac{\sigma_p^2}{\langle p^2 \rangle} \quad (1)$$

where $\langle p^2 \rangle$ is the squared mean of the fluorescence distribution and σ^2 is the variance of the distribution. Noise in the reporter protein containing the stem-loop will be characterized by comparing to the other reporter protein lacking the stem-loop.

I also propose to measure the intrinsic and extrinsic contribution of noise for each stem-loop construct by using a two-reporter system where two different reporter proteins will be expressed from the same stem-loop in their 5'UTR RNA. These two components of the noise can be measured based on these formulas [81]:

$$\eta_{extrinsic}^2 = \frac{\langle (RFP.CFP) \rangle}{\langle RFP \rangle \langle CFP \rangle} - 1 \quad (2)$$

$$\eta_{intrinsic}^2 = \frac{\langle (RFP - CFP)^2 \rangle}{\langle RFP \rangle \langle CFP \rangle} \quad (3)$$

where the total noise can be obtained by summing these two components ($\eta_{total}^2 = \eta_{intrinsic}^2 + \eta_{extrinsic}^2$).

Along with flow cytometry, RT-qPCR will be utilized to measure the mRNA level in cells. Briefly, in this method, RNAs in the mid-exponential growth phase is extracted, followed by DNase treatment and reverse transcription. Then, qPCR is performed by an appropriate probe and primers to quantify the mRNA level [82].

e. smFRET microscopy

To understand the molecular mechanisms involved in translation initiation from a structured RNA, I propose to utilize smFRET microscopy (Figure 3,d). This technique can detect the structural changes during time (structural dynamics). It can also have the ability to detect the intermediate long/short-lasting states. This method might help to get a close insight into how RNA's overall dynamics regulate the translation initiation. To do so, we will use selected RNAs from the RNA library that can generate bursty initiation. We will design two different smFRET assays:

1. smFRET assay to detect a free RNA's structural dynamics: For the designed stem-loop RNAs, I will label stem-loop RNA from two spots on upstream and downstream of the stem-loop by a FRET pair to get information about the kinetics of folding/unfolding of the free stem-loop.
2. smFRET assay to detect the RNA structural dynamics during translation initiation events: For this assay, we will label RNA from two spots upstream and downstream of the structured RNA. The position of the two fluorophores (donor to acceptor molecules) must be in a range that FRET energy transfer from the donor to the acceptor could happen.
3. smFRET assay to detect the ribosomes dynamics during initiation from the structured UTR: for this assay, I propose to label the RNA at a single spot downstream of the structured motif and one of the ribosomal proteins close to the RNA entry channel so that it can detect each single ribosome that passes the structured motif and reaches the labeled spot. I propose placing a stop codon downstream of the structured motif at an appropriate distance to make sure that it will not affect the translation initiation events. Moreover, the labeled spot will be after the stop codon.

f. Stochastic modeling approach

To model the stochastic nature of translation and its consequences in expression level, I will use a combination of Gillespie simulation [79] and the chemical master equation (CME) treatment [83]. The goal here is to understand

how different translational bursting characteristics, generated by different RNA stem-loop 5' UTR, can affect the mean and noise level in gene expression during time and also across a population of cells.

g. Using Noise-characterized structured RNAs as a part of synthetic circuits

I propose to utilize the noise-characterized stem-loops in genetic circuits. A simple and well-studied circuit is the toggle switch [84]. This circuit incorporates two genes where each one expresses a protein that the represses the expression of the other gene. Noise in the expression level of both genes plays the main role in switching between the expression of one gene to the other one. I believe that if we insert the stem-loop RNAs into the 5' UTR of each one of these genes, we will observe different functionalities than simple on-off switch. Thus, these noise-characterized stem-loop RNAs with different burst size and frequencies will be a great tool to control the functionality of the toggle switches.

h. Translational initiation bursting from synthesized riboswitches

From a different point of view, the stem-loop at 5'UTR region of the RNA, which is of interest of this study, can be considered as a riboswitch in which the on-state (off-state) is when the stem-loop is in the unfolded state (folded state). The only difference with conventional riboswitches is the point that here the stem-loop does not need any input signal (metabolites) to switch on or off.

Here, I propose to investigate the translation initiation bursting in the RNAs containing synthetic riboswitches [?]. The experimental pathway for riboswitches will be the same as the stem-loop case. The only difference is the mechanisms of on/off switch. While the hypothesis for the stem-loop is that the sequence-structure features of the stem-loop trigger on/off switches, in riboswitch systems the off/switch is controlled by the concentration of an input signal.

i. Computationally designing 5' UTR structured RNA library for generating eukaryotic translational bursts

Eukaryotic translation follows a different mechanism than prokaryotes. There is no strong SD:aSD base-pairing in eukaryotes. Though the Kozak sequence (A or G/C rich region) upstream the AUG start codon can modulate the initiation event, no RNA:RNA base pairing has been seen between mRNA and ribosome within this region [85]. Translation initiation efficiency in this cell type mostly depends on the assembly of the 40S subunit on the 5' cap at the end of the 5' UTR and also on the secondary structure motifs along the UTR region [86]. Studies showed that 5'UTR structural stability inhibits translation initiation [87, 88]. The more stable the structural motifs in this

region, the higher the inhibitory effects [86,89]. There are a few works in eukaryotes where they have used stem-loop motifs to fine-tune gene expression. For example, it has been shown that the increasing the number of G-C pair in the stem-loop placed in the 5' UTR can decrease the expression level [90]. Also in a recent work, Weenink et al. designed a library of RNA constructs equivalent to the bacterial RBS Calculator for tuning expression in eukaryotes [91]. However, these works mostly didn't consider the effect of the non-equilibrium states or intermediate states of the structured motifs. Here I propose to incorporate the states of the structured 5' UTR during initiation to predict the initiation level and use the competitive kinetic race between ribosomal subunit binding to 5' cap and RNA unfolding/refolding to predict the dynamics of bursting or variation in translation initiation.

2.2 Noise in the translation initiation from an IRES element (CrPV IRES) in a single-molecule level

Note: I will work on IRES noise after finishing the first project (SL), so I consider this part as a side project.

To our knowledge, translational noise in non-canonical translation from an IRES structure has not been scrutinized yet. IRESes experience conformational changes during initiation [67] (figure 4). The hypothesis is that this deformation might act as an on-off switch or, in other words, it can cause bursting with its frequencies, durations, and sizes. The primary question one might ask is whether an IRES structure is capable of initiating succeeding rounds of translation or not. The immediate next interesting question is if the initiation events follow a Poissonian distribution which can be attributed to randomly independent initiation events or it is occurring in a burst-like mechanism. Answering these questions can provide an insight on how an IRES element can affect the noise in protein level.

Here, I aim to quantify the kinetics and dynamics of translation initiation from IGR IRES of the cricket paralysis virus. The experimental approaches will be the same as stem-loop explained above. I propose to use smFRET to detect structural changes in IRES structure during initiation events and also to detect succeeding translation initiation event from an IRES element. To this end, designing and synthesizing IGR IRES RNA and preparation of essential eukaryotic translation factors and ribosomes will be carried out. The other step will be integrating the IRES into appropriate plasmids or the genome for expression analysis.

A simple modeling approach to obtain the dwell time of translation initiation from an IGR IRES

If the IRES structure is under conformation changes, as a consequence, the capability of IRES to recruit the ribosomal machineries might be affected by this change. That means the initiation rate from the IRES element will be high when it is in its compact structural form, and the rate might be low when it is under non-productive conformational changes. Let's assume that we can have two conformational states, one with a high rate of ribosome recruitment (on state), and the other one with a low rate (off state).

To model initiation mechanism by IRES_{IGR} of CrPV, we need to divide it into different compartments: The direct ribosomal recruitment is the first step where the 80S ribosomal assembly binds to specific sequences on the domain I and II of the IRES. From the cryo-EM structural analysis, as shown in figure 4, during pre- and post-translocation of the domain three on the ribosomal sites the other domains do not experience noticeable changes. Thus we can assume that the next ribosome recruitment might not be affected. However, the question might be if the conformation changes of the domain three can affect the efficiency of ribosome initiation processes. That might allow us to consider still having on/off states with low and high rates. Primary translocation of IRES Domain 3 on ribosomal sites has been shown to occur at different rates. We can consider this as another compartment (figure 7). Knowing the biochemical and structural pathways during initiation phase, we can use the method of chemical master equation (CME) to get an analytical expression for the initiation rate.

2.3 RNase E scanning mechanism (side-project)

Note that I started this project with Dylan G. for the independent study course. I will continue working on RNase E project as a side project. The initial goal was to understand the kinetics of cleavage by this endonuclease using a rapid kinetic method (stopped-flow experiment). The next step on this project is investigating the molecular mechanism of cleavage using smFRET to detect possible scanning mechanism.

RNase E is an essential bacterial endonuclease with roles in the degradation of mRNAs and the processing of different RNA species such as tRNA, rRNA and sRNAs [92, 93]. RNase E forms a core of a multienzyme assembly called the RNA degradosome [94]. RNase E encoding gene (Rne) has 1061 amino acids residues where its N-terminal domain contains the catalytic domain (1-529 residues) while the C-terminal domain (530-1061) contains specific binding sites for various proteins including polynucleotide phosphorylase (PNPase), enolase, and the helicase RhIB. In vivo, RNase E exists as a tetramer composed of a dimer of dimers [95]. However, it has been designated

that each dimer of RNase E is capable of cleaving and processing RNA [96]. The catalytic domain of RNase E has been shown to be capable of cleaving and processing RNA in the absence of the C-terminal domain and is utilized to study the enzyme *in vitro* as it does not require the other components of the degradosome to be stable [97].

RprA and ArcZ are small non-coding RNAs recruited by Hfq to mRNA to perform post-transcriptional gene regulation [98, 99]. They are being used as to study the mechanism of RNase E as they have been shown to be processed to shorter stable species by the enzyme [100]. In Chao *et al.*, they performed *in vitro* experiments to show that both RNA are cleaved by RNase E and that this cleavage is dependent on the upstream uracil. Upon mutation of the uracil, the RNA molecules were no longer processed. Therefore, we are using these two RNAs as a model system to study the kinetics and mechanism of the upstream uracil dependent mechanism of RNase E RNA cleavage. RNase E catalyzes the cleavage [100]. RNase E distorts the RNA backbone that opens up these bonds and expands the angle of these two bonds allowing the water molecule to access a phosphate group to attach. Then, nucleophilic substitution happens leading to cleavage of the RNA. This distortion is necessary for recognition of the cleavage site. Besides, unpublished simulation results have suggested that RNase E may scan along an RNA molecule until it recognizes the right upstream uracil for cleavage. Once RNase E recognizes the upstream uracil the protein will stall on this site allowing for the protein to perform RNA cleavage.

Here, the question is how RNase E recognizes the cleavage site. Whether it is performing a scanning mechanism or not is of interest of this work.

3 Significance and conclusion

Though valuable works have been done to characterize the source of noise, there were mostly considered transcription as being the foremost source of noise and paid less attention to the contribution of the translation. In this work, we aim to look into noise in the initiation phase of translation as it is critical where most *cis* and *trans*-regulatory motifs are present and acting on this phase. One can think of this phase as a gate to not only control the expression level of a gene but also to regulate the noise of the expression. We argue that the structured RNA in the 5' UTR region is exceedingly contributing to the noise generation in the translation initiation. Therefore, we want to understand the interplay between structural dynamics of RNA and noise in underlying mechanism and function of translation initiation by utilizing the smFRET technique which can quantify real-time dynamics in single-molecule scale.

From a different perspective, ribosomes as complex biological machines play a significant role in paving the way for designing synthetic molecular machines. Noise characteristics of a ribosome either in its self-organization

in an away-from-equilibrium condition (cell environment) or its functionality over a structured RNA will be highly beneficial in the domain of the nanotechnology. The interaction of the substrate (here RNA) with the molecular machine in the molecular level is accompanied by stochasticity, and the strategy by which the machine is dealing with substrate's specific properties and functions with high fidelity is still under intense research. Interestingly, substrate RNA properties (sequences and the structure) can regulate ribosome's functionality, and ribosome can control its assembly and processing to interact with the RNA.

The significance of this work can be argued in several folds: (i) It will provide insight into the molecular mechanisms underpinning the translation initiation from a structured mRNA. In particular, we will understand how succeeding ribosome assemblies occur and how they slide thorough/interact with an inhibitory motif in 5' untranslated region of the mRNA at single-molecular level. (ii) A library of synthetic RNA stem-loops will be designed to generate translation initiation events with predictable burst characteristics. This library as a standard part can be used in synthetic devices as we know that well-characterized standard units with predictable functionality are significantly crucial in synthetic biology. (iii) We will get insight into the contribution of translational initiation noise in overall gene expression noise. We will shed light on how translational initiation bursting can be used as an imperative tool to control the level of noise in gene expression.

References

- [1] Minoru SH Ko. Problems and paradigms: Induction mechanism of a single gene molecule: Stochastic or deterministic? *Bioessays*, 14(5):341–346, 1992.
- [2] Ian L Ross, Catherine M Browne, and David A Hume. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunology and cell biology*, 72(2):177, 1994.
- [3] Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997.
- [4] Jonathan M Raser and Erin K O’shea. Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013, 2005.
- [5] Mukund Thattai and Alexander Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.

- [6] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.
- [7] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.
- [8] Wiep Klaas Smits, Oscar P Kuipers, and Jan-Willem Veening. Phenotypic variation in bacteria: the role of feedback regulation. *Nature Reviews Microbiology*, 4(4):259, 2006.
- [9] Jian Liu, Jean-Marie François, and Jean-Pascal Capp. Use of noise in gene expression as an experimental parameter to test phenotypic effects. *Yeast*, 33(6):209–216, 2016.
- [10] Alan Basset, Keith H Turner, Elizabeth Boush, Sabina Sayeed, Simon L Dove, and Richard Malley. Expression of the type 1 pneumococcal pilus is bistable and negatively regulated by the structural component rrga. *Infection and immunity*, 79(8):2974–2983, 2011.
- [11] Supreet Saini, Santosh Koirala, Emily Floess, Patrick J Mears, Yann R Chemla, Ido Golding, Christine Aldridge, Phillip D Aldridge, and Christopher V Rao. Fliz induces a kinetic switch in flagellar gene expression. *Journal of bacteriology*, 192(24):6477–6481, 2010.
- [12] Mads Kaern, Timothy C Elston, William J Blake, and James J Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature reviews. Genetics*, 6(6):451, 2005.
- [13] Guilhem Chalancon, Charles NJ Ravarani, S Balaji, Alfonso Martinez-Arias, L Aravind, Raja Jothi, and M Madan Babu. Interplay between gene expression noise and regulatory network architecture. *Trends in genetics*, 28(5):221–232, 2012.
- [14] Alvaro Sanchez and Ido Golding. Genetic determinants and cellular constraints in noisy gene expression. *Science*, 342(6163):1188–1193, 2013.
- [15] Shasha Chong, Chongyi Chen, Hao Ge, and X Sunney Xie. Mechanism of transcriptional bursting in bacteria. *Cell*, 158(2):314–326, 2014.
- [16] William J Blake, Mads Kærn, Charles R Cantor, and James J Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633, 2003.
- [17] Christine Vogel. Translation's coming of age. *Molecular systems biology*, 7(1):498, 2011.

- [18] Christine Vogel, Raquel de Sousa Abreu, Daijin Ko, Shu-Yun Le, Bruce A Shapiro, Suzanne C Burns, Devraj Sandhu, Daniel R Boutz, Edward M Marcotte, and Luiz O Penalva. Sequence signatures and mrna concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular systems biology*, 6(1):400, 2010.
- [19] Björn Schwankhäusser, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, 2011.
- [20] Rinat Arbel-Goren, Asaf Tal, Tamar Friedlander, Shiri Meshner, Nina Costantino, Donald L. Court, and Joel Stavans. Effects of post-transcriptional regulation on phenotypic noise in *Escherichia coli*. *Nucleic Acids Research*, 41(9):4825–4834, May 2013.
- [21] Long Cai, Nir Friedman, and X Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358, 2006.
- [22] Anthony M Mustoe, Steven Busan, Greggory M Rice, Christine E Hajdin, Brant K Peterson, Vera M Ruda, Neil Kubica, Razvan Nutiu, Jeremy L Baryza, and Kevin M Weeks. Pervasive regulatory functions of mrna structure revealed by high-resolution shape probing. *Cell*, 2018.
- [23] Joe H Levine, Yihan Lin, and Michael B Elowitz. Functional roles of pulsing in genetic circuits. *Science*, 342(6163):1193–1200, 2013.
- [24] Jeremy E Purvis, Kyle W Karhohs, Caroline Mock, Eric Batchelor, Alexander Loewer, and Galit Lahav. p53 dynamics control cell fate. *Science*, 336(6087):1440–1444, 2012.
- [25] Antje Jilg, Sandra Lesny, Natalie Peruzki, Herbert Schwegler, Oliver Selbach, Faramarz Dehghani, and Jörg H Stehle. Temporal dynamics of mouse hippocampal clock gene expression support memory processing. *Hippocampus*, 20(3):377–388, 2010.
- [26] Colin Echeverría Aitken and Jon R Lorsch. A mechanistic overview of translation initiation in eukaryotes. *Nature structural & molecular biology*, 19(6):568–576, 2012.
- [27] Thomas E Dever and Rachel Green. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harbor perspectives in biology*, 4(7):a013706, 2012.

- [28] Alan G Hinnebusch, Ivaylo P Ivanov, and Nahum Sonenberg. Translational control by 5-untranslated regions of eukaryotic mrnas. *Science*, 352(6292):1413–1416, 2016.
- [29] Nicola K Gray and Marvin Wickens. Control of translation initiation in animals. *Annual review of cell and developmental biology*, 14(1):399–458, 1998.
- [30] Virginia M Pain. Initiation of protein synthesis in eukaryotic cells. *The FEBS Journal*, 236(3):747–771, 1996.
- [31] Morgan L Truitt and Davide Ruggero. New frontiers in translational control of the cancer genome. *Nature Reviews Cancer*, 16(5):288–304, 2016.
- [32] Jerry Pelletier and Nahum Sonenberg. Insertion mutagenesis to increase secondary structure within the 5 noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell*, 40(3):515–526, 1985.
- [33] Kateryna Feoktistova, Enkhee Tuvshintogs, Angelie Do, and Christopher S Fraser. Human eif4e promotes mRNA restructuring by stimulating eif4a helicase activity. *Proceedings of the National Academy of Sciences*, 110(33):13339–13344, 2013.
- [34] Andrew L Wolfe, Kamini Singh, Yi Zhong, Philipp Drewe, Vinagolu K Rajasekhar, Viraj R Sanghvi, Konstantinos J Mavrakis, Man Jiang, Justine E Roderick, Joni Van der Meulen, et al. RNA g-quadruplexes cause eif4a-dependent oncogene translation in cancer. *Nature*, 513(7516):65–70, 2014.
- [35] A Ralston. Simultaneous gene transcription and translation in bacteria. *Nature Education*, 1(1):4, 2008.
- [36] Marilyn Kozak. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361:13–37, 2005.
- [37] T Martin Schmeing and V Ramakrishnan. What recent ribosome structures have revealed about the mechanism of translation. *Nature*, 461(7268):1234, 2009.
- [38] Amin Espah Borujeni and Howard M Salis. Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. *Journal of the American Chemical Society*, 138(22):7016–7023, 2016.
- [39] Amin Espah Borujeni, Anirudh S Channarasappa, and Howard M Salis. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic acids research*, 42(4):2646–2659, 2013.

- [40] Maarten H de Smit and Jan van Duin. Translational initiation on structured messengers: another role for the Shine-Dalgarno interaction. *Journal of molecular biology*, 235(1):173–184, 1994.
- [41] Mélodie Duval, Alexey Korepanov, Olivier Fuchsbaier, Pierre Fechter, Andrea Haller, Attilio Fabbretti, Laurence Choulier, Ronald Micura, Bruno P Klaholz, Pascale Romby, et al. Escherichia coli ribosomal protein s1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS biology*, 11(12):e1001731, 2013.
- [42] Yue Wan, Michael Kertesz, Robert C Spitale, Eran Segal, and Howard Y Chang. Understanding the transcriptome through RNA structure. *Nature Reviews Genetics*, 12(9):641, 2011.
- [43] Elizabeth A Dethoff, Jeetender Chugh, Anthony M Mustoe, and Hashim M Al-Hashimi. Functional complexity and regulation through RNA dynamics. *Nature*, 482(7385):322, 2012.
- [44] Krishnarjun Sarkar, Duc A Nguyen, and Martin Gruebele. Loop and stem dynamics during RNA hairpin folding and unfolding. *RNA*, 16(12):2427–2434, 2010.
- [45] T Glen Lawson, BK Ray, JT Dodds, JA Grifo, RD Abramson, WC Merrick, DF Betsch, HL Weith, and RE Thach. Influence of 5' proximal secondary structure on the translational efficiency of eukaryotic mRNAs and on their interaction with initiation factors. *Journal of Biological Chemistry*, 261(30):13979–13989, 1986.
- [46] Joseph A Curran and Benjamin Weiss. What is the impact of mRNA 5' UTR heterogeneity on translational start site selection and the mammalian cellular phenotype? *Frontiers in genetics*, 7:156, 2016.
- [47] MARILYN Kozak. Circumstances and mechanisms of inhibition of translation by secondary structure in eukaryotic mRNAs. *Molecular and cellular biology*, 9(11):5134–5142, 1989.
- [48] Xue-hai Liang, Hong Sun, Wen Shen, Shiyu Wang, Joyee Yao, Michael T Migawa, Huynh-Hoa Bui, Sagar S Damle, Stan Riney, Mark J Graham, et al. Antisense oligonucleotides targeting translation inhibitory elements in 5' UTRs can selectively increase protein levels. *Nucleic acids research*, 45(16):9528–9546, 2017.
- [49] Jonathan Jagodnik, Claude Chiaruttini, and Maude Guillier. Stem-loop structures within mRNA coding sequences activate translation initiation and mediate control by small regulatory RNAs. *Molecular cell*, 68(1):158–170, 2017.
- [50] George W Rogers, Nancy J Richter, Walt F Lima, and William C Merrick. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *Journal of Biological Chemistry*, 276(33):30914–30922, 2001.

- [51] Francis Robert and Jerry Pelletier. Perturbations of rna helicases in cancer. *Wiley Interdisciplinary Reviews: RNA*, 4(4):333–349, 2013.
- [52] Armen Parsyan, David Shahbazian, Yvan Martineau, Emmanuel Petroulakis, Tommy Alain, Ola Larsson, Geraldine Mathonnet, Gritta Tettweiler, Christopher U Hellen, Tatyana V Pestova, et al. The helicase protein dhw29 promotes translation initiation, cell proliferation, and tumorigenesis. *Proceedings Of The National Academy Of Sciences*, 106(52):22217–22222, 2009.
- [53] Vera P Pisareva and Andrey V Pisarev. Dhw29 reduces leaky scanning through an upstream aug codon regardless of its nucleotide context. *Nucleic acids research*, 44(9):4252–4265, 2016.
- [54] Vera P Pisareva, Andrey V Pisarev, Anton A Komar, Christopher UT Hellen, and Tatyana V Pestova. Translation initiation on mammalian mrnas with structured 5 utrs requires dhw-box protein dhw29. *Cell*, 135(7):1237–1250, 2008.
- [55] Philip Aisen, Caroline Enns, and Marianne Wessling-Resnick. Chemistry and biology of eukaryotic iron metabolism. *The international journal of biochemistry & cell biology*, 33(10):940–959, 2001.
- [56] Yoshiaki Tsuji, Hitoshi Ayaki, Susan P Whitman, Charles S Morrow, Suzy V Torti, and Frank M Torti. Coordinate transcriptional and translational regulation of ferritin in response to oxidative stress. *Molecular and cellular biology*, 20(16):5818–5827, 2000.
- [57] Hans J Lipps and Daniela Rhodes. G-quadruplex structures: in vivo evidence and function. *Trends in cell biology*, 19(8):414–422, 2009.
- [58] Andrew N Lane, J Brad Chaires, Robert D Gray, and John O Trent. Stability and kinetics of g-quadruplex structures. *Nucleic acids research*, 36(17):5482–5515, 2008.
- [59] Sunita Kumari, Anthony Bugaut, Julian L Huppert, and Shankar Balasubramanian. An rna g-quadruplex in the 5 utr of the nras proto-oncogene modulates translation. *Nature chemical biology*, 3(4):218, 2007.
- [60] James R Williamson, MK Raghuraman, and Thomas R Cech. Monovalent cation-induced structure of telomeric dna: the g-quartet model. *Cell*, 59(5):871–880, 1989.
- [61] Amit Arora, Mariola Dutkiewicz, Vinod Scaria, Manoj Hariharan, Souvik Maiti, and Jens Kurreck. Inhibition of translation in living eukaryotic cells by an rna g-quadruplex motif. *Rna*, 14(7):1290–1296, 2008.

- [62] Jerry Pelletier and Nahum Sonenberg. Internal initiation of translation of eukaryotic mrna directed by a sequence derived from poliovirus rna. *Nature*, 334(6180):320, 1988.
- [63] Dennis G Macejak and Peter Sarnow. Internal initiation of translation mediated by the 5 leader of a cellular mrna. *Nature*, 353(6339):90, 1991.
- [64] Lauriane Gross, Quentin Vicens, Evelyne Einhorn, Audrey Noireterre, Laure Schaeffer, Lauriane Kuhn, Jean-Luc Imler, Gilbert Eriani, Carine Meignin, and Franck Martin. The ires5 utr of the dicistrovirus cricket paralysis virus is a type iii ires containing an essential pseudoknot structure. *Nucleic acids research*, 45(15):8993–9004, 2017.
- [65] Joan E Wilson, Tatyana V Pestova, Christopher UT Hellen, and Peter Sarnow. Initiation of protein synthesis from the a site of the ribosome. *Cell*, 102(4):511–520, 2000.
- [66] Joan E Wilson, Marguerite J Powell, Susan E Hoover, and Peter Sarnow. Naturally occurring dicistronic cricket paralysis virus rna is regulated by two internal ribosome entry sites. *Molecular and cellular biology*, 20(14):4990–4999, 2000.
- [67] Christian MT Spahn, Eric Jan, Anke Mulder, Robert A Grassucci, Peter Sarnow, and Joachim Frank. Cryo-em visualization of a viral internal ribosome entry site bound to human ribosomes: the ires functions as an rna-based translation factor. *Cell*, 118(4):465–475, 2004.
- [68] Eric Jan and Peter Sarnow. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *Journal of molecular biology*, 324(5):889–902, 2002.
- [69] Damien Nicolas, Nick E Phillips, and Felix Naef. What shapes eukaryotic transcriptional bursting? *Molecular BioSystems*, 13(7):1280–1290, 2017.
- [70] Stuart A Sevier, David A Kessler, and Herbert Levine. Mechanical bounds to transcriptional noise. *Proceedings of the National Academy of Sciences*, 113(49):13983–13988, 2016.
- [71] Judith B Zaugg and Nicholas M Luscombe. A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome research*, 22(1):84–94, 2012.
- [72] Ertugrul M Ozbudak, Mukund Thattai, Iren Kurtser, Alan D Grossman, and Alexander Van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature genetics*, 31(1):69–73, 2002.

- [73] François Bertaux, Samuel Marguerat, and Vahid Shahrezaei. Impact of division rate and cell size on gene expression noise. *bioRxiv*, page 209593, 2017.
- [74] Michael T Wolfinger, W Andreas Svrcek-Seiler, Christoph Flamm, Ivo L Hofacker, and Peter F Stadler. Exact folding dynamics of rna secondary structures. *J. Phys. A: Math.*, 37:4731–4741, 2004.
- [75] Howard M Salis. The ribosome binding site calculator. In *Methods in enzymology*, volume 498, pages 19–42. Elsevier, 2011.
- [76] Benjamin Reeve, Thomas Hargest, Charlie Gilbert, and Tom Ellis. Predicting Translation Initiation Rates for Designing Synthetic Biology. *Frontiers in Bioengineering and Biotechnology*, 2, January 2014.
- [77] Howard M Salis, Ethan A Mirsky, and Christopher A Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nature biotechnology*, 27(10):946, 2009.
- [78] RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression | Bioinformatics | Oxford Academic.
- [79] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.
- [80] Fabien Pitoiset, Michèle Barbié, Guillaume Monneret, Cécile Braudeau, Pierre Pochard, Isabelle Pellegrin, Jacques Trauet, Myriam Labalette, David Klatzmann, and Michelle Rosenzwajg. A standardized flow cytometry procedure for the monitoring of regulatory t cells in clinical trials. *Cytometry Part B: Clinical Cytometry*, 2018.
- [81] Arren Bar-Even, Johan Paulsson, Narendra Maheshri, Miri Carmi, Erin O'Shea, Yitzhak Pilpel, and Naama Barkai. Noise in protein expression scales with natural protein abundance. *Nature genetics*, 38(6):636, 2006.
- [82] Weimin Gao, Weiwen Zhang, and Deirdre R Meldrum. Rt-qpcr based quantitative analysis of gene expression in single bacterial cells. *Journal of microbiological methods*, 85(3):221–227, 2011.
- [83] Daniel T Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, 1992.
- [84] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339, 2000.

- [85] So Nakagawa, Yoshihito Niimura, Takashi Gojobori, Hiroshi Tanaka, and Kin-ichiro Miura. Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic acids research*, 36(3):861–871, 2007.
- [86] Shlomi Dvir, Lars Velten, Eilon Sharon, Danny Zeevi, Lucas B Carey, Adina Weinberger, and Eran Segal. Deciphering the rules by which 5'-utr sequences affect protein expression in yeast. *Proceedings of the National Academy of Sciences*, 110(30):E2792–E2801, 2013.
- [87] A Robbins-Pianka, Michael D Rice, and Michael P Weir. The mrna landscape at yeast translation initiation sites. *Bioinformatics*, 26(21):2651–2655, 2010.
- [88] Michael Kertesz, Yue Wan, Elad Mazor, John L Rinn, Robert C Nutter, Howard Y Chang, and Eran Segal. Genome-wide measurement of rna secondary structure in yeast. *Nature*, 467(7311):103, 2010.
- [89] Markus Ringnér and Morten Krogh. Folding free energies of 5'-utrs impact post-transcriptional regulation on a genomic scale in yeast. *PLoS computational biology*, 1(7):e72, 2005.
- [90] Erwin Lamping, Masakazu Niimi, and Richard D Cannon. Small, synthetic, gc-rich mrna stem-loop modules 5 proximal to the aug start-codon predictably tune gene expression in yeast. *Microbial cell factories*, 12(1):74, 2013.
- [91] Tim Weenink, Robert M McKiernan, and Tom Ellis. Rational design of rna structures that predictably tune eukaryotic gene expression. *bioRxiv*, page 137877, 2017.
- [92] L Aravind and Eugene V Koonin. A natural classification of ribonucleases. *Methods in enzymology*, 341:3, 2001.
- [93] Jonathan AR Worrall and Ben F Luisi. Information available at cut rates: structure and mechanism of ribonucleases. *Current opinion in structural biology*, 17(1):128–137, 2007.
- [94] Andras Miczak, Vladimir R Kaberdin, Chia-Li Wei, and Sue Lin-Chao. Proteins associated with rnase e in a multicomponent ribonucleolytic complex. *Proceedings of the National Academy of Sciences*, 93(9):3865–3869, 1996.
- [95] Anastasia J Callaghan, Maria Jose Marcaida, Jonathan A Stead, Kenneth J McDowall, et al. Structure of escherichia coli rnase e catalytic domain and implications for rna turnover. *Nature*, 437(7062):1187, 2005.

- [96] Xunqing Jiang and Joel G Belasco. Catalytic activation of multimeric rnase e and rnase g by 5-monophosphorylated rna. *Proceedings of the National Academy of Sciences of the United States of America*, 101(25):9211–9216, 2004.
- [97] Jonathan AR Worrall, Maria Górná, Nicholas T Crump, Lara G Phillips, Alex C Tuck, Amanda J Price, Vassiliy N Bavro, and Ben F Luisi. Reconstitution and analysis of the multienzyme escherichia coli rna degradosome. *Journal of molecular biology*, 382(4):870–883, 2008.
- [98] Pierre Mandin and Susan Gottesman. Integrating anaerobic/aerobic sensing and the general stress response through the arcZ small rna. *The EMBO journal*, 29(18):3094–3107, 2010.
- [99] Kai Papenfort and Carin K Vanderpool. Target activation by regulatory rnas in bacteria. *FEMS microbiology reviews*, 39(3):362–378, 2015.
- [100] Yanjie Chao, Lei Li, Dylan Girodat, Konrad U Förstner, Nelly Said, Colin Corcoran, Michał Śmiga, Kai Papenfort, Richard Reinhardt, Hans-Joachim Wieden, et al. In vivo cleavage map illuminates the central role of rnase e in coding and non-coding rna pathways. *Molecular cell*, 65(1):39–51, 2017.
- [101] Kenneth C. Keiler. Mechanisms of ribosome rescue in bacteria. *Nature Reviews Microbiology*, 13(5):285–297, May 2015.
- [102] Jennifer S Pfingsten, David A Costantino, and Jeffrey S Kieft. Structural basis for ribosome recruitment and manipulation by a viral ires rna. *Science*, 314(5804):1450–1454, 2006.
- [103] YASUSHI KANAMORI and NOBUHIKO NAKASHIMA. A tertiary structure model of the internal ribosome entry site (ires) for methionine-independent initiation of translation. *Rna*, 7(2):266–274, 2001.
- [104] Hilda H Au, Gabriel Cornilescu, Kathryn D Mouzakis, Qian Ren, Jordan E Burke, Seonghoon Lee, Samuel E Butcher, and Eric Jan. Global shape mimicry of trna within a viral internal ribosome entry site mediates translational reading frame selection. *Proceedings of the National Academy of Sciences*, 112(47):E6446–E6455, 2015.
- [105] Martin Schüler, Sean R Connell, Aurelie Lescoute, Jan Giesebricht, Marylena Dabrowski, Birgit Schroer, Thorsten Mielke, Paweł A Penczek, Eric Westhof, and Christian MT Spahn. Structure of the ribosome-bound cricket paralysis virus ires rna. *Nature structural & molecular biology*, 13(12):1092–1096, 2006.

- [106] Israel S Fernández, Xiao-Chen Bai, Garib Murshudov, Sjors HW Scheres, and V Ramakrishnan. Initiation of translation by cricket paralysis virus ires requires its translocation in the ribosome. *Cell*, 157(4):823–831, 2014.
- [107] Margarita Muhs, Tarek Hilal, Thorsten Mielke, Maxim A Skabkin, Karissa Y Sanbonmatsu, Tatyana V Pestova, and Christian MT Spahn. Cryo-em of ribosomal 80s complexes with termination factors reveals the translocated cricket paralysis virus ires. *Molecular cell*, 57(3):422–432, 2015.

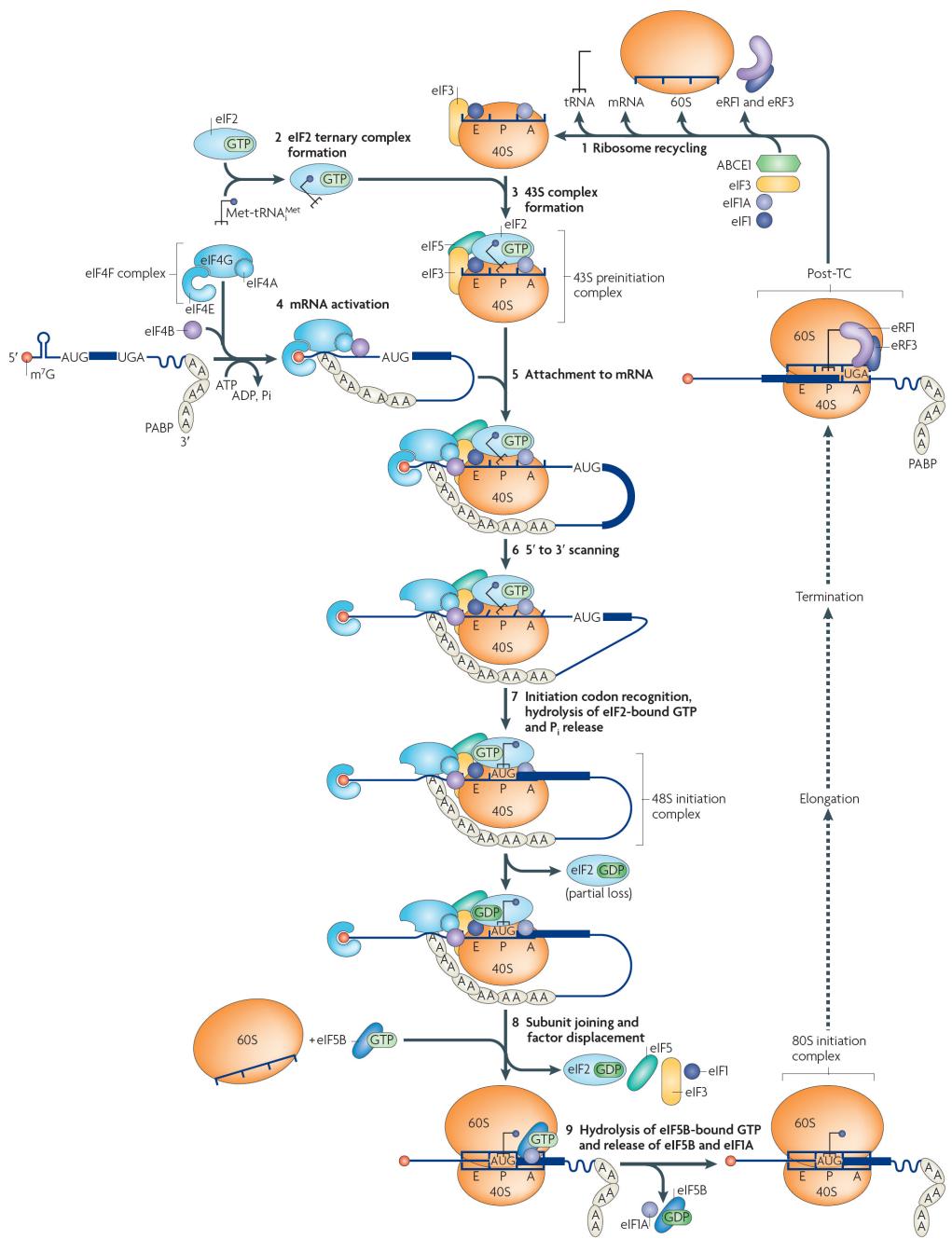


Figure 1: Canonical translation initiation from 5' cap in eukaryotes [28].

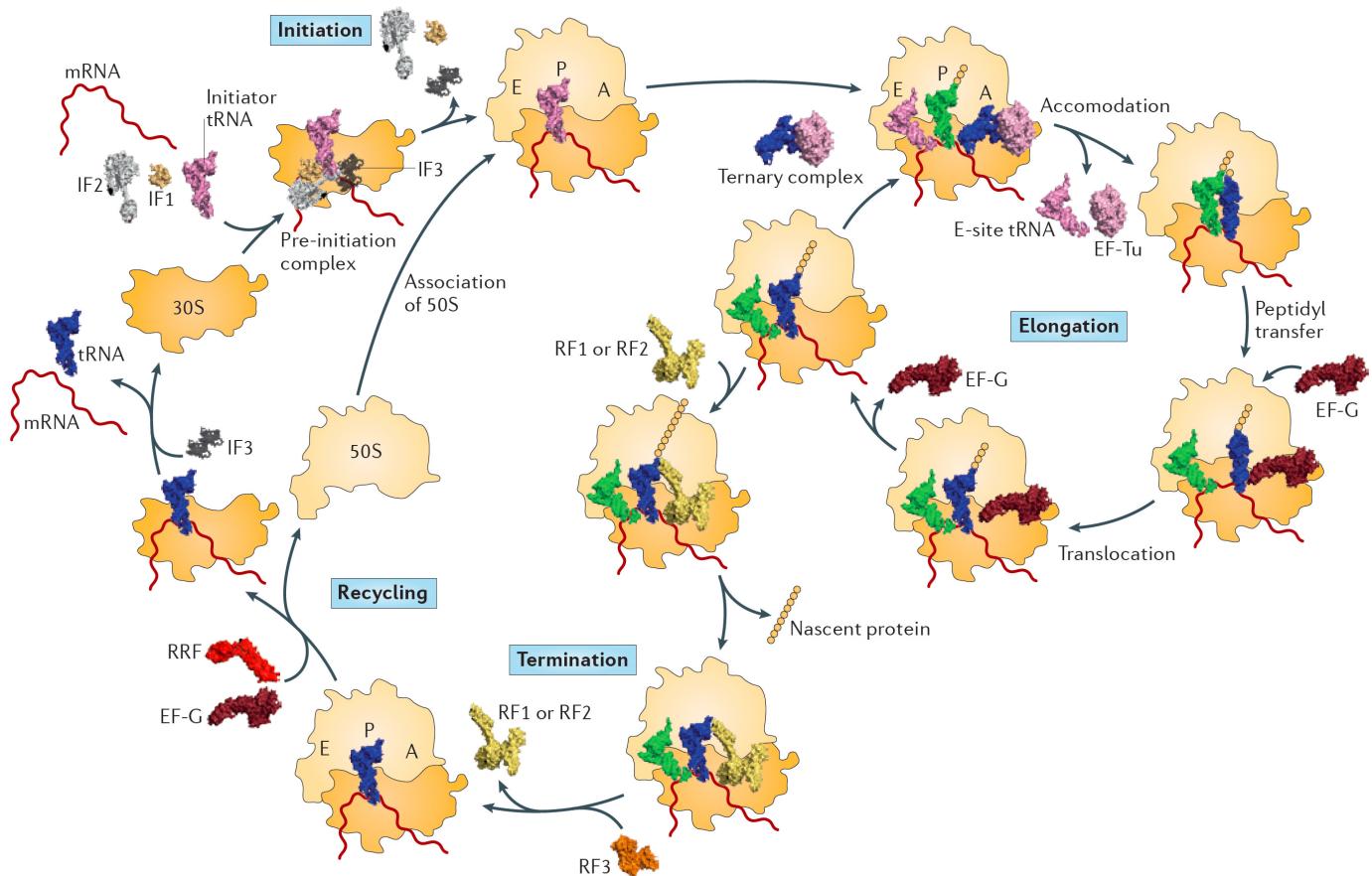


Figure 2: Translation in bacteria [101]

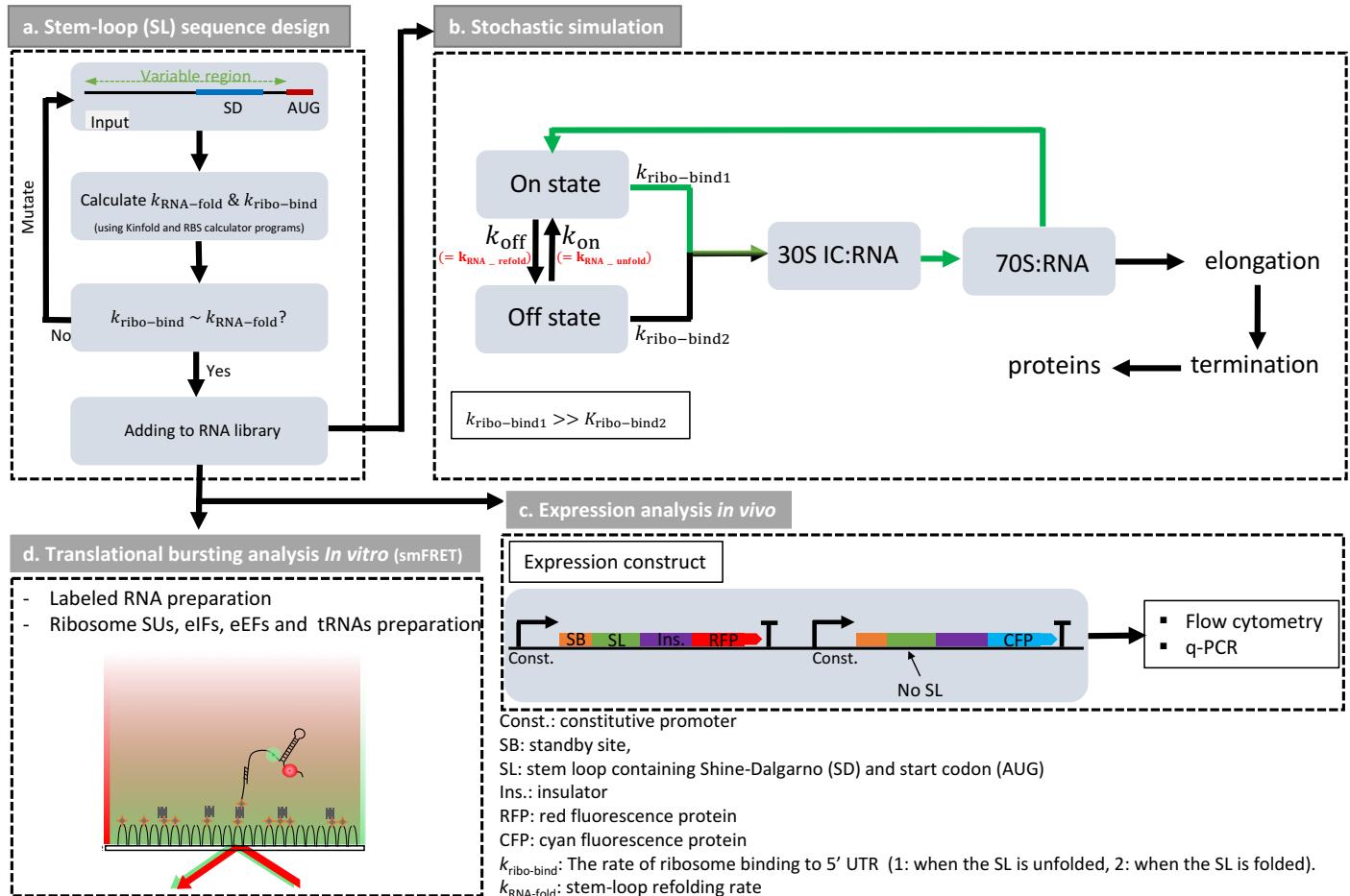


Figure 3: An overview of the approach for analysing the noise generation in translation initiation from a stem-loop structure in the 5' untranslated region.

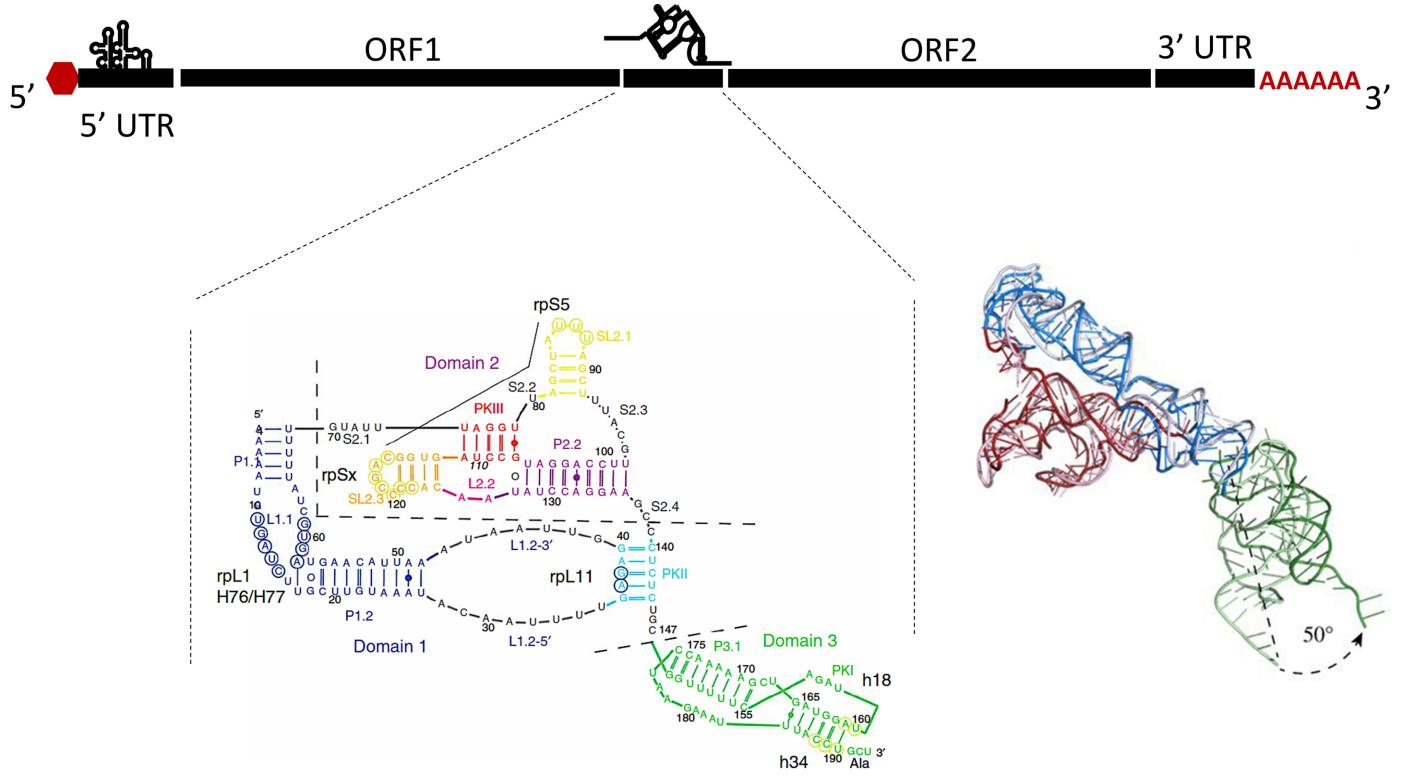


Figure 4: The cricket pralysis virus. Its structure is composed of three domains as shown with dashed lines in figure 4. Domains 1 and 2 are responsible for recruiting 40S and 60S subunits respectively [68, 102]. Domain 3 is folded independently from domain 1 and 2 without any contact with them [103] and is essential during initiation as it places the start codon into the A-site after 80S ribosome formation [104, 105]. It has been determined that in the 80S:IRES complex, the pseudo-knot of the domain 3 (PK1) of the IRES first occupies the A-site of the ribosome, then it is translocated to the p-site by incorporation of an elongation factor (EF2) [106]. Removing the A-site blockage facilitates the recruitment of the first tRNA [106]. According to cryo-EM structural studies [105–107], the translocation of IRES from A- site to P-site is co-occurred by conformational changes on both ribosomal subunits and the IRES structures by the action of EF2. Pre-translocation (faded color) and post-translocation states (dark color) of the IRES is shown in figure 4 right.

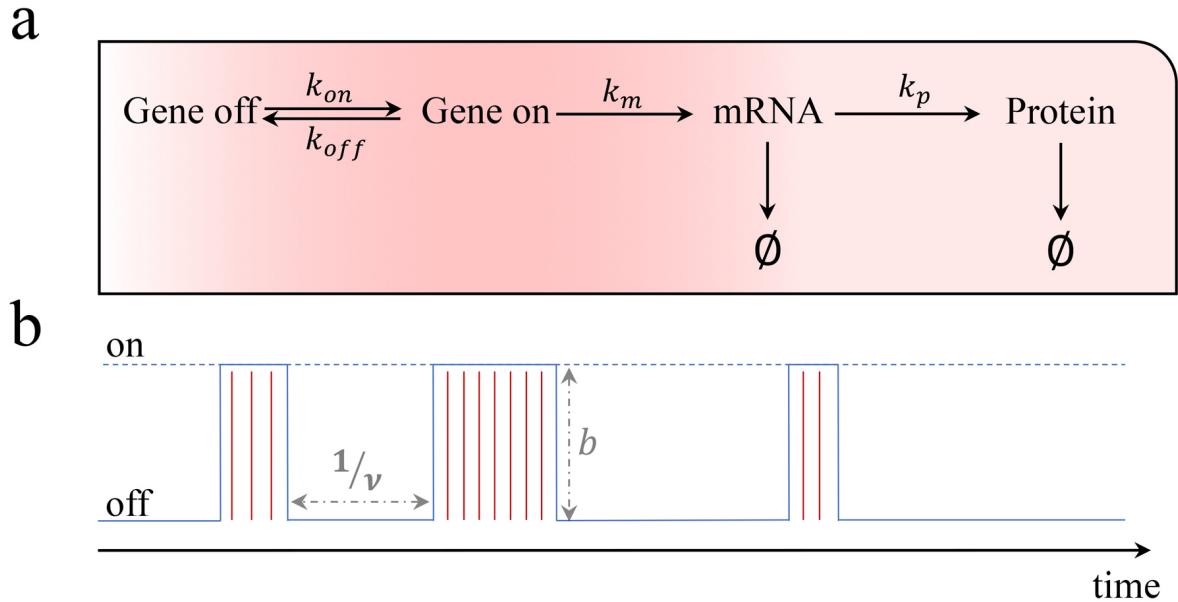


Figure 5: a) A classic coarse-grained off-switch model of bursting in the transcription level. b) The outcome of bursty transcription where mRNAs with different frequencies(ν) and size are produced.

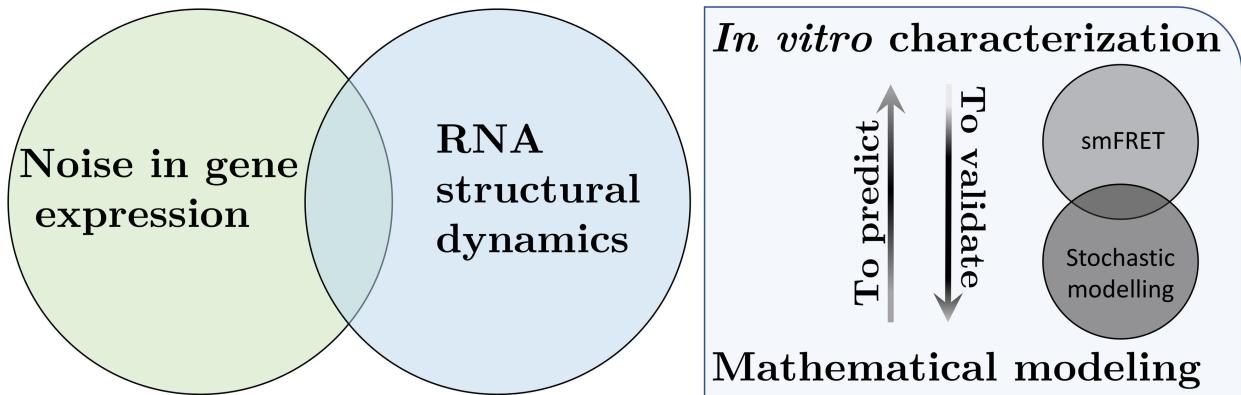


Figure 6: The general objectives of the Ph.D. proposal: the interplay between noise in gene expression and RNA structural dynamics. smFRET and stochastic modeling will be the main approaches to deal with the defined questions in this framework.

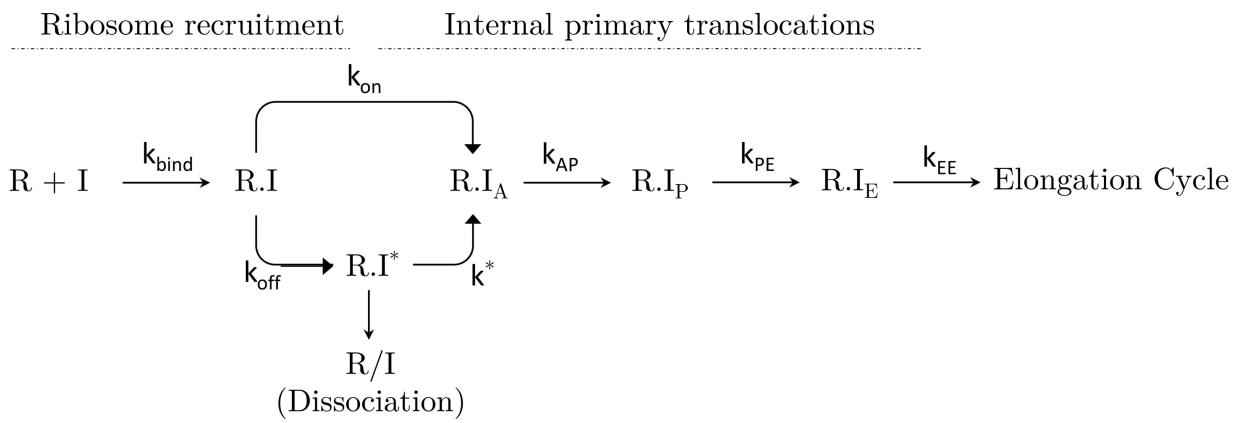


Figure 7: IRES initiation model. The ribosome (R) binds to the IRES structure (I) and forms $R \cdot I$ complex. Then it splits to two pathways: (i) The on state where the IRES is in its compact active structure can initiate translation by subsequent translocation of the domain 3 through the A, P and E sites with corresponding rates. (ii) The off state where the IRES is in a form that is incapable of performing the initiation event. That results in the dissociation of the ribosome or initiation event with the meager rate. The two compartments and also the rates for each transition are further shown in the figure.

Tasks	Year Month	2018		2019			2020			2021		
		Summer	Fall	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer	Fall
1. Molecular mechanisms of translational bursting from a structured stem-loop in <i>E. coli</i>												
a. Determine expression noise generated by 5' UTR stem-loops												
Computational design of 5' UTR stem-loops												
Design and synthesize 5' UTR stemloop DNA												
Assembling stem-loop reporter constructs												
Flow cytometry and qPCR- initial and confirmation analysis												
b. smFRET to determine structural dynamics of free stem-loops												
Preparation of fluorescently labeled stem-loop												
smFRET experiment and analysis												
c. smFRET to determine structural dynamics of stem-loops during translation initiation												
Preparation of fluorescently labeled stem-loop												
Preparation of Ribosome and translation factors												
smFRET experiment and analysis												
d. smFRET to determine the kinetics of sequential initiation events												
Preparation of fluorescently labeled stem-loop RNA and ribosome												
Preparation of stem-loops and translation factors												
smFRET experiment and bursting analysis												
e. developing a mathematical model for translational bursting and its application in circuits												
2. design toggle switches with burst-characterized RNAs												
3. Molecular mechanisms of translational bursting from a structured stem-loop in yeast												
4. Translational bursting in the translation from an IGR IRES												
Apply to scholarships												
1. AIHS												
2. Vanier												
3. AB-Cancer Foundation												
Comprehensive Exam												
Attending conferences												
1. Ribowest												
2. SMB conference in Montreal												
3. EMBL conference												
4. RNA society/ 25th Annual meeting												
Writing thesis/Defence												

1

Figure 8: PhD timeline.