Biochimica et Biophysica Acta 1809 (2011) 369–378

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Epigenetic control of gene regulation in plants ☆

Massimiliano Lauria [a], Vincenzo Rossi [b,*]

[a] Consiglio Nazionale delle Ricerche, Istituto di Biologia e Biotecnologia Agraria, Via Bassini 15, I-20133 Milano, Italy
[b] Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Unità di Ricerca per la Maiscoltura, Via Stezzano 24, I-24126 Bergamo, Italy

ARTICLE INFO

Article history:
Received 15 February 2011
Received in revised form 3 March 2011
Accepted 5 March 2011
Available online 15 March 2011

Keywords:
Epigenetics
Gene regulation
Chromatin
Transcription factor
Transposable element

ABSTRACT

In eukaryotes, including plants, the genome is compacted into chromatin, which forms a physical barrier for gene transcription. Therefore, mechanisms that alter chromatin structure play an essential role in gene regulation. When changes in the chromatin states are inherited trough mitotic or meiotic cell division, the mechanisms responsible for these changes are defined as epigenetic. In this paper, we review data arising from genome-wide analysis of the epigenetic landscapes in different plant species to establish the correlation between specific epigenetic marks and transcription. In the subsequent sections, mechanisms of epigenetic control of gene regulation mediated by DNA-binding transcription factors and by transposons located in proximity to genes are illustrated. Finally, plant peculiarities for epigenetic control of gene regulation and future perspectives in this research area are discussed. This article is part of a Special Issue entitled: Epigenetic Control of cellular and developmental processes in plants.

1. Introduction

Fifty years ago the Nobel Prize Laureates Francois Jacob and Jacques Monod published their seminal paper proposing a general model for bacterial gene regulation [1]. They provided evidence that genes are controlled at the level of transcription by the products of other genes encoding regulators and that these regulators act through binding of specific sites on DNA near the genes they control. A few years later it was demonstrated that these regulators are DNA-binding proteins able to activate or repress transcription, although more recent emergence of RNAi-related mechanisms indicates that RNA can also directly act as regulator of gene expression [2–5]. The key principles arising from these original works are still valid. However, a modern view of gene regulation must take into account that eukaryotic DNA is tightly packaged around a core of structural proteins, the histones, to generate the chromatin nucleosome array [6]. The nucleosome is composed of a histone octamer containing two copies of histones H2A, H2B, H3, and H4 and around which 147 bp of DNA is wrapped [7]. The discovery that nucleosomes inhibit transcription *in vitro* [8], that deletion of histones leads to a global increase in gene transcription *in vivo* [9], and that the first biochemically characterized chromatin remodeling complexes and histone-modification enzymes act as co-activators or co-repressors of gene transcription [10–14], clearly demonstrated that chromatin represents a physical barrier to transcription and factors able to modify chromatin structure and composition can modulate gene activation or silencing. A negative correlation with gene expression was also reported for cytosine methylation (mC), which can repress transcription by altering chromatin structure [15–17]. Therefore, although basic principles of gene regulation are universal, the DNA packaging into chromatin determines a more sophisticated regulatory option, which is essential for eukaryotic organisms to express genes in the incredibly diverse patterns required for a major biological complexity [18]. Since the mechanisms leading to modulation of gene expression due to alteration of chromatin structure do not affect primary DNA sequence, they can be defined as epigenetic if the change in the newly formed chromosomal state is inherited even after that the provoking stimulus is removed [19].

The seminal studies reported above were performed in yeast and animal systems; however, plants conserve all of the characteristics of the eukaryotic gene regulation previously described [5]. In this review, we will focus on the role of epigenetic mechanisms in plant gene regulation by specifically referring to low copy-number genes, whose ground state is usually characterized by a chromatin environment relatively permissive for transcription (for epigenetic regulation of heterochromatic high copy-number repeated sequences see the review of Saze and Kakutani [20]. Data from genome-wide analysis of epigenetic landscapes will be illustrated to identify their correlation with gene transcription. The role of chromatin in gene expression will be subsequently addressed by considering the cross-talk between chromatin modifications and *trans*-acting DNA-binding transcription factors (TFs). Gene regulation in *cis*, through diffusion of specific chromatin states from neighboring repeated sequences (e.g. transposons), will be also described. Finally, plant peculiarities in

☆ This article is part of a Special Issue entitled: Epigenetic Control of cellular and developmental processes in plants.

* Corresponding author. Tel.: +39 35 313132; fax: +39 35 316054.
E-mail address: Vincenzo.Rossi@entecra.it (V. Rossi).

1874-9399/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagrm.2011.03.002

epigenetic control of gene regulation with respect to other eukaryotes and future perspectives in this fascinating field will be discussed.

2. Epigenetic landscapes and their relationship with gene transcription

Mechanisms that regulate chromatin structure include: mC, factors affecting nucleosome positioning and composition, and post-translational modifications (PTMs) of histones; all of which are often interlaced with the small RNA (sRNA) pathway that acts in establishing and in maintaining specific chromatin states [21]. Recent advances, which make it possible to couple methods for epigenetic marks analysis with microarrays, or next generation sequencing-based techniques, have provided us with a genome-wide picture of the epigenetic landscapes. Importantly, in most of these studies, changes in the mRNA and/or sRNA transcriptome were concomitantly investigated in order to associate transcriptome and epigenome alterations. In this section, we will review findings from these studies to establish, where possible, general rules correlating distinct epigenetic modifications with plant gene transcription (Fig. 1).

2.1. Cytosine methylation

In plants, mC occurs at cytosine bases in all sequence contexts: the symmetric CG and CHG contexts (where H is A, T, or C) and the asymmetric CHH context, with specific enzymes that establish and successively maintain mC patterns during DNA replication; removal of mC also take places by means of DNA glycosylase activity [22]. Because CG methylation is coupled to DNA replication, it is a classical example of a stably inherited epigenetic mark and it was correlated with repression of gene transcription due to its ability to recruit, through interaction with methyl domain binding (MBD) proteins, chromatin remodeling complexes that establish a more condensed chromatin structure [23]. However, indications from methylome analysis provide new and sometime surprising information.

Genome-wide analysis of mC was performed in a number of flowering plant species (Arabidopsis, rice, maize, and poplar) using different strategies. Digestion with the methylation sensitive McrBC enzyme [24–27] or immunoprecipitation with a mC specific antibody (MeDIP) [28,29] was coupled with hybridization to high density genome arrays. Alternatively, DNA treated with sodium bisulfite, which converts unmethylated cytosine to uracil, but does not affect methylated cytosine, has also been employed for high-throughput sequencing, followed by mapping of the sequenced reads to a reference genome (BS-seq) [30–33]. The first two methods are useful to analyze changes of the mC level, but only BS-seq allows to generate a methylation map at single-base-pair resolution. Collectively, these studies indicate that mC occurs predominantly at repeats and transposons (more than 90% are methylated), but approximately the 20% of genes also exhibit a certain degree of mC. Overall, the levels of mC in the Arabidopsis genome at CG, CHG, and CHH are about 24%, 6.7%, and 1.7% respectively, but methylation within the genes is primarily restricted to CG sites and was predominantly observed in the transcribed coding region (the so called gene body). Comparison with the transcriptome data revealed that modestly expressed genes are more likely to be methylated within gene body, while genes expressed at high and low levels are usually less methylated [28–31]. The mC level at the promoters is quite low (less than 5% in Arabidopsis) and is associated with increased tissue specificity of gene expression. Analysis of the Arabidopsis and rice endosperm methylome indicates that CG hypomethylation upstream of the transcriptional start site (TSS) is related to endosperm specific gene expression [34–36]. Therefore, while the level of mC within promoters is negatively correlated with gene activation, the correlation becomes ambiguous if gene body mC is considered.

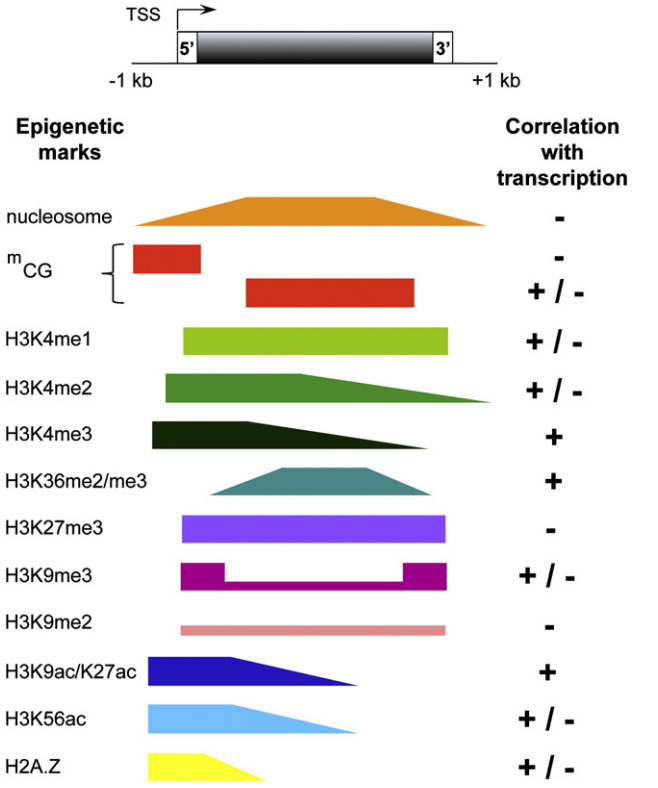

Fig. 1. Correlation between epigenetic landscapes and gene transcription. The distribution of epigenetic marks is mapped on an arbitrary gene, by considering the data from genome-wide approaches. ± indicates not precise association with active transcription.

The biological significance of gene body mC is still not clear. It was proposed that it can suppress aberrant transcription from cryptic promoters inside the genes [29,37]. Aberrant transcription occurs in moderately expressed genes, due to RNA polymerase II (RNAPII) transit during elongation, which disrupts chromatin and allows pre-initiation complex (PIC) assembly at non canonical promoters. Since aberrant transcripts can lead to methylation of homologous DNA trough sRNA pathway [22,29,37], their production may provoke mC accumulation within the gene body, thus impairing further cycles of aberrant transcripts formation. Highly and weakly transcribed genes would not necessitate gene body mC, due to a fast or slow RNAPII elongation rate, respectively, which reduce chances of PIC assembly. Gene body mC may also be correlated with definition of exon–intron boundaries. Indeed, Chodavarapu et al., [38] reported that, in Arabidopsis, nucleosomes and mC are enriched in exons and at exon–intron junctions and that RNAPII specifically accumulates in exons rather than introns, suggesting that the high density of methylated nucleosomes in exons provokes RNAPII stalling to enhance accurate splicing of upstream introns. The study of Chodavarapu et al., [38] also provides clues with respect to the mechanism by means of which mC can repress transcription initiation. The authors showed that nucleosomal DNA is preferentially methylated compared to flanking linker DNA, suggesting that nucleosomal DNA can be a preferential substrate for DNA methyltransferases (DMTs). This observation is in agreement with the proposal that mC interferes with gene transcription initiation, acting as a signal for nucleosome targeting of histone modifiers that, in turn, deposit repressive histone PTMs inhibiting nucleosome displacement for PIC formation [29].

2.2. Nucleosome positioning and composition

The work of Chodavarapu et al., [38] not only provides important information regarding the possible mechanisms of interaction