Chromatin architecture and the orchestration of gene expression: cell systems to explore epigenetic gene control

Brink, M.C.

Publication date
2009

Document Version
Final published version

Chapter 1

Introduction

Maartje C. Brink & Pernette J. Verschure

Introduction

The functional organization of nuclear architecture

The eukaryotic cell is faced with the challenge of packaging its genetic material in the nucleus while maintaining accessibility of its genes for DNA-associated processes. The answer that evolution has come up with is both elegant and effective: the genetic material is functionally organized as chromatin. Efficient folding of the about two meters of DNA inside the nucleus, which is only 10-20 μm in diameter, occurs by wrapping the DNA around an octamer of histone proteins, forming a 10 nm diameter fiber (Carter, 1978; Finch *et al.*, 1977; Kornberg, 1977; Luger *et al.*, 1997). Coiling of this fiber results in a compacted fiber of 30 nm diameter, which is folded further in an unknown way. The ensemble of DNA and proteins constituting chromatin plays an essential role in transcriptional regulation. Detailed understanding of the molecular systems that underlie the functional architecture of chromatin, both at the nucleosomal and the higher-order level (*i.e.* above the 30 nm fiber), remains elusive. Current research in this field is focused on uncovering the molecular mechanisms that are responsible for the dynamic higher-order structure of chromatin, controlling gene expression.

In order to orchestrate the level and timing of the expression of thousands of genes, the eukaryotic genome employs various methods of regulation. The most direct type of regulation is found at the DNA sequence level, by the interaction between transcription factors and regulatory sequences. Functionally related genes are sometimes clustered on the linear genome so they can be coregulated, as can be seen for the HoxB and β-globin loci for example (Chambeyron and Bickmore, 2004; Tolhuis *et al.*, 2002). Additionally, gene clusters of unrelated function were also identified on a genome-wide scale by integrating the human genome map with messenger RNA expression data (Caron *et al.*, 2001). The resulting human transcriptome map revealed the existence of genomic stretches of several mega basepairs containing functionally unrelated genes with similar levels of transcriptional activity. Regions of increased gene expression (ridges) and low gene expression (anti-ridges) were identified on the human genome (Lercher *et al.*, 2002; Versteeg *et al.*, 2003). Interestingly, insertion of a GFP reporter gene within a ridge resulted in an average fourfold upregulation of the GFP transcription compared to integration in an anti-ridge, demonstrating the importance of local genomic organization at the linear DNA sequence level (Gierman *et al.*, 2007).

Another form of gene regulation occurs at the level of the higher-order folding of chromatin, *i.e.* above the 30 nm fiber. On the basis of classic cytological studies, the genome can be divided into two major types of chromatin: compact and open chromatin, named heterochromatin and euchromatin, respectively. These types of chromatin are considered to reflect gene activity: the often gene-poor heterochromatin domains are generally compact and transcriptionally silent, while gene-rich

euchromatin consists of open, transcribed chromatin domains (Felsenfeld and Groudine, 2003). However, more recent data suggest that a more subtle discrimination between the functionality of compact and less compact chromatin domains should be made. Using sucrose gradient sedimentation, chromatin fibers were separated based on their compactness, demonstrating that transcriptionally active as well as inactive sites are found in both such compact and less compact isolated chromatin fibers (Gilbert *et al.*, 2004). In addition, Gilbert and colleagues (2004) found that less compact chromatin fibers are more gene dense than compact chromatin fibers. These *in vitro* results indicate that the distinction between heterochromatin and euchromatin is complex and probably based on more parameters than transcriptional activity alone.

An interesting feature of chromosome architecture is the existence of chromosome territories (CTs) in interphase nuclei (for a review, see Cremer and Cremer, 2001). According to fluorescent *in situ* hybridization (FISH) experiments, each chromosome occupies a distinct territory within the nucleus that was initially thought not to be intruded by other chromosomes (Cremer *et al.*, 1993). However, upon increasing resolution by imaging 150 nm thick cryosections (cryo-FISH) with the electron microscope, some intermingling between adjacent chromosomes was demonstrated (Branco and Pombo, 2006). CTs consist of domains of different degrees of compaction: 100-500 nm domains are surrounded by interchromatin areas containing little or no chromatin (Cremer *et al.*, 2004). Active transcription sites can be found throughout the CT near the surface of the compact subchromosomal domains (Fakan, 1994; Fakan and van Driel, 2007; Verschure *et al.*, 1999). An interesting finding is that, upon transcriptional activation, genomic loci may loop out, away from their CT (Mahy *et al.*, 2002a; Volpi *et al.*, 2000). A striking example of looping of transcriptionally active genes was found for the HoxB locus (Chambeyron and Bickmore, 2004). Here, induction of the genes within the cluster caused sequential looping out of the CT, in the same order as their transcriptional activation. Furthermore, introduction of the human β-globin locus control region (LCR) into a gene-dense region of the mouse genome caused transcriptional upregulation combined with more frequent positioning away from the CT (Noordermeer *et al.*, 2008). These findings argue that transcription occurs near the surface of the compact chromatin domains, possibly on chromatin that loops out into the interchromatin domain. It should be noted however, that looping is not necessary to enable transcription of a chromatin fiber, since transcriptionally active genes can also be situated inside a CT (Mahy *et al.*, 2002b; Verschure *et al.*, 1999).

The spatial organization of chromatin in the nucleus may also play a role in gene regulation. For instance, in some cell types gene-poor chromatin is clustered at the periphery of the cell nucleus. This has been illustrated by FISH experiments, demonstrating that in human primary lymphocytes and lymphoblasts, the gene-poor chromosome 18 is positioned more to the periphery of the nucleus than the gene-rich chromosome 19 (Croft *et al.*, 1999). Furthermore, Goetze *et al.* (2007a) demonstrated

in five different human cell lines that certain gene-rich, highly expressed ridges are generally located more towards the nuclear interior than certain gene-sparse, low expressed anti-ridges of the same chromosomes. This radial positioning of chromosomal (sub)domains leads to compartmentalization of the nucleus, resulting in a gene-rich, transcriptionally active interior. However, radial positioning of chromosomes according to gene density seems to be restricted to specific cell types. In human skin fibroblasts for instance, chromosome size rather than gene content dictates radial positioning (Bolzer *et al.*, 2005). Approaching the issue of radial positioning from a different perspective, Guelen *et al.* (2008) identified genomic regions that are naturally recruited to the nuclear periphery in human lung fibroblasts, using the DNA adenine methyltransferase (Dam) ID technique. The authors demonstrated that these lamina-associated genomic domains are generally gene poor and depleted of RNAPII, allowing little transcriptional activity (Guelen *et al.*, 2008). Taken together, these data suggest a model in which gene-poor inactive chromatin is located more towards the periphery of the nucleus in some cell types, whereas transcriptionally active chromatin is more concentrated in the center of the nucleus.

Whether positioning of genes near the nuclear periphery is caused or influenced by the transcriptional status, has been addressed in several recent studies using a bacterial lactose (lac) operator integrated into mammalian chromatin (Robinett *et al.*, 1996). Because of the high affinity of the bacterial lac repressor (lacR) for the lac operator (lacO), proteins fused to the lacR can be effectively targeted to an integrated lacO repeat (Robinett *et al.*, 1996). By fusing lacR to a lamin-interacting protein, a lacO locus could be targeted to the nuclear lamina (Finlan *et al.*, 2008; Kumaran and Spector, 2008; Reddy *et al.*, 2008). Finlan *et al.* (2008) demonstrated that transcription of some, but not all, of the endogenous genes located in or near the targeted site was suppressed. Yet another study showed that after targeting of a chromosome arm containing a lacO repeat and a reporter gene to the nuclear periphery, transcription of the reporter gene could be induced (Kumaran and Spector, 2008). These data indicate that the nuclear domain close to the lamina, *i.e.* the nuclear periphery, is primarily a region harbouring repressed genes, but that gene expression can be induced, suggesting a complex functional relationship between nuclear localization and gene regulation. Taken together, it has become clear that the spatial organization of chromatin in the nucleus is functionally linked to the transcriptional status of chromatin, although the molecular mechanisms remain to be uncovered.

Epigenetic regulation: the orchestration of gene expression

The insight that not only the genetic code but also other factors, such as chromatin architecture, influence transcriptional activity has led to many questions regarding the extent and mechanisms of this regulation. The field of epigenetics focuses on identifying these factors and unraveling their mechanistic role in genome regulation. Over the past decade a large number of studies has furthered our understanding of

epigenetic gene regulation and a view has emerged in which chromatin is the key player. As reviewed by Taverna and colleagues (2007), histone- and DNA-modifying proteins can heritably change gene expression by covalently adding or removing methyl groups on DNA and/or a variety of chemical groups on the histone proteins of chromatin, most notably phosphate, methyl, ubiquitin and acetyl groups. The interplay of histone modifications in controlling genome activity is often referred to as the 'histone code' (Jenuwein and Allis, 2001; Turner, 2007). The histone code is thought to mark parts of the genome as eligible for transcription or for repression. Exactly how this is encoded is still largely unknown, but two models are currently considered (Taverna *et al.*, 2007). The first is a 'direct' model in which specific histone modifications, including phosphorylation and acetylation, change the positive histone charge thereby disturbing interactions with the negatively charged phosphate groups of DNA. These altered interactions induce conformational changes in chromatin structure, rendering it more compatible with transcription. The second 'effector-mediated' model depends on the existence of effector proteins that translate the histone code into meaningful downstream events. For instance, specific methylated sites of histones are recognized by proteins containing a chromodomain, such as heterochromatin protein 1 (HP1; Bannister *et al.*, 2001). Acetylated sites are recognized by bromodomains and several phosphorylated sites by 14-3-3 proteins (Kouzarides, 2007). These effector proteins can in turn recruit transcription factors and/or chromatin remodeling complexes, affecting changes in transcription.

However, the histone code functions in a more complex way than a simple standard conversion from an epigenetic mark to the gene expression state. For instance, the location of the modification in the promoter or coding region of a gene exerts a different effect on the transcriptional state. Histone 3 lysine 9 methylation has a negative effect on transcription when present at the promoter region, but when present throughout the coding region of genes it is associated with transcriptional elongation through the recruitment of HP1γ (Vakoc *et al.*, 2005). The same epigenetic mark may therefore lead to a different response, depending on its location.

Furthermore, the abundance of available histone modifications presents the possibility for a vast array of transcriptional responses as a result of epigenetic cross-talk. For instance, opposing histone modifications may be present at the same time on a single regulatory site. The existence of these so-called bivalent chromatin domains, containing both active and repressive marks simultaneously, forms a particularly challenging puzzle (Azuara *et al.*, 2006; Bernstein *et al.*, 2006). The enrichment of these opposing histone modifications in mouse embryonic stem cells is present on chromatin regions that are mostly silent. Upon differentiation, the cells preserve only one of the two histone modifications and change gene expression levels accordingly, dependent on the cell type (Bernstein *et al.*, 2006). Supposedly, the function of these opposing histone modifications is to poise the genetic locus, preparing it for transcription, but not yet activating it.

One approach to unraveling the complex interplay between epigenetic marks is through the characterization of the proteins that interact with chromatin. By interfering with the protein of interest, its effect on epigenetic modifications, chromatin structure and transcriptional status can be deduced. For example, HP1 protein function was inhibited by means of a dominant-negative approach to characterize the function of HP1 in maintaining heterochromatin structure. HP1 binds to histone H3 methylated at lysine 9, present in silent chromatin such as mouse chromocenters. By over-expressing a chromodomain-deficient HP1 mutant, we depleted HP1 from pericentromeric heterochromatin (*i.e.* mouse chromocenters; Mateos-Langerak *et al.*, 2007; chapter 4 of this thesis). Remarkably, in the absence of functional HP1 pericentromeric heterochromatin domains retained their condensed chromatin structure, demonstrating that heterochromatin maintenance is independent of the binding of HP1 (Mateos-Langerak *et al.*, 2007).

By studying the effects of both gene knock-out and over-expression, important discoveries have also been made recently for methyl-CpG-binding protein 2 (MeCP2). MeCP2 has long been considered a transcriptional silencer, because of its affinity for methylated DNA (a mark for silent chromatin) and recruitment of HDAC silencing complexes (Meehan *et al.*, 1992; Nan *et al.*, 1997; Nan *et al.*, 1998b). However, a recent report studied gene expression patterns after up- or downregulation of MeCP2 in the mouse hypothalamus (Chahrour *et al.*, 2008). Surprisingly, ~85% of the genes that bound over-expressed MeCP2 were upregulated. Conversely, knockout of the MeCP2 gene caused these genes to be downregulated, suggesting a role in transcriptional activation for MeCP2. These results are in line with our finding that the targeting of MeCP2 to compact chromatin domains results in decondensation (chapter 2 of this thesis).

A true advance in the study of epigenetic regulators and their *in vivo* effects on chromatin was made with the introduction of engineered targeting systems. Such systems consist of protein-binding arrays that associate with DNA-binding proteins with a high affinity. Fluorescent tagging of the DNA-binding protein allows one to visualize the genomic binding array in living cells. Moreover, these systems can be exploited to demonstrate the effect of binding of epigenetic regulatory proteins on the compaction state of the integrated chromosomal array. The initial targeting system, devised in mammalian cells, contained a highly amplified chromosomal domain consisting of a stretch of bacterial lac operator (lacO) repeats that bind a chimeric fluorescent lac repressor (lacR) protein with high affinity (Robinett *et al.*, 1996; figure 1). Subsequent studies established the value of this tool, demonstrating that the *in vivo* targeting of the lacR-tagged viral activation domain VP16 to a lacO array results in the decondensation of the chromosomal array and its movement to the interior of the nucleus (Tumbar and Belmont, 2001; Tumbar *et al.*, 1999). Targeting of the lacR-tagged epigenetic silencer HP1 to the chromosomal array produced an opposite effect; lacO chromatin condensed locally, concomittant with the recruitment of histone

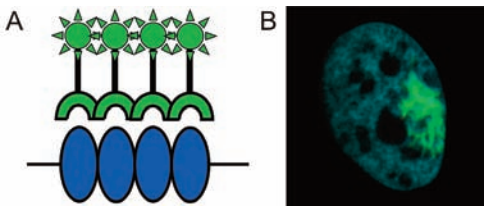

Figure 1. The engineered lac operator/lac repressor targeting system. (A) Schematic representation of the lac operator/lac repressor targeting system. Integrated lacO repeats (blue) are recognized and bound by the lacR-EGFP fusion protein (green). (B) A eukaryotic cell with an amplified lacO array that has been targeted by lacR-EGFP. The three-dimensional structure of chromatin is visible due to its fluorescent tagging.

methyltransferase SETDB1 and the enhancement of trimethylation of histone H3 lysine 9 in that domain (Brink *et al.*, 2006; Verschure *et al.*, 2005; chapter 3 of this thesis). The lacO/lacR technology has been further developed over the years, adding inducibility of the lacR fusion protein and incorporation of a fluorescent reporter gene and MS2 repeats for measurements of transcript concentration *in vivo*, further expanding its usefulness (Janicki *et al.*, 2004; Tsukamoto *et al.*, 2000). Studies using engineered lacO/lacR systems are described in chapters 2, 3 and 5 of this thesis.

In conclusion, eukaryotic gene regulation is achieved at several levels. A multitude of studies have established various correlations between chromatin organization, *e.g.* higher-order folding of the chromatin fiber and positioning of chromosomal domains within the nucleus, and gene expression. However, whether the level of gene expression is a cause or consequence of chromatin organization is still unclear. The use of engineered targeting systems allows us to manipulate and visualize changes at the chromatin level in relation to changes in gene expression and hence systematically unravel cause-and-effect relationships. Uncovering these details will advance our understanding of the principles of genome biology.

Thesis outline

The objective of the studies described in this thesis was to define causal relationships between reputed epigenetic chromatin silencers, chromatin structure and changes in gene expression. We have used several approaches to achieve this aim.

In **chapter 2** and **3** we have utilized an amplified lacO array targeting system in Chinese Hamster Ovary cells. Targeting the gene silencer MeCP2 (**chapter 2**) we observed, contrary to our expectations, that chromatin of the lacO-containing amplified chromosome region is decondensed upon MeCP2 binding. This decondensation requires the C-terminal domain of MeCP2. While HP1α and HP1β remain bound to the lacO, HP1γ is depleted after MeCP2 binding. Interestingly, transcriptional silencing is maintained, despite the induced chromatin decondensation.

Chapter 3 examines targeting of a truncated HP1α or HP1β protein, lacking its chromodomain, which normally mediates binding to histone H3 methylated at lysine 9. Like full-length HP1, local binding of chromodomain-deficient HP1 results in heterochromatinization of the amplified chromosome region as shown by three-

dimensional quantitative measurements of chromatin structure, elevated levels of H3K9me3 and recruitment of endogenous HP1α and HP1β (Brink *et al.*, 2006). These results suggest that recruitment of protein factors by the chromoshadow and hinge domain is sufficient to induce heterochromatinization.

In **chapter 4** we present the results of depletion of HP1 from heterochromatin by a dominant-negative approach (Mateos-Langerak *et al.*, 2007). We utilize the self-binding properties of HP1 through its chromoshadow domain. By over-expressing truncated HP1 without its chromodomain, we inhibit the binding of full-length HP1 to its chromatin binding sites, depleting HP1 from mouse chromocenters. Pericentromeric heterochromatin domains remain intact after HP1 displacement, demonstrating that HP1 is not essential to heterochromatin maintenance.

In **chapter 5** we describe the setting up of an advanced chromatin targeting system in human cells and discuss its benefits and points for improvement. A key feature of this engineered system is its single-copy integration into predefined well-characterized genomic locations in different functional domains. This enables us to study the effect of the genomic environment on chromatin structure and on gene expression of a reporter gene. Furthermore, two different small arrays that are incorporated in the engineered construct can be used to simultaneously target different regulatory factors, providing information about their functional interactions and their effect on gene expression. Pilot studies demonstrate the potential of our novel engineered targeting system.

Finally, **chapter 6** is a perspective in which we discuss novel approaches to describe and analyze eukaryotic gene networks. Among others, this should give insight into the role of epigenetics in gene networks.