Topology of the exon definition model

Splicing reactions are catalyzed by the spliceosome, which recognizes splice site sequences and forms a catalytically active higher-order complex across introns. To model this process, we considered that human spliceosomes frequently operate via a so-called exon definition mechanism, in which the pioneering spliceosome subunits U1 and U2 cooperatively bind to splice sites flanking an exon before the final cross-intron complex is formed during spliceosome splicing outcomes as described further below.

In the model, the pre-mRNA (species P0_0_0) is synthesized at a constant rate s. In the pre-mRNA, none of the three exons is bound ('defined') by the spliceosome (white boxes), therefore this state is called P0_0_0 in the model (**Figure S7**) with the notation '_' indicating the presence of an intron. The spliceosome may bind reversibly to each of the exons with on-rates k1, k2 and k3. For instance, from P000 we can obtain P1_0_0, P0_1_0 and P0_0_1 through binding to the first, second and third exon, respectively. Subsequently, further binding is possible, e.g., P1_0_1 can be generated from P1_0_0 with the rate constant k3. In total, there are eight spliceosomal binding states, including the fully bound state (P1_1_1), in which all exons are defined. All binding reactions are assumed to be reversible, i.e., k4, k5, and k6 are the dissociation rate constants and the reverse of k1, k2, and k3 respectively. For example, in P1_1_0 spliceosome dissociation from exon 1 with the rate constant k4 yields the species P0_1_0.

Depending on the exon definition states, splicing decisions are made, and irreversible splicing reactions are possible. For a splicing event to occur, we consider that both exons flanking a future splice junction must be defined. For instance, skipping of exon 2 is possible from P1_0_1 and occurs with the rate constant i12. Likewise, splicing of the first intron occurs from the species P1_1_0 and P1_1_1 (rate constant i1), and splicing of the second intron from P0_1_1 and P1_1_1 (rate constant i2). The inclusion isoform is generated in two steps, i.e., from the subsequent removal of intron 1 and intron 2 in random order: From the binding state P1_1_1, intron splicing generates two alternative intermediates in which either of the introns is already spliced (P1_11 or P11_1) and the retained intron can be further spliced in a subsequent reaction. Splicing of the partially defined species P1_1_0 and P0_1_1 yields the species P11_0 and P0_11; in these, the spliceosome can further reversibly bind exons 3 and 1, respectively, and undergo a second splicing reaction towards inclusion.

In the model, all terminal splice products are subject to degradation (kincl: degradation rate constant of inclusion, kskip: skipping, kdr1: first intron retention, kdr2: second intron retention). The degradation rate constant of the full intron retention isoform is the sum of kdr1 and kdr2, reflecting that either intron may contain a destabilizing premature stop codon. Model species that can be bound or spliced further (P0_0_0, P1_0_0, P0_1_0, P0_0_1, P1_1_0, P1_0_1, P0_1_1, P1_1_1,P0_11,P1_11,P11_0,P11_1) are not subject to degradation, but they may be exported from the nucleus with the rate constant kret. This reaction reflects that there is a limited time window for splicing to occur, the intermediates otherwise being terminally frozen in the corresponding intron retention state.

The ordinary differential equations of the model are given **Table S7**.