Supplemental Material for:

Splicing of designer exons informs a biophysical model for exon definition

Mauricio A. Arias, Ashira L. Lubkin and Lawrence A. Chasin, 2014

Contents

Supplemental Box 1: Model assumptions and definitions

General equations were obtained based on the following conditions and assumptions:

1. The cell or system studied is at steady state.
2. We consider all the pre-mRNA molecules of interest that have started transcription within the same negligibly small time interval as "tagged"; it is this "pulse-tagged" cohort of molecules whose fate will be analyzed.
3. For simplicity, we assume that the transcription rate is the same for all of these molecules.
4. We define time zero as the time at which the exon of interest has been synthesized and made available for splicing.
5. Each tagged molecule contains at least one internal exon, one of which is the exon of interest. We will assess inclusion and skipping of this exon in all the tagged molecules.
6. There is a complex that forms on the exon of interest that is an obligatory intermediate for exon inclusion. We consider this complex to be an exon definition complex. At any given point in time we define P as the number of tagged uncommitted molecules with this complex and I as the number that are committed to inclusion, taken to be splicing to the committed exon that lies immediately upstream. Additionally, we define S as the number of molecules that are committed to skipping, taken to be splicing of the downstream exon directly to the upstream exon, effectively removing the exon of interest as being part of a long intron between these two.
7. The exons flanking the exon of interest are constitutive. In particular, we assume that by the time the exon of interest is made available for splicing, the upstream exon is already committed.
8. Exon definition can commit an internal exon to inclusion whether or not the downstream exon has been synthesized.
9. We assume first order kinetics for all transitions between states. In particular, $dI/dt = \rho_I * P$, where ρ_I is the rate at which these molecules commit to the included pathway.
10. All tagged molecules will follow one of two pathways: inclusion or skipping; we will consider only the decision between these 2 possibilities.

The Model: Approach

The general assumptions made to analyze splicing of an internal exon in a three-exon system are shown in Supplemental Box 1. The resulting state diagram is shown in Fig. 7 and described in the main text of this article. Importantly, two time periods were defined with respect to a time τ: the period after the exon of interest, but before the downstream exon, is made available for splicing ($t \leq \tau$) and the period after the downstream exon is made available for splicing ($t > \tau$).

Initially, a solution was sought relating psi to the various transition rates as shown in Fig. 7. In the following two sections an overview of the derivation of the general solution is presented; this solution, which was based on the assumptions and definitions in Supplemental Box 1, was then simplified by introducing a non-trivial simplifying assumption. The corresponding simplified solution was obtained for both the pre- and post- τ periods. The final equation defines the fraction of molecules that skip the exon of interest and it is used as the foundation of a statistical mechanical approach to splicing. Psi can be easily derived from this equation; however, the form of the equation for the skipped molecules was chosen to present the analysis because it is simpler to understand and manipulate (see main text).

The subsequent two sections present statistical mechanical proposals to model the transition rates a and d in Fig. 7, describing mechanisms for the formation and dissociation of the exon definition complex. A brief description of how the results of these four sections are combined into a final equation is then presented. For a version of this derivation that includes all algebraic steps and explanations of the approximations used, please see the Appendix.

Solving the system of differential equations for $t \leq \tau$

At time point t, let L(t) be the number of uncomplexed (naked) pre-mRNA molecules, P(t) be the number of molecules in an exon definition complex, and I(t) be the number of molecules committed to inclusion. The equations for L = L(t), P = P(t) and I = I(t) according to the state diagram depicted in Fig. 7A are

S1. $dL/dt = d\,P - a\,L$

S2. $dP/dt = a\,L - (d+\rho_I)\,P$

S3. $dI/dt = \rho_I\,P$

where a and d are association and dissociation constants, respectively, and ρ_I is the rate at which complexed molecules commit to the included pathway.

Defining F as the number of uncommitted molecules, F = L + P, we solved this simple system of differential equations for F(t) using Laplace transformations and partial fractions. The result is

S4. $F(t) = [(r_2\,F_0 - \rho_I\,P_0)\,e^{r1\,t} - (r_1\,F_0 - \rho_I\,P_0)\,e^{r2\,t}] \,/\, (\,r_2 - r_1\,)$

where F_0 and P_0 represent initial values for F(t) and P(t) respectively and r_1 and r_2 ($r_1 \geq r_2$) are the roots of the quadratic equation used to obtain the inverse Laplace transformation

S5. $x^2 + (d+a+\rho_I)\, x + \rho_I\, a = 0$

Solving the system of differential equations for P(t) yields

S6. $P(t) = [-(r_1\, P_0 + r\, F_0)\, e^{r1\, t} + (r_2\, P_0 + r\, F_0)\, e^{r2\, t}] / (r_2 - r_1)$

Evaluating these equations at time τ and noting that $I(t) = L_0 - F(t)$ where L_0 is the initial value for L(t), we obtain the following general solutions for the pre-τ period

S7. $F_\tau = [(r_2\, F_0 - \rho_I\, P_0)\, e^{r1\, \tau} - (r_1\, F_0 - \rho_I\, P_0)\, e^{r2\, \tau}] / (r_2 - r_1)$

S8. $P_\tau = [(a\, F_0 + r_2\, P_0)\, e^{r2\, \tau} - (a\, F_0 + r_1\, P_0)\, e^{r1\, \tau}] / (r_2 - r_1)$

S9. $I_\tau = L_0 - F_\tau$

where the notation X_τ represents X(t) at time τ.

At the beginning of the observation period no complexes have formed, so $P_0 = 0$ and $F_0 = L_0$. If we assume that the assembly or the dissociation of the complex occurs much faster than commitment, so that $d+a >> \rho_I$, which leads to $|r_2| >> |r_1|$. We then obtain

S10. $r_2 \approx -(d+a)$,

S11. $r_1 \approx -\rho_I\, a / (d+a)$ and

S12. $r_2 - r_1 \approx -(d+a)$.

Defining p_I as

S13. $p_I = \rho_I / (1+d/a)$

we get

S14. $F_\tau \approx L_0\, e^{-pI\, \tau}$

Therefore the system can now be approximated by the state diagram shown in Fig 7B.

Solving the system of differential equations for t > τ

The presence of the downstream exon defines several new states depending on the formation or dissociation of the exon definition complex on this exon (see Fig. 7C). Importantly a new end state is defined, S. This state represents the molecules that are committed to splicing the downstream exon directly to the upstream exon, effectively skipping the exon of interest. To minimize the complexity of notation below, we define a new reference time t' that sets time τ to zero: $t' = t - \tau$. From the state diagram shown in Fig. 7C, the following equations are obtained for $t' > 0$

S15. $dL/dt' = d\, P + d'\, b - (a+a')\, L$

S16. $dP/dt' = a\, L + d'\, B - (d+a'+\rho_I)\, P$

S17. $db/dt' = a'\, L + d\, B - (d'+a+\rho_S)\, b$

S18. $dB/dt' = a'\, P + a\, b - (d+d'+\rho_I+\rho_S)\, P$

S19. $dI/dt' = \rho_I\, (P+B)$

S20. $dS/dt' = \rho_S\, (b+B)$

where S represents molecules committed to skipping (i.e., the joining of exon 1 to exon 3), ρ_S is the rate at which complexed molecules commit to the skipped pathway, B represents molecules with both exons in EDCs, b represents molecules with a downstream exon in an EDC but with the exon of interest not in an EDC, and a’ and d’ are the association and dissociation constants, respectively, for the formation of b.

Although we are most interested in the probability of exon inclusion, I, it is easier to calculate exon skipping, S, and its final expression actually provides more insight into the roles of the different parameters. I becomes simply all the tagged molecules not included in S. Therefore we will focus on an expression for S(t’) as t’ → ∞, S_∞. The value of S(t') for $t' \leq 0$ equals 0 if commitment to skipping requires the presence a downstream exon. Similarly, no tagged molecules contain an EDC on exon 3 at t' = 0, since the downstream exon has not yet been synthesized. Using Laplace transforms and the final value theorem, an expression can be obtained for S_∞

S21. $S_\infty = a'\rho_S\,[\beta F_\tau - \gamma\rho_I P_\tau]\,/\,\{\alpha\,[(d'+a')\,a\rho_I + (d+a)\,a'\rho_S] + (a+a')\,(a\rho_I^2+\gamma\rho_I\rho_S+a'\rho_S^2) +$

$(d'a\rho_I+da'\rho_S)\,(\rho_I+\rho_S)\}$

where $\alpha = d+d'+a+a'$, $\beta = \alpha\,(d+a) + (\alpha+d)\,\rho_I + (d+a+a')\,\rho_S + (\rho_I+\rho_S)\,\rho_I$ and $\gamma = \alpha+\rho_I+\rho_S$.

This, along with the equations for F_τ and P_τ (equations S7 and S8), provide the general solution for S_∞. However a more useful expression can be obtained if we make the simplifying assumption that assembly or dissociation of the complexes on both exons occurs much faster than commitment for either pathway: i.e., $d+a \gg \rho_I$, $d+a \gg \rho_S$, $d'+a' \gg \rho_I$ and $d'+a' \gg \rho_S$. This assumption is essentially the same simplifying assumption made for the pre-τ period. Using p_I as defined previously and defining p_S analogously as

S22. $p_S = \rho_S / (1 + d' / a')$

yields

S23. $S_\infty \approx L_0\, e^{-p_I \tau}\, p_S/(p_S+p_I)$

This situation can be summarized with the state diagram shown in Fig. 7D for t > τ, with the initial condition $L_\tau = L_0\, e^{-p_I \tau}$. To model the system at all times requires only three constants, namely τ, p_I and p_S (see Fig. 7B and 7D).

Collision of tethered exon ends

To model the rate of formation of an exon definition complex, a, a physical interaction between the two ends of the exon is proposed to be a crucial event. For this interaction to occur the exon ends would have to find each other: i.e., collide. A productive collision across the exon and involving its ends occurs when both ends of the exon are suitably occupied and they approach each other in the correct orientation through thermal movements. The ends will then be at a fitting distance, y_i, from each other as shown in Fig. 9A. Assuming the RNA behaves as a worm-like chain with contour length much greater than persistence length, the probability for a given end-to-end distance as a function of exon size can be obtained using a Gaussian approximation (Becker et al. 2010). Using this approximation, the ends of the molecule while inside the range of distances within which attractive and repulsive forces become important can be modeled. Taking this range to be small with respect to the fitting distance, y_i, and applying the mean-value theorem for integrals, the collision probability can be estimated with the formula

S24. $P(Y_i,x) \approx k_i Y_i^2 Z^{-3/2} e^{-3Y_i^2/Z}$

Here Z is the size of the exon in nt figuring 2 nt per nm, the index i refers to the splice sites used (4 sets, Table 1: sets 2, 3, 5 and 7), Y_i is the distance y_i divided by the square root of the Kuhn length for an RNA molecule, assuming a cationic concentration equivalent to ~300 mM and a Kuhn length of ~3.0 nm (Chen et al. 2012). The catch-all constant k_i depends on the Kuhn length, the range of distances within which attractive and repulsive forces become important and the chance that a collision will result in an association; k_i is independent of the length of the exon in question. Although the values of these parameters are unknown, we consider them as constant for any set of splice sites. The association rate constant a is proportional to $P(Y_i,x)$. The key constant D, the ratio of the disassociation and assembly rate constants (a/d) of the exon definition complex, determines the efficiency of splicing. D is inversely proportional to a and so will be inversely proportional to $P(Y_i,x)$, as will be seen in equation S29 below.

Stability of the exon definition complex

We propose that enhancers act by increasing the stability of the exon definition complex. In this case, the rate of dissociation, d, should be proportional to the rate at which random collisions transfer kinetic energy greater than a threshold, $E_{threshold}$, to the complex; i.e., the complex is broken through collisions with other molecules. The addition of a single ESE was taken to increase this energy threshold by a fixed constant amount $\Delta E = E_{enh}$. Any additional copies of this ESE will increase this energy threshold by an additional ΔE.

Let's obtain an equation relating the rate of dissociation of the complex and the energy increment brought about by the enhancer. For a simplified analysis, we considered the collision between the complex on the exon of interest, C, and a molecule, M, in the absence of the enhancer. This collision transfers enough kinetic energy to cause dissociation of C if the collision is head-on and the relative kinetic energy of M is higher than a threshold. However, if the collision is not head-on, then the geometry of the collision should be taken into account. As an approximation, C and M were modeled as spheres; the angles between the collision trajectory and the tangent plane at the site of contact determine the energy that is transferred. An analogous situation is found when modeling reactive encounters (Atkins and de Paula 2002): following a traditional analysis of such situations, an equation for the rate of dissociation d_o was obtained

S25. $d_o \approx \alpha\, e^{-Ethreshold/(kT)}$

where α is a proportionality constant that takes into account all speeds and collision angles, k is the Boltzmann constant, T is the absolute temperature and $E_{threshold}$ is the energy necessary to cause dissociation of C. If the situation is modified by adding an enhancer with its corresponding activator, the energy required to cause the complex to dissociate becomes $E'_{threshold} = E_{threshold} + E_{enh}$, making the new dissociation constant, d_E,

S26. $d_E \approx \alpha\, e^{-(Ethreshold+Eenh)/(kT)} = \alpha\, e^{-Ethreshold/(kT)}\, e^{-Eenh/(kT)} = d_o\, e^{-Eenh/(kT)}$

Comparing equations S25 and S26, we observe that the addition of a single enhancer modified the dissociation rate by a factor of $c_E = e^{-Eenh/(kT)} < 1$, and

S27. $d_E \approx d_o c_E$

Repeating the analysis to account for the addition of n identical enhancers yielded

S28. $d_{En} \approx d_o c_E^n$

Notice that these results could be generalized by making $c_E = \gamma\ e^{-Eenh/(kT)}$, which allows γ to account for other parameters such as occupancy.

Consequently, each ESE affects D (see equation S29) by the factor c_E and n of those sequences affect it by c_E^n. To be consistent with the results observed for the ESE under consideration, this effect was taken to be independent of position. In this simple scenario, multiple enhancers were modeled as independent, leading to an exponential dependence of D on the number of enhancers present. Note that although the effect of increasing the number of ESE copies affects D exponentially, splicing is affected in a more complex fashion, and ends up increasing sigmoidally (see Supplemental Fig. S8).

Combining all the effects

Since D is either proportional or inversely proportional to each effect and these effects are assumed to be independent, their combined effect should be given by simple multiplication of the individual effects. Assuming stability effects for the ESE, the ESS and the reference sequence, equation S29 (identical to equation 7 in the text) was obtained

S29. $D \approx K_i\ Y_i^{-2}\ c_E^{nE}\ c_R^{nR}\ c_{SF}^{nF}\ c_{SL}^{nL}\ c_{SI}^{nI}\ Z^{3/2}\ e^{3Yi^2/Z}$

Modeling recruitment as an alternative to stability

To evaluate recruitment as an alternative to stability as the mode of action of ESEs and ESSs, we assumed that the probability (and thus the rate) of association would be affected by the number of ESEs and ESSs in a linear manner (Hertel and Maniatis 1998), generating the approximation

S30. $D \approx K_i\ Y_i^{-2}\ c_R^{nR}\ Z^{3/2}\ e^{3Yi^2/Z} / (1+c_E\ n_E+ c_{SF}\ n_{SF} +c_L\ n_{SL} +c_{SI}\ n_{SI})$

Preliminary attempts using this equation gave values for T, C, K_2, K_3, K_5, and K_7 of the order of thousands or more, suggesting a rate of dissociation that was much greater than the rate of association and convergence was difficult to achieve. To solve this issue, a was assumed to be negligible compared to d in equation 6

S31. $pso \approx 100\ e^{-T/(1+D)}/(1+C/(1+D))$

to generate

S32. $pso \approx 100\ e^{-T/D}/(1+C/D)$

where pso is proportion spliced out.

The data available cannot be used to separate the contributions of T, C and K_i (as part of D) in equation S32. However, if we assume a value for T, C and K_i can be optimized. We decided to retain the value for T obtained with the stability model: 5.24 (see Table 3). After a first round of optimization, a second round was performed using as input only DEs for which a positive prediction was obtained in the first round. Additionally, in order to mimic the effects of saturation, any predicted value above 100% was taken to be 100% and any negative value was taken to be 0%. The optimized values for the model are shown in Supplemental Table S3. We considered the possibility that these regulatory sequences affect both splice sites by squaring each contribution (c_E, etc.); this modification did not improve R^2 for either the input data itself or the more complex DEs (data not shown).

Detailed Materials and Methods

Double stranded oligomers

Sense and antisense oligomers were purchased from either Invitrogen or Fisher Scientific and annealed by mixing them together at a concentration of 40 µM each in 300 mM potassium acetate. These mixtures were placed in a 500 ml boiling water bath for 5 min and allowed to slowly cool down to room temperature in the bath. The annealed oligomers were phosphorylated at a final concentration of 100 nM with T4 polynucleotide kinase from New England Biolabs (NEB) by following the manufacturer's protocol. We call these molecules phosphorylated double stranded oligomers or P-ds-oligos.

Removable Adapters

Removable adapters or RAs are sequences that contain recognition sites for type IIS restriction enzymes (REs) that cut at both ends of the adapter. Due to the nature of type IIS enzymes, the sequence of the overhangs generated can be chosen essentially without restrictions. Two kinds of removable adapters were designed. RAs of the first kind (RA-I) are removed by a single type IIS RE that cuts on both sides of the adapter. RAs of the second kind (RA-II) allow independently controlled cuts on either end: one type IIS RE cuts on one side while a different type IIS RE cuts on the other side.

Plasmids

Supplemental Fig. S1 shows the features of the modified dhfr minigene used to harbor the DEs.

All modifications performed on plasmids were verified by sequencing the appropriate regions (Genewiz).

A "drafting" plasmid, pAL-SB, was derived from pEGFP-C3 (Addgene) to facilitate the construction of DEs. This plasmid contains an adapter that allows the use of type IIS enzymes BsmBI and BsaI to add building blocks at either flank of the DE in progress, but it does not contain a *dhfr* minigene. The finished DEs can be copied and pasted into any of the receiving plasmids (see below). In order to provide flexibility for future extensions, BfuAI sites were removed from pEGFP-C3. For this purpose, nested PCR was performed using two primer pairs: oligo36 and oligo37, and oligo38 and oligo39; the oligo36 and oligo39 primers were used for the final amplification, which appended temporary BsaI sites at both ends to generate the appropriate overhangs. The products were cut with BsaI and ligated into pEGFP-C3 which was previously digested with BfuAI. This was followed by transformation of DH5-alpha competent cells and selection in kanamycin (Sigma-Aldrich). Successful clones were selected by evaluating digestion patterns with BfuAI. The intended use of BsaI for DE construction required removal of the BsaI site from pEGFP-C3. To remove it, a PCR fragment was obtained using primers oligo40 and oligo41; this PCR fragment and the previously modified pEGFP-C3 plasmid were digested with

BsaI and EcoO109I, mixed and ligated together. After these preparations and in order to add the appropriate adapter, the oligo42 primer was designed. Along with oligo43, it was used to amplify a fragment from pEGFP-C3. Both the adapter-containing PCR fragment and the plasmid were digested with PstI (NEB) and HindIII (NEB), mixed and ligated together to obtain pAL-SB.

As a starting point for all the dhfr minigene containing plasmids, pMA-URA was made from plasmid pUHD10-3 (Gossen and Bujard 1992). The whole dhfr minigene was copied from a DE-containing plasmid derived from the pD12 plasmid (Zhang et al. 2009) and integrated into pUHD10-3 by placing it under the control of the tet-responsive promoter with a SV40 polyA signal for cleavage and polyadenylation. During the transfer, all the ATGs in exon 1 were eliminated, the first out-of-frame ATG in exon 3 was eliminated, and the following in-frame ATG was modified to conform to the Kozak sequence. These modifications were performed to reduce possible translation effects of modifying the middle exon. Additionally, the DE was substituted with an RA-II. The RA-II employed relies on BfuAI and BtgZI for its function. Therefore, the BtgZI site present in pUHD10-3 was removed. We call the dhfr minigene in pMA-URA the modified dhfr minigene; its sequence is included below. The details for its generation follow. For the removal of the BtgZI site, the plasmid was cut with BtgZI (NEB) and NgoMIV (NEB) and dephosphorylated; P-ds-oligo oligo1/oligo2 was ligated to this plasmid using T4 ligase (NEB) by following the manufacturer’s protocol. The dhfr minigene was transferred from a DE-containing plasmid derived from the pD12 plasmid (Zhang et al. 2009) and simultaneously modified in five stages using PCR and P-ds-oligo ligations as described below. An intermediate plasmid, piMA-F5, was obtained by PCR amplification of fragment F5 (oligo3 and oligo4), digestion with XbaI and MluI and ligation into the modified pUHD10-3 after its digestion with XbaI and BtgI and dephosphorylation. This was followed by transformation of DH5-alpha competent cells and selection in ampicillin (Sigma-Aldrich). A clone with piMA-F5 was chosen by verification of the expected sizes of appropriate PCR products. Similarly, fragment 4 (oligo5 and oligo6) and fragment 3 (oligo7 and oligo8) were sequentially added using BfuAI (NEB) and SphI (NEB) for the digestions of the PCR products and BtgZI (NEB) followed by SphI for the plasmids. Fragment 2 was added as P-ds-oligo oligo9/oligo10 to the previous plasmid digested with BtgZI followed by SphI. Fragment 1 was added as P-ds-oligo oligo11/oligo12 to the resulting plasmid after digestion with BtgZI and BsiWI. This new plasmid was digested with NotI (NEB) and NheI (NEB) and ligation with P-ds-oligo oligo13/oligo14 generated pMA-URA.

A series of plasmids containing an RA-II were generated: the construction plasmids. These plasmids contain the modified dhfr minigene and an RA-II surrounded by an appropriate SS set to allow “on-site” construction of DEs (see below). These plasmids are derived from pMA-URA. For SS Set 7, the 5’SS, the polypyrimidine tract, and the 3’SS were added by three sequential rounds of ligation, transformation and selection using the restriction sites for NheI, for NotI and SphI, and for BtgZI and SphI, respectively, and three pairs of P-ds-oligos: oligo15/oligo16 for the 5’SS, oligo17/oligo18 for the polypyrimidine tract, and oligo19/oligo20 for the 3’SS. A similar procedure was used for SS Set 3, but oligo21/oligo22 was used for the second ligation. For SS Set 5, the 5’SS, and the polypyrimidine tract together with the 3’SS were added sequentially using the restriction sites for NheI, and for NotI and SphI respectively and two pairs of P-ds-oligos: oligo23/oligo24 for the 5’SS, and oligo25/oligo26 for the rest. For SS Set 6, a similar approach was used but oligo27/oligo28 was used for the second ligation.

The receiving pMA plasmids contain an RA-I and were used for incorporating the DEs made in the pAL-SB plasmid into an modified dhfr minigene. Each receiving plasmid contains a SS set. For SS Set 5, P-ds-oligo oligo32/oligo33 was ligated into the pMA-URA plasmid after digesting the latter with NheI and NotI. The intermediate plasmid containing the 5'SS of SS Set 7 described in the previous paragraph was digested with SphI and NotI and ligated to P-ds-oligo oligo34/oligo35 to generate the receiving pMA plasmid for SS Set 3. The RA-I used in the receiving pMA plasmids is different from the RA-II in the plasmids that allow stepwise construction of the DE and it leaves different overhangs upon its removal.

The pMA-FW plasmid provided the basis for incorporation of the modified dhfr minigenes into the genome. It contains a kanamycin resistance gene for initial selection of the cell line, a promoterless puromycin gene for subsequent selection of site-specific recombinations with DE-containing plasmids, an attP site for site-specific recombination and only the downstream half of the modified dhfr minigene. This plasmid was derived from pEGFP-C3. The CMV promoter and the EGFP gene were cut out with AseI and BamHI and in its stead a promoterless puromycin resistance gene was ligated by amplification from ptTA (a kind gift from Jim Manley) using primers oligo44 and oligo45 and digestion with AseI and BamHI. This new plasmid was digested with XhoI (NEB), dephosphorylated and ligated to P-ds-oligo oligo46/oligo47, which provides an attP site for PhiC31 recombinase (Groth et al. 2000). Several clones were sequenced and, of the two orientations possible for the attP site, the one in which oligo46 was on the sense strand of the puromycin gene was chosen. This intermediate plasmid was digested with AseI and XhoI. The downstream half of the minigene starting in the middle of intron 2 (1 bp downstream from the EcoRI site) and including 100 bp downstream from the polyA site was amplified from pMA-URA using the primers oligo48 and oligo49, digested with AseI and XhoI, mixed with the digested intermediate plasmid and ligated to obtain pMA-FW.

Plasmid pMA-IC allows reconstitution of a fully functional puromycin resistance gene and a DE-containing modified dhfr minigene upon site-specific recombination with the sequence from pMA-FW (Supplemental Fig. S10). The DE-containing plasmids for site-specific recombinations contained a CMV promoter to drive the puromycin resistance gene after site-specific recombination, the upstream half of the modified dhfr minigene including the DE for reconstitution of the modified dhfr minigene, and an attB site for site-specific recombination. An "empty" pMA-IC plasmid was constructed from a pMA-URA derived plasmid which contained an irrelevant sequence between the NotI and the NheI sites. The CMV promoter was amplified from pEGFP-C3 using oligo50 and oligo51; both the pMA-URA derived plasmid and the PCR product were cut with XbaI (NEB) and EcoRI (NEB) and ligated together. This step removed the downstream half of the minigene. An attB site for PhiC31 recombinase (Groth et al. 2000) was ligated into the XhoI site of the modified plasmid as P-ds-oligo oligo52/oligo53. Of the two orientations possible, the one in which oligo52 was on the sense strand of the partial dhfr minigene was chosen. The BtgZI site was removed to enable future extensions by digesting the previous plasmid with NcoI (NEB) and BsaAI (NEB) and ligating P-ds-oligo oligo54/oligo55.

To serve as the basis for the coupled-standards, the plasmid piS-Std was generated, which contained the skipped cDNA for the modified dhfr minigene. The cDNA of a transient transfection with a DE of 110nt (SS Set 7) composed exclusively of reference sequences was used for PCR amplification using primers oligo56 and oligo6. The PCR fragments obtained were digested with BfuAI and BsiWI and ligated into the plasmid piMA-F5 previously digested sequentially with BsiWI and BtgZI. Plasmid piS-Std was selected by the size of the products in appropriately chosen PCR amplifications. An adapter to facilitate subsequent ligations was added to generate piS-StdwAd by digestion with NcoI and XbaI, dephosphorylation and ligation of P-ds-oligo oligo57/oligo58. For generating the Gamma Actin coupled-standard, piSActin-Std, cDNA generated from MA-tTA cells by reverse transcription with primer oligo61 was amplified using primers oligo62 and oligo63. The PCR product and plasmid piS-StdwAd were digested with EcoRI and NotI and ligated together. For generating the coupled-standard for SS Sets 1, 2 and 4, piSI-CAG-Std, cDNA from a transient transfection using a DE of 110nt composed exclusively of reference sequences and SS Set 1 was amplified using primers oligo59 and oligo60. This PCR product and plasmid piS-StdwAd were digested with EcoRI and NotI and ligated together. The coupled-standard for SS Sets 3, 5, 6 and 7, piSI-CAA-Std, was made analogously from a transient transfection using a DE with SS Set 3: a mutation of A to G at position 64 of the DE was deemed innocuous and accepted.

DE construction

Most DEs were constructed in a stepwise fashion by ligating P-ds-oligos oligo64/oligo65 (RR), oligo66/oligo67 (EE), oligo68/oligo69 (ER), oligo70/oligo71 (RE), oligo72/oligo73 (SS), oligo74/oligo75 (SR), and oligo76/oligo77 (RS) into pAL-SB or the RA-II-containing construction plasmids (previous section). For the pAL-SB plasmids, the appropriate plasmids were digested with either BsmBI (to add a building block upstream of the DE in progress) or BsaI (to add a building block downstream). The final DEs were amplified with primers oligo78 and oligo42, digested with BbvI and ligated to the appropriate receiving pMA plasmid after removing its RA-I by digestion with BfuAI or its isoschizomer BveI (Fermentas). For the RA-II-containing construction plasmids, appropriate plasmids were digested with BfuAI (to add a building block downstream of the RA), BtgZI (to add a building block upstream) or both (to remove the RA or replace it with a building block). RA-II-containing construction plasmids with SS Sets 3 and 7 were digested with NheI and BtgZI to incorporate 22 bp DEs by ligating a P-ds-oligo oligo79/oligo80 or oligo81/oligo82 as appropriate. Constructs using SS Set 1 and SS Set 2 were made by amplifying the corresponding DEs from plasmids with SS Sets 3 and 7, respectively, using primers oligo29 and oligo30, digesting both the PCR products and pMA-URA with NheI and NotI and ligating them together. By following this protocol, DEs using SS Set 4 were made by amplifying the corresponding DEs from plasmids with SS Set 3 using primers oligo29 and oligo31.

For generating the DE-containing pMA-IC plasmids, DEs were amplified by PCR from the appropriate modified dhfr minigenes using oligo29 and oligo83, digested with NotI and EcoRI and ligated into the pMA-IC plasmid, which was previously digested with NotI and EcoRI and dephosphorylated.

Psi measurement

RNA was extracted from transiently transfected cells using the RNA Spin Mini kit (GE Healthcare) and quantified using a Nanodrop 1000 (Thermo Scientific). Lack of degradation was assessed by gel electrophoresis. Reverse transcription (RT) was performed using an Omniscript kit (QIAGEN) in 10 µl reactions with 400 ng of RNA for each sample using the primer oligo84 at 100 nM. To measure the ratio between the mRNA molecules that skip the DE and those that contain it for SS Sets 3, 5, 6 and 7, the appropriate coupled-standard was prepared from plasmid piSI-CAA-Std by digestion with EcoO109I (NEB) followed by inactivation. The concentration of this digested plasmid was approximately 10^{10} plasmid molecules per µl based on absorbance measurements. This solution was diluted to approximately 10^8 molecules per µl and a dilution series was prepared: 10-fold dilution per step. The starting solution was labeled as having exactly 10^8 arbitrary units/µl. Given that the coupled-standard plasmid concatenates a molecule that skips the DE and a molecule that includes it (see Supplemental Fig. S11), each diluted solution contains equimolar amounts of each, which enables accurate calibration by QPCR of one type of molecule relative to the other. Furthermore, all coupled-standards were calibrated to each other by means of the common "skipped mRNA" region to further allow comparisons among standards. QPCR was performed in 20 µl reactions that included 400 nM of forward and reverse primers, 2 µl of a 1:5 dilution of the RT product for each sample and 10 µl of 2X Power Green QPCR Master Mix (Applied Biosystems) using a 7300 PCR System (Applied Biosystems) according to the manufacturer's protocol. The data was analyzed using the software provided by the manufacturer. The primer sets used in QPCR reactions share the reverse primer oligo83 and include either oligo85 as the forward primer to detect the molecules that contain the DE or oligo86 to detect molecules that skip it. Each experimental QPCR Ct result was compared to a dilution series of the coupled-standard, comparing included to included and skipped to skipped. Because the coupled-standards contain equimolar amounts of the included and skipped, and since background cross-detection is negligible (<1%, data not shown), the ratios of skipped to included (SOI) are automatically calibrated. The psi was obtained by the formula psi=100/(1+SOI). For SS Sets 1, 2 and 4, the coupled-standard derived from the plasmid piSI-CAG-Std was used along with oligo87 as the forward primer for the detection of inclusion. Importantly, because of the placement of the QPCR primers all amplified products consist of identical sequences for each SS set and in particular are independent of the E, S and N combinations used, thus ensuring equal PCR efficiencies.

To assess the expression levels of the minigenes in stable transfections, the Gamma Actin primer oligo61 was added to the RT reaction. To quantify the mRNA levels for Gamma Actin, the coupled-standard derived from piSActin-Std was used in QPCR reactions. Comparisons between Gamma Actin mRNA and mRNA for the minigene are affected by the relative efficiency of the two reverse transcription primers, disallowing a direct comparison. However, normalization to Gamma Actin mRNA enables direct comparisons for the transcription levels of the minigene between samples.

QPCR coupled-standard calibration

Ten-fold dilutions of the coupled standard were quantified using primers to detect either included (blue curve) or skipped (red curve) molecules in two or more QPCR reactions (see Supplemental Fig. S13, upper panel). For quantification of the samples, a separate plate was used for each type of reaction (included, skipped, and an actin mRNA control when appropriate); each plate included a dilution series of the coupled-standards. Running the standards in each plate avoids small interplate differences. The same layout was used for each set of samples plus standards to avoid positional differences.

The two types of reactions have similar efficiencies as shown by the slopes of the curves in Supplemental Fig. S13: 88% using the primers to detect the included molecules and 87% using those for the skipped molecules. The QPCR software was used to automatically set the intercept used for Ct determination. The gap between the lines depends among other things on this intercept choice. All these differences are intrinsically taken into account by using relative quantification with respect to the standard dilution series. The high correlation confirms that the dilutions were precise and that QPCR is an adequate tool for these measurements. The measured psi of the coupled standard was verified to be 50.2% ±1.5 and was independent of the initial concentration used (Supplemental Fig. S13, lower panel). As a final test, a set of samples covering a psi range from ~7% to ~85% (n=24) was measured twice using separate QPCR runs. Good reproducibility (R^2 = 0.99; slope = 0.99; intercept = 0.71%) was observed between the 2 sets of measurements (see Supplemental Fig. S13C), the standard deviation of the differences in psi being only 2%.

Transfections

For transient transfections, cMA-HEK293-tTA cells were grown in 10 cm dishes to ~80% confluence. Cells from each dish were plated in 6 wells of a 6-well plate and incubated at 37°C for 24 hours. Transient transfections were performed using 600 ng of plasmid and 4 µl of Lipofectamine 2000 (Invitrogen) using Opti-MEM I (Invitrogen) according to the manufacturer's protocol. Cells were incubated at 37°C for 25 hours before RNA extraction.

For stable transfections, cMA-FW cells were grown in 10 cm dishes to ~80% confluence. Transfections were performed using 2.4 µg of the DE-containing pMA-IC plasmid, 15 µg of pPGKPhiC31obpA plasmid (Addgene) and 30 µl of Lipofectamine 2000 using Opti-MEM I according to the manufacturer's protocol. Successful PhiC31 site-specific recombinations (see Supplemental Fig. S10) were selected after 72 hours of incubation at 37°C by adding puromycin (Sigma-Aldrich) to a final concentration of 4.2 µg/ml. In effect, only site-specific recombination allows reconstitution of the minigene (Supplemental Fig. S10). The puromycin containing medium was changed every 5 days. After ~3 weeks of puromycin exposure, the surviving clones were pooled and allowed to grow in 6-well dishes before RNA extraction.

Cell lines

HEK 293 cells were modified to express the tet-Off trans-activator (Gossen and Bujard 1992) by co-transfecting 1 µg of pUHD15-1 plasmid and 0.1 µg of pLi082 plasmid, which provides hygromycin resistance. Clones were grown in 100 µg/ml hygromycin and recloned. Individual clones were chosen and expression of a tetracycline-response-element controlled minigene was evaluated. A clone, cMA-HEK293-tTA, that displayed adequate expression levels and a good response to doxycycline was chosen (data not shown). This clone was used for all transient transfections.

This cell line was used to generate the cMA-FW by incorporation of pMA-FW digested with MluI and transfected by electroporation using Nucleofector II (Lonza) according to the manufacturer's protocol. Cells were incubated for 48 hours at 37°C and successful genomic incorporations of the transfected DNA were selected by adding G418 (Invitrogen) at a final concentration of 500 µg/ml. Site-specific recombination into these cells was evaluated with a pMA-IC plasmid containing the DE RRRERE. One clone cMA-FW was selected that provided adequate levels of expression for the reconstituted minigene and generated an acceptable number of colonies after puromycin selection (data not shown). The presence of a single genomic copy of pMA-FW was evaluated by verifying the full disruption of attP sites in puromycin surviving colonies by PCR using primers oligo83 and oligo88 (data not shown): full disruption in multiple independent site-specific recombinations, evidenced by the absence of PCR products, is expected only if a single attP site is present since reconstitution of a single puromycin resistance gene suffices for survival. This result was confirmed by using a Southern blot (data not shown). Also genomic DNA was digested with NspI, diluted and ligated to obtain DNA circles; inverse PCR was then performed using nested primer pairs: oligo83 and oligo44 in the first PCR reaction and oligo89 and oligo90 in the second. These products were cloned into the Not I site of pMA-URA and sequenced. This information allowed mapping of the genomic integration point to PLEKHG1 in chromosome 6, specifically 141 bp before its 23 nt exon (i.e., intron 14 in NM_001029884.1). The location was verified by detection of PCR products that crossed the 2 ends of the integration site in the genomic DNA using primer pairs oligo91 with oligo89 and oligo92 with oligo93. Additionally, the size profile observed in the Southern Blot coincided with that predicted from integration at this genomic location.

Since the minigene was integrated into the sense strand of the PLEKHG1gene, we were concerned about the possibility that fusion transcripts would be synthesized in which a PLEKHG1 exon was spliced to a DE, leading to a counterfeit measurement of inclusion. However, no such fused mRNAs were detected by PCR using oligo94 (in the PLEKHG1 sequence) and oligo83 (in dhfr exon 3) probably due to the presence of a SV40 polyA site in pMA-FW upstream of the minigene.

Model Optimization

A Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) adapted from Press et al. (Press et al. 2007) implementing walls to force all optimized values to be non-negative and using explicit gradient was written in Perl for minimizing the sum of the squared differences between observed and predicted pso (equation 6). All values for the model were seeded as 1 except for T, C and Y_i. T was allowed to vary between 0 and 10 in steps of 1, C between 10^{-8} and 100 with a factor of 10 between steps and Y_3 between 1 and 25 in steps of 1. Y_2 and Y_3 were started with the same seed but subsequently allowed to vary independently; any other Y_i was assumed to be equal to either Y_3 or Y_2 as indicated in the Results. To improve convergence, the routine was modified to reset the direction for line minimization to that of steepest descent if the vector of values for the model did not change when a full step in the updated BFGS direction was taken. This modification reduces the number of seed sets for which the program crawls to a stop without reaching a convergent solution (a stalled run) and practically increases the number of convergent solutions by allowing otherwise stalled runs to converge. So as not to include the results of stalled runs, a set of values was taken as a solution if and only if, for a set of input data, considering the full set of seeds, it provided the minimum sum of the squared differences. This same minimum sum had to be obtained several times with all optimized values identical to at least 5 significant figures (using different seeds). The magnitude of the gradient had to be smaller than 10^{-7} in at least 2 cases. If no such solution was found the program was said to have failed to find a convergent solution. These criteria were met for all optimized values except for C, which was so low as to be negligible. In this case the value yielding the minimum sum with the minimum gradient was used. In fact, setting C equal to zero did not affect the results.

Using all the data points available (Supplemental Table S1), the program failed to converge on a set of values for the model. We reasoned that the multidimensional surface was too complex and relatively flat causing the program to crawl to a stop when exploring it. To address this issue, we simplified the data by using our observations that ESEs are position independent and that, using single-ESS DEs, the effects of multiple-ESS DEs can be predicted. We thus condensed all ESE results corresponding to a given number of ESEs by their average and removed the 36 data points corresponding to multiple-ESS DEs. (The data points for ESEs exclusively with SS Set 7 were used.) We also found it necessary to remove a single outlying point (SS Set 3, length = 206 in Fig. 2) from the 19 size perturbation points in order to achieve reproducible convergence. This point also did not agree with the data from permanent transfections (Supplemental Fig. S2). While this reduced and condensed set was used for optimizing the values for the model, the single-parameter-perturbation evaluation of the model was performed using all the 112 points available.

Sequence of pMA-URA

Shown below is the sequence inserted into pUHD10-3. Regions of exons 1 and 3 are shown in blue. The regions of introns 1 and 2 that were used are shown in gray. The restriction sites used for incorporation of DEs are indicated: the NotI site is highlighted in blue and the NheI site is highlighted in yellow. The removable adapter is highlighted in green. The first and last four nucleotides of the entire sequence correspond to the overhangs added for cloning. The 3 nucleotides, TAC, that follow the first four were added to facilitate the transfer.

```
CGCGTACGGTTCGACCGCTGAACTGCATCGTCGCCGTGTCCCAGAATAAGGGCATCGGCAAGAACGGAGACCTTCCC
TGGCCAAAGCTCAGgtactggctggattgggttagggaaaccgaggcggttcgctgaatcgggtcgagcacttggcg
gagacgcgcgggccaactacttagggacagtcatgaggggtaggcccgccggctgcagcccttgcccatgcccgcgg
tgatccccatgctgtgccagcctttgcccagaggcgctctagctgggagcaaagtccggtcactgggcagcaccacc
ccccggacttgcatgggtagccgctgagatggagcctgagcacacgcggccgccgcatgcaacatcgcacctgctag
ctggccagtgagatccaagaatcttcctgtctctgctgatccactgataggattacaagtacatgccaccaagccca
gcttcctcttaccaggtgctggggaccaaacttaggccctcattcctacacagtgaatacttgactttgttatcacc
caaccctaataaataactcactatccaaacaagttgaaacccttagaattctgtgttgctccagcatgatgttgtgg
taaacgttaatacaataagatgcacaggtcataagtgcacattagctaagtgttgacaaagacttagacctacataa
cttaaccctattagccctccagaaagttcctcattctccattccaggcaactttcatcacaccacatcatgtacaac
tactattgaagttgttttccactatagatacaatgagatgtcacatacggctttgtgttttgatttgcaagtaccaa
tcgagtatgaaatatggagtggatattggacattggccaccatctaaatactttgtgttaaaagaattggttttcat
aatttgttttgtactgactgctggctagtcagattacctgactagtatggacaggattttgcaataatcataattct
tttttcagGGAACCACCACAAGGAGCTCATTTTCTTGCCAAAAGTCTGGACGAAGCCTTAAAACTTATTGAACAACC
AGAGTTAGCAGATAAAGTGGAGCTGTCATGGTTTGGATAGTTGGAGGCAGTTCCGTTTACAAGGAAGCCATGAATCA
GCCAGGCCATCTCAGACTCTTTGTGACAAGGATCATGCAGGAATTTGAAAGTGACACGTTCTTCCCAGAAATTGATT
TGGAGAAATATAAACTTCTCCCAGAGTACCCAGGGGTCCTTTCTGAAGTCCAGGAGGAAAAAGGCATCAAGTATAAA
TTTGAAGTCTATGAGAAGAAAGGCTAACAGAAAGATACTTGCTGATTGACTTCAAGTTCTACTGCTTTCCTCCTAAA
ATTATGCATTTTTACAAGACCATGGGACTTGTGTTGGCTTTAGATCTATGAGTTATTCTTTCTTTAGAGAGGGATAG
TTAGGAAGATGTATTTGTTTTGTGGTACCAGAGATGGAACCTGGGATCCTGTGCATCCTGGGCAACTGTTGTACTCT
AAGCCACTCCCCAAAGTCATGCCCCAGCCCCTGTATAATTCTAAACAATTAGAATTATTTTCATTTTCATTAGTCTA
ACCAGGTTATCTAG
```

Below is a typical DE (REERRR, using SS Set 7) that would be cloned into the above pMA-URA using the NotI and NheI sites. The NotI site is highlighted in blue and the NheI site is highlighted in yellow, the regions of introns 1 and 2 that were used are shown in gray and the exon is shown in green, and splice site sequences are **bolded**. The restriction sites used for incorporation of DEs are indicated:

NotI
gcggccgctgttaacgcagtgtt**tctctaactttcagG**ccaaacaaCCAAACAAccaaacaaUCCUCGAAccaaac
aaUCCUCGAAccaaacaaCCAAACAAccaaacaaCCAAACAAccaaacaaCCAAACAAccaaacaaCA**CAAgtaag**
tgctagc
NheI

Supplemental discussion

DEs as a model for exon definition

Many systems used to study splicing, especially *in vitro* splicing, use 2-exon substrates or substrates with short (<200 nt) introns, favoring intron definition rather than exon definition (Talerico and Berget 1994; Fox-Walsh et al. 2005). In contrast, we sought here to focus on exon definition, and so studied splicing of an internal exon and used longer introns (~300 and ~600 nt). Importantly, weakening either splice site of the internal exon resulted only in increased skipping, as expected with exon-definition, with no signs of intron retention (data not shown).

Tethered end collisions across the intron

Several similarities have been noted between the size restrictions for exons in exon definition and those for introns in intron definition. For example, introns longer than ~300 nt are disfavored in organisms relying mostly on intron definition (*D. melanogaster*) and exons longer than ~300 nt are disfavored in organisms relying mostly on exon definition (humans) (Sterner et al. 1996; Fox-Walsh et al. 2005; Xiao et al. 2007). Moreover, Garcia-Blanco and colleagues presented evidence supporting pairing of the ends of introns via three dimensional diffusion (Pasman and Garcia-Blanco 1996), a mechanism similar to that proposed here for exon end pairing. Interestingly, the size distributions of short introns in human and Drosophila are greatly disjoint (Fig. 6 in (Fox-Walsh et al. 2005)). The optimum size for splicing is greater in human (90 nt) than in Drosphila (75 nt) nuclear extracts (Guo et al. 1993). These observations suggest that a size dependence similar to that in Fig 2 could explain this species difference by assuming tethered end collisions across the intron with a different y_i for each organism. This difference could be dictated by differences in the size and/or number of the proteins involved.

Recruitment model vs. stabilization model

Changes in stability, expressed as the rate of dissociation (d within equation 5), should respond exponentially to the number of ESEs. This stability model predicts a sigmoidal curve but with a near linear relationship between psi and the number of ESEs over much of the range examined and accounts for the saturation effect when more than 4 ESEs are used (Fig. 4). In contrast, recruitment depends on a change in binding probability of the splicing machinery, which is expected to be linear with respect to ESE number. This linearity could be incorporated in a (the association rate constant) within equation 5 (Supplemental Materials and Methods). The resulting recruitment model led to a fit for predicting the results on the single-parameter-perturbation data (an R^2 of 0.92, a slope of 0.90, and an intercept of 5.62%) that was nearly as good as the stability model (Supplemental Fig. S5). However, it produced a negative exponential shaped curve that did not fit the ESE data as well (See Supplemental Fig. S8) and unlike the stability model it performed poorly for the complex DEs (R^2 of 0.37 compared to 0.86). In particular, for the constant size class of 110 nt which isolates the effect of ESE/ESS combinations, even though an acceptable R^2 of 0.84 was obtained, a slope of 1.78 and an intercept of -63% revealed a flawed performance compared to the stability model (compare Supplemental Fig. S14 and Supplemental Fig. S6).

ESS number and position effect

We showed that the effect of multiple ESSs could be predicted by their linear combination as long as the particular characteristics of positions 1 and 6 were taken into account (Fig. 6B). For modeling, we contented ourselves with considering the action of ESSs to be opposite that of ESEs; that is, as destabilizing elements. Although only the data for single-ESS DEs were used to optimize the model, the effect of multiple ESSs (which included the saturating case of 6 ESSs) was accurately predicted (Supplemental Fig. S15). These predictions were in fact more highly correlated ($R^2 = 0.82$) than simply summing the effects of the individual ESSs (Eq. 1 and Fig. 6B, $R^2 = 0.73$). The position effect seen for ESSs suggests that ESSs may act by destabilizing bound U1 snRNP or even blocking its binding. Further studies using different ESS/splice site combinations and incorporating competition between RBPs and spliceosomal complexes into the model could be used to explore these ideas.

Supplemental Tables

Supplemental Table S1. *Exon inclusion of DEs: Data used for model optimization*

SS Set No.	3’SS	5’SS	Exon size (nt)	Internal Name	Code	psi	Std error
Size perturbation							
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	14	i6u	0R	42	4
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	46	i6uRR	2R	97	1
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	78	i6u(R)x4	4R	96	2
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	110	i6u(R)x6	6R	76	5
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	142	i6u(R)x8	8R	33	11
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	174	i6u(R)x10	10R	13	5
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	206	i6u(R)x12	12R	12	6
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	238	i6u(R)x14	14R	4	1
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	270	i6u(R)x16	16R	7	4
3	UCUCUUUUUUUCAG/G	CAA/GUAAGU	302	i6u(R)x18	18R	5	2
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	22	irm1I3	½R	4	2
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	46	iRRm1I3	2R	58	14
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	78	i(R)x4m1I3	4R	94	3
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	110	i(R)x6m1I3	6R	94	3

2	UCUCUAACUUUCAG/G	CAG/GUAAGU	142	i(R)x8m1I3	8R	78	2
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	174	i(R)x10m1I3	10R	31	1
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	206	i(R)x12m1I3	12R	10	4
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	238	i(R)x14m1I3	14R	8	3
2	UCUCUAACUUUCAG/G	CAG/GUAAGU	270	i(R)x16m1I3	16R	4	0
			ESE Perturbation				
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i555	RRRRRR	7	0
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i255	ERRRRR	30	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i455	RERRRR	21	3
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i525	RRERRR	34	11
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i545	RRRERR	27	5
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i552	RRRRER	30	7
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i554	RRRRRE	18	3
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i155	EERRRR	34	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i225	ERERRR	38	8
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i245	ERRERR	57	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i252	ERRRER	59	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i254	ERRRRE	35	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i425	REERRR	47	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i445	RERERR	50	4
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i452	RERRER	41	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i454	RERRRE	46	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i515	RREERR	46	5
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i522	RRERER	39	3
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i524	RRERRE	34	4
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i542	RRREER	32	4
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i544	RRRERE	31	4
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i551	RRRREE	29	5
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i125	EEERRR	74	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i145	EERERR	80	7
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i152	EERRER	75	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i154	EERRRE	56	8
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i215	EREERR	81	8
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i222	ERERER	79	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i224	ERERRE	57	3
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i242	ERREER	79	5
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i244	ERRERE	65	3
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i251	ERRREE	60	4
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i415	REEERR	73	4
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i422	REERER	74	5
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i424	REERRE	71	5
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i442	REREER	76	6
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i444	RERERE	67	5
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i451	RERREE	78	4
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i512	RREEER	74	7

7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i514	RREERE	72	11
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i521	RREREE	70	11
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i541	RRREEE	63	13
7	UCUCUAACUUUCAG/G	CAA/GUAAGU	110	i4u555	EEEEEE	96	0
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u255	ERRRRR	90	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u455	RERRRR	90	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u525	RRERRR	90	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u545	RRRERR	89	3
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u552	RRRRER	89	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u554	RRRRRE	86	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u111	EEEEEE	98	0
ESS Perturbation							
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u555	RRRRRR	49	3
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u855	SRRRRR	47	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u655	RSRRRR	38	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u585	RRSRRR	38	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u565	RRRSRR	38	9
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u558	RRRRSR	34	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u556	RRRRRS	30	5
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u955	SSRRRR	37	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u885	SRSRRR	35	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u865	SRRSRR	46	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u858	SRRRSR	38	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u856	SRRRRS	30	3
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u685	RSSRRR	35	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u665	RSRSRR	33	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u658	RSRRSR	34	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u656	RSRRRS	24	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u595	RRSSRR	32	3
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u588	RRSRSR	20	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u586	RRSRRS	22	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u568	RRRSSR	27	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u566	RRRSRS	19	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u559	RRRRSS	15	3
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u985	SSSRRR	27	5
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u965	SSRSRR	19	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u958	SSRRSR	18	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u956	SSRRRS	16	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u895	SRSSRR	18	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u888	SRSRSR	22	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u886	SRSRRS	13	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u868	SRRSSR	18	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u866	SRRSRS	17	6
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u859	SRRRSS	16	6
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u695	RSSSRR	16	5

5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u688	RSSRSR	15	5
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u686	RSSRRS	8	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u668	RSRSSR	16	5
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u666	RSRSRS	5	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u659	RSRRSS	8	1
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u598	RRSSSR	13	4
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u596	RRSSRS	11	3
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u589	RRSRSS	11	2
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u569	RRRSSS	14	3
5	UCUCUAUUUUUCAG/G	CAA/GUAAGU	110	i4u999	SSSSSS	5	0

Supplemental Table S2. *Oligomers used*

Oligomer	Sequence	Primary purposes
oligo1	CGCGCGACCTTCAGCATTG	Removal of BtgZI site on pUHD10-3
oligo2	CCGGCAATGCTGAAGGTCG	Removal of BtgZI site on pUHD10-3
oligo3	AAACGCGTACGGCGATGCCGCATGCAAGCTGTCATGGTTTGGATAGTTGG	Primer for fragment 5 of the modified dhfr minigene
oligo4	CCTCTAGATAACCTGGTTAGACTAATG	Primer for fragment 5 of the modified dhfr minigene
oligo5	CCGCATGCAAAGCCTTAAAACTTATTGAACAACC	Primer for fragment 4 of the modified dhfr minigene
oligo6	CCCCCCACCTGCAAAACAGCTCCACTTTATCTGCTAACTCTGG	Primer for fragment 4 of the modified dhfr minigene
oligo7	CCGCATGCAAAGCTCAGGTACTGGCTGGATTGGG	Primer for fragment 3 of the modified dhfr minigene
oligo8	CCCCCCACCTGCAAAAGGCTTCGTCCAGACTTTTGGCAAG	Primer for fragment 3 of the modified dhfr minigene
oligo9	CAAAGGGCATCGGCAAGAACGGAGACCTTCCCTGGCCAA	Primer for fragment 2 of the modified dhfr minigene
oligo10	AGCTTTGGCCAGGGAAGGTCTCCGTTCTTGCCGATGCCCTTTGCATG	Primer for fragment 2 of the modified dhfr minigene
oligo11	GTACGGTTCGACCGCTGAACTGCATCGTCGCCGTGTCCCAGAATA	Primer for fragment 1 of the modified dhfr minigene
oligo12	CCCTTATTCTGGGACACGGCGACGATGCAGTTCAGCGGTCGAACC	Primer for fragment 1 of the modified dhfr minigene
oligo13	GGCCGCCGCATGCAACATCGCACCTG	Addition of the universal RA (URA)
oligo14	CTAGCAGGTGCGATGTTGCATGCGGC	Addition of the universal RA (URA)
oligo15	CTAGGAAACAACACAAGTAAGTG	Addition of the 5’SS for SS Set 7 to pMA-URA
oligo16	CTAGCACTTACTTGTGTTGTTTC	Addition of the 5’SS for SS Set 7 to pMA-URA
oligo17	GGCCGCTGTTAACGCAGTGTTTCTCTAACTTTAAGCATG	Addition of the polypyrimidine tract for SS Set 7 to pMA-URA
oligo18	CTTAAAGTTAGAGAAACACTGCGTTAACAGC	Addition of the polypyrimidine tract for SS Set 7 to pMA-URA
oligo19	CTTTCAGGCCAAACGGGCATG	Addition of the 3’SS for SS Set 7 to pMA-URA
oligo20	CCCGTTTGGCCTG	Addition of the 3’SS for SS Set 7 to pMA-URA
oligo21	GGCCGCTGTTAACGCAGTGTTTCTCTTTTTTTAAGCATG	Addition of the polypyrimidine tract for SS Set 3 to pMA-URA
oligo22	CTTAAAAAAAGAGAAACACTGCGTTAACAGC	Addition of the polypyrimidine tract for SS Set 3 to pMA-URA
oligo23	CTAGGAAACAACACAAGTAAGTG	Addition of the 5’SS for SS Set 5 to pMA-URA
oligo24	CTAGCACTTACTTGTGTTGTTTC	Addition of the 5’SS for SS Set 5 to pMA-URA
oligo25	GGCCGCTGTTAACGCAGTGTTTCTCTATTTTTCAGGCCAAACGGGCATG	Addition of the polypyrimidine tract and 3’SS for SS Set 5 to pMA-URA
oligo26	CCCGTTTGGCttTGAAAAATAGAGAAACACTGCGTTAACAGC	Addition of the polypyrimidine tract and 3’SS for SS Set 5 to pMA-URA

oligo27	GGCCGCTGTTAACGCAGTGTTTCTCTAATTTTCAGGCCAAACGGGCATG	Addition of the polypyrimidine tract and 3’SS for SS Set 6 to pMA-URA
oligo28	CCCGTTTGGCCTGAAAATTAGAGAAACACTGCGTTAACAGC	Addition of the polypyrimidine tract and 3’SS for SS Set 6 to pMA-URA
oligo29	GCCAACTACTTAGGGACAGT	Common primer to transfer DEs
oligo30	CACTGGCCAGCTAGCACTTACCTGTGTTGTTTG	Primer to transfer DEs while adding a consensus 5’SS
oligo31	CACTGGCCAGCTAGCACTCACTTGTGTTGTTTG	Primer to transfer DEs while adding a wild-type 5’SS
oligo32	GGCCGCTGTTAACGCAGTGTTTCTCTATTTTTCAGGCCAAACAGGGCGCAGGTGCATGCACCTGCTAGGAAACAACACAAGTAAGTG	Generation of the receiving plasmid with SS Set 5
oligo33	CTAGCACTTACTTGTGTTGTTTCCTAGCAGGTGCATGCACCTGCGCCCTGTTTGGCCTGAAAAATAGAGAAACACTGCGTTAACAGC	Generation of the receiving plasmid with SS Set 5
oligo34	GGCCGCTGTTAACGCAGTGTTTCTCTTTTTTTCAGGCCAAACAGGGCGCAGGTGCATG	Generation of the receiving plasmid with SS Set 3
oligo35	CACCTGCGCCCTGTTTGGCCTGAAAAAAAGAGAAACACTGCGTTAACAGC	Generation of the receiving plasmid with SS Set 3
oligo36	AAAAAAGGTCTCGGATTGCACGCTGGTTCTCCGGCCGCTTGGGT	Primer from set 1 in nested PCR to remove BfuAI sites from EGFP-C3 plasmid
oligo37	TCGAATGGGCACGTAGCCGGATCAAGCGTATGCA	Primer from set 1 in nested PCR to remove BfuAI sites from EGFP-C3 plasmid
oligo38	TGATCCGGCTACGTGCCCATTCGACCACCAAGCGAAACA	Primer from set 2 in nested PCR to remove BfuAI sites from EGFP-C3 plasmid
oligo39	AAAAAAGGTCTCGTCGTGATGCCAGGTTGGGCGTCGCTTGGT	Primer from set 2 in nested PCR to remove BfuAI sites from EGFP-C3 plasmid
oligo40	AAAAAAGGTCTCCCCACCCAGACCCCATTGGGGCCAATA	Primer for PCR to remove the BsaI site from EGFP-C3 plasmid
oligo41	TATGGCAGGGCCTGCCGCCCCGA	Primer for PCR to remove the BsaI site from EGFP-C3 plasmid
oligo42	GCAAAGACCCCAACGAGAAGCGCGA	Addition of the linker to generate the pAL-SB plasmid
oligo43	CCCCCTGCAGCAGCCGTCTCCAAACAGAGACCAGCTGCAAGCTTGAGCTCGAGATCTGAGTA	Addition of the linker to generate the pAL-SB plasmid
oligo44	CCCCCCATTAATCCCCCCTCGAGCCACCATGACCGAGTACAAGCCCA	Amplification of the promoterless puromycin gene
oligo45	GGGGATCCTCAGGCACCGGGCTTGCGGGT	Amplification of the promoterless puromycin gene
oligo46	TCGAGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG	Incorporation of the attP site to generate pMA-FW
oligo47	TCGACCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGC	Incorporation of the attP site to generate pMA-FW
oligo48	CCCTCGAGTGTTGCTCCAGCATGATGTTGT	Amplification and transfer of the second half of the modified dhfr minigene to generate pMA-FW

oligo49	GGGGGGATTAATAGACGACGAGGCTTGCAGGATCAT	Amplification and transfer of the second half of the modified dhfr minigene to generate pMA-FW
oligo50	CCCCCCTCTAGATCATAGCCCATATATGGAGTTCCGCGT	Amplification and transfer of the CMV promoter
oligo51	TGAATTCCGGATCTGACGGTTCACTAAACCA	Amplification and transfer of the CMV promoter
oligo52	AATTCGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACC	Incorporation of the attB site to generate pMA-IC
oligo53	AATTGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCG	Incorporation of the attB site to generate pMA-IC
oligo54	CATGGTAATAGCCATGACTAATAC	Removal of BtgZI site to generate pMA-IC plasmid
oligo55	GTATTAGTCATGGCTATTAC	Removal of BtgZI site to generate pMA-IC plasmid
oligo56	CCCGTACGGTTCGACCGCTGAACTGCATCG	Amplification of cDNA to generate the standard plasmids
oligo57	CATGGACGAATTCCCCAAAGCGGCCGCAA	Addition of an adapter to generate the coupled-standard plasmids
oligo58	CTAGTTGCGGCCGCTTTGGGGAATTCGTC	Addition of an adapter to generate the coupled-standard plasmids
oligo59	CCCCCGCGGCCGCAAAGATCCAGCCTCCGCGTA	Amplification of cDNA to generate the coupled-standard plasmids
oligo60	CCCCCGAATTCAAAACACAAGTCCCATGGTCTTGTA	Amplification of cDNA to generate the coupled-standard plasmids
oligo61	GCATTTGCGGTGGACG	Reverse transcription primer for Gamma Actin
oligo62	AAACCGCGGCCGCTCGTGCGTGACATTAAGGAGA	Amplification of Gamma Actin cDNA for coupled-standard generation
oligo63	AAACCGAATTCGCATTTGCGGTGGACG	Amplification of Gamma Actin cDNA for coupled-standard generation
oligo64	AAACAACCAAACAACCAAACAACCAAACAACC	Generation of a building block containing two reference sequences (RR)
oligo65	GTTTGGTTGTTTGGTTGTTTGGTTGTTTGGTT	Generation of a building block containing two reference sequences (RR)
oligo66	AAACAATCCTCGAACCAAACAATCCTCGAACC	Generation of a building block containing two enhancer sequences (EE)
oligo67	GTTTGGTTCGAGGATTGTTTGGTTCGAGGATT	Generation of a building block containing two enhancer sequences (EE)
oligo68	AAACAATCCTCGAACCAAACAACCAAACAACC	Generation of a building block containing an enhancer and a reference sequence (ER)
oligo69	GTTTGGTTGTTTGGTTGTTTGGTTCGAGGATT	Generation of a building block containing an enhancer and a reference sequence (ER)
oligo70	AAACAACCAAACAACCAAACAATCCTCGAACC	Generation of a building block containing a reference and an enhancer sequence (RE)
oligo71	GTTTGGTTCGAGGATTGTTTGGTTGTTTGGTT	Generation of a building block containing a reference and an enhancer sequence (RE)
oligo72	AAACAACACATGGTCCAAACAACACATGGTCC	Generation of a building block containing two silencer sequences (SS)

oligo73	GTTTGGACCATGTGTTGTTTGGACCATGTGTT	Generation of a building block containing two silencer sequences (SS)
oligo74	AAACAACACATGGTCCAAACAACCAAACAACC	Generation of a building block containing a silencer and a reference sequence (SR)
oligo75	GTTTGGTTGTTTGGTTGTTTGGACCATGTGTT	Generation of a building block containing a silencer and a reference sequence (SR)
oligo76	AAACAACCAAACAACCAAACAACACATGGTCC	Generation of a building block containing a reference and a silencer sequence (RS)
oligo77	GTTTGGACCATGTGTTGTTTGGTTGTTTGGTT	Generation of a building block containing a reference and a silencer sequence (RS)
oligo78	GCGGTACCGTCGACTTCAGCAGCCGT	Transfer of the finished DEs to a receiving plasmid
oligo79	AAACAACCAAACAACACAGGTAAGTG	Generation of 22nt DEs: SS Set 3
oligo80	CTAGCACTTACCTGTGTTGTTTGGTT	Generation of 22nt DEs: SS Set 3
oligo81	AAACAACCAAACAACACAAGTAAGTG	Generation of 22nt DEs: SS Set 7
oligo82	CTAGCACTTACTTGTGTTGTTTGGTT	Generation of 22nt DEs: SS Set 7
oligo83	GGAACTGCCTCCAACTATCCAA	Transfer of DEs to pMA-IC for stable transfections and shared QPCR reverse primer for the modified dhfr minigene
oligo84	AGAGTCTGAGATGGCCTGGCT	Reverse transcription primer for the modified dhfr minigene
oligo85	CAAACAACACAAGGAACCACCA	QPCR forward primer for molecules including DEs with wild type 5’SS
oligo86	GCCAAAGCTCAGGGAACCA	QPCR forward primer for molecules skipping the DEs
oligo87	CAAACAACACAGGGAACCACC	QPCR forward primer for molecules including DEs with consensus 5’SS
oligo88	TCATGGTGGTTCGACCCCCAA	Primer for verification of disruption of attP sites
oligo89	AAAAGCGGCCGCGTGCCTGAGGATCGGATCTA	Primer for the second PCR amplification in the nested PCR used to map the genomic incorporation of pMA-FW
oligo90	AAAAGCGGCCGCGGCGGTAATACGGTTATCCA	Primer for the second PCR amplification in the nested PCR used to map the genomic incorporation of pMA-FW
oligo91	TCATCTTTACATAATTGTCATGGCAT	Primer for verification of the genomic location of pMA-FW
oligo92	CAACACTCAACCCTATCTCGGTCTA	Primer for verification of the genomic location of pMA-FW
oligo93	CTGAAGTGAACATTTCCAAGTAAGAA	Primer for verification of the genomic location of pMA-FW
oligo94	GCTCTAAAGAAGGTTCTGCTCCAT	Primer for detection of hybrid mRNA molecules

Supplemental Table S3. *Values of coefficients for the recruitment model*

T	C	K_2	K_3	K_5	K_7	c_E	c_R	c_{SF}	c_{SL}	c_{SI}	y_2	y_3
5.24	4.60 x 10^{-13}	6.44 x 10^{-3}	1.90 x 10^{-2}	3.49 x 10^{-2}	0.946	5.56	-4.53 x 10^{-2}	-6.76 x 10^{-2}	-0.242	-0.175	15.5	8.73

Supplemental Figures

Fig. S1.

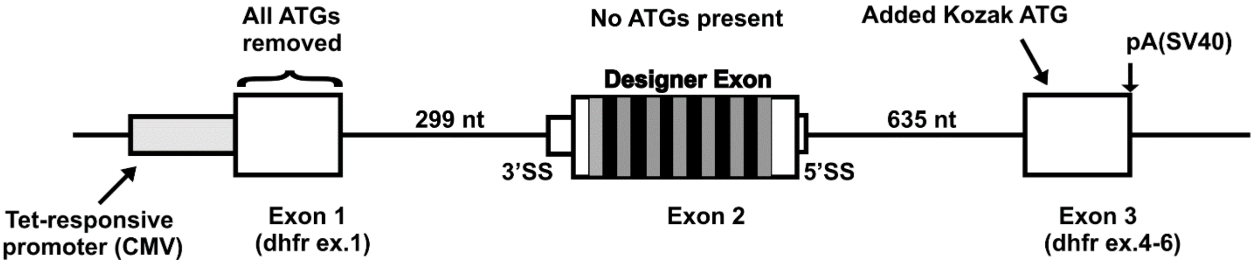

Figure S1. Cartoon of a typical DE-containing minigene.

Fig. S2.

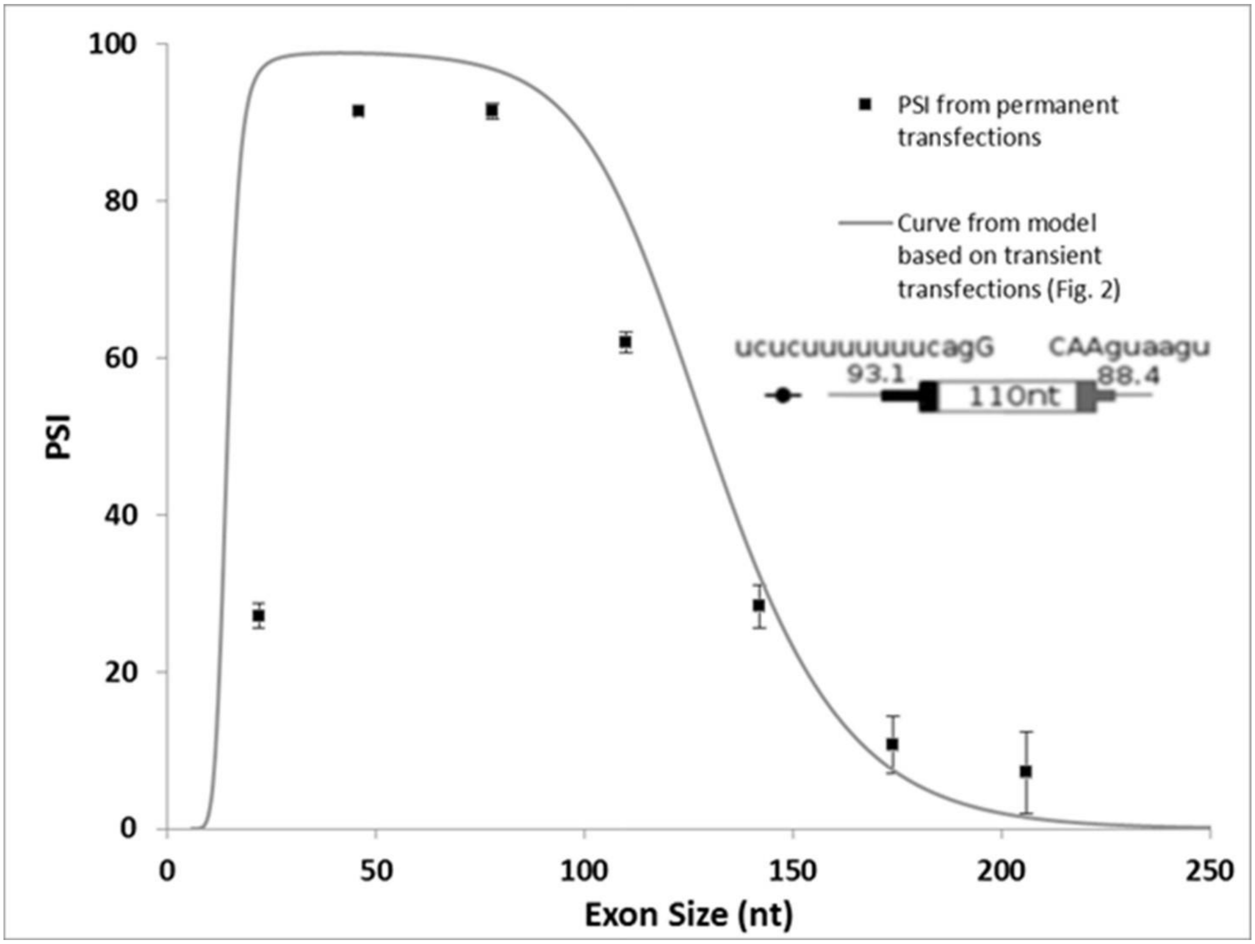

Figure S2. Exon inclusion also exhibits an optimum size range at a chromosomal location. Points: Inclusion levels (psi) of DEs of various sizes in a chromosomal context. DEs consist of reference sequences and have a strong 3'SS (SS set 3). Error bars: SEM, n≥3. Curve: Predicted by the model based on transient transfection data, taken from Fig. 2.

Fig. S3.

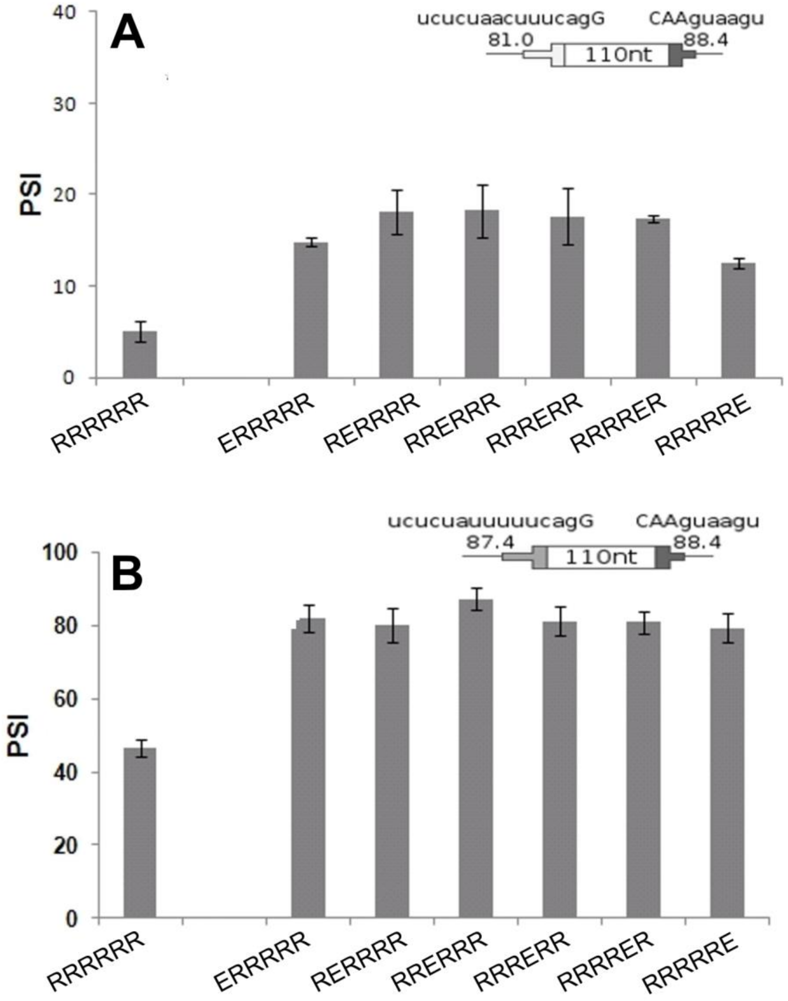

Figure S3. Addition of a single ESE enhances inclusion level and is position independent in a chromosomal context. The cartoons show the consensus values for splice site strengths used. Error bars: SEM, $n \geq 3$. In all panels the psi of DEs with an ESE are significantly different from that without an ESE (t-test, $p < 0.01$).

Fig. S4.

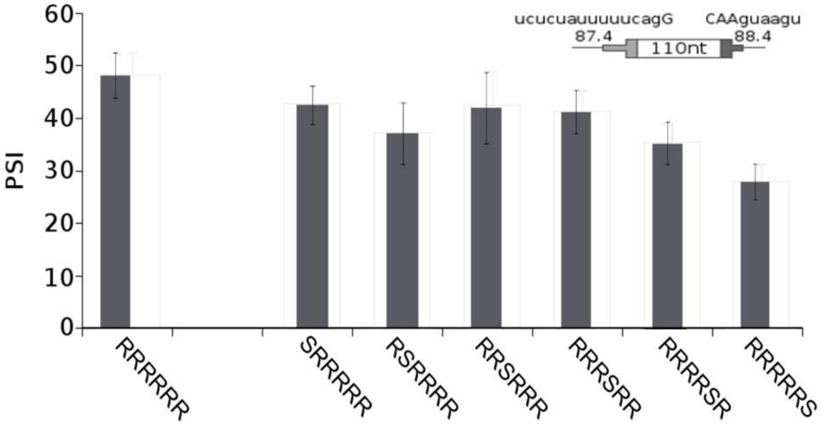

Figure S4. Addition of a single ESS decreases inclusion level and shows some position dependence in a chromosomal context. The psi for DEs with a single ESS are shown for stable transfections. Error bars: SEM, n≥3.

Fig. S5.

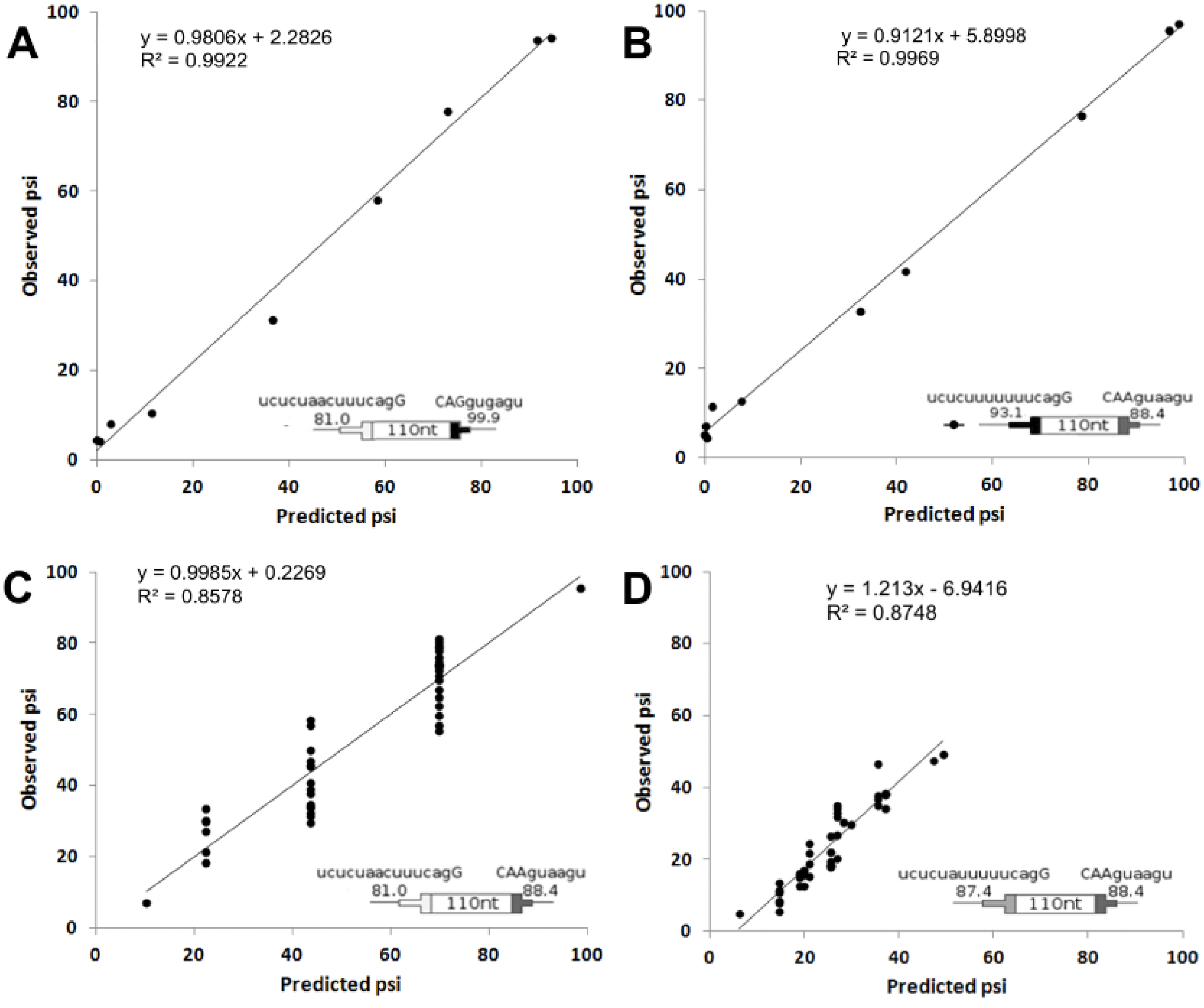

Figure S5. The model accurately predicts the inclusion levels of DEs for each parameter examined. The values in Table 3 were used to predict the psi; these values were optimized using a condensed and abridged version of the data presented here. A. Exons of different sizes using SS Set 2. B. Exons of different sizes using SS Set 3. C. Exons with 0, 1, 2, 3 or 6 ESEs in all positional combinations. SS Set 7 was used. D. Exons with 0, 1, 2, 3 or 6 ESSs in all positional combinations. SS Set 5 was used.

Fig. S6.

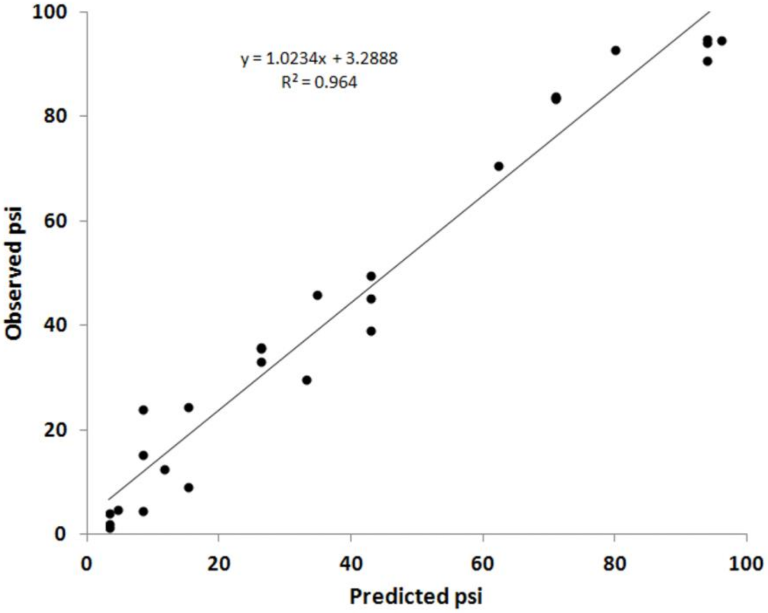

Figure S6. The combinations of ESEs and ESSs are accurately modeled in 110 nt exons in the complex designer exon set. The predictions of complex DEs of 110 nt were assessed separately to evaluate the performance of the model for combining ESEs and ESSs while removing the effect of size.

Fig. S7.

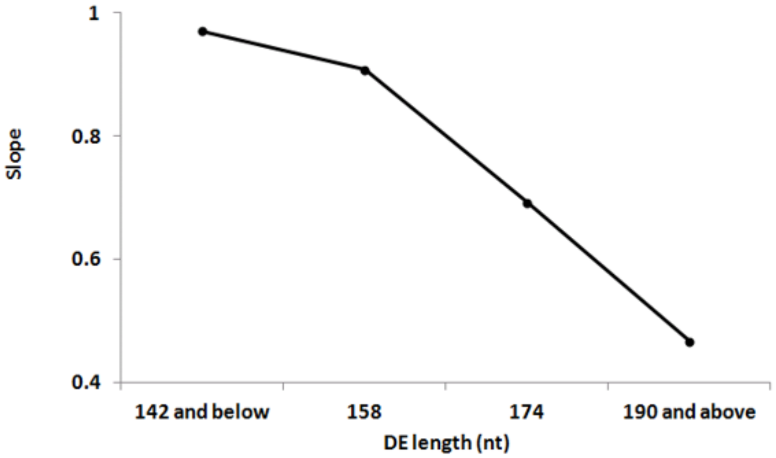

Figure S7. The observed psi for complex DEs progressively falls short of prediction as sizes increase above 142 nt. A best fit was performed comparing observed vs. predicted psi for complex DEs of the indicated sizes. Although R^2 values were high in all (not shown), accuracy, as reflected in the slope of the fit, decreased with size. This artifact would be expected if there is a decrease in the efficiency of PCR with size for the included (but not the skipped) mRNAs in the end point PCR measurements that were used for the complex DEs.

Fig. S8

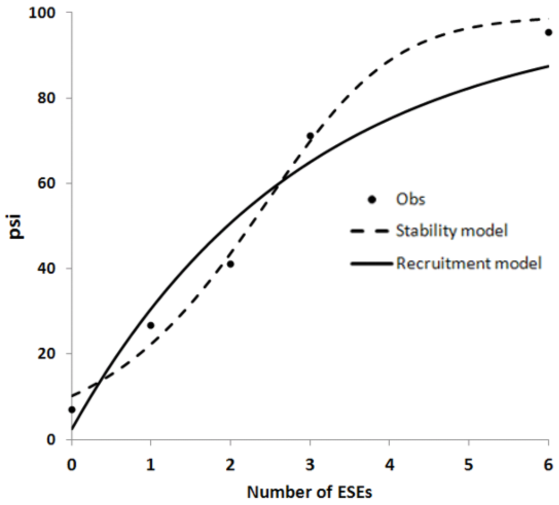

Figure S8. The stability and the recruitment models yield different curves for psi dependence on ESE number. Adding ESEs generates a sigmoidal curve according to the stability model and generates a negative exponential curve according to the recruitment model. The experimental observations follow the stability model more closely.

Fig. S9.

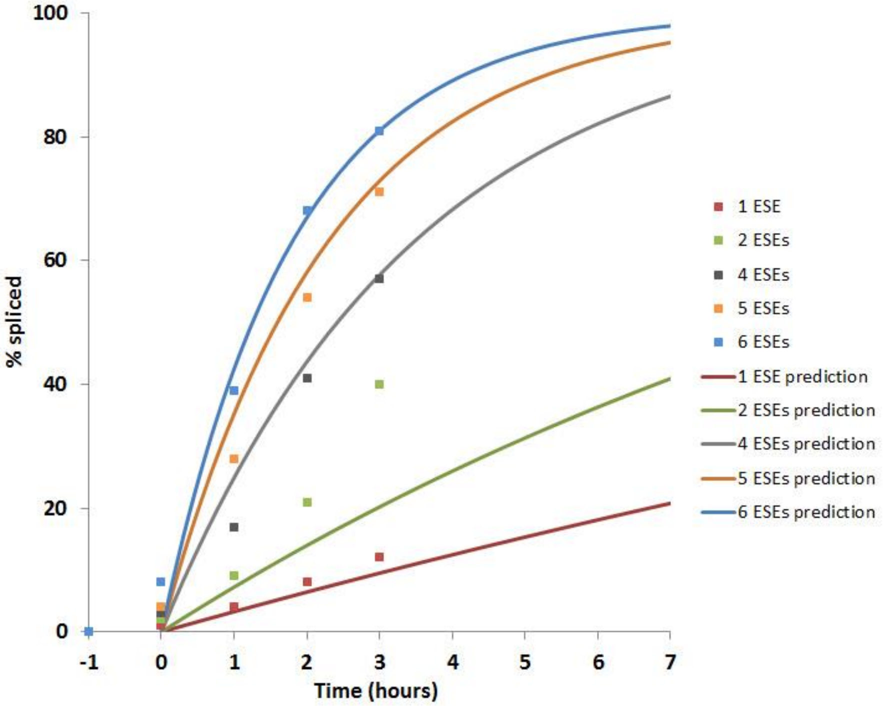

Figure S9. A stabilization model can explain *in vitro* splicing kinetic measurements. Comparison of cell-free splicing time course experiments (Hertel and Maniatis 1998) showing the effect of increasing ESE numbers with time course predictions using the model described in the text. The observed data (points) were extracted from Fig. 2D of (Hertel and Maniatis 1998) and plotted assuming a splicing delay of 1 hr. The coefficients in Table 3 were used for the model (curves) with the exception of T, which was set to a chosen constant times time to account for the slower kinetics of the *in vitro* reactions. The points and the curve corresponding to 1 ESE are shown in red, 2 ESEs in green, 4 ESEs in gray, 5 ESEs in orange, and 6 ESEs in blue.

Fig. S10.

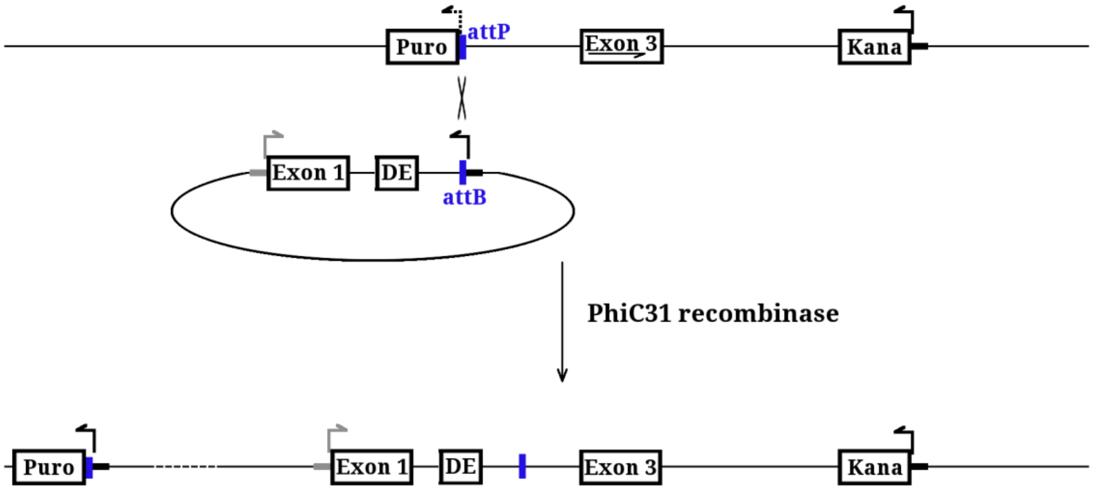

Figure S10. Site-specific recombination reconstitutes the minigene in a specific location of the genome. Using the kanamycin resistance gene (Kana) through selection with G418, an attP site has been incorporated in the genome of HEK 293 cells along with the downstream half of the modified dhfr minigene and a promoterless copy of a gene conferring puromycin resistance (Puro). After transient transfection with a plasmid incorporating the upstream half of the minigene as well as a promoter for the puromycin resistance gene, along with a gene for PhiC31 recombinase, puromycin-resistant site-specific recombinants can be isolated that have reconstituted the minigene as well as the puromycin resistance gene. Exons are indicated with boxes while introns and intergenic regions are indicated by thin lines. The promoters are indicated with thick horizontal lines: gray for the minigene and black for the puromycin and G418 resistance genes. The direction of transcription is indicated by bent arrows; the dashed arrow indicates the nominal direction of transcription for the promoterless puromycin resistance gene. For exon 3 of the minigene, the nominal direction of transcription is indicated with a horizontal arrow. The PhiC31 recombination sites are indicated by blue vertical lines.

Fig. S11.

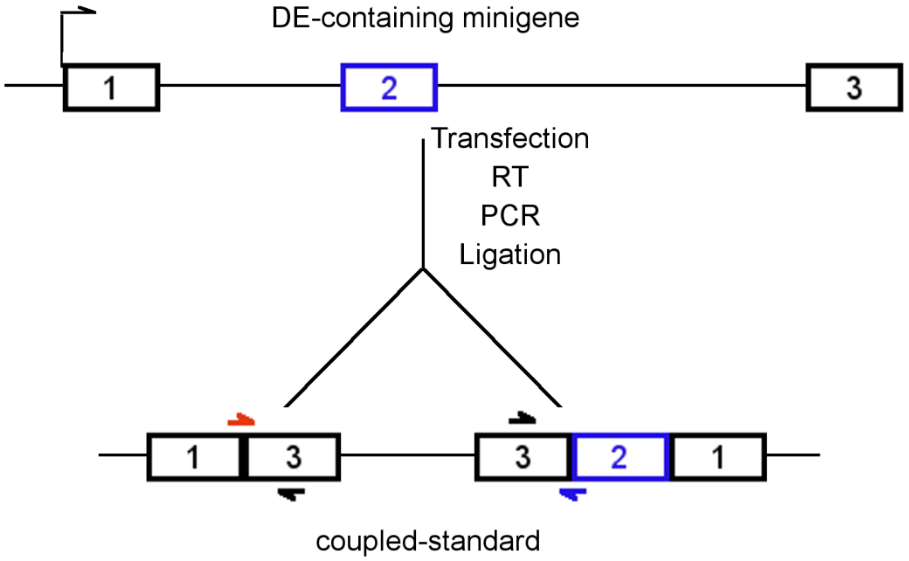

Figure S11. Coupled-standards incorporate two cDNAs into a single molecule. A reverse transcribed copy of the mRNA with the DE spliced in (included) as well as a copy without it (skipped) have been incorporated into the same plasmid molecule by sequential ligations. Digestions of this plasmid are therefore guaranteed to have equimolar amounts of both species. A dilution series of these molecules was used as a standard in QPCR reactions. The primers used for QPCR of the standards and the experimental samples are indicated with arrows: black, shared primer; blue, joint primer for detection of included molecules; and red, joint primer for detection of skipped molecules. See Detailed Materials and Methods above for details.

Fig. S12.

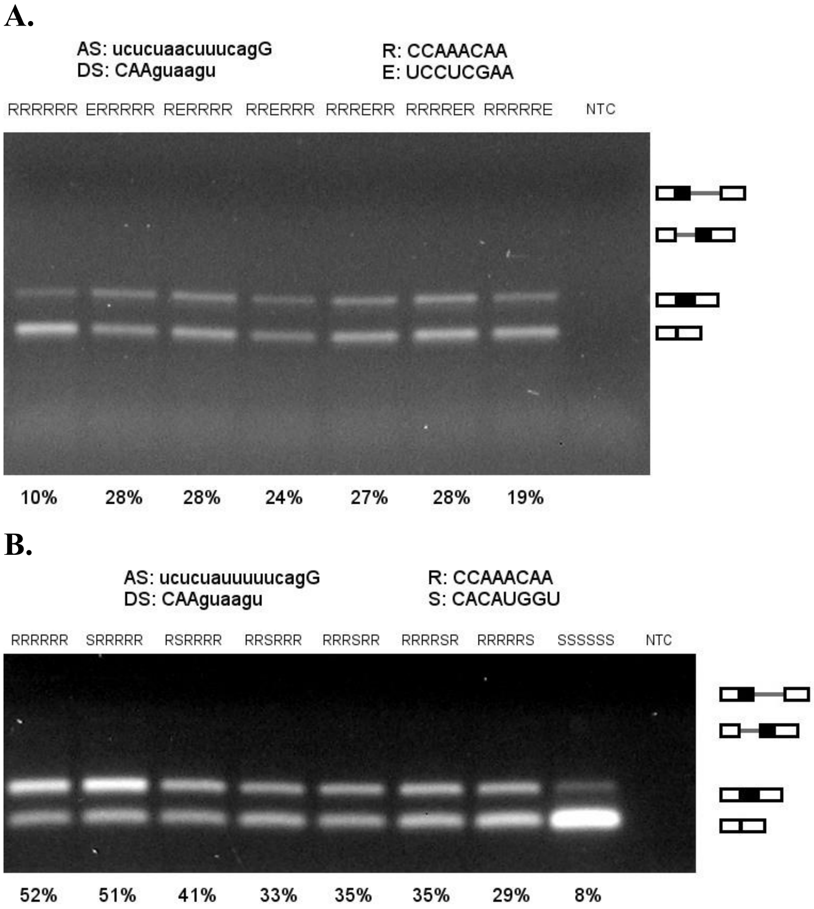

Figure S12. End point RT-PCR results. The indicated RNA samples from the experiments shown in Figures 3 and 5 (A and B respectively) were subjected to end point RT-PCR for 21-22 cycles. The products were separated by agarose gel electrophoresis and stained with ethidium bromide. The bands were quantified using Image J software. AS, acceptor site sequence; DS, donor site; R, Reference Sequence; E, ESE; S, ESS; NTC, no RNA template control.

Fig. S13.

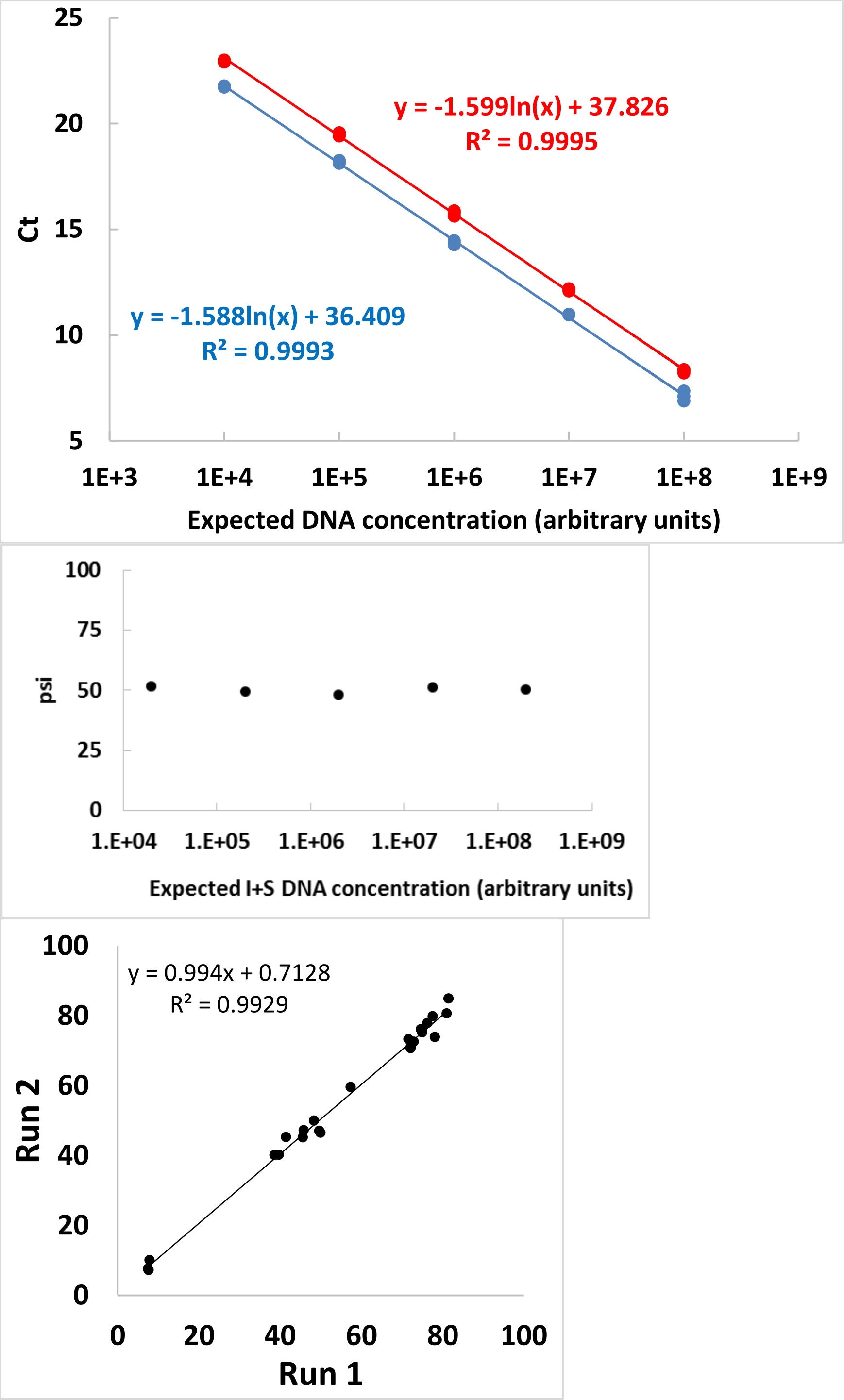

Figure S13. QPCR measurements. Upper: Detection of both the included and the skipped molecules is precise. The expected concentration of both included (blue) and skipped (red) molecules in the standard varies linearly with the Ct at which detection occurs. Middle: The measured ratio of included (I) to skipped (S) in the coupled-standard is independent of DNA concentration used. Lower: Agreement between two independent QPCR runs of the same 24 samples.

Fig. S14.

Figure S14. The recruitment model fails to accurately predict the effect of combining ESEs and ESSs in 110 nt exons. The predictions of complex DEs of 110 nt were assessed separately to evaluate the performance of the model for combining ESEs and ESSs while removing the effect of size. Although a high R^2 was achieved the slope and intercept deviate markedly from the expected values of 1 and 0%, respectively.

Fig. S15.

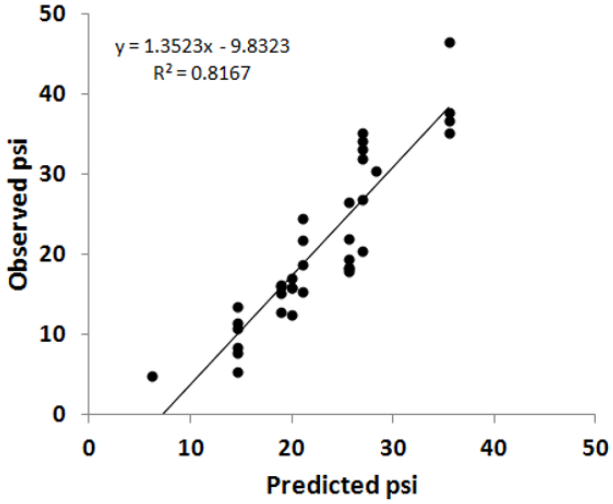

Figure S15. The psi of DEs with multiple ESSs as predicted by the model. The values in Table 3 were used to predict the psi for all constructs with multiple ESSs. These values were obtained using only single-ESS DEs. The predictions show a good level of correlation: $R^2 = 0.82$. Although the R^2 is high, the slope and intercept deviate somewhat from what is expected (1 and 0%, respectively) suggesting that the model could be further refined.

Appendix

The following sections describe a detailed derivation of equation 5 (also known as equation S23). The analysis presented here has as its goal a clear presentation of the derivation to put in evidence its reasonableness. Therefore, some assumptions have been made slightly more rigorous than they need to be; this has been noted in the text below. Additionally, the quality of the approximations was not evaluated rigorously. A more thorough analysis to show that the assumptions used in the main text are sufficient would be lengthier, more complex and would not add any new insights. We have therefore not included it here.

Exact solutions and approximate solutions for both time periods were obtained. The exact solution presented for the second time period incorporated the exact solution for the first time period. Due to the complexity of the expressions involved, the approximate solution (equation 5) was used to build the model.

Notes: All the rate constants (a, a', d and d') are expected to be positive. The equations corresponding to those in the Supplemental Material are referenced using **bold characters**.

First time period: $t \leq \tau$

We let L(t) be the number of uncomplexed (naked) pre-mRNA molecules, P(t) be the number of molecules in an exon definition complex, and I(t) be the number of molecules committed to inclusion. A set of differential equations relates the number of tagged P, I and L molecules starting at t = 0:

A1. $dL/dt = d\,P - a\,L$

A2. $dP/dt = a\,L - (d+\rho_I)\,P$

A3. $dI/dt = \rho_I\,P$

These equations are a reiteration of **equations S1-S3** (also **equations 2-4**).

Let's define F as the number of uncommitted molecules, F = L + P.

The main tool used to solve this set of differential equations is the Laplace transform: indicated with italic letters as $X = X(s)$ for any function X(t). Taking the Laplace transform for equations A1 and A2, we get

A4. $(s+a)\,F = (s+d+a)\,P + F_0 - P_0$

A5. $(s+a+d+\rho_I)\,P = a\,F + P_0$

where F_0 and P_0 represent initial values for F(t) and P(t) respectively.

Solving for *F*,

A6. $[s^2 + (a+d+\rho_I)\,s + a\rho_I]\,F = (s+a+d+\rho_I)\,F_0 - \rho_I\,P_0$

Factoring the second degree polynomial above to allow the use of partial fractions, we obtain

A7. $F = [(r_2\,F_0 - \rho_I\,P_0) / (s - r_1) - (r_1\,F_0 - \rho_I\,P_0) / (s - r_2)] / (r_2 - r_1)$

with r_1 & r_2 ($r_1 \geq r_2$) being the roots of the quadratic equation

A8. $s^2 + (a+d+\rho_I)\,s + a\rho_I = 0$

A9. $r_1 = \frac{-(a+d+\rho I)+\sqrt{(a+d+\rho I)^2-4a\rho I}}{2}$

A10. $r_2 = \frac{-(a+d+\rho I)-\sqrt{(a+d+\rho I)^2-4a\rho I}}{2}$

Notice that $a+d+\rho_I$ is greater than $a+\rho_I$, which is itself greater than 0, and that $a\rho_I > 0$. Therefore the discriminant is greater than $(a-\rho_I)^2$, which is non-negative: $(a+d+\rho_I)^2 - 4\,a\rho_I > (a+\rho_I)^2 - 4\,a\rho_I = (a-\rho_I)^2 \geq 0$. Hence, both roots are real and since the square root of the discriminant is smaller than $a+d+\rho_I$, both roots are negative. Hence, r_1 is closer to zero than r_2 is, or they are both equal. Equation A8 uses s to denote the variable and it is equivalent to **equation S5**, which uses x to denote the variable.

Taking the inverse Laplace transform of equation A7, we obtain

A11. $F(t) = [(r_2\,F_0 - \rho_I\,P_0)\,e^{r1\,t} - (r_1\,F_0 - \rho_I\,P_0)\,e^{r2\,t}] / (r_2 - r_1)$

This is **equation S4**.

Now, solving for P in an analogous manner as A6 to A7,

A12. $[s^2 + (a+d+\rho_I)\, s + a\rho_I]\, P = a\, F_0 + s\, P_0$

and

A13. $P = [-(r_1\, P_0 + r\, F_0) / (s - r_1) + (r_2\, P_0 + r\, F_0) / (s - r_2)] / (\, r_2 - r_1\,)$

Taking the inverse Laplace Transform yields

A14. $P(t) = [-(r_1\, P_0 + r\, F_0)\, e^{r1\; t} + (r_2\, P_0 + r\, F_0)\, e^{r2\; t}] / (\, r_2 - r_1\,)$

This is **equation S6**.

These equations describe the kinetics of this system up to time τ. Importantly, they also set up the initial conditions for the next time period, t > τ, where competition between two fates, I and S, occurs. Evaluating these equations at time τ to obtain these initial conditions and noting that $I(t) = L_0 - F(t)$ where L_0 is the initial value for L(t), we get

A15. $F_\tau = [(r_2\, F_0 - \rho_I\, P_0)\, e^{r1\; \tau} - (r_1\, F_0 - \rho_I\, P_0)\, e^{r2\; \tau}] / (r_2 - r_1)$

A16. $P_\tau = [-\, (a\, F_0 + r_1\, P_0)\, e^{r1\; \tau} + (a\, F_0 + r_2\, P_0)\, e^{r2\; \tau}] / (r_2 - r_1)$

A17. $I_\tau = L_0 - F_\tau$

where the notation X_τ represents X(t) at time τ. These equations are **equations S7-S9** and represent the exact solution for the first time period.

At the beginning of the observation period no complexes have formed, so $P_0 = 0$ and $F_0 = L_0$. If we **assume** that the assembly or the dissociation of the complex occurs much faster than commitment, so that $a+d \gg \rho_I$, r_1 can be simplified as follows:

A18. $r_1 = \frac{-(a+d+\rho I)+\sqrt{(a+d+\rho I)^2-4a\rho I}}{2}$

A19. $r_1 = \frac{-(a+d+\rho I)+(a+d+\rho I)\sqrt{1-4a\rho I/(a+d+\rho I)^2}}{2}$

A20. $r_1 = \frac{-(a+d+\rho I)}{2}[1-\sqrt{1-\frac{4a\rho I}{(a+d+\rho I)^2}}]$

Taking $w = \frac{4a\rho I}{(a+d+\rho I)^2}$

A21. $r_1 = \frac{-(a+d+\rho I)}{2}[1-\sqrt{1-w}]$

If we further assume that $a+d+\rho_I \gg 4\rho_I$, then $(a+d+\rho_I)^2 \gg 4\rho_I\, (a+d+\rho_I)$ and, therefore, $(a+d+\rho_I)^2 \gg 4a\rho_I$. From these considerations $1 \gg w$, making the first order Taylor polynomial a good approximation, so $\sqrt{1-w} \approx 1-\frac{w}{2}$, and we get

A22. $r_1 \approx \frac{-(a+d+\rho I)}{4}\, w$

Generating

A23. $r_1 \approx \frac{-(a+d+\rho I)}{4} \frac{4a\rho I}{(a+d+\rho I)^2}$

A24. $r_1 \approx \frac{-a\rho I}{a+d+\rho I}$

Since a+d >> ρ_I, we can approximate the denominator as a+d and we obtain

A25. $r_1 \approx \frac{-a\rho I}{a+d}$

This is **equation S11**.

Actually, the assumption that a+d+ρ_I >> 4ρ_I is not required in order to derive A25; the more conservative assumption that a+d >> ρ_I suffices, but that more complex derivation has been omitted here for clarity.

Rather than performing the analogous derivation for r_2, we used the approximation for r_1 to obtain an approximation for r_2. The product of the roots generates the constant term in the quadratic equation (equation A8), i.e., aρ_I. Therefore

A26. $r_1\, r_2 \approx a\, \rho_I$

A27. $r_2 \approx a\, \rho_I / r_1$

using equation A25 this yields

A28. $r_2 \approx -(a+d)$

This is **equation S10**.

Considering equations A25 and A28, notice that $|r1| < \rho_I << a+d \approx |r_2|$, so $|r_1| << |r_2|$ which makes $r_2 - r_1 \approx r_2$ or

A29. $r_2 - r_1 \approx -(a+d)$

This is **equation S12**.

Going back to F_τ, we note that there are two contributing exponential terms in equation A15. One decays by r_1t while the other decays by r_2t. However, we showed that $|r_2| >> |r_1|$. Therefore, the e^{r2t} term decays vastly faster than the e^{r1t} term, causing the former to be a negligible contributor after a relatively short time. Therefore we disregard the $e^{r2\, t}$ term. With these results and defining p_I as

A30. $p_I = \rho_I / (1+d/a)$

we get

A31. $r_1 \approx \frac{-\rho I}{1+\frac{d}{a}}$

A32. $r_1 \approx -p_I$

It is p_I that is used in the main text rather than r_1.

So equation A15 reduces to

A33. $F_\tau \approx (r_2\, F_0 - \rho_I\, P_0)\, e^{-pI\, \tau} / (r_2 - r_1)$

Since P_0 is 0 and so $L_0 = F_0$, we get

A34. $F_\tau \approx r_2\, L_0\, e^{-pI\, \tau} / (r_2 - r_1)$

and since $|r_1| << |r_2|$

A35. $F_\tau \approx L_0\, e^{-pI\, \tau}$

This is **equation S14**.

Second time period: $t > \tau$

For times starting at time τ the molecules can consider splicing the upstream exon to the downstream exon, i.e., skipping the exon of interest (Fig. 7C and E). To minimize the complexity of notation below, we define a new reference time t' that sets time τ to zero: $t' = t - \tau$. From the state diagram shown in Fig. 7C, the following equations, a reiteration of **equations S15-S20**, are obtained for $t' > 0$:

A36. $dL/dt' = d\,P + d'\,b - (a+a')\,L$

A37. $dP/dt' = a\,L + d'\,B - (d+a'+\rho I)\,P$

A38. $db/dt' = a'\,L + d\,B - (d'+a+\rho S)\,b$

A39. $dB/dt' = a'\,P + a\,b - (d+d'+\rho I+\rho S)\,P$

A40. $dI/dt' = \rho I\,(P+B)$

A41. $dS/dt' = \rho_S\,(b+B)$

where S represents molecules committed to skipping (i.e., the joining of exon 1 to exon 3), ρ_S is the rate at which complexed molecules commit to the skipped pathway, B represents molecules with both exons in EDCs, b represents molecules with a downstream exon in an EDC but with the exon of interest not in an EDC, and a' and d' are the association and dissociation constants, respectively, for the formation of b. Note that the characters P, b, and B can be viewed as schematics for molecules with and EDC for exon 2, exon 3, or both, respectively.

The Laplace transform of the first four equations, indicated by italics as $X = X(s)$ for any function X(t'), was taken yielding

A42. $sL - L\tau = d\,P + d'\,b - (a+a')\,L$

A43. $sP - P\tau = a\,L + d'\,B - (d+a'+\rho_I)\,P$

A44. $sb - b\tau = a'\,L + d\,B - (d'+a+\rho_S)\,b$

A45. $sB - B\tau = a'\,P + a\,b - (d+d'+\rho_I+\rho_S)\,B$

where the notation X_τ represents X(t') at $t' = 0$ (i.e., at $t = \tau$).

Although we are most interested in the probability of exon inclusion, it is easier to calculate S, and its final expression actually provides more insight into the roles of the different parameters. I becomes simply all the tagged molecules not included in S. Therefore we will focus on an expression for S_∞. Let's define $\text{Б} = b+B$. The value of $S(t') = 0$ for $t' \leq 0$ if commitment to skipping requires the presence a downstream exon. According to the final value theorem and equation A41, as $t' \rightarrow \infty$, S(t') approaches $S_\infty = \rho_S \lim_{s\rightarrow 0} \textit{Б}(s) = \rho_S\,\textit{Б}_0$, where the notation X_0 represents X(s) evaluated at s = 0. Substituting $L = F - P$ and $b = \text{Б} - B$, and since no tagged molecules contain the second complex at $t' = 0$, $b_\tau = B_\tau = \text{Б}_\tau = 0$, we obtain

A46. $s\,(F - P) - (F\tau - P\tau) = d\,P + d'\,(\textit{Б} - B) - (a+a')\,(F - P)$

A47. $sP - P\tau = a\,(F - P) + d'\,B - (d+a'+\rho_I)\,P$

A48. $s\,(\textit{Б} - B) = a'\,(F - P) + d\,B - (d'+a+\rho_S)\,(\textit{Б} - B)$

A49. $sB = a'P + a\,(\textit{Б} - B) - (d+d'+\rho_I+\rho_S)\,B$

Taking s = 0, these equations become

A50. $(a+a')\,F_0 = (d+a+a')\,P_0 + d'\,\textit{Б}_0 - d'\,B_0 + F_\tau - P_\tau$

A51. $(d+a+a'+\rho I)\,P_0 = a\,F_0 + d'\,B_0 + P\tau$

A52. $(d'+a+\rho S)\,\textit{Б}_0 = a'\,F_0 + (d+d'+a+\rho_S)\,B_0 - a'\,P_0$

A53. $(d+d'+a+\rho_I+\rho_S)\,B_0 = a'\,P_0 + a\,\textit{Б}_0$

Substituting F_0 from equation A51 into equations A50 and A52, we get

A54. $[a'\,(d+a+a'+\rho_I) + a\,\rho I]\,P_0 = d'\,a\,\textit{Б}_0 + d'\,a'\,B_0 + a\,F\tau + a'\,P\tau$

A55. $(d'+a+\rho S)\,a\,\textit{Б}_0 = (d+a'+\rho_I)\,a'\,P_0 + [(d+d'+a+\rho_S)\,a - d'\,a']\,B_0 - a'\,P_\tau$

Substituting P_0 from equation A53 into these equations, they become

A56. $\{[a'\,(d+a+a'+\rho_I) + a\,\rho_I]\,(d+d'+a+\rho_I+\rho_S) - d'\,a'^2\}\,B_0 = [a'\,(d+d'+a+a'+\rho_I) + a\,\rho_I]\,a\,\textit{Б}_0 + a\,a'\,F\tau$
$+ a'^2\,P\tau$

A57. $(d+d'+a+a'+\rho_I+\rho_S)\,a\,\textit{Б}_0 = [(d+d'+a+\rho_I+\rho_S)\,(d+a'+\rho_I) + (d+d'+a+\rho_S)\,a - d'\,a']\,B_0 - a'\,P\tau$

Defining $\alpha = a+a'+d+d'$ and $\beta = \alpha\,(a+d) + (\alpha+d)\,\rho_I + (a+a'+d)\,\rho_S + (\rho_I+\rho_S)\,\rho_I$, these equations simplify to

A58. $\{\beta\,a' + a\,[\alpha\,\rho_I+(\rho_I + \rho_S)\,\rho_I]\}\,B_0 = [\alpha\,a' + (a+a')\,\rho_I]\,a\,\textit{Б}_0 + a\,a'\,F\tau + a'^2\,P\tau$

A59. $(\alpha+\rho I+\rho_S)\,a\,\textit{Б}_0 + a'\,P_\tau = \beta\,B_0$

Substituting B_0 from equation A59 into equation A58, taking $\gamma = \alpha+\rho_I+\rho_S$ and substituting $S_\infty = \rho_S\,\textit{Б}_0$, we get

A60. $\{\alpha\,[(a'+d')\,a\rho_I + (a+d)\,a'\rho_S] + (a+a')\,(a\rho_I^2+\gamma\rho_I\rho_S+a'\rho_S^2) + (ad'\rho_I+ a'd\rho_S)\,(\rho_I+\rho_S)\}\,S_\infty = a'\rho_S$
$[\beta F_\tau - \gamma\rho_I P_\tau]$

This is **equation S21** and along with equations A15 (equation S7) and A16 (equation S8) provide the general solution for S_∞. However a more compact and useful expression can be obtained if we **assume** that assembly or dissociation of the complexes on both exons 2 and 3 occurs much faster than commitment to either the S or I pathway: i.e., $a+d \gg \rho_I$, $a+d \gg \rho_S$, $a'+d' \gg \rho_I$ and $a'+d' \gg \rho_S$. This is in essence the same assumption made in the previous section. Using equations A15 and A16, along with the fact that $F_0 = L_0$ and $P_0 = 0$, the right term in equation A60 can be expanded as follows:

A61. Right term $= \frac{a'\rho s}{(r2 - r1)}\,[\beta\,(r_2\,L_0\,e^{r1\,\tau} - r_1\,L_0\,e^{r2\,\tau}) - \gamma\rho_I\,(a\,L_0\,e^{r2\,\tau} - a\,L_0\,e^{r1\,\tau})]$

As was mentioned before, the contributions generated by the slow decaying exponential ($e^{r1\tau}$) are taken to be dominant in this situation. Hence, the right term becomes

A62. Right term $= a'\rho_S\,L_0\,e^{r1\,\tau}\,\{\beta\,r_2 + \gamma\rho_I\,a\}\,/\,(r_2 - r_1)$

Substituting β and γ, we obtain

A63. Right term = $\frac{a'\rho s}{(r2 - r1)} L_0 \, e^{r1\,\tau} \, \{[(a+a'+d+d')(a+d) + (d+d')\rho_I + (a+a'+d+\rho_I)(\rho_I+\rho_S)]\, r_2 +$

$(a+a'+d+d'+\rho_I+\rho_S)\, \rho_I \, a\}$

Since ρ_I and ρ_S are both << a+d and a'+d' we ignore terms that contain either ρ_I or ρ_S as a factor, yielding

A64. Right term ≈ $\frac{a'\rho s}{(r2 - r1)} L_0 \, e^{r1\,\tau} \, (a+a'+d+d')\, (a+d)\, r_2$

Since $r_2 \approx r_2 - r_1$, we can further simplify this to

A65. Right term ≈ $a'\rho_S \, L_0 \, e^{r1\,\tau} \, (a+a'+d+d')\, (a+d)$

For the left side of equation A60 a similar procedure can be used.

A66. Left term = $\{(a+a'+d+d')\, [(a'+d')\, a\rho_I + (a+d)\, a'\rho_S] + (a+a')\, [a\rho_I^2 + (a+a'+d+d'+\rho_I+\rho_S)$

$\rho_I\rho_S + a'\rho_S^2] + (ad'\rho_I + a'd\rho_S)\, (\rho_I+\rho_S)\}\, S_\infty$

Here all the terms have either ρ_I or ρ_S, but some have only one of those factors while others have two multiplied or one of them raised to at least the second power. The contribution of the latter terms is therefore small compared to that of the former. An approximation can then be obtained by disregarding these smaller terms to obtain

A67. Left term ≈ $(a+a'+d+d')\, [(a'+d')\, a\rho_I + (a+d)\, a'\rho_S]\, S_\infty$

With these simplifications, solving equation A60 for S_∞ yields

A68. $S_\infty \approx a'\rho_S \, L_0 \, e^{r1\,\tau} \frac{(a+a'+d+d')\,(a+d)}{(a+a'+d+d')\,[(a'+d')\, a\rho I + (a+d)\, a'\rho s]}$

Rearranging, replacing r_1 with $-p_I$ and cancelling common factors in the numerator and denominator, we get

A69. $S_\infty \approx L_0 \, e^{-pI\,\tau} \frac{a'\rho s\,(a+d)}{(a'+d')\, a\rho I + (a+d)\, a'\rho s}$

Dividing the numerator and the denominator by the product (a'+d') (a+d), using p_I as defined previously and defining p_S analogously as

A70. $p_S = \rho_S / (1 + d' / a')$

yields

A71. $S_\infty \approx L_0 \, e^{-pI\,\tau} \, p_S/(p_S+p_I)$

This is **equation S23,** which is also **equation 5**. Regarding the quality of this approximation, lengthier analyses show that the assumptions made provide a good approximation for the left and right terms of equation A60. Additionally, deviations introduced by these two approximations at least partially cancel each other out when solving for S_∞ further improving its approximation.

References

Atkins P, de Paula J. 2002. *Physical Chemistry*. W. H. Freeman and Company, New York.

Becker NB, Rosa A, Everaers R. 2010. The radial distribution function of worm-like chains. *Eur Phys J E Soft Matter* **32**: 53-69.

Chen H, Meisburger SP, Pabit SA, Sutton JL, Webb WW, Pollack L. 2012. Ionic strengthdependent persistence lengths of single-stranded RNA and DNA. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 799-804.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 16176-16181.

Gossen M, Bujard H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 5547-5551.

Groth AC, Olivares EC, Thyagarajan B, Calos MP. 2000. A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5995-6000.

Guo M, Lo PC, Mount SM. 1993. Species-specific signals for the splicing of a short Drosophila intron in vitro. *Mol Cell Biol* **13**: 1104-1118.

Hertel KJ, Maniatis T. 1998. The function of multisite splicing enhancers. *Mol Cell* **1**: 449-455.

Pasman Z, Garcia-Blanco MA. 1996. The 5' and 3' splice sites come together via a three dimensional diffusion mechanism. *Nucleic Acids Res* **24**: 1638-1645.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 2007. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, New York.

Sterner DA, Carlo T, Berget SM. 1996. Architectural limits on split genes. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 15081-15085.

Talerico M, Berget SM. 1994. Intron definition in splicing of small Drosophila introns. *Mol Cell Biol* **14**: 3434-3445.

Xiao X, Wang Z, Jang M, Burge CB. 2007. Coevolutionary networks of splicing cis-regulatory elements. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 18583-18588.

Zhang XH, Arias MA, Ke S, Chasin LA. 2009. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA* **15**: 367-376.