

For reprint orders, please contact: reprints@futuremedicine.com

When epigenetics meets alternative splicing: the roles of DNA methylation and GC architecture

“We hypothesize that the increase in GC content allowed other players to take part in recognition of exons in the leveled GC regions ... one such possible player is DNA methylation.”

KEYWORDS: DNA methylation ■ GC content ■ splicing

The process of pre-mRNA splicing has been studied for more than 30 years, yet it is far from being fully understood. Accumulating evidence suggests that many splicing events occur cotranscriptionally and that the mRNA precursor remains associated with the chromatin until all of the introns have been removed. Cotranscriptional splicing adds many more factors that might take part in the complex and highly regulated process of exon recognition. If cis-acting regulatory factors, such as splice-site sequences and splicing factors binding domains, did not provide enough complexity, splicing researchers are now realizing that the chromatin structure itself might also affect the exon selection process [1]. The amazing advances of the last several years in sequencing technologies have commenced a new era for studying genome-wide epigenetic factors, as well as new layers of splicing regulation. The available single-nucleotide-resolution data has made it possible to observe that exons, rather than flanking introns, are already marked at the DNA level by higher occupancy of nucleosomes and certain histone modifications [2–4]. DNA methylation is more abundant in coding sequences than noncoding regions [5,6]. Confounding interpretation of these observations is the fact that exons have a higher GC content than flanking introns. Genomic regions with high GC content have enhanced bendability, which may facilitate nucleosome binding. In addition, DNA methylation occurs at CG dinucleotides and, thus, GC rich exons are bound to contain higher levels of DNA methylation than noncoding regions. So the question remains: do high levels of DNA methylation on exons occur circumstantially, as an outcome of high exonic GC content, or does DNA methylation have a biological function in the exon selection process? To answer this question, it is necessary to delve

into the evolutionary changes that have occurred in gene structure and to determine whether these changes are related to the splicing process.

GC content architecture on the exon–intron structure has changed during evolution, especially during the transition from cold- to warm-blooded organisms. The ancestral genome had short introns of low GC content and exons with a much higher GC content. During the evolution of warm-blooded organisms, two gene structures evolved: one located in high GC content regions and the other in low GC content regions. During mammalian and avian evolution GC-rich regions underwent a GC increase that resulted in even higher GC values. Genes located in these high GC content regions have a much less pronounced difference in GC content between exons and introns than those in low GC content regions, and the flanking introns are short, as were their ancestor’s introns, probably as a result of purifying selection [7]. We refer to these regions as having leveled GC content. The regions that did not increase in GC content underwent an increase in intron length that was made possible by the strong splicing signals that characterize exon–intron structures in these regions [8]. Exons located in these regions have a significantly higher GC content than their flanking introns and we refer to them as differential GC exons. Notably, mRNAs encoded by genes characterized by a high differential GC content between exons and introns are more likely to be alternatively spliced through exon skipping, whereas intron retention is the hallmark of genes that have a similar GC content in exons and introns. This, and other observations, implies that splicing regulation differs between these two gene structures. Indeed, exons with a higher GC content than the flanking introns are more efficiently recognized by the splicing machinery



Sahar Gelfman

Department of Human Molecular Genetics & Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv 69978, Israel



Gil Ast

Author for correspondence:
Department of Human Molecular Genetics & Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv 69978, Israel
Tel.: +972 3 640 6893
Fax: +972 3 640 9900
gilast@post.tau.ac.il

than exons in leveled GC content regions when splicing signal strength and intron length parameters are examined [7].

Using cases from these two distinct GC architectures can further our understanding of the possible effect of GC content on chromatin structure and alternative splicing regulation. Differential GC exons are more efficiently marked by nucleosomes, as well as several histone modifications, such as H3K36Me3, than exons in leveled GC content regions [2–4]. Epigenetic factors may mediate the recognition of differential GC exons in several ways: one possibility is that the nucleosome has an effect on RNA polymerase II elongation (the ‘bumper’ theory), which allows ample time for the splicing machinery to recognize exons during cotranscriptional splicing [9–11]. It is also possible that splicing factors are recruited to certain histone modifications, thereby directing the spliceosome to the small exonic islands in the large intronic oceans, which is the case with the PTB splicing factor and H3K36Me3 modification [12].

“...exons with a higher GC content than the flanking introns are more efficiently recognized by the splicing machinery than exons in leveled GC content regions when splicing signal strength and intron length parameters are examined.”

But what happens when the GC content is similar between exons and introns? In this case, nucleosomes have no preference for binding of exons versus introns, and RNA-polymerase II confronts ‘bumpers’ scattered along both exons and introns. Remarkably, all minigenes generated from genes of high GC content fail to splice properly when transfected into mammalian cells [7]. What then helps the splicing machinery to recognize the exons in these ‘evolutionarily new’ regions? It is likely that regulatory systems for RNA processing evolved along with changes in base content. Supporting evidence for this theory can be found in two recent papers that show that alternative splicing patterns are species specific, whereas gene expression patterns are tissue specific and highly correlated between species [13,14]. We hypothesize that the increase in GC content allowed other players to take part in the recognition of exons in the leveled GC regions and that this provided added regulation complexity. One such possible player is DNA methylation.

Analysis of methylation patterns in regions with both differential and leveled GC architectures shows that this modification strongly marks

the exons and is not biased by GC content as is nucleosome occupancy [15]. This type of analysis may have a bias, however, that is related to the abundance of the CpG dinucleotide. The CpG dinucleotide is under-represented in the human genome compared with other dinucleotide steps but is much more prevalent in coding sequences than noncoding regions.

“DNA methylation can participate in chromatin remodeling and is also found in its highest abundance at exon boundaries.”

Thus, even in exons that have the same GC content as their flanking introns, the CpG dinucleotide is more common in exons than introns. This elevated CpG abundance causes, in turn, elevated methylation. When we normalize for this by dividing methylation by CpG abundance, we observe that CpGs in leveled GC environments are more frequently methylated in exons than introns. This finding is accentuated by a drop in the probability of methylated CpGs in the intronic regions close to the exons (0–100 nt) compared with the rest of the intron. This is not the case for CpGs in differential GC exons, which only have a slightly better chance of being methylated than flanking intronic CpGs [15].

Is the difference in methylation abundance in exons in the leveled GC regions biologically significant? Does DNA methylation contribute to the recognition of exons in a leveled GC architecture? CpGs in alternatively spliced exons are less frequently methylated than those in constitutively spliced exons suggesting that, in the absence of the methylation mark, exon recognition is diminished. As the splicing reaction is performed on RNA transcripts, an active mechanism must link DNA methylation to promotion or oppression of exon inclusion. The recently described action of the CTCF protein provides that link. The binding of CTCF to exon 5 of the *CD45* gene is methylation sensitive; exon 5 is included in the mRNA only when it is not methylated [16]. Since CTCF binds to a specific sequence in a small subset of exons, there must be other splicing regulatory proteins that are recruited by methyl binding proteins and deposited on the mRNA precursor at the right time and place. Chromatin structure might also play a role in translating methylation information to splicing. DNA methylation can participate in chromatin remodeling [17,18] and is also found in its highest abundance at exon boundaries. Since CpG dinucleotides are correlated with higher inclusion levels [15], an

indirect mechanism might affect splicing: there may be a cross-talk between nucleosomes and DNA methylation that directs nucleosomes to the DNA strand near splice sites. This type of mechanism would ensure that the nucleosome would be near an exon when it is transcribed to affect the elongation rate of RNA polymerase II.

In conclusion, splicing regulation is clearly complex and involves many factors. By mining the extensive sequencing of RNA, DNA and epigenetic data that are currently available, we are unraveling its secrets strand by strand. In mammalian genomes, there are two different mechanisms of splicing regulation that are governed by the regional GC architecture. Although differences were masked in whole genome analyses of the GC-heterogeneous genome, nucleosome occupancy and DNA methylation clearly differ

between differential and leveled GC regions. It will be interesting to discover whether other epigenetic factors also participate in splicing regulation of one group or the other.

Financial & competing interests disclosure

This work was supported by grants from the Israel Science Foundation (ISF 61/09, ISF Bikura 838/10 and ISF Morasha 64/12) and the Deutsche-Israel Project (DIP 6403). This study was supported in part by a fellowship from the Edmond J Safra Center for Bioinformatics at Tel Aviv University. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

- Kornblihtt AR, Schor IE, Allo M, Blencowe BJ. When chromatin meets splicing. *Nat. Struct. Mol. Biol.* 16(9), 902–903 (2009).
- Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16(9), 990–995 (2009).
- Spies N, Nielsen CB, Padgett RA, Burge CB. Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* 36(2), 245–254 (2009).
- Tilgner H, Nikolaou C, Althammer S *et al.* Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16(9), 996–1001 (2009).
- Brenet F, Moh M, Funk P *et al.* DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One* 6(1), e14524 (2011).
- Lister R, Pelizzola M, Dowen RH *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271), 315–322 (2009).
- Amit M, Donyo M, Hollander D *et al.* Differential GC Content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* 1(5), 543–556 (2012).
- Gelfman S, Burstein D, Penn O *et al.* Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res.* 22(1), 35–50 (2012).
- Batsche E, Yaniv M, Muchardt C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* 13(1), 22–29 (2006).
- de la Mata M, Alonso CR, Kadener S *et al.* A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell* 12(2), 525–532 (2003).
- Hodges C, Bintu L, Lubkowska L, Kashlev M, Bustamante C. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* 325(5940), 626–628 (2009).
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. *Science* 327(5968), 996–1000 (2010).
- Barbosa-Morais NL, Irimia M, Pan Q *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338(6114), 1587–1593 (2012).
- Merkin J, Russell C, Chen P, Burge CB. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* 338(6114), 1593–1599 (2012).
- Gelfman S, Cohen N, Yearim A, Ast G. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res.* 23(5), 789–799 (2013).
- Shukla S, Kavak E, Gregory M *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479(7371), 74–79 (2011).
- Choy JS, Wei S, Lee JY, Tan S, Chu S, Lee TH. DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.* 132(6), 1782–1783 (2010).
- Sarraf SA, Stancheva I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* 15(4), 595–605 (2004).