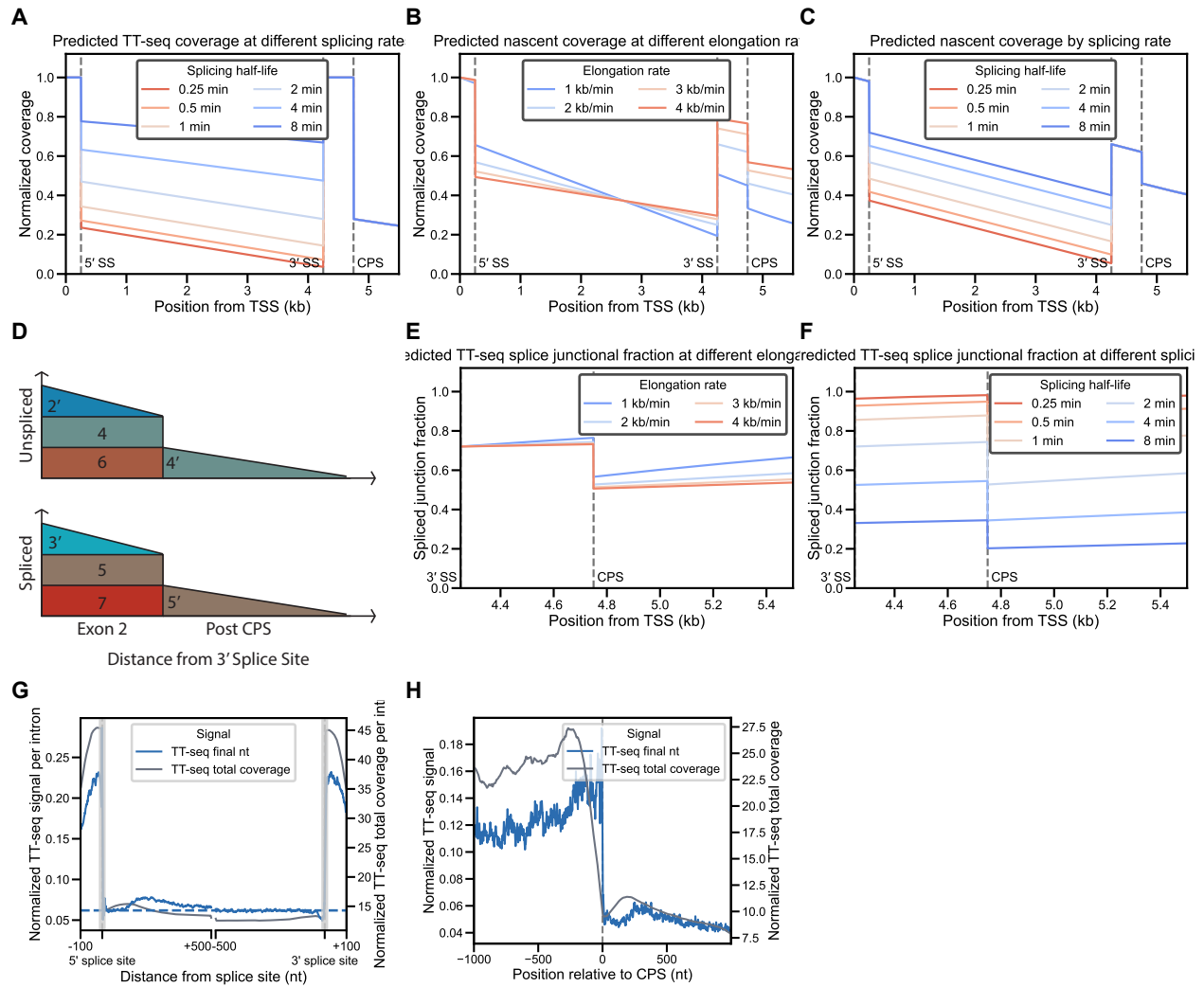
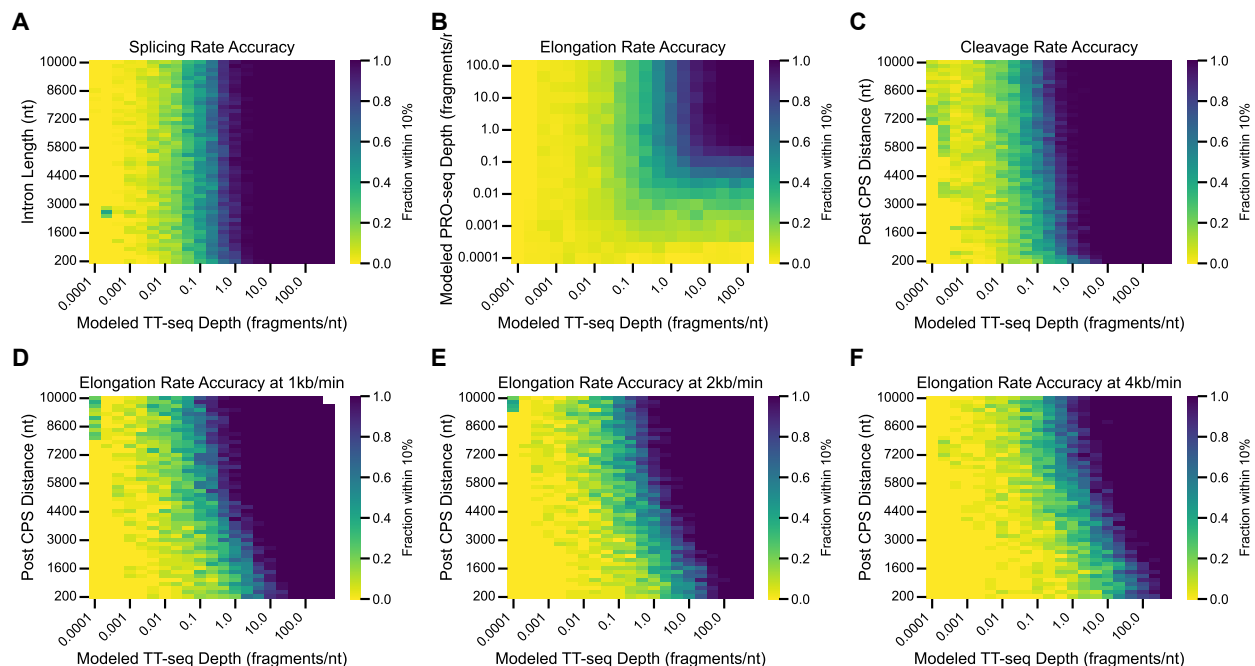
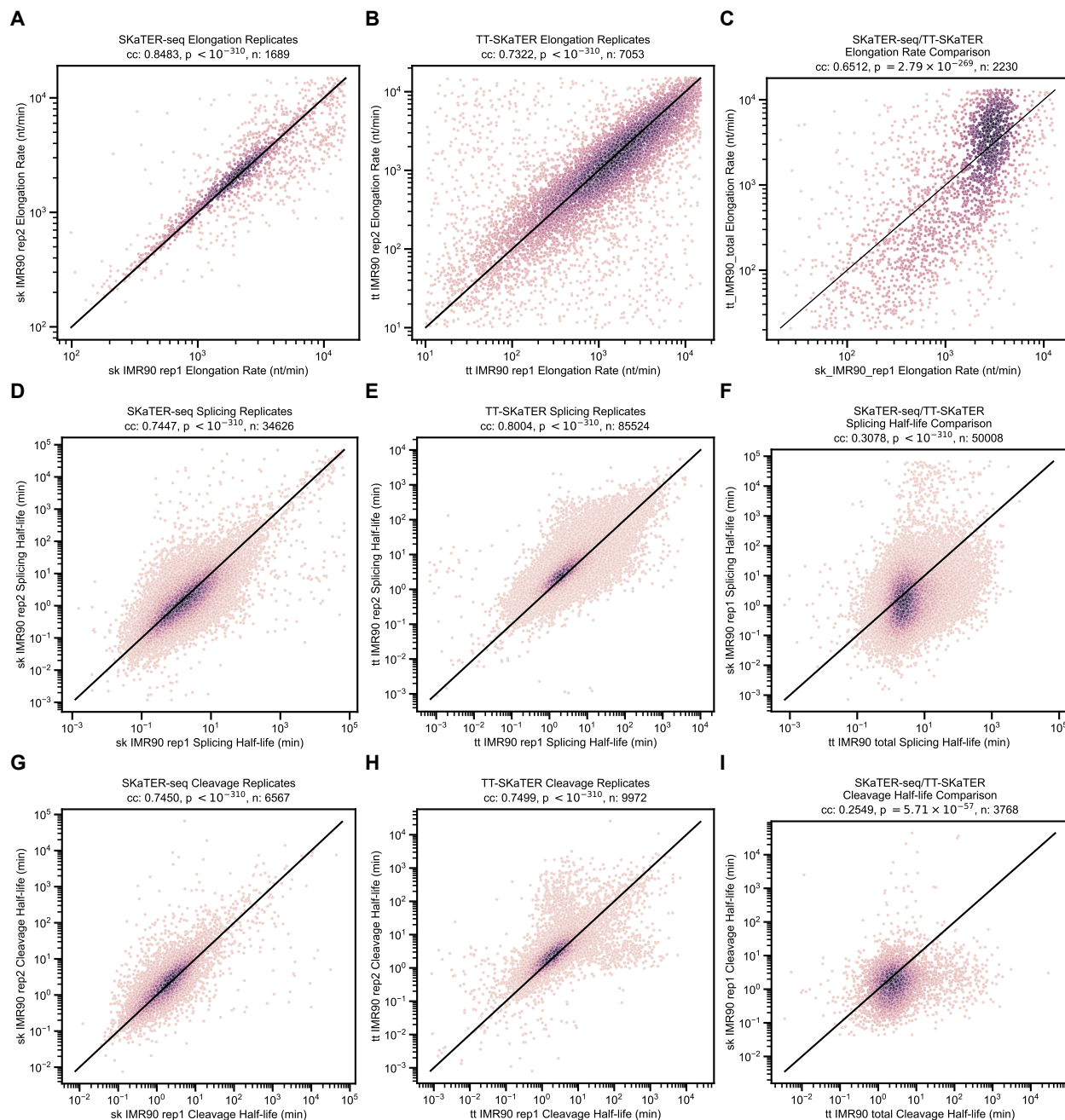




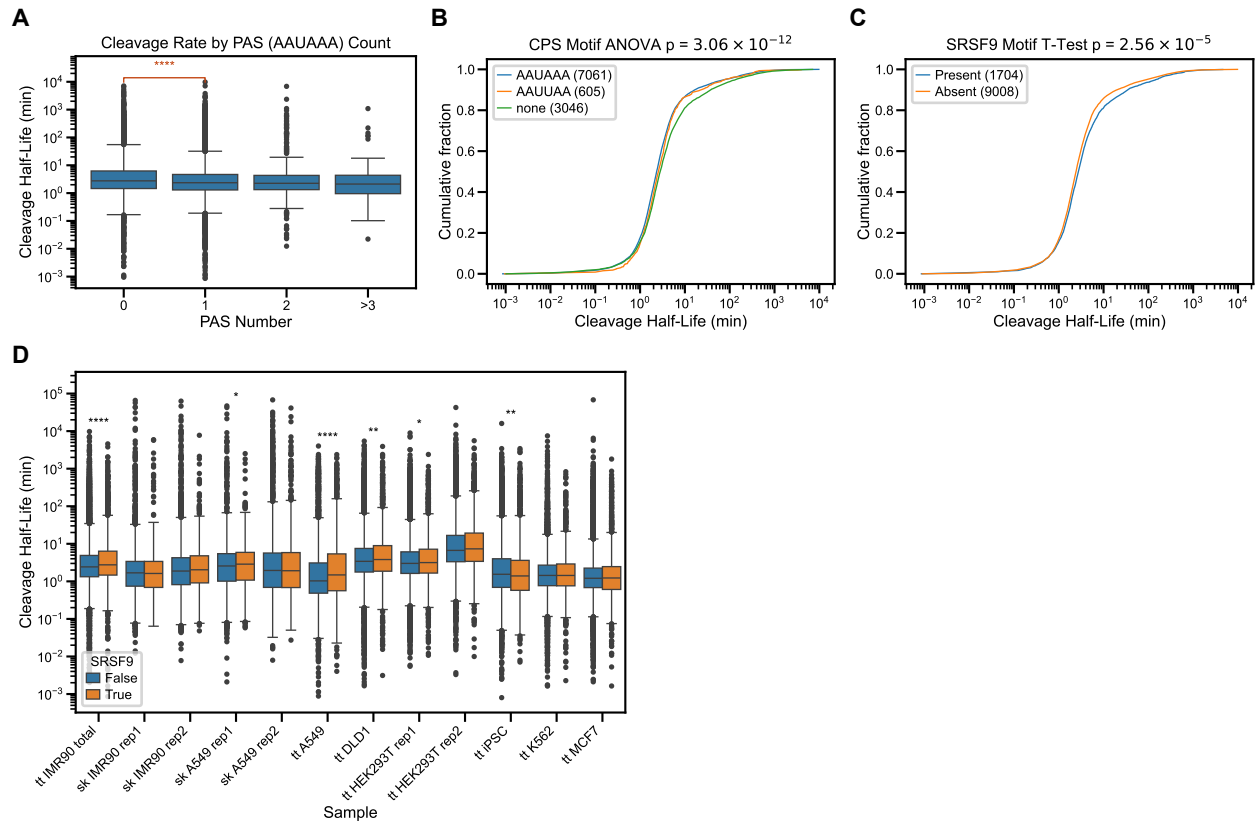
Supplemental Figure 1. Nascent transcription contains kinetic information: (A) Plot of changes in TT-seq coverage at constant elongation rate with variable splicing rates shows that higher splicing rates cause lower intronic coverage. (B-C) Plots of changes in nascent RNA coverage at different elongation or splicing rates respectively. These plots show that the slope of coverage within the intron should vary with elongation rate and the total amount of intronic coverage should vary with splicing rates. (D) A subset of states corresponding to unspliced or spliced mRNAs. (E-F) Plots of predicted splice junction fractions in TT-seq data at different elongation rates or splicing rates respectively. Nascent coverage yields identical images. (G) Intronic coverage of final nucleotides or total coverage in TT-seq. Final nucleotides increase emphasis on intron rather than flanking exons. (H) Coverage of final nucleotides or whole reads in TT-seq near the CPS. Final nucleotides reduce distortion right before the CPS.



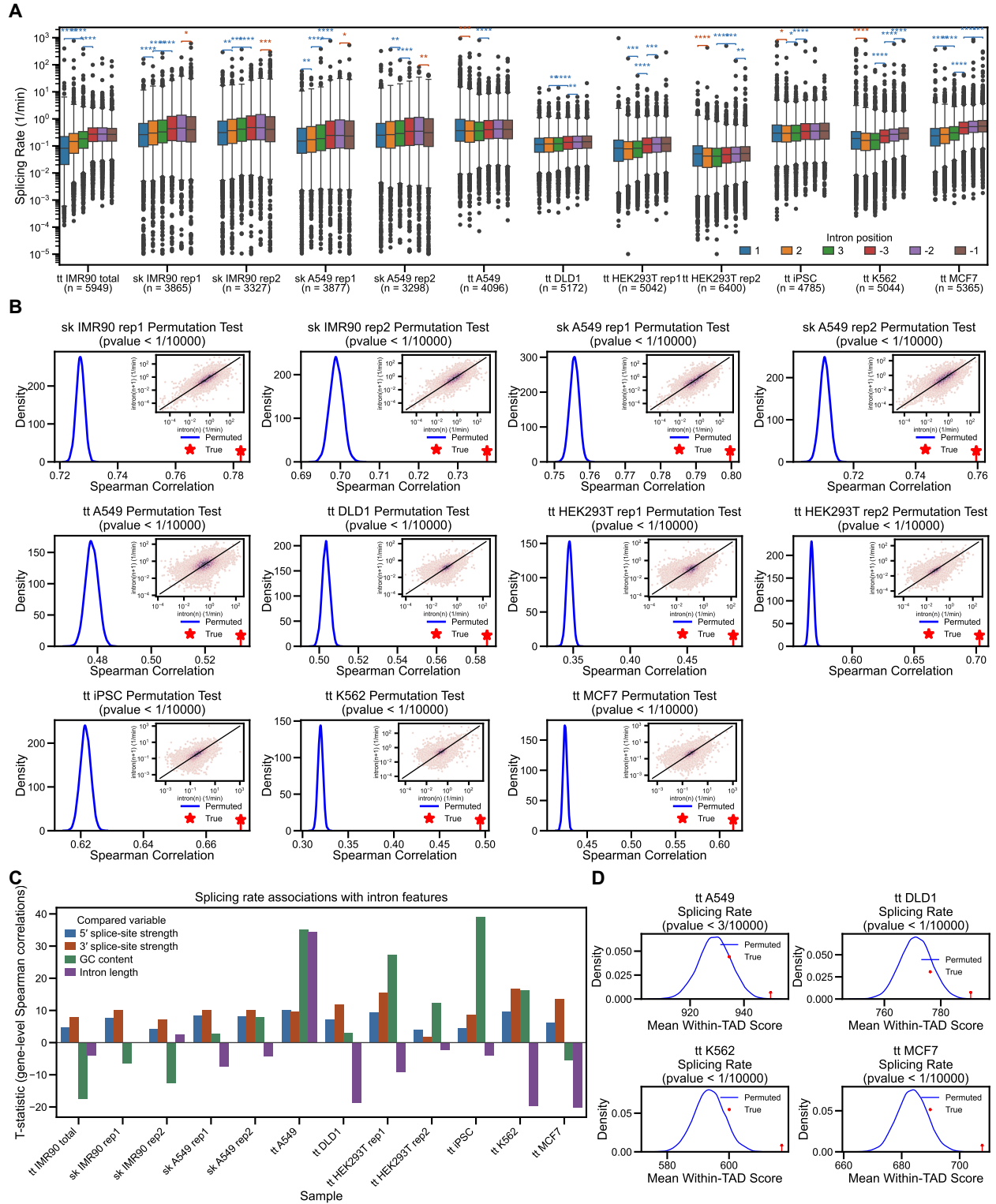
Supplemental Figure 2. Stochastic modeler identifies structural requirements to accurately fit kinetics.: (A) Splicing rate accuracy plotted as a function of TT-seq depth and intron length shows introns are accurately fit with increasing depth, regardless of length. (B) Elongation rate accuracy plotted as a function of TT-seq depth and PRO-seq depth shows TT-seq depth of 1 fragment/nt and PRO-seq depth of 0.1 fragment/nt is needed to accurately fit elongation rates. (C) Cleavage rate accuracy as a function of depth and distance past the PAS included in the model show cleavage rate are accurately fit with increasing depth, regardless of post PAS distance. (D-F) Elongation rate accuracy plotted against post CPS distance for a gene with a single 2 kb intron at elongation rates of 1kb/min, 2kb/min or 4kb/min respectively. At higher elongation rates, higher depth is required to fit rates.

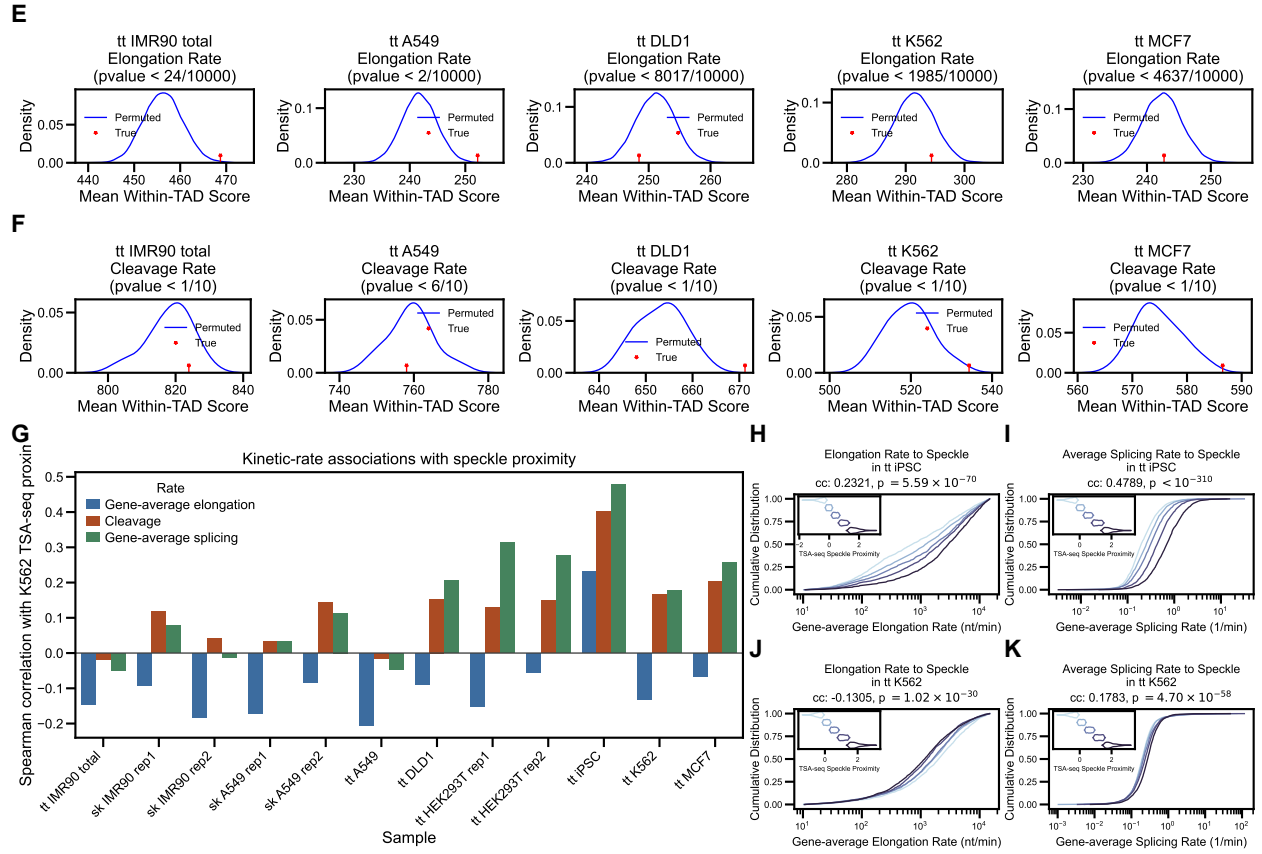


Supplemental Figure 3. SKaTER reliably measures kinetics from both SKaTER-seq and TT-seq/PRO-seq: (A) Correlation between two replicates of SKaTER-seq (B) and TT-SKaTER. (C) Correlation of elongation rates in 5 minute windows between SKaTER-seq replicate 1 and merged TT-SKaTER datasets. While data is highly correlated, we note elongation rates from TT-SKaTER are notably lower than SKaTER-seq rates. (D) Correlations between SKaTER-seq replicates (E) and TT-SKaTER replicates. (F) Correlation of splicing rates between SKaTER-seq rep 1 and TT-SKaTER showing a more loose correlation. (G) Correlations between SKaTER-seq replicates (H) and TT-SKaTER replicates. (I) Correlation of cleavage rates between SKaTER-seq rep 1 and TT-SKaTER showing a more loose correlation.

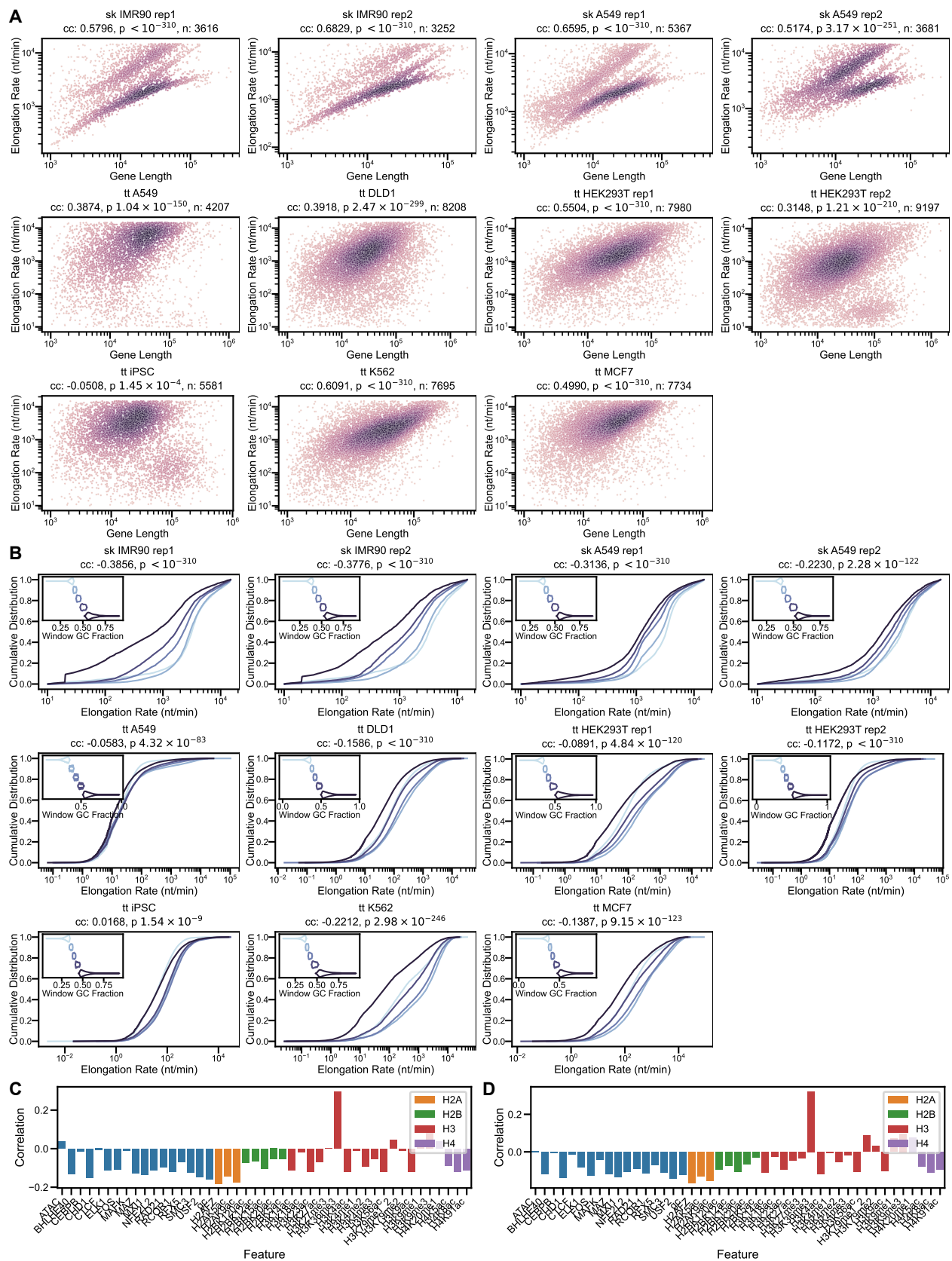


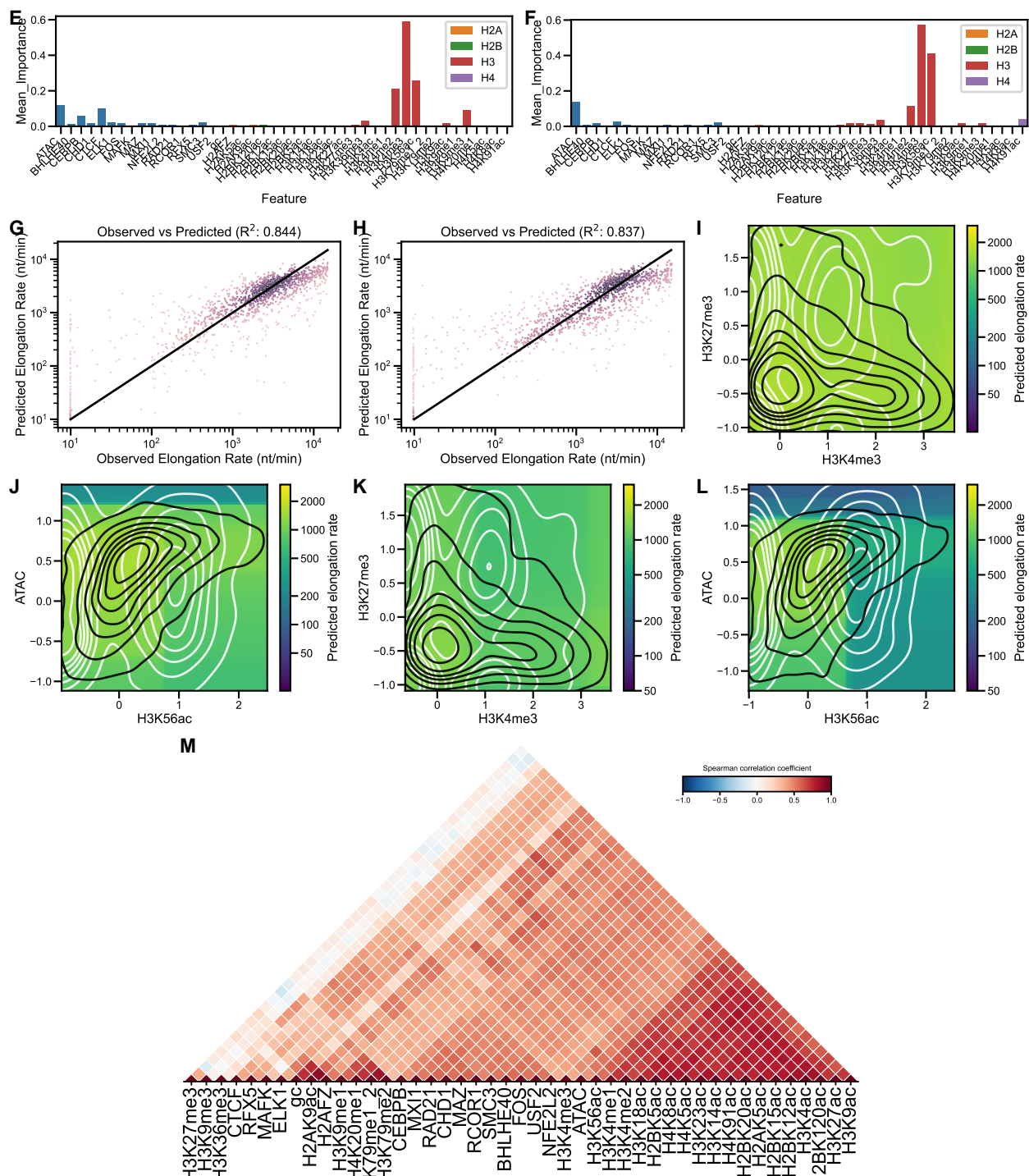
Supplemental Figure 4. Cleavage rates are associated with RNA motifs: (A) Cleavage rates are faster in the presence of a single PAS in IMR90 cells, but there is no improvement with additional PAS motifs. (B) CDF showing cleavage rates at sites with the canonical PAS, a secondary PAS (AAUUAA) or neither sequence show both canonical and secondary PAS result in faster cleavage. (C) CDF showing cleavage rates at sites with or without the SRSF9 binding motif. (D) Cleavage half-lives across samples, stratified by the presence of an SRSF9 binding motif.





Supplemental Figure 5. Splicing-rate correlations exceed chance across cell lines: (A) Splicing rates by relative intron position across datasets. (B) Intron-to-intron permutation tests across SKaTER-seq and TT-seq datasets. In each dataset, the observed correlation between adjacent introns is compared with the distribution obtained by randomly permuting intron pairings. (C) T-statistic correlations of splice site strengths, GC content and intron length to splicing rates across datasets. (D) TAD correlations for splicing rates in A549, DLD1, K562, and MCF7. (E) TAD correlations for elongation rates in IMR90, A549, DLD1, K562, and MCF7, respectively. (F) TAD correlations for cleavage rates in the same cell-lines. (G) Spearman correlations between K562 TSA-seq speckle proximity and gene-average elongation, cleavage, and gene-average splicing rates across datasets. (H–K) Gene-average elongation and splicing rates relative to K562 speckle proximity in iPSC and K562 datasets.





Supplemental Figure 6. Elongation rate associations across datasets: (A) Elongation rate versus gene length for the SKaTER-seq and TT-SKaTER datasets. (B) Elongation rate versus GC content for the same datasets. (C–D) Feature correlations with elongation rate in log-normalized IMR90 SKaTER-seq replicates 1 and 2. (E–F) Random-forest permutation importance for the same models. (G–H) Observed-versus-predicted rates for each replicate. (I–L) Two-dimensional partial dependence plots of chromatin marks. Overlaid on this heatmap in black is a KDE of the histone mark intensity of windows with fast elongation rates (over 3 kb/min) and a second KDE

of slow elongation rates (under 50 nt/min) in white. (M) Chromatin marks incorporated into the random forest model are highly correlated with each other.

Derivation of State Model

The Model

Nomenclature

Before beginning to discuss how we model a gene, we first need to define a set of terms to mathematically describe a gene. To do this, we start from the transcription start site and annotate each 5' splice site using F , each 3' splice site using T , and each cleavage site using C . For example, a gene with a single constitutive intron and single cleavage site could be written as FTC . However, this method is ambiguous if a gene has multiple splicing sites. For example, a gene with two constitutive introns and a cassette exon would both be written as $FTFTC$. Therefore, we can number each site and add the letter $J_{x,y}$ to the end of the name, here indicating that 5' splice site x and 3' splice site y can be sliced together. Now, a gene with a single constitutive intron would be written $F_0T_0C_0J_{0,0}$, a gene with two constitutive introns would be $F_0T_0F_1T_1C_0J_{0,0}J_{1,1}$ and a gene with a cassette exon would be $F_0T_0F_1T_1C_0J_{0,0}J_{0,1}J_{1,1}$.

Events

One of the major assumptions made for this model is the independence of splicing events. In other words, the splicing and cleavage status of introns do not affect each other as long as they are not alternative events (when splicing of one intron physically prevents the splicing of another). With this assumption, we can divide a gene into smaller mock genes called events. Each event is able to describe coverage for a particular region of a gene. For example, the gene $F_0T_0F_1T_1C_0J_{0,0}J_{1,1}$ be split into 3 events:

1. C_0 upstream of F_0 , between T_0 and F_1 , and downstream of T_1

2. $F_0T_0C_0J_{0_0}$ between F_0 and T_0

3. $F_1T_1C_0J_{1_1}$ between F_1 and T_1

Both $F_0T_0C_0J_{0_0}$ and $F_1T_1C_0J_{1_1}$ can be described by the same equations, so we will term their event $F_0T_0C_0J_{0_0}$. This modification is performed purely to increase floating point precision and reduce computation time.

States

mRNAs progress from newly initialized to fully spliced mature mRNA through multiple potential steps. Each of these steps is termed as a state. For example, event C_0 would have 3 states:

1. Nascent mRNA upstream of the cleavage site
2. Nascent mRNA downstream of the cleavage site
3. Cleaved mature mRNA

Meanwhile, event $F_0T_0C_0J_{0_0}$ would have 7 states:

1. Nascent mRNA upstream of 3' splice site
2. Unspliced nascent mRNA upstream of the cleavage site
3. Spliced nascent mRNA upstream of the cleavage site
4. Unspliced nascent mRNA downstream of the cleavage site
5. Spliced nascent mRNA downstream of the cleavage site
6. Unspliced but cleaved mRNA
7. Spliced and cleaved mature mRNA

As we can see, the number of states associated with a give event increases dramatically with each additional splicing or cleavage event.

mRNAs can move between states through elongation (transcribing a new 3' splice site or cleavage site), splicing of an intron, or cleavage at a CPS. In our model, mRNAs are assumed to take a fixed amount of time to reach each position in the gene. When that amount of time has elapsed since the mRNA initiates at the TSS, it moves into that new state. Meanwhile, splicing and cleavage are both modeled as an exponential decay. Once a mRNA elongates into a state where it becomes eligible for splicing (by transcribing the 3' splice site) or cleavage (by transcribing the CPS), it will exit that state at a fixed probability (unless it exits that state by splicing a different intron or transcribing another 3' splice site or cleavage site).

State Equations

The change in the number of mRNAs in any given state N at any time can be calculated as the sum of all possible ways to enter and exit the state. For a state with no entries or exits by splicing, this would be:

$$dN(t) = P_{\text{enter}} * dr(t - \tau_{\text{enter}}) - P_{\text{exit}} * dr(t - \tau_{\text{exit}})$$

Where t is the time since the gene activation (start of the experiment), $dN(t)$ is the change of number of mRNAs in state N as a function of time t , P_{enter} is the probability of entering state N by elongation, P_{exit} is the probability of exiting state N by elongation, τ_{enter} is time it takes to reach the entry position for state N from the TSS, τ_{exit} is the time it takes to reach the exit position for state N from the TSS, and $dr(t)$ is the rate of initiation of new RNA molecules at the TSS as a function of time t

Suppose state N can be accessed by splicing/cleavage from states N_i for $i \in I$ using splicing/cleavage rates of λ_i and can access other states via splicing or cleavage rates $j \in J$. The change in number of mRNAs would be combination of the number of mRNAs entering and

exiting by elongation, the number of mRNAs entering by splicing from a different state, and the number of mRNAs exiting by splicing to any of the states accessible from state N . This produces the following formula:

$$dN(t) = P_{\text{enter}} dr(t - \tau_{\text{enter}}) - P_{\text{exit}} dr(t - \tau_{\text{exit}}) + \sum_{i \in I} (\lambda_i N_i) - N * \sum_{j \in J} \lambda_j$$

To solve for this solution, we need to develop a formula for the probabilities P_{enter} and P_{exit} . This will be discussed in section 2. Additionally, we need a method of integrating for $N(t)$. In particular, we need to solve for two separate classes of states:

1. States with exits by splicing/cleavage (first order non-homogeneous differential equation)
2. States without exits by splicing/cleavage (solvable with standard integration)

The formulas for the integrated forms in DRB released (non-steady state) conditions will be derived in section 3 while the steady state versions will be derived in section 4.

Probabilities

Purely constitutive

Suppose we have a state N_0 that can be entered from state N'_0 with a given probability P'_0 . If state N_0 has no exits by splicing or cleavage, then the only path a mRNA that initially entered N_0 by elongation from N'_0 can take is to exit by elongation. We can define the probability of a mRNA taking this path within time range δ as $P(\emptyset)$. Thus, the probability P_0 of leaving state N_0 by elongation can be defined as:

$$P_0 = P_{0'} * P(\emptyset) = P'_0$$

States with exits by splicing or cleavage

Suppose instead we have state N_1 that has one exit by splicing, but only can be entered by elongation from N'_1 . We can define the splicing rate for the splicing of intron 1 as λ_1 . The splicing of this intron occurs at a random time x_1 , so to remain in N_1 we need to find the probability that x_1 is greater than δ . In this case, the probability P_1 of remaining in N_1 to exit by elongation is:

$$P_1 = P'_1 * P(x_1 > \delta)$$

As discussed above, the rate of splicing is modeled as an exponential decay. Therefore, we can simplify the above notation as:

$$P(x_1 > \delta) = \int_{\delta}^{\infty} \lambda_1 e^{-\lambda_1 x_1} dx_1 = e^{-\lambda_1 \delta}$$

So we have:

$$P_1 = P'_1 e^{-\lambda_1 \delta}$$

Entrance by splicing

Suppose we have a state N_2 that cannot be exited by splicing or cleavage, and we want to calculate the probability of exiting N_2 by elongation. If N_2 can only be entered by splicing from N_1 with rate λ_1 , we would need to calculate the probability of x_1 occurring before δ and would have:

$$P_2 = P'_1 * P(x_1 < \delta)$$

$$P(x_1 < \delta) = \int_0^{\delta} \lambda_1 * e^{-\lambda_1 x_1} dx_1 = 1 - e^{-\lambda_1 \delta}$$

$$P_2 = P'_1 (1 - e^{-\lambda_1 \delta})$$

If we also suppose that N_2 can be entered by elongation from N'_2 , then the probability would be the sum of both entering in from N_1 by splicing and remaining in N_2 .

$$P_2 = P'_1(1 - e^{-\lambda_1\delta}) + P_{2'}$$

As these probabilities are independent, we can say that the total probability of leaving N_2 by elongation is the sum of all the paths that can be taken to enter N_1 and remain there until the end of the interval δ .

Multiple introns

Now lets move on to discussing states with multiple splicing or cleavage connections. We can define a variable Λ_i as the sum of all the individual splicing or cleavage rates able to exit state N_i . For example, consider state N_1 which can be exited by splicing rates $\lambda_1, \dots, \lambda_n$. The splicing of each of these introns i occurs at random time x_i , so to remain in N_1 we need to find the probability that none of these times are greater than δ . In this case, the probability of remaining in N_1 to exit by elongation is:

$$P_1 = P'_1 * P(x_1, \dots, x_n > \delta)$$

$$P(x_1, \dots, x_n > \delta) = \int_{\delta}^{\infty} \dots \int_{\delta}^{\infty} \prod_{i=1}^n (\lambda_i * e^{-\lambda_i x_i}) dx_n \dots dx_1 = \prod_{i=1}^n e^{-\lambda_i \delta} = e^{-\delta \sum_{i=1}^n \lambda_i} = e^{-\Lambda_1 \delta}$$

So we have:

$$P_1 = P'_1 e^{-\Lambda_1 \delta}$$

In general, we can define the probability of remaining in a state N_a as follows:

$$p_{[a]} = e^{-\Lambda_a \delta}$$

Where Λ_a is the sum of all splicing or cleavage rates leaving state N_a .

Independent Introns

With these multiple introns in mind, now let's revisit state N_2 , which can only be entered by splicing intron 1 from N_1 . Suppose an mRNA in N_1 can leave by processing multiple introns of which none are in direct conflict. We would like to calculate the probability of x_1 occurring before δ while all other splicing states occur after δ . Note that if any other splicing event occurred, then the mRNA would move to a different state, not N_2 . The probability of this transition occurring before time δ is:

$$P_2 = P'_1 * P(x_1 < \delta < x_2, \dots, x_n)$$

Where we can define $P(x_1 < \delta < x_2, \dots, x_n)$ as:

$$\begin{aligned} P(x_1 < \delta < x_2, \dots, x_n) &= \left(\int_0^\delta \right) \dots \left(\int_\delta^\infty \right) \prod_{i=1}^n (\lambda_i e^{-\lambda_i x_i}) dx_n \dots dx_1 \\ &= \prod_{i=2}^{n(e^{-\lambda_i \delta})} \left(\int_0^\delta \right) \lambda_1 e^{-\lambda_1 x_1} dx_1 = (1 - e^{-\lambda_1 \delta}) e^{-(\Lambda_1 - \lambda_1) \delta} \end{aligned}$$

Substituting back in to the original equations gives:

$$P_2 = P'_1 (1 - e^{-\lambda_1 \delta}) e^{-(\Lambda_1 - \lambda_1) \delta}$$

However, as all the above introns are independent, we can simplify as follows:

$$\Lambda_1 = \Lambda_2 + \lambda_1$$

$$P_2 = P'_1 (1 - e^{-\lambda_1 \delta}) e^{-\Lambda_2 \delta}$$

$$P_2 = P'_1 (1 - e^{-(\Lambda_1 - \Lambda_2) \delta}) e^{-\Lambda_2 \delta}$$

Conflicting Introns

Now, let's assume that introns $k_1, \dots, k_m$ conflict with the splicing of intron 1. We can define set K_1 as the set of all introns that conflict with intron 1, and L as the set of all introns that do not conflict. We can then revise the equation as follows:

$$P_2 = P'_1 * P(x_1 < \delta, x_1 < x_i \forall i \in K_1, \delta < x_i \forall i \in L)$$

To simplify the above general expression, let's define a new variable $p_{[\text{start}, \dots, \text{end}]}$ representing the probability of starting in the first state and ending in the second state.:

$$p_{[1,2]} = P(x_1 < \delta, x_1 < x_i \forall i \in K_1, \delta < x_i \forall i \in L)$$

We can then solve for $p_{[1,2]}$ as follows:

$$\begin{aligned} p_{[1,2]} &= \int_0^\delta \int_{x_1}^\infty \dots \left(\int_\delta^\infty \right) \left(\prod_{i=1}^n \lambda_i e^{-\lambda_i x_i} \right) dx_n \dots dx_1 \\ &= \left(\prod_{i \in L} e^{-\lambda_i \delta} \right) \left(\int_0^\delta \right) \lambda_1 e^{-\lambda_1 x_1} \prod_{i \in K_1} (e^{-\lambda_i x_1}) dx_1 \\ &= \left(\prod_{i \in L} e^{-\lambda_i \delta} \right) \frac{\lambda_1}{\lambda_1 + \sum_{i \in K_1} \lambda_i} \left(1 - e^{-(\lambda_1 + \sum_{i \in K_1} \lambda_i) \delta} \right) \end{aligned}$$

Substituting this back in to the original equation for P_2 gives:

$$P_2 = P'_1 \left(\prod_{i \in L} e^{-\lambda_i \delta} \right) \frac{\lambda_1}{\lambda_1 + \sum_{i \in K_1} \lambda_i} \left(1 - e^{-(\lambda_1 + \sum_{i \in K_1} \lambda_i) \delta} \right)$$

However, all introns in L are still accessible from N_2 as they do not conflict with intron

1. Thus:

$$\Lambda_2 = \sum_{i \in L} \lambda_i$$

$$\Lambda_1 - \Lambda_2 = \lambda_1 + \sum_{i \in K_1} \lambda_i$$

As such we can simplify as follows:

$$p_{[1,2]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} (1 - e^{-(\Lambda_1 - \Lambda_2)\delta}) e^{-\Lambda_2\delta}$$

$$P_2 = P'_1 \frac{\lambda_1}{\Lambda_1 - \Lambda_2} (1 - e^{-(\Lambda_1 - \Lambda_2)\delta}) e^{-\Lambda_2\delta}$$

Note how this formula becomes identical to the solution derived in the previous section if all introns are independent (i.e. $\Lambda_0 = \Lambda_1 + \lambda_0$).

Expanding this out to arbitrary states a and b connected by splicing rate λ_a , we have a general formula for connection by one 'jump' as:

$$p_{[a,b]} = \frac{\lambda_a}{\Lambda_a - \Lambda_b} (1 - e^{-(\Lambda_a - \Lambda_b)\delta}) e^{-\Lambda_b\delta}$$

This form can similarly be rewritten as follows:

$$p_{[a,b]} = \frac{\lambda_a}{\Lambda_a - \Lambda_b} (e^{-\Lambda_b\delta} - e^{-\Lambda_a\delta})$$

$$p_{[a,b]} = \frac{\lambda_a}{\Lambda_a - \Lambda_b} (p_{[b]} - p_{[a]})$$

Splice twice without conflicting introns

Now suppose we want to access state N_3 from state N_1 by first splicing with rate λ_1 into N_2 and then by splicing with rate λ_2 into N_3 . First, let's assume that there are no conflicting introns (all other events are independent).

$$P_3 = P'_1 * P(x_1 < x_2 < \delta, \delta < x_3, \dots, x_n)$$

We can then solve for $P(x_1 < x_2 < \delta, \delta < x_3, \dots, x_n)$ as follows:

$$\begin{aligned}
P(x_1 < x_2 < \delta, \delta < x_3, \dots, x_n) &= \int_0^\delta \int_0^{x_2} \int_\delta^\infty \dots \int_\delta^\infty \left(\prod_{i=1}^n \lambda_i e^{-\lambda_i x_i} \right) dx_n \dots dx_1 dx_2 \\
&= \left(\prod_{i=3}^n e^{-\lambda_i \delta} \right) \int_0^\delta \int_0^{x_2} \lambda_1 \lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} dx_1 dx_2 \\
&= \left(\prod_{i=3}^n e^{-\lambda_i \delta} \right) \int_0^\delta \lambda_2 (1 - e^{-\lambda_1 x_2}) e^{-\lambda_1 x_2} dx_2 \\
&= \left(\prod_{i=3}^n e^{-\lambda_i \delta} \right) \int_0^\delta \lambda_2 (e^{-\lambda_2 x_2} - e^{-(\lambda_1 + \lambda_2) x_2}) dx_2 \\
&= \left(\prod_{i=3}^n e^{-\lambda_i \delta} \right) \left((1 - e^{-\lambda_2 \delta}) - \frac{\lambda_2}{\lambda_1 + \lambda_2} (1 - e^{-(\lambda_1 + \lambda_2) \delta}) \right) \\
&= e^{-\Lambda_3 \delta} \left((1 - e^{-\lambda_2 \delta}) - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (1 - e^{-(\Lambda_1 - \Lambda_3) \delta}) \right) \\
&= \left((e^{-\Lambda_3 \delta} - e^{-\Lambda_2 \delta}) - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (e^{-\Lambda_3 \delta} - e^{-\Lambda_1 \delta}) \right)
\end{aligned}$$

We can then substitute this formula back in to solve for P_3 :

$$\begin{aligned}
P_3 &= P_1' e^{-\Lambda_3 \delta} \left((1 - e^{-\lambda_2 \delta}) - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (1 - e^{-(\Lambda_1 - \Lambda_3) \delta}) \right) \\
&= P_1' \left((e^{-\Lambda_3 \delta} - e^{-\Lambda_2 \delta}) - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (e^{-\Lambda_3 \delta} - e^{-\Lambda_1 \delta}) \right)
\end{aligned}$$

Splice twice with conflicting introns

Now to increase complexity, let's assume that each intron λ_i conflicts with introns of set K_i . Remaining introns that are not conflicting with any state are in set L . Some splicing events can be mutually exclusive with other splicing events, so they are assigned to the set of their FIRST mutually exclusive event in this path. As a reminder, the following formula still holds:

$$\Lambda_i = \lambda_i + \sum_{j \in K_i} \lambda_j + \sum_{j \in L} \lambda_j$$

Rewriting the above formula with these added categories gives the following: $P_3 = P'_1 *$

$$P(x_1 < x_2 < \delta, x_1 < x_i \forall i \in K_1, x_2 < x_i \forall i \in K_2, \delta < x_i \forall i \in L)$$

Using the simplified notation from above, we can collapse this into the following:

$$p_{[1,2,3]} = P(x_1 < x_2 < \delta, x_1 < x_i \forall i \in K_1, x_2 < x_i \forall i \in K_2, \delta < x_i \forall i \in L)$$

Writing out the equations gives:

$$\begin{aligned}
p_{[1,2,3]} &= \int_0^\delta \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \int_{x_1}^\infty \dots \int_{x_1}^\infty \int_\delta^\infty \dots \int_\delta^\infty \prod_{i \in L} (\lambda_i e^{-\lambda_i x_i} dx_i) \prod_{j=1}^2 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= \prod_{i \in L} (e^{-\lambda_i \delta}) \int_0^\delta \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \int_{x_1}^\infty \dots \int_{x_1}^\infty \prod_{j=1}^2 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= e^{-\Lambda_3 \delta} \int_0^\delta \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \int_{x_1}^\infty \dots \int_{x_1}^\infty \prod_{j=1}^2 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= e^{-\Lambda_3 \delta} \int_0^\delta \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \lambda_1 e^{-(\lambda_1 + \sum_{i \in K_1} \lambda_i) x_1} dx_1 \prod_{i \in K_2} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_2 e^{-\lambda_2 x_2} dx_2 \\
&= e^{-\Lambda_3 \delta} \int_0^\delta \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \lambda_1 e^{-(\Lambda_1 - \Lambda_2) x_1} dx_1 \prod_{i \in K_2} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_2 e^{-\lambda_2 x_2} dx_2 \\
&= e^{-\Lambda_3 \delta} \int_0^\delta \int_{x_2}^\infty \dots \int_{x_2}^\infty \frac{\lambda_1}{\Lambda_1 - \Lambda_2} (1 - e^{-(\Lambda_1 - \Lambda_2) x_2}) \prod_{i \in K_2} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_2 e^{-\lambda_2 x_2} dx_2 \\
&= e^{-\Lambda_3 \delta} \int_0^\delta \frac{\lambda_1}{\Lambda_1 - \Lambda_2} (1 - e^{-(\Lambda_1 - \Lambda_2) x_2}) \lambda_2 e^{-(\lambda_2 + \sum_{i \in K_2} \lambda_i) x_2} dx_2 \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} e^{-\Lambda_3 \delta} \int_0^\delta (1 - e^{-(\Lambda_1 - \Lambda_2) x_2}) \lambda_2 e^{-(\Lambda_2 - \Lambda_3) x_2} dx_2 \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} e^{-\Lambda_3 \delta} \int_0^\delta \lambda_2 (e^{-(\Lambda_2 - \Lambda_3) x_2} - e^{-(\Lambda_1 - \Lambda_3) x_2}) dx_2 \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} e^{-\Lambda_3 \delta} \left(\frac{\lambda_2}{\Lambda_2 - \Lambda_3} (1 - e^{-(\Lambda_2 - \Lambda_3) \delta}) - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (1 - e^{-(\Lambda_1 - \Lambda_3) \delta}) \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(\frac{\lambda_2}{\Lambda_2 - \Lambda_3} (1 - e^{-(\Lambda_2 - \Lambda_3) \delta}) e^{-\Lambda_3 \delta} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (1 - e^{-(\Lambda_1 - \Lambda_3) \delta}) e^{-\Lambda_3 \delta} \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(\frac{\lambda_2}{\Lambda_2 - \Lambda_3} (e^{-\Lambda_3 \delta} - e^{-\Lambda_2 \delta}) - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (e^{-\Lambda_3 \delta} - e^{-\Lambda_1 \delta}) \right)
\end{aligned}$$

Using the general equations above, we can simplify as follows:

$$p_{[1,2,3]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(\frac{\lambda_2}{\Lambda_2 - \Lambda_3} (p_{[3]} - p_{[2]}) - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (p_{[3]} - p_{[1]}) \right)$$

$$p_{[1,2,3]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (p_{[3]} - p_{[1]}) \right)$$

Multiple splicing with conflicting introns

We can repeat this process for N_4 as follows:

$$\begin{aligned}
p_{[1,2,3,4]} &= \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \int_{x_1}^\infty \dots \int_{x_1}^\infty \int_\delta^\infty \dots \\
&\quad \int_\delta^\infty \prod_{i \in L} (\lambda_i e^{-\lambda_i x_i} dx_i) \prod_{j=1}^2 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= \prod_{i \in L} (e^{-\lambda_i \delta}) \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \int_{x_1}^\infty \dots \int_{x_1}^\infty \prod_{j=1}^3 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= e^{-\Lambda_3 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \int_{x_1}^\infty \dots \int_{x_1}^\infty \prod_{j=1}^3 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= e^{-\Lambda_3 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} \int_{x_2}^\infty \dots \int_{x_2}^\infty \int_0^{x_2} \lambda_1 e^{-(\Lambda_1 - \Lambda_2)x_1} dx_1 \prod_{j=2}^3 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} e^{-\Lambda_3 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} \int_{x_2}^\infty \dots \int_{x_2}^\infty (1 - e^{-(\Lambda_1 - \Lambda_2)x_2}) \prod_{j=2}^3 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - e^{-\Lambda_4 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} \int_{x_2}^\infty \dots \int_{x_2}^\infty e^{-(\Lambda_1 - \Lambda_2)x_2} \prod_{j=2}^3 \prod_{i \in K_j} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_j e^{-\lambda_j x_j} dx_j \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - e^{-\Lambda_4 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} e^{-(\Lambda_1 - \Lambda_2)x_2} \lambda_2 e^{-(\Lambda_2 - \Lambda_3)x_2} dx_2 \prod_{i \in K_3} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_3 e^{-\lambda_3 x_3} dx_3 \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - e^{-\Lambda_4 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty \int_0^{x_3} \lambda_2 e^{-(\Lambda_1 - \Lambda_3)x_2} dx_2 \prod_{i \in K_3} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_3 e^{-\lambda_3 x_3} dx_3 \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} e^{-\Lambda_4 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty (1 - e^{-(\Lambda_1 - \Lambda_3)x_3}) \prod_{i \in K_3} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_3 e^{-\lambda_3 x_3} dx_3 \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4]} - e^{-\Lambda_4 \delta} \int_0^\delta \int_{x_3}^\infty \dots \int_{x_3}^\infty e^{-(\Lambda_1 - \Lambda_3)x_3} \prod_{i \in K_3} (\lambda_i e^{-\lambda_i x_i} dx_i) \lambda_3 e^{-\lambda_3 x_3} dx_3 \right) \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4]} - e^{-\Lambda_4 \delta} \int_0^\delta \lambda_3 e^{-(\Lambda_1 - \Lambda_4)x_3} dx_3 \right) \right) \\
&= \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4]} - \frac{\lambda_3}{\Lambda_1 - \Lambda_4} e^{-\Lambda_4 \delta} (1 - e^{-(\Lambda_1 - \Lambda_4)\delta}) \right) \right)
\end{aligned}$$

We can repeat this process for N_5 results in the following:

$$p_{[1,2,3,4,5]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4,5]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4,5]} - \frac{\lambda_3}{\Lambda_1 - \Lambda_4} \left(p_{[4,5]} - \frac{\lambda_4}{\Lambda_1 - \Lambda_5} (p_{[5]} - p_{[1]}) \right) \right) \right)$$

To summarize, we have the following probabilities associated particular paths from 1 up to 5 states:

$$p_{[1]} = e^{-\Lambda_1 \delta}$$

$$p_{[1,2]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} (p_{[2]} - p_{[1]})$$

$$p_{[1,2,3]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (p_{[3]} - p_{[1]}) \right)$$

$$p_{[1,2,3,4]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4]} - \frac{\lambda_3}{\Lambda_1 - \Lambda_4} (p_{[4]} - p_{[1]}) \right) \right)$$

$$p_{[1,2,3,4,5]} = \frac{\lambda_1}{\Lambda_1 - \Lambda_2} \left(p_{[2,3,4,5]} - \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4,5]} - \frac{\lambda_3}{\Lambda_1 - \Lambda_4} \left(p_{[4,5]} - \frac{\lambda_4}{\Lambda_1 - \Lambda_5} (p_{[5]} - p_{[1]}) \right) \right) \right)$$

To try and simplify these formulas, we can define a new variable $q_{a:[b,\dots,c]}$ as follows:

$$q_{1:[2,3]} = \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (p_{[3]} - p_{[1]})$$

$$q_{1:[2,3,4]} = \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4]} - \frac{\lambda_3}{\Lambda_1 - \Lambda_4} (p_{[4]} - p_{[1]}) \right) = \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (p_{[3,4]} - q_{1:[3,4]})$$

$$q_{1:[2,3,4,5]} = \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4,5]} - \frac{\lambda_3}{\Lambda_1 - \Lambda_4} \left(p_{[4,5]} - \frac{\lambda_4}{\Lambda_1 - \Lambda_5} (p_{[5]} - p_{[1]}) \right) \right)$$

$$= \frac{\lambda_2}{\Lambda_1 - \Lambda_3} \left(p_{[3,4,5]} - \frac{\lambda_3}{\Lambda_1 - \Lambda_4} (p_{[4,5]} - q_{1:[4,5]}) \right)$$

$$= \frac{\lambda_2}{\Lambda_1 - \Lambda_3} (p_{[3,4,5]} - q_{1:[3,4,5]})$$

General formulas

Based on the above equations, we can derive a general formula for the probability of taking particular path $[a, \dots]$ starting from state N_a as follows:

$$p_{[a]} = e^{-\Lambda_a \delta}$$

$$p_{[a,b]} = \frac{\lambda_a}{\Lambda_a - \Lambda_b} (p_{[b]} - p_{[a]})$$

$$p_{[a,b,\dots,n]} = \frac{\lambda_a}{\Lambda_a - \Lambda_b} (p_{[b,\dots,n]} - q_{a:[b,\dots,n]})$$

$$q_{a:[b,c]} = \frac{\lambda_b}{\Lambda_a - \Lambda_c} (p_{[c]} - p_{[a]})$$

$$q_{a:[b,c,\dots,n]} = \frac{\lambda_b}{\Lambda_a - \Lambda_c} (p_{[c,\dots,n]} - q_{a:[c,\dots,n]})$$

The probability P_n of entering any state N_n by elongation after time δ is the sum of all the paths $x \in \mathbb{X}$ able to reach that state multiplied by the probability of initially entering the state of origin N_{x_0} by elongation P'_{x_0} .

$$P_n = \sum_{x \in \mathbb{X}} (P'_{x_0} * p_x)$$

State Integration

RNA Production Rate

Suppose we have a number of TSSs G , all of which are off and blocked by DRB at the start of the experiment. At time $t = 0$, the DRB is washed out, and the TSSs will randomly reactivate at rate Λ_0 . Once reactivated, new transcripts will initiate at the TSS with rate λ_n . Therefore, we can define a rate of new mRNA production $dr(t)$ after time $t = 0$ as:

$$dr(t) = G\lambda_n * (1 - e^{-\Lambda_0 t}) + G\Lambda_0 e^{-\Lambda_0 t}$$

Rearranging this formula gives:

$$dr(t) = \begin{cases} 0 & t < 0 \\ G\lambda_n - G(\lambda_n - \Lambda_0)e^{-\Lambda_0 t} & t \geq 0 \end{cases}$$

As discussed above, the state model produces equations of the following form:

$$dN(t) = P_{\text{enter}}dr(t - \tau_{\text{enter}}) - P_{\text{exit}}dr(t - \tau_{\text{exit}}) + \sum_{i \in I} (\lambda_i N_i) - N * \sum_{j \in J} \lambda_j$$

These equations are first order non-homogeneous differential equations, which have a general solution that we can use to develop a set of formulas for all possible state equations, similar to what we did with the probabilities. However, there is also a special case where there are no exits from state N via splicing/cleavage, meaning the equation can be solved directly by integration. This case will be discussed towards the end of the section.

Event with exits by splicing

Suppose we have state N_1 which can be exited by splicing with a total rate of Λ_1 , where Λ_1 is the sum of all splicing/cleavage rates for events exiting N_1 . Suppose that state N_1 can be entered by elongation at time τ_0 and exited by elongation at time τ_1 . The probability of entering N_1 by elongation can be defined as P_{enter} and exiting as P_{exit} , which are derived above in section 2. The state equation for N_1 would be:

$$\frac{dN_1}{dt} = P_{\text{enter}}dr(t - \tau_0) - P_{\text{exit}}dr(t - \tau_1) - \Lambda_1 N_1$$

This is a first order non-homogeneous differential equation of the form $y'(x) + p(x)y = f(x)$. Equations of this type can be solved as:

$$y(x) = e^{-\int p(x)} \left(A + \int (e^{\int p(x)}) f(x) \right)$$

Substituting in for our equation, we have:

$$N_1 = e^{-\Lambda_1 t} \left(A + \int (e^{\Lambda_1 t}) (P_{\text{enter}} dr(t - \tau_0) - P_{\text{exit}} dr(t - \tau_1)) \right)$$

As $dr(t = 0) = 0$, we have

$$N_1(0) = A$$

. Since there are no mRNAs present when transcription has not yet started,

$$A = 0$$

. Thus, we have:

$$N_1 = e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) (P_{\text{enter}} dr(t - \tau_0) - P_{\text{exit}} dr(t - \tau_1))$$

We can simplify this further as :

$$N_1 = P_{\text{enter}} f_{[1]}(t, \tau_0) - P_{\text{exit}} f_{[1]}(t, \tau_1)$$

$$f_{[1]}(t, \tau) = e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) dr(t - \tau)$$

If instead there was no exit to state N_1 by elongation, then $P_{\text{exit}} = 0$. We have a similar result if there is no entrance by elongation.

If we instead take state N_2 that can be reached by state N_1 by splicing with rate λ_1 , we would have:

$$N_2 = P_{\text{enter } 2} f_{[2]}(t, \tau_0) - P_{\text{exit } 2} f_{[2]}(t, \tau_1) + e^{-\Lambda_2 t} \int (e^{\Lambda_2 t}) \lambda_1 N_1$$

Simplifying would give:

$$N_2 = P_{\text{enter } 2} f_{[2]}(t, \tau_0) - P_{\text{exit } 2} f_{[2]}(t, \tau_1) + P_{\text{enter } 1} f_{[1,2]}(t, \tau_0) - P_{\text{exit } 1} f_{[1,2]}(t, \tau_1)$$

$$\begin{aligned}
f_{[1,2]}(t, \tau) &= e^{-\Lambda_2 t} \int (e^{\Lambda_2 t}) \lambda_1 f_{[1]}(t, \tau) \\
&= e^{-\Lambda_2 t} \int (e^{\Lambda_2 t}) \lambda_1 e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) dr (t - \tau)
\end{aligned}$$

This can continue to extrapolate out for even longer chains of connections. However, to solve all of these equations, we need to develop a solution for the equation $f_{[a,\dots]}(t, \tau)$.

First step

We start with trying to solve for equations with only one integration $f_{[1]}(t, \tau) = e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) dr (t - \tau)$

First, we substitute in $dr(x)$, which requires splitting the integration into two separate integrals

$$f_{[1]}(t, \tau) = e^{-\Lambda_1 t} \int_0^\tau (e^{\Lambda_1 t})(0) + e^{-\Lambda_1 t} \int_\tau^t (e^{\Lambda_1 t})(G\lambda_n - G(\lambda_n - \Lambda_0)e^{-\Lambda_0(t-\tau)})$$

We can then solve as follows:

$$\begin{aligned}
f_{[1]}(t, \tau) &= G\lambda_n e^{-\Lambda_1 t} \int_\tau^t (e^{\Lambda_1 t}) - e^{\Lambda_0 \tau - \Lambda_1 t} G(\lambda_n - \Lambda_0) \int_\tau^t (e^{(\Lambda_1 - \Lambda_0)t}) \\
&= \frac{G\lambda_n}{\Lambda_1} e^{-\Lambda_1 t} (e^{\Lambda_1 t} - e^{\Lambda_1 \tau}) - \frac{G(\lambda_n - \Lambda_0)}{\Lambda_1 - \Lambda_0} e^{\Lambda_0 \tau - \Lambda_1 t} (e^{(\Lambda_1 - \Lambda_0)t} - e^{(\Lambda_1 - \Lambda_0)\tau}) \\
&= \frac{G\lambda_n}{\Lambda_1} e^{-\Lambda_1 t} (e^{\Lambda_1 t} - e^{\Lambda_1 \tau}) - \frac{G(\lambda_n - \Lambda_0)}{\Lambda_1 - \Lambda_0} e^{\Lambda_0 \tau - \Lambda_1 t} (e^{(\Lambda_1 - \Lambda_0)t} - e^{(\Lambda_1 - \Lambda_0)\tau}) \\
&= \frac{G\lambda_n}{\Lambda_1} (1 - e^{\Lambda_1(\tau-t)}) - \frac{G(\lambda_n - \Lambda_0)}{\Lambda_1 - \Lambda_0} (e^{\Lambda_0(\tau-t)} - e^{\Lambda_1(\tau-t)}) \\
&= \frac{G\lambda_n}{\Lambda_1} - \left(\frac{G\lambda_n}{\Lambda_1} - \frac{G(\lambda_n - \Lambda_0)}{\Lambda_1 - \Lambda_0} \right) e^{\Lambda_1(\tau-t)} - \frac{G(\lambda_n - \Lambda_0)}{\Lambda_1 - \Lambda_0} e^{\Lambda_0(\tau-t)}
\end{aligned}$$

We can now define a variable

C

as follows:

$$C_{[0]} = G(\lambda_n - \Lambda_0)$$

$$C_{[0,1]} = \frac{G\lambda_n}{\Lambda_1} - \frac{C_{[0]}}{\Lambda_1 - \Lambda_0}$$

This simplifies the solution for $f_{[1]}(t, \tau)$ where $t > \tau$ as follows:

$$f_{[1]}(t, \tau) = \frac{G\lambda_n}{\Lambda_1} - C_{[0,1]}e^{\Lambda_1(\tau-t)} - \frac{C_{[0]}}{\Lambda_1 - \Lambda_0}e^{\Lambda_0(\tau-t)}$$

Second step

Above, we defined $f_{[1,2]}$ as:

$$f_{[1,2]}(t, \tau) = e^{-\Lambda_2 t} \int (e^{\Lambda_2 t}) \lambda_1 f_{[1]}(t, \tau)$$

We can then solve as follows:

$$\begin{aligned} f_{[1,2]}(t, \tau) &= e^{-\Lambda_2 t} \int (e^{\Lambda_2 t}) \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1} - C_{[0,1]}e^{\Lambda_1(\tau-t)} - \frac{C_{[0]}}{\Lambda_1 - \Lambda_0}e^{\Lambda_0(\tau-t)} \right) \\ &= \lambda_1 e^{-\Lambda_2 t} \left(\int (e^{\Lambda_2 t}) \frac{G\lambda_n}{\Lambda_1} - C_{[0,1]}e^{\Lambda_1(\tau-t)} - \frac{C_{[0]}}{\Lambda_1 - \Lambda_0}e^{\Lambda_0(\tau-t)} \right) \end{aligned}$$

As above, the state must be empty when $t < \tau$. Thus, we simplify the integration

$$\begin{aligned}
f_{[1,2]}(t, \tau) &= \lambda_1 e^{-\Lambda_2 t} \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} (e^{\Lambda_2 t} - e^{\Lambda_2 \tau}) - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{\Lambda_1 \tau} (e^{\Lambda_2 t} e^{-\Lambda_1 t} - e^{\Lambda_2 \tau} e^{-\Lambda_1 \tau}) \right. \\
&\quad \left. - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{\Lambda_0 \tau} (e^{\Lambda_2 t} e^{-\Lambda_0 t} - e^{\Lambda_2 \tau} e^{-\Lambda_0 \tau}) \right) \\
&= \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} (e^{\Lambda_2 t} - e^{\Lambda_2 \tau}) e^{-\Lambda_2 t} - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{\Lambda_1 \tau} e^{-\Lambda_2 t} (e^{\Lambda_2 t} e^{-\Lambda_1 t} - e^{\Lambda_2 \tau} e^{-\Lambda_1 \tau}) \right. \\
&\quad \left. - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{\Lambda_0 \tau} e^{-\Lambda_2 t} (e^{\Lambda_2 t} e^{-\Lambda_0 t} - e^{\Lambda_2 \tau} e^{-\Lambda_0 \tau}) \right) \\
&= \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{G\lambda_n}{\Lambda_1 \Lambda_2} e^{-\Lambda_2(t-\tau)} - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{-\Lambda_1(t-\tau)} + \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{-\Lambda_2(t-\tau)} \right. \\
&\quad \left. - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{-\Lambda_0(t-\tau)} + \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{-\Lambda_2(t-\tau)} \right) \\
&= \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} \right) e^{-\Lambda_2(t-\tau)} - \right. \\
&\quad \left. \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{-\Lambda_1(t-\tau)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{-\Lambda_0(t-\tau)} \right)
\end{aligned}$$

Similarly, we can define:

$$C_{[0,1,2]} = \frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)}$$

Thus, we get:

$$f_{[1,2]}(t, \tau) = \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - C_{[0,1,2]} e^{-\Lambda_2(t-\tau)} - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{-\Lambda_1(t-\tau)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{-\Lambda_0(t-\tau)} \right)$$

Third step

Similarly, we can define $f_{[1,2,3]}$ as:

$$f_{[1,2,3]}(t, \tau) = e^{-\Lambda_3 t} \int (e^{\Lambda_3 t}) \lambda_2 f_{[1,2]}(t, \tau)$$

We can solve similar to the previous step (again only integrating over $t > \tau$)

$$\begin{aligned} f_{[1,2,3]}(t, \tau) &= e^{-\Lambda_3 t} \int_{\tau}^{t(e^{\Lambda_3 t})} \lambda_2 \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - C_{[0,1,2]} e^{-\Lambda_2(t-\tau)} \right. \\ &\quad \left. - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{-\Lambda_1(t-\tau)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{-\Lambda_0(t-\tau)} \right) dt \\ &= \lambda_2 \lambda_1 e^{-\Lambda_3 t} \left(\int_{\tau}^{t(e^{\Lambda_3 t})} \frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \int_{\tau}^{t(e^{\Lambda_3 t})} C_{[0,1,2]} e^{-\Lambda_2(t-\tau)} \right. \\ &\quad \left. - \int_{\tau}^{t(e^{\Lambda_3 t})} \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} e^{-\Lambda_1(t-\tau)} - \int_{\tau}^{t(e^{\Lambda_3 t})} \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)} e^{-\Lambda_0(t-\tau)} \right) \\ &= \lambda_2 \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C_{[0,1,2]}}{\Lambda_3 - \Lambda_2} \right. \right. \\ &\quad \left. \left. - \frac{C_{[0,1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)(\Lambda_3 - \Lambda_0)} \right) e^{-\Lambda_3(t-\tau)} \right. \\ &\quad \left. - \frac{C_{[0,1,2]}}{\Lambda_3 - \Lambda_2} e^{-\Lambda_2(t-\tau)} - \frac{C_{[0,1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(t-\tau)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)(\Lambda_3 - \Lambda_0)} e^{-\Lambda_0(t-\tau)} \right) \end{aligned}$$

We can similarly define:

$$C_{[0,1,2,3]} = \frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C_{[0,1,2]}}{\Lambda_3 - \Lambda_2} - \frac{C_{[0,1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)(\Lambda_3 - \Lambda_0)}$$

$$\begin{aligned} f_{[1,2,3]}(t, \tau) &= \lambda_2 \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - C_{[0,1,2,3]} e^{-\Lambda_3(t-\tau)} - \frac{C_{[0,1,2]}}{\Lambda_3 - \Lambda_2} e^{-\Lambda_2(t-\tau)} \right. \\ &\quad \left. - \frac{C_{[0,1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(t-\tau)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)(\Lambda_3 - \Lambda_0)} e^{-\Lambda_0(t-\tau)} \right) \end{aligned}$$

General Formula

With a few repetitions, we can see a pattern emerging.

$$C_{[0]} = G(\lambda_n - \Lambda_0)$$

$$C_{[0,1]} = \frac{G\lambda_n}{\Lambda_1} - \frac{C_{[0]}}{\Lambda_1 - \Lambda_0}$$

$$C_{[0,1,2]} = \frac{G\lambda_n}{\Lambda_1\Lambda_2} - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)}$$

$$C_{[0,1,2,3]} = \frac{G\lambda_n}{\Lambda_1\Lambda_2\Lambda_3} - \frac{C_{[0,1,2]}}{\Lambda_3 - \Lambda_2} - \frac{C_{[0,1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)(\Lambda_3 - \Lambda_0)}$$

$$C_{[0,\dots,m]} = \frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^{m-1} \left(\frac{C_{[0,\dots,i]}}{\prod_{j=i}^m (\Lambda_j - \Lambda_i)} \right)$$

Similarly, we can derive a pattern for the integrations:

$$f_{[1]}(t, \tau) = \frac{G\lambda_n}{\Lambda_1} - C_{[0,1]}e^{\Lambda_1(\tau-t)} - \frac{C_{[0]}}{\Lambda_1 - \Lambda_0}e^{\Lambda_0(\tau-t)}$$

$$f_{[1,2]}(t, \tau) = \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1\Lambda_2} - C_{[0,1,2]}e^{-\Lambda_2(t-\tau)} - \frac{C_{[0,1]}}{\Lambda_2 - \Lambda_1}e^{-\Lambda_1(t-\tau)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)}e^{-\Lambda_0(t-\tau)} \right)$$

$$f_{[1,2,3]}(t, \tau) = \lambda_2\lambda_1 \left(\frac{G\lambda_n}{\Lambda_1\Lambda_2\Lambda_3} - C_{[0,1,2,3]}e^{-\Lambda_3(t-\tau)} - \frac{C_{[0,1,2]}}{\Lambda_3 - \Lambda_2}e^{-\Lambda_2(t-\tau)} \right.$$

$$\left. - \frac{C_{[0,1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)}e^{-\Lambda_1(t-\tau)} - \frac{C_{[0]}}{(\Lambda_1 - \Lambda_0)(\Lambda_2 - \Lambda_0)(\Lambda_3 - \Lambda_0)}e^{-\Lambda_0(t-\tau)} \right)$$

$$f_{[1,\dots,m]}(t, \tau) = \left(\prod_{i=1}^{m-1} (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^m \left(\frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \right)$$

Proof by Induction

Suppose the above formulas hold true for $f_{[1,\dots,m]}$. We can then write:

$$\begin{aligned}
f_{[1,\dots,m+1]}(t, \tau) &= e^{-\Lambda_{m+1}t} \int (e^{\Lambda_{m+1}t}) \lambda_m f_{[1,\dots,m]}(t, \tau) \\
&= e^{-\Lambda_{m+1}t} \int_{\tau}^t (e^{\Lambda_{m+1}t}) \lambda_m \left(\prod_{i=1}^{m-1} (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^m \left(\frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \right) \\
&= \left(\prod_{i=1}^m (\lambda_i) \right) \left(e^{-\Lambda_{m+1}t} \int_{\tau}^t (e^{\Lambda_{m+1}t}) \frac{G\lambda_m}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^m e^{-\Lambda_{m+1}t} \int_{\tau}^t (e^{\Lambda_{m+1}t}) \left(\frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \right) \\
&= \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} (\Lambda_i)} (1 - e^{-\Lambda_{m+1}(t-\tau)}) \right. \\
&\quad \left. - \sum_{i=0}^m \frac{C_{[0,\dots,i]}}{(\Lambda_m + 1 - \Lambda_i) \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_{m+1}t} e^{\Lambda_i\tau} (e^{\Lambda_{m+1}t} e^{-\Lambda_i t} - e^{\Lambda_{m+1}\tau} e^{-\Lambda_i\tau}) \right) \\
&= \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} (\Lambda_i)} - \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} (\Lambda_i)} - \sum_{i=0}^m \frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} \right) e^{-\Lambda_{m+1}(t-\tau)} \right. \\
&\quad \left. - \sum_{i=0}^m \frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^{n+1} (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \\
&= \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} (\Lambda_i)} - C_{[0,\dots,m+1]} e^{-\Lambda_{m+1}(t-\tau)} - \sum_{i=0}^m \frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \\
&= \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} (\Lambda_i)} - \sum_{i=0}^{m+1} \frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right)
\end{aligned}$$

Therefore, we see that the formula also holds for $f_{[1,\dots,m+1]}$ and prove the formula is general for all integrations of this form.

General Framework

With the general formula for integration solved, we can return to the state functions for an arbitrary state N_m where I is the set of all states that enter state m by splicing or cleavage, P_{m,τ_0}

is the probability of entering state m by elongation at time τ_0 and P_{m,τ_1} is the probability of exiting state m by elongation at time τ_1

$$\frac{dN_m}{dt} = P_{m,\text{enter}} dr(t - \tau_0) - P_{m,\text{exit}} dr(t - \tau_1) - \Lambda_m N_m + \sum_{i \in I} \lambda_i N_i$$

Each state $i \in I$ can potentially take multiple paths from to reach state m . We can define a variable y containing:

1. The path v_y from entering/exiting a given state i by elongation to reaching state m where the first value in v_y is i and the last value in v_y is m .
2. The time τ_y when the first elongation event occurs.
3. The probability P_y of a given mRNA entering state i at time τ_y . If instead the mRNAs are leaving state i by elongation at time τ_y , P_y would give a negative number.

For example, there are two paths y_0 and y_1 entering state m by elongation:

1. $v_{y_0} = [m]; \tau_{y_0} = \tau_0; P_{y_0} = P_{m,\text{enter}}$
2. $v_{y_1} = [m]; \tau_{y_1} = \tau_1; P_{y_1} = -P_{m,\text{exit}}$

If another state a enters state m by splicing, the paths that enter and exit state a by elongation both need to be accounted for separately. However, paths that enter state a by splicing or cleavage should not be accounted for through state a but rather from their state of origin. This set of possible paths can be defined by variable $\mathbb{Y}$ resulting in the following general formula:

$$N_m = \sum_{y \in \mathbb{Y}} (P_y f_{v_y}(t, \tau_y))$$

Event without exits by splicing

Like we discussed above, suppose we have state N_1 which cannot be exited by splicing. It can only be entered by elongation at time τ_0 and exited by elongation at time τ_1 . The probability of entering N_1 by elongation can be defined as P_{enter} and exiting as P_{exit} , which are derived above in section 2. The state equation for N_1 would be:

$$\frac{dN_1}{dt} = P_{\text{enter}} dr(t - \tau_0) - P_{\text{exit}} dr(t - \tau_1)$$

This equation can be solved by direct integration over $dr(t)$.

$$N_1 = \int (P_{\text{enter}} dr(t - \tau_0) - P_{\text{exit}} dr(t - \tau_1)) dt$$

As $dr(t = 0) = 0$ and we know that $N_1(t = 0) = 0$, we have:

$$N_1 = P_{\text{enter}} \int_{\tau_0}^t dr(t - \tau_0) dt - P_{\text{exit}} \int_{\tau_1}^t dr(t - \tau_1) dt$$

We can define a function

$$F(t, \tau) = \int_{\tau}^t dr(t - \tau) dt$$

and solve as follows:

$$\begin{aligned} F_{\emptyset}(t, \tau) &= \int_{\tau}^t (G\lambda_n - G(\lambda_n - \Lambda_0)e^{-\Lambda_0(t-\tau)}) dt \\ &= G\lambda_n(t - \tau) - \frac{G(\lambda_n - \Lambda_0)}{\Lambda_0} (1 - e^{-\Lambda_0(t-\tau)}) \\ &= G\lambda_n(t - \tau) - \frac{C_{[0]}}{\Lambda_0} (1 - e^{-\Lambda_0(t-\tau)}) \end{aligned}$$

Suppose instead we were working with state m with possible entrances by splicing or cleavage from states $i \in I$ with the following differential equation.

$$\frac{dN_m}{dt} = P_{m,\text{enter}} dr(t - \tau_0) - P_{m,\text{exit}} dr(t - \tau_1) + \sum_{i \in I} \lambda_i N_i$$

We can substitute for N_i using the formula from above:

$$\frac{dN_m}{dt} = P_{m,\text{enter}} dr(t - \tau_0) - P_{m,\text{exit}} dr(t - \tau_1) + \sum_{i \in I} \lambda_i \left(\sum_{y \in \mathbb{Y}_i} P_y f_y(t, \tau_y) \right)$$

Each state $i \in I$, for each path $y \in \mathbb{Y}$, we can define a variable $z \in \mathbb{Z}$ containing:

1. The path $v_z = v_y$
2. The time $\tau_z = \tau_y$
3. The probability $P_z = P_y$
4. The rate $\lambda_z = \lambda_i$

This allows us to simplify the equation as follows:

$$\begin{aligned} N_m &= \int \left(P_{m,\text{enter}} dr(t - \tau_0) - P_{m,\text{exit}} dr(t - \tau_1) + \sum_{z \in \mathbb{Z}} (\lambda_z P_z f_{v_z}(t, \tau_z)) \right) dt \\ &= P_{m,\text{enter}} F_{\emptyset}(t, \tau_0) - P_{m,\text{exit}} F_{\emptyset}(t, \tau_1) + \sum_{z \in \mathbb{Z}} P_z \int_{\tau_z}^t \lambda_z f_{v_z}(t, \tau_z) dt \end{aligned}$$

To solve this equation, we can define a variable

$$F_{v_z}(t, \tau) = \int_{\tau_z}^t \lambda_z f_{v_z}(t, \tau_z) dt$$

We can substitute in the formula for $f_{v(t, \tau)}$ defined above:

$$\begin{aligned}
F_{[1,\dots,m]}(t, \tau) &= \int_{\tau}^t \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^m \left(\frac{C_{[0,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \right) \\
&= \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} (t - \tau) - \sum_{i=0}^m \left(\frac{C_{[0,\dots,i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i(t-\tau)}) \right) \right)
\end{aligned}$$

We can then substitute back in to the expression for N_m to get:

$$N_m = P_{m,\text{enter}} F_{\emptyset}(t, \tau_0) - P_{m,\text{exit}} F_{\emptyset}(t, \tau_1) + \sum_{z \in \mathbb{Z}} P_z F_{v_z}(t, \tau_z) dt$$

As the λ_z no longer appears separately in this formula, we can switch back to variable x .

If we accept that the path v_y can be the empty set, we can collapse the formula for states exited by elongation only as:

$$N_m = \sum_{y \in \mathbb{Y}} P_y F_{v_y}(t, \tau_y)$$

Where $\mathbb{Y}$ contains all possible paths for an mRNA to enter or exit state m , including the paths where an mRNA entered state m through elongation or exited state m through elongation in addition to those defined in the previous section.

Summary

For each state, we have a list of possible paths $y \in \mathbb{Y}$ containing:

1. The path v_y from entering/exiting a given state i by elongation to reaching state m where the first value in v_y is i and the last value is m .
2. The time τ_y when the elongation event occurs.
3. The probability P_y of a given mRNA entering/exiting state i at time τ_y . If instead the mRNAs are leaving state i by elongation at time τ_y , P_y would give a negative number. Formulas to calculate probabilities are derived above in section 2.

The total amount of mRNA in a particular state m at a particular time can be defined as:

$$N_m = \begin{cases} \sum_{y \in \mathbb{Y}} P_y F_{v_y}(t, \tau_y) & \Lambda_m = 0 \\ \sum_{y \in \mathbb{Y}} P_y f_{v_y}(t, \tau_y) & \Lambda_m > 0 \end{cases}$$

Where $f_{v_y}(t, \tau_y)$ and $F_{v_y}(t, \tau_y)$ are defined as:

$$f_{[1, \dots, m]}(t, \tau) = \left(\prod_{i=1}^{m-1} (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^m \left(\frac{C_{[0, \dots, i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \right)$$

$$F_{\emptyset}(t, \tau) = G\lambda_n(t - \tau) - \frac{C_{[0]}}{\Lambda_0} (1 - e^{-\Lambda_0(t-\tau)})$$

$$F_{[1, \dots, m]}(t, \tau) = \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} (t - \tau) - \sum_{i=0}^m \left(\frac{C_{[0, \dots, i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i(t-\tau)}) \right) \right)$$

Where Λ_0 is the DRB release rate, G is the number of transcription sites, λ_n is the initiation rate at unblocked transcription sites, and Λ_i is the sum of all splicing rates leaving state i . λ_{v_i} defined locally as the splicing/cleavage rate for moving between states v_i and v_{i+1} or, in other words, the rate of splicing used to leave state i in the particular path, and C is defined as follows:

$$C_{[0, \dots, m]} = \frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^{m-1} \left(\frac{C_{[0, \dots, i]}}{\prod_{j=i}^m ((\Lambda_j - \Lambda_i))} \right)$$

Steady State

Steady State Assumptions

As discussed above, we defined the number of transcripts at time t after DRB release as:

$$dr(t) = \begin{cases} 0 & t < 0 \\ G\lambda_n - G(\lambda_n - \Lambda_0)e^{-\Lambda_0 t} & t \geq 0 \end{cases}$$

where G is the number of transcription start sites, λ_n is the spawn rate, and $\Lambda_0 = \lambda_b$ represents the DRB release rate. In steady state, there is no DRB release required, so transcripts are immediately at maximum level after the gene activates at time $t = 0$.

$$dr(t) = \begin{cases} 0 & t < 0 \\ G\lambda_n & t > 0 \end{cases}$$

All remaining assumptions for modeling transcription elongation using a time delay and splicing/cleavage using exponential distributions remains the same. Therefore, derivations for probability distributions will also remain the same. While in the previous section, we needed to only derive equations for two cases, here there are 3 situations we need to deal with:

1. States with exits by splicing/cleavage AND elongation
2. States with exits only by elongation
3. States with exits only by splicing/cleavage

Exits by splicing/cleavage and elongation

Suppose we have a state N_1 with exits by splicing with a rate of Λ_1 . The amount of mRNAs in the state can be defined by the following differential equation:

$$\frac{dN_1}{dt} = P_{\text{enter}}dr(t - \tau_0) - P_{\text{exit}}dr(t - \tau_1) - \Lambda_1 N_1$$

As discussed above first order differential equations can be solved using:

$$y(x) = e^{-\int p(x)} \left(A + \int (e^{\int p(x)}) f(x) \right)$$

Substituting in for our equation, we have:

$$N_1 = e^{-\Lambda_1 t} \left(A + \int (e^{\Lambda_1 t}) (P_{\text{enter}}dr(t - \tau_0) - P_{\text{exit}}dr(t - \tau_1)) \right)$$

As $dr(t=0) = 0$, we have $N_1(t=0) = A$. Since there are no mRNAs present when transcription has not yet started, $A = 0$. Thus, we have:

$$N_1 = e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) (P_{\text{enter}} dr(t - \tau_0) - P_{\text{exit}} dr(t - \tau_1))$$

Separating out probabilities gives:

$$\begin{aligned} N_1 &= P_{\text{enter}} e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) (dr(t - \tau_0) - p_{[1]} dr(t - \tau_1)) \\ &= P_{\text{enter}} e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) (dr(t - \tau_0) - e^{-\Lambda_1(\tau_1 - \tau_0)} dr(t - \tau_1)) \end{aligned}$$

First step

We can now define a new variable $f'_{[1]}(t, \tau_0, \tau_1)$ as:

$$f'_{[1]}(t, \tau_0, \tau_1) = e^{-\Lambda_1 t} \int (e^{\Lambda_1 t}) (dr(t - \tau_0) - e^{-\Lambda_1(\tau_1 - \tau_0)} dr(t - \tau_1))$$

As $dr(t)$ is a piecewise function, we need to split it into three separate integrals.

1. $(t \leq \tau_0)$:

$$f'_{[1]}(t, \tau_0, \tau_1) = e^{-\Lambda_1 t} \int_0^{\tau_0} (e^{\Lambda_1 t} (0 - e^{-\Lambda_1(\tau_1 - \tau_0)} 0)) dt f'_{[1]}(t, \tau_0, \tau_1) = 0$$

2. $(\tau_0 < t \leq \tau_1)$:

$$\begin{aligned}
f'_{[1]}(t, \tau_0, \tau_1) &= e^{-\Lambda_1 t} \int_0^{\tau_0} (e^{\Lambda_1 t} (0 - e^{-\Lambda_1(\tau_1 - \tau_0)} 0)) dt + e^{-\Lambda_1 t} \int_{\tau_0}^t (e^{\Lambda_1 t} (G\lambda_n - e^{-\Lambda_1(\tau_1 - \tau_0)} 0)) dt f'_{[1]}(t, \tau_0, \tau_1) \\
&= 0 + G\lambda_n e^{-\Lambda_1 t} \int_{\tau_0}^t e^{\Lambda_1 t} dt f'_{[1]}(t, \tau_0, \tau_1) \\
&= \frac{G\lambda_n}{\Lambda_1} e^{-\Lambda_1 t} (e^{\Lambda_1 t} - e^{\Lambda_1 \tau_0}) f'_{[1]}(t, \tau_0, \tau_1) = \frac{G\lambda_n}{\Lambda_1} (1 - e^{\Lambda_1(t - \tau_0)}) f'_{[1]}(t, \tau_0, \tau_1) \\
&= \frac{G\lambda_n}{\Lambda_1} - \frac{G\lambda_n}{\Lambda_1} e^{\Lambda_1(t - \tau_0)}
\end{aligned}$$

3. ($t > \tau_1$) :

$$\begin{aligned}
f'_{[1]}(t, \tau_0, \tau_1) &= e^{-\Lambda_1 t} \int_0^{\tau_0} (e^{\Lambda_1 t} (0 - e^{-\Lambda_1(\tau_1 - \tau_0)} 0)) dt + e^{-\Lambda_1 t} \int_{\tau_0}^{\tau_1} (e^{\Lambda_1 t} (G\lambda_n - e^{-\Lambda_1(\tau_1 - \tau_0)} 0)) dt \\
&\quad + e^{-\Lambda_1 t} \int_{\tau_1}^t (e^{\Lambda_1 t}) (dr(t - \tau_0) - e^{-\Lambda_1(\tau_1 - \tau_0)} dr(t - \tau_1)) dt f'_{[1]}(t, \tau_0, \tau_1) \\
&= 0 + G\lambda_n e^{-\Lambda_1 t} \int_{\tau_0}^{\tau_1} e^{\Lambda_1 t} dt + G\lambda_n e^{-\Lambda_1 t} \int_{\tau_1}^t (e^{\Lambda_1 t}) dt - G\lambda_n e^{-\Lambda_1 t} \int_{\tau_1}^t (e^{\Lambda_1 t}) e^{-\Lambda_1(\tau_1 - \tau_0)} dt f'_{[1]}(t, \tau_0, \tau_1) \\
&= G\lambda_n e^{-\Lambda_1 t} \left(\int_{\tau_0}^t e^{\Lambda_1 t} dt - e^{-\Lambda_1(\tau_1 - \tau_0)} \int_{\tau_1}^t (e^{\Lambda_1 t}) dt \right) f'_{[1]}(t, \tau_0, \tau_1) \\
&= \frac{G\lambda_n}{\Lambda_1} e^{-\Lambda_1 t} ((e^{\Lambda_1 t} - e^{\Lambda_1 \tau_0}) - e^{-\Lambda_1(\tau_1 - \tau_0)} (e^{\Lambda_1 t} - e^{\Lambda_1 \tau_1})) f'_{[1]}(t, \tau_0, \tau_1) \\
&= \frac{G\lambda_n}{\Lambda_1} e^{-\Lambda_1 t} (e^{\Lambda_1 t} - e^{\Lambda_1 \tau_0} - e^{-\Lambda_1(\tau_1 - \tau_0)} e^{\Lambda_1 t} + e^{\Lambda_1 \tau_0}) f'_{[1]}(t, \tau_0, \tau_1) \\
&= \frac{G\lambda_n}{\Lambda_1} - \frac{G\lambda_n}{\Lambda_1} e^{-\Lambda_1(\tau_1 - \tau_0)}
\end{aligned}$$

We can define a constant:

$$C'_{[1]} = \frac{G\lambda_n}{\Lambda_1}$$

Thus we have:

$$f'_{[1]}(t, \tau_0, \tau_1) = \begin{cases} 0 & t \leq \tau_0 \\ \frac{G\lambda_n}{\Lambda_1} - C'_{[1]}e^{-\Lambda_1(t-\tau_1)} & \tau_0 < t \leq \tau_1 \\ \frac{G\lambda_n}{\Lambda_1} - C'_{[1]}e^{-\Lambda_1(\tau_1-\tau_0)} & t > \tau_1 \end{cases}$$

Since we particularly care about steady state, we can take $t \rightarrow \infty$ and substitute $\delta = \tau_1 - \tau_0$ to get:

$$N_1 = P_{\text{enter}} f'_{[1]}(\delta)$$

$$f'_{[1]}(\delta) = \frac{G\lambda_n}{\Lambda_1} - C'_{[1]}e^{-\Lambda_1\delta}$$

Something we can also note here is that the probabilities of moving between each state by elongation are constant with time t . This means that flux of mRNAs out of each state by elongation is constant with time when $t > \tau_1$. Therefore, the amount of mRNAs in each state must be constant once $t > \tau_1$, as if there were more mRNAs present in the state, then they would eventually have to leave by elongation. For future states, we will be just integrating until $t = \tau_1$. We could still manually solve by including the probability of leaving each state in the equations, as we did above, but we will take this shortcut to dramatically simplify the derivations.

Second Step

Suppose we have a state N_2 that can be entered from state N_1 by splicing with a rate of λ_1 and can be exited by elongation at time τ_1 or splicing with total splicing rate Λ_2 . Writing this differential equation, we would have:

$$\frac{dN_2}{dt} = P_{\text{enter},1}\lambda_1 N_1 - P_{[0,1]}dr(t - \tau_1) - \Lambda_2 N_2$$

Substituting in would give:

$$N_2 = P_{\text{enter},1} f'_{[1,2]}(t, \tau_0, \tau_1)$$

$$f'_{[1,2]}(t, \tau_0, \tau_1) = \lambda_1 e^{-\Lambda_2 t} \int (e^{\Lambda_2 t}) (f'_{[1]}(t, \tau_0, \tau_1) - P_{\text{stay in } 2} dr(t - \tau_1))$$

As $f'_{[1]}(t, \tau_0, \tau_1)$ is a piecewise function, we need to split it into three separate integrals

like above.

1. ($t \leq \tau_0$) :

$$f'_{[1,2]}(t, \tau_0, \tau_1) = \lambda_1 e^{-\Lambda_2 t} \int_0^{\tau_0} (e^{\Lambda_2 t} (0 - P_{\text{stay in } 2} 0)) dt f'_{[1,2]}(t, \tau_0, \tau_1) = 0$$

2. ($\tau_0 < t \leq \tau_1$) :

$$\begin{aligned} f'_{[1,2]}(t, \tau_0, \tau_1) &= \lambda_1 e^{-\Lambda_2 t} \int_0^{\tau_0} (e^{\Lambda_2 t} (0 - P_{\text{stay in } 2} 0)) dt \\ &+ e^{-\Lambda_2 t} \int_{\tau_0}^t \left(e^{\Lambda_2 t} \left(\frac{G\lambda_n}{\Lambda_1} - C'_{[1]} e^{-\Lambda_1(t-\tau_0)} - P_{\text{stay in } 2} 0 \right) \right) dt f'_{[1,2]}(t, \tau_0, \tau_1) \\ &= 0 + \lambda_1 e^{-\Lambda_2 t} \int_{\tau_0}^t \left(e^{\Lambda_2 t} \left(\frac{G\lambda_n}{\Lambda_1} - C'_{[1]} e^{-\Lambda_1(t-\tau_0)} \right) \right) dt f'_{[1,2]}(t, \tau_0, \tau_1) \\ &= \lambda_1 e^{-\Lambda_2 t} \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} (e^{\Lambda_2 t} - e^{\Lambda_2 \tau_0}) - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{\Lambda_1 \tau_0} (e^{(\Lambda_2 - \Lambda_1)t} - e^{(\Lambda_2 - \Lambda_1)\tau_0}) \right) f'_{[1,2]}(t, \tau_0, \tau_1) \\ &= \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} (e^{\Lambda_2 t} e^{-\Lambda_2 t} - e^{\Lambda_2 \tau_0} e^{-\Lambda_2 t}) - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} (e^{-\Lambda_2 t} e^{\Lambda_1 \tau_0} e^{(\Lambda_2 - \Lambda_1)t} - e^{-\Lambda_2 t} e^{\Lambda_1 \tau_0} e^{(\Lambda_2 - \Lambda_1)\tau_0}) \right) f'_{[1,2]}(t, \tau_0, \tau_1) \\ &= \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} (1 - e^{-\Lambda_2(t-\tau_0)}) - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} (e^{-\Lambda_1(t-\tau_0)} - e^{-\Lambda_2(t-\tau_0)}) \right) f'_{[1,2]}(t, \tau_0, \tau_1) \\ &= \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(t-\tau_0)} - \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} \right) e^{-\Lambda_2(t-\tau_0)} \right) \end{aligned}$$

3. ($t > \tau_1$) :

Using the shortcut we discussed above, $f'_{[1,2]}(t > \tau_1, \tau_0, \tau_1)$ is constant and $f'_{[1,2]}(t, \tau_0, \tau_1)$

is continuous, so we have:

$$f'_{[1,2]}(t, \tau_0, \tau_1) = \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1\Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(\tau_1 - \tau_0)} - \left(\frac{G\lambda_n}{\Lambda_1\Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} \right) e^{-\Lambda_2(\tau_1 - \tau_0)} \right)$$

We can define a constant $C'_{[1,2]}$ and substitute it in as follows:

$$C'_{[1,2]} = \frac{G\lambda_n}{\Lambda_1\Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)}$$

Thus we have:

$$f'_{[1,2]}(t, \tau_0, \tau_1) = \begin{cases} 0 & t \leq \tau_0 \\ \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1\Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(t - \tau_0)} - C'_{[1,2]} e^{-\Lambda_2(t - \tau_0)} \right) & \tau_0 < t \leq \tau_1 \\ \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1\Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(\tau_1 - \tau_0)} - C'_{[1,2]} e^{-\Lambda_2(\tau_1 - \tau_0)} \right) & t > \tau_1 \end{cases}$$

Since we particularly care about steady state, we can take $t \rightarrow \infty$ and substitute $\delta = \tau_1 -$

τ_0 to get:

$$f'_{[1,2]}(\delta) = \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1\Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1\delta} - C'_{[1,2]} e^{-\Lambda_2\delta} \right)$$

Third Step

Let's repeat this for a third state N_3 accessed from state N_2 by splicing with a rate of λ_2 . As

$f'_{[1,2]}(t, \tau_0, \tau_1)$ is a piecewise function, we need to split it into three separate integrals like above.

1. ($t \leq \tau_0$) :

$$f'_{[1,2,3]}(t, \tau_0, \tau_1) = \lambda_1 \lambda_2 e^{-\Lambda_3 t} \int_0^{\tau_0} (e^{\Lambda_3 t} (0 - P_{\text{stay in } 3} 0)) dt f'_{[1,2,3]}(t, \tau_0, \tau_1) = 0$$

2. ($\tau_0 < t \leq \tau_1$) :

$$\begin{aligned}
f'_{[1,2,3]}(t, \tau_0, \tau_1) &= \lambda_1 \lambda_2 \left(e^{-\Lambda_3 t} \int_0^{\tau_0} (e^{\Lambda_3 t} (0 - P_{\text{stay in } 3} 0)) dt \right. \\
&\quad \left. + e^{-\Lambda_3 t} \int_{\tau_0}^t \left(e^{\Lambda_3 t} \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(t-\tau_0)} - C'_{[1,2]} e^{-\Lambda_2(t-\tau_0)} \right) \right) dt \right) f'_{[1,2,3]}(t, \tau_0, \tau_1) \\
&= 0 + \lambda_1 \lambda_2 e^{-\Lambda_3 t} \int_{\tau_0}^t \left(e^{\Lambda_3 t} \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(t-\tau_0)} - C'_{[1,2]} e^{-\Lambda_2(t-\tau_0)} \right) \right) dt f'_{[1,2,3]}(t, \tau_0, \tau_1) \\
&= \lambda_1 \lambda_2 e^{-\Lambda_3 t} \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} (e^{\Lambda_3 t} - e^{\Lambda_3 \tau_0}) - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{\Lambda_1 \tau_0} (e^{(\Lambda_3 - \Lambda_1)t} - e^{(\Lambda_3 - \Lambda_1)\tau_0}) \right. \\
&\quad \left. - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{\Lambda_2 \tau_0} (e^{(\Lambda_3 - \Lambda_2)t} - e^{(\Lambda_3 - \Lambda_2)\tau_0}) \right) f'_{[1,2,3]}(t, \tau_0, \tau_1) \\
&= \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} (e^{-\Lambda_3 t} e^{\Lambda_3 t} - e^{-\Lambda_3 t} e^{\Lambda_3 \tau_0}) \right. \\
&\quad \left. - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} (e^{-\Lambda_3 t} e^{\Lambda_1 \tau_0} e^{(\Lambda_3 - \Lambda_1)t} - e^{-\Lambda_3 t} e^{\Lambda_1 \tau_0} e^{(\Lambda_3 - \Lambda_1)\tau_0}) \right. \\
&\quad \left. - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} (e^{-\Lambda_3 t} e^{\Lambda_2 \tau_0} e^{(\Lambda_3 - \Lambda_2)t} - e^{-\Lambda_3 t} e^{\Lambda_2 \tau_0} e^{(\Lambda_3 - \Lambda_2)\tau_0}) \right) f'_{[1,2,3]}(t, \tau_0, \tau_1) \\
&= \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} (1 - e^{-\Lambda_3(t-\tau_0)}) - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} (e^{-\Lambda_1(t-\tau_0)} - e^{-\Lambda_3(t-\tau_0)}) \right. \\
&\quad \left. - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} (e^{-\Lambda_2(t-\tau_0)} - e^{-\Lambda_3(t-\tau_0)}) \right) f'_{[1,2,3]}(t, \tau_0, \tau_1) \\
&= \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(t-\tau_0)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{-\Lambda_2(t-\tau_0)} \right. \\
&\quad \left. - \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} \right) e^{-\Lambda_3(t-\tau_0)} \right)
\end{aligned}$$

3. ($t > \tau_1$) :

Using the shortcut we discussed above, $f'_{[1,2,3]}(t > \tau_1, \tau_0, \tau_1)$ is constant and

$f'_{[1,2,3]}(t, \tau_0, \tau_1)$ is continuous, so we have:

$$f'_{[1,2,3]}(t, \tau_0, \tau_1) = \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(\tau_1 - \tau_0)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{-\Lambda_2(\tau_1 - \tau_0)} \right. \\ \left. - \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} \right) e^{-\Lambda_3(\tau_1 - \tau_0)} \right)$$

We can define a constant $C'_{[1,2,3]}$ and substitute it in as follows:

$$C'_{[1,2,3]} = \frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)}$$

Thus we have:

$$f'_{[1,2,3]}(t, \tau_0, \tau_1) = \begin{cases} 0 & t \leq \tau_0 \\ \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(t - \tau_0)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{-\Lambda_2(t - \tau_0)} - C'_{[1,2,3]} e^{-\Lambda_3(t - \tau_0)} \right) & \tau_0 < t \leq \tau_1 \\ \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(\tau_1 - \tau_0)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{-\Lambda_2(\tau_1 - \tau_0)} - C'_{[1,2,3]} e^{-\Lambda_3(\tau_1 - \tau_0)} \right) & t > \tau_1 \end{cases}$$

Since we particularly care about steady state, we can take $t \rightarrow \infty$ and substitute $\delta = \tau_1 -$

τ_0 to get:

$$f'_{[1,2,3]}(\delta) = \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1 \delta} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{-\Lambda_2 \delta} - C'_{[1,2,3]} e^{-\Lambda_3 \delta} \right)$$

General Formula

With a few repetitions, we see a pattern emerging:

$$\begin{aligned}
C'_{[1]} &= \frac{G\lambda_n}{\Lambda_1} C'_{[1,2]} = \frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} C'_{[1,2,3]} \\
&= \frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} C'_{[1,\dots,m]} \\
&= \frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=1}^{m-1} \left(\frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} \right)
\end{aligned}$$

Similarly, we can derive a pattern for the integrations:

$$f'_{[1]}(t, \tau_0, \tau_1) = \begin{cases} 0 & t \leq \tau_0 \\ \frac{G\lambda_n}{\Lambda_1} - C'_{[1]} e^{-\Lambda_1(t-\tau_1)} & \tau_0 < t \leq \tau_1 \\ \frac{G\lambda_n}{\Lambda_1} - C'_{[1]} e^{-\Lambda_1(\tau_1-\tau_0)} & t > \tau_1 \end{cases}$$

$$f'_{[1,2]}(t, \tau_0, \tau_1) = \begin{cases} 0 & t \leq \tau_0 \\ \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(t-\tau_0)} - C'_{[1,2]} e^{-\Lambda_2(t-\tau_0)} \right) & \tau_0 < t \leq \tau_1 \\ \lambda_1 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)} e^{-\Lambda_1(\tau_1-\tau_0)} - C'_{[1,2]} e^{-\Lambda_2(\tau_1-\tau_0)} \right) & t > \tau_1 \end{cases}$$

$$f'_{[1,2,3]}(t, \tau_0, \tau_1) = \begin{cases} 0 & t \leq \tau_0 \\ \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(t-\tau_0)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{-\Lambda_2(t-\tau_0)} - C'_{[1,2,3]} e^{-\Lambda_3(t-\tau_0)} \right) & \tau_0 < t \leq \tau_1 \\ \lambda_1 \lambda_2 \left(\frac{G\lambda_n}{\Lambda_1 \Lambda_2 \Lambda_3} - \frac{C'_{[1]}}{(\Lambda_2 - \Lambda_1)(\Lambda_3 - \Lambda_1)} e^{-\Lambda_1(\tau_1-\tau_0)} - \frac{C'_{[1,2]}}{(\Lambda_3 - \Lambda_2)} e^{-\Lambda_2(\tau_1-\tau_0)} - C'_{[1,2,3]} e^{-\Lambda_3(\tau_1-\tau_0)} \right) & t > \tau_1 \end{cases}$$

$$f'_{[1,\dots,m]}(t, \tau_0, \tau_1) = \begin{cases} 0 & t \leq \tau_0 \\ \left(\prod_{i=1}^{m-1} \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau_0)} \right) & \tau_0 < t \leq \tau_1 \\ \left(\prod_{i=1}^{m-1} \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(\tau_1-\tau_0)} \right) & t > \tau_1 \end{cases}$$

Or if we take $t \rightarrow \infty$ and substitute $\delta = \tau_1 - \tau_0$ we get:

$$f'_{[1,\dots,m]}(\delta) = \left(\prod_{i=1}^{m-1} \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i \delta} \right)$$

Proof by induction

To prove the general formula, we need to show

$$f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) = e^{-\Lambda_m t} \int_0^t \lambda_m e^{\Lambda_m t} f'_{[1,\dots,m]}(t, \tau_0, \tau_1) dt$$

This integration has three cases:

1. $(t \leq \tau_0)$:

$$\begin{aligned} f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) &= e^{-\Lambda_{m+1} t} \int_0^{\tau_0} \lambda_m e^{\Lambda_{m+1} t} f'_{[1,\dots,m]}(t, \tau_0, \tau_1) dt f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\ &= e^{-\Lambda_{m+1} t} \int_0^{\tau_0} \lambda_m e^{\Lambda_{m+1} t} 0 dt f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) = 0 \end{aligned}$$

2. $(\tau_0 < t \leq \tau_1)$:

$$\begin{aligned}
f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) &= e^{-\Lambda_{m+1}t} \int_{\tau_0}^t \lambda_m e^{\Lambda_{m+1}t} f'_{[1,\dots,m]}(t, \tau_0, \tau_1) dt f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\
&= e^{-\Lambda_{m+1}t} \int_{\tau_0}^t \lambda_m e^{\Lambda_{m+1}t} \left(\prod_{i=1}^{m-1} \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau_0)} \right) dt f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\
&= \left(\prod_{i=1}^m \lambda_i \right) e^{-\Lambda_{m+1}t} \int_{\tau_0}^t \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} e^{\Lambda_{m+1}t} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{(\Lambda_{m+1}-\Lambda_i)t} e^{\Lambda_i\tau_0} \right) dt f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\
&= \left(\prod_{i=1}^m \lambda_i \right) e^{-\Lambda_{m+1}t} \left(\frac{G\lambda_n}{\Lambda_{m+1} \prod_{i=1}^m \Lambda_i} (e^{\Lambda_{m+1}t} - e^{\Lambda_{m+1}\tau_0}) \right. \\
&\quad \left. - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{(\Lambda_{m+1} - \Lambda_i) \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (e^{(\Lambda_{m+1}-\Lambda_i)t} e^{\Lambda_i\tau_0} - e^{(\Lambda_{m+1}-\Lambda_i)\tau_0} e^{\Lambda_i\tau_0}) \right) f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\
&= \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} \Lambda_i} (1 - e^{-\Lambda_{m+1}(t-\tau_0)}) - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} (e^{-\Lambda_i(t-\tau_0)} - e^{-\Lambda_{m+1}(t-\tau_0)}) \right) f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\
&= \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} \Lambda_i} - \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} \right) e^{-\Lambda_{m+1}(t-\tau_0)} \right) \\
&\quad - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau_0)} f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\
&= \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} \Lambda_i} - C'_{[1,\dots,m+1]} e^{-\Lambda_{m+1}(t-\tau_0)} \right) - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau_0)} f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) \\
&= \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} \Lambda_i} - \sum_{i=1}^{m+1} \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau_0)} \right)
\end{aligned}$$

3. ($t > \tau_1$) :

Using the shortcut we discussed above, $f'_{[1,\dots,m+1]}(t > \tau_1, \tau_0, \tau_1)$ is constant and

$f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1)$ is continuous, so we have:

$$f'_{[1,\dots,m+1]}(t, \tau_0, \tau_1) = f'_{[1,\dots,m+1]}(\tau_1, \tau_0, \tau_1) = \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^{m+1} \Lambda_i} - \sum_{i=1}^{m+1} \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^{m+1} (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(\tau_1 - \tau_0)} \right)$$

All of these formulas match the general formula, concluding the proof by induction.

Exits by elongation only

Consider a state N_{m+1} that can be accessed by splicing from state N_m but can only be exited by elongation at time τ_1 . The differential equation describing this state would be:

$$\frac{dN_{m+1}}{dt} = \lambda_m N_m - P_{\text{exit}, m+1} dr(t - \tau_1)$$

$$N_{m+1} = F'_{[1,\dots,m]}(t, \tau_0, \tau_1) = \int_0^t (\lambda_m f'_{[1,\dots,m]}(t, \tau_0, \tau_1) - P_{\text{exit}, m+1} dr(t - \tau_1)) dt$$

$F'_{[1,\dots,m]}(t, \tau_0, \tau_1)$ similarly has 3 cases:

1. ($t \leq \tau_0$) :

$$F'_{[1,\dots,m]}(t, \tau_0, \tau_1) = \lambda_m \int_0^{\tau_0} (0 - P_{\text{exit}, m+1} 0) dt F'_{[1,\dots,m]}(t, \tau_0, \tau_1) = 0$$

2. ($\tau_0 < t \leq \tau_1$) :

$$\begin{aligned} F'_{[1,\dots,m]}(t, \tau_0, \tau_1) &= \lambda_m \int_{\tau_0}^t \left(\prod_{i=1}^{m-1} \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t - \tau_0)} \right) dt \\ &= \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} (t - \tau_0) - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i(t - \tau_0)}) \right) \end{aligned}$$

3. ($t > \tau_1$) :

Using the shortcut we discussed above, $F'_{[1,\dots,m]}(t > \tau_1, \tau_0, \tau_1)$ is constant and

$F'_{[1,\dots,m]}(t, \tau_0, \tau_1)$ is continuous, so we have:

$$F'_{[1,\dots,m]}(t, \tau_0, \tau_1) = F'_{[1,\dots,m]}(\tau_1, \tau_0, \tau_1) = \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} (\tau_1 - \tau_0) - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i(\tau_1 - \tau_0)}) \right)$$

We can take $t \rightarrow \infty$ and substitute $\delta = \tau_1 - \tau_0$ to get:

$$F'_{[1,\dots,m]}(\delta) = \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} \delta - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i \delta}) \right)$$

Exits by Splicing/cleavage only

To calculate the amount of mRNA in states with exits by splicing and cleavage only, we can use the same formulas derived above in 4.b, but take $\delta \rightarrow \infty$ as follows:

$$f'_{[1,\dots,m]}(\delta \rightarrow \infty) = \left(\prod_{i=1}^{m-1} \lambda_i \right) \frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i}$$

Summary

For each state, we have a list of possible paths $y' \in \mathbb{Y}'$ containing:

1. The path $v_{y'}$ from entering a given state i by elongation to reaching state m where the first value in $v_{y'}$ is i and the last value in $v_{y'}$ is m .
2. The time δ between entering and exiting the state by elongation
3. The probability $P_{y'}$ of a given mRNA entering state i . Formulas to calculate probabilities are derived above in section 2.

The total amount of mRNA in a particular state

$$m$$

at a particular time can be defined as:

$$N_m = \begin{cases} \sum_{y' \in \mathbb{Y}'} P_{y'} F'_{v_{y'}}(\delta_{y'}) & \Lambda_m = 0 \\ \sum_{y' \in \mathbb{Y}'} P_{y'} f'_{v_{y'}}(\delta_{y'}) & \Lambda_m > 0 \end{cases}$$

Where $f'_{v_{y'}}(\delta_{y'})$ and $F'_{v_{y'}}(\delta_{y'})$ are defined as:

$$f'_{[1,\dots,m]}(\delta) = \left(\prod_{i=1}^{m-1} \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i \delta} \right)$$

$$f'_{[1,\dots,m]}(\delta \rightarrow \infty) = \left(\prod_{i=1}^{m-1} \lambda_i \right) \frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i}$$

$$F'_{[1,\dots,m]}(\delta) = \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} \delta - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i \delta}) \right)$$

Where Λ_0 is the DRB release rate, G is the number of transcription sites, λ_n is the initiation rate at unblocked transcription sites, and Λ_i is the sum of all splicing rates leaving state i . λ_{v_i} defined locally as the splicing/cleavage rate for moving between states v_i and v_{i+1} or, in other words, the rate of splicing used to leave state i in the particular path, and $C_{v_{y'}}$ is defined as follows:

$$C'_{[1,\dots,m]} = \frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=1}^{m-1} \left(\frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} \right)$$

PRO-seq to time conversion

Assume PRO-seq density can be used as a relative measure of polymerase dwell-time within a particular gene. We assume that the total number of PRO-seq reads separating two positions in a given gene is proportional to the amount of time it takes to transcribe that region multiplied by a gene specific polymerase dwell time $\lambda_{p,0}$ defined as the time per PRO-seq read.

$$\lambda_{p,0} = \frac{\text{CPS}_0}{\bar{v} \sum_{i=0}^{\text{CPS}_0-1} R(i)}$$

where $R(i)$ is the number of PRO-seq reads at position i and CPS_0 is the position of the first CPS, and $\bar{v}$ is the gene average elongation rate. Before the first CPS, the amount of time taken to reach a particular position $\tau(n)$ can be defined as:

$$\tau(n) = \sum_{i=0}^n R(i) \lambda_{p,0}$$

Past the CPS, this coverage will be an underestimate due to the cleavage and termination of some portion of polymerases. Assuming that termination is fast relative to cleavage, the polymerase flux in an interval past the cleavage site CPS_0 is proportional to the input flux, the amount of time since transcribing the cleavage site t , and the cleavage rate of the CPS $\lambda_{c,0}$. The input flux is the rate at which uncleaved polymerases enter the interval, measured in PRO-seq signal per unit time; at the start of the first post-CPS interval it is $r(0) = \frac{1}{\lambda_{p,0}}$:

$$r(t) = \frac{1}{\lambda_{p,0}} e^{-\lambda_{c,0}t}$$

Therefore, the total amount of PRO-seq reads Q observed over a given interval can be defined as

$$Q = \int_0^{\Delta t} \frac{1}{\lambda_{p,0}} e^{-\lambda_{c,0}t} dt = \frac{1 - e^{-\lambda_{c,0}\Delta t}}{\lambda_{c,0}\lambda_{p,0}}$$

where Δt is the amount of time since transcribing the CPS_0 . Rearranging to solve for Δt gives

$$\Delta t = -\frac{1}{\lambda_{c,0}} \ln(1 - \lambda_{c,0}\lambda_{p,0}Q)$$

For additional cleavage sites, let $\Lambda_{C,j} = \sum_{a=0}^j \lambda_{c,a}$ be the cumulative cleavage rate for all CPSs transcribed by polymerases in interval j . The new baseline polymerase dwell time $\lambda_{p,j+1}$ at CPS_{j+1} is calculated from the total time Δt_j spent between transcribing CPS_j and CPS_{j+1} :

$$\lambda_{p,j+1} = \lambda_{p,j} e^{\Lambda_{C,j} \Delta t_j} = \frac{\lambda_{p,j}}{1 - \Lambda_{C,j} \lambda_{p,j} \sum_{i=\text{CPS}_j}^{\text{CPS}_{j+1}-1} R(i)}$$

These expressions require $\Lambda_{C,j} \lambda_{p,j} \sum_{i=\text{CPS}_j}^{\text{CPS}_{j+1}-1} R(i) < 1$; otherwise the assumed cleavage process exhausts the polymerase flux before the observed PRO-seq signal is accumulated.

More generally, for $\text{CPS}_j \leq i < \text{CPS}_{j+1}$, the elapsed time from cleavage site CPS_j to nucleotide i is

$$\Delta t(i) = -\frac{1}{\Lambda_{C,j}} \ln \left(1 - \Lambda_{C,j} \lambda_{p,j} \sum_{k=\text{CPS}_j}^i R(k) \right)$$