**Position-dependent effects of RNA-binding proteins in the context of co-transcriptional splicing**

Timur Horn, Raymond Gosliga, Mihaela Enculescu\* and Stefan Legewie\*

**Abstract**

**Introduction**

1) What is splicing? What is its relevance? Mechanism of spliceosomes? Types of splicing

2) Complexity of splicing is complex from a systems point of view:

-Splice sites need to be defined and then correctly joined

-Splice site recognition controlled by RNA-binding proteins (RBPs) that determine spliceosome recruitment

-Further complexity (or complexity reduction) from co-transcriptional dynamics of splicing

3) Splicing is a co-transcriptional process

-deposition of splice factors to splice sites by RNAP

-yeast: splicing instantaneously occurs after exit from RNAP

-human: splicing co-transcriptional but likely delayed

-RNAP elongation rate affects splicing outcomes

-different types of relationships

-one reason may be binding of inhibitory RBPs

4) Position-dependent effects of RBP – Switch from activator to inhibitor depending on binding position relative to splice sites

-genome-wide: RNA maps relate RBP position to splice site usage and/or exon inclusion

-m RNAP

-human: splicing co-transcriptional but likely delayed

**Results**

***Modeling of co-transcriptional splicing***

We investigated the splicing behavior of a minimal splicing system, in which an alternatively spliced cassette exon is flanked by introns and outer constitutive exons (Fig. 1A). Alternative splicing in this system involves inclusion or exclusion of the cassette exon.

To model co-transcriptional splicing, we implemented a system of ordinary differential equations (ODEs) and considered a simplified mechanism of splicing decision making, in which the transcript can commit to either skipping or inclusion (Fig. 1B). Moreover, we considered that the commitment reactions towards splicing only occur with a delay after the transcript has been initiated (Supplement). Immediately after transcript initiation, no splicing is possible, since none of the introns has yet been synthesized. Subsequently, splicing can occur, but only after transcription of the corresponding introns and exons is complete.

In our model, we neglected the initial, splicing-less phase in our model, and started our simulations only after the synthesis of the first intron has been completed (Fig. 1B). It was assumed that commitment to inclusion can occur throughout the lifetime of a simulated transcript, whereas commitment to skipping is possible only after a delay τ, once the cassette exon and its flanking introns have been transcribed. Thereby, we reflect that early splicing of the first intron excludes a later skipping reaction and commits splicing to inclusion even before the second intron has been synthesized.

To implement the time shift of skipping relative to inclusion, we did not explicitly model polymerase progression, but implemented a time delay for the skipping reaction (Fig. 1B and C). Specifically, the rate of commitment to skipping (kS) is initially zero and then increases in a step-like manner, whereas the commitment rate to inclusion (ki) is time-invariant.

The solution of this system with time-dependent reaction rates was computed by integrating the ODEs in a step-wise manner: until the time delay where skipping starts, the integration was performed considering only the commitment reaction to inclusion (skipping was assumed zero), and both reactions were taken into account afterwards. By calculating the integral of the corresponding time courses, we obtain the probability of commitment to skipping and inclusion. Under the assumption of a steady state and the same degradation rate for skipping and inclusion, this probability is proportional to the concentration of the isoforms (see Supplement for details).

***Slow transcript elongation favors exon inclusion in the basic model***

We analyzed the concentrations skipping and inclusion isoforms in our models of co-transciptional splicing and focused on the impact of varying transcript elongation rates. Specifically, we asked whether variations in RNA polymerase speed in the model affect the inclusion of the cassette exon. Such a dependency had been reported in the published experimental literature (REF).

To simulate altered transcript elongation rates, we varied the delay parameter τ and monitored the PSI metric (PSI = inclusion / (inclusion + skipping)) as a measure of inclusion frequency (Fig. 1C). In line with an earlier modeling study and experimental work on post-transcriptional splicing, we find that the inclusion frequency decreases with increasing polymerase elongation rate (Ref). At very low polymerase speed, inclusion is the only splicing outcome, i.e., all transcripts commit to inclusion before the transcript is elongated throughout the second intron where skipping can occur (Fig. 1C and D). Fast transcript elongation eliminates this kinetic advantage of inclusion, and the splicing outcome is a mixture of skipping and inclusion, which is directly proportional the ratio of the corresponding commitment rates (Supplement).

Using analytical calculations, we show that between these two extreme cases of pure co- and post-transcriptional splicing, respectively, the PSI-elongation curve decreases monotonically irrespective of the kinetic parameter values (Supplement). In quantitative terms, the PSI metric starts to decline as soon as the elongation speed outcompetes commitment to inclusion (i.e., as soon as ki≤τ). Hence, co-transcriptional splicing has an effect on splicing outcomes whenever polymerase elongation is slow enough to prevent early commitment to any of the isoforms (skipping), thereby establishing a kinetic advantage of other isoforms (inclusion).

We quantitatively confirmed these simulation results using an alternative implementation of co-transcriptional splicing dynamics: in the multistep model variant, we described the progression of RNA polymerase using multiple elongation states, and in similarity to the delay model assumed that skipping only occurs after several elongation steps have been completed (Fig. 1C). Also in this model, the PSI metric was used as a measure of splicing outcome and alterations in transcript elongation were simulated by changing the progression parameter between states (kelong). This parameter was normalized by the total number of elongation states to calculate an overall polymerase speed (velong = kelong / number of states).

For large number of elongation steps, the simulation results of the multistep model quantitatively agreed with the delay model, whereas the agreement was qualitative - but not quantitative - for a lesser number of steps (Fig. 1C). This is due to the fact that the time delay in skipping is not sufficiently sharp in a multistep model with few intermediates (≤20), rendering the Psi-velong curve slightly less switch-like compared to the delay model.

***Non-canonical splicing responses to elongation encoded by position of RBP binding***

Genome-wide measurements of RNA polymerase mutant cells revealed that the transcript elongation rate affects splicing outcomes in a gene-specific manner (REF): a large number of genes shows the canonical dependency on the polymerase elongation rate as in the simulations above, i.e., fast elongation promotes skipping. However, in the experiments, hundreds of genes show non-canonical behavior, i.e., an inversed response, where faster RNA polymerization shifts splicing towards inclusion. Moreover, two additional gene classes exist, in which the relationship between transcript elongation and PSI is non-monotonous, resulting in the bell- or U-shaped curves in the experimental velong-PSI-diagram (Fig. 2A).

Recent experimental work suggested that the inverse non-canonical splicing response to alterations in the transcript elongation speed is established by an inhibitory RNA-binding protein that prevents exon inclusion when deposited by RNA polymerase during elongation (REF). We extended our model of co-transcriptional splicing to mimic this type of regulation: we considered an inhibitory reaction which prevents commitment to inclusion by shifting the mRNA into an inhibited state (mRNAinh) between the delay times τinh,1 and τinh,2 (Fig. 2B). In the inhibited state mRNAinh, skipping can still occur, but only after the delay time τ. In similarity to the basic model, τ reflects the time it takes for polymerase to complete the synthesis of the second intron (see above).

In essence, the inhibitor introduces early commitment to skipping, while preventing inclusion. Importantly, we neglected inhibitor binding to the transcript independent of RNA polymerase. Accordingly, no transition to the inhibited transcript state mRNAinh is possible before τinh,1 or after τinh,2. Thus, the period τinh,1τinh,2 reflects a window-of-opportunity for inhibition of inclusion. Moreover, it is assumed that the inhibitor binds irreversibly and has a long residence time on the transcript at least at the time scale of splicing commitment. Thus, inhibitor dissociation does not occur in our model.

Interestingly, this extended model of co-transcriptional splicing exhibits all experimentally observed types of velong-PSI-diagram including monotonically increasing or decreasing, as well as bell- and U-shaped curves (Fig. 2C). Which type of curve is retrieved in the model critically depends on the position of the time window τinh,1τinh,2 during the transcript elongation cycle: If the inhibitor binds at early stages of elongation - concomitantly or before commitment to inclusion is possible –inclusion increases with elongation speed and eventually a bell-shaped curve may be observed (Fig. 2C). In contrast, later inhibition after the possibility of commitment to inclusion leads to a decreasing PSI-velong diagram and potentially a U-shaped curve (Fig. 2C). Hence, the position RBP binding in the transcript has strong effects on splicing outcomes and the qualitative behavior of the model.

To intuitively understand these distinct types of behaviors, it is important to consider how splicing commitment rates change during elongation of the transcript. For instance, for the early inhibitor binding case, three phases can be distinguished (Fig. 2D): (i) early in the elongation cycle, commitment to inclusion and inhibitor-mediated commitment to skipping is possible; (ii) subsequently, only commitment to inclusion occurs (since the inhibitor has not previously been deposited by RNA polymerase); (iii) finally, commitment to both inclusion and skipping is possible once the second intron has been completely synthesized.

In Fig. 2D, it is schematically depicted how this translates into bell-shaped splicing response for changing transcript elongation rates. For very slow elongation, early commitment to splicing outcompetes transcript elongation in an uncommitted state, resulting in an early-determined mixture of skipping and inclusion (Fig. 2D, i). At medium elongation rates, commitment on average occurs later in the elongation cycle when inclusion is the only possible splicing outcome (Fig. 2D, ii). Finally, very fast elongation prevents commitment during ongoing transcription, and results in essentially post-transcriptional splicing commitment towards skipping and inclusion (Fig. 2D, iii). As a result, the PSI-velong diagram exhibits a bell shape, as shown in Fig. 2C alongside with the corresponding splicing commitment rates.

The precise behavior depends on the exact choice of parameter values. For instance, the PSI-velong curve may also monotonically increase in the early inhibition scenario as shown in Fig. 2C. In this simulation, the late decision between skipping and inclusion is still dominated by the inclusion reaction. In essence, this system therefore experiences only two commitment regimes during elongation, one early dominated by inhibitor-mediated skipping and one late dominated by inclusion.

In general, splicing outcomes in the PSI-velong diagrams are confined by the ratios of skipping and inclusion commitment rates observed in the three commitment regimes along the elongating transcript (Supplement). PSI values predicted by these ratios are shown as horizontal dash-dotted lines i in Fig. 2C and correspond well to the plateaus in the PSI-velong curves. Whether the maximum of the bell-shaped PSI-velong relationship actually reaches the value predicted by the commitment ratio of the middle regime depends on: (i) the dimension of the middle commitment regime, i.e., the length of the corresponding intron and/or exon sequences (Supplement); (ii) the difference of commitment rates between the three regimes. Specifically, a qualitative change in the PSI-velong curve in x-direction can be expected as soon as the elongation outcompetes one or multiple commitment rates, as then the splicing decision will be shifted from one regime to another. To illustrate this point, we show these outcompetition points as vertical lines in the PSI-velong diagram….

Taken together, parameter values shape the qualitative and quantitative splicing behavior for varying transcript elongation. However, certain types of behaviors cannot be achieved by tuning commitment rates, but require a structural change in the model topology. Specifically, a U-shaped PSI-velong diagram is not accessible if the inhibitor acts early during elongation, but requires later inhibitor action (Supplement and Fig. 2C). This ensures that predominant commitment to inclusion may occur early and late during elongation, i.e., at low and high polymerase velocities in the PSI-velong diagram. In the middle of the elongating transcript and thus, inhibitor-mediated commitment to skipping dominates, thus lowering the PSI metric at intermediate elongation rates and resulting in a U-shape.

In conclusion, qualitatively distinct relationships between transcript elongation and splicing outcomes may be realized if a single inhibitory controls splicing commitment at different position in the elongating transcript.

***Modeling of local RBP effects on spliceosome recruitment to splice sites***

So far we assumed that binding of an inhibitory RBP suppresses recognition of the alternative exon, thereby preventing inclusion and promoting commitment to skipping. In order to further verify our conclusions concerning the relationship between transcript elongation and splicing, we implemented a more detailed model which more accurately describes mechanisms of splicing decision making and RBP-dependent regulation (Fig. 3A).

Splicing decision making was modelled based on the recognition of splice sites by U1 and U2 subunits of the spliceosome. Spliceosome binding was assumed to occur by a so-called exon definition mechanism, which implies highly cooperative subunit binding on both sides of an exon, before splicing across introns occurs. To implement exon definition, we considered only complete (but not partial) definition

that considered that the inhibitory RBP may block spliceosome recruitment.

To describe local RBP action, we assumed that the inhibitor only affects recognition of local splice sites, thus modulates the exon definition parameter kx → kx\_inh(Fig. 3B)

is likely to affect splicing by preventing the recruitment of the spliceosome to nearby splice sites. Hence, depending on its binding position, the RBP may affect the commitment to inclusion, skipping or both. We sought to compare our simulation results using a mechanistic model, in which we describe the recognition of functional splicing units by the spliceosome and its RBP-dependent modulation.

***Position-dependence of RBP effects on PSI***

***Noise in alternative splicing follows a binomial distribution***

Cellular RNAs are frequently expressed at low levels, often summing up to a total concentration of only a few molecules per cell (REF). At such low concentrations, biochemical reaction do not occur deterministically, but involve a probabilistic component. Thus, alternative splicing may be a stochastic process with uncertainty in the exon inclusion frequency, as opposed to a deterministic system where the fraction of the inclusion isoform is predictable and completely determined by the kinetic rate constants (REF – Dan Larsson).

To quantify uncertainties in splicing outcomes, we performed stochastic simulations using our co-transcriptional splicing models and compared all previously introduced model variants. For stochastic simulations, delay times, commitment rates and inhibitor/spliceosome binding rates to mRNA were sampled from exponential distributions to determine the order of reaction steps in a manner reminiscent of the Gillespie algorithm (Supplement). This way, we take into account stochastic variation in transcript elongation (delay times) and biochemical steps of splicing decision making. The stochastic simulation was repeated xxx times for each parameter combination and the PSI metric was calculated to obtain splicing distributions across xxx cells (Fig. 5A).

Splicing uncertainty was characterized by relating the standard deviation of the PSI metric as a function of its mean value (Fig. 5B). As discussed previously (REF), this std-mean relationship exhibits a bell-shape with zero noise at PSI close to one or zero. At intermediate PSI values, the noise becomes peaks, the maximal value of the std being determined by the mRNA expression level, i.e., the ratio of mRNA synthesis and degradation. At very low molecule numbers, the splicing outcome is very unstable, whereas it approaches the deterministic solution for a total expression of >xxx molecules per cell (Fig. 5B).

Interestingly, the noise curves of all model variants are congruent after a correction for the (average) total number of molecules. This implies that the dynamics of co-transcriptional splicing and a modulation by an inhibitory RBP have little impact on splicing fluctuations, but the system always behaves like a simple binary decision between two alternative isoforms. In fact, all models follow the expectation derived from a binomial distribution, in which two categorical outcomes are drawn from a random distribution. Thus, even complex alternative splicing regulation and stochastic heterogeneity in transcript elongation (pausing) seems to add little intrinsic stochastic noise to gene expression outcomes. This explains why a large part of cell-to-cell variability in two splicing decisions could be explained by a purely binomial model (Waks et al., MSB).

To investigate the role of extrinsic fluctuations, we considered that RBP expression and the transcript elongation rate may vary between cells, and therefore sampled the corresponding model parameters from a log-normal distribution with a CV = 0.3 before performing individual stochastic splicing simulations…. (Fig. 5C)