In search for multifunctional lncRNAs

Bharat Ravi Iyengar

Institute for Evolution and Biodiversity, University of Münster,
Hüfferstraße 1, 48149 Münster, Germany

b.ravi@uni-muenster.de

Abstract

Long non-coding RNAs (lncRNAs) were so named because at the time of their discovery, no corresponding protein products were known. Despite the lack of evidence for translation, many lncRNAs perform essential cellular functions such as regulation of gene expression. Recent studies show that many lncRNAs, including those with known regulatory functions, bind to ribosomes and are translated, suggesting that RNAs can perform two different kinds of functions (a phenomenon known in proteins, called moonlighting). Using a formal mathematical model, we show that the two functions are usually mutually exclusive. However, an RNA can transition from one function to the other, simply by a spatiotemporal regulation of the expression of the RNA's binding partners. These findings may encourage further exploration of moonlighting lncRNAs, their regulation, and their role in the evolution of new protein coding genes.

Main

A well known function of ribonucleic acid (RNA) in all living cells, is that it serves as a template for protein synthesis in the form of messenger RNA (mRNA). Ribosomal RNA (rRNA) and transfer RNAs (tRNAs) perform the vital cellular function of facilitating protein synthesis. RNAs, that are not known to give rise to any proteins, are labeled as non-coding RNAs (ncRNAs, Mattick and Makunin, 2006). Long non-coding RNAs (lncRNAs) are a loosely defined class of ncRNAs that are longer than 200nt, that are frequently involved in regulation of gene expression (Statello *et al.*, 2020; Mattick *et al.*, 2023). Recent studies have shown that many lncRNAs bind to ribosomes and are translated to synthesize proteins (Ruiz-Orera *et al.*, 2014; Ingolia *et al.*, 2014; Patraquim *et al.*, 2022). Thus their classification as a non-coding RNA may be inaccurate, and needs reassessment (Jalali *et al.*, 2016). Interestingly, among these ribosome-associated (and translated) lncRNAs are several lncRNAs that perform other distinct functions. For example, a ribosome associated lncRNA, *bxd* is involved in repressing the transcription of the developmental gene Ultrabithorax (*Ubx*) in *Drosophila melanogaster*

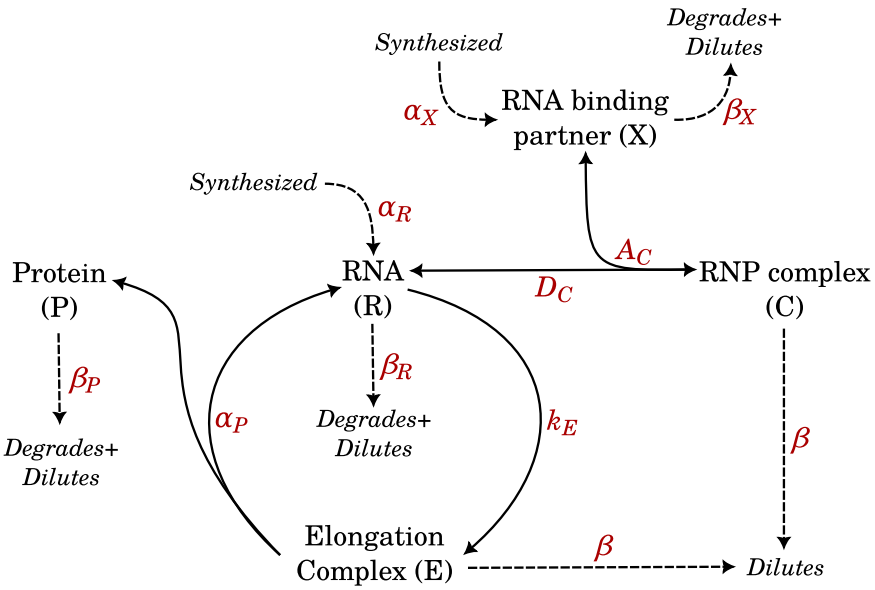

Figure 1: Illustration of the competition model, that is formally described in Equations 1 – 5. The parameters that denote the various biochemical reactions are described in Table 1. Here, dilution refers to dilution of biomolecules due to cell division.

(Patraquim *et al.*, 2022). Similarly, the mammalian lncRNA *Malat1*, that is known to regulate gene expression and cell cycle, can also associate with ribosomes and synthesize a protein (Ingolia *et al.*, 2014). It is yet to be determined if the proteins produced by translation of different lncRNAs are beneficial to the cell. Nonetheless, it is clear that an RNA molecule can perform two distinct functions. Such a non-singular functionality, called "moonlighting", is well known in proteins (Jeffery, 2017). An important question about moonlighting is whether the two functions of a biomolecule can be simultaneously executed by the same molecule. Specifically in case of lncRNAs, it is important to understand if engagement of an lncRNA in protein translation, would compromise its other cellular function, for example, gene regulation. One can indeed expect this because binding to a ribosome may prevent an lncRNA to associate with its other binding partners, and its localization to specific cellular compartments where it performs its functions. Conversely, an lncRNA's cognate function and binding partners, may prevent it from binding to the ribosome and being translated to a protein.

We developed a a simple model of competition to explain the conflict between the two possible functions of an lncRNA. In this model, a hypothetical RNA (R) binds to an RNA binding protein (RBP, denoted by X) to form a ribonucleoprotein complex (C). The lncRNA can also bind to a ribosome to form a productive translation elongation complex (E), that translates an open reading frame (ORF) in the RNA. The ribosome disengages from the RNA when the elongation complex reaches a stop codon, while releasing a molecule of the translated protein. Because the free ribosome is usually in surplus, its concentration would not play a major role in the dynamics of lncRNA binding. This model of lncRNA partitioning illustrated in Figure 1, can be described by the following set of ordinary differential equa-

Parameter (rate constant)	Symbol	Default value	Estimated from
RNA length	-	1000 nt	assumption: variable
ORF length	-	100 codons	assumption: variable
ORF position	-	100 nt	assumption: variable
RNA (R) synthesis	α_R	$\frac{2.5}{\text{RNA length}}$ nM . min^{-1}	Veloso *et al.* (2014)
Synthesis of RBP (X)	α_X	4.78 nM . min^{-1}	assumption: variable
Translation of Protein (P) from RNA	α_P	$\frac{300}{\text{ORF length}}$ min^{-1}	Gerashchenko *et al.* (2020)
Translation initiation	k_E	$\frac{60}{6 \times \text{ORF position}}$ min^{-1}	Vassilenko *et al.* (2011)
Complex formation	A_C	10^{-3} nM^{-1} . min^{-1}	Jain *et al.* (2017)
Complex dissociation	D_C	0.3 min^{-1}	Jain *et al.* (2017)
Degradation/dilution of RNA	β_R	0.0017 min^{-1}	division time of 10h
Degradation/dilution of Protein (P)	β_P	0.0017 min^{-1}	division time of 10h
Degradation/dilution of RBP	β_X	0.0017 min^{-1}	division time of 10h
General dilution	β	0.0017 min^{-1}	division time of 10h

Table 1: Values of different parameters (rate constants) used in the Equations 1 – 5. These values are estimated from published data in the corresponding references, and are biologically reasonable. The standard abbreviations, nM and min, stand for nanomolars and minutes, respectively.

tions:

$$\frac{dR}{dt} = \alpha_R + D_C C + \alpha_P E - R(\beta_R + k_E + A_C X) \tag{1}$$

$$\frac{dX}{dt} = \alpha_X + D_C C - X(A_C R + \beta_X) \tag{2}$$

$$\frac{dC}{dt} = A_C X R - C(D_C + \beta) \tag{3}$$

$$\frac{dE}{dt} = k_E R - E(\alpha_P + \beta) \tag{4}$$

$$\frac{dP}{dt} = \alpha_P E - P\beta_P \tag{5}$$

Solutions of these equations at steady state (equating each equation to zero), describe the long term behavior of the system. The steady state is thus dependent on the value of different parameters (Table 1). The competition between the ribosome and the RBP can be described by four parameters. The first parameter describes the rate at which the RBP (X) is synthesized (α_X) that determines its total amount in the cell. The second parameter (A_C) describes the rate at which RNA binds to the RBP to form the complex (C). The third parameter (D_C) denotes the rate at which the complex dissociates into its two individual constituents. The fourth parameter (k_E) describes the rate at which translation of an ORF in the RNA is initiated.

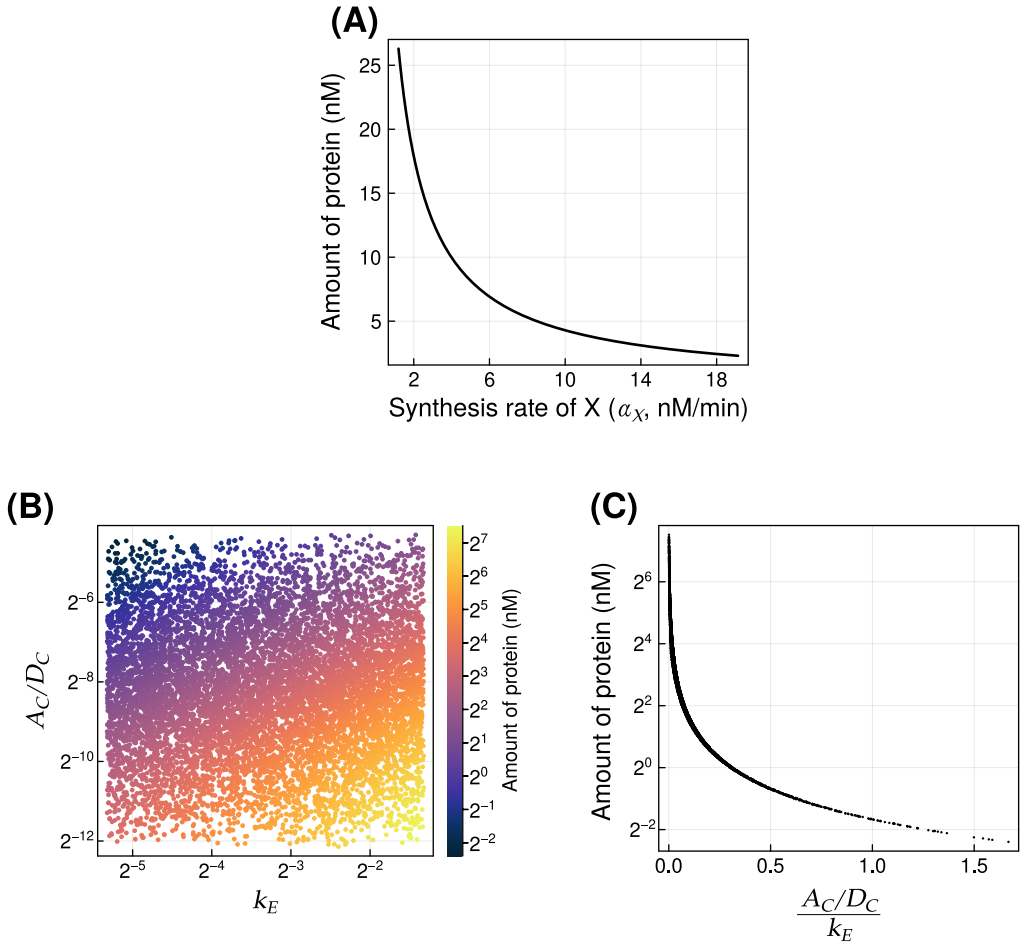

Figure 2: The RNP (X) and the ribosome compete for binding the RNA. **(A)** A higher rate of RNP expression (α_X, horizontal axis) causes lower translation of the protein P (vertical axis). **(B)** Stronger binding to RNP ($A_C/D_C = K_{AC}$, vertical axis, $\log_2$-scaled), and slower initiation of translation (k_E, horizontal axis, $\log_2$-scaled), leads to lower translation of protein P (color gradient, $\log_2$-scaled), and *vice versa*. The scatter plot displays 10000 data points that describe the steady state solutions with 10000 parameter sets, where we randomly sampled the parameters A_C, D_C and k_E from a multivariate log-uniform distribution ranging from $1/4$ to 4 times their default values. We kept the other parameters at their default values. **(C)** is a summary of panel **(B)**. Here, the horizontal axis ($\log_2$-scaled) shows the amount of protein and the vertical axis shows the K_{AC}/k_E ratio.

A high expression of the RBP (X) leads to a smaller amount of translated protein (P), because more RNP is available to bind to the RNA, and prevent it from being translated (Figure 2A). A higher expression of the RBP will cause lower translation of the RNA as long as all the RBP molecules are not already bound by the RNA, or if there are no regulatory mechanisms such as self-inhibition. More specifically, the rate of RNA-RBP binding depends on the amount of free RBP. A stable ribonucleoprotein complex can form due to a fast association of the constituents and a slow dissociation of the complex. Specifically, the ratio of the association and dissociation rate constants (A_C/D_C), that is the equilibrium association constant (K_{AC}), is a good indicator of the complex's stability. A more stable complex would sequester the RNA better, which in turn would prevent the translation of the RNA. The rate of translation initiation k_E determines how quickly the RNA commits to translation, from which it would be released only when the entire ORF is translated. The competition

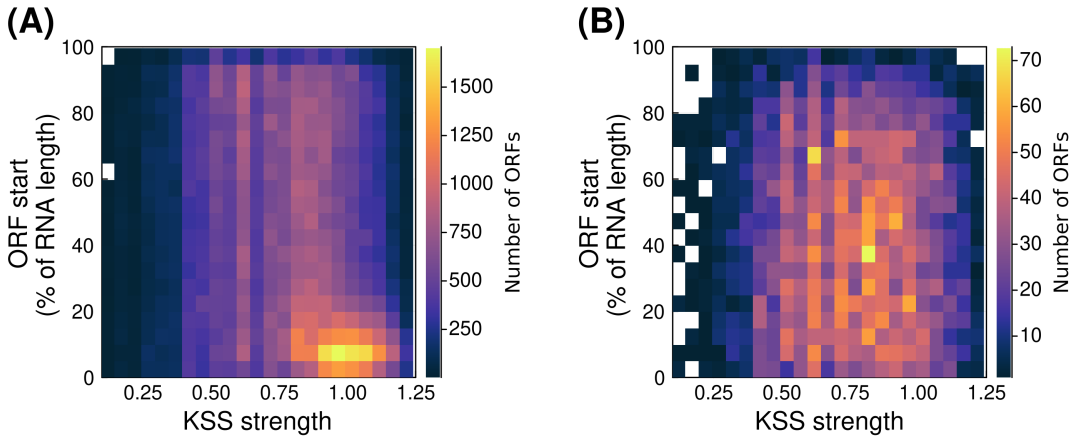

Figure 3: Kozak-optimal ORFs are frequently located in the beginning of *Drosophila melanogaster* **(A)** mRNAs, but not **(B)** ncRNAs. In both panels, the horizontal axes denote the strength of Kozak consensus sequence (KSS, Acevedo *et al.*, 2018), the vertical axes denote the start of the ORF relative to the length of the parent RNA, and the color scale denotes the number of ORFs.

between RBP and ribosome can thus be described by the value of K_{AC} relative to k_E. Specifically, the amount of translated protein is minimum when the K_{AC} is high and k_E is low (Figure 2B). Conversely, the protein's expression is highest when K_{AC} is low and k_E is high. More specifically, the amount of translated protein decreases super-exponentially with increasing values of K_{AC}/k_E ratio (Figure 2C). While the association and dissociation rate constants completely depend on the biochemical properties of the RBP and the RNA (including their 3D structures), the rate of translation initiation can depend on the location of the ORF in the RNA. This assumption follows the widely accepted model of eukaryotic translation initiation, where the ribosome binds to a capped RNA and slides along its length (at an approximate rate of 6-8nt per second) until it encounters a start codon (Vassilenko *et al.*, 2011). Thus mRNAs, that are optimized for being translated would contain the ORF close to their 5′ end. Conversely, an lncRNA would lack such an optimal placement of the ORF. Analyses of translated lncRNAs in *Drosophila* show that highly efficiently translated ORFs are indeed located in the beginning (5′ end) of the RNA (Patraquim *et al.*, 2022). Nucleotide composition around the start codon also has an influence on translation initiation efficiency. Specifically, a sequence of seven nucleotides including the start codon itself, known as Kozak consensus sequence (KSS), determine translation initiation efficiency (Kozak, 1986; Acevedo *et al.*, 2018). In an RNA with many ORFs, the KSS determines which one of them is most efficiently translated (Patraquim *et al.*, 2022). To understand if mRNAs are more optimized for translation than lncRNAs, we analysed the position of ORFs in *Drosophila melanogaster* mRNAs and lncRNAs, and the KSS strength around their start codons. Our analysis revealed that *Drosophila* ORFs with most optimal KSS are frequently present in the beginning of the mRNAs (KSS strength versus relative ORF position, Pearson $\rho = -0.16$, $P < 10^{-15}$; Figure 3A). This is not the case with Drosophila lncRNAs (Pearson $\rho = -0.012$, $P = 0.2$; Figure 3B). Thus most lncRNAs are indeed not optimized for translation, probably because their translation can interfere with their functions as a regulatory RNA. This leads back to the original ques-

tion of whether an RNA can perform both the translational and non-translational functions efficiently. The model suggests that this is unlikely. However, it also shows that by tuning the expression of the RBP (or any other binding partner), a cell can switch a translating RNA to a non-translating RNA. Such a switching behavior indeed exists in Drosophila where the some lncRNAs are translated only in specific developmental stages (Patraquim *et al.*, 2022). LncRNAs can also "moonlight" in other ways. For example, they can associate with different binding partners under different conditions, and can thus perform different functions (Cheng and Leung, 2018; Balcerak *et al.*, 2019). In these cases too, the competition between two binding partners will impose a trade-off between the two functions. A spatio-temporal segregation of the two functions, by regulated expression of the binding partners can resolve this trade-off. Moonlighting lncRNAs would be particularly interesting on the topic of *de novo* gene emergence (Ruiz-Orera *et al.*, 2014). In this phenomena, new protein coding genes (proto-genes) emerge from non-genic DNA sequences, by evolving features essential for transcription and translation (Van Oss and Carvunis, 2019; Schmitz and Bornberg-Bauer, 2017). Transcriptional features may also erode in these emerging proto-genes before translational features emerge. In functional lncRNAs, translational features would be preserved via selection. Therefore emergence of proto-genes in lncRNAs may be more likely than in intergenic regions.

To summarize, bifunctional (moonlighting) lncRNAs indeed exist; identifying them and understanding their functional regulation is an important avenue for future research.

Methods

Mathematical model

We calculated the steady state solutions to the mathematical model by analytically solving the ordinary differential equations (Equations 1 – 5).

Analysis of *D. melanogaster* RNAs

We downloaded *D. melanogaster* mRNAs and ncRNAs from FlyBase (release dmel-6.46, Gramates *et al.*, 2022). We obtained the ORFs from both classes of RNAs using the `getorf` program from the EMBOSS suite (Rice *et al.*, 2000). We calculated the strength of Kozak consensus sequences for the different ORFs using the data from Acevedo *et al.* (2018).

References

Acevedo, J. M., Hoermann, B., Schlimbach, T., and Teleman, A. A. 2018. Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Scientific Reports*, 8(1): 4018.

Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., and Grzybowska, E. A. 2019.

RNA–protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biology*, 9(6).

Cheng, L. and Leung, K.-S. 2018. Identification and characterization of moonlighting long non-coding RNAs based on RNA and protein interactome. *Bioinformatics*, 34(20): 3519–3528.

Gerashchenko, M. V., Peterfi, Z., Yim, S. H., and Gladyshev, V. N. 2020. Translation elongation rate varies among organs and decreases with age. *Nucleic Acids Research*, 49(2): e9–e9.

Gramates, L. S., Agapite, J., Attrill, H., and others 2022. FlyBase: a guided tour of highlighted features. *Genetics*, 220(4).

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., and others 2014. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Reports*, 8(5): 1365–1379.

Jain, N., Lin, H.-C., Morgan, C. E., Harris, M. E., and Tolbert, B. S. 2017. Rules of RNA specificity of hnRNP a1 revealed by global and quantitative analysis of its affinity distribution. *Proceedings of the National Academy of Sciences*, 114(9): 2206–2211.

Jalali, S., Gandhi, S., and Scaria, V. 2016. Navigating the dynamic landscape of long noncoding RNA and protein-coding gene annotations in GENCODE. *Human Genomics*, 10(1).

Jeffery, C. J. 2017. Protein moonlighting: what is it, and why is it important? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1738): 20160523.

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44(2): 283–292.

Mattick, J. S. and Makunin, I. V. 2006. Non-coding RNA. *Human Molecular Genetics*, 15(suppl_1): R17–R29.

Mattick, J. S., Amaral, P. P., Carninci, P., and others 2023. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology*.

Patraquim, P., Magny, E. G., Pueyo, J. I., Platero, A. I., and Couso, J. P. 2022. Translation and natural selection of micropeptides from long non-canonical RNAs. *Nature Communications*, 13(1).

Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The european molecular biology open software suite. *Trends in Genetics*, 16(6): 276–277.

Ruiz-Orera, J., Messeguer, X., Subirana, J. A., and Alba, M. M. 2014. Long non-coding RNAs as a source of new peptides. *eLife*, 3.

Schmitz, J. and Bornberg-Bauer, E. 2017. Fact or fiction: updates on how protein-coding genes might emerge de novo from previously non-coding DNA . *F1000Research*, 6(57).

Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. 2020. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, 22(2): 96–118.

Van Oss, S. B. and Carvunis, A.-R. 2019. De novo gene birth. *PLOS Genetics*, 15(5): 1–23.

Vassilenko, K. S., Alekhina, O. M., Dmitriev, S. E., Shatsky, I. N., and Spirin, A. S. 2011. Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. *Nucleic Acids Research*, 39(13): 5555–5567.

Veloso, A., Kirkconnell, K. S., Magnuson, B., and others 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Research*, 24(6): 896–905.