Alternative splicing: current perspectives

Eddo Kim,[†] Amir Goren,[†] and Gil Ast*

Summary

Alternative splicing is a well-characterized mechanism by which multiple transcripts are generated from a single mRNA precursor. By allowing production of several protein isoforms from one pre-mRNA, alternative splicing contributes to proteomic diversity. But what do we know about the origin of this mechanism? Do the same evolutionary forces apply to alternatively and constitutively splice exons? Do similar forces act on all types of alternative splicing? Are the products generated by alternative splicing functional? Why is ''improper'' recognition of exons and introns allowed by the splicing machinery? In this review, we summarize the current knowledge regarding these issues from an evolutionary perspective. ***BioEssays*** **30:38–47, 2008.**

Introduction

Splicing is the process by which introns are removed from an mRNA precursor (pre-mRNA) and exons are ligated to form a mature mRNA.[1] Most types of splicing, in organisms ranging from yeast to human, take place within the spliceosome—a large complex composed of five ribonucleoproteins (RNPs) containing the small nuclear RNAs (snRNAs) U1, U2, U4, U5 and U6 and as many as 150 proteins.[2–5] The splicing machinery recognizes exons and introns by using multiple signals, which presumably results in a network of interactions across exons and/or introns; this recognition is known as exon definition and intron definition, respectively.[6] The four main splice signals that delineate the proper exon–intron boundaries are (1) the 5′ and (2) the 3′ splice sites (5′ss and 3′ss), located at the upstream and downstream exon–intron junctions, respectively, (3) the branch site (BS), and (4) the polypyrimidine tract, which is located upstream of the 3′ss[1,7] (Fig. 1A).

In metazoans, these four splice signals are not sufficient for the recognition of exons and introns by the splicing machinery; it has been estimated that these splicing signals provide only one-half of the information required.[8] So what other factors assist the splicing machinery in precisely recognizing the exon–intron junctions? Studies of the molecular basis of splicing revealed the existence of exonic and intronic cis-acting regulatory sequences that bind trans-acting factors and thus influence splice-site selection. These cis-acting elements are relatively short, usually 4–18 nucleotides, and are classified as exonic or intronic splicing enhancers or silencers. These regulatory elements are required for constitutive splicing and for the regulation of alternative splicing.[9–14] Specific binding of splicing regulatory proteins (such as SR and hnRNP proteins) to these splicing regulatory elements assists in the placement of the spliceosome on the appropriate splice sites.[15,16]

Alternative splicing is a mechanism, by which more than one mRNA transcripts are generated from the same mRNA precursor[17] due to variations in the incorporation of coding regions, giving rise to functionally different proteins.[1,18] Alternative splicing of untranslated (UTR) regions can also determine mRNA localization and stability, as well as efficiency of translation.

Types of alternative splicing

Alternative splicing events are classified into four main subgroups: (1) exon skipping (cassette exons), where the exon can be spliced out of the transcript together with its flanking introns, (2) alternative 5′ss and (3) 3′ss selection, which are the results of the recognition of two or more splice sites at one end of an exon, and (4) intron retention, in which an intron can remain in the mature mRNA molecule[19,20] (Fig. 1B). Finally, there are other, less frequent, complex events that give rise to alternative transcript variants, including mutually exclusive events, alternative transcription start sites and multiple polyadenylation sites.[1,21–24]

Bioinformatic analyses indicate that higher eukaryotes exhibit a higher proportion of alternatively spliced genes

Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv, Israel.
Funding agency: This work was supported by a grant from the Israel Science Foundation (1449/04 and 40/05), MOP Germany-Israel, GIF, DIP, the chief scientist of the Health Ministry; ICA through the Ber-Lehmsdorf Memorial Fund and EURASNET to GA. EK is a fellow of the Clore Scholars Programme. AG is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.
[†]Both authors contributed equally to this work.
*Correspondence to: Gil Ast, Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv 69978, Israel. E-mail: gilast@post.tau.ac.il
DOI 10.1002/bies.20692
Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: 5′ss, 5′ splice site; 3′ss, 3′ splice site; PTC, premature termination codons; NMD, nonsense-mediated mRNA decay.

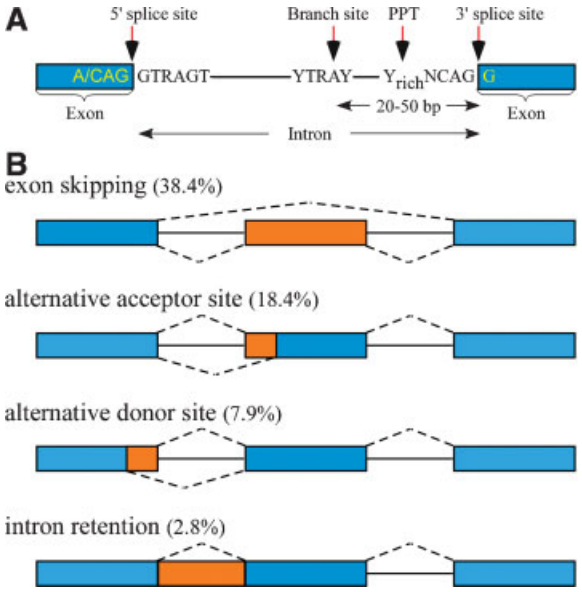

Figure 1. Types of alternative splicing. **A:** The four basal splice signals are depicted: 5′ splice site, branch site, polypyrimidine tract (PPT), and 3′ splice site. **B:** The four main types of alternative splicing are illustrated: exon skipping, alternative acceptor site selection, alternative donor site selection, and intron retention. The relative prevalence of each type in alternative exons conserved in human and mouse is shown in parenthesis. The remaining 32.5%, which are not shown, represent more complex alternative splicing events. Constitutive exons are shown in blue; alternatively spliced regions in orange; introns are represented by solid lines; and dashed lines indicate splicing options.

(genes that generate more than one type of mRNA molecule), and alternatively spliced exons than do lower metazoans.[19,20] In particular, more than 70* of human genes undergo alternative splicing (reviewed in Ref. 25). Alternative splicing does exist in lower metazoans, as well as in fungi[26–29] and in the protozoan *Dictyostelium discoideum*.[30] However, these studies indicate that alternative splicing events in these species are extremely rare. In these organisms, the most-prevalent type of alternative splicing was found to be intron retention, which was found to be the rarest alternative splicing event in vertebrates and invertebrates (<5%)[19,20,31] and may indicate that these are cases of mis-splicing.[25] However, exon skipping, which is the most-prevalent type of alternative splicing in vertebrates and invertebrates (~30–40%),[19,20] is the rarest (if not absent) form in these organisms. Thus, alternative splicing is believed to be a major source for the phenotypic complexity in higher eukaryotes. Indeed, exon skipping exhibits a gradual increase in its relative prevalence along the eukaryotic tree, suggesting it is the more-important event in shaping phenotypic complexity. Following this line, alternative splicing might conceivably explain part of the discrepancy between the number of human protein-coding genes (~25,000), which is only slightly higher than the numbers in nematode (~19,000 genes) and lower than in rice (~40,000 genes), and the phenotypic complexity.[32–34]

Plants exhibit low levels of alternatively spliced genes in general, but exhibit high relative levels of intron retention (~30%) and a very low level of exon skipping (<5%).[19,20,35] From an evolutionary perspective, this suggests that alternative splicing played a less-prominent role in plant evolution than in vertebrate and invertebrate evolution where alternative splicing greatly enhances transcriptomic and proteomic diversity. This could be explained by the fact that plants exhibit extensive whole/partial genome duplication events. Gene duplication gives rise to a state of genetic redundancy, in which one of the newly formed gene copies enters a period of reduced evolutionary pressure. Selective constraints ensure that one of the duplicates retains its original function, while the second copy is free from these constraints and, thus, accumulates mutations, which in turn may lead to a different expression pattern or to neofunctionalization that advances organism speciation.[36–38] Indeed, it was shown that whole genome duplication in plants is associated with speciation.[37,39] This presumably led to a reduced need for alternative splicing and, in particular, for exon skipping, as an essential mechanism for genomic innovations leading to speciation. Indeed, Yanai and colleagues found an inverse correlation between the size of a gene's family and its use of alternatively spliced isoforms.[40] Moreover, Gu and coworkers revealed loss of alternative splicing after gene duplication.[41] This suggests that exon duplication and alternative splicing are interchangeable evolutionary mechanisms and that the requirement for diversification may be satisfied by either of the two mechanisms.[40]

Although the importance of alternative splicing in higher eukaryotes is well established, the prevalence and presence of alternative splicing (and regulated splicing in general) in lower eukaryotes is unclear. Recent findings in yeast (*C. neoformans*) revealed evidence for a variety of alternative splicing events in 4.2% of the genes, including exon skipping and selection of alternative 5′ and 3′ splice sites.[42] This finding supports reports regarding the presence of alternative splicing in several yeasts and protozoa.[26–30] Such studies shed light on the origin of alternative splicing, with recent estimates suggesting that the origin of multi-intron genes dates back to ancient eukaryotes.[43]

Different evolutionary constraints act on alternative cassette and constitutive exons

Exon skipping is the most-prevalent type of alternative splicing in higher eukaryotes, and hence its regulation has been the subject of many studies over the last decade. One of the main methods used for detection of the regulatory mechanisms of alternative splicing is comparative genomics. The idea underlying a

computational search for regulatory mechanisms via comparative genomics is that selective pressure causes functional sequences to evolve at a slower rate or to even remain intact through evolution. Comparative genomics of human and mouse is often used because both genomes are fully sequenced and their transcriptomes are represented by millions of partially sequenced mRNA sequences called expressed sequence tags (ESTs) and fully sequenced mRNA sequences called cDNAs. Furthermore, the human and mouse lineages diverged between 96 and 110 million years ago, providing sufficient time for mutations to accumulate, so that functional sequences are revealed by conservation.[44–46]

Based on human–mouse comparative analyses of constitutive and alternative cassette exons, it has been demonstrated that alternatively spliced exons are under different selective pressures than constitutive ones. These constraints are evolutionary forces that are related to the mechanisms that ensure the regulation of alternative splicing. To date, more than ten characteristics are known to distinguish alternatively from constitutively spliced exons (Fig. 2). Exons that are alternatively spliced in both human and mouse are more conserved than constitutive exons. This probably reflects the need of cassette exons to rely on exonic splicing regulatory sequences for exon selection and alternative splicing regulation.[14,47,48] This conservation is higher toward exon edges, presumably to assist in placement of the basal splicing machinery on the correct exon–intron junctions. The sequence conservation also extends into the flanking introns of conserved alternative cassette exons for 80 to 100 nucleotides, compared to constitutive ones, where the conservation level is significantly lower. This presumably indicates that cassette exons rely on regulatory sequences in their flanking introns for the proper regulation of alternative splicing as well.[20,49–52] Another important distinguishing characteristic is the weak splice sites of cassette exons compared to the strong ones in constitutive exons. In the case of the 5′ss, it was shown that strong base pairing with the U1 snRNP leads to constitutive splicing, whereas weak or suboptimal binding gives rise to alternative splicing. Also, different levels of inclusion are dependent on the binding affinity to U1. Thus, the type of the 5′ss dictates constitutive or alternative splicing and also the level of inclusion/skipping in alternative splicing.[53,54] Alternative cassette exons are also shorter[55,56] and are flanked by longer introns than constitutively spliced ones.[19,57] The longer intronic sequences flanking these short cassette exons presumably obstruct the ability of the splicing machinery to recognize these exons. Altogether these observations imply that cassette exons are suboptimal for recognition by the splicing machinery.

These characteristics also shed light on the evolutionary history of cassette exons. The suboptimal signals of these exons (short, weak splice sites, located between long flanking

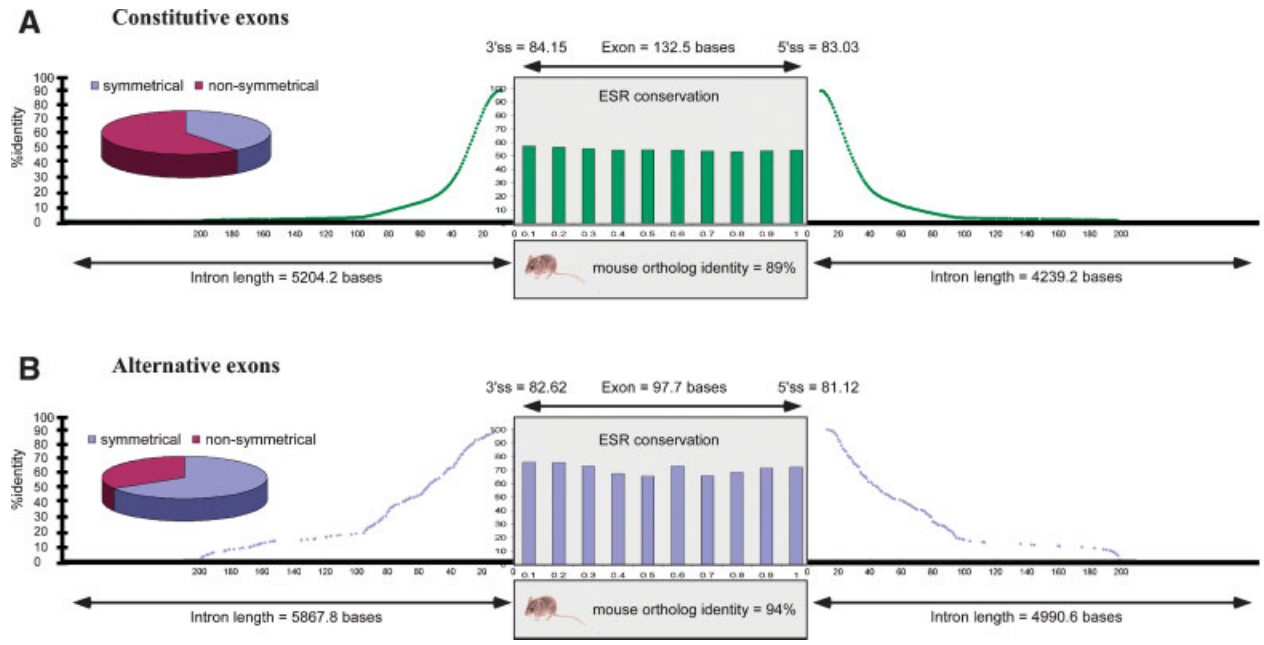

Figure 2. Certain characteristics distinguish conserved alternative exons from constitutively spliced ones. The main features that differ between **A:** constitutively and **B:** alternatively spliced exons that are conserved in human and mouse are illustrated, namely, exon length, splice site strength, exonic splicing regulatory sequence (ESR) conservation, percent identity between human and mouse, length of flanking introns and their conservation level between human and mouse, and the fraction of symmetrical exons.

introns) presumably implies that the splicing machinery must rely more on exonic and intronic regulatory sequences for exon selection and alternative splicing regulation. Indeed, exonic splicing regulatory sequences were found to be significantly more conserved in alternatively spliced exons compared with constitutively spliced ones.[58] This could explain why alternative exons are more conserved than constitutive ones, especially in the wobble position, since maintaining the protein sequence is not the only selective pressure acting on these exons.[48] This suggests another characteristic that distinguishes alternative cassette exons from constitutive ones—the ratio between non-synonymous (Ka) and synonymous mutations (Ks). This ratio was previously proposed for detecting exons in genomic regions through the Ka/Ks ratio test,[59] based on the assumption that a strong purifying selection is acting on most of the protein-coding regions in the human genome. Thus, when comparing human–mouse coding regions, the rates of synonymous mutations should be much higher than non-synonymous ones, yielding Ka/Ks ratios that are significantly lower than one. However, since alternative exons exhibit a low Ks rate due to the high conservation level dictated by regulatory sequences, this ratio test might not detect alternative exons effectively, as was indeed discovered by Xing and Lee.[60] Hence, this ratio test may be used as a distinguishing characteristic between the two exon types. It is worth noting that the literature shows contradicting results. A recent publication indicates that the level of Ks is actually higher in alternative exons compared to constitutive ones, indicating a fast evolution rate in alternatively spliced coding regions.[61]

Another characteristic that distinguishes alternative from constitutive exons is the percentage of symmetrical exons.[56,62] Symmetrical exons are defined as exons comprising a divisible-by-three number of base pairs. Hence, exclusion or inclusion of such exons would maintain the reading frame of the downstream exons. This selective force is probably the strongest one acting on alternative splicing in general, and on exon skipping, in particular, and is further discussed below in detail.

Origin and characteristics of alternative 3′ss and 5′ss selection and intron retention

The regulatory factors that govern alternative 3′ss and 5′ss selection and intron retention have not been as extensively studied as those governing alternative use of cassette exons. Yet, a recent study revealed that alternative 3′ss and 5′ss selection emerged from constitutive exons. Like alternative cassette exons, alternatively used splice sites have weak splicing signals along with a high conservation of the flanking introns. This supports the assumption that weakening or loss of one of the splicing elements (here, weakening of splice signals) is partially compensated for by an increased dependence on additional splicing signals (for example, regulatory elements in the flanking introns). The constitutive side of exons with alternatively used splice sites, on the other hand, exhibits characteristics similar to constitutive exons, namely, strong splice signals and lower level of conservation in the flanking intron.[63] These suggest that alternative splice site use is an intermediate between constitutive and alternative cassette exons.

Intron retention is the least common of the four major types of alternative splicing in higher eukaryotes. Intron retention is believed to be the result of intron, rather than exon, definition, associated with failure of the splicing machinery to recognize weak splice sites flanking short introns. A recent study found that the weaker splice sites, which are associated with events of intron retention, cannot be the only explanation for this phenomenon. Specifically, retained introns not only exhibit weaker splice sites, but are also associated with genes harboring short introns, and exhibit higher levels of expression and lower density of exonic splicing silencers and of the intronic splicing enhancer GGG.[31]

These results suggest that alternative splicing due to selection of alternative splice sites is a subgroup of exon skipping, whereas intron retention might reflect mis-splicing.[25]

Origin of alternative cassette exons

Until recently, only two mechanisms were suggested to be responsible for the origin of alternatively spliced exons. Both mechanisms describe the appearance of new exons, which are then spliced alternatively. One of these mechanisms is known as exon shuffling, in which a new exon is inserted into an existing gene or an exon is duplicated within the same gene, and becomes alternatively spliced (Fig. 3A).[64,65] About 10% of all genes contain tandemly duplicated exons and about 10% of mutually exclusive alternatively spliced exons originated from tandemly duplicated exons.[2,66]

The second mechanism for the origin of alternatively spliced exons involves the emergence of alternatively spliced exons following exonization of intronic sequences (Fig. 3B). For example, the primate-specific *Alu* retroelement, which is highly abundant in intronic sequences,[67] contains multiple sites that are similar, but not identical, to real splice sites.[68,69] Several mutations that change pseudo splice sites to real ones can result in the recognition of a part of the *Alu* element as a *bona-fide* exon.[53,70,71] About 4% of human genes contain transposable element motifs in their coding regions, indicating that exons might have originated from the exonization of these elements.[68,69,72–75] A recent analysis indicated that up to 3.6 times more exons originated from transposable elements in the human genome than in the mouse (1824 and 506 exons, respectively), and this difference is primarily attributable to exonization events of the primate-specific *Alu* element.[67] This phenomenon is not restricted to the human and mouse genomes, as other studies revealed similar trends in other species as well.[76–78]

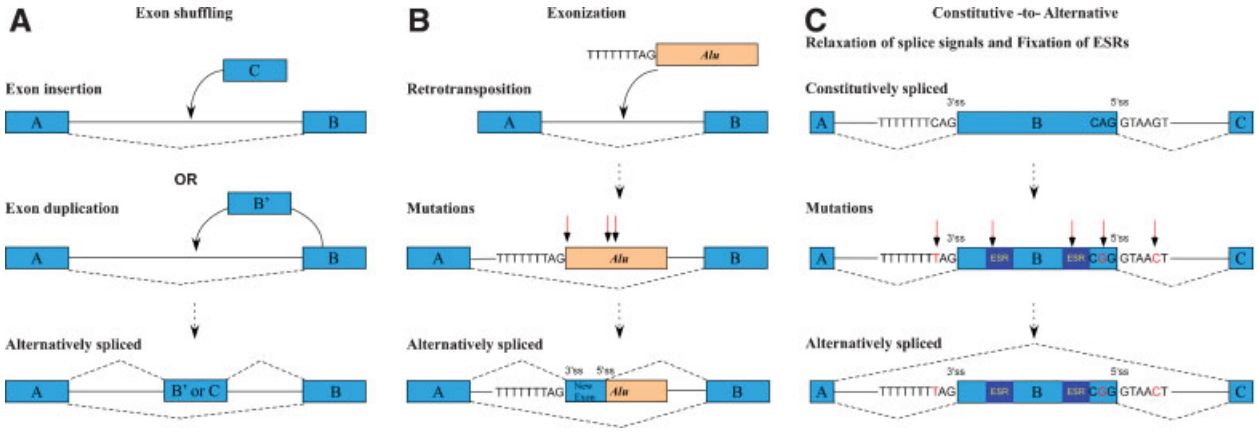

Figure 3. The origin of alternative splicing. Three possible mechanisms have been suggested for the origin of alternatively spliced exons: **A:** exon shuffling, **B:** exonization of transposable elements, such as primate-specific *Alu*s, and **C:** shift in the splicing mode of exons from constitutive to alternative as a result of relaxation of splice signals during evolution, which is coupled to fixation of ESRs.

A recent study provides evidence for a third mechanism for generation of alternative cassette exons.[25,58] While in the two above-described mechanisms, the alternatively spliced exons are generated de novo, in this third mechanism, alternative cassette exons are derived from constitutively spliced ones (Fig. 3C). In fact, the hypothesis that constitutively spliced exons became alternatively spliced is not limited to the creation of alternative cassette exon. As was discussed above, it was recently shown that alternative 5′ss and 3′ss are derived from ancestral constitutive exons. Mutations inside an exon or along the flanking introns were shown to be responsible for the creation of new splice signals that compete with the authentic ones, leading to alternative splice site selection.[63]

Evolution of newly born alternatively spliced exons

Exons that originated from exonization events are alternatively spliced more than could be expected by chance.[67,78] In human, almost all exonized *Alu* elements are alternatively spliced,[67,69] a phenomenon also observed for retroelement exonization events in rodents.[77] In addition, these exons exhibit low inclusion levels; namely, they appear in only a small fraction of the mRNAs transcribed from the corresponding gene.[69,78,79] Thus, the ancestral transcript is maintained as the major form, leaving the minor transcript form almost free of selective pressure and ready for exaptation (acquirement of a new function).[80] Presumably, some of these exons can confer an evolutionary advantage to the organism and therefore become exapted. In such a scenario, the exaptation process is expected to be accompanied by an increase in the inclusion levels of these exons.[75,81,82]

Hence, one could expect alternatively spliced exons that originated from previously constitutively spliced ones to exhibit a high inclusion level in order to maintain the ancestral transcript as the main product. In fact, a recent computational study revealed that alternative cassette exons exhibiting high inclusion levels are typically conserved between human and other mammalian genomes, whereas alternatively spliced exons with low inclusion levels are mostly not.[79] Moreover, previous studies demonstrated that homologous exons that splice differently between human and mouse, termed species-specific exons, exhibit a high inclusion level, which presumably reflects their evolutionary history.[83] Therefore, alternative cassette exons with high inclusion levels probably originated from exons that were previously constitutively spliced, whereas alternative cassette exons with low inclusion levels probably originated from de-novo exonization of intronic sequences.[58] This suggests that the inclusion level may serve as an indicator of the age of alternatively spliced exons and, consequently, of their origin and evolution.

The evolutionary forces that shift exons from constitutive to alternative splicing

A new evolutionary mechanism, by which constitutively spliced exons became alternatively spliced, was recently shown.[58] This mechanism involves mutations that accumulated in the splice sites, leading to their suboptimal recognition by the splicing machinery, and thus skipping of the corresponding exon (Fig. 3C). This mechanism is associated with mutations in exonic and intronic sequences that generate splicing regulatory sequences. These in turn assist in placing the splicing machinery on the correct splice junctions and also regulate the inclusion–skipping ratio of the alternatively spliced exon. We will now examine this mechanism in detail.

Relaxation of 5′ss selection is thought to be a major selective force in alternative splicing. The strength of the 5′ss is measured based on the binding affinity to U1 snRNA. While U1

snRNA is highly conserved, the 5′ss have changed considerably during evolution. The 5′ss of the yeast *S. cerevisiae* (and of Hemiascomycetous fungi, in general) is highly conserved in the intronic portion. On average, it is composed of 6.5 constitutive base pairs, all of which interact with U1. Conversely, the mammalian 5′ss is degenerate and covers nine nucleotides (the last three positions of the exon and the first six intronic positions), seven of which typically pair with U1.[25] The rigid 5′ss site in yeast provides a strong binding site for U1 and presumably explains why *S. cerevisiae* introns are efficiently removed and alternative splicing is rare. However, in mammals, degeneracy at the 5′ss results in many regions that have comparable affinity for U1. In some cases, the binding platforms for U1 are weak, and thus in one mRNA precursor U1 binds to that site and supports splicing of the corresponding exons, whereas in another pre-mRNA molecule generated from the same gene, U1 binds to that site inefficiently or does not bind at all and thus the exon is skipped—which is the major form of alternative splicing in higher eukaryotes.[53,54,84–86] Hence, a strong U1/5'ss is associated with constitutive splicing and a weaker 5′ss (also called a suboptimal site) is associated with alternative splicing.[25] It was therefore proposed that these changes led constitutive exons to evolve into alternative ones and vice versa.

But what was the strength of the 5′ss of the first spliceosomal intron? Irima et al. studied the 5′ss of spliceosomal introns in 49 genomes from a wide variety of eukaryotic lineages. They suggested that the eukaryotic ancestors had relatively weak 5′ss and that some lineages independently underwent 5′ss strengthening. Recent reconstructions of gene structure in the ancestral eukaryote also imply the presence of multi-intron genes.[43] These observations suggest a much earlier emergence of widespread alternative splicing than previously appreciated.[87] This may mean that in lineages where relaxation of the 5′ss led to suboptimal recognition and alternative splicing, the ancestors might have been subjected to opposing evolutionary forces.

The pressure on alternative exons to maintain the reading frame

Many of the human cassette exons with high inclusion levels are also alternatively spliced in the mouse orthologous gene. This indicates that the alternative form emerged before the human/mouse lineages diverged and that the alternative state remained (fixated) during evolution. This conservation implies that there are functional roles for both the exon inclusion and skipping isoforms and that the alternative form is not merely a splicing error. About 66% of alternative cassette exons conserved between human and mouse are symmetrical—that is, the total number of nucleotides in the exon is divisible by three—compared with only 40% of constitutively spliced ones.[56,62] Hence, functional alternative exons tend to maintain the open reading frame.

A similar trend was observed for alternative 5′ and 3′ splice sites selection—the alternative extension was usually symmetrical (63–72%), whereas the exon as a whole exhibited symmetry levels identical to constitutive exons.[63] But why does such selection act mainly on the alternative regions and less on the constitutive ones? Part of the answer is revealed when the origin of these exons is considered. Selection of alternative 5′ss and 3′ss emerged from previously constitutive exons that gained a competing splice site through point mutations.[63] The original splice site is maintained as the major site, whereas the new one is the minor site. One of the major selective forces on fixation of such new events is that the selection of the new splice site will not change the reading frame. Following a possible fixation and gain of function, the inclusion level of the minor form is expected to increase. Thus, splice site selection and inclusion level might tell us something about the evolution of alternative splicing.

Most alternative cassette exons that are conserved between human and mouse exhibit high inclusion levels (~64%).[58] This suggests that these exons originated from constitutive exons, since the inclusion level of newly born exons is very low. Examination of these exons revealed that inclusion level negatively correlates with symmetry.[58] This is further indicative of the selective pressure to maintain the open reading frame, since the lower the inclusion level of the major form, the more alternative transcripts lacking the exon are formed. Thus, there is selective pressure to prevent frameshifts that lead to premature termination codons (PTCs) when inclusion levels are high (Fig. 4). In contrast, recently created alternative cassette exons that were only recently created, as in events of exonization of intronic sequences, exhibit very low inclusion levels (~19%). These newly born human alternative cassette exons do not exhibit a tendency to be symmetrical and 79% of them disrupt the reading frame or introduce PTCs.[55,67]

The majority of functional alternative exons are symmetrical. But what about the ~30% of alternative splicing events, which are conserved between human and mouse, that are non-symmetrical? Such alternative non-symmetrical exons lead to frameshifts and consequently to PTCs and truncated proteins that might be deleterious, for example, by acting in a dominant negative manner (Fig. 4). Since these alternative cassette exons introduce transcripts that contain PTCs, it is somewhat surprising that these splicing events have been maintained for the 96–110 million years since the divergence of the human and mouse lineages. One would expect such transcript variants to disappear in the course of evolution and the alternative exon to shift back to a constitutive state. One mechanism that might relax purifying selection against alternative splicing events that introduce PTCs is the nonsense-mediated mRNA decay (NMD) pathway.[88,89] PTC-containing transcripts are targeted and marked for degradation by NMD and thus their concentration in the cytoplasm is

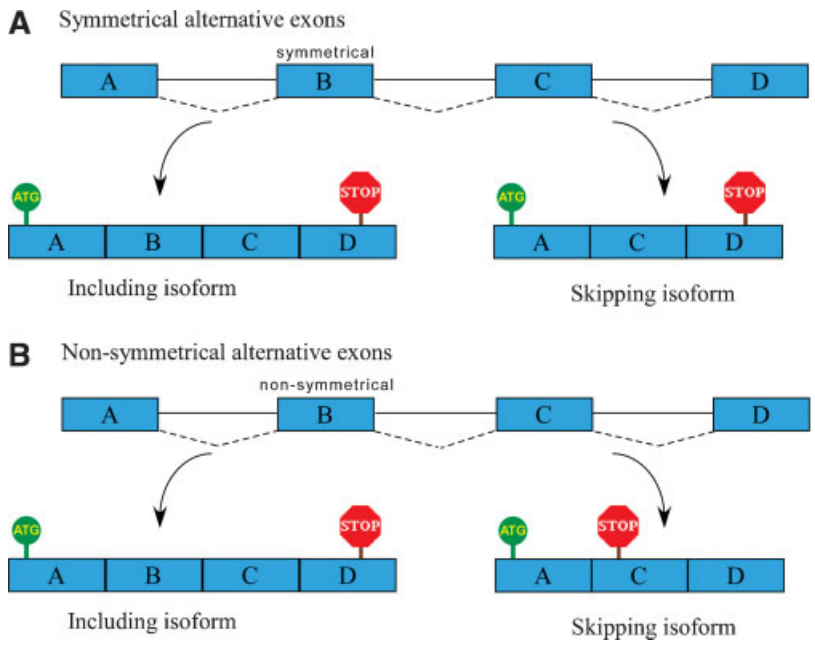

Figure 4. Effects of symmetrical and non-symmetrical alternatively spliced exons on the reading frame. **A:** A schematic of a gene structure consisting of four exons, where the second exon is a symmetrical cassette exon, is depicted. Inclusion or exclusion of the symmetrical exon does not alter the open reading frame of the downstream exons. **B:** A gene consisting of four exons, where the second exon is a non-symmetrical cassette exon, is illustrated. Inclusion or exclusion of the non-symmetrical exon results in a frame-shift and a PTC downstream of the alternative exon.

expected to be low or non-existent.[90] One example for that is the case of polypyrimidine tract binding protein (PTB), where an alternative transcript lacking the non-symmetrical exon 11 is removed by NMD.[91] Also, recent studies from the Ares and Brenner labs revealed that human SR genes have alternatively spliced isoforms that contain PTCs and are degraded by NMD.[92,93] These non-symmetrical and conserved exons of the skipping type are usually observed near the 5′ end of the pre-mRNA, presumably due to selection against similar exons that were located downstream (nearer the 3′ end).[94] This implies that alternatively spliced exons that fail to activate the NMD degradation and thus generate 'poison' transcripts were selected against during evolution.[94] Moreover, recent studies, and the fact that these transcript are present in mRNA datasets, suggest that the role of NMD in targeting these PTC-containing transcripts is not as common as previously estimated.[95-97]

So, what important role do these alternative transcripts play that enabled their fixation through evolution?

Why do cells tolerate noise instead of turning down the volume?

Recently, an ambitious project to identify all functional elements in the human genome was launched. The ENCODE (ENCyclopedia Of DNA Elements) consortium began with a pilot phase that focused on a specified 30 megabases (approximately 1%) of the human genome sequence.[98] The results imply that the identification of all protein-coding genes is nearly completed, but raise many questions regarding random use of off-site promoters to generate non-protein coding transcripts.[99] Moreover, alternative splicing in human genes was found to be much more frequent than has commonly been suggested. However, many of the potential alternative gene products exhibit substantial rearrangements and will have markedly different structures and functions compared with their constitutively spliced counterparts.[100,101] Moreover, the vast majority of these alternative isoforms exhibit low inclusion levels, and little evidence exists to indicate whether they have a role as functional proteins, or whether they are merely "noise" of the splicing machinery.[101]

Several studies have shown that truncated mRNA molecules, which were generated as a result of alternative splicing, are translated into proteins. As a result of use of an alternative splice acceptor site, the MUC1 gene encodes a truncated mRNA transcript, introducing a frame-shift, that is translated into protein.[102] The FMR1 gene exon 14 is non-symmetrical and when spliced out introduces a frame-shift. Interestingly, the absence of exon 14 changed the localization of the transcript from the cytoplasm to the nucleus.[103] Exon skipping is also responsible for a truncated ASBT mRNA that

is translated.[104] However, no biological function for these truncated proteins has yet been demonstrated.

If these alternatively spliced transcripts are indeed noise, are transcripts that contain non-symmetrical alternative exons the only ones that produce such noise? Probably not—several chimeric transcripts that represent gene fusion events were previously reported in the human genome. These very long transcripts originate from the fusion of two consecutive genes as a result of misidentification of the transcription termination site and subsequent continuation of transcription into the following adjacent gene.[105,106] These fused long transcripts generally exhibit unique patterns of splicing and often (>50%) contain PTCs. In-frame PTCs can also be created by non-linear mRNA processing, such as exon repetition (RNA tandem exon(s) duplication) or exon scrambling.[107–110]

So, why do cells tolerate the noise of PTC containing transcripts? In some cases, these transcripts might not be noise at all, but rather these mRNAs might play a regulatory role, balancing between mRNA transcripts that produce a functional protein and mRNA transcripts producing non-functional proteins. This would provide a regulatory layer at the splicing level rather than at the transcription or translation level. Other PTC-producing isoforms are presumably noise and might be tolerated as long as levels are low (low inclusion or skipping level).[101]

Conclusions

We are at the beginning of an exciting era in which sequencing of different genomes and their transcriptomes allows us to examine biological mechanisms and to compare them among different organisms along a known evolutionary tree. The publication of the genomes of human and many other vertebrates, invertebrates, yeast, protozoa and plants, as well as sequences of large numbers of transcripts generated from each organism provides the data for comparative genomics and transcriptomics analyses. We are starting to elucidate the forces that shape the evolution of alternative exons and beginning to identify their origins. Sequencing of genomes and transcripts from more organisms from a broad range of lineages will further our understanding of the alternative splicing mechanism.

Acknowledgments

We thank S. Schwartz for critically reading the manuscript.

References

1. Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72:291–336.
2. Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, et al. 2006. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol Cell Biol 26:5528–5543.
3. Hartmuth K, Urlaub H, Vornlocher HP, Will CL, Gentzel M, et al. 2002. Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci USA 99:16719–16724.
4. Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12:5–14.
5. Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. Nature 419:182–185.
6. Berget SM. 1995. Exon recognition in vertebrate splicing. J Biol Chem 270:2411–2414.
7. Graveley BR. 2001. Alternative splicing: increasing diversity in the proteomic world. Trends Genet 17:100–107.
8. Lim LP, Burge CB. 2001. A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci USA 98:11193–11198.
9. Caceres JF, Kornblihtt AR. 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet 18:186–193.
10. Woodley L, Valcarcel J. 2002. Regulation of alternative pre-mRNA splicing. Brief Funct Genomic Proteomic 1:266–277.
11. Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25: 106–110.
12. Graveley BR. 2000. Sorting out the complexity of SR protein functions. Rna 6:1197–1211.
13. Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298.
14. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. Science 297:1007–1013.
15. Sanford JR, Ellis J, Caceres JF. 2005. Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem Soc Trans 33:443–446.
16. Singh R, Valcarcel J. 2005. Building specificity with nonspecific RNA-binding proteins. Nat Struct Mol Biol 12:645–653.
17. Chabot B. 1996. Directing alternative splicing: cast and scenarios. Trends Genet 12:472–478.
18. Maniatis T, Tasic B. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418:236–243.
19. Kim E, Magen A, Ast G. 2007. Different levels of alternative splicing among eukaryotes. Nucleic Acids Res 35:125–131.
20. Sugnet CW, Kent WJ, Ares M Jr, Haussler D. 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. Pac Symp Biocomput:66–77.
21. Breitbart RE, Andreadis A, Nadal-Ginard B. 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu Rev Biochem 56:467–495.
22. Letunic I, Copley RR, Bork P. 2002. Common exon duplication in animals and its role in alternative splicing. Hum Mol Genet 11:1561–1567.
23. Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. 2000. Patterns of variant polyadenylation signal usage in human genes. Genome Res 10:1001–1010.
24. Suzuki Y, Taira H, Tsunoda T, Mizushima-Sugano J, Sese J, et al. 2001. Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. EMBO Rep 2:388–393.
25. Ast G. 2004. How did alternative splicing evolve? Nat Rev Genet 5:773–782.
26. Okazaki K, Niwa O. 2000. mRNAs encoding zinc finger protein isoforms are expressed by alternative splicing of an in-frame intron in fission yeast. DNA Res 7:27–30.
27. Yatzkan E, Yarden O. 1999. The B regulatory subunit of protein phosphatase 2A is required for completion of macroconidiation and other developmental processes in Neurospora crassa. Mol Microbiol 31:197–209.
28. Ebbole DJ, Jin Y, Thon M, Pan H, Bhattarai E, et al. 2004. Gene discovery and gene expression in the rice blast fungus, Magnaporthe grisea: analysis of expressed sequence tags. Mol Plant Microbe Interact 17:1337–1347.
29. Vilardell J, Chartrand P, Singer RH, Warner JR. 2000. The odyssey of a regulated transcript. Rna 6:1773–1780.
30. Escalante R, Moreno N, Sastre L. 2003. Dictyostelium discoideum developmentally regulated genes whose expression is dependent on MADS box transcription factor SrfA. Eukaryot Cell 2:1327–1335.

31. Sakabe NJ, de Souza SJ. 2007. Sequence features responsible for intron retention in human. BMC Genomics 8:59.
32. Consortium IHGS. 2004. Finishing the euchromatic sequence of the human genome. Nature 431:931–945.
33. Consortium TCeS. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282:2012–2018.
34. Goff SA, Ricke D, Lan TH, Presting G, Wang R, et al. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100.
35. Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, et al. 2004. Intron retention is a major phenomenon in alternative splicing in Arabidopsis. Plant J 39:877–885.
36. Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in Arabidopsis. Science 290:2114–2117.
37. Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3:827–837.
38. Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, et al. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol 23:469–478.
39. Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.
40. Kopelman NM, Lancet D, Yanai I. 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat Genet 37:588–589.
41. Su Z, Wang J, Yu J, Huang X, Gu X. 2006. Evolution of alternative splicing after gene duplication. Genome Res 16:182–189.
42. Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, et al. 2005. The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans. Science 307:1321–1324.
43. Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, et al. 2005. Vertebrate-type intron-rich genes in the marine annelid Platynereis dumerilii. Science 310:1325–1326.
44. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.
45. Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. Nature 392:917–920.
46. Nei M, Xu P, Glazko G. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc Natl Acad Sci USA 98:2497–2502.
47. Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev 18:1241–1250.
48. Goren A, Ram O, Amit M, Keren H, Lev-Maor G, et al. 2006. Comparative analysis identifies exonic splicing regulatory sequences. The complex definition of enhancers and silencers. Mol Cell 22:769–781.
49. Sorek R, Ast G. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13: 1631–1637.
50. Sugnet CW, Srinivasan K, Clark TA, O'Brien G, Cline MS, et al. 2006. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. PLoS Comput Biol 2:e4.
51. Fagnani M, Barash Y, Ip J, Misquitta C, Pan Q, et al. 2007. Functional coordination of alternative splicing in the mammalian central nervous system. Genome Biol 8:R108.
52. Yeo GW, Nostrand EL, Liang TY. 2007. Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. PLoS Genet 3:e85.
53. Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, et al. 2004. Minimal conditions for exonization of intronic sequences: 5′ splice site formation in *Alu* exons. Mol Cell 14:221–231.
54. Roca X, Sachidanandam R, Krainer AR. 2005. Determinants of the inherent strength of human 5′ splice sites. Rna 11:683–698.
55. Sorek R, Shamir R, Ast G. 2004. How prevalent is functional alternative splicing in the human genome? Trends Genet 20:68–71.
56. Sorek R, Shemesh R, Cohen Y, Basechess O, Ast G, et al. 2004. A non-EST-based method for exon-skipping prediction. Genome Res 14: 1617–1623.
57. Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, et al. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci USA 102:16176–16181.
58. Lev-Maor G, Goren A, Sela N, Kim E, Kerem H, Doron-Faigenboim A, Leibman-Barak S, Pupko T, Ast G. The ''Alternative'' Choice of Constitutive Exons throughout Evolution. PLoS Genet 2007;3(11):e203.
59. Nekrutenko A, Makova KD, Li WH. 2002. The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. Genome Res 12:198–202.
60. Xing Y, Lee C. 2005. Assessing the application of Ka/Ks ratio test to alternatively spliced exons. Bioinformatics 21:3701–3703.
61. Ermakova EO, Nurtdinov RN, Gelfand MS. 2006. Fast rate of evolution in alternatively spliced coding regions of mammalian genes. BMC Genomics 7:84.
62. Resch A, Xing Y, Alekseyenko A, Modrek B, Lee C. 2004. Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. Nucleic Acids Res 32:1261–1269.
63. Koren E, Lev-Maor G, Ast G. 2007. The emergence of alternative 3′ and 5′ splice site exons from constitutive exons. PLoS Comput Biol 3:e95.
64. Gilbert W. 1978. Why genes in pieces? Nature 271:501.
65. Kondrashov FA, Koonin EV. 2001. Origin of alternative splicing by tandem exon duplication. Hum Mol Genet 10:2661–2669.
66. Kondrashov FA, Koonin EV. 2003. Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. Trends Genet 19:115–119.
67. Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. Comparative analysis of transposed element insertion within human and mouse genes reveals *Alu*'s unique role in shaping the human transcriptome. Genome Biol 2007;8(6):R127.
68. Makalowski W, Mitchell GA, Labuda D. 1994. *Alu* sequences in the coding regions of mRNA: a source of protein variability. Trends Genet 10:188–193.
69. Sorek R, Ast G, Graur D. 2002. Alu-containing exons are alternatively spliced. Genome Res 12:1060–1067.
70. Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. Science 300:1288–1291.
71. Krull M, Brosius J, Schmitz J. 2005. Alu-SINE exonization: en route to protein-coding function. Mol Biol Evol 22:1702–1711.
72. Nekrutenko A, Li WH. 2001. Transposable elements are found in a large number of human protein-coding genes. Trends Genet 17:619–621.
73. Singer SS, Mannel DN, Hehlgans T, Brosius J, Schmitz J. 2004. From ''junk'' to gene: curriculum vitae of a primate receptor isoform gene. J Mol Biol 341:883–886.
74. Gotea V, Makalowski W. 2006. Do transposable elements really contribute to proteomes? Trends Genet 22:260–267.
75. Zhang XH, Chasin LA. 2006. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. Proc Natl Acad Sci USA 103:13427–13432.
76. Wang W, Kirkness EF. 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res 15:1798–1808.
77. Wang W, Zheng H, Yang S, Yu H, Li J, et al. 2005. Origin and evolution of new exons in rodents. Genome Res 15:1258–1264.
78. Alekseyenko AV, Kim N, Lee CJ. 2007. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. Rna 13:661–670.
79. Modrek B, Lee CJ. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet 34:177–180.
80. Brosius J, Gould SJ. 1992. On ''genomenclature'': a comprehensive (and respectful) taxonomy for pseudogenes and other ''junk DNA''. Proc Natl Acad Sci USA 89:10706–10710.
81. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441:87–90.
82. Xing Y, Lee C. 2005. Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc Natl Acad Sci USA 102:13526–13531.

83. Pan Q, Bakowski MA, Morris Q, Zhang W, Frey BJ, et al. 2005. Alternative splicing of conserved exons is frequently species-specific in human and mouse. Trends Genet 21:73–77.
84. Stamm S, Zhang MQ, Marr TG, Helfman DM. 1994. A sequence compilation and comparison of exons that are alternatively spliced in neurons. Nucleic Acids Res 22:1515–1526.
85. Carmel I, Tal S, Vig I, Ast G. 2004. Comparative analysis detects dependencies among the 5′ splice-site positions. Rna 10:828–840.
86. Freund M, Hicks MJ, Konermann C, Otte M, Hertel KJ, et al. 2005. Extended base pair complementarity between U1 snRNA and the 5′ splice site does not inhibit splicing in higher eukaryotes, but rather increases 5′ splice site recognition. Nucleic Acids Res 33:5112–5119.
87. Irimia M, Penny D, Roy SW. 2007. Coevolution of genomic intron number and splice sites. Trends Genet 23:321–325.
88. Hillman RT, Green RE, Brenner SE. 2004. An unappreciated role for RNA surveillance. Genome Biol 5:R8.
89. Lejeune F, Maquat LE. 2005. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. Curr Opin Cell Biol 17:309–315.
90. Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci USA 100:189–192.
91. Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW. 2004. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell 13:91–100.
92. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. 2007. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446:926–929.
93. Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, et al. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev 21:708–718.
94. Magen A, Ast G. 2005. The importance of being divisible by three in alternative splicing. Nucleic Acids Res 33:5574–5582.
95. Pan Q, Saltzman AL, Kim YK, Misquitta C, Shai O, et al. 2006. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. Genes Dev 20:153–158.
96. Baek D, Green P. 2005. Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. Proc Natl Acad Sci USA 102:12813–12818.
97. Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. Cell 126:37–47.
98. Consortium EP. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306:636–640.
99. Consortium TEP. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816.
100. Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, et al. 2007. Prominent use of distal 5′ transcription start sites and discovery of a large number of additional exons in ENCODE regions. Genome Res 17:746–759.
101. Tress ML, Martelli PL, Frankish A, Reeves GA, Wesselink JJ, et al. 2007. The implications of alternative splicing in the ENCODE protein complement. Proc Natl Acad Sci USA 104:5495–5500.
102. Levitin F, Baruch A, Weiss M, Stiegman K, Hartmann ML, et al. 2005. A novel protein derived from the MUC1 gene by alternative splicing and frameshifting. J Biol Chem 280:10655–10663.
103. Sittler A, Devys D, Weber C, Mandel JL. 1996. Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. Hum Mol Genet 5:95–102.
104. Lazaridis KN, Tietz P, Wu T, Kip S, Dawson PA, et al. 2000. Alternative splicing of the rat sodium/bile acid transporter changes its cellular localization and transport properties. Proc Natl Acad Sci USA 97: 11092–11097.
105. Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, et al. 2006. Transcription-mediated gene fusion in the human genome. Genome Res 16:30–36.
106. Parra G, Reymond A, Dabbouseh N, Dermitzakis ET, Castelo R, et al. 2006. Tandem chimerism as a means to increase protein complexity in the human genome. Genome Res 16:37–44.
107. Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, et al. 1991. Scrambled exons. Cell 64:607–613.
108. Cocquerelle C, Daubersies P, Majerus MA, Kerckaert JP, Bailleul B. 1992. Splicing with inverted order of exons occurs proximal to large introns. Embo J 11:1095–1098.
109. Frantz SA, Thiara AS, Lodwick D, Ng LL, Eperon IC, et al. 1999. Exon repetition in mRNA. Proc Natl Acad Sci USA 96:5400–5405.
110. Dixon RJ, Eperon IC, Hall L, Samani NJ. 2005. A genome-wide survey demonstrates widespread non-linear mRNA in expressed sequences from multiple species. Nucleic Acids Res 33:5904–5913.