Title
Combinatorial control of exon recognition.

Permalink
https://escholarship.org/uc/item/0xm2x8p4

Journal
The Journal of biological chemistry, 283(3)

ISSN
0021-9258

Author
Hertel, Klemens J

Publication Date
2008

DOI
10.1074/jbc.r700035200

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed
Combinatorial Control of Exon Recognition*

Klemens J. Hertel

From the Department of Microbiology and Molecular Genetics, University of California, Irvine, California 92697-4025

Pre-mRNA splicing is a fundamental process required for the expression of most metazoan genes. It is carried out by the spliceosome, which catalyzes the removal of noncoding intronic sequences to assemble exons into mature mRNAs prior to export and translation. Given the complexity of higher eukaryotic genes and the relatively low level of splice site conservation, the precision of the splicing machinery in recognizing and pairing splice sites is impressive. Introns ranging in size from <100 up to 100,000 bases are removed efficiently. At the same time, a large number of alternative splicing events are observed between different cell types, during development, or during other biological processes. This extensive alternative splicing implies a significant flexibility of the spliceosome to identify and process exons within a given pre-mRNA. To reach this flexibility, splice site selection in higher eukaryotes has evolved to depend on multiple parameters such as splice site strength, the presence or absence of splicing regulators, RNA secondary structures, the exon/intron architecture, and the process of pre-mRNA synthesis itself. The relative contributions of each of these parameters control how efficiently splice sites are recognized and flanking introns are removed.

Of the ~25,000 genes encoded by the human genome (1), >70% are believed to produce transcripts that are alternatively spliced. Thus, alternative splicing of pre-mRNAs results in the production of multiple protein isoforms from a single pre-mRNA, significantly enriching the proteomic diversity of higher eukaryotic organisms (2, 3). Because regulation of this process can determine when and where a particular protein isoform is produced, changes in alternative splicing patterns modulate many cellular activities.

Exon/intron boundaries are defined by direct interactions between the spliceosome and pre-mRNA signature elements. The formation of the spliceosome requires the activity of >300 distinct protein factors and the U1, U2, U4, U5, and U6 snRNAs (4), which assemble onto the pre-mRNA in a stepwise manner (5). After initial splice site recognition and pairing (6, 7), the catalytic components of the spliceosome are activated through extensive structural rearrangements, ultimately resulting in intron removal (8). This minireview will focus on the basic principles that control initial exon recognition and how the interplay between these parameters results in the generation of differentially spliced mRNA isoforms.

Components Influencing Exon Recognition

Splice Site Strength

A critical step in pre-mRNA splicing is the recognition and pairing of 5'- and 3'-splice sites (Fig. 1). Whereas the 5'-splice site is defined by a single element of 8 nt, the 3'-splice site is defined by three sequence elements usually contained within 40 nt upstream of the exon/3'-intron junction (6). These elements are the branch-point sequence, the polyypyrimidine tract, and the exon/3'-intron junction. Initial recognition of exon/intron junctions is based on direct interactions between U1 snRNP with the 5'-splice site, the U2 auxiliary factor with the polyypyrimidine tract, and U2 snRNP with the branch-point sequence. Because the sequence specificity of these interactions is driven by pre-mRNA/snRNA interactions and the U2 auxiliary factor binding preference for polypyrimidines, splice sites are classified by their complementarity to U1 snRNA (5'-splice site) and the extent of the polyypyrimidine tract (3'-splice site). Greater complementarity with U1 snRNA and longer polypyrimidine tracts translate into higher affinity binding sites for these spliceosomal components and thus more efficient exon recognition.

Splicing Enhancers and Silencers

For classical cases of alternative splicing, it was shown that cis-acting RNA sequence elements increase exon inclusion by serving as binding sites for the assembly of multicomponent splicing enhancer complexes. Because these sequence elements were located within the regulated exon, they were defined as exonic splicing enhancers (ESEs) (Fig. 1) (5). ESEs are usually recognized by at least one member of the essential serine/arginine-rich protein (SR protein) family that recruits the splicing machinery to the adjacent intron (5, 9). Interestingly, SR protein-binding sites are present not only within alternatively spliced exons, but also within the exons of constitutively spliced pre-mRNAs (10). It is therefore expected that SR proteins bind to sequences found in most, if not all, exons.

However, regulation of pre-mRNA splicing is much more complex than the simple ESE recruitment model. Intronic splicing enhancers and splicing silencers, either exonic (ESS) or intronic (ISS), occur frequently and influence splice site selection (Fig. 1). The best characterized examples of ESSs are recognized by members of the family of heterogeneous nuclear RNP complexes, whereas ISSs usually bind the polypyrimidine track-binding protein (5). Several mechanisms have been proposed for ESS- or ISS-mediated splicing repression. Heterogeneous nuclear RNP-bound splicing silencers have been shown to repress spliceosomal assembly through multimerization along
Exon/Intron Architecture

Intron size has been correlated with rates of evolution (20) as well as the regulation of genome size (21). In addition, the exon/intron architecture has also been shown to have an influence on splice site recognition (22). For example, increasing the size of mammalian exons results in exon skipping. However, the same enlarged exons were included when the flanking introns were small (23). Thus, splice site recognition is more efficient when introns or exons are small. These early observations suggested that splice sites are recognized across an optimal nucleotide length and predicted that intron length significantly influences the efficiency of pre-mRNA splicing and alternative splice site choice (Fig. 1). This is an important hypothesis because of the divergent distribution of intron length in the human genome and because it had been proposed that the spliceosome uses two modes of recognition to define splice sites (22). An “intron definition” model has been formulated in which the 5’- and 3’-splice sites of introns are directly identified as the splicing unit. In the alternative model, the “exon definition” model, splice sites flanking the exon are initially recognized and subsequently paired (Fig. 2A).

Using a kinetic approach, it was recently demonstrated that splice site preference across introns ceases when intron size is between 200 and 250 nt (24). Beyond this threshold, splice sites were recognized across the exon. Splice site preference across the intron was significantly more efficient than splice site preference across the exon, resulting in enhanced inclusion of exons with weak splice sites. These observations demonstrated that intron size profoundly influences the likelihood that an exon is alternatively spliced. These experimental results were further supported using computational analyses of expressed sequence tag data bases. For example, Drosophila exons flanked by long introns display an up to 90-fold higher probability to be alternatively spliced compared with exons flanked by two short introns, demonstrating that the exon/intron architecture is a major determinant in governing the frequency of alternative splicing in Drosophila (Fig. 2B). In agreement with experimental predictions, the most dramatic change in the alternative splicing probability was observed within the transition from intron definition to exon definition. Exon skipping is also more likely to occur when exons are flanked by long introns in the human genome. Interestingly, experimental and computational analyses showed that the length of the upstream intron is more important in inducing alternative splicing than the length of the downstream intron (Fig. 2B), most likely reflecting the influence RNA transcription exerts on pre-mRNA splicing. These results showed that the exon/intron architecture defines mechanisms of splice site recognition and influences the frequency of alternative pre-mRNA splicing (24).

RNA Secondary Structure

Single-stranded RNA is likely to adopt local secondary folds and tertiary interactions that may involve up to hundreds of nucleotides. Although pre-mRNAs are typically depicted in a linear fashion, we have to assume that higher order structures exist that maintain a good portion of the RNA double-stranded. Depending on the thermodynamic stability, these structures may persist long enough to interfere or modulate splice site recognition (Fig. 1). In principle, these local structures can be inhibiting or activating spliceosomal assembly. This is because the recognition of splice sites, enhancers, and silencers usually depends on interactions between protein factors and a single-stranded portion of the pre-mRNA. Local RNA structures can interfere with spliceosomal assembly if they conceal splice sites or enhancer-binding sites within stable helices. On the other hand, local RNA structures can also promote spliceosomal assembly by masking splicing repressor-binding sites.

The importance of RNA secondary structure in modulating splice site selection has been documented frequently. For example, two classes of conserved RNA elements have been...
MINIREVIEW: Combinatorial Control of Exon Recognition

Over the last few years, it has become increasingly clear that exon selection is influenced by a number of activating and inhibitory elements. Given the divergent sequence and architecture of genes, every exon has its specific set of identity elements that permit its recognition by the spliceosome. Each exon is flanked by a unique pair of splice enhancers and silencers and secondary structures. The sum of contributions from each of identified in the Dscam (Down syndrome cell adhesion molecule) exon 6 cluster, which contains 48 alternative exons: a common docking site and selector sequences unique to each exon 6 variant. Each selector sequence can base pair with the docking site to form a secondary structure, thereby activating and directing mutually exclusive exon pairing (25). An inhibitory role of RNA secondary structure was demonstrated for splice site recognition of SMN2 exon 7. The formation of an RNA hairpin close to the 5′-splice site of SMN2 exon 7 interfered with its interaction with U1 snRNP, resulting in reduced exon inclusion levels (26). In agreement with these observations are computational analyses showing that ~5% of alternative splicing events strongly correlate with the presence of stable secondary structures. These examples support the idea that local RNA secondary structures play a more significant role in modulating splice site recognition than perhaps currently appreciated.

Pre-mRNA Synthesis by RNA Polymerase II

Recent studies have demonstrated that the process of splice site recognition can occur co-transcriptionally (27–29), i.e. the splice sites of an exon can be identified by the spliceosome while downstream exons still await their synthesis by RNA polymerase II (pol II). Thus, like 5′-capping and 3′-polyadenylation, intron removal is physically and temporally linked to RNA transcription. For example, alternative splicing of a reporter gene varied depending on the pol II promoter structure from which the transcript originated (30), suggesting a model in which splicing factors associate with pol II close to the promoter. As a consequence, differences in promoter structure could lead to differences in the splicing factors recruited to the transcription machinery (31).

In a complementary model, the kinetics of pol II transcription is proposed to influence alternative splicing. As pol II polymerizes in a strict 5′–3′-direction, alternative exons are made prior to or after the synthesis of competing neighboring exons. Accordingly, the relative timing of producing competing exons can bring about changes in the splicing pattern. Strong support for the kinetic proposal stems from experiments testing the effects of increased intron length, different classes of transcription activators, and pol II elongation mutants (32, 33).

Combinatorial Control of Exon Recognition

Over the last few years, it has become increasingly clear that exon selection is influenced by a number of activating and inhibitory elements. Given the divergent sequence and architecture of genes, every exon has its specific set of identity elements that permit its recognition by the spliceosome. Each exon is flanked by a unique pair of splice enhancers and silencers and secondary structures. The sum of contributions from each of
MINIREVIEW: Combinatorial Control of Exon Recognition

Given the variation in splice site strength, exon/intron architecture, number of splicing enhancers and silencers, the likelihood of local secondary structures, and fluctuations in pre-mRNA synthesis, exons will display a wide spectrum of recognition potential. Exons with high levels of recognition potential represent mainly constitutively spliced exons (yellow), whereas exons with low levels are predominantly skipped (blue). The strength of interactions between the splicing machinery and exons within the median category (gray) is within an energetic range that can trigger increased or decreased inclusion of an exon upon a slight shift in the concentration of splicingosomal components.

These identity elements then define the overall recognition potential of an exon or the overall binding affinity for the spliceosome. Considering the variation in splice sites, exon/intron architecture, number of enhancers and silencers, and secondary structures, the recognition potential of exons is expected to span a wide range (Fig. 3). The spectrum of exon recognition potential ranges from exons that are constitutively spliced and always included in the final transcript to exons that are rarely included in the mature mRNA. The center of the spectrum represents exons that are not always included into the final mRNA isoform, alternative exons. Both extreme exon classes are expected to be impervious toward subtle changes in the splicing environment because their affinity for the spliceosome is so great or low that minor changes will not significantly alter overall exon definition efficiencies. On the other hand, the exon class within the center of the recognition spectrum is expected to be most sensitive to even minor changes in splicing efficiency. For example, a small drop in the concentration of the spliceosome or an SR protein could trigger a change from preferential inclusion of an exon to preferential exclusion.

What are the most important parameters for exon recognition? Although not tested systematically, it is likely that the strength of the splice sites and their relative proximity (within 200 nt) are the most crucial aspects of efficient splicing. Because the spliceosome assembles around splice sites, the binding potential of splice sites builds the foundation for efficient exon definition. The contributions of the other parameters will vary significantly from case to case, augmenting or reducing the overall affinity of the splicing machinery. As a consequence of the expected fluctuations in the concentrations of spliceosomal components and splicing activator/repressors between different cell types or between distinct biological processes such as the cell cycle and development, it is anticipated that the same exon may display variable exon recognition potentials in these scenarios. As a result, exons that are alternatively included in one cell type can be alternatively excluded in another.

Promoting Alternative Splicing

In the literature, alternative splicing is attributed mainly to the activities of splicing enhancers and repressors that allow transient interactions with splicing regulators (5). In most cases, the presence or absence of splicing regulatory proteins modulates the overall exon recognition to significantly tilt the balance between exon inclusion and exclusion. Similarly, protein interactions within the pre-mRNA may induce or interfere with the formation of RNA secondary structures that modulate efficient spliceosomal assembly. However, invariant elements such as splice site sequences and exon/intron architecture have also the potential to mediate differential splicing. Based on the principle of mass action, fluctuations in snRNP levels can induce changes in the efficiency of splice site recognition, thus altering exon inclusion ratios. Such changes in the concentration of the general splicing factors could account for many of the observed alternative splicing events observed between different cell types.

In the cell, alternative splicing has also been attributed to promoter-dependent recruitment of specific splicing regulators or to changes in the kinetics of pre-mRNA synthesis (28). Thus, modulating the recruitment of specific splicing factors or modulating the relative synthesis of competing splice sites can influence the selection of alternative splicing patterns. Alternatively, changes in the kinetics of RNA synthesis are able to influence the likelihood that local RNA secondary structures form that induce alternative splice site selection (34).

Regulation of alternative splicing can be achieved through modulating any one of the exon recognition components (35). However, specific regulation requires the selected targeting of splicing activators or repressors unique to particular exons. This is often mediated through perturbations of post-translational modifications that are essential for optimal activity of many splicing regulatory factors, such as alterations in the phosphorylation state of specific SR proteins (36).

Perspectives

In higher eukaryotes, the splicing machinery is confronted with multiple cues that promote or interfere with the identification of exon/intron boundaries. Compared with yeast, where splice site sequences adhere to a strict consensus, higher eukaryotes have evolved to accept much greater sequence variability within splice sites. To compensate for the accompanying loss in information content, additional mechanisms for spliceosomal recruitment have emerged that, when combined, ensure proper exon identification. One evolutionary benefit of such combinatorial control is the ability of an organism to generate and test alternative gene products to increase survival fitness. However, mechanisms such as nonsense-mediated decay (37) must also exist to dispose of alternatively spliced mRNA products that encode for potentially harmful gene products.

Given the expected spread of exon recognition potentials (Fig. 3), it is likely that all exons participate in alternative splic-
ing at some level. This raises the important question of when and how many alternative splicing events are biologically significant. As of today, biological functions for mRNA isoforms have been demonstrated in a large number of the cases studied (38). However, it is also possible that a significant number of the mRNA isoforms that survive quality control steps, such as nonsense-mediated decay, are ultimately translated without an obvious biological function. Thus, increased proteomic output through alternative splicing may come at the cost of generating isoforms that initially do not have biological functions in the cell (39).

In the post-genomic world, research into the mechanisms of splice site selection is leading toward the establishment of rules that will allow splice patterns to be predicted on the basis of sequence information. Computational methods combined with laboratory experiments have already generated algorithms that predict splicing regulatory sequences (18, 19, 40, 41). This significant progress suggests the exciting possibility of crafting a “splicing code” that permits the prediction of exons and the probability of their inclusion in the most abundant mRNA isoform. Ideally, a splicing code should be able to differentiate between alternative splicing events in different tissues and different biological processes. Whereas the current expressed sequence tag coverage may allow classifications only between a few cell types, new powerful pyrosequencing techniques will prove to be an extremely useful tool in predicting mutations data base of human disease alleles (www.hgmd.org) are located within splice sites. A splicing code should predict the effect these mutations have on pre-mRNA splicing and could also prove to be an extremely useful tool in predicting mutations that disrupt splicing.

REFERENCES
1. International Human Genome Sequencing Consortium (2004) Nature 431, 931–945
2. Maniatis, T., and Tasic, B. (2002) Nature 418, 236–243
3. Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loehr, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R., and Shoemaker, D. D. (2003) Science 302, 2141–2144
4. Jurica, M. S., and Moore, M. J. (2003) Mol. Cell 12, 5–14
5. Black, D. L. (2003) Annu. Rev. Biochem. 72, 291–336
6. Reed, R. (1996) Curr. Opin. Genet. Dev. 6, 215–220
7. Lim, S. R., and Hertel, K. J. (2004) Mol. Cell 15, 477–483
8. Staley, J. P., and Guthrie, C. (1998) Cell 92, 315–326
9. Graveley, B. R. (2000) RNA 6, 1197–1211
10. Schaal, T. D., and Maniatis, T. (1999) Mol. Cell. Biol. 19, 261–273
11. Zhu, J., Mayeda, A., and Krainer, A. R. (2001) Mol. Cell 8, 1351–1361
12. Tange, T. O., Damgaard, C. K., Guth, S., Valcarcel, J., and Kjems, J. (2001) EMBO J. 20, 5748–5758
13. Martinez-Contreras, R., Fisette, I. F., Nasim, F. U., Madden, R., Cordeau, M., and Chabot, B. (2006) PloS Biol. 4, e21
14. Lorson, C. L., and Androphy, E. J. (2000) Hum. Mol. Genet. 9, 259–265
15. Cartegni, L., and Krainer, A. R. (2002) Nat. Genet. 30, 377–384
16. Kashima, T., and Manley, J. L. (2003) Nat. Genet. 34, 460–463
17. Singh, N. N., Androphy, E. J., and Singh, R. N. (2004) RNA 10, 1291–1305
18. Zhang, X. H., and Chasin, L. A. (2006) Mol. Cell. Biol. 18, 1241–1250
19. Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., and Burge, C. B. (2004) Cell 119, 831–845
20. Marais, G., Nouvellet, P., Kightley, P. D., and Charlesworth, B. (2005) Genetics 170, 481–485
21. Deutsch, M., and Long, M. (1999) Nucleic Acids Res. 27, 3219–3228
22. Berget, S. M. (1995) J. Biol. Chem. 270, 2411–2414
23. Sterner, D. A., Carlo, T., and Berget, S. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 15081–15085
24. Fox-Walsh, K. L., Dou, Y., Lam, B. J., Hung, S. P., Baldi, P. F., and Hertel, K. I. (2005) Proc. Natl. Acad. Sci. U. S. A. 102, 16176–16181
25. Graveley, B. R. (2005) Cell 123, 65–73
26. Singh, N. N., Singh, R. N., and Androphy, E. J. (2007) Nucleic Acids Res. 35, 371–389
27. Gornemann, J., Kotovic, K. M., Hujer, K., and Neugebauer, K. M. (2005) Mol. Cell 19, 53–63
28. Kornbluth, A. R. (2005) Curr. Opin. Cell Biol. 17, 262–268
29. Tardiff, D. F., Lacadie, S. A., and Rosbash, M. (2006) Mol. Cell 24, 917–929
30. Cramer, P., Pesce, C. G., Baralle, F. E., and Kornbluth, A. R. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 11564–11560
31. de la Mata, M., and Kornbluth, A. R. (2006) Nat. Struct. Mol. Biol. 13, 973–980
32. de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornbluth, A. R. (2003) Mol. Cell 12, 525–532
33. Hicks, M. J., Yang, C. R., Kotlajich, M. V., and Hertel, K. J. (2006) PLoS Biol. 4, e147
34. Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (1988) Cell 54, 393–401
35. House, A. E., and Lynch, K. W. (2008) J. Biol. Chem. 283, 1217–1221
36. Stamms, S. (2008) J. Biol. Chem. 283, 1223–1227
37. Chang, Y. F., Imam, J. S., and Wilkinson, M. F. (2001) Annu. Rev. Biochem. 70, 51–74
38. Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001) Nucleic Acids Res. 29, 2850–2859
39. Boue, S., Letunic, I., and Bork, P. (2003) BioEssays 25, 1031–1034
40. Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. (2002) Science 297, 1007–1013
41. Ben-Dov, C., Hartmann, B., Lundgren, J., and Valcarcel, J. (2008) J. Biol. Chem. 283, 1229–1233
42. Bentley, D. R. (2006) Curr. Opin. Genet. Dev. 16, 545–552
Minireviews:
Combinatorial Control of Exon Recognition

Klemens J. Hertel
J. Biol. Chem. 2008, 283:1211-1215.
doi: 10.1074/jbc.R700035200 originally published online November 16, 2007

Access the most updated version of this article at doi: 10.1074/jbc.R700035200

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 42 references, 18 of which can be accessed free at http://www.jbc.org/content/283/3/1211.full.html#ref-list-1