Transcription regulation in time and space: Engineered cell systems to modulate the epigenetic chromatin structure: The role of Methyl-CpG-binding protein 2

Piebes, D.G.E.

Publication date
2016

Document Version
Final published version

Introduction

D.G.E Piebes, P.J. Verschure

The simplest paradigm of transcription regulation dictates that at the end of a signalling cascade towards the nucleus single molecules bind to regulatory proteins which alter their affinity for the DNA. The interaction of the DNA-protein complex with RNA polymerase thereby triggers transcription. Transcription changes instantaneously upon fluctuations in the concentration of regulatory molecules. In differentiated cell types of multicellular organisms the intricate task of gene regulation is achieved by combined actions of multiple different transcription regulatory proteins including proteins that affect DNA and chromatin structure.

Several components of gene expression regulation including transcription factors, epigenetic enzymes and chromatin binding proteins are well known, however insight on dynamic and subsequent interactions of such proteins with DNA and chromatin is still unresolved. The epigenetic composition exhibits a regulatory layer that establishes heritable changes in gene activity that are not encoded in the DNA sequence. Alterations in the epigenetic state can be obtained via changes in chemical modifications of the DNA and chromatin and via conformational changes of the chromatin. Transcription patterns illustrating which genes are transcribed and to what extent, differ from cell to cell and they can change throughout the life span of an organism. Cell type specific gene expression patterns in higher eukaryotes confer stability of cellular phenotypes, while allowing changes in expression in response to environmental or developmental cues. Derangements in gene regulation have severe effects on cell behaviour and contribute to development of a diseased state.

Below a highlight is given of the actors and molecular events of gene expression regulation. In this thesis the intricate composition of chromatin including the role of core and variant histone proteins, posttranslational histone modifications and the nucleosomal and chromosomal structure are discussed. Moreover, the sequence of events when a gene is transcribed or repressed is discussed and we highlight some involved proteins, specifically focusing on the epigenetic regulatory protein methyl-CpG binding protein 2 (MeCP2). Finally we discuss the use of reporter gene cassette containing cells to perform systematic and quantitative, spatial and time resolved, single cell and single gene

transcription measurements and to determine the role of targeted alterations in the epigenetic chromatin state.

Chromatin

Chromatin consists of DNA, enzymes and proteins together exhibiting a very dynamic interplay in the confined space of the cell nucleus. To enable the folding of the DNA in the cell nucleus, DNA is wrapped around nucleosomes consisting of 2 of each H2A, H2B, H3 and H4 core histone proteins and linker histone H1 connector proteins. Besides these canonical histones which are incorporated into chromatin during replication of the genome, there are several variants of each histone, which are built into nucleosomes allowing defined activities. There are several histone variants known to be associated with transcription and to facilitate or prevent transcription regulatory proteins to bind at promoter regions. Some histone variants are more associated with the structure of the chromatin. For instance H3CENPA occurs only at centromeres, H3.3, H2AZ and H2ABdb are associated with transcriptionally active chromatin and macroH2A is found at inactive chromatin being abundantly present at the inactive X chromosome. The presence of conventional or variant histones in the nucleosome represents a layer of information controlling transcriptional gene activity (Gurard-Levin and Almouzni, 2014; Li et al., 2007).

Histone proteins consist of N-terminal tails that stick out of the nucleosomal structure. The amino acids at these histone tails can be modified by enzymes that act either as ‘writer’ or ‘eraser’ enzymes inducing the addition or removal of small chemical groups such as acetyl, methyl, phospho and more (Kouzarides, 2007a). The posttranslational histone modification state also provides binding sites for so-called epigenetic ‘reader’ proteins and enzymes that attract additional regulatory proteins. The specific amino acid as well as its position at the N-terminal tail or globular domain that is chemically modified largely determines the outcome of posttranslational histone modifications. Also the addition of one, two or three methyl groups can have an opposite effect. For example histone H3 lysine 4 monomethylation involves transcriptional activation whereas histone H3 lysine 9 trimethylation gene repression. These posttranslational histone modifications,

together with chemical modification of the DNA, e.g. DNA methylation or hydroxymethylation, form an important regulatory layer known as the epigenome that guides processes in the nucleus, i.e. the folding of the chromatin and hence gene expression patterning (Cutter and Hayes, 2015; Li and Zhu, 2015; Li et al., 2007; Luger and Hansen, 2005).

Posttranslational histone modifications create binding sites for transcription regulatory proteins allowing the establishment of protein complexes at defined genomic loci thereby altering the folding of the chromatin structure. Moreover, histone acetylation is able to directly alter the compaction of the chromatin structure since it neutralizes the positive charge of histones thereby decreasing the interaction of the histone N-terminal tail with negatively charged phosphate groups of DNA (Iizuka and Smith, 2003; Kouzarides, 2007b).

The folding of DNA into nucleosomal chromatin structures provides chromatin a dynamic behavior, allowing the chromatin to 'loop' and nucleosomes to 'breath'. Another advantage of nucleosomal chromatin packaging is that genes or regulatory sequences can be transcribed from the chromatin structure via local chromatin unwinding preventing the need to unwind the entire genome (Luger and Hansen, 2005). In principle genes with high nucleosomal density are less accessible for the transcription machinery, whereas a low nucleosomal density has a more open and accessible composition (Li et al., 2007). Although detailed information of the nucleosomal structure and the positioning of nucleosomes exist, knowledge of higher levels of chromatin structure remains incomplete. High-resolution transmission electron microscopy studies revealed that the fraction of the nuclear volume that is occupied with chromosomal material is small and that the majority of the genome is packed into large-scale chromatin structures. Several observations relating transcriptional regulation to interphase chromosomal folding have emerged, (1) interphase chromosomes occupy chromosome territories, (2) gene-rich chromosomes are spatially segregated from gene-poor chromosomes and show an increased chromosomal decondensation and a more nuclear internal position compared to gene-poor chromosome territories, (3) megabase-sized gene loci are able to loop outside of chromosome territories upon transcriptional activation of such loci, (4)

transcriptionally active or silent gene loci preferentially position at defined nuclear structures including 'nuclear bodies', (5) the association of cis and trans chromosomal regions is likely mediated by the association with nuclear domains. (Hemmerich et al., 2011; van Steensel, 2011; Verschure et al., 1999; Zhao et al., 2009)The use of engineered chromosome regions using bacterial lac operator/lac repressor or tet operator/tet repressor systems allowing targeted protein-DNA biding has a large impact on exploring the functional large-scale chromatin organization. Genome folding plays an important role in gene regulation. The causal relationship between gene folding and gene expression representing various hierarchical levels of transcription regulation is still largely unresolved.

Transcription

Gene expression is accomplished by the activity of RNA polymerase. Polymerization of the RNA polymerase enzyme requires other proteins to make the DNA accessible for RNA polymerase and to direct the enzyme to the site of transcription. RNA polymerase II is one of the proteins of the transcription machinery. RNA polymerase II transcribes a large set of genes, which is largely determined by the epigenetic composition of the chromatin. RNA polymerase forms a complex with proteins that guide the polymerase to correct genomic sites, whereas RNA polymerase II itself performs polymerisation of the mRNA based upon an intact DNA sequence. Polymerization has a certain reaction rate and seems to either take place or not following a large chain of events. The reaction rate of the synthesis of an entire mRNA can be regulated by the stalling of subsequent events of polymerisation. mRNA synthesis can also be regulated by the number of RNA polymerases that are progressing over a genomic sequence and transcribing the respective gene. The reaction rate of polymerizing the entire mRNA is determined by the epigenetic chromatin state including the functioning of DNA regulatory elements such as enhancer looping, the CpG content of the promoter and insulator elements (Svejstrup, 2004; Zhou et al., 2012).The initiation and elongation of RNA polymerase II producing a few transcripts is a rather inefficient process. In an active chromatin context only 10% of the RNA polymerases that are attracted to the transcription site will initiate and of these

initiated RNA polymerases, only 10% will successfully elongate and produce a mRNA molecule (Darzacq et al., 2007). Upon RNA polymerase initiation, proteins involved in RNA capping are required to ensure the creation of mature mRNA. RNA polymerase II progression is driven by posttranslational modifications such as phospho groups on RNA polymerase subunits, i.e. phosphorylation of serine 2 or serine 5 of the C-terminal domain. It is speculated that phosphorylation of RNA polymerase II functions as a checkpoint to prevent mRNA malformation by attracting the capping machinery before RNA polymerase II starts the elongation phase (Svejstrup, 2004). Of interest histone acetylation indirectly drives RNA polymerase from pausing to elongating. For instance, acetyl groups attract bromodomain reader proteins such as bromodomain-containing protein 4 (BRD4), which attracts positive elongation factor b (P-tefb) a cyclin dependent kinase which phosphorylates proteins that induce RNA polymerase pausing, as well as proteins that induce RNA polymerase elongation (Zhou et al., 2012).

Gene repression

Similar as transcriptional activation, transcription repression involves an intricate regulation of a variety of regulatory proteins and involves changes in the chromatin structure. Several transcription repression regulatory proteins are involved in recruiting enzymes that act as histone posttranslational ‘eraser’ enzymes, e.g. a histone deacetylases or histone demethylases.

Transcription repression involves DNA methylation. Genes with a high CpG content in their DNA sequence upstream of their transcription start site and promoter, i.e. CpG islands, generally exhibit a low amount of cytosine methylation and exhibit a transcriptionally active default state, whereas genes with low CpG content promoter regions generally consist of cytosine methylated sites and exhibit a transcriptional ‘off’ state. When a CpG island in the promoter region of a gene becomes methylated its expression is repressed. The addition of methyl groups is carried out by DNA methyltransferase enzymes that are responsible for the maintenance of established DNA

methylation or new or de novo methylation patterns.(Bodnar and Spector, 2013; Nan et al., 1998a)

Methyl CpG binding proteins play an important role as biological mediator of DNA methylation. Methyl CpG binding Protein 2 (MeCP2) contains both a methylated DNA and a transcription repression binding domain and exhibits chromatin remodeling activities. Depending on its protein folding MeCP2 attracts other proteins that are able to trigger a variety of functions in the nucleus. Transcription repression involving the presence of MeCP2 is often accompanied by the Heterochromatin Protein 1 (HP1) α, β and γ isoforms. HP1 is a chromatin-binding protein that binds with its chromo- and chromoshadow domain to H3K9-methylated histones and to proteins having histone modifying activities thereby advancing the 'spreading' of heterochromatin in cis (Verschure et al., 2005). HP1 proteins are involved in a wide variety of chromatin regulatory processes.

MeCP2 and MeCP2-related diseases

Each cell type exhibits a unique epigenetic chromatin make-up allowing it to maintain its cell type-specific gene expression pattern and hence cell identity. MeCP2 is a chromatin regulatory protein that is expressed in every cell type, but most abundantly in neuronal cells (Song et al., 2014). MeCP2 does not have a catalytic domain but it exists of domains that enable to trigger a variety of functions (Nan et al., 1993, 1998b). MeCP2 is able to attract transcription regulatory proteins and to direct them to a target gene increasing the probability to bind distinct sites (Hager et al., 2009).

MeCP2 mutations, deletion or amplification of the MeCP2 protein are the cause of two progressive neurodevelopmental disorders, i.e. Rett syndrome and Xq28 duplication syndrome. Rett syndrome typically affects the binding properties of MeCP2 whereas Xq28 duplication syndrome affects the concentration of MeCP2 in the nucleus. Both MeCP2-related disorders involve an altered chromatin structure thereby distorting genomic transcription patterns and neuronal behavior (Amir et al., 1999; Ramocki et al., 2010).

Reporter gene cassette containing cells

The composition of the epigenome has in many instances been mapped genome-wide and is actively pursued in large consortia. These efforts have resulted in impressive databases of epigenetic information. However, the rules that dictate how the structural, mechanistic and kinetic aspects of epigenetic transcription regulation concertedly determine gene activity are still poorly understood. This is in part due to the dynamic nature of the epigenome and due to the lack of techniques to resolve the structural aspects of epigenetic regulation. Moreover, a drawback of genome-wide biochemical analysis of the epigenome and transcriptome is that such methods do not allow the detection of cell-to-cell variation. The insertion of a single, non-endogenous engineered reporter gene cassette into the eukaryotic genome provides unique opportunities to directly influence the epigenetic state compared to the transcriptional state and offers a number of advantages to visually monitor the dynamics of epigenetic gene regulation. Such reporter gene cassette genomic integration approaches enable to obtain single cell and single gene analysis quantitative time-resolved data as input for computational simulations thereby advancing our understanding.