Computer simulation of a translational roadblock model

Gabin ROUSSEAU

2024-05-21

Table of contents

1 Abstract

Eukaryotic translation begins by initiation: the ribosome loads onto the 5' end of the mRNA and scans its 5'UTR to find a start codon. While the scanning ribosome is sensitive to contextual cues as well as secondary structures in the mRNA sequence, RNA-binding proteins are also able to control translation. Ssd1 is a fungal protein that binds a specific motif in 5'UTRs. It is understood to inhibit protein production in cell wall biogenesis and was proposed to do so by directly blocking scanning ribosomes. We wrote a stochastic model on Python, rini, that belongs to the model family of TASEPs. After adjusting the basic framework of the TASEP with targeted, dynamic defects or roadblocks to simulate the proposed role of Ssd1, rini proved to function as expected of a TASEP. Between roadblock independence and positive cooperativity, only cooperativity could explain the experimental data available. Projecting the explored parameter space with randomised block binding probabilities further illustrated the difference between the hypotheses. Results remain preliminary despite highlighting Ssd1 cooperativity: parameters lack significance with real equivalents, and more should be tested. We recommend repeating the flow cytometry experiment with a varied Ssd1 motif landscape to improve the model and predictions.

2 Abbreviations

- RNA: ribonucleic acid
- mRNA: messenger RNA
- tRNA: transfer RNA
- UTR: untranslated region
- TASEP: totally asymmetric simple exclusion process
- eIF: eukaryotic initiation factor
- PIC: pre-initiation complex
- SSU: small sub-unit
- LSU: large sub-unit
- ORF: open reading frame
- RBP: RNA-binding protein
- IBP: iron-binding protein

- PABP: poly(A)-binding protein
- CSD: cold shock domain
- SBM: Ssd1 binding motif
- ATP: adenosine triphosphate
- rini: roadblock in initiation
- corini: cooperative roadblock in initiation
- LD: low density
- MC: maximum current
- HD: high density

3 Introduction

3.1 Translation initiation and mRNA scanning

Protein synthesis results from translation, where ribosomes read mRNAs to assemble amino-acids into specific peptide chains. In eukaryotes, a ribosomal small subunit (SSU) first interacts with 11 eukaryotic initiation factors (eIFs) to form a pre-initiation complex (PIC), and load onto the mRNA from the 5' m7G cap (Dever and Pavitt, 2016). The PIC then scans the 5' UTR of the mRNA towards its 3' end until a start codon (AUG) is recognised, which triggers a series of reactions to release the eIFs and enable a large subunit (LSU) to form the 80S ribosome complex with the SSU, now primed to begin elongation by reading the translated region that follows (Kozak, 1986).

The scanning ribosome moves in the 5'-3' direction with a footprint of 20-30nt, passing the single mRNA strand through the SSU's mRNA channel until a start codon is detected at the recognition site where the methionyl initiator tRNA (Met-tRNAiMet) is held in place within the PIC by eIF1 (Hinnebusch, 2011; Figure 1 A). Secondary structures in the mRNA such as double-stranded hairpins would prevent the ribosome from advancing and can be melted by a set of helicases, mainly eIF4A, that benefit from the overall structure of the PIC and hydrolyse ATP to function (García-García *et al.*, 2015; Pisareva and Pisarev, 2016; Gupta *et al.*, 2018). The exact nature of the scanning motion appears to be the object of various models, but the consensus seems to forward Brownian motion as the key thermodynamic principle behind scanning. ATP hydrolysis energy expended by eIF4A likely supports translocation further, which remains directed towards the 3' end thanks to the PIC's structural

characteristics and possibly dynamic interactions between eIF4 subunits downstream of the PIC (Spirin, 2009; Marintchev *et al.*, 2009).

In addition, the scanning complex is sensitive to context cues in the 5' UTR that will affect recognition of AUG codons, and therefore which open reading frame (ORF) will be read. For example, preproinsulin mRNAs in vertebrates had significantly impaired initiation with mutations in the consensus optimal sequence 5'-GCCGCC(A/G)CCAUGG-3' about the cognate start codon (Kozak, 1987). Therefore, how long scanning lasts, whether it is carried out successfully and where it ends inherently rely on the 5'UTR sequence itself as well as its secondary structures. However, there exists a variety of other mechanisms which control translation and initiation.

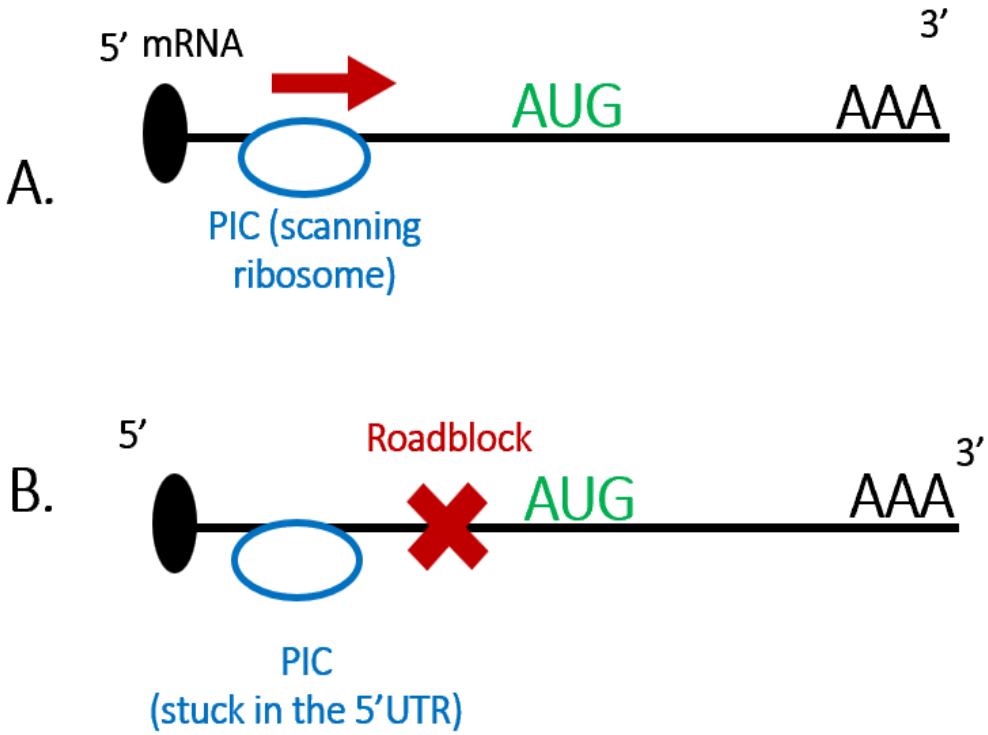


Figure 1: ***Ribosome scanning and roadblocking*** *- A: The PIC scans along the 5' UTR until a start codon AUG is recognised to trigger initiation; B: RNA-binding factors may affect translation dynamics, such as roadblocks binding the 5'UTR, sterically blocking the scanning PIC on its path*

3.2 RNA-binding proteins and translational control

RNA-binding proteins (RBPs) may control translation initiation in specific pathways. Ferritin is a ubiquitous protein able to carry iron Fe3+ ions within ferritin complexes. Its subunits' translation can

be suppressed by iron regulatory protein (IRPs) to control iron homeostasis. Ferritin mRNAs indeed contain an iron responsive element (IRE) which forms a stem-loop secondary structure on which IRPs can individually bind, sterically blocking loading of the PIC on the mRNA by interacting with the eIF4F complex located on the 5' cap (Muckenthaler *et al.,* 1998).

RBPs also control translation from the 3'UTR. Poly(A)-binding proteins (PABPs) ubiquitously bind 3' poly(A) tails in eukaryotes and are notably involved the 3'-5' recruitment of initiation factors to enhance translation (Mangus *et al.,* 2003). However, PABP is also understood to create a negative-feedback loop with itself by inhibiting translation of its own mRNA (Bag and Wu, 1995). This negative feedback loop is possible due to poly-A motifs in the PABP mRNA's 5'UTR; this unusual 5'UTR binding leads to steric hindrance preventing ribosomes from scanning through the 5' UTR (de Melo Neto *et al.,* 1995).

Similarly to IRP and PABP, our protein of interest inhibits translation by binding a motif in 5' UTRs. Indeed, Ssd1 targets a bipartite sequence (18-24nt footprint) concentrated in the 5'UTRs of fungal cell wall synthesis mRNAs (Uesono *et al.*, 1997; Hogan *et al.*, 2008; Bayne *et al.*, 2022). Ssd1 is an RNAse II homolog in which the nuclease catalytic site is inactive and blocked by loop structures; instead of relying on its nuclease site, Ssd1 binds its motif with high affinity *via* specific positively charged regions on the outer face of both of its β-barrel cold shock domains (CSDs; Bayne *et al.*, 2022; Figure 2). The binding motif thus contains two different consensus binding sequences separated by a short gap (4-10nt) that are both key to high affinity binding of Ssd1. The Ssd1 binding motif (SBM) often appears in multiple copies. The SUN4 5'UTR contains two SBMs separated by 23 nucleotides, while the 5' SRL1 gene features 5 adjacent SBMs near the start codon, with intervals of 20-30 nucleotides between each motif (Cherry, 2015). Up to 9 SBMs have been observed in a single 5'UTR of SUN4 homologs across ascomycete fungi.

What is Ssd1's mode of action as an RBP? The kinase Cbk1 was identified as an Ssd1 inhibitor in a reporter expression assay. Cbk1 deletion mutants were only viable if they also had Ssd1 deletions or loss of function mutations (Jansen *et al.*, 2009). These results forwarded Ssd1 as a translation inhibitor which is itself regulated by Cbk1. Due to its properties as a highly specific RBP that strongly binds its target motif, Ssd1 has been proposed to function as a roadblock, sterically preventing PICs from progressing through the 5'UTR (Bayne *et al.,* 2022).

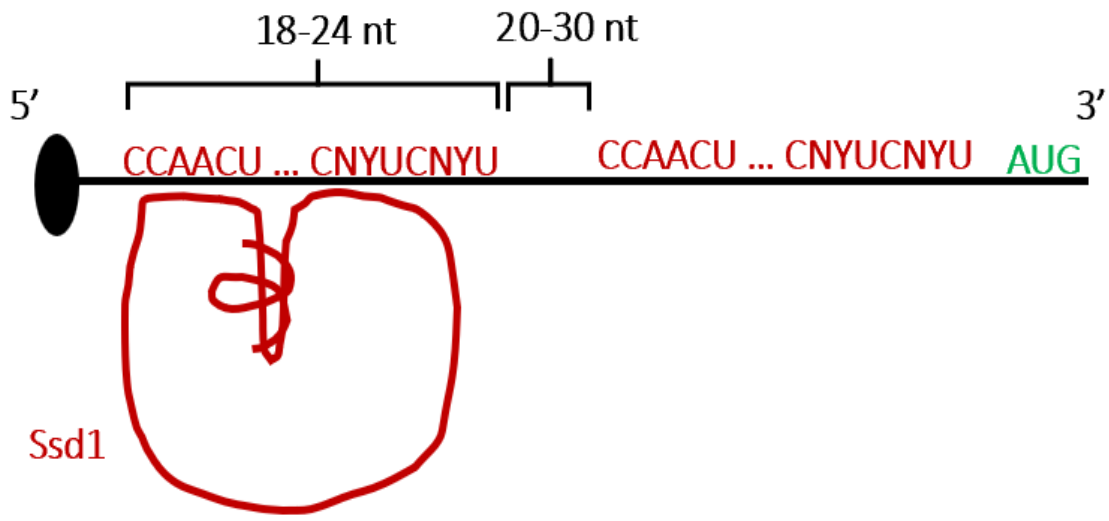
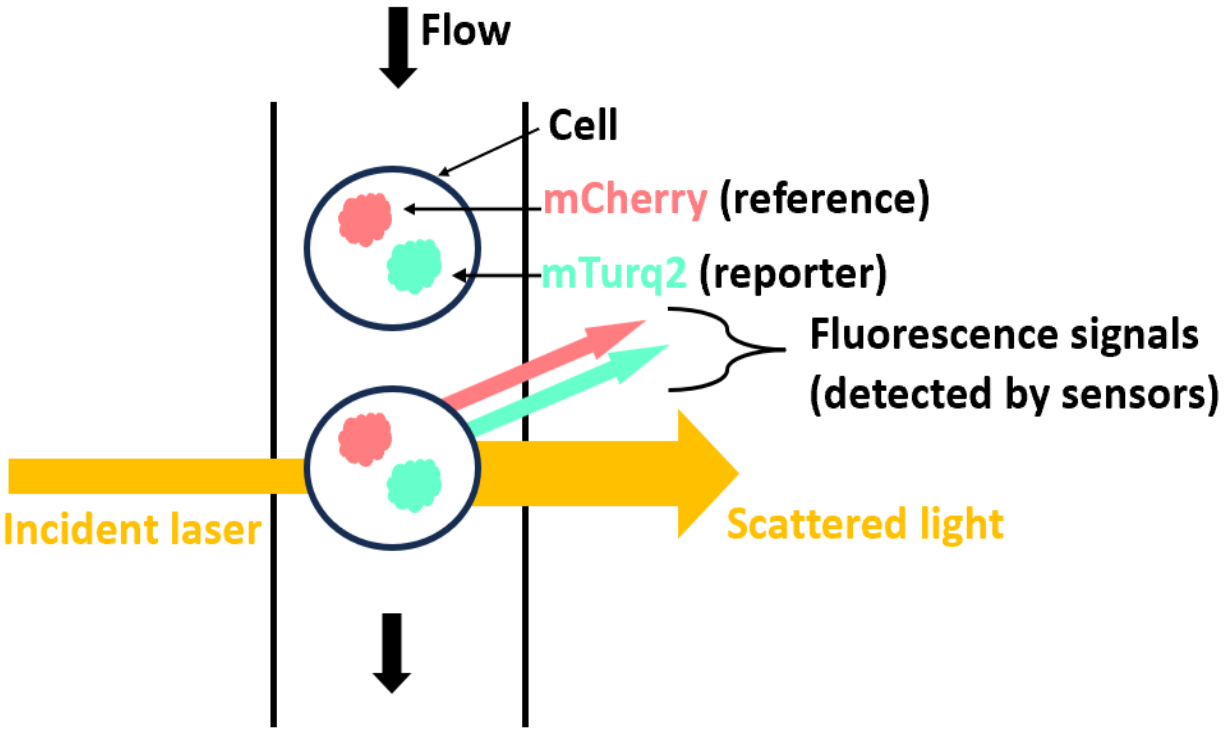


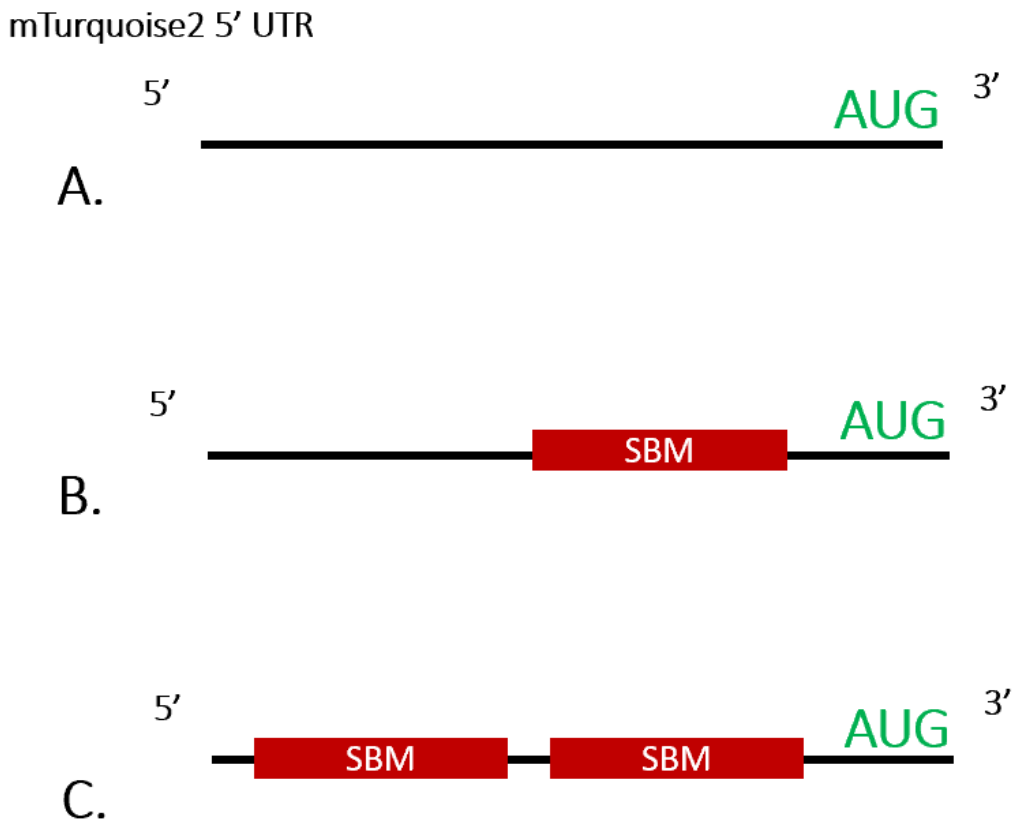
Figure 2: ***Ssd1 binds a conserved bipartite motif in 5' UTRs*** *- Ssd1 binds its motif along the surface of its two CSDs and with high affinity, both nucleotide sequences are needed to achieve this high affinity; the first part isn't necessary to observe a minimum level of Ssd1 binding. As more mutations are introduced in the consensus sequences, Ssd1 affinity for a binding motif can be altered accordingly (Bayne et al., 2022). In the SUN4 5' UTR, two motifs are present, separated by a 23 nt gap. Neighbouring SBMs are commonly observed within 20-30 nt of one another.*

3.3 The importance of multiple Ssd1 binding sites

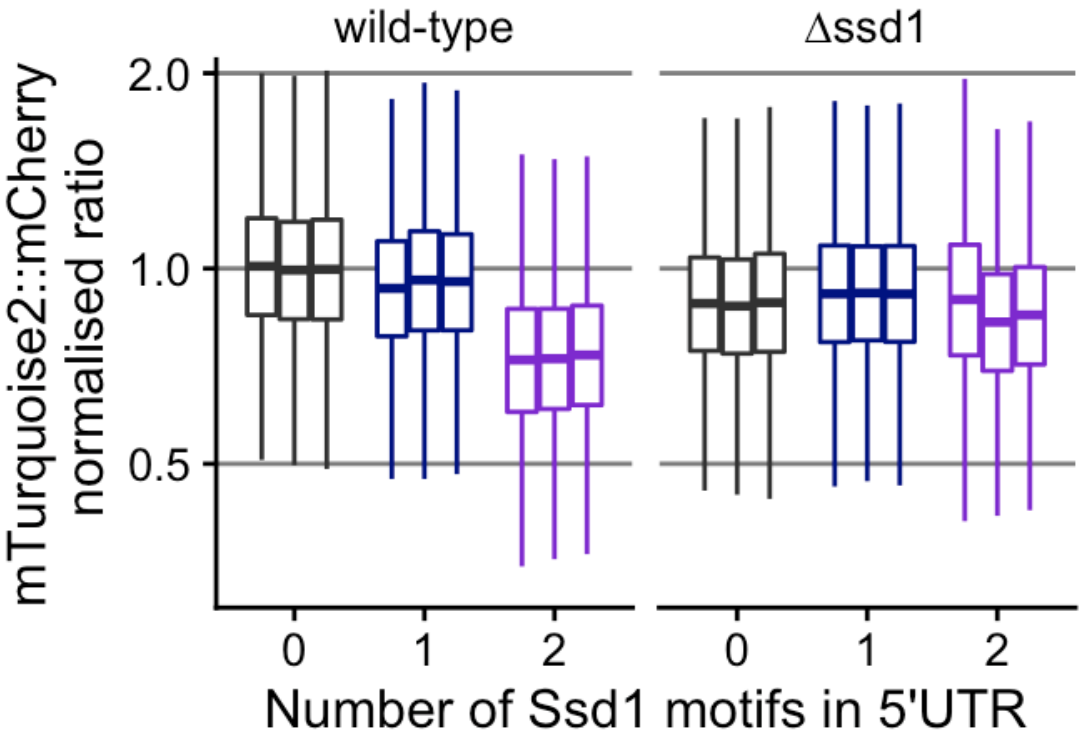
In 2021, Edward Wallace analysed results from a flow cytometry experiment in *S. cerevisiae* that measured the production of a fluorescent reporter, mTurquoise2, which had a variable number of Ssd1 binding motifs in its mRNA's 5'UTR, from 0 to 2 (Figure 3). The reporter's signal was normalised with the signal of a reference protein (mCherry) to use the flow cytometry measurements as proxies for protein production. The data from the experiment showed a small decrease in reporter expression in the presence of 1 binding motifs, whereas two motifs were associated with a much stronger effect: the loss of nearly a third of the standard protein production without Ssd1 motifs. With these results in mind, what experiments could help makes sense of this pattern to unveil Ssd1's mode of action? This raised the idea of designing a computational model to simulate translational roadblocking, with respect to the proposed mode of action, to help understand Ssd1 functionality as well as generate quantitative hypotheses that could serve as the basis for future experiments.



(a) ***Flow cytometry design to standardise a protein reporter with a reference signal*** *- The signal of both proteins was captured in separate channels. Organism: Saccharomyces Cerevisiae*



(b) ***Reporter 5’ UTRs contained 0 to 2 SBMs*** *- The mTurquoise2 reporter was associated with three 5’ UTRs with different numbers of SBMs, from 0 to 2 adjacent motifs, similar to the SUN4 5’ UTR in S. cerevisiae. The aim of the experiment was to approximate the relationship between the presence of SBMs and protein production.*



(c) ***Key flow cytometry results*** *- shows data from 3 replicates; mTurquoise2 is the reporter signal associated with up to 2 Ssd1 binding sites in its mRNA's 5'UTR, its signal serves as the proxy for protein production, normalised by the reference mCherry signal and the WT median mean of the signal ratio at 0 binding sites; if we average replicate data, the WT standardised median ratio (protein production efficiency) with 1 and 2 binding sites are respectively 0.959 and 0.741*

Figure 3: ***Early flow cytometry results link Ssd1 binding site number with protein production*** *- Unpublished results from the Edward Wallace lab, University of Edinburgh 2021.*

3.4 Computational models of translation initiation

Good computational models directly based on experimental data should be able to efficiently generate hypothetical explanations for the data and establish a dialogue between past and future experiments to streamline the research process (Möbius and Laan, 2015). Part of this efficiency stems from isolating the key elements investigated in the system and their expected interactions to balance simplicity with interpretation accuracy.

For this model, we therefore need to take into consideration the essential elements involved: if Ssd1 controls translation initiation by blocking ribosomes while they scan an mRNA's 5'UTR, then our system contains three essential elements. We need an object (ribosome) to follow a one-directional trajectory (mRNA) with a start (5'cap) and an end (AUG codon). Finally, a secondary object (Ssd1) must be able to stand in the way of the first dynamically. If ribosomes reach the end of the 5'UTR, then we could assume successful protein production and track these exits. Fortunately, these conditions are well adapted to a type of stochastic, one-directional and open-ended diffuse transport model referred to as TASEP (Totally Asymmetric Simple Exclusion Process; Popkov and Schütz, 1999; Bonnin *et al.,* 2017).

TASEPs have been frequently used to model translation processes involving ribosome traffic. Such TASEPs find their own individual variations. For example, one TASEP was designed to study the effects of ribosome drop-off mid-way through translation, adding the possibility of having ribosomes (referred to as 'particles' in TASEPs) suddenly leaving the mRNA (the 'lattice') before reaching the end; another interested in initiation simulated the interaction between scanning PICs and elongating ribosomes that formed due to an upstream ORF in the 5'UTR, involving dynamic and random changes in the ribosomes' behaviour (Bonnin *et al.,* 2017; Andreev *et al.,* 2018). However, these examples only model specific ribosome behaviours and dynamics that are not directly featured in the basic TASEP concept, and lack a secondary object that could correspond to another RNA-binding factor. Waclaw *et al.,* suggested an extension to TASEP involving "site-wise dynamic disorder" (2019). Disorder in the lattice-particle system would be created by having obstacles ('defects') dynamically and randomly attach to sites on the lattice, blocking the way until they detach. Different sets of rules were suggested for obstacle behaviour, which included constrained defect dynamics. Constrained defect dynamics not only allow defects to block particles, but also particles to prevent defects from binding the site they stand on, which would better reflect the behaviour of physical objects like ribosomes and Ssd1.

3.5 Aims and objectives

This project aimed to:

- Simulate translational control by roadblock RBPs with the development of an in-house TASEP model with roadblocking, specifically intended to help generate predictions regarding the effect of Ssd1 landscapes on translation initiation.

- Generate simulated output that can be used as proxy to be compared with the flow cytometry experiment.
- Test two hypotheses on Ssd1 activity: whether Ssd1 units bind SBMs independently or cooperatively.
- Use simulation results to offer directions and predictions for future experiments.

4 Methods

4.1 Writing the model: rini

We wrote rini (short for "roadblock in initiation") in Python 3.12 to model the effect of dynamic roadblocks on translation initiation efficiency. The model and all functions that rely on its output were written and run in a Jupyter notebook (Kluyver *et al.*, 2016; see Appendix for the final version used in this report, Section 7). rini dependencies include the NumPy package for array programming (Harris *et al.,* 2020). Figures were generated with the packages Matplotlib and Seaborn (Hunter, 2007; Waskom, 2021). Data was processed in pandas data frames (Pandas development team, 2024).

Rini is a TASEP model in which we consider a one-dimensional lattice of length L, the length corresponding to the total number of sites that can be occupied. Particles may enter the lattice, *i.e.* load onto the first site, with a loading rate α. Any particle in the lattice with an unoccupied site ahead of it may hop forward onto that site with a progression rate p (set to the constant value of 1 in this project). Finally, a particle that reaches the end of the lattice can exit the lattice with an unloading rate β (Figure 4 A). Loading (entrance) and unloading (exit) rates used ranged from 0 to 1. These arbitrary boundaries were chosen to provide an easily interpretable framework to the project. In translation initiation, particles would correspond to scanning ribosomes, and consequently the lattice would be the mRNA 5'UTR. When a particle exits the lattice, we assume the PIC to have detected a cognate start codon, flagging the start of elongation, which is assumed to result in successful protein production.

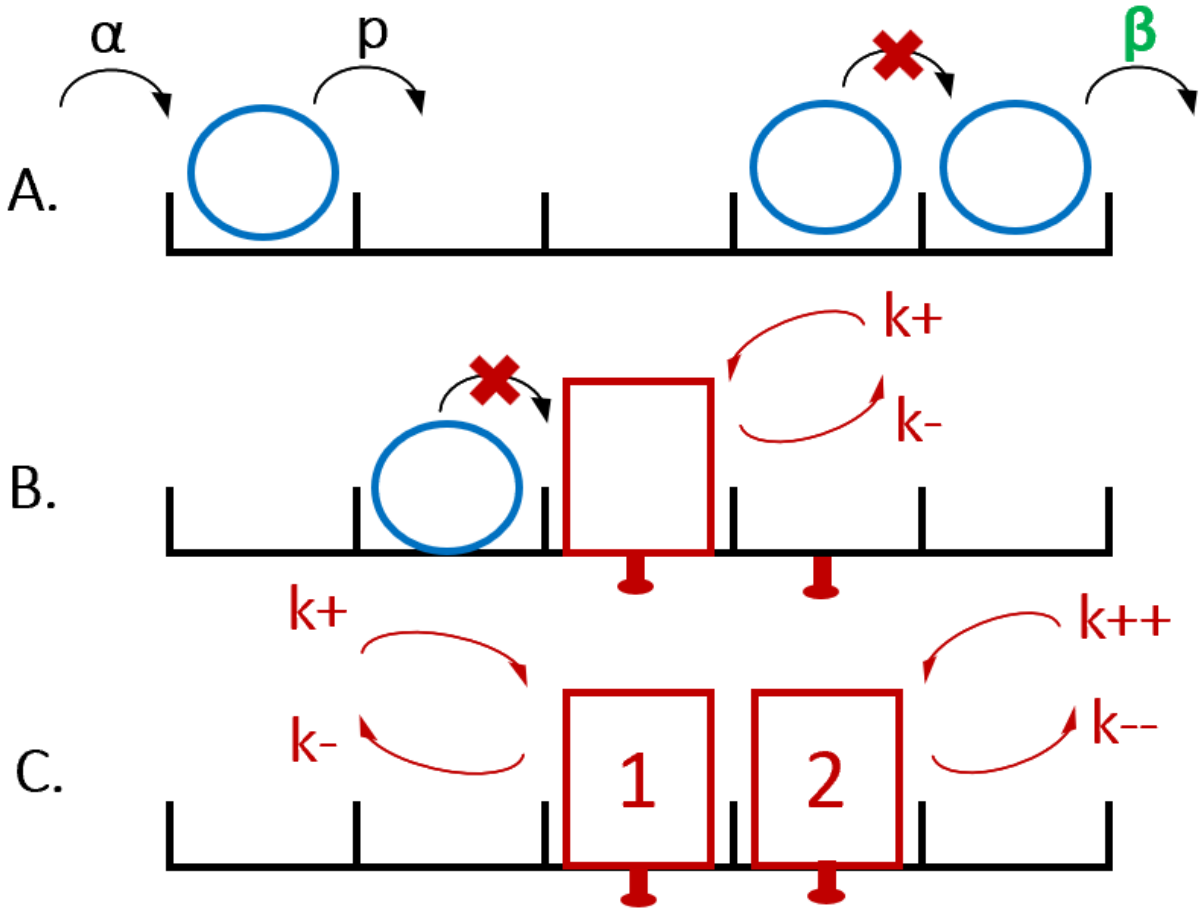


Figure 4: ***Structure of the rini model*** *- A: Key features of a TASEP model, particles (blue) stochastically enter a lattice containing n sites (black), and hop from one site to the next until reaching the exit. Movements are associated with set probabilities. Particles can only move in one direction and can only proceed to the next site if it is unoccupied. B: In rini, sites can be flagged as targets for roadblock binding (also known as "defects", red). Every flagged site has a probability to get bound or unbound by a roadblock. In the first version, these values are fixed and independent from the roadblock landscape. C: the cooperativity version corini features an additional rule that adapts block dynamics according to the roadblock landscape. The probabilities of block binding and unbinding are modified in the presence of a neighbouring block in favour of the blocks staying on the lattice.*

The model simulates this stochastically with respect to the Gillespie algorithm formalism (Gillespie, 1977). For the model to move forward over time, the user needs to set a time limit t. Indeed, rini will create time steps until t is reached, starting from runtime $rt = 0$. Each time step revisits the same series of key events:

- Checking all possible actions on the lattice and recording the corresponding probabilities P as a separate set of propensities A, *e.g.* the probability of an action such as particle loading to occur at the current time step.
- Calculating R, the sum of the propensities.
- Selecting a random number $r1$ from an exponential distribution that scales on R to calculate how long the current time step lasts.
- Selecting a random number $r2$ from a uniform distribution within the bounds (0, 1) that scales on R to determine which of the actions will be executed. This is done by tracing which action appearing in the R distribution corresponds to r2. For example, if in that distribution $r2 = 0.15$ and the first two propensities a_1 and a_2 each weigh 0.1, then $a_1 < r_2 \leq a_1 + a_2$: action 2 will be performed.
- Executing the chosen action and making appropriate changes to the system as a whole (*i.e.* the lattice, propensities).

In addition to particles traveling from start to end, rini features targeted, dynamic roadblocks that represent RNA-binding factors such as Ssd1 here. Specific sites B_i (i: lattice index of the site) can be defined as block binding targets. When such a site is unoccupied, a block may bind with a probability k_{on}. A bound block may unbind at a probability k_{off}. When bound by a block, a site is considered "occupied" the same as when a particle is present, blocking passage (Figure 4 B).

4.2 Corini: a second model for the cooperativity hypothesis

After gaining confidence in the model, we observed in the original cell cytometry data that protein production appeared to be exponentially inhibited by the presence of 2 Ssd1 binding sites in the reporter protein's 5'UTR, compared to the effect linked to the presence of 1 site. In other words, experimental data displays a non-linear change in protein production from 0 to 2 Ssd1 binding sites. This difference approximately goes from a drop in protein production efficiency of 4.1% with 1 site, to 25.9% with 2 sites (Figure 3 (c)). The nearly 5-fold effect increase would suggest that the presence of adjacent binding sites for Ssd1 multiplies the proteins inhibitory effect on translation. If Ssd1 acts as a direct roadblock for the pre-initiation ribosome, then cooperativity between Ssd1 units could explain this pattern.

However, rini alone does not emulate cooperativity between roadblocks, as k_{on} and k_{off} remain strictly constant for any blockable site, independently of the surrounding roadblock landscape. This being likely insufficient to explain the experimental data, rini became the 'independence' model.

We wrote an alternate 'cooperativity' version of the program, corini, which assumes the presence of a block may improve block binding or decrease block unbinding in its direct neighbourhood, creating simple cooperativity between adjacent roadblock units. If cooperativity becomes active at a block site and at any point in time, then the rates for block binding or unbinding, where applicable, are modified by fixed values in favour of the roadblocks: $coop_p$ is added to k_{on}, and $coop_m$ is substracted from k_{off} (read: "coop-plus" and "coop-minus"; Figure 4 C).

These two model versions enabled the testing of two hypotheses, roadblock independence or cooperativity.

5 Results

5.1 Testing the model's implementation with a density phase profile

To assess the implementation of rini as a TASEP model, we investigated its phase density profile with the rini_AvB function (Alpha *vs.* Beta, see Appendix in Section 7). Phase density profiles are a characteristic of TASEP models where particles dynamics are expected to follow specific rules according to the combined values of α, p and β (Table 1; Bonnin *et al.,* 2017). If rini is able to satisfy these rules, then we can be confident that it serves as a correct implementation of TASEP.

The AvB function creates 10 independent rini runs with random combinations of α and β (values vary between 0.1 and 1); it returns the phase diagram, plotting α against β, alongside the average occupation density d against the expected bulk density for the corresponding density phase (Table 1; Bonnin *et al.,* 2017). This randomisation may occasionally lead some points to fall into a less predictable phase: Shock Phase, where $\alpha = \beta$ and both probabilities are below 0.5; in that case, densities are unpredictable at specific points in time, interchangably high or low along sections of the lattice.

Table 1: Main density phases expected in TASEP models where p = 1. Adapted from Bonnin et al, 2017.

Density Phase	Conditions	Expected density
Low Density	$\alpha < 0.5, \alpha < \beta$	α
Maximum Current	$\alpha > 0.5, \beta > 0.5$	0.5
High Density	$\beta < 0.5, \alpha > \beta$	$1 - \beta$

In TASEP models, low density (LD) means the lattice will on average be occupied by few particles as α limits particle entry. High density (HD) represents the opposite scenario, where β significantly limits particle exits, yielding high occupation densities. Otherwise, a TASEP model is said to undergo maximum current (MC) when both α and β are high enough for the particle flow, and thus occupation density, to be capped by the progression probability rather than entrance or exit.

We are considering the occupation density as $d = X/L$, with X as the number of occupied sites.

We ran AvB with $t = 10000$ and $L = 100$ with no block sites to approach the steady state and reduce size effects respectively, and selected an output featuring 10 independent points and at least 3 points per density phase (Figure 5). 3 points were expected to be under low density (LD), another 3 in high density (HD), and finally 4 in maximum current (MC). Their expected densities ranged from less than 0.2, to nearly 0.9. The average occupation density calculated over time for each point consistently remains very close to the steady-state bulk density expected from the associated density phases in a regular TASEP model, and don't deviate more than about 5% away from the expected value (Table 1, Figure 5); the point expected to have the highest density, nearly 0.9, yielded the biggest error of 5%, however bigger deviations are not necessarily tied to the highest densities, as a point in LD and another in MC are also noticeably away from the segmented line that represents result-expectation equivalence.

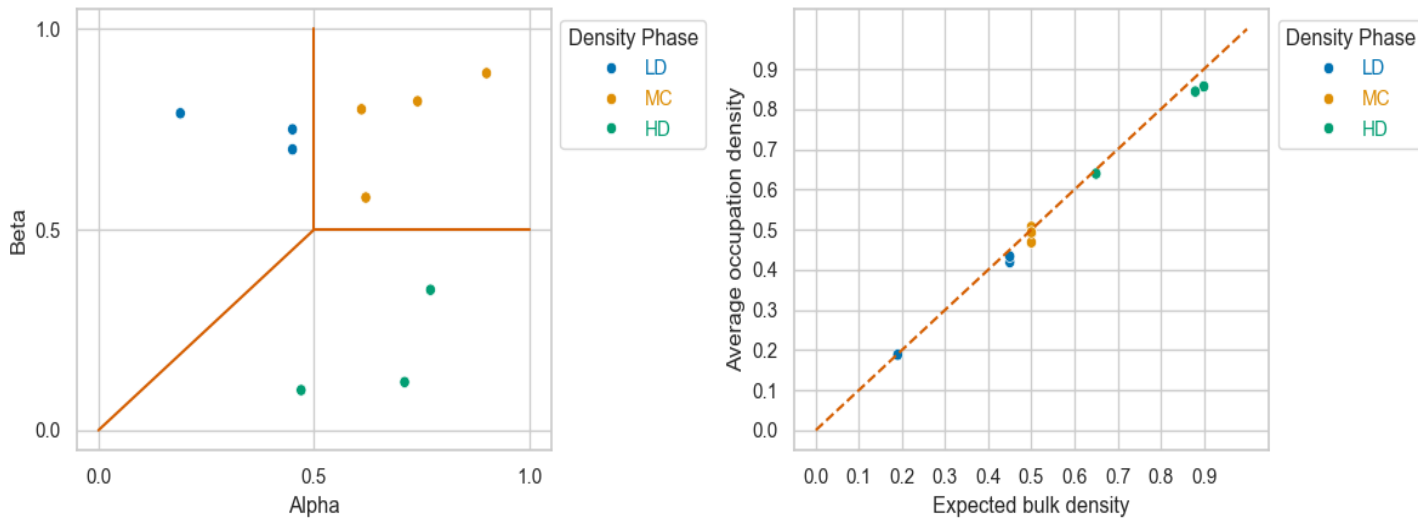


Figure 5: ***Density phase profiling of rini*** *- t=1000, L=100, n=10; Left: density phases expected of each run according to their randomised alpha and beta parameters (LD: Low density, MC: Maximum Current; HD: High Density); Right: average occupation density over each run against the densities expected from density phases (Table 1).*

5.2 Comparing simulated data to flow cytometry data

For the model to contribute to further experiments and generate hypotheses on Ssd1 function, we explored fitting the model to the experimental data available. We have data about the link between block sites and protein production from the flow cytometry experiment, and chose the WT standardised median signal ratios at 1 and 2 block sites as our experimental values of references (0.959 and 0.741 respectively, Figure 3 (c)).

With these values, we can assess how each model version compares to biological data. From a simulation, we can extract the particle exit rate e, which we define as the total number of particles that have exited the lattice over a run, divided by t. Exiting particles represent ribosomes that succesfully complete initiation and proceed to elongation, which we assume results in protein production. So, we choose e to be the initial proxy for the raw protein production. However, we want a comparison with medians of standardised values. For a parameter set ϕ, we calculate median e values over $n = 10$ sampled simulations using this parameter set for an appropriate range of block sites (from 0 to 2). If we use the median at 0 block sites as our model's standard value, we can divide e by this standard to get standardised exit rates ϵ, representation *initiation efficiency*, and plot the data as a boxplot to see the corresponding medians alongside their experimental counterpart.

We wrote the function rini_INIvBLOCK to apply this method on excellent parameter sets we established for each model version (Table 2; Appendix in Section 7). t, L and targetable sites were chosen for simplicity to maximise efficiency when testing multiple parameter sets at once, as lattice size effects appeared to have no notable effect on exit rates once a certain time has elapsed (hence the seemingly high time limit).

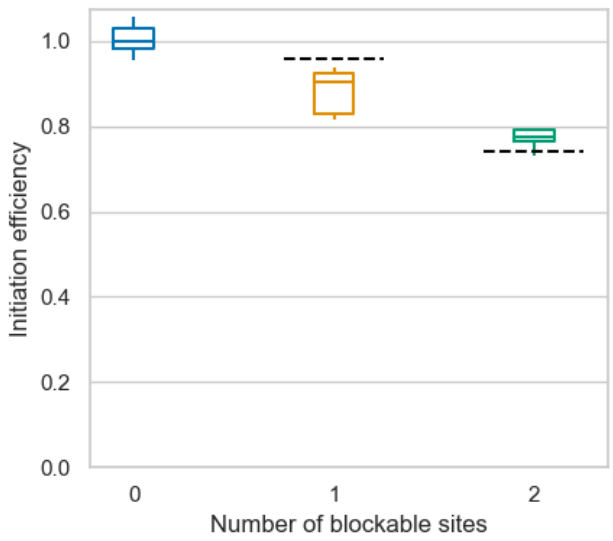
Table 2: Key parameter set details used to generate exemplar fits for each model version

Model	t	n	L	α	β	p	k_on	k_off	coop_p	coop_m	Targetable sites (first and second added)
rini	1000	10	5	0.75	0.75	1	0.227	0.5	0.4	0.4	2 and 3
corini	1000	10	5	0.75	0.75	1	0.043	0.5	0.4	0.4	2 and 3

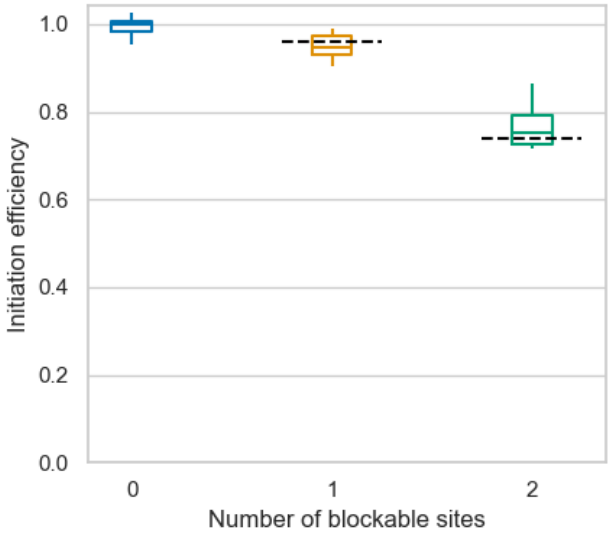
After inspecting whether the model can behave as expected at different density phases, we generated data to be compared with our experiment of reference, and plotted result comparisons for both rini and corini in Figure 6. Both figures display excellent fits from the model versions that resulted from changing the block binding rate k_{on}.

Adding a second blockable site in rini yielded a seemingly linear decrease in initiation efficiency, doubling the effect of a single blockable site. The data found little to no overlap with the experimental medians and with its linearity failed to capture the trend observed in the flow cytometry experiment; with 1 site, all data points stand between the experimental expectation of 0.959 and 0.8, whereas the model mostly stands between 0.8 and the experimental median 0.741 with 2 sites.

The corini version of the model was able to produce output that stays consistently close to biological data (Figure 6 (b)); the experimental medians overlap well with the data and remain close to the model medians. With 1 site, the experimental median indeed lies between the model's first quartile and median (maximum close to 1.0, minimum over 0.90); with 2 sites, it lies between the model's median and third quartile (maximum close to 0.85, minimum over 0.7). Alongside this similarity between the results, we notice that the model follows the same non-linear trend found in the experiment when roadblock cooperativity is taken into consideration.



(a) rini - independent roadblocks



(b) corini - cooperative roadblocks

Figure 6: ***Excellent model fits to experimental data*** *- Black segmented lines: experimental medians; n=10*

5.3 Exploring the parameter space and improving model fits

To understand how well each model is able to explain biological results and what parameters matter the most, we began exploring the model's input parameter space with the function rini_paramscore (Appendix, Section 7).

Exploring a parameter space requires analysing many semi-randomised ϕ possibilities and scoring each parameter set based on how close the corresponding data is to the experiment. We tested $N = 1000$ parameter sets keeping the default values from Table 2 while randomising k_{on} from a uniform distribution ranging from 0 to 1 (sampling for median calculation: $n = 10$).

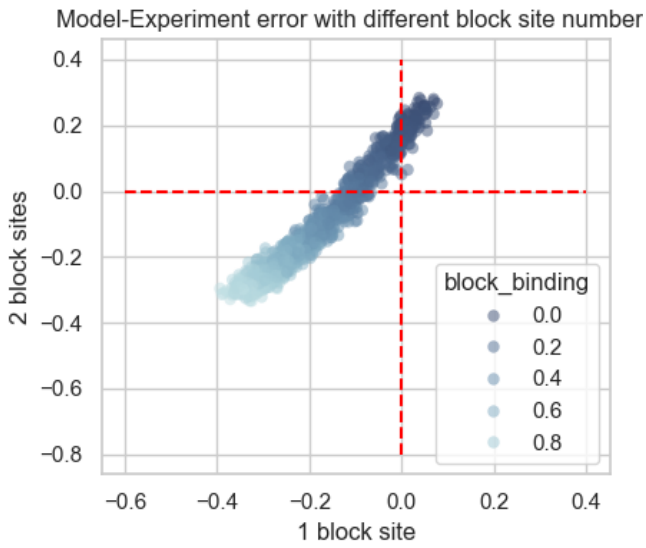
To score each ϕ, we calculate three values:

1. $s1$ (score1) is the difference (or raw error) between the model's median ϵ and the experimental median at 1 site: $\epsilon_{model} - \epsilon_{experiment}$.
2. $s2$ is the equivalent at 2 sites.
3. S is the quadratic mean of s1 and s2: $\sqrt{(s1^2 + s2^2)/2}$.

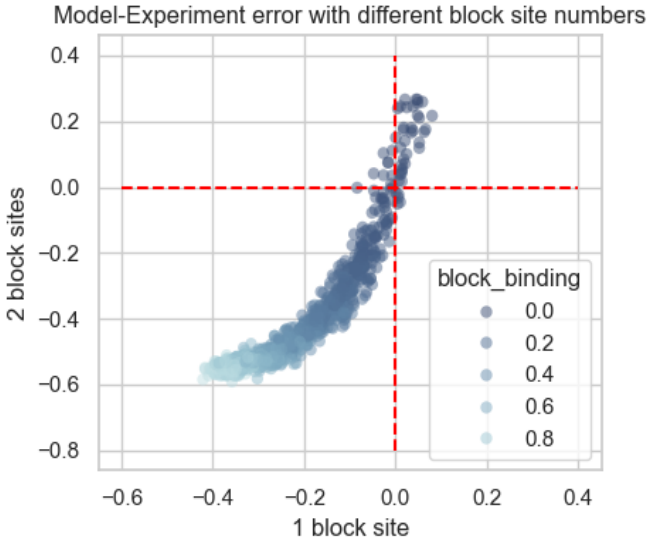
S scales with the difference between s1 and s2 to give an overall score of how close we are to biological data. The closer $s1$ and $s2$ are both to 0, the better the fit. Moreover, the sign of $s1$ and $s2$ gives us an indication of whether the model's median output is above or below the expectation. While S is the main score for a model fit, we also plotted $s1$ and $s2$ against one another to visualise the parameter space.

To test a wide-range of parameter sets and observe how well the results fit biological data, we began preparing a function to explore the parameter space and plotted results for 1000 different data sets where we randomly varied the block binding probability in rini and corini (Figure 7). In both versions, reducing k_{on} increased initiation efficiency, as seen in the trend going towards positive raw errors. However, as fits improve and S scores decrease, we notice a constant bias between s1 and s2 in rini that prevents the data from crossing the (0,0) coordinates that represent a perfect fit. Indeed, the independent model's closest point stands at the coordinates (-0.0394, 0.0177; same k_{on} as in Figure 6 (a)). On the other hand, corini is seemingly free from this bias and is able to reach these coordinates, with its best point having the coordinates (-0.0032, 0.0039) and its S score being 10-fold lower than the best point from rini (Figure 7 (b); same k_{on} as in Figure 6 (b)).

A key difference between both models is therefore the relationship between $s1$ and $s2$, the results at 1 site and those at 2 sites. For both rini and corini, $s1$ ranges between -0.4 and about 0.05. The skew towards negatives associated with better block binding shows that for most block binding probabilities in the default parameter set used, block binding would be too strong to fit the biological data. As block binding decreases in the independence model, which makes the model's median increase closer to the experiment's at 1 site, the raw error also increases linearly at 2 sites (Figure 7 (a)). In the cooperativity model, the 2-blocks data increases following an exponential trend alongside 1-block data, rather than linearly. Indeed, when k_{on} becomes especially low, cooperativity has an increasingly important effect in the model's dynamics as it enables strong binding when multiple blocks are present, even though blocks otherwise rarely bind the lattice alone. This effect enables the cooperative model to generate much stronger fits than the independent version, by at least an order of magnitude.



(a) rini - independent roadblocks



(b) corini - cooperative roadblocks

Figure 7: ***Model-Experiment errors over 1000 randomised parameter sets*** *- Red segmented lines highlight axes with a 0 coordinate; N=1000, n=10*

6 Discussion

Rini is a model that aimed to efficiently emulate the essential dynamics expected of the system mRNA-ribosome-Ssd1, and simulate the effect on protein production seen with multiple Ssd1 binding motifs (Figure 3 (c)). The density phase profiling step helped build confidence that the model fundamentally behaved as expected of a model of its kind (Popkov and Schütz, 1999), and the values for t and L chosen were enough to mitigate size effects and observe consistent behaviour within a 95% confidence interval.

Simulation results show rini can explain experimental data when roadblocks cooperate. Indeed, cooperativity resulted in an exponential decrease in output production (exits) as the number of block binding sites increases, similar to the flow cytometry data, whereas roadblock independence was limited to a linear effect (Figure 6; Figure 7 (b)). Consequently, we expect Ssd1 to act as a roadblock to

pre-initiation scanning, where multiple Ssd1 units bind to adjacent motifs to cooperatively block translation. The results seem furthermore consistent with the frequent recorded presence of multiple SBMs with short gaps in-between that approximately match the footprint of Ssd1 itself (Bayne *et al.*, 2022).

Indeed, we managed to get good simulations with low S scores upon randomising k_{on} among a set of default parameters that we selected to keep the model efficient and minimalist (L), while assuming Maximum Current. In these conditions, simulations were most succesful when the base binding affinity of blocks for their sites was very low (less than 0.05), paired to a relatively high unbinding rate and strong affinity improvements in a cooperativity context.

This only represents one type of parameter set with which the model can accurately simulate biological data, and is already questionable since Ssd1 is supposed to bind its target with high affinity. Initiation is expected to be rate-limiting in translation, and the configuration of the 5' UTR sequence (*e.g.,* likelihood to form secondary structures) seems to play an important part (Cambray *et al.*, 2018). Therefore, maximum current appears to be an interesting assumption, where the key limiting rate would be p. However, the difference in specific parameter values and the range of randomisation chosen to explore the parameter space lack a clear relationship with the reality we aim to simulate; time is unique to the model and specific rates lie on the arbitrary range of (0,1).

In addition to questioning the limits of the parameters we worked with to produce this report, we must also consider whether the system we modelled is overly simplistic. Ssd1 is phosphorylated and suppressed by Cbk1, a regulation mechanism that is essential to yeast cells (Jansen et al., 2009); therefore, Cbk1 may be a key factor that should have been included in rini to interfere with block dynamics. Would implementing Cbk1 solve the problem of low k_{on} values in good simulations?

Nonetheless, rini in its current version benefits from a level of universality as it broadly models roadblock mechanisms in scanning. For example, rini may also be found relevant to the study of PABP, or riboswitches, primary mRNA sequences that are stabilised into a secondary structure when bound to a ligand (Garst *et al.,* 2011); the binding site would be the primary sequence while the roadblock could correspond to the effector in this case. This idea raises a new question for alternate applications: are the relative sizes of the roadblock, the lattice and the particles essential to correct biological interpretations? While Ssd1 and the PIC have comparable footprints between 20 and 30 nucleotides (which is also the usual gap between adjacent SBMs), which favours the coarse-grained approach of

our simulations where particles occupy the same space as blocks on a tight lattice, the same may not be true for all scanning roadblock mechanisms, including riboswitches.

Variable particle length has been integrated in TASEPs modelling translation (Ciandrini *et al.,* 2023), and rini in fact successfully allows for customisable particle and roadblock lengths following similar concerns in the early design of the model. We note that the common gap between Ssd1 sites was not accounted for in this project.

To summarise, the observations made in the project essentially highlight the relevance of cooperativity to the mode of action of Ssd1, but how could we harness rini to make predictions on the relevance of each parameter to obtain optimal simulations, and what would that mean biologically?

We recommend conducting further parameter space explorations that focus on testing multiple parameters at once, and most importantly to isolate parameter value ranges that would be nonsense biologically, such as excessively low k_{on} values. Evaluating the significance of time and probability values based on real ribosome scanning rates and effective particle progression rates would be the starting point of these efforts.

Yet, we're now interested in whether Ssd1 units undergo positive cooperativity, likely through direct protein-protein interaction if SBM proximity is required to obtain the flow cytometry results (Figure 3). Although we focused on modelling roadblock dynamics for up to 2 binding sites, the model would be able to generate data beyond just 2 sites and help predict results for various scenarios. However, to generate data for 3 block sites or more requires a revision in the way cooperativity is programmed. The current rini simply detects cooperativity and modifies block dynamics by set values, $coop_p$ and $coop_m$. If the decrease in protein production exponentially decreases alongside an increase in SBMs beyond two motifs, as in this project, then the current version of rini will be insufficient, which is likely the case considering cooperative RNA-binding factors that directly interact with one another see their overall affinity for RNA increase as the number of elements bound increases (Stitzinger *et al.*, 2023). Expanding rini simulations will consequently require an update of the cooperativity logic to make the cooperativity effects functions of the number of participating blocks.

Knowing this, what would be the next step to elucidate the role of Ssd1 in translational control? While we think that Ssd1 may indeed undergo positive cooperativity when two motifs are juxtaposed in a 5'UTR, the model should easily generate predictions for a wide range of SBM landscapes and their role

in scanning, with the help of updated code for cooperativity. New experimental data to accompany further model simulations should help shed light on the action of Ssd1 on scanning and translation overall. We encourage further projects to follow the flow cytometry experiment design to explore 3 axes that define the SBM landscape: the number of SBMs in the reporter's 5'UTR, the gaps that separate them, and finally the affinity of SBMs (which can be mutated for a range of reduced affinities, see Figure 2 from Bayne *et al.*, 2022; Figure 8 A).

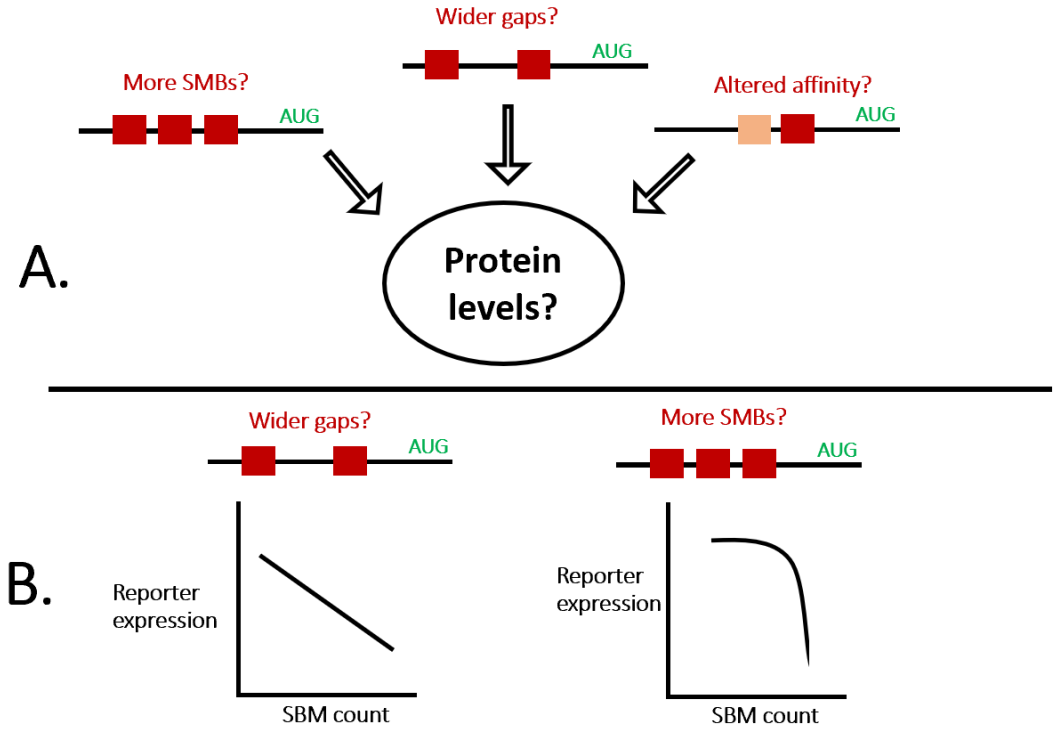


Figure 8: ***Future points of interest for Ssd1 function that can be modelled and predicted by rini*** *- A: Axes to be tested experimentally alongside rini simulations, red: SMBs, orange: SMB with reduced Ssd1 affinity; B: possible predictions based on our results, large gaps between two SMBs in the 5' UTR may yield results consistent with the independence model, whereas more adjacent sites should be the continuation of the past flow cytometry results with an exponentially stronger effect on protein production.*

We can already formulate predictions on what the outcome of these follow-up experiments might be: from the loss of the cooperativity effect with a sufficiently large gap between 5'UTR SBMs, to an even stronger exponential effect as SBMs are added (Figure 8 B). If problems are encountered to explain data with later versions of rini, we suggest considering the implementation of changes with respect to

some of the variables mentioned here, such as Cbk1, which remains an indirect scanning regulator if Ssd1 is confirmed as a roadblock.

7 Appendix

Code for the model and data analysis can be accessed in the rini v0.81 GitHub repository.

8 Acknowledgements

I would like to thank my supervisors Edward Wallace, and Ramon Grima, for their kind guidance and extremely constructive advice over the course of this project. Writing and experimenting with rini was an incredible opportunity to learn and practice modelling, and I sincerely hope the work and tools discussed in this report will fuel a future project!

I am thankful for being able to work alongside Edward and Ramon's lab groups for the past 4 months; lab meetings were an interesting environment to have interdisciplinary discussions, since this project meshes together biology concepts with a lot of mathematics and algorithm writing.

I would also like to thank Juraj Szavits-Nossan who patiently explained Gillespie algorithms, sending me in many good directions to explore the project!

Finally, many thanks to the Biotechnology teaching team here at the University of Edinburgh, and to my family, for supporting my studies and offering counsel as well as material along the way.

WORD COUNT: 5527 words

9 References

Andreev, D.E. et al. (2018) 'TASEP modelling provides a parsimonious explanation for the ability of a single uORF to derepress translation during the integrated stress response', eLife. Edited by N. Sonenberg, 7, p. e32563. Available at: https://doi.org/10.7554/eLife.32563.

Bag, J. and Wu, J. (1996) 'Translational Control of Poly(A)-Binding Protein Expression', European Journal of Biochemistry, 237(1), pp. 143–152. Available at: https://doi.org/10.1111/j.1432-1033.1996.0143n.x.

Bayne, R.A. et al. (2022) 'Yeast Ssd1 is a non-enzymatic member of the RNase II family with an alternative RNA recognition site', Nucleic Acids Research, 50(5), pp. 2923–2937. Available at: https://doi.org/10.1093/nar/gkab615.

Bonnin, P. et al. (2017) 'Novel mRNA-specific effects of ribosome drop-off on translation rate and polysome profile', PLOS Computational Biology, 13(5), p. e1005555. Available at: https://doi.org/10.1371/journal.pcbi.1005555.

Cambray, G., Guimaraes, J.C. and Arkin, A.P. (2018) 'Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli', Nature Biotechnology, 36(10), pp. 1005–1015. Available at: https://doi.org/10.1038/nbt.4238.

Cherry, J.M. (2015) 'The Saccharomyces Genome Database: A Tool for Discovery', Cold Spring Harbor Protocols, 2015(12), p. pdb.top083840. Available at: https://doi.org/10.1101/pdb.top083840.

Ciandrini, L., Crisostomo, R.L. and Szavits-Nossan, J. (2023) 'TASEPy: a Python-based package to iteratively solve the inhomogeneous exclusion process'. arXiv. Available at: https://doi.org/10.48550/arXiv.2308.00847.

García-García, C. et al. (2015) 'Factor-dependent processivity in human eIF4A DEAD-box helicase', Science (New York, N.Y.), 348(6242), pp. 1486–1488. Available at: https://doi.org/10.1126/science.aaa5089.

Garst, A.D., Edwards, A.L. and Batey, R.T. (2011) 'Riboswitches: Structures and Mechanisms', Cold Spring Harbor Perspectives in Biology, 3(6), p. a003533. Available at: https://doi.org/10.1101/cshperspect.a003533.

Gillespie, D.T. (1977) 'Exact stochastic simulation of coupled chemical reactions', The Journal of Physical Chemistry, 81(25), pp. 2340–2361. Available at: https://doi.org/10.1021/j100540a008.

Gupta, N., Lorsch, J.R. and Hinnebusch, A.G. (2018) 'Yeast Ded1 promotes 48S translation pre-initiation complex assembly in an mRNA-specific and eIF4F-dependent manner', eLife, 7, p. e38892. Available at: https://doi.org/10.7554/eLife.38892.

Harris, C.R. et al. (2020) 'Array programming with NumPy', Nature, 585(7825), pp. 357–362. Available at: https://doi.org/10.1038/s41586-020-2649-2.

Hinnebusch, A.G. (2011) 'Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes', Microbiology and Molecular Biology Reviews, 75(3), pp. 434–467. Available at: https://doi.org/10.1128/mmbr.00008-11.

Hunter, J.D. (2007) 'Matplotlib: A 2D Graphics Environment', Computing in Science & Engineering, 9(3), pp. 90–95. Available at: https://doi.org/10.1109/MCSE.2007.55.

Kluyver, T. et al. (2016) 'Jupyter Notebooks – a publishing format for reproducible computational workflows', in Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, pp. 87–90. Available at: https://doi.org/10.3233/978-1-61499-649-1-87.

Kozak, M. (1986) 'Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes', Cell, 44(2), pp. 283–292. Available at: https://doi.org/10.1016/0092-8674(86)90762-2.

Kozak, M. (1987) 'An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs', Nucleic Acids Research, 15(20), pp. 8125–8148. Available at: https://doi.org/10.1093/nar/15.20.8125.

Mangus, D.A., Evans, M.C. and Jacobson, A. (2003) 'Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression', Genome Biology, 4(7), p. 223. Available at: https://doi.org/10.1186/gb-2003-4-7-223.

Marintchev, A. et al. (2009) 'Topology and Regulation of the Human eIF4A/4G/4H Helicase Complex in Translation Initiation', Cell, 136(3), pp. 447–460. Available at: https://doi.org/10.1016/j.cell.2009.01.014.

de Melo Neto, O.P., Standart, N. and de Sa, C.M. (1995) 'Autoregulation of poly(A)-binding protein synthesis in vitro', Nucleic Acids Research, 23(12), pp. 2198–2205. Available at: https://doi.org/10.1093/nar/23.12.2198.

Möbius, W. and Laan, L. (2015) 'Physical and Mathematical Modeling in Experimental Papers', Cell, 163(7), pp. 1577–1583. Available at: https://doi.org/10.1016/j.cell.2015.12.006.

Muckenthaler, M., Gray, N.K. and Hentze, M.W. (1998) 'IRP-1 Binding to Ferritin mRNA Prevents the Recruitment of the Small Ribosomal Subunit by the Cap-Binding Complex eIF4F', Molecular Cell, 2(3), pp. 383–388. Available at: https://doi.org/10.1016/S1097-2765(00)80282-8.

Spirin, A.S. (2009) 'How Does a Scanning Ribosomal Particle Move along the 5′-Untranslated Region of Eukaryotic mRNA? Brownian Ratchet Model', Biochemistry, 48(45), pp. 10688–10692. Available at: https://doi.org/10.1021/bi901379a.

Stitzinger, S.H., Sohrabi-Jahromi, S. and Söding, J. (2023) 'Cooperativity boosts affinity and specificity of proteins with multiple RNA-binding domains', NAR Genomics and Bioinformatics, 5(2), p. lqad057. Available at: https://doi.org/10.1093/nargab/lqad057.

Pandas development team (2024) 'pandas-dev/pandas: Pandas'. Zenodo. Available at: https://doi.org/10.5281/zenodo.10957263.

Waclaw, B., Cholewa-Waclaw, J. and Greulich, P. (2019) 'Totally asymmetric exclusion process with site-wise dynamic disorder', Journal of Physics A: Mathematical and Theoretical, 52(6), p. 065002. Available at: https://doi.org/10.1088/1751-8121/aafb8a.

Waskom, M.L. (2021) 'seaborn: statistical data visualization', Journal of Open Source Software, 6(60), p. 3021. Available at: https://doi.org/10.21105/joss.03021.