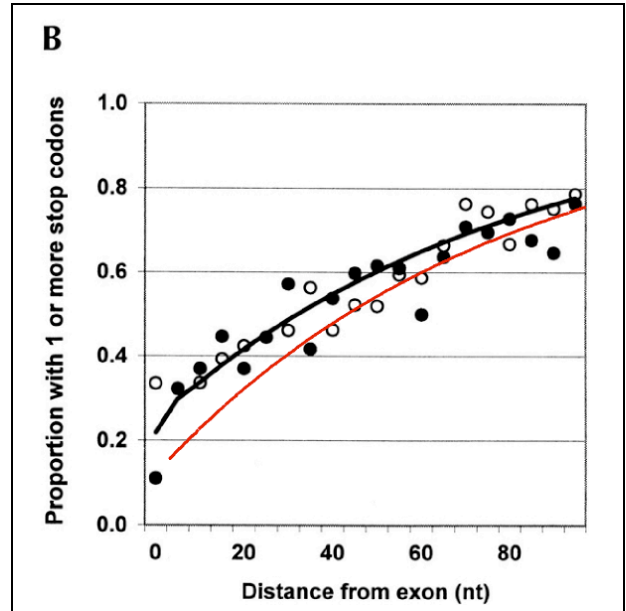


Regulation of splicing: The importance of being translatable

APPENDIX to E. Miriami, R. Sperling, J. Sperling and U. Motro. 2004. *RNA* 10:1-4.

In this appendix we explain our claim that Zhang et al. (2003) erred in plotting their data (Figure 1B in that paper), which led them to wrongly conclude that the frequency of in-frame stop codons upstream of latent 5' splice sites is not greater than that expected by chance.

The x-axis in Figure 1B of Zhang et al. (2003) (see box) represents introns, grouped in windows of 5, by the distance at which their most 5' latent splice site is found downstream of the authentic one. The y-axis represents the proportion of introns in each group, which have at least one in-frame stop codon upstream of the latent site (dots). Because the probability of having at least one in-frame stop codon in sequences longer than 100 nt approaches unity, only introns in which the distance between the latent and the authentic 5' splice sites is smaller than 100 nt were chosen for this analysis. The black curve represents the expected proportion of introns that have at least one stop codon. We claim that this curve was wrongly drawn and should be replaced by the red curve.



To calculate the black curve, Zhang et al. (2003) divided the 100 nt region into two components (1-6 and 7-100). The first 6 nts of all introns comprise part of the splice site itself, which has a high probability of harboring a stop codon due to the URA embedded in the consensus sequence AG/GURAGU. Based on the matrix of bases underlying this consensus sequence, they calculated that the probability of a stop codon sequence occurring by chance is 0.74. To obtain the probability for an in-frame occurrence, they divided 0.74 by 3, assuming equal probabilities for all 3 possible reading frames. This was a wrong assumption, because the frequency at which the URA triplet of the consensus is in the reading frame of upstream exon is not 0.333, but significantly lower. For our human database, we have found this frequency to be 0.186 (Miriami et al., 2002), which is compatible with 0.182 for mammals (Tomita et al., 1996), and 0.20 and 0.21 for human databases (Long et al., 1998; Long & Deutsch, 1999), respectively. Thus, the appropriate probability of finding an in-frame stop codon in the +1 to +6 region is at most 0.155 (0.74 x 0.21), rather than 0.274 (0.74 x 0.333).

To obtain the expected fraction (E) of introns harboring at least one in-frame stop codon (black curve), Zahng and Chasin used the 0.247 value in the equation:

$$E = 1 - e^{-0.247} - e^{-m(L-6)/3},$$

where m , the probability of finding an in-frame stop codon in the +7 to +100 nt region, is 0.038 per triplet, and L is the distance from the exon border to the lower limit of the latent site. Whereas using the correct value of 0.155 in the correct^a equation:

$$E = 1 - (1 - 0.155) - (1 - m)^{(L-6)/3}$$

gives the red curve. It can be seen that most data points fall above this curve, meaning that the proportion of intron sequences having at least one in-frame nonsense codon is greater than that expected by chance.

It should be pointed out that in the present letter, as well as in our previous publication (Miriami et al., 2002) we based our conclusion on the finding that in-frame stop codons occur more frequently in introns having latent 5' splice sites than in introns lacking such sites. It is therefore reassuring that a different approach, where the actual frequency of in-frame stop codons is shown to be higher than that expected by chance alone, leads to the same conclusion; namely, that the occurrence of intronic in-frame stop codon(s) upstream of latent 5' splice sites is a general factor in the suppression of splicing at such sites.

REFERENCES

Long, M., de Souza, S. J., Rosenberg, C., and Gilbert, W. 1998. Relationship between "proto-splice sites" and intron phases: evidence from dicodon analysis. *Proc. Natl. Acad. Sci. USA* 95:219-223.

Long, M., and Deutsch, M. 1999. Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. *Mol. Biol. Evol.* 16:1528-1534.

Miriami, E., Motro, U., Sperling, J., and Sperling, R. 2002. Conservation of an open-reading frame as an element affecting 5' splice site selection. *J. Struct. Biol.* 140:116-122.

Tomita, M., Shimizu, N., and Brutlag, D. L. 1996. Introns and reading frames: correlation between splicing sites and their codon positions. *Mol. Biol. Evol.* 13:1219-1223.

Zhang, X., Lee, J., and Chasin, L. A. 2003. The effect of nonsense codons on splicing: A genomic analysis. *RNA* 9:637-639.

Footnote

^aFor their equation of the expected fraction of introns harboring at least one in-frame stop codon Zhang et al. (2003) used a Poisson probability distribution. We use a binomial probability distribution, which is more accurate, although both curve almost overlap (see below).

