Alternative splicing of mutually exclusive exons—A review

Martin Pohl¹, Ralf H. Bortfeldt², Konrad Grützmann³, Stefan Schuster³

¹ Department of Bioinformatics, Friedrich Schiller University of Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany
² Department of Crop and Animal Sciences, Humboldt University Berlin, Invalidenstrasse 42, 10115 Berlin, Germany

ABSTRACT

Alternative splicing (AS) of pre-mRNAs in higher eukaryotes and several viruses is one major source of protein diversity. Usually, the following major subtypes of AS are distinguished: exon skipping, intron retention, and alternative 3’ and 5’ splice sites. Moreover, mutually exclusive exons (MXEs) represent a rare subtype. In the splicing of MXEs, two (or more) splicing events are not independent anymore, but are executed or disabled in a coordinated manner. In this review, several bioinformatics approaches for analyzing MXEs are presented and discussed. In particular, we revisit suitable definitions and nomenclatures, and bioinformatics tools for finding MXEs, adjacent and non-adjacent MXEs, clustered and grouped MXEs. Moreover, the molecular mechanisms for splicing MXEs proposed in the literature are reviewed and discussed.

© 2013 Elsevier Ireland Ltd. All rights reserved.

Keywords:
Alternative splicing
Exon clusters
Mutually exclusive exons
Non-sense mediated decay
Splicing mechanisms

Contents

1. Introduction .......................................................................................................................... 31
  1.1. Molecular biology background ......................................................................................... 31
  1.2. Bioinformatics resources for analyzing MXE splicing ...................................................... 32
2. Nomenclatures and definitions .............................................................................................. 33
  2.1. Established approaches .................................................................................................... 33
  2.2. Adjacent MXEs and non-adjacent dependencies ............................................................ 34
  2.3. A Boolean nomenclature .................................................................................................. 34
3. Detection .................................................................................................................................. 35
4. Mechanisms leading to mutual exclusion of exons .............................................................. 36
5. Evolutionary conservation .................................................................................................... 36
6. Conclusions ............................................................................................................................ 36
7. Acknowledgements ................................................................................................................. 36
8. References .............................................................................................................................. 36

1. Introduction

1.1. Molecular biology background

In addition to the genetic code, several other codes are used by the living cell at the molecular level, for example, the calcium oscillation code and the code used for signaling among plants by volatile chemicals. In eukaryotes one of these is the splicing code, by which the cell decides which sequence parts are finally used (Choudhary and Krithivasan, 2007; Barbieri, 2008; Barash et al., 2010; Reddy et al., 2012).

In the post-genomic era, alternative splicing (AS) of pre-mRNAs in higher eukaryotes got in the focus of research as one major source of protein diversity (Black, 2000; Graveley, 2001; Kim et al., 2008; Nilsen and Graveley, 2010; Chen et al., 2012a). AS was discovered in adenoviruses (Berget et al., 1977) and also occurs in several other viruses such as cytomegalovirus (Gatherer et al., 2011). Protein variability contributes to a high complexity of higher eukaryotes while keeping the numbers of genes relatively low. AS is a means to change proteins, in dependence on gender, developmental stage or environmental conditions and can affect binding...
properties, intracellular localization, enzymatic activity and many more properties of proteins (Stamm et al., 2005; Yap and Makeyev, 2013). Estimations raised from one third to 95% of human genes affected by AS, with other mammals showing similar high AS levels (Florea, 2006; Pan et al., 2008; Wang et al., 2008). Alternative splicing and splicing in general is a major problem in gene finding in eukaryotes because it may disrupt ORFs (Pohl et al., 2012).

The potential for variability is enormous. For instance, the human calcium-activated potassium channel subunit alpha-1 gene and the three neurexin genes could potentially generate 500 and more than 2000 different protein isoforms, respectively, by different ways of splicing (Black, 1998; Tabuchi and Südhof, 2002). The Drosophila Down Syndrome Cell Adhesion molecule gene (Dscam) has several sets of cassette exons with one of them involving 48 alternative exons among which one is selected (Graveley, 2005; Anastassiou et al., 2006; Mejiers et al., 2007; Olson et al., 2007; Hemani and Soller, 2012; Wang et al., 2012). This leads to 38,016 theoretical splicing variants.

AS is thought to lower selective pressure on gene sequences allowing a higher trial and error rate by mutations in one of the isoforms without compromising the acquired functionality of the other isoform (Boué et al., 2003; Chen et al., 2006; Noh et al., 2006). The apparent evolutionary advantages of AS require, however, significant energetic and metabolic costs because the splicesome, which performs the splicing reaction, is a large complex of proteins and RNA including up to several hundreds of constituents (Jurica and Moore, 2003; Kielbassa et al., 2009; Bortfeldt et al., 2010; Hoskins et al., 2011). Given the enormous effort to assemble such complicated molecular machinery it can be assumed that the benefit of transcript flexibility outweighs the biochemical costs. In contrast, some organisms such as many plants, seem to have achieved their level of protein variability mainly by gene duplications i.e., an increase in genome length (Kopelman et al., 2005).

The ability to cope with stress is widely enhanced via transcriprome plasticity (Mastrangelo et al., 2012). Moreover, the involvement and prevalence of AS in many diseases is becoming increasingly clear. Hence, protein variability as generated by alternative splicing is of great medical and biotechnological importance because different isoforms are often associated with diseases such as cancer (Hernandez-Lopez and Graham, 2012) or with the distinction between intracellular and extracellular enzymes (Andreasri and Riccio, 2009). This renders AS and its regulation a potential therapeutic target (Mount and Pandey, 2005; Garcia-Blanco, 2006; He et al., 2009; Tazi et al., 2009, 2010; Douglas and Wood, 2011; Germann et al., 2012; Hernandez-Lopez and Graham, 2012; Sanchez-Pla et al., 2012).

Several attempts for general AS annotations have been presented (Xing et al., 2004; Nagasaki et al., 2006; Sammeth et al., 2008; Kroll et al., 2012). Among the well-known subtypes of AS are exon skipping (Sorek et al., 2004b), intron retention (Wang et al., 2006), alternative 5’ splice sites (Dou et al., 2006; Bortfeldt et al., 2008; Hiller and Platzer, 2008), alternative 3’ splice sites (Bortfeldt et al., 2008; Hiller and Platzer, 2008). A less abundant subtype of AS is represented by mutually exclusive exon (MXE) splicing.

MXEs are characterized by splicing of exons in a coordinated manner such that two or more splicing events are not independent. As the name “mutually exclusive” indicates, exactly one out of two exons (or one group out of two exon groups) is retained, while the other one is spliced out. Sammeth (2009) applies the term in a less strict way, allowing the case that none or all of the exons under consideration are retained. In contrast to other variants of alternative splicing, mutually exclusive splicing can leave the size of the protein unchanged provided that the exchanged sequence is of the same length and does not introduce a premature stop codon. Depending on the similarity of exchanged exon sequences, minor changes as in subtle alternative 5’ and 3’ splicing events or major changes of whole protein domains as in exon skipping are possible. In case of minor protein sequence changes, MXEs may provide an advantage to many types of proteins, such as ion channels, because the spatial structure is preserved, while the protein exhibits an altered function (Birzele et al., 2008a). Interestingly, another RNA processing mechanism, RNA editing, can also occur in a mutually exclusive manner as shown for the TPP2 gene (Grohmann et al., 2010) resulting in a similar effect as mutually exclusive exon splicing.

A common assumption is that MXEs have originated from exon duplication and, hence, are highly similar (Letunic et al., 2002; Copley, 2004; Sorek, 2009; Pillmann et al., 2011). Accordingly, some authors (Stephan et al., 2007; Pillmann et al., 2011) define MXEs based on similar length and sequence. In our opinion, these criteria are not necessary. The term “mutually exclusive” only implies that exons do not occur together but does not refer to length, sequence or exon numbers. In general, also a group (cluster) of exons can be mutually exclusive with respect to another group (cluster) of exons. Such cases should be distinguished from exon cassettes where exactly one out of several exons is retained in the mature transcript, such as in the Dscam gene in Drosophila. However, the terminology is not used consistently among researchers, MXE were previously also termed as “exon clusters” (Pillmann et al., 2011) or “cassette exons” (Stephan et al., 2007).

MXEs turned out to be very promising candidates for generation of highly diverse but specific processes (Anastassiou et al., 2006; Soom et al., 2008). The alternative selection of exons enables the encoding of a whole class of proteins with similar scaffold and similar length but with highly specific functionality. Beside the above-mentioned Drosophila Dscam gene, examples of biological relevance are provided by the voltage dependence of ion channels (Soom et al., 2008) and calcium sensitivity of muscle proteins in higher animals (Waites et al., 1992). Like other AS types, MXEs proved to be of medical relevance, e.g., at regulation of expression levels of the mammalian pyruvate kinase M isoforms (Chacko and Ranganathan, 2009b; Chen et al., 2012b). Examples of MXEs have been described in human (Soom et al., 2008), mouse (Chacko and Ranganathan, 2009a), rat (Gustafson et al., 1993), chicken (Waites et al., 1992; Chacko and Ranganathan, 2009a), cow (Chacko and Ranganathan, 2009b), nematode (Johnson et al., 2003) and other species.

1.2. Bioinformatics resources for analyzing MXE splicing

As biochemical analyses are expensive and time consuming, computational approaches have attracted an ever increasing interest. Accordingly, AS is an important topic in bioinformatics (Dou et al., 2006; Zavolan and van Nimwegen, 2006; Hiller et al., 2007; Bortfeldt et al., 2008; Hiller and Platzer, 2008; Sammeth et al., 2008; Busch and Hertel, 2012; Chen et al., 2012a; Sanchez-Pla et al., 2012).

To date many resources on AS emerged thanks to the growing amount of sequence and alignment data, in spite of incomplete- ness and considerable noise within the data (Black, 2003; Lareau et al., 2004; Chen et al., 2012a). Relevant databases that emerged in the context of MXE are MAASE (Zheng et al., 2005), HOLLYWOOD (Holste et al., 2006), ASAP II (Kim et al., 2007), ECGene (Lee et al., 2007), Ensembl (including former ASD/ATD/ASTD/AEdb projects (Koscielny et al., 2009), SPLOOCE (Kroll et al., 2012).

Also, the assembly of the splicesome has been described by bioinformatics approaches (Kielbassa et al., 2009; Bortfeldt et al., 2010; Hoskins et al., 2011). Different types of the splicesome were suggested to produce MXE splicing patterns (see Section 4). Beside the major splicesome, a minor splicesome can process splice sites that have distinct consensus sequences and are incompatible with the major splicesome (Will and Lührmann, 2005).

In this review, we discuss several bioinformatics approaches for analyzing MXE splicing. In particular, we will focus on appropriate
**2. Nomenclatures and definitions**

### 2.1. Established approaches

A widely used graphical representation of AS events shows the alignments of transcripts as boxes representing exons connected by individual links for each isoform (Fig. 1). Early in the analysis of AS, it became clear that standardization of the nomenclature for AS forms is important (Zavolan and van Nimwegen, 2006). Since then, some attempts have been made without leading to a broadly used and accepted nomenclature. Nagasaki et al. (2006) introduced number vectors based on bit arrays where exonic regions are denoted by ‘1’ and intronic regions are denoted by ‘0’. Sammeth et al. (2008) suggested a general nomenclature for AS by representing transcripts as sequence of splice sites connected by symbols for exons and exon–exon junctions. Recently, Kroll et al. (2012) introduced a character-based syntax to describe results achieved with the analysis of bit arrays by regular expressions. An overview of these notations is given in Fig. 1.

Even more notations have been suggested in the literature. Malko et al. (2006) used strings of one-character-codes for basic alternative events. Based on the exon–intron structure of isoforms, Riva and Pesole (2009) computed unique signature strings as a basis for an unambiguous nomenclature which facilitates database searches.

One drawback of such nomenclatures is that transcripts are considered and represented only individually, such that possible dependencies between splicing events remain hidden. The resulting codes describing the splice patterns make it presently cumbersome to detect such dependencies because they have to be decoded and compared first. Additionally, for a more comprehensive splicing picture more than two transcript variants must be considered. One solution to this is the search for subgraphs within splicing graph representations of transcript isoforms (Sammeth, 2009).

In the strict definition, MXEs should be perfectly mutually exclusive, as the name suggests. In many studies, only two transcripts are considered, e.g., in two tissues or developmental stages. However, MXEs found in this manner need not be MXEs when taking more abundant transcript data into account. Thus, the term MXE is relative with respect to the abundance of known transcripts at a gene locus, which in turn depends on how many different conditions are studied. This is the case in the example of TCL6, where specific tissues show indeed exclusive patterns while on the basis of all known transcripts the pattern is lost (Sammeth et al., 2008). We suppose that this also applies to KCNMA1 (Soom et al., 2008). The more detailed analysis by Nilson and Graveley (2010) shows that on the basis of more transcripts, the exons of KCNMA1 are not perfectly exclusive, because there are transcripts containing both mutually exclusive exons.

Summarizing these considerations, it might be worthwhile relaxing the strict definition of MXEs in that a certain percentage of non-exclusive events are allowed. In living organisms in general, many exceptions and deviations occur, for example in the number of teeth in humans. The clover plant usually has three leaves with a rare deviation showing four leaves. These deviations do not prevent the individual to be classified in the general type (e.g., human, clover, etc.). Analogously, AS events could be defined as context specific MXEs, allowing a certain percentage of cases, that are not mutually exclusive when compared across different conditions, e.g., developmental stages, tissues or disease states. However,
it is difficult to define a biologically well-founded value for such a threshold.

2.2. Adjacent MXEs and non-adjacent dependencies

In a genome wide study of MXE events in humans, we detected a number of non-adjacent MXEs (Pohl et al., 2009). This contradicts the intuitive assumption that MXEs – which are expected to originate from exon duplication – should usually be in direct genomic neighborhood (Letunic et al., 2002; Copley, 2004). Glauser et al. (2011) found dependencies between individual sites even in the case of non-adjacent alternatively spliced exons in Caenorhabditis elegans slo-1, which is in line with our findings. This is also supported by results presented in the ASTALAVISTA study (Foissac and Sammeth, 2007). Non-adjacent, mutually dependent exons have also been reported recently by Kroll et al. (2012) (see the pattern “−s-E−s−” in Table 3 in that reference). That indicates that only approaches and nomenclatures considering mutual (including long-ranging) dependencies among exons that include a constitutive exon in between will have a chance of success in predicting the full splicing picture, all the more as the line between different AS types is not clear-cut (Sammeth et al., 2008). For example, distal exclusions are a combination of skipped exons, and “Twintrons” are formed by co-occurring alternative 5′ and 3′ ends (Fig. 1).

2.3. A Boolean nomenclature

In case of clustered and/or non-adjacent MXEs, it may be difficult to grasp the dependencies among the exons. Then, it is helpful to formalize the notation. For example, if either an exon A is used or a cluster of two exons B and C, we may write: A ⊕ (B ∧ C) (with ⊕ denoting the exclusive disjunction, XOR). However, it is more convenient to attach indices to splice sites rather than to exons, because the possible overlap of exons (such as in the cases of intron retention or alternative ends) is easier to recognize in that way.

Denoting the splice sites in the above example by 1–6, the notation is E_{1,2} ⊕ (E_{3,4} ∧ E_{5,6}), where E_{ij} stands for an exon between splice sites i and j. The usual case of one pair of MXEs would then be written as E_{1,2} ⊕ E_{3,4}.

Also alternative 5′ and 3′ ends can now be described in a unique way by this notation. Assume, for example, that B includes A and three further nucleotides at the 3′ end, then A ⊕ B describes a tandem donor splice site AS event, not an MXE. Applying splice site indices we obtain the distinct E_{1,2} ⊕ E_{1,3} for the alternative exon end (Fig. 1C–E).

When these Boolean expressions get rather long and complex, it is useful to condense them to the so-called disjoint normal forms. For example, the above-mentioned case of non-adjacent MXEs might be written first as (E_{1,2} ⊕ E_{3,4}) ⊕ E_{5,6}. This can be simplified to the expression (E_{1,2} ⊕ E_{3,4}) ⊕ E_{5,6}.

It is worth noting that not only MXEs, but also other AS events can be described in this way. For example, the notation (E_{1,2} ⊕ E_{5,6}) ⊕ (E_{1,2} ⊕ E_{3,4} ∧ E_{5,6}) describes the usual exon skipping event (Fig. 1A). Interestingly, there are cases that are neither perfect MXEs nor classical exon skipping, but more complex dependencies. Sammeth (2009) considered such dependencies earlier, using another notation, and included more complex dependencies into the class of MXEs (Fig. 6 in Sammeth, 2009), while we advocate to use this term for the “exclusive or” case only (Fig. 1F). For example, E_{1,2} ⊕ E_{3,4} describes the situation where exon A or B can be skipped but also the case may occur where both exons are used. It excludes, however, that both exons are skipped. Conversely, there may be the case ¬(E_{1,2} ⊕ ¬(E_{3,4})). Another notation for this case is ⊕ ∧ (E_{1,2} ⊕ E_{3,4}).

In doing so, Sammeth et al. (2008) proposed a Boolean-like notation in which the splice sites rather than exons are used as elementary units, denoted by numbers. For example, “1−2−, 3−4−” describes two mutually exclusive exons (Fig. 1F), where the hyphens stand for exons, the caret stands for introns and the comma separates isoforms. Such annotation will perfectly describe each isoform, but dependencies among splicing events are included only indirectly. This holds for other approaches like binary array based number vectors (Nagasaki et al., 2006) or isoform signatures (Riva and Pesole, 2009). Dependencies that span constitutive positions (Glauser et al., 2011) will get lost. The intuitive description used by SPLOOCe (Kroll et al., 2012) is inappropriate for cases with more than two dependent isoforms (Fig. 1G). Our suggestion is more general in that it describes more than two possibilities (e.g. E_{1,2} ⊕ E_{3,4} ⊕ E_{5,6}) at once and enables the explicit inclusion of mutual dependencies between splicing events.

3. Detection

The detection of AS can be well distinguished by the three data sources they depend on, namely microarrays, RNA sequence alignments and mere DNA or RNA sequences. Microarrays as a long consolidated methodology still remain useful and accurate for transcriptomic analysis with low input requirements, while RNA-seq technology complements and extends microarray measurements for novel discoveries. Sequence based approaches utilize existing knowledge about splicing and its regulation, e.g. sequence and secondary structure motifs, for investigation of pure genomic sequence data (Hallegger et al., 2010; Raghavachari et al., 2012; Sanchez-Pla et al., 2012).

- **Microarrays** are useful for identification and expression analysis of predicted exons (Exon arrays), in scanning areas for reported and novel exon usage (tiling arrays) and for measurement of known exon junctions (splicing arrays). Exon junctions are sequences after joining at splice sites. Microarray approaches proved to be useful although being strongly dependent on previous knowledge on gene transcription (Castle et al., 2008; Gonzalez-Porta et al., 2012; Raghavachari et al., 2012; Sanchez-Pla et al., 2012).
- **Alignment based analysis** is widely applied in identification of novel splicing events de novo (Sacomoto et al., 2012) or by alignment to genomes (Wang et al., 2008, 2010; Zhou et al., 2012) and in combination with a variety of data reduction techniques like number strings (Sammeth et al., 2008), isoform signatures, Boolean arrays (Nagasaki et al., 2005, 2006) and regular expressions (Kroll et al., 2012).
- **De novo sequence analysis**
  - **Similarity based search** depends on assumptions from knowledge about former described splice events like associated sequence motifs. One assumption is that mutually exclusive exons encode regions of the same structural part of the protein product. This precondition provides restrictions to the search for candidate exons concerning their length, splice site conservation and reading frame preservation, and overall homology. Mutually exclusive exons that are not homologous and not of about the same length will not be found (Sorek et al., 2004b; Stephan et al., 2007; Pillmann et al., 2011).
  - **RNA secondary structure prediction** applies the principles of sequence similarity search on the structural level, such shifting the focus from sequence level to the regulatory relevant structural formation (Washietl et al., 2005; Raker et al., 2009; Reuter and MatheWs, 2010; Yang et al., 2011).
4. Mechanisms leading to mutual exclusion of exons

Several general mechanisms for realizing mutually exclusive splicing have been proposed (Smith, 2005; Nilsen and Graveley, 2010; Jin et al., 2011; Pervouchine et al., 2012; Hemani and Soller, 2012) (Fig. 2):

- **Spliceosome incompatibility** (Fig. 2A): combinations of alternative splice sites can imply that the mutually exclusive exons are recognized and spliced by different spliceosomes, i.e. the U1 or U12 spliceosome (Burge et al., 1998; Letunic et al., 2002; Will and Lührmann, 2005). Beside the major spliceosome, a minor spliceosome, in which the subunits U1 and U2 are replaced by U11 and U12, can process splice sites that have distinct consensus sequences and are incompatible with the major spliceosome (Will and Lührmann, 2005). Thus, each type of the two different spliceosomes is compatible with only one of the MXEs.

- **Coordinated splicing control coupled to NMD** (Fig. 2B): if exons are regulated by trans-acting splicing factors, which suppress splicing of one exon while enhancing splicing of the other one, the mature mRNA can be mutually exclusive (Jones et al., 2001; Spellman et al., 2005). To avoid mis-splicing in the absence of splicing factors, isoforms that contain both exons incorporate premature stop codons and are potentially degraded by nonsense mediated mRNA decay (NMD), as was suggested also for subtle tandem donor splicing (Bortfeldt et al., 2008). This mechanism has been described for the growth factor receptor FGFR2 (Jones et al., 2001), α-tropomyosin (Spellman et al., 2005), and the channel CaV1.2 (Tang et al., 2011).

- **Steric hindrance** (Fig. 2C): for introns shorter than 60 bp the spliceosome cannot perform the necessary structural arrangements, hence, one splice site is used and the other one is skipped due to steric hindrance (Smith and Nadal-Ginard, 1989; Mullen et al., 1991; Kennedy and Berget, 1997). Thus, by skipping one exon an intron is generated that is long enough to become properly spliced.

- **Formation of secondary structure elements** (Fig. 2D):
  - **Induced steric hindrance**: loop formation within long introns may bring splice sites in close proximity preventing splicing in a similar manner as steric hindrance at short introns (Jin et al., 2011).
  - **Dockers-selector pairing. Looping out**: a group of exons, each possessing a similar upstream selector site, can compete in forming a secondary structure by binding. Forming intronic RNA pairings, to a complementary docker site upstream of all these exons (Miriami et al., 2003; Graveley, 2005; Anastassiou et al., 2006; Yang et al., 2011). Each exon is normally bound by a repressor, which inhibits the inclusion of the exon. Base pairing of the docker with one selector site activates the exon downstream adjacent to the selector by releasing the repressor. As only one loop can be formed there will be one exon included exclusively. This is known from Drosophila as docker-selector principle (Schmucker et al., 2000; Graveley, 2005; Anastassiou et al., 2006; Olson et al., 2007). This mechanism may have occurred by exon duplication and subsequent mutations.
  - **Approximation of cis elements**: RNA pairing directs control elements flanking an exon into close physical distance, thus forming a splicing activating complex (Muh et al., 2002; Yang et al., 2011).

Many of these mechanisms suggest that MXEs are adjacent to each other in the genome. Moreover, it has been proposed...
that regulatory microRNAs and epigenetic mechanisms such as chromatin structure and nucleosome organization play a role in exon choice (Luco and Misteli, 2011; Huang et al., 2012). One may argue that this implies a splicing code at a higher level.

Findings about non-adjacent and clustered MXEs are of interest in view of the relevance of the above-mentioned mechanisms of MXE splicing proposed in the literature (Fig. 2). Spliceosome incompatibility seems unlikely in the case of non-adjacent MXEs. This is because the entire region between the two MXEs would have to be removed (a contradiction to the assumption that it includes constitutive exons), unless the constitutive exons would be recognized by both spliceosomes. Steric hindrance is prevented if constitutive exons exist between the MXEs because these MXEs would then be sufficiently distant. Also the mechanism via NMD and trans-acting splicing factors is unlikely, since such a mechanism would require the following conditions: (a) retention of both alternative exons should cause a frame shift leading to NMD; (b) removal of both alternative exons should cause a frame shift leading to NMD (conditions a and b would guarantee mutual exclusivity); (c) splicing of the first alternative exon should not cause a frame shift in order to preserve functionality of the constitutive exons in between the alternative exons. However, these conditions contradict each other, since they imply that the second alternative exon would cause a frame shift both when absent and present. The docker–selector principle would not work either because constitutive exons must not be situated in the loop occurring in that mechanism.

As mentioned in the Introduction, the Dscam gene has a set involving 48 alternative exons (Graveley, 2005). This led to hypotheses about the possibility of more than two mutually exclusive exons (or exon groups) within one and the same gene in the human genome. Such cases would be an indication of docker–selector pairing or related mechanisms. However, in men and mice, we found no more than two mutually exclusive parties (Pohl et al., 2009). Such cases had not been detected by other groups either, nor have they been considered in splicing databases for the human genome. Therefore, databases like ASAP or HOLLYWOOD are designed for mutually exclusive splicing with exactly two exclusive exons and do not cover mutually exclusive exon groups. Nonetheless, the docker–selector mechanism cannot be excluded as hypothetical splicing mechanism for adjacent MXEs. As for non-adjacent MXEs, in our opinion, the existence of a hitherto unknown mechanism has to be considered.

5. Evolutionary conservation

In general, functional AS events conserved across species tend to preserve the reading frame (Sorek et al., 2004a; Kim et al., 2008). Events shifting the reading frame imply NMD and are hardly conserved between humans and mice (Zhang et al., 2009). By utilizing AS, evolution may currently be working on adoption of further functions for the concerned genes.

In future investigations, it would be interesting to analyze the difference in conservation between more frequent and less frequent MXE isofoms. From an evolutionary point of view, one can expect that one of the two MXE isofoms shows a high sequence conservation on the DNA level to maintain the biological function of the protein (Boué et al., 2003). The other form (probably the minor form) may be less conserved, to serve as a “playground of evolution”. Similar observations were made for skipped exons (Xing and Lee, 2006) and subtle tandem donor AS (Borthfeldt et al., 2008). Note that a partially low EST coverage might have prevented the detection of some orthogonal isofoms possibly serving this purpose. Since non-adjacent MXEs are less conserved and tend to shift the reading frame, they are more likely to be a “playground of evolution” than adjacent MXEs. Analogous hypotheses have been put forward for intron retention and alternative ends (Sorek, 2007; Tarriño et al., 2008). In contrast, in situations where both MXEs are equally important, as it appears to be the case for ion channels, the degree of conservation is not likely to differ. In fact, the instances of adjacent MXEs within the human genome we found within our studies, were predominantly assigned to ion channels (Pohl et al., 2009).

A following step would be to investigate the impact of MXE splicing on protein function. To do so, it is useful to compare the protein structures resulting from the alternative transcripts (Birzele et al., 2008a, b). However, at present for many proteins no structural information is available.

6. Conclusions

Here we have reviewed approaches to analyze a rare subtype of alternative splicing (AS), termed mutually exclusive exon (MXE) splicing. We have discussed various approaches and nomenclatures in this context. MXE splicing is the only type of AS that can maintain the size of the protein introducing a quasi-exchange, provided that the exons are (nearly) of the same length.

It is often assumed that MXEs – which are expected to originate from exon duplication – should usually be in direct genomic neighborhood (Letunic et al., 2002; Copley, 2004). However, there is no reason to exclude non-adjacent MXEs. Hence, only approaches and nomenclatures considering mutual (perhaps long-ranging) dependencies within complete genes will have a chance of success in deciphering the full splicing picture.

In many studies often only two transcripts (e.g. in two tissues or developmental stages) are considered. MXEs found in this manner need not to be MXEs when taking more abundant transcript data into account. Thus, the term MXE is relative; it depends on how many different conditions are studied and the regulation mechanism might not work in a perfectly strict manner (Tang et al., 2011). Summarizing these considerations, it may be worth relaxing the rigid definition of MXEs in that a certain fraction of non-exclusive events is allowed. Nevertheless, as many different tissues as possible should be analyzed as this for example the detection of a switch like skipped exons (Wang et al., 2008). In clustered MXEs, a strict dependence between the exons within each cluster occurs. Such a dependence can also occur in other types of AS. For example, in some events of exon skipping, a strict correlation between two consecutive exons was found (Sammeth et al., 2008).

The analysis of MXEs by various bioinformatics techniques is likely to be very helpful in diagnosing diseases and tailoring pharmaceuticals in personalized medicine. For example, the knowledge of receptor variants can help identify new drug targets. The knowledge of variants of exoenzymes degrading polymers can be useful in optimizing biotechnological processes.

Acknowledgements

We thank Ines Heiland, Stefan Heinemann, Dirk Holste and Günter Theifzen for stimulating discussions and Ina Weiß for valuable assistance in the literature search. S.S. acknowledges financial support by the German Ministry of Education and Research (BMBF) in the Virtual Liver program.

References


