Topology of the intron definition model

Since a subset of human genes is spliced by an intron definition mechanism (PMID: 35182478), we have also considered this scenario in a modified version of our splicing model. In contrast to the exon definition model, the 5‘ and 3‘ splice sites of an exon can be bound independently of one another in the intron definition model. Furthermore, splicing of an intron is possible as soon a both splice sites flanking this intron are defined. Hence, definition of two splice sites is sufficient for splicing, whereas in the exon definition model four splice sites need to be defined (3‘ and 5‘ splice sites of the flanking exons)

For the intron definition model, we used a binding state notation that is similar to the exon definition model. For instance, we consistently assigned the state where no spliceosome component is bound as P0_0_0(**Figure S7**). For spliceosome binding to exons 1 and 3, we again considered a single binding reaction, as only the splice sites flanking the considered introns matter for splicing. Hence, a transition from '0' to '1' in the first position (e.g., P0_0_0 to P1_0_0) represents spliceosome binding state downstream of exon 1 (5' of the first intron), while '0' to '1' in the third position indicates binding upstream of exon 3 (3' of the second intron). For exon 2, we considered to separate splice site binding events. We used '0' for no binding, 'a' for upstream binding (e.g., P0_a_0), 'b' for downstream binding (e.g., P0_b_0),, and '1' for both U2 and U1 being simultaneously bound (e.g., P0_1_0). Similar to exon definition, the presence or absence of '_' indicates whether the intron is removed or not. We have the same parameter notation k1/k4 and k3/k6 to describe binding/unbinding to exons 1 and 3, respectively. New parameters k2a/k5a and k2b/k5b were introduced to represent spliceosome binding/unbinding around exon 2, k2a/k5a denoteding the upstream site and k2b/k5b denoteding the downstream site. There are a total of 16 spliceosomal binding states in the intron defintion model, with the following additional states that had not been considered in the exon definition model: P0_a_0, P0_b_0, P1_a_0, P1_b_0, P0_a_1, P0_b_1, P1_a_1, P1_b_1.

Once both splice sites flanking a future splice junction are defined, splicing decisions, implemented as irreversible splicing reactions in the model,can occur. Skipping of exon 2 is possible from P1_0_1 and occurs with the rate i12. Splicing of the first intron occurs from the species P1_a_0, P1_1_0,P1_a_1 and P1_1_1 (rate i1), and splicing of the second intron occurs from P0_b_1, P0_1_1,P1_b_1, and P1_1_1 (rate i2). The inclusion isoform is generated in two steps: first, intron 1 or 2 is spliced from P1_1_1, generating P1_11 or P11_1, respectively. Second, the retained intron can be further spliced in a subsequent reaction. Splicing of the partially defined species P1_a_0,P1_1_0.P0_b_1 and P0_1_1 yields the species,P1a_0, P11_0,P0_b1 and P0_11, respectively. To these, the spliceosome can bind further reversibly with the rate constants k1,k2a,k2 and k3 (depending on the site of binding) and once the species P1_11 or P11_1 are formed, a second splicing reaction towards inclusion can occur.

All terminal splice products are subject to degradation, with the the same assumptions and degradation rate constant notation as in the exon definition model. Furthermore, model species that can be bound or spliced further (P0_0_0, P1_0_0, P0_a_0, P0_b_0, P0_1_0, P0_0_1, P1_1_0, P1_a_0, P1_b_0, P1_0_1, P0_a_1, P0 _b_1, P0_1_1, P1_a_1, P1_b_1, P1_1_1) may again be exported from the nucleus with a rate constant of kret.

The ordinary differential equations of the model are given **Table S8**