How do RNA sequence, DNA sequence, and chromatin properties regulate splicing?
Leonie Ringrose

Address: IMBA - Institute of Molecular Biotechnology, Dr Bohr-Gasse 3, 1030 Vienna, Austria
Email: leonie.ringrose@imba.oeaw.ac.at


Abstract
Recent genome-wide studies have revealed a remarkable correspondence between nucleosome positions and exon-intron boundaries, and several studies have implicated specific histone modifications in regulating alternative splicing. In addition, recent progress in cracking the 'splicing code' shows that sequence motifs carried on the nascent RNA molecule itself are sufficient to accurately predict tissue-specific alternative splicing patterns. Together, these studies shed light on the complex interplay between RNA sequence, DNA sequence, and chromatin properties in regulating splicing.

Introduction and context
The fact that the majority of pre-mRNAs are spliced while they are still being transcribed [1] has led to the proposal that local chromatin structure and histone modifications may play direct roles in regulating splicing. Indeed, several studies have shown that specific histone modifications can affect the association of splicing factors with chromatin and the efficiency of the splicing process [2-4]. A link between nucleosome positioning and exon-intron boundaries was first proposed in 1991, long before the functional link between splicing and transcription was established [5]. Recent advances in the computational prediction and experimental verification of nucleosome positioning, based on underlying DNA sequence features [6-8], in combination with recently published high-resolution genome-wide maps of nucleosome positions [9,10] and histone modifications [11,12] in several organisms, have now enabled genome-wide comparisons of nucleosome positioning, histone modifications, and intron-exon architecture, revealing surprising correlations between these features and throwing light on the question of how chromatin features may help the splicing machinery to distinguish between exons and introns.

A distinct question is that of alternative splicing. Alternative splicing can generate many different transcripts from a single gene. Recent high-throughput transcriptome analyses have detected alternative splicing in approximately 95% of multi-exon human genes [13-15] and also extensively in mice [16], plants [17], flies [18,19], and yeast [20,21]. Alternative splicing is often regulated by trans-acting factors that are differentially expressed in different tissues or metabolic states and bind specific sequence or structural motifs on the pre-mRNA, resulting in alternative splicing [22-24]. It has long been a goal of the splicing field to crack the 'splicing code': to identify the pre-mRNA sequence features that can explain and predict not only the exact sites of constitutive splicing but also the features that determine tissue-specific alternative splicing patterns [25].

Previous work has successfully identified pre-mRNA-encoded features that define the precise boundaries of certain classes of constitutively spliced exons [26-28] and additional splicing enhancers and silencers [28,29], some of which have been shown to correlate with exon inclusion levels in specific tissues [16,28]. However, it has been suggested that pre-mRNA-encoded information alone is not sufficient to explain the recognition of short exons in a desert of long introns or the modulation of this process in alternative splicing [30]. The idea that local chromatin structure may add an additional regulatory layer to splicing, particularly to alternative
splicing, is currently gaining momentum, and both correlative evidence and functional evidence are emerging [30-35]. Additional insights into the role of RNA sequence come from the recent demonstration that by taking hundreds of RNA features into account, a ‘splicing code’ based on pre-mRNA features is able to qualitatively predict tissue-specific alternative splicing for thousands of vertebrate exons [36]. Here, recent advances and current ideas on how chromatin and pre-mRNA sequence contribute to constitutive and alternative splicing are reviewed.

**Major recent advances**

In 2009 and 2010, several groups [37-42] used bioinformatic approaches to uncover patterns from published genome-wide maps of nucleosome positioning in human T cells and in Caenorhabditis elegans, derived from deep sequencing of micrococcal nuclease-digested chromatin [9,10]. These recent bioinformatic analyses [37-42] now show that nucleosomes sit preferentially on exon sequences whereas introns are relatively depleted of nucleosomes (Figure 1). This remarkable arrangement was observed in several metazoan species and was found to be independent of transcriptional activity, suggesting that the arrangement is an inherent property of chromatin [37-39,41]. Strikingly, the average length observed for metazoan exons (140-150 base pairs) corresponds neatly with the length of DNA required to wrap a single nucleosome (147 base pairs) [39,41]. Further correlations to the splicing process were observed by several authors (e.g., exons flanked by weak splice sites or by long introns have a higher tendency to be bound by nucleosomes) [40,41], raising the idea that nucleosomes may help the splicing machinery by ‘marking’ exons that may otherwise be difficult to recognize (reviewed in [30-32]).

By analysis of chromatin immunoprecipitation (ChIP) combined with high-throughput sequencing (ChIP-seq) and ChIP followed by microarray analysis (ChIP-chip) data sets [9,11,12], several studies have examined correlations between histone modifications and exon/intron boundaries and have reported conflicting results. Several authors conclude that specific histone modifications are enriched on exons compared with introns, suggesting an active marking mechanism [12,37,40,43], whereas others argue that these apparent enrichments are due mostly to nucleosome positioning, which is independent of modification status [38,39,41]. It has been suggested that these discrepancies may be due mainly to difficulties of normalization of ChIP data for nucleosome occupancy, compounded by the fact that occupancy studies and modification studies were performed with different techniques (micrococcal nuclease digestion versus ChIP) [30]. It is also possible that gene-specific differences exist but go undetected in global analyses.

To what extent can we understand the preference of nucleosomes for exonic sequences in terms of known
pre-mRNA splicing signals? Figure 1 summarizes the RNA, DNA, and chromatin features at exon-intron boundaries. The preference of nucleosomes for exons appears to be a consequence not only of the higher guanine and cytosine content of exons [38,39,41] but also of a high density of sequences that repel nucleosomes [6] exactly at the intron-exon boundary and a depletion of these sequences within the exon itself relative to the adjacent intron [39] (Figure 1). Particularly interesting is the polypyrimidine tract (PPT), typically a long (10-20 nucleotides) run of uracil bases at the 3’ intron end in the pre-mRNA, that is specifically recognized by spliceosome components (Figure 1). At the DNA level, the PPT corresponds to poly T, one of the strongest nucleosome-repelling sequences identified by Kaplan et al. [6]. Thus, it is clear that splicing signal sequences play a role both at the RNA level, for recognition by RNA-binding proteins of the splicing machinery, and at the DNA level, in determining the positions of nucleosomes. An additional role of splice site sequences in increasing RNA flexibility has been proposed [42]. Is nucleosome positioning thus merely a coincidence of the RNA sequence or does it have a causal role in splicing? Experimental evidence beyond correlation is currently lacking but several possible roles have been proposed (summarized in Figure 2).

The models proposed so far fall into one of two non-mutually exclusive classes – the ‘recruitment’ models and the ‘kinetic coupling’ models [32] – and are summarized in Figure 2. Recruitment models favor the idea that nucleosome positioning or nucleosome modifications on exons guide the splicing machinery to the right place (Figure 2), whereas kinetic coupling proposes that the presence of a nucleosome in the path of the polymerase decreases the speed of transcription and thus allows more time for splicing to occur (Figure 2, arrow 5). In favor of such ‘speed bump’ models [30], single-molecule in vitro experiments have shown that the speed of RNA polymerase II (Pol II) transcription is modulated by nucleosomal barriers [44]. On the other hand, a recent in vivo study of Pol II transcription rates measured similar speeds on exonic and intronic sequences and showed that although splicing does indeed occur cotranscriptionally, it lags substantially behind the transcription process, being approximately twice as slow [45]. This raises the question of how much the chromatin features ahead of the polymerase can influence splicing events that take place far behind it (Figure 2). However, this study was limited to a small number of genes, and it is not known whether they contain exons that are regulated in a chromatin or Pol II elongation-dependent manner (or both).

If nucleosome positions do affect splicing, then how could such a system permit the vast amount of flexibility observed in alternative splicing? Several authors have proposed that nucleosome remodelers, a relaxation of nucleosome positioning at alternatively spliced exons, or tissue-specific histone modifications may account for tissue-specific differences in nucleosome positioning or properties and thus facilitate alternative splicing [30-32,39]. In support of this idea, ectopic recruitment of heterochromatin modifications to the vicinity of an alternative exon was found to reduce Pol II speed and to
Future directions

In the future, it will be essential to determine the relative contributions of RNA sequence features, nucleosome positioning, and histone modifications to constitutive and alternative splicing. The data so far correlating nucleosome positioning to exon-intron architecture have been limited to few cell types. It will be important to see whether nucleosome positions are indeed different in different tissues, as has been proposed via thermodynamic competition with tissue-specific transcription factors [7] and observed for regulatory regions of specific loci [46]. If this is also the case on a genome-wide scale, then do alternative nucleosome positions correlate with alternatively spliced exons?

Furthermore, it would be extremely interesting to investigate whether any of the RNA sequence features that are predictive for alternative splicing [36] could play an additional role at the DNA level (e.g., as nucleosome positioning or repelling sequences [6] or as binding sites for site-specific DNA binding proteins that may compete with nucleosomes in a tissue-specific manner [7]).

Finally, it will be essential to go beyond global correlations and to determine cause and effect for specific loci. To unravel the relative contributions of RNA sequence, DNA sequence, and chromatin architecture, it will be essential to overcome the inherent difficulty in such experiments (i.e., that any change in DNA sequence changes the RNA sequence). Thus, it will be essential to devise strategies by which nucleosome positions can be modulated without affecting the underlying DNA sequence (e.g., by manipulating levels of remodelers or competing transcription factors).

Abbreviations

ChiP, chromatin immunoprecipitation; ChIP-chip, chromatin immunoprecipitation followed by microarray analysis; ChIP-seq, chromatin immunoprecipitation followed by high throughput sequencing; FGFR2, fibroblast growth factor receptor 2; H3K36me3, histone 3 lysine 36 trimethylation; Pol II, RNA polymerase II; PPT, polypyrimidine tract.

Competing interests

The author declares that she has no competing interests.

Acknowledgments

Work in the author’s laboratory is supported by the Austrian Academy of Sciences, the FWF (Fonds zur Förderung der wissenschaftlichen Forschung [Austrian Scientific Research Fund]), Gen-AU (Genomforschung in Österreich [Genome Research in Austria], The Epigencode Network of Excellence (EU- FP6), and the FCT (Fundação para a Ciência e Tecnologia [Portuguese Foundation for Science and Technology]).

References

2. Loomis RJ, Naoe Y, Parker JB, Savic V, Bozovsky MR, Macfarlan T, Manley JL, Chakravarti D: Chromatin binding of Srp20 and ASF/SF2 and dissociation from mitotic chromosomes is


F1000 Factor 8.2 Exceptional Evaluated by Steven Henikoff 07 Jan 2009, Lin Chen 27 Feb 2009, Jesper Sveistrup 02 Apr 2009


F1000 Factor 9.0 Exceptional Evaluated by Paul Liu 29 Apr 2008


F1000 Factor 3.0 Recommended Evaluated by Charles Auffray 02 Feb 2009


F1000 Factor 6.4 Must Read Evaluated by Ken Irvine 20 Nov 2008, Donald Rio 01 Dec 2008


F1000 Factor 3.2 Recommended Evaluated by Julin Maloof 17 Feb 2010, Daniel Chamovitz 06 Jul 2010


F1000 Factor 3.0 Recommended Evaluated by Miriam Zolan 09 Jan 2009


F1000 Factor 6.0 Must Read Evaluated by Anne Baranger 20 Aug 2008


F1000 Factor 8.0 Exceptional Evaluated by Donald Rio 23 Dec 2004, Reinhard Luehrmann 05 Jan 2005


F1000 Factor 6.0 Must Read
Evaluated by Andrea Barta 16 Jul 2009


F1000 Factor 6.9 Must Read


F1000 Factor 6.8 Must Read


F1000 Factor 6.0 Must Read
Evaluated by Reinhard Luehrmann 07 Sep 2009


F1000 Factor 3.0 Recommended
Evaluated by I King Jordan 07 Jan 2010


F1000 Factor 3.2 Recommended


F1000 Factor 8.2 Exceptional
