Seminars in Cell & Developmental Biology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Reciprocal regulatory links between cotranscriptional splicing and chromatin

Sérgio Fernandes de Almeida*, Maria Carmo-Fonseca*

Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Splicing regulation
RNA polymerase II
CTD
Chromatin
Nucleosome

ABSTRACT

Here we review recent findings showing that chromatin organization adds another layer of complexity to the already intricate network of splicing regulatory mechanisms. Chromatin structure can impact splicing by either affecting the elongation rate of RNA polymerase II or by signaling the recruitment of splicing regulatory proteins. The C-terminal domain of the RNA polymerase II largest subunit serves as a coordination platform that binds factors required for adapting chromatin structure to both efficient transcription and processing of the newly synthesized RNA. Reciprocal interconnectivity of steps required for gene activation plays a critical role ensuring efficiency and fidelity of gene expression.

Contents

1. Introduction

Gene expression is a multi-step process that is regulated at multiple levels. In eukaryotic cells, expression of protein-coding genes starts with transcription of precursor messenger RNAs (pre-mRNAs) by RNA Polymerase II (Pol II). Before initiating RNA synthesis, Pol II must first gain access to the promoter and unwind the DNA. Next, Pol II proceeds elongating through the entire length of the gene until the RNA is released and the enzyme is recycled for further rounds of transcription. Concomitant with transcription, nascent pre-mRNAs are modified by conversion of the 5′-terminus into a 7-methyl guanosine cap structure, splicing and addition of a poly(A) tail at the 3′-end. Each phase of the transcription cycle, i.e., initiation, elongation and termination is coupled to a specific step of pre-mRNA processing. The 5′-end of the RNA is capped during transcription initiation, and intronic sequences are excised by the spliceosome during elongation, before polyadenylation and transcription termination (reviewed in [1]).

For the vast majority of human protein-coding genes, nascent transcripts contain up to 90% of non-coding sequences in the form of introns that must be spliced with single nucleotide precision. This task is carried out by the spliceosome, an elaborate macromolecular machine composed of uridine-rich small nuclear RNAs

* Corresponding authors at: Instituto de Medicina Molecular, Faculdade de Medicina, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal. Tel.: +351 21 7999411; fax: +351 21 7999412.
E-mail addresses: sergioalmeida@fm.ul.pt (S.F. de Almeida), carmo.fonseca@fm.ul.pt (M. Carmo-Fonseca).

http://dx.doi.org/10.1016/j.semcdb.2014.03.010
1084-9521/

(U snRNAs) packed as ribonucleoprotein particles (snRNPs) that function in conjunction with over 100 distinct non-snRNP auxiliary proteins to recognize evolutionary conserved sequences present at intron boundaries (reviewed in [2,3]). Depending on the combinatorial effect of proteins that either promote or repress the recognition of these conserved splicing sequences, splice sites in pre-mRNA can be differentially selected to produce multiple mRNA isoforms by alternative splicing (reviewed in [4–6]). Recent analysis of human RNA sequencing (RNA-Seq) data revealed that approximately 95% of human pre-mRNAs that contain more than one exon are spliced to yield multiple mRNAS, and that most isoforms display variable expression across tissues [7,8]. Genes with few exons typically encode a small number of mRNA isoforms, while more diverse mRNA repertoires can be produced from genes containing numerous exons. Distinct splicing patterns of a given pre-mRNA can be observed pending on the cellular environment. For example, some mRNA isoforms are specifically expressed in certain tissues or developmental stages and others are triggered in response to external stimuli such as signaling pathways [9,10]. How human cells control more than 100,000 alternative splicing decisions remains a topic that attracts much attention. For many years, splicing was studied in isolation from other stages of gene expression but this approach was challenged by several studies indicating that pre-mRNA splicing is tightly coupled to transcription [11]. More recently, another paradigm shift was started by a series of observations suggesting that chromatin modifications can guide the spliceosome in the process of alternative splicing regulation. In this review we provide an overview of recent advances addressing interconnections of splicing with transcription and chromatin structure. We first refer to the different strategies used by the transcriptional machinery to overcome the nucleosome barrier, and we highlight the role of the carboxyl terminus of the largest Pol II subunit as a binding platform that recruits factors required for both adapting chromatin structure to efficient transcription and processing the newly synthesized RNA. Next, we focus on the concept that coupling between splicing, transcription and chromatin is reciprocal, and we discuss how such interwoven connections might contribute for efficiency and fidelity of gene expression.

2. Pol II overcomes a physical barrier imposed by chromatin

Following the discovery that packaging promoters in nucleosomes impedes the initiation of transcription in vitro [12,13], it has become progressively evident that chromatin imposes obstacles to transcription. Early studies using micrococcal nuclease digestion showed that enhancers and promoters were “hypersensitive” to digestion most probably due to loss of nucleosomes. These regions were thought to consist of naked DNA associated with specific transcription factors (reviewed in [14]). More recent genome-wide studies confirmed that genes transcribed by Pol II typically contain a histone-depleted area of approximately 200 bp upstream of the transcription start site (TSS) flanked by two highly positioned nucleosomes containing the histone variant H2A.Z, directly downstream of the TSS (reviewed in [15]). It has been suggested that the nucleosome-free region facilitates the binding of activators to the promoter, which in turn triggers a cascade of recruitment of coactivator complexes and general transcription factors (GTFs) that assist positioning of Pol II. Pol II interacts first with TFIID, TFIIA, and TFIIB to form the closed preinitiation complex (PIC). TFIIH then melts 11–15 bp of DNA and the resulting single-strand template is positioned in the Pol II cleft (open PIC) to initiate RNA synthesis [16]. Some of the coactivators recruited to the promoter (for example, SAGA) have chromatin-remodeling and histone acetyltransferase activities. As a consequence of histone acetylation, nucleosome fluidity increases at the promoter thereby facilitating concomitant ATP-dependent chromatin remodeling activities (for example, by SWI/SNF). Thus, the binding of activators to the promoter is thought to make the chromatin more accessible to accommodate the massive PIC. Remarkably, some promoters assemble partial PICs devoid of Pol II and TFIIH and, in this case, nucleosomes are not lost [17] implying that Pol II itself is required for promoter chromatin remodeling [16].

In contrast to low nucleosome density at promoters, nucleosome occupancy is high within gene bodies. To transcribe efficiently through this barrier, the elongating Pol II makes use of specific factors that are recruited by interaction with the carboxy-terminal domain (CTD) of its largest subunit, as we shall discuss in the next section.

3. Pol II contains a unique CTD that is dynamically phosphorylated

The CTD is a unique feature of Pol II that distinguishes it from all other polymerases. The Pol II CTD is a flexible, tail-like extension that lies below the RNA exit channel. This allows for factors bound to the CTD to act on nascent transcripts as soon as they emerge from the Pol II globular structure. The CTD consists of highly conserved heptapeptide repeats (consensus Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7), varying in number from 26 in yeast to 52 in vertebrates [18]. Each CTD heptapeptide contains five putative phosphorylation sites: three Serines (Ser2; Ser5 and Ser7), one Tyrosine (Tyr1) and one Threonine (Thr4). The differential activity of specific kinases and phosphatases create a CTD phosphorylation code that is used to establish specific interactions with distinct proteins as Pol II proceeds through the transcription cycle (reviewed in [19,20]) (Fig. 1). During PIC formation, the CTD is not phosphorylated, a condition that facilitates the interaction between Pol II and GTFs. Upon assembly of the last GTF, TFIIH, its cyclin-dependent kinase activity (CDK7) phosphorylates the CTD on Ser5 and Ser7. This occurs during the first 30 bp of transcription and as a result Pol II looses its contacts with the PIC and clears off the promoter. Further phosphorylation of the CTD on Ser2 and Thr4 is required to switch Pol II into an active elongating complex [21]. Ser2 phosphorylation results mainly from CDK9 kinase activity, which associates with the positive transcription elongation factor b (P-TEFb) and is also responsible for Thr4 phosphorylation [22]. During elongation, CDK12 and Polo-like kinase 3 are also implicated in CTD Ser2 and Thr4 phosphorylation, respectively [21,23,24]. Interestingly, the different CTD phosphorylation marks are not interdependent: the Ser5 mark is independent of Ser2, Ser7, and Thr4, and the Ser2 mark is independent of Ser5, Ser7 and Thr4 [25].

The current view is that the CTD acts as a binding platform to which various proteins are sequentially recruited and released, pending on the phosphorylation state. Most likely, the modifications created upon phosphorylation of the CTD lead to the selective binding of factors that help adapting chromatin structure to efficient transcription as well as proteins required for the maturation of the newly synthesized RNA. Next, we will discuss in more detail the CTD functions in RNA processing.

3.1. The role of Pol II CTD in RNA maturation

Transcription initiation starts with the formation of the first phosphodiester bound between RNA nucleotides. The nascent RNA passes through a positively charged exit channel within the Pol II structure, reaching the exterior once it is approximately 18 nucleotides long. Concomitantly, DSIF (DRB-sensitivity inducing factor) binds to Pol II and recruits NELF (negative elongation factor),

Please cite this article in press as: de Almeida SF, Carmo-Fonseca M. Reciprocal regulatory links between cotranscriptional splicing and chromatin. Semin Cell Dev Biol (2014), http://dx.doi.org/10.1016/j.semcdb.2014.03.010