ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Structural Biology 140 (2002) 116–122

Journal of Structural Biology

www.academicpress.com

Conservation of an open-reading frame as an element affecting 5′ splice site selection

Elana Miriami,[a] Uzi Motro,[b] Joseph Sperling,[c] and Ruth Sperling[a,*]

[a] *Department of Genetics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel*
[b] *Department of Evolution Systematics and Ecology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel*
[c] *Department of Organic Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel*

Received 9 July 2002, and in revised form 17 September 2002

Abstract

Splice site selection is a key element of pre-mRNA splicing and involves specific recognition of consensus sequences at the 5′ and 3′ splice sites. Evidently, the compliance of a given sequence with the consensus 5′ splice site sequence is not sufficient to define it as a functional 5′ splice site, because not all sequences that conform with the consensus are used for splicing. We have previously hypothesized that the necessity to avoid the inclusion of premature termination codons within mature mRNAs may serve as a criterion that differentiates normal 5′ splice sites from unused (latent) ones. We further provided experimental support to this idea, by analyzing the splicing of pre-mRNAs in which in-frame stop codons upstream of a latent 5′ splice site were mutated, and showing that splicing using the latent site is indeed activated by such mutations. Here we evaluate this hypothesis by a computerized survey for latent 5′ splice sites in 446 protein-coding human genes. This data set contains 2311 introns, in which we found 10 490 latent 5′ splice sites. The utilization of 10 045 (95.8%) of these sites for splicing would have led to the inclusion of an in-frame stop codon within the resultant mRNA. The validity of this finding is confirmed here by statistical analyses. This finding, together with our previous experimental results, invokes a nuclear scanning mechanism, as part of the splicing machine, which identifies in-frame stop codons within the pre-mRNA and prevents splicing that could lead to the formation of a prematurely terminated protein.

1. Introduction

Pre-mRNA splicing is a critical step in gene expression and involves two transesterification reactions through which intron sequences are removed and exon sequences are ligated, yielding mRNA that can exit from the nucleus to the cytoplasm. The accuracy and efficiency of the splicing reaction were attributed to several *cis* sequence elements and *trans*-acting factors, which are required for splicing. The *trans*-acting factors include spliceosomal U snRNPs and a multitude of protein splicing factors which interact specifically, in a dynamic fashion, with the pre-mRNA *cis* elements. The *cis* elements include consensus sequences at the 5′ and 3′ splice sites, a branch point, a polypyrimidine tract, and splicing enhancer and silencer sequence elements. Base pairing interactions between conserved pre-mRNA and U snRNA sequences play an important role in splice site selection and in the chemical transformation of the splicing reaction. A key step in these interactions involves the recognition and selection of a consensus sequence AG/GTRAGT (in mammals, where R denotes purine and "/" denotes the splice junction) at the 5′ splice site (Black, 1995; Burge et al., 1999; Krämer, 1996; Staley and Guthrie, 1998).

Frequently, however, sequences that comply with the consensus are not selected for splicing (Green, 1991). For example, an internal intron in the CAD gene [encoding the multifunctional protein carbamoyl-phosphate synthetase, aspartate *trans*-carbamylase, dihydroorotase (Padgett et al., 1982)] of Syrian hamster cells contains the octanucleotide sequence AG/GUGGGU, which complies with the 5′ splice site consensus sequence better than the upstream normal 5′ splice site of the same intron (AG/GUGGCA). Nonetheless, the downstream 5′ splice site is not used for splicing under normal

* Corresponding author. Fax: +972-2-658-6975.
E-mail address: sperling@vms.huji.ac.il (R. Sperling).

1047-8477/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S1047-8477(02)00539-7

growth conditions. However, it can be activated for splicing by mild heat shock treatment, indicating its capability to serve as a legitimate 5′ splice site (Miriami et al., 1994). We refer to such 5′ consensus sequences as "latent 5′ splice sites"; to splicing events in which they are used as "latent splicing"; and to the resulting mRNA as "latent RNA." Examination of the intronic sequence between the latent and the upstream normal 5′ splice sites revealed the occurrence of stop codons in the reading frame of the gene. We have therefore speculated that the downstream site was rendered latent because, otherwise, its usage for splicing would have led to the inclusion of stop codons in the reading frame of CAD mRNA and thereby to premature termination of the protein encoded by the CAD gene (Miriami et al., 1994).

Our hypothesis predicted that mutations that eliminate such stop codons would render the downstream latent site active in splicing (Fig. 1). We verified this prediction in two gene systems: (i) CAD and (ii) IDUA—the gene encoding the human lysosomal enzyme α-L-iduronidase, mutations in which are the cause of a group of lysosomal disorders termed mucopolysaccharidosis I; the most severe of these is the Hurler syndrome (Neufeld, 1991). We have thus shown that splicing of pre-mRNAs transcribed in vivo from mutant CAD and IDUA minigenes, in which the intronic in-frame stop codons were replaced by sense codons or eliminated by frame-shift mutations, occurred at both the normal and the latent sites, while splicing of the wild-type pre-mRNAs occurred exclusively at the normal 5′ splice site (Li et al., 2002).

Here we performed a computer search for 5′ splice site consensus sequences within nonredundant complete human gene sequences from GenBank (http://www.fruitfly.org/seq_toools/datasets/Human/coding_data/) and show that latent 5′ splice sites are highly abundant.

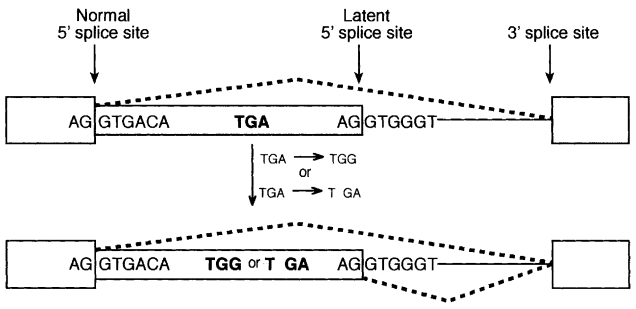

Fig. 1. Activation of latent splicing by mutations that remove in-frame stop codons. A schematic drawing of pre-mRNAs (boxes, exon 1, exon 2; narrow box, extension of exon 1 resulting from splicing at the latent 5′ splice site; solid line, introns) and their expected splicing patterns (heavy dotted lines). Splicing at the latent 5′ splice site of the in-frame stop codon-containing pre-mRNA (upper scheme, TGA) is suppressed. Nonsense to sense mutation of this stop codon by either point or frame-shift mutation (lower scheme, TGA to TGG or T GA) results in activation of splicing from the downstream latent 5′ splice site.

Using statistical analyses we further demonstrate that the abundance of intronic in-frame stop codons upstream of latent 5′ splice sites is significantly higher than that found in introns without latent 5′ splice sites. This finding indicates that the occurrence of in-frame stop codons upstream of latent 5′ splice sites is a general phenomenon that may be involved in the prevention of splicing at latent 5′ splice sites. This general finding conforms with the idea that the cell's nucleus harbors a checking mechanism that is capable of recognizing premature stop codons in pre-mRNAs, resulting in the suppression of splicing at downstream latent 5′ splice sites (Li et al., 2002; Miriami et al., 1994).

2. Methods

2.1. Database

The data were derived from a nonredundant complete human gene sequences data base (http://www.fruitfly.org/seq_toools/datasets/Human/coding_data/). The data were further filtered and 16 genes were excluded based on the criteria that they were described as pseudogenes or partial (Burge et al., 1999). After this filtering, the data set included 446 spliced genes.

2.2. Identification of 5′ splice site signals

A program in C language was written to search for 5′ splice site signals within the database. The criteria for the search for 5′ splice site consensus sequences are detailed under Section 3. Intronic in-frame stop codons refer to the reading frame of the preceding exon.

The identified 5′ splice site consensus sequences were represented as weight matrices and presented in a logo representation produced by the PICTOGRAM program [http://genes.mit.edu/pictogram.html (Lim and Burge, 2001)]. The representation of signals as a weight matrix allows also for the evaluation of its degree of conservation by computing its information content.

3. Results

We have previously shown that splicing at latent 5′ splice sites could be activated upon removal of in-frame stop codons that are located between the normal and the latent sites (Li et al., 2002). This finding provided an experimental proof for our hypothesis that intronic in-frame stop codons play a role in the maintenance of an open-reading frame in mRNAs by rendering downstream 5′ splice sites as latent. To assess the generality of our hypothesis, we first asked whether 5′ splice site consensus sequences are abundant in protein-coding

Table 1
Occurrence of 5′ splice site (SS) consensus sequences[a] in 446 human genes

	Total	Sequences with consensus 5′SS[b]	Number of consensus 5′SS[a]	Sequences without consensus 5′SS[c]
Introns	2311	1601	10 626	710
Exons	2311[d]	429	551	1882

[a] Consensus 5′SSs, excluding the normal 5′SSs.
[b] Number of introns or exons that contain at least one consensus 5′SS.
[c] Number of introns or exons that do not contain consensus 5′SS.
[d] Last exons were excluded from this count.

pre-mRNAs. We thus performed a computer search for 5′ splice site consensus sequences within a nonredundant complete human gene sequences from GenBank. This data set comprises 446 multiexon genes having a total of 2311 introns. The search algorithm allowed for a maximum of two mismatches in the consensus octanucleotide AG/GTRAGT, excluding the dinucleotide GT (underlined), which invariably appears at the 5′ end of U2 introns (Burge et al., 1999; Mount, 1982; Ohshima and Gotoh, 1987; Shapiro and Senapathy, 1987). The use of this criterion is justified because 85.5% of the normal 5′ splice sites in this data set deviate from the consensus by no more than two nucleotides, thereby allowing each of most of the latent 5′ splice sites to compete with the respective normal 5′ splice site for a downstream 3′ splice site. Of the 2311 normal 5′ splice sites, a total of 1977 sites were picked up by this search. The remaining 334 normal 5′ splice sites deviate from the consensus by three (305) or more (29) mismatches. In addition to these 2311 normal 5′ splice sites, the search revealed 10 626 intronic 5′ splice sites consensus sequences which are located within 1601 introns (Table 1), whereas 710 introns are devoid of internal 5′ splice sites consensus sequences. It should be pointed out, that 5′ splice site consensus sequences are significantly less abundant in exon sequences, as only 551 were scored in 429 of the 2311 exons (Table 1).

Next, we show that introns having internal 5′ splice site consensus sequences do not differ from introns lacking such sites with respect to the weight matrices of their normal 5′ splice sites (Figs. 2a and b). To ascertain that these two matrices are not different we applied a χ^2 test to compare the nucleotide distributions throughout the aligned positions. These two weight matrices are not statistically different (P-value 0.95). A χ^2 test to compare these two weight matrices and the weight matrix of the internal 5′ splice site sequences (Fig. 2c) revealed that the nucleotide distributions throughout the aligned positions were not different, except at position +3 which exhibited a preference to G at the latent sites. We also show that introns having internal 5′ splice site consensus sequences do not differ from introns lacking such sites with respect to their phases (zero, one, or two, according to whether the first nucleotide in the given intron is the first, second or third in a coding triplet, respectively). Similarly to previous analysis (Tomita et al., 1996), we show that introns from both groups have a preference to phase zero (Table 2), whereas introns of phase two are less abundant. However, the introns with and without internal 5′ splice site consensus sequences vary significantly with respect to the presence of stop codons, as described below.

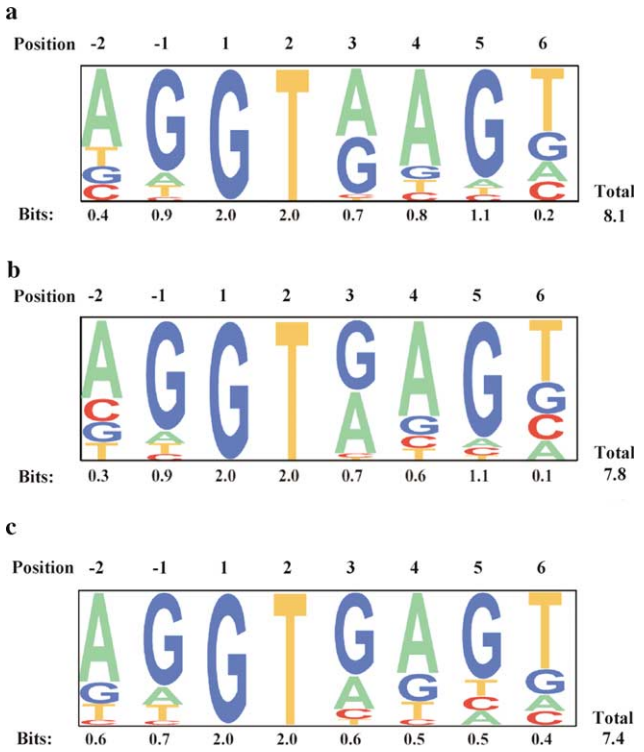

Fig. 2. 5′ Splice site sequence patterns. The sequence pattern at the normal 5′ splice sites for introns with internal 5′-splice site consensus sequences (a), and for introns without internal 5′ splice site consensus sequences (b), as well as the sequence pattern at the internal intronic 5′ splice site consensus sequences (c), are represented as weight matrices. The sequence patterns were displayed by using the PICTOGRAM program [http://genes.mit.edu/pictogram.html; (Lim and Burge, 2001)]. The height of each letter is proportional to the frequency of the corresponding base at the given position, and bases are listed in descending order of frequency from top to bottom. The information content (IC in bits) relative to the background of the intron base composition is also shown.

Table 2
Percentage of intron phases

	Number of introns	Phase		
		Zero	One	Two
With consensus 5′SS	1601	44.72	36.67	18.61
Without consensus 5′SS	710	46.62	35.35	18.03

To consider the possibility that the occurrence of in-frame stop codons upstream of internal 5′ splice site consensus sequences is a significant general phenomenon, which may be involved in the prevention of splicing at such sites, we scored the occurrences of in-frame stop codons upstream to all internal 5′ splice site consensus sequences (Table 3). Splicing at 136 internal 5′ splice site consensus sequences lacking upstream in-frame stop codons, which are located in 105 introns, would not change the reading frame of the resulting mRNAs. These sites are thus potential candidates for alternative splicing, and have therefore been excluded from the statistical analysis described below. The remaining 1496 introns contain 10 490 latent 5′ splice sites—10 045 of which (95.8%) have at least one upstream in-frame stop codon.

Next we asked whether in introns having latent 5′ splice sites, stop codons upstream of the latent sites occur more frequently than in introns lacking such sites. For this analysis we examined a total of 1496 introns that have latent 5′ splice sites and 710 introns that do not. The introns examined are, of course, of different lengths. If N is the length of an intron in nucleotides, its *effective* number of codons (L) is the integral part of $N/3$, $(N-2)/3$, or $(N-1)/3$ (according to the reading phase being zero, one, or two, respectively). The number of in-frame stop codons in each intron of effective length L was scored, and their density was calculated. Then, estimates and standard errors of the density of stop codons were calculated for each group using ratio estimate techniques (Cochran, 1953). For introns having latent 5′ splice sites, the estimate of the density of stop codons was 0.04840 ± 0.00064 stop codons per effective number of codons, whereas for introns that do not have latent 5′ splice sites, the estimate of the density was only 0.03382 ± 0.00110. This difference is highly significant ($P < 0.001$). If, for introns having latent 5′ splice sites, we consider only the upstream part (up to the most 3′ latent site), the estimate of the density is 0.04883 ± 0.00069, which makes the difference between the two types of introns even more pronounced.

A slightly different approach is to consider whether or not each intron has any stop codons at all. If stop codons are distributed independently, the probability that an intron, whose effective number of codons is L, will not have any stop codons is a negative exponential function $e^{-\beta L}$. For any given L, the probability of not having any stop codons is smaller, the larger β is. Out of 1496 introns that have latent 5′ splice sites, 137 do not have any stop codons upstream of the most 3′ latent site (Table 3); and out of 710 introns that do not have latent 5′ splice sites, 167 introns do not have any stop codons. From the data, we derived maximum-likelihood estimates (MLE) of β. For introns having latent 5′ splice sites, the MLE of β was 0.04485 ± 0.00007 (estimate $\pm$ standard error), compared to only 0.03019 ± 0.00006 for introns without latent 5′ splice sites This difference is highly significant ($P < 0.001$).

4. Discussion

In a survey of a database consisting of 446 human protein-coding-genes we identified 10 490 latent 5′ splice sites which are located within 1496 introns. The high abundance of latent 5′ splice sites in the genome indicates that the splicing machine must frequently and efficiently discriminate between normal 5′ splice sites and latent ones, and emphasizes the necessity for a mechanism that identifies such latent sites and renders them as latent. How this discrimination is achieved is not clear yet. However, a clue for a possible mechanism arose from a set of experiments we have previously conducted, showing that splicing involving intronic latent 5′ splice sites could be activated when stop codons, which are located upstream of the latent sites and in-frame with

Table 3
Occurrence of in-frame stop codons upstream of intronic 5′ splice site consensus sequences in 446 human genes

	Number of introns	Number of 5′SS
Introns with stop codons	1359[a]	10 045[b]
Introns without stop codons	137[c]	445[d]
Total introns with latent 5′SSs	**1496**	**10 490**
Introns with potential alternative 5′SSs	**105**	**136**
Total	**1601**	**10 626**

[a] Introns with in-frame stop codon upstream of the most 3′ latent site.
[b] Latent 5′SSs with upstream in-frame stop codons.
[c] Introns without in-frame stop codon upstream of the most 3′ latent site.
[d] Latent 5′SSs without upstream in-frame stop codons.

the preceding exons, were eliminated (Li et al., 2002). We interpreted these results by invoking a nuclear scanning mechanism as an integral component of the pre-mRNA splicing machine, and hypothesized that this mechanism is capable of recognizing in-frame stop codons and suppress splicing at downstream intronic 5′ splice sites, such that the inclusion of premature translation termination codons (PTCs) and their possible deleterious effect are eliminated prior to splicing.

To evaluate the general nature of our hypothesis we first show here that of the 10 490 latent 5′ splice sites scored in our survey, 10 045 (95.8%) have at least one in-frame stop codon in the upstream intronic sequence. Subsequent statistical analyses of the data confirmed that in-frame stop codons are significantly over represented in introns that harbor latent 5′ splice sites, in comparison to introns that lack such sites. Thus, our studies strongly suggest a functional linkage between 5′ splice site selection and the necessity to maintain an open-reading frame in the mRNA. Further, because the occurrence of intronic latent 5′ splice sites, having upstream in-frame stop codons, appears to be widespread in the genome, our study invokes a general mechanism whose role is to identify in-frame stop codons and suppress splicing at latent downstream 5′ splice sites so that the deleterious effects of including PTCs in apparently wild-type transcripts are avoided. We have referred to this mechanism as suppression of splicing, or SOS (Li et al., 2002). Clearly, not all parameters that control splice site selection are known, and the relative importance of those that have already been characterized cannot be assessed as yet. Nonetheless, the relevance of the SOS mechanism for accurate gene expression is highlighted by the fact that 413 genes out of the 446 genes in the data base examined here have at least one intron with latent 5′ splice sites. Of these 413 genes, 404 (98%) have at least one intron with a latent 5′ splice site whose usage for splicing would introduce an in-frame stop codon into the resultant mRNA.

What is the nature of this mechanism? Transcripts arising from genes that underwent somatic or intentional nonsense mutations in *bonafide* exons are frequently subjected to splicing. The resulting PTC-bearing mRNAs are characteristic of many genetic diseases (Culbertson, 1999) and, in most cases, elicit a mechanism known as nonsense-mediated mRNA decay (NMD), whereby the abundance of the aberrant mRNAs is reduced to 1–30% of the normal mRNA level (Culbertson, 1999; Frischmeyer and Dietz, 1999; Hentze and Kulozik, 1999; Hilleren and Parker, 1999; Jacobson and Peltz, 1996; Li and Wilkinson, 1998; Maquat, 1995). It has also been established that, in addition to splicing, NMD is dependent on mRNA translation (reviewed by Maquat, 2002). Using the latter criterion, we have shown that SOS is not affected when translation is inhibited, or upon treatment with drugs that allow translational stop codon readthrough. We have thus excluded the possibility that splicing events involving intronic latent 5′ splice sites occurred in the wild-type transcripts, but the PTC-containing mRNAs thus obtained were subjected to rapid degradation by NMD (Li et al., 2002; Wachtel et al., unpublished).

Another way of eliminating PTCs, introduced by mutations in *bonafide* exons, is by exon skipping or intron retention. These phenomena are also typical of many human genetic diseases (Aoufouchi et al., 1996; Bach et al., 1993; Dietz and Kendzior, 1994; Dietz et al., 1993; Gersappe et al., 1999; Gibson et al., 1993; Lozano et al., 1994; Naeger et al., 1992; Naylor et al., 1993; Sherratt et al., 1993), though several cases of exon skipping have been attributed to the effect of PTCs on splicing control elements (Caputi et al., 2002; Liu et al., 2001; Valentine, 1998). We have also ruled out the possibility that the observed effect of mutating intronic stop codons on splice site selection could have been attributed to modulations of splicing control elements (Li et al., 2002). It seems therefore plausible that SOS requires the recognition of potential PTCs prior to splicing, which invokes the requirement for a nuclear scanning mechanism. Such a mechanism has been proposed before for several cases of exon skipping and intron retention (Aoufouchi et al., 1996; Dietz and Kendzior, 1994; Urlaub et al., 1989), as well as for SOS (Miriami et al., 1994). However, while exon skipping and intron retention eliminate PTCs caused by nonsense mutations in exon sequences, the issue addressed by the SOS mechanism is how stop codons, which occur naturally within intron sequences, can affect splice site selection in such a way that prevents altogether the production of PTC-bearing mRNAs. This issue is of a broad scope because almost every gene harbors a large number of sequences that conform with the consensus for 5′ splice sites but, yet, splicing involving these sites is suppressed. Though PTC-dependent exon skipping, intron retention, and suppression of splicing are phenotypically different, it is plausible that a common nuclear surveillance mechanism underlies these phenomena. Such a surveillance mechanism implies that the reading frame of mRNAs can be recognized at the level of pre-mRNA before splicing commences.

The destabilization of fully spliced nonsense mRNAs by the NMD pathway is dependent on protein synthesis and may involve the recognition of PTCs by ribosomes. The view that intranuclear ribosomes may be involved in nuclear NMD is supported by the occurrence in the nucleus of: (i) charging of tRNA (Lund and Dahlberg, 1998); (ii) translation initiation factors (Dostie et al., 2000); (iii) coupled transcription and translation (Iborra et al., 2001), and by cap-binding complex (CBC)-supported pioneer round of translation (Fortes et al., 2000; Ishigaki et al., 2001). In contrast, the SOS mechanism does not seem to involve ribosomes because it is not

affected by the protein synthesis inhibitor cycloheximide, neither by the aminoglycoside antibiotic that allows stop codon readthrough (G-418) (Li et al., 2002; Wachtel et al., unpublished), which have been shown to abrogate NMD. Nonetheless, the involvement of ribosomal components, or subcomplexes, in SOS cannot be excluded at this stage.

Whatever the nature of the nuclear scanning machine is, it would be difficult to envisage its function on a pre-mRNA because the presence of introns, whose length is not an integer multiple of three, alters, the reading frame. We proposed that the large nuclear ribonucleoprotein (lnRNP) particles, in which nuclear pre-mRNAs are assembled and processed (Medalia et al., 2002; Müller et al., 1998; Raitskin et al., 2001; Sperling et al., 1997), might perform this task (Li et al., 2002; Miriami et al., 1994). The lnRNP particle, which functions as a supraspliceosome complex, serves as a frame onto which the pre-mRNA is folded to align exons about to be spliced, while introns are looped out of each of the respective spliceosomes. This configuration allows exon sequences to be scanned consecutively even when introns are still part of the pre-mRNA. Further characteristics of this mechanism, which should constitute a part of the pre-mRNA splicing machine, have yet to be resolved.

Acknowledgments

We thank Dr. Hanah Margalit for her advice and Ms. Anna Baron for assistance. This research was supported by grants to R.S. from the Israel Science Foundation, to J.S. from the Leo and Julia Forchheimer Center for Molecular Genetics at the Weizmann Institute, and to E.M from the Lady Davis fellowship fund. Correspondance and requests for materials should be addressed to R.S.

References

Aoufouchi, S., Yelamos, J., Milstein, C., 1996. Nonsense mutations inhibit RNA splicing in a cell-free system: Recognition of mutant codon is independent of protein synthesis. Cell 85, 415–422.

Bach, G., Moskowitz, S.M., Tieu, P.T., Matynia, A., Neufeld, E.F., 1993. Molecular analysis of Hurler syndrome in Druze and Muslim Arab patients in Israel: Multiple allelic mutations of the IDUA gene in a small geographic area. Am. J. Hum. Genet. 53, 330–338.

Black, D.L., 1995. Finding splice sites within a wilderness of RNA. RNA 1, 763–771.

Burge, C.B., Tuschl, T.H., Sharp, P.A., 1999. Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 525–560.

Caputi, M., Kendzior Jr., R.J., Beemon, K.L., 2002. A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. Genes Dev. 16, 1754–1759.

Cochran, W.G., 1953. Sampling Techniques. Wiley, New York.

Culbertson, M.R., 1999. RNA surveillance. Trends Genet. 15, 74–80.

Dietz, H.C., Kendzior, R.J., 1994. Maintenance of an open reading frame as an additional level of scrutiny during splice site selection. Nat. Genet. 8, 183–188.

Dietz, H.C., Valle, D., Francomano, C.A., Kendzior, R.J., Pyeritz, R.E., Cutting, G.R., 1993. The skipping of constitutive exons in vivo induced by nonsense mutations. Science 259, 680–683.

Dostie, J., Lejbkowicz, F., Sonenberg, N., 2000. Nuclear eukaryotic initiation factor 4E (eIF4E) colocalizes with splicing factors in speckles. J. Cell Biol. 148, 239–247.

Fortes, P., Inada, T., Preiss, T., Hentze, M.W., Mattaj, I.W., Sachs, A.B., 2000. The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. Mol. Cell 6, 191–196.

Frischmeyer, P.A., Dietz, H.C., 1999. Nonsense-mediated mRNA decay in health and disease. Hum. Mol. Genet. 8, 1893–1900.

Gersappe, A., Burger, L., Pintel, D.J., 1999. A premature termination codon in either exon of minute virus of mice P4 promoter-generated pre-mRNA can inhibit nuclear splicing of the intervening intron in an open reading frame-dependent manner. J. Biol. Chem. 274, 22452–22458.

Gibson, R.A., Hajianpour, A., Murer, O.M., Buchwald, M., Mathew, C.G., 1993. A nonsense mutation and exon skipping in the Fanconi anaemia group C gene. Hum. Mol. Genet. 2, 797–799.

Green, M.R., 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. 7, 559–599.

Hentze, M.W., Kulozik, A.E., 1999. A perfect message: RNA surveillance and nonsense-mediated decay. Cell 96, 307–310.

Hilleren, P., Parker, R., 1999. Mechanisms of mRNA surveillance in eukaryotes. Annu. Rev. Genet. 33, 229–260.

Iborra, F.J., Jackson, D.A., Cook, P.R., 2001. Coupled transcription and translation within nuclei of mammalian cells. Science 293, 1139–1142.

Ishigaki, Y., Li, X., Serin, G., Maquat, L.E., 2001. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell 106, 607–617.

Jacobson, A., Peltz, S.W., 1996. Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. Annu. Rev. Biochem. 65, 693–739.

Krämer, A., 1996. The structure and function of proteins involved in pre-mRNA splicing. Annu. Rev. Biochem. 65, 367–409.

Li, B., Wachtel, C., Miriami, E., Yahalom, G., Friedlander, G., Sharon, G., Sperling, R., Sperling, J., 2002. Stop codons affect 5′ splice site selection by surveillance of splicing. Proc. Natl. Acad. Sci. USA 99, 5277–5282.

Li, S., Wilkinson, M.F., 1998. Nonsense surveillance in lymphocytes? Immunity 8, 135–141.

Lim, L.P., Burge, C.B., 2001. A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci USA 98, 11193–11198.

Liu, H.X., Cartegni, L., Zhang, M.Q., Krainer, A.R., 2001. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat. Genet. 27, 55–58.

Lozano, F., Maertzdorf, B., Pannell, R., Milstein, C., 1994. Low cytoplasmic mRNA levels of immimoglobulin κ light chain genes containing nonsense codons correlate with inefficient splicing. EMBO J. 13, 4617–4622.

Lund, E., Dahlberg, J.E., 1998. Proofreading and aminoacylation of tRNAs before export from the nucleus. Science 282, 2082–2085.

Maquat, L.E., 1995. When cells stop making sense: Effects of nonsense codons on RNA metabolism in vertebrate cells. RNA 1, 453–465.

Maquat, L.E., 2002. NASty effects on fibrillin pre-mRNA splicing: Another case of ESE does, but proposals for translation-dependent splice site choice live on. Genes Dev. 16, 1743–1753.

Medalia, O., Typke, D., Hegerl, R., Angenitzki, M., Sperling, J., Sperling, R., 2002. Cryoelectron microscopy and cryoelectron tomography of the nuclear pre-mRNA processing machine. J. Struct. Biol. 138, 74–84.

Miriami, E., Sperling, J., Sperling, R., 1994. Heat shock affects 5′ splice site selection, cleavage and ligation of CAD pre-mRNA in hamster cells, but not its packaging in InRNP particles. Nucleic Acids Res. 22, 3084–3091.

Mount, S.M., 1982. A catalogue of splice junction sequences. Nucleic Acids Res. 10, 459–472.

Müller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., Sperling, R., 1998. A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. J. Mol. Biol. 283, 383–394.

Naeger, L.K., Schoborg, R.V., Zhao, Q., Tullis, G.E., Pintel, D.J., 1992. Nonsense mutations inhibit splicing of MVM RNA in cis when they interrupt the reading frame of either exon of final spliced product. Genes Dev. 6, 1107–1119.

Naylor, J.A., Green, P.M., Rizza, C.R., Giannelli, F., 1993. Analysis of factor VIII mRNA reveals defects in everyone of 28 haemophilia A patients. Hum. Mol. Genet. 2, 11–17.

Neufeld, E.F., 1991. Lysosomal storage diseases. Annu. Rev. Biochem. 60, 257–280.

Ohshima, Y., Gotoh, Y., 1987. Signals for the selection of a splice site in pre-mRNA computer analysis of splice junction sequences and like sequences. J. Mol. Biol. 195, 247–259.

Padgett, R.A., Wahl, G.M., Stark, G.R., 1982. Structure of the gene for CAD, the multifunctional protein that initiates UMP synthesis in Syrian hamster cells. Mol. Cell. Biol. 2, 293–301.

Raitskin, O., Cho, D.S., Sperling, J., Nishikura, K., Sperling, R., 2001. RNA editing activity is associated with splicing factors in lnRNP particles: The nuclear pre-mRNA processing machinery. Proc. Natl. Acad. Sci. USA 98, 6571–6576.

Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15, 7155–7174.

Sherratt, T.G., Vulliamy, T., Dubowitz, V., Sewry, C.A., Strong, P.N., 1993. Exon skipping and translation in patients with frameshift deletions in the dystrophin gene. Am. J. Hum. Genet. 53, 1007–1015.

Sperling, R., Koster, A.J., Melamed-Bessudo, C., Rubinstein, A., Angenitzki, M., Berkovitch-Yellin, Z., Sperling, J., 1997. Three-dimensional image reconstruction of large nuclear RNP (lnRNP) particles by automated electron tomography. J. Mol. Biol. 267, 570–583.

Staley, J.P., Guthrie, C., 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. Cell 92, 315–326.

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

Urlaub, G., Mitchell, P.J., Ciudad, C.J., Chasin, L.A., 1989. Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. Mol. Cell. Biol. 9, 2868–2880.

Valentine, C.R., 1998. The association of nonsense codons with exon skipping. Mutat. Res. 411, 87–117.

Wachtel, C., Li, B., Sperling, J., Sperling, R. Stop codon mediated suppression of latent splicing is a nuclear scanning mechanism not affected by elements of protein synthesis and NMD, in preparation.