Alternative Splicing and Genomic Stability

Kevin Cahill

cahill@unm.edu

http://dna.phys.unm.edu/

Abstract

In a cell that uses alternative splicing, the total length of all the exons is far less than in a cell that encodes the same set of proteins without using alternative splicing. A genome with a shorter exon length is more stable. For a broad class of alternatively spliced genes, a formula is given for the increase in their stability.

Alternative Splicing

A procaryote transcribes one or more genes into mRNA and immediately translates the mRNA into protein. But a eucaryote first transcribes a single gene into pre-mRNA, and then, using spliceosomes, turns the pre-mRNA into mRNA by splicing out most or all of its introns and often many of its exons.

The eucaryote then exports the mRNA out of its nucleus into its cytosol, where its ribosomes translate the mRNA into protein.

A eucaryote often can make different proteins from the same pre-mRNA transcript by splicing it in different ways. This trick is called alternative splicing.

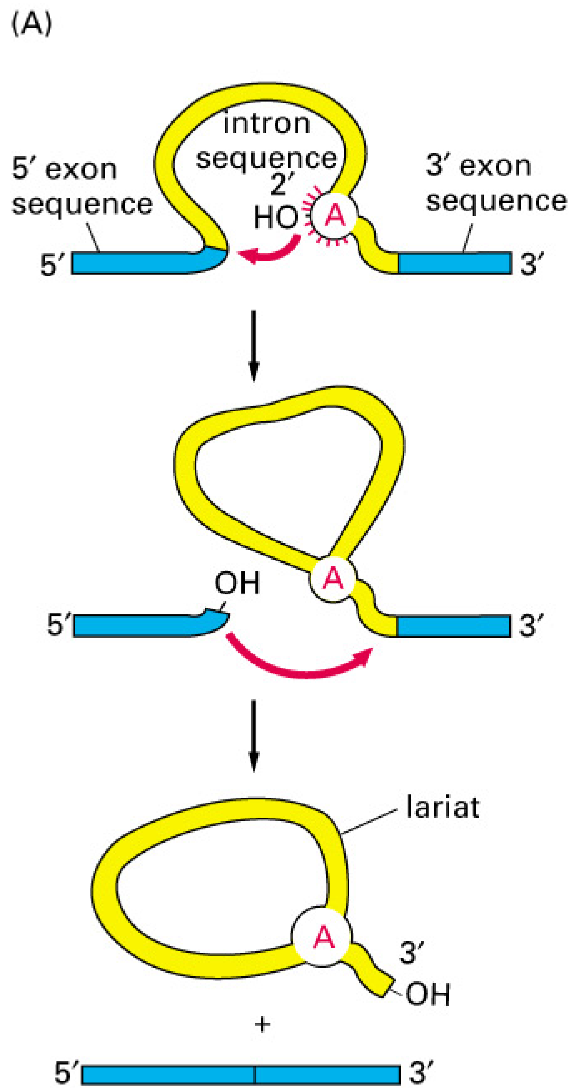


Figure 6–26 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

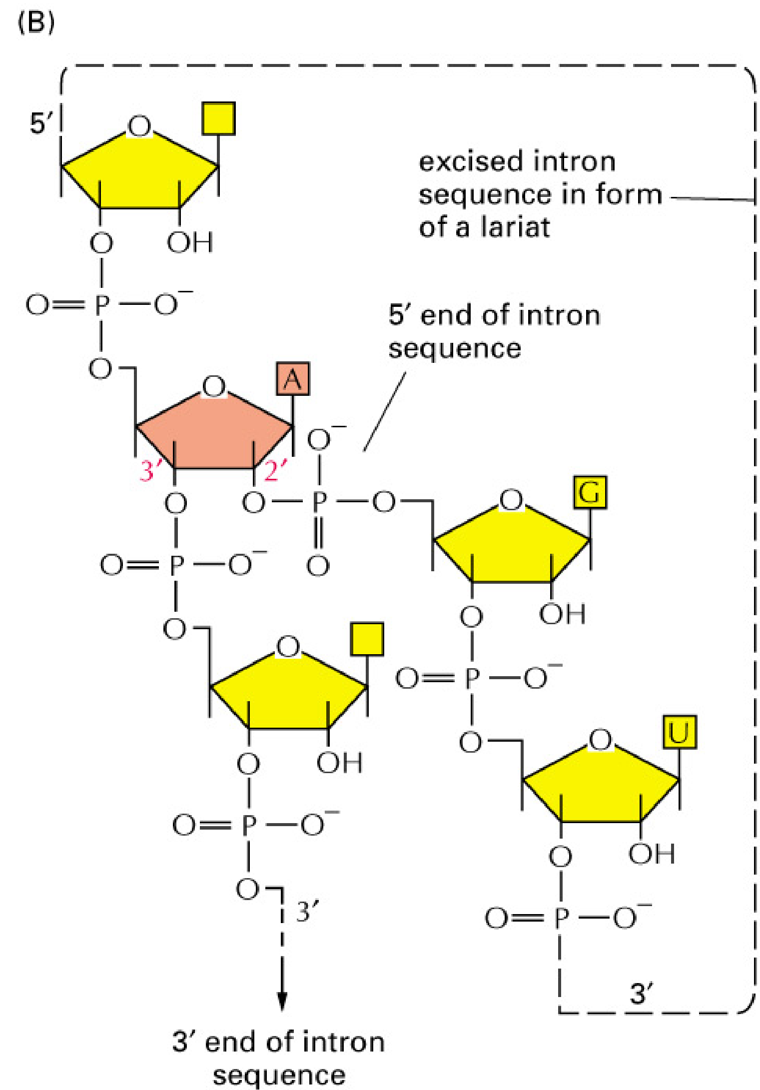


Figure 6–26 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

A spliceosome cuts out an intron. Part B illustrates the yellow lariat near the bottom of part A.

Why Splice Alternatively?

Alternative splicing allows an organism to make different proteins in different cells at different times, all from the same gene, by poorly understood regulatory devices. But this diversity of proteins could also be produced by several different genes controlled by promoters and enhancers --- in fact, that is how biologists thought genes worked until they discovered alternative splicing.

The advantage of alternative splicing is that its economical use of exons makes genes more stable during reproduction and development.

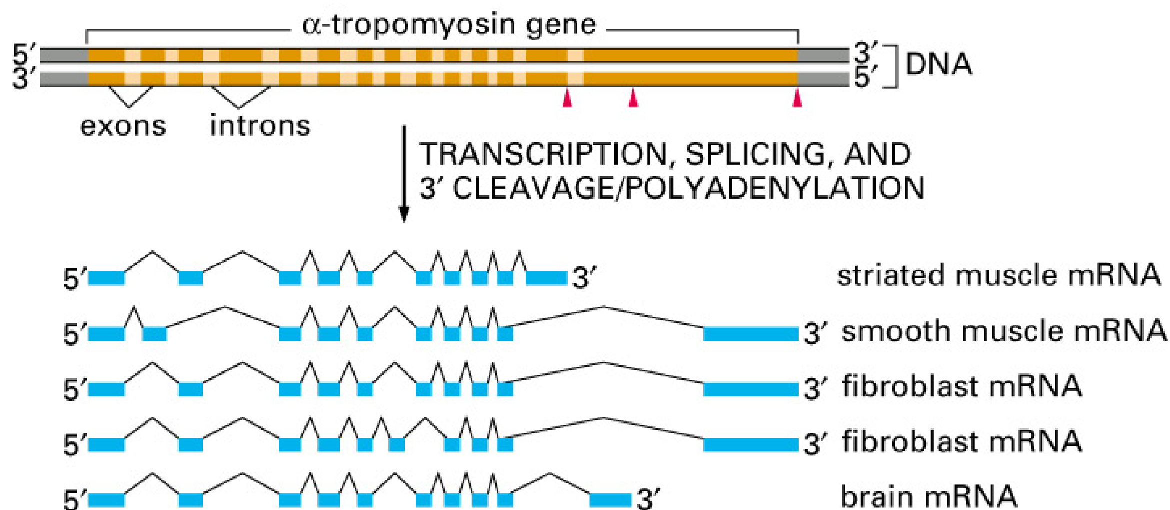


Figure 6–27. Molecular Biology of the Cell, 4th Edition.

An Example

Consider, for instance, a gene that has a long exon of 1000 base pairs (b) and two short ones, each 100 b long. The total length of its exons is 1200 b. Alternative splicing allows the cell to make two different RNAs, each of 1100 b.

Without alternative splicing, the cell would need two genes, each 1100 b long, for a total exon length of 2200 b.

Thus in this example, alternative splicing reduces the length of the exons in the DNA by 45%. This reduction in the length of exonic DNA implies a reduction of 45% in the error rate during the replication of this gene. In effect, the gene is nearly twice as stable due to alternative splicing.

Since an error in the replication of critical exonic DNA is potentially lethal, this extra genomic stability is biologically significant and is one of the reasons why higher eucaryotes use alternative splicing.

How Does the Spliceosome Know What to Excise?

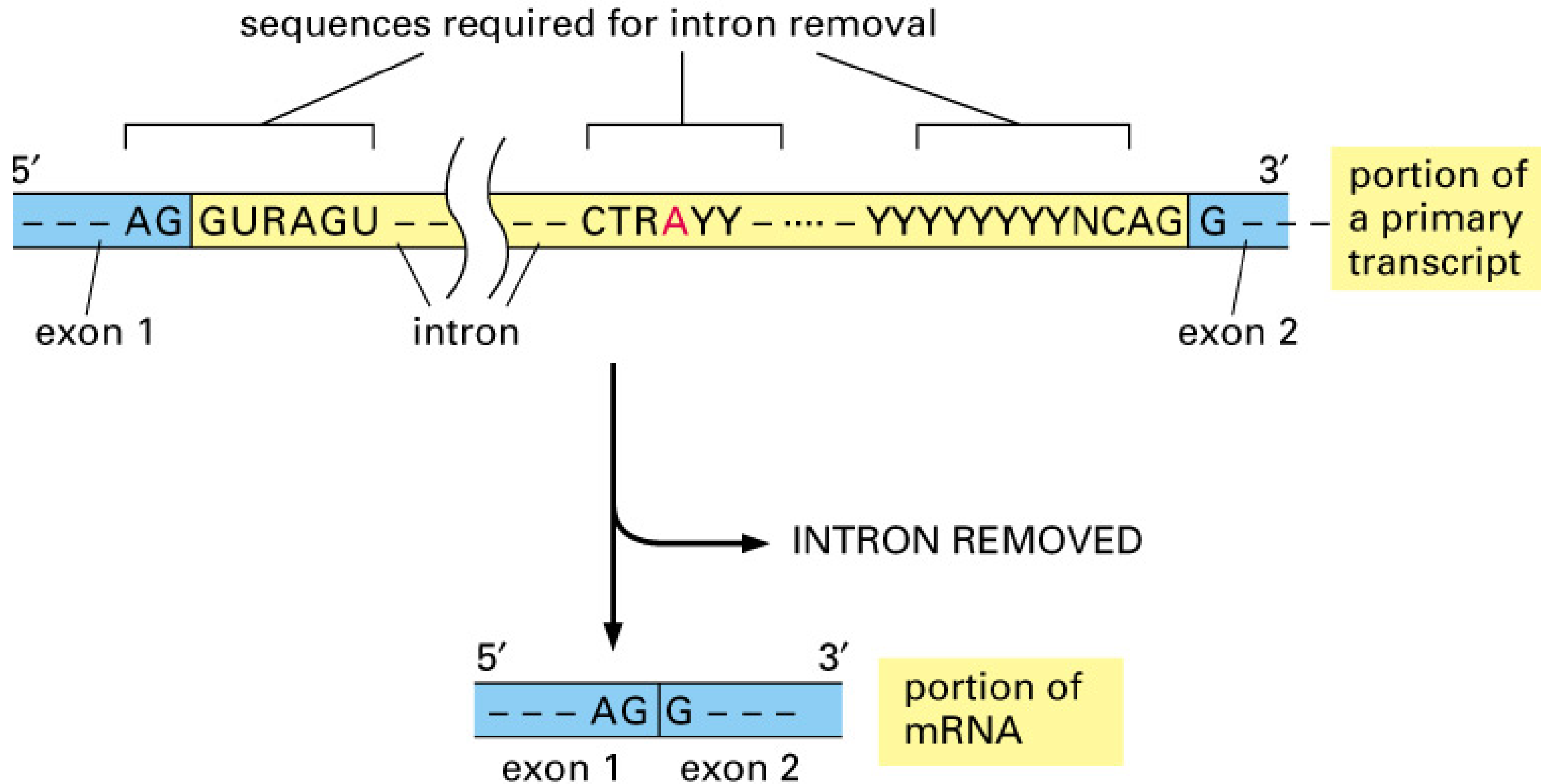


Figure 6–28. Molecular Biology of the Cell, 4th Edition.

Here R = A or G; Y = C or U; N = R or Y; and in DNA T = U in RNA.

The RNA Splicing Mechanism

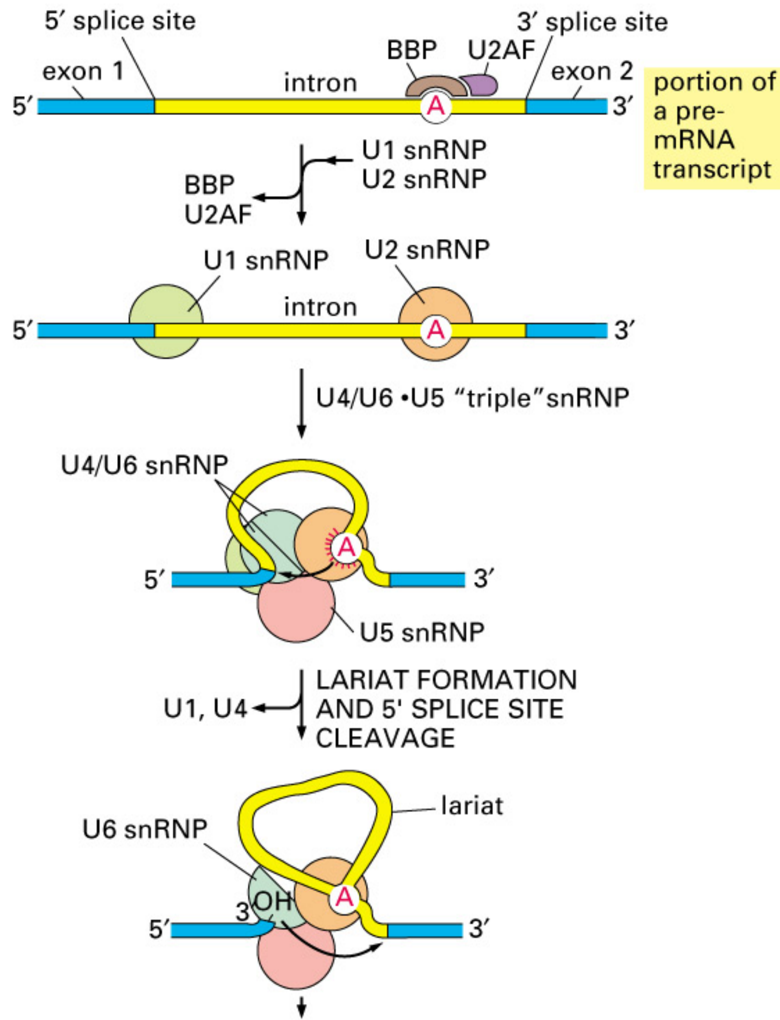


Figure 6–29 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

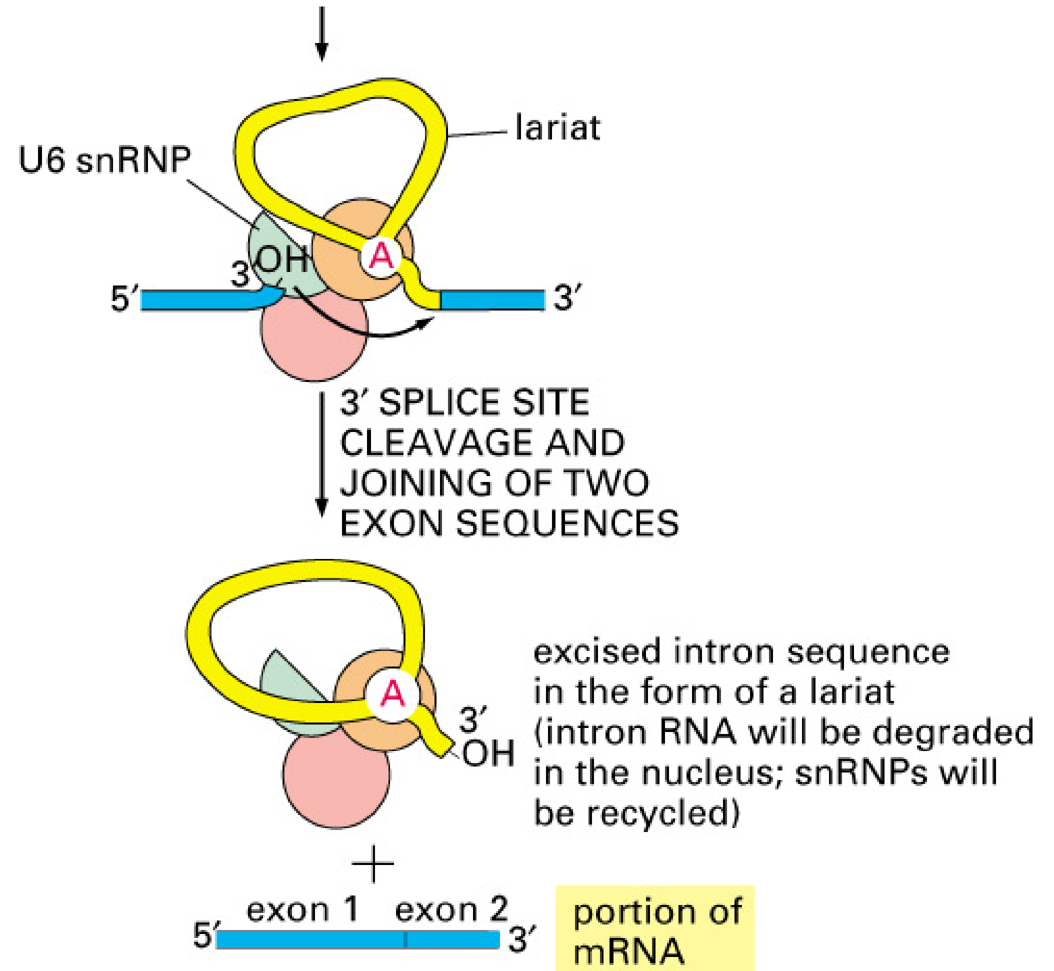


Figure 6–29 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

A spliceosome is a team of snRNPs (small nuclear ribonucleoproteins).

How the snRNAs Work

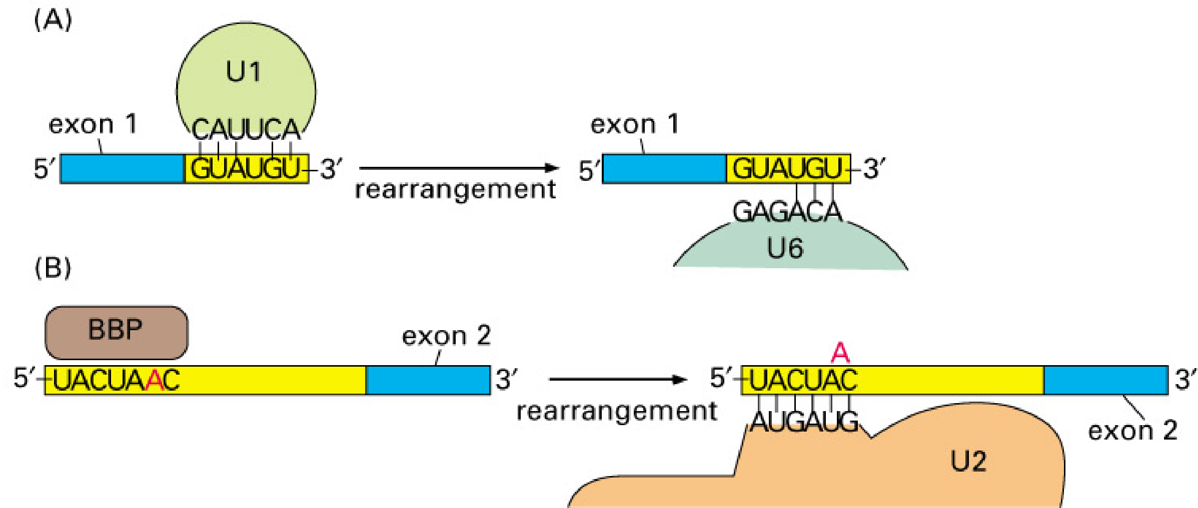


Figure 6–30 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

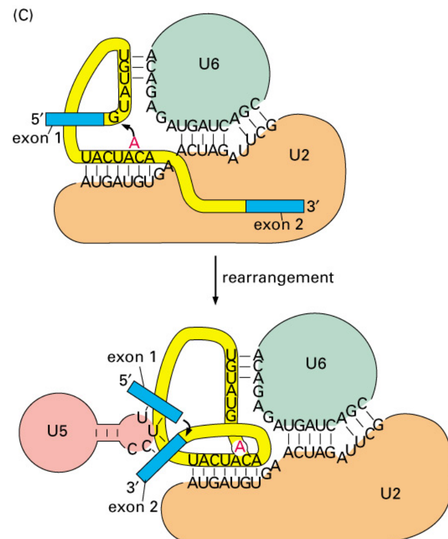


Figure 6–30 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

Four Patterns of Alternative RNA Splicing

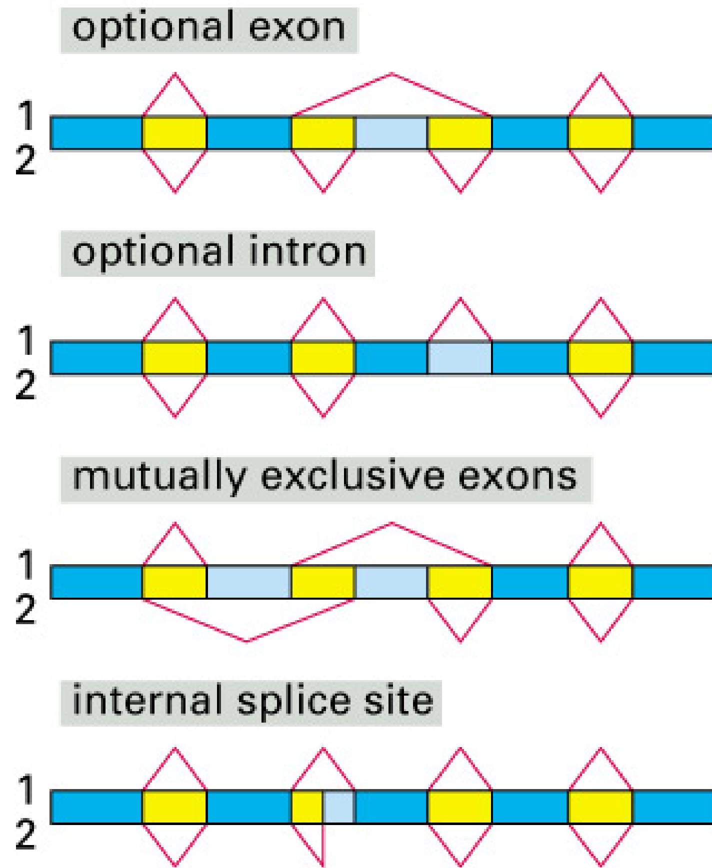


Figure 7–88. Molecular Biology of the Cell, 4th Edition.

Assumptions

Let us consider a gene that has M groups of mutually exclusive exons in addition to the constitutively spliced exons. For each group i $(i = 1, \ldots M)$, let N_i denote the number of mutually exclusive exons in the group (including the null exon of length zero if the spliceosome may splice out all the exons of the group).

Assume that the spliceosome always selects at most one exon and no introns from each of the M groups of N_i exons with no shuffling. Assume that the organism expresses all

$$\mathcal{N} = \prod_{j=1}^{M} N_j \tag{1}$$

possible proteins at some time in some cell.

Without Alternative Splicing

If L_r is the total length in nucleotides of the exons that are always kept, then without alternative splicing, these L_r nucleotides would be repeated in each of the $\mathcal{N}$ proteins for a total length of $\mathcal{N}\, L_\mathrm{r}$.

Let L_{ik} be the length of exon k of group i. Then without alternative splicing, each of these N_i exons of length L_{ik} would be repeated $\mathcal{N}/N_i$ times. So the total length devoted to group i without alternative splicing is

$$\frac{\mathcal{N}}{N_i} \sum_{k=1}^{N_i} L_{ik} = \left(\prod_{\substack{j=1 \\ j \neq i}}^{M} N_j \right) \sum_{k=1}^{N_i} L_{ik}. \tag{2}$$

Thus the number of nucleotides that would be needed to encode for all $\mathcal{N}$ proteins and that would have to be copied correctly each time a cell divides is

$$N_\mathrm{nas} = \mathcal{N} \left(L_\mathrm{r} + \sum_{i=1}^{M} \frac{1}{N_i} \sum_{k=1}^{N_i} L_{ik} \right) \tag{3}$$

without alternative splicing.

With Alternative Splicing

But with alternative splicing, the number of needed nucleotides is only the length of all the exons,

$$N_{\rm as} = L_{\rm r} + \sum_{i=1}^{M} \sum_{k=1}^{N_i} L_{ik}. \tag{4}$$

The Advantage of Having $N_{\rm as} < N_{\rm nas}$

Since the error rate in the replication of DNA is 10^{-9} per base pair [Alberts *et al.*, 2002], the probability of an exonic error in the gene during replication is $N_{\rm nas} \times 10^{-9}$ without alternative splicing, but only $N_{\rm as} \times 10^{-9}$ with alternative splicing. So if we ignore the critical control sequences in the introns, then the ratio

$$I = \frac{N_{\rm nas}}{N_{\rm as}} \tag{5}$$

is the increase in the stability of the gene due to alternative splicing. The intron control sequences probably boost I slightly.

Drosophila DSCAM

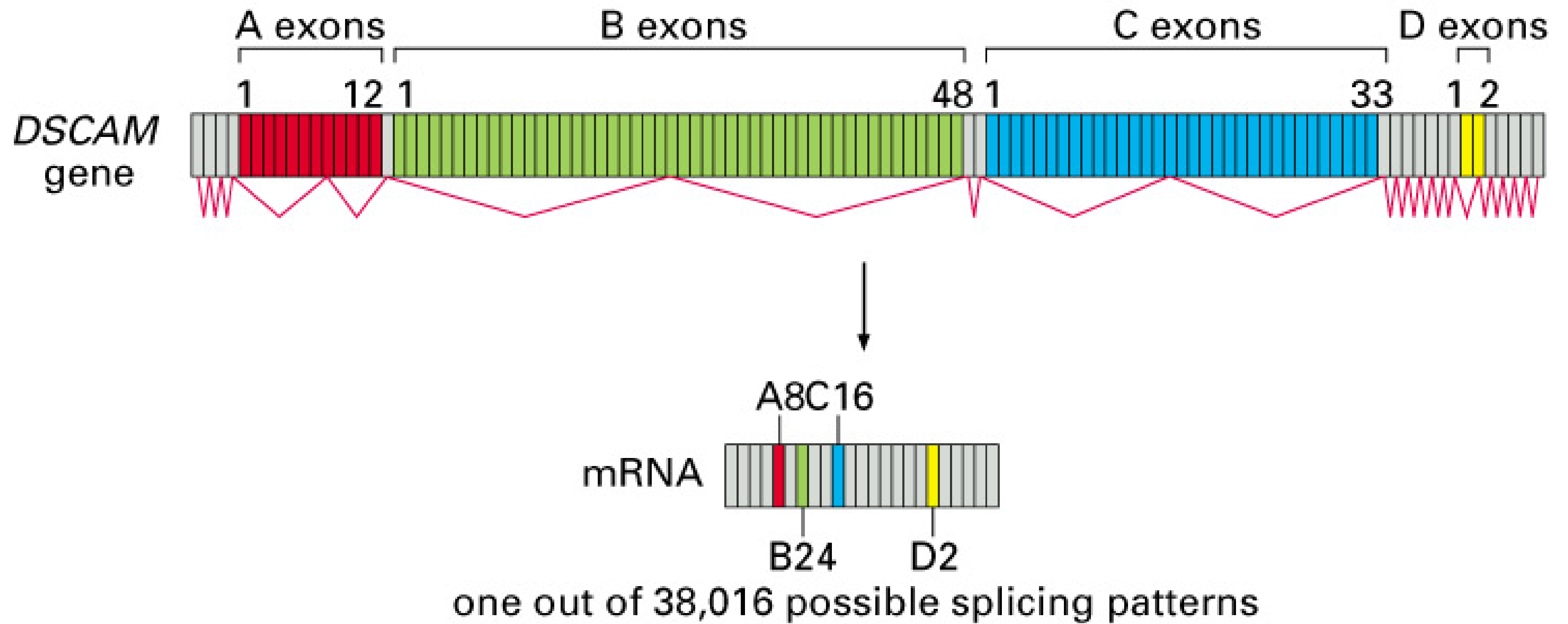


Figure 7–89. Molecular Biology of the Cell, 4th Edition.

Negative and Positive Control of Alternative Splicing

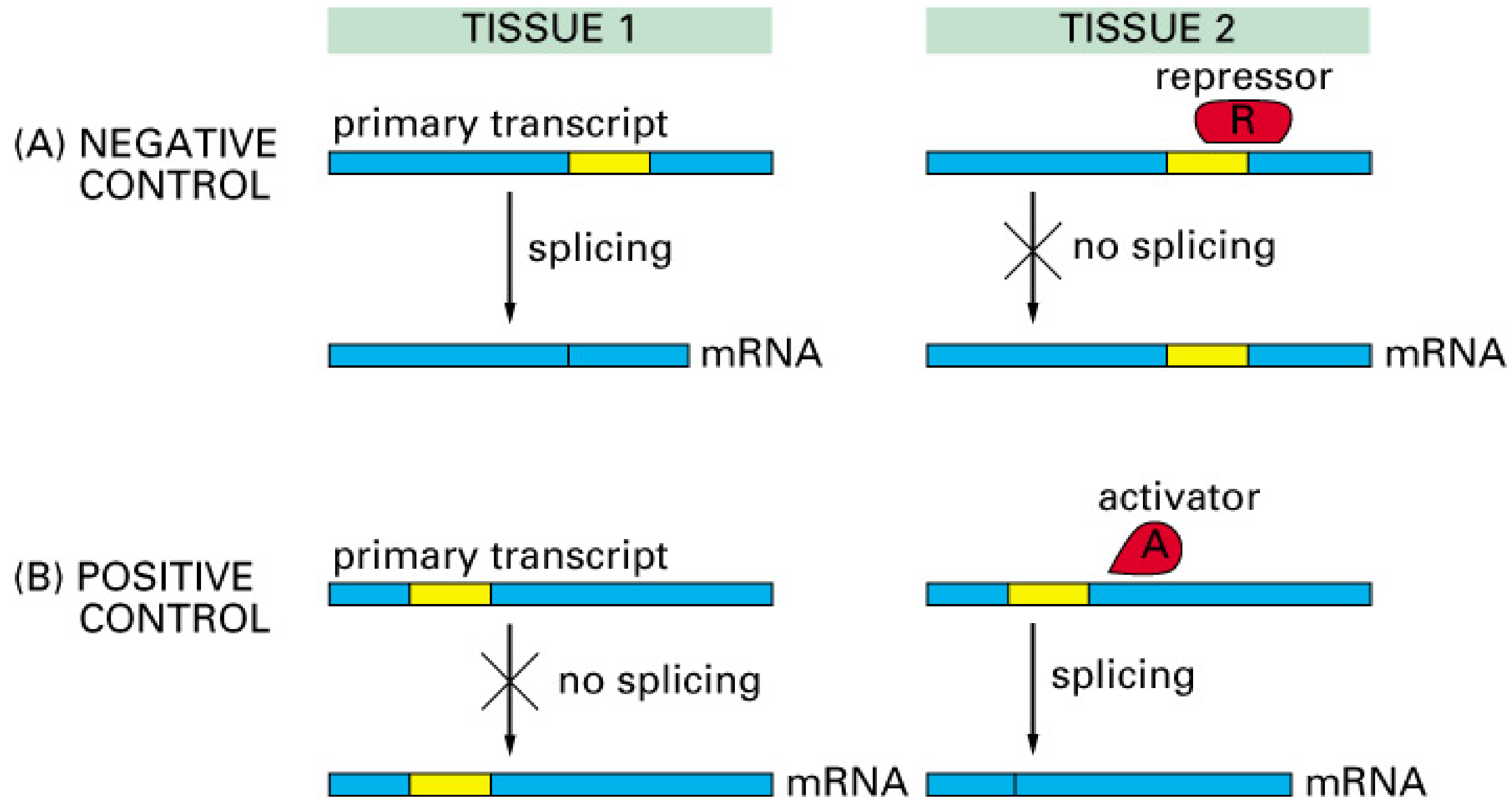


Figure 7–90. Molecular Biology of the Cell, 4th Edition.

Here R and A are proteins that bind to the preRNA and Repress or Activate a splice site.

Drosophila DSCAM

The DSCAM gene of Drosophila provides a striking example of alternative splicing. This gene encodes receptors that guide the growth of the axon of Bolwig's nerve in the fly embryo [Schmucker *et al.*, 2000]. It has $M = 4$ groups of $N_1 = 12$, $N_2 = 48$, $N_3 = 33$, and $N_4 = 2$ exons [Schmucker *et al.*, 2000, Black, 2000]. The exons in each group are mutually exclusive, and the total number of possible proteins is $\mathcal{N} = 12 \times 48 \times 33 \times 2 = 38,016$. The DSCAM gene, including introns, is 61.2 kb long, and its mRNA, after transcription and splicing, contains 24 exons and is 7.8 kb long [Schmucker *et al.*, 2000, Black, 2000].

The ratio $N_{\text{nas}}/N_{\text{as}}$ depends explicitly upon the lengths L_{r} and L_{ik}. Since most internal exons are between 50 and 300 nucleotides in length [Smith & Valcárcel, 2000], let us simplify the bookkeeping by setting $L_{ik} = 200$ b. The spliced DSCAM mRNA is 7.8 kb long and contains 4 alternatively spliced exons and 20 constitutively spliced exons. So the set of constitutively spliced exons is of length

$$L_{\text{r}} = 7800 - 4 \times 200 = 7000 = 7\text{kb}. \tag{6}$$

$$N_{\rm as} < N_{\rm nas}$$

Thus by Eq.(4), the exonic length required with alternative splicing is

$$\begin{aligned} N_{\rm as} &= L_{\rm r} + 200 \sum_{i}^{4} N_i \\ &= 7000 + 200\,(12 + 48 + 33 + 2) = 26000 = 26\text{kb}. \end{aligned} \tag{7}$$

But by Eq.(3), the exonic length required without alternative splicing is

$$\begin{aligned} N_{\rm nas} &= \mathcal{N}\,(L_{\rm r} \,+\, 200\,M) \\ &= 38016 \times 7800 = 296524800 = 297\text{Mb}, \end{aligned} \tag{8}$$

which, incidentally, is nearly twice the length of the entire Drosophila genome and about six times the length of all the exons in the human genome.

Error Reduction

With these assumptions, the chance of a crucial error in the DSCAM gene during replication is 0.30 without alternative splicing, but only 2.6×10^{-5} with alternative splicing. The ratio I

$$I = \frac{N_{\text{nas}}}{N_{\text{as}}} = 11400 = 1.1 \times 10^4 \tag{9}$$

is the increase in genetic stability due to alternative splicing.

Flies without alternative splicing would accumulate about 10,000 exonic DSCAM errors in 30,000 generations (2,500 years), and each fly would have its own set of 10,000 errors. Over this period, the DSCAM gene of the fly population gradually would become uniformly dysfunctional with relatively small differences in fitness among individual flies.

But flies with alternative splicing would accumulate less than one exonic DSCAM error in 30,000 generations.

Alternative Splicing and Natural Selection

Moreover, the probability that the one error would occur in the L_r exons that are constitutively expressed would be $L_\mathrm{r}/N_\mathrm{as} = 7/26 = 0.27$, and that unlucky fly would be distinctly unfit.

Thus alternative splicing helps natural selection weed out unfit individuals, and so alternative splicing and natural selection cooperate to preserve the integrity of the genome.

If I Were 5

In most genes, the increase in genomic stability due to alternative splicing might be more like 5 or 10 than 10^4, but even a 500% increase in genetic stability during reproduction and development is worth the trouble of alternative splicing. For if the average gene were 5 times longer without alternative splicing, then 7.5% rather then 1.5% of the genomes of higher vertebrates would consist of exons. The DNA of a human diploid cell has 6.4 billion base pairs. The error rate of 10^{-9} per base pair implies that on average there will be 6.4 errors per cell division. With alternative splicing, only 1.5% of these errors occur in exons and are potentially deleterious, so the probability of a daughter cell with perfect exons is

$$P_{\text{as}} = (1 - 10^{-9})^{0.015 \times 6.4 \times 10^9} \approx e^{-0.096} = 0.908. \tag{10}$$

Without alternative splicing, 7.5% of the errors would occur in exons, and so the probability of a daughter cell with perfect exons would be

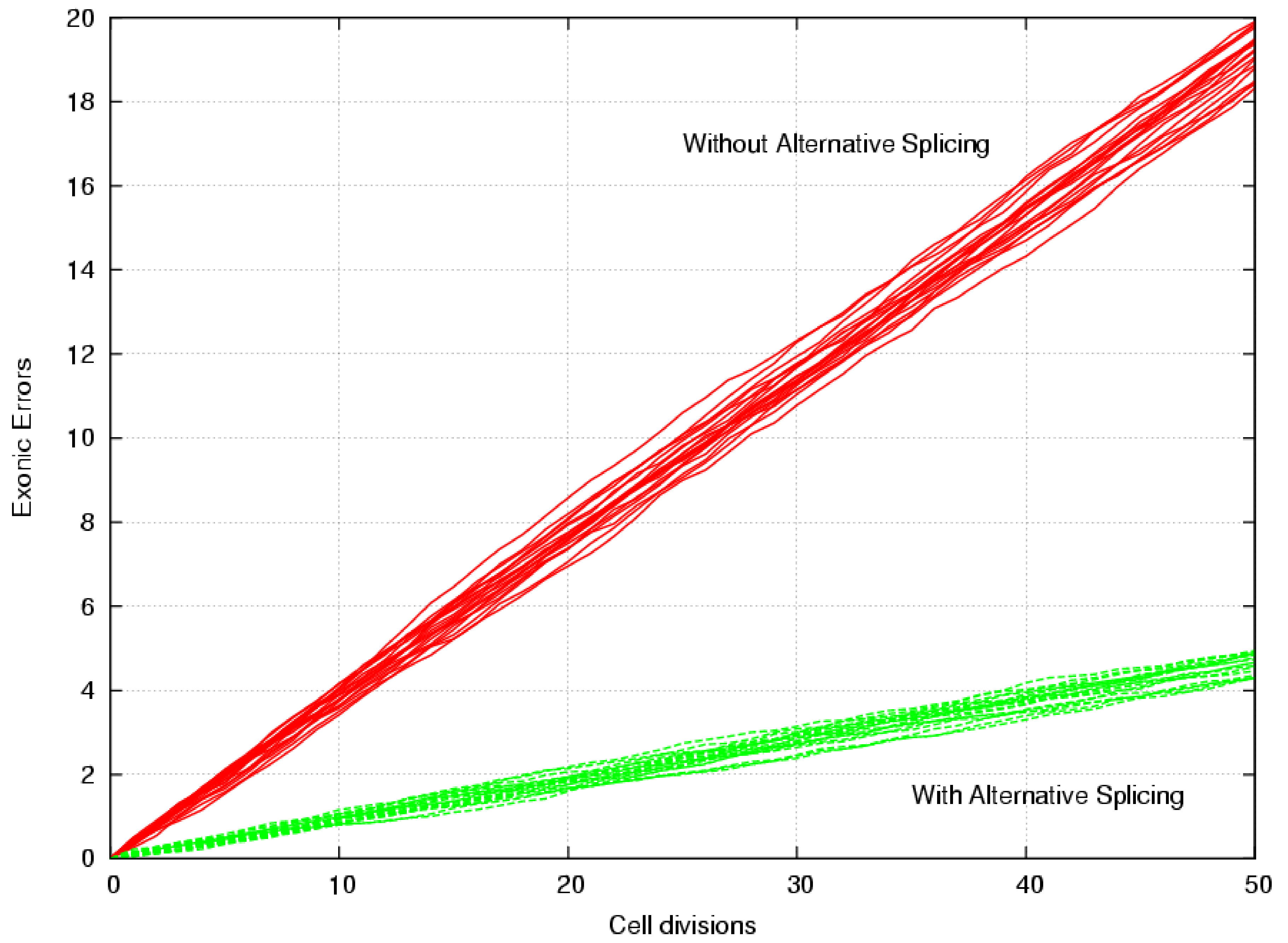
$$P_{\text{nas}} = (1 - 10^{-9})^{0.075 \times 6.4 \times 10^9} \approx e^{-0.48} = 0.619. \tag{11}$$

Human Development

The adult human arises from about 46 cell divisions, so the probability that any given adult cell has perfect exons is $(P_{\mathrm{as}})^{46} = 0.012$ with alternative splicing, but only $(P_{\mathrm{nas}})^{46} = 2.6 \times 10^{-10}$ without alternative splicing.

To estimate the effects of the probabilities P_{as} and P_{nas} on evolution and development, I let 1000 cells divide 50 times with errors accumulating according to the probabilities $(1 - P_{\mathrm{as}})$ and $(1 - P_{\mathrm{nas}})$. I divided the 1000 cells into 20 groups of 50 cells each and plotted in the figure the average number of exonic errors per cell for each of the 20 groups with and without alternative splicing.

As shown in the figure, the average number of defective exonic base pairs per daughter cell after 46 cell divisions is about 4 with alternative splicing (lower, green lines) and more than 17 without alternative splicing (upper, red lines).


Without Alternative Splicing
With Alternative Splicing
Exonic Errors
Cell divisions
0
2
4
6
8
10
12
14
16
18
20
0
10
20
30
40
50

Exon Errors

Since with alternative splicing, cells free of exonic error produce daughter cells that also are free of exonic error at a rate of 91%, apoptosis followed by division of adjacent cells can correct the 1 or 2 of the 4 exonic errors that are troublesome. But because without alternative splicing, cells free of exonic error produce daughter cells free of exonic error at a rate of only 62%, it is hard to see how apoptosis could cope with 17 exonic errors per adult cell.

Estimate of Increase in Stability

We may derive a rule of thumb for the increase in genetic stability by noting that $\langle L_\mathrm{s} \rangle$ defined by

$$\langle L_\mathrm{s} \rangle = \sum_{i=1}^{M} \frac{1}{N_i} \sum_{k=1}^{N_i} L_{ik} \tag{12}$$

is an effective average length of the alternative exons that are spliced into the mRNA and that $\langle N \rangle$ defined by

$$\langle N \rangle \langle L_\mathrm{s} \rangle = \sum_{i=1}^{M} \sum_{k=1}^{N_i} L_{ik} \tag{13}$$

is a kind of average of the numbers N_i of alternative exons in the M groups. Let us further use r for the ratio of the average length $\langle L_\mathrm{s} \rangle$ of the selected exons to the length L_r of the exons that are always expressed,

$$r = \frac{\langle L_\mathrm{s} \rangle}{L_\mathrm{r}}. \tag{14}$$

Rule of Thumb

Then with these definitions, the increase I in genetic stability is

$$I = \frac{N_{\rm nas}}{N_{\rm as}} = \mathcal{N}\,\frac{1+r}{1+\langle N\rangle\, r}. \tag{15}$$

The fraction that multiplies the total number $\mathcal{N}$ of possible proteins is less than unity. But it is generally not tiny because the ratio r usually is small and because $\langle N\rangle$ usually is less then 30. In the case of Drosophila DSCAM and with the assumptions $L_{\rm r} = 7.0$ kb and $L_{ik} = 200$ b, the four selected exons are of length $\langle L_{\rm s}\rangle = 800$ b; the ratio r is $r = 800/7000 = 0.114$; and the effective average number $\langle N\rangle$ of exons per group is $\langle N\rangle = 95 \times 200/\langle L_{\rm s}\rangle = 95/4 = 23.7$. The fraction $(1+r)/(1+\langle N\rangle\, r) = 3/10$, and the increase in genetic stability is $I = 0.3\,\mathcal{N} = 11,400$.

Hearing in Chickens

Hearing in chickens provides another example of the contribution of alternative splicing to genetic stability. The cSlo gene of the chicken cochlea encodes the membrane proteins that form Ca^{2+}-activated K^+ channels that determine the resonant frequency of each hair cell in the basilar papilla. Alternative splicing provides some $\mathcal{N} = 576$ variants of the mRNA for each of the four components of this tetramer membrane protein [Rosenblatt *et al.*, 1997, Navaratnam *et al.*, 1997, Black, 1998], resulting in a huge number possible resonant frequencies. In cSlo, the ratio $r = 0.1$, and the mean number $\langle N \rangle$ of exons in each of the eight groups is about 2.6 [Rosenblatt *et al.*, 1997]. So by the rule of thumb (15), alternative splicing increases the genetic stability of cSlo by a factor of about

$$I = 576 \, \frac{1.1}{1.26} = 503. \tag{16}$$

The tetrameric structure of the functional membrane protein effectively boosts I by another factor.

Conclusions

We have seen that the exonic economy of alternative splicing increases the stability of the genome. The mammalian immune system achieves a similar exonic economy by employing site-specific genetic recombination in developing B cells [Alberts *et al.*, 2002]. But here genetic stability is less of a goal because diversity in antibody chains is the key to immunity.

References

[Alberts *et al.*, 2002] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell*. New York, NY: Garland Science, fourth edition.

[Black, 1998] Black, D. L. (1998). *Neuron*, **20**, 165--168.

[Black, 2000] Black, D. L. (2000). *Cell*, **103**, 367--370.

[Navaratnam *et al.*, 1997] Navaratnam, D. S., Bell, T. J., Tu, T. D., Cohen, E. L., & Oberholtzer, J. C. (1997). *Neuron*, **19**, 1077--1085.

[Rosenblatt *et al.*, 1997] Rosenblatt, K. P., Sun, Z.-P., Heller, S., & Hudspeth, A. J. (1997). *Neuron*, **19**, 1061--1075.

[Schmucker *et al.*, 2000] Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E., & Zipursky, S. L. (2000). *Cell*, **101**, 671--684.

[Smith & Valcárcel, 2000] Smith, C. W. J. & Valcárcel, J. (2000). *Trends in Biochemical Science*, **25**, 381--388.

Acknowledgments

Thanks to Stephanie Ruby and James Thomas for helpful conversations. Peter J. Steinbach kindly hosted the author during his sabbatical at NIH.