Reading the maps: Organization and function of chromatin types in Drosophila

Braunschweig, U.

Citation for published version (APA):
Amsterdam: Nederlands Kanker Instituut - Antoni van Leeuwenhoek Ziekenhuis

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
General Discussion
General Discussion

In this thesis, genome-wide binding studies of chromatin proteins in Drosophila Kc cells have been used to derive general rules of chromatin organization. The results describe determinants for binding of linker histone H1, functionally significant domains of chromatin protein enrichment on multiple scales, targeting interactions among chromatin proteins, and the principal chromatin types in Kc cells. Taken together, they provide insights into the molecular morphology of chromatin domain organization, the mechanisms underlying differences in chromatin types, and how different chromatin types are linked to the regulation of gene expression.

Domain organization of chromatin

One of the most striking results from the studies in this thesis is that the Drosophila genome is systematically organized into chromatin domains. Previously, domain organization of chromatin has been defined molecularly in different ways and based on different data: nuclease sensitivity (Lawson et al. 1982), gene expression state of endogenous genes (Caron et al. 2001) or randomly inserted reporter genes (Gierman et al. 2007), histone modifications and binding of a chromatin protein (Noma et al. 2001), and combinations of different data types (Thurman et al. 2007), among others. There is a consensus that chromatin domains are functionally significant because expression of genes in them is linked to their other molecular characteristics.

While the data presented in both chapter 2 and chapter 4 underline a strong domain organization of the Drosophila genome, the structure of these domains differs between the two studies. The domainogram analyses in chapter 2 indicate a domain organization on many scales in a nested fashion, where most but not all of the genome is encompassed in domains. The view presented in chapter 4 is quite different, as it regards the whole genome as segmented into domains with similar protein composition. But are chromatin domains discrete entities or do they rather correspond to diffuse regions of enrichment of some component, and does this matter?

There are several technical reasons for the distinctions. First and foremost, the different statistical algorithms used necessarily lead to different types of segmentation. The BRICKS algorithm (chapter 2) looks for significant protein enrichment independently at different scale levels, resulting in the nested structure of BRICKS (chapter 2, Fig. 4). In chapter 4, the Hidden Markov Model (HMM) produces the most likely segmentation of the whole genome into five different chromatin types (chapter 4, Fig. 1).

The latter approach does not take into account if these discrete domains are clustered in a chromosomal region; However, visual inspection suggests that such clustering exists (chapter 4, Supplementary Fig. S2). Another technical difference is the resolution, which is much higher in chapter 4 due to the use of tiling microarrays instead of cDNA arrays. It is easily possible that the continuum of binding values at low resolution is due to contributions from several small, discrete domains at high resolution.

A local clustering of small, discrete chromatin domains results in a relatively higher signal at low resolution. Furthermore, within a discrete domain as identified by the HMM, there may be local, even nested enrichment of one or several proteins that does not contribute enough to the overall variance of protein binding to cause splitting or merging of the discrete domains.

The two views do not exclude one another. The high-resolution study shows that there indeed are relatively homogeneous chromatin domains at a scale of ~10 kb (chapter 4, Fig. 2B). Given the high significance of enrichment found in chapter 2, these are probably clustered in a non-random fashion giving rise to enrichments at higher scales. This notion is supported by the fact that proteins with uniform binding throughout a domain at high resolution (e.g. H1, SuUR, Pc) also show nested enrichments at larger scales (chapter 4, Supplementary Fig. S4).
The binding pattern of one particular protein can similarly be in discrete, homogenous domains or in - possibly nested - clusters of focal binding sites. Some proteins like histone H1 (chapter 1) or histone modifications like H3K27me3 (Tolhuis et al. 2006) can cover tens or even hundreds of kb rather uniformly, suggesting that most nucleosomes in this region are bound respectively modified. This most generally reflects non-sequence specific binding and in some cases a mode of directional spreading as e.g. for H3K9me3 (Noma et al. 2001). Such domains are usually well delineated and can be restricted by boundary elements (Noma et al. 2001). Changes of the chromatin type within the domain can be effected by deleting boundary elements, or by manipulating organizing elements like polycomb response elements that nucleate Polycomb binding in a region (Comet et al. 2006).

A ‘domain’ of enrichment of a transcription factor (TF) with focal binding means that there is a clustering of binding sites that is stronger than expected by chance. This can indicate that multiple genes in that area are regulated by this same TF. Because the genes in such domains tend to work together in similar biological functions more often than expected (chapter 2, Fig. SC-D and Supplementary Fig. S8), even small influences of chromatin proteins on TF binding at individual genes would cooperate in the regulation of the common function.

Is linear domain structure linked to spatial organization of chromatin?

An interesting possibility is that the nested cluster structure may be linked to the spatial folding of the chromosome fiber. In vivo crosslinking experiments indicate that any locus is frequently in physical contact with other loci. These contacts are most frequent over short distances, but can also occur over megabases (Simonis et al. 2007).

This phenomenon could be linked to the hierarchical enrichment of chromatin components as reported in chapter 2: When reasonably distant loci with similar chromatin composition make chance contacts, the interactions could be stabilized through shared chromatin components. Due to the hierarchical structure of protein enrichment, the whole region may thus engage in nested hierarchical interactions between genomic regions of the same chromatin type.

Interaction frequencies would decrease with genomic distance, but distant regions of the same chromatin type would interact more than equally distant regions of different types (Fig. 1). Thus, by (nested) clustering of chromatin domains or focal binding sites for specific proteins, the higher-order folding of the chromatin fiber and the formation of distinct regulatory compartments could be facilitated.
In a recent whole-genome study of human chromatin, the interaction pattern supports the notion of a fractal globule folding of chromosomes in interphase (Lieberman-Aiden et al. 2009).

The same study showed that at a resolution on the megabase scale, the genome consists of two groups of loci interacting preferentially with loci in the same group. We currently do not know if the five chromatin types identified in chapter 4 correspond to different groups of loci interacting in 3D.

The evolutionary implications of interaction domains are twofold: On the one hand, neighborhood of genes with contrasting chromatin composition and thus regulatory programs may be disfavored in the whole interacting region because a high probability of direct contact with inappropriate chromatin proteins may lead to misregulation.

On the other hand, frequent contacts of genes coregulated initially only due to their similar chromatin may provoke the genesis of more elaborate coregulation mechanisms that depend on the contact. The net effect on genome organization in both cases is evolutionary pressure for clustering of chromatin proteins and coregulation.

What is a chromatin type?

Some groups of proteins in chapter 2 have largely overlapping domains of enrichment: apart from the PcG proteins and the classical heterochromatin proteins which are each known to colocalize, overlap is seen between SuUR, D1, H1, and LAM; as well as between JRA, BCD, DSP1, GAF (Trl), MBD-like, BRM, GRO and SIR2 (chapter 2, Fig. 4 and Supplementary Fig. S4). This extensive overlap indicates that the number of combinations of proteins is far smaller than the number of proteins. This observation is confirmed and extended in chapter 4, where 5 chromatin types were defined based on typical combinations of 53 bound proteins.

The groups of overlapping proteins listed above correspond well to the BLACK and RED chromatin types (chapter 4, Fig. 3A). Also the similarity of BLACK and BLUE and to a lesser extent BLACK and GREEN chromatin is reflected in considerable, but mutually exclusive overlap of PcG proteins and the heterochromatin protein group with the BLACK chromatin components.

The five chromatin types defined in chapter 4 are based on recurrent, typical combinations of proteins binding to genomic regions. Genomic fragments with the same type have some common chromatin components as well as optional ones that typically co-occur.

These are groups of proteins that either promote or at least do not interfere with each other’s binding and probably share regulatory or structural functions. Due to the need to express different sets of genes, the location of chromatin type domains are expected to differ between tissues and cell types. Centromere-proximal GREEN domains may be more constant at least in dividing cells because the HP1 chromatin type is known to be important for genome stability (Peters et al. 2001; Peng and Karpen 2009).

Even in different cell types, it is unlikely that new chromatin types arise by completely different combinations of proteins, as this would require many proteins to strongly change their preference of binding partners. Instead, new types may result from subsetting or merging of the five types defined in the current study.

Colocalization of many proteins in RED chromatin

The chromatin type characterized by the largest combination of proteins is RED chromatin, which has up to 40 of the tested 53 proteins bound at some locations (chapter 4, Supplementary Fig. S3B). How can dozens of proteins bind to the same site?

First, the resolution of DamID with microarray readouts in the current implementation is probably about 1-2 kb (data not shown), so signal overlap does not require physical contact of proteins.

Second, it should be kept in mind that techniques like DamID or chromatin immunoprecipitation give end-point readouts averaged over a cell population. In extreme cases, overlapping signals could derive from mutually exclusive binding in different cells (or when using DamID, in the same cell at different time points). For instance, colocalization the histone deacetylase SIR2 and the acetyltransferase TIP60 does not necessarily indicate simultaneous binding, but could be the result of cycles of histone acetylation and deacetylation (Clayton et al. 2006).
Equally importantly, binding of chromatin proteins and TFs is very dynamic. Photobleaching and laser crosslinking experiments show that for diverse types of proteins including transcription factors, high mobility group proteins and nucleosome remodelers, residence times on a binding site is only on the scale of seconds or minutes at most, although core histones are an exception (McNally et al. 2000; Phair et al. 2004; Meshorer et al. 2006). In the light of these facts, extensive overlap of binding sites should probably in most cases not be interpreted as formation of a complex, unless protein-protein interaction is demonstrated directly.

Third, several of the proteins that accumulate in RED chromatin are TFs. Preliminary analyses indicate that the accumulation of these TFs cannot be simply explained by the distribution of their cognate motifs in the DNA (unpublished results). Rather, RED chromatin seems to facilitate the binding of many (but not all) TFs. Interestingly, it was previously found that a TF with a mutated DNA binding domain is still targeted to TF binding hotspots (Moorman et al. 2006), indicating that protein-protein interactions contribute substantially to the accumulation of TFs in RED chromatin. To some extent, interactions between factors bound sequence-specifically at distant sites by way of looping may contribute to overlapping binding signals.

**Specific gene regulation in spite of broad binding of regulatory factors**

Overlap of a large number of proteins may be taken to indicate the possibility of functional binding in the same region, e.g. the same promoter. Many protein-DNA interactions are not functionally productive, but may be rendered productive locally upon a specific signal. It has been shown that nuclear hormone receptors bind to a target promoter in the absence of hormone with short residence time, and addition of signal leads to retarded dynamics (Elbi et al. 2004). Thus, ‘binding’ of a protein to many sites may reflect probing rather than productive binding, so genes near which the same set of TFs or chromatin proteins can bind can potentially be regulated by the same pathways. This is not to say that genes in the same chromatin environment are always coregulated, because the precise arrangement of binding sites and contact with enhancers or silencers impose different regulatory logic. Rather, the chromatin type (e.g., BLACK or RED) may determine if the underlying genes are deaf to any activating signals or open to be activated in their gene-specific ways. Converting a BLACK domain into RED that allows probing by a set of regulatory proteins would thus be a mechanism of unlocking a group of genes for certain types of regulation.

In chapter 4, RED and YELLOW regions were identified as active chromatin types with different sets of proteins, suggesting that at least two such mechanisms are at work in Kc cells.

**BLACK chromatin: The default state?**

BLACK chromatin covers about half of the non-repetitive genome in *Drosophila* Kc cells. It is characterized by broad binding of histone H1, and many sites additionally acquire SuUR, D1 and other BLACK chromatin components which likely have at least some sequence preferences (Levinger 1985; Spana et al. 1988). H1, D1 and SuUR have phenotypes in polytene chromosome morphology, suggesting that BLACK chromatin is important for proper chromosome architecture, especially somatic pairing and polytene endoreplication (Belyaeva et al. 2006; Lu et al. 2009; Smith and Weiler 2010). PcG silencing and HP1 association outside of active genes are compatible with H1 and might even require it to form condensed nucleosomal structures.

How is BLACK chromatin specified? For the establishment of PcG domains and pericentric HP1 domains, PREs and repetitive sequences play crucial roles, respectively (Simon et al. 1993; de Wit et al. 2005). One possibility is that also BLACK chromatin requires sequence elements that recruit its components, or at least nucleate it at some sites from where it can spread.
There are several indications that this is unlikely to be the case. BLACK chromatin spans heterogeneous genomic regions from gene deserts in contact with the nuclear lamina, to regions with silent genes that will be activated in other tissues. Any sequence elements thus would have to be interspersed in all these different contexts.

Additionally, mechanisms would have to exist to overrule BLACK chromatin when genes are activated. The most ubiquitously binding protein in this chromatin type, histone H1, shows uninterrupted, featureless binding in BLACK regions without clear signs of nucleation sites (chapter 4, Fig. 4AB). H1 association with chromatin can occur by self-assembly without the need for ATP in vitro, and does not require specific sequence elements because it binds to nucleosomes (Meersseman et al. 1991). It is therefore conceivable that H1 associ- ates broadly with chromatin ‘by default’, i.e. in the absence of specific opposing signals. Once bound, it presumably stabilizes regular nucleosomal arrays (Pennings et al. 1994 and Fig. 2, top).

At sites where DNA has to be accessed, H1 binding has to be counteracted by either site-specific or global mechanisms. As described in chapter 1, almost all sites with low H1 occupancy in Kc cells are enriched for the replication-independently positioned histone variant H3.3, and H3.3 contributes to inhibition of H1 binding. H3.3 is incorporated at sites of histone exchange (Deal et al. 2010). These sites are active transcription start sites as well as intergenic sites that overlap with cis-regulatory elements, and are classified as RED, YELLOW or GREEN regions in chapter 4. Nucleosome remodeling has a well-documented role in overcoming H1 association for activation of cis-regulatory elements. H1-containing nucleosome arrays can be remodeled by the remodeling ATPase ACF, but not by CHD1 (Maier et al. 2008).

Glucocorticoid signaling at a responsive enhancer in rat or at the mouse mammary tumor virus results in chromatin remodeling by SWI/SNF type enzymes and H1 dissociation (Deroo and Archer 2001; Flavin et al. 2004). Interestingly, RED chromatin is devoid of H1, highly bound by the SWI/SNF homologue BRM and highly accessible as measured by FAIRE (chapter 4), suggesting a crucial role for BRM in this chromatin type.

BRM has also emerged as a central factor in the targeting of several transcription factors as well as H3.3 in the Bayesian network analysis (chapter 3).

Considering these facts, BLACK chromatin may be a default chromatin state that automatically assembles in the absence of activating signals. For activation, nucleosome stabilization by H1 needs to be overcome by ATP-dependent mechanisms. At nucleosomal sites where sequence-specific factors cooperate with remodelers for activation, nucleosomes are destabilized, H3.3 is incorporated, and H1 does not bind anymore (Fig. 2, bottom). The presence of H3.3 further contributes to antagonizing H1 binding (chapter 1). This model is in line with the observation that upon knockdown of BRM, accessibility is selectively reduced in RED regions (data not shown). It is currently not known if in this situation H1 binding is reinforced.

H1 binding can be globally regulated by competition with a number of abundant, non-sequence specific chromatin proteins, particularly high mobility group (HMG) proteins (Catez et al. 2006). In early Drosophila embryos, the levels of H1 are low, whereas the HMG protein HMG-D is abundant and competes with H1 (Ner et al. 2001). This probably represents an adaptation to the special requirements on chromatin before the mid-blas- tula transition when zygotic expression has not yet commenced.

Many multicellular organisms similarly substitute their somatic linker histones with specialized ones in germ line cells in order to support chromatin remodeling in spermatocytes and fertilized oocytes (Govin et al. 2004; Becker et al. 2005; Jedrusik and Schulze 2007). Binding of H1 appears to be regulatable by phosphorylation:

Some H1 isoforms become hyperphosphorylated at specific residues in M phase and remain bound to mitotic chromosomes, whereas phosphorylation at other residues correlates with gene activity (Talasz et al. 2009; Zheng et al. 2010). Thus, specific phosphorylation marks may mediate either reinforced or diminished binding.
The predicted network of targeting interactions (chapter 3, Fig. 2) is in reasonably good agreement with the groups of proteins most strongly overlapping with the five chromatin types outlined in chapter 4 in spite of the much lower resolution and lack of information in intergenic sequences (chapter 4, Fig. 3). Nevertheless, the good rate of confirmation of the tested predictions suggests that the approach is useful for generating hypotheses.

The targeting interaction network may be further refined by applying Bayesian Network Inference (BNI) to the high resolution dataset, and by inclusion of antagonistic interactions (exemplified by the inhibitory effect of H3.3 on H1 binding) into the network modeling approach.

A central question remains to be tested: How are chromatin types established? The composition of the five chromatin types suggests that the BLUE type is based upon the BLACK type, and that active genes in GREEN chromatin are more similar to YELLOW than RED chromatin.

**Figure 2**

Proposed model how ‘default’ BLACK chromatin is converted into red chromatin during transcriptional activation. Upper panel, self-assembly of histone H1 containing structures results in formation of BLACK chromatin in the absence of activating signals. Transcription factor (TF) access is prevented. Lower panel, upon activating signal, nucleosome remodeling ATPases like BRM destabilize nucleosomes (blurred nucleosomes). Remodelers may be targeted by the activating signal or are present constitutively, resulting in constitutively accessible sites. Remodeling and incorporation of H3.3 prevent H1 binding. TFs and cofactors (X) can access all remodeled regions, but exchange rapidly and only lead to transcriptional activation at sites where all required factors colocalize and cooperate productively.
These similarities may reflect in part positive targeting interactions between many proteins in similar types, and antagonistic interactions between highly dissimilar types.

This could provide an explanation for the striking preferences of adjacency between the types (chapter 4, Fig. 7). By targeting core factors of one chromatin type to sites within another chromatin type, it can be tested whether its protein-protein interactions with its fellow type members are sufficient to bring about a conversion of chromatin type. Good candidate proteins for this approach may be obtained by BNI.

Genes in similar chromatin environments are coregulated. Very often (this thesis is no exception), the importance of chromatin-mediated gene regulation is emphasized, but it is not known how much of the variance in gene regulation is caused by chromatin as opposed to the activity of sequence-specific TFs.

Promising approaches to address this question include random insertions of reporter genes in different chromatin environments using transposons. If individual transposons are tagged with unique sequence 'barcodes', it is possible to correlate the chromatin state with the expression state of individual insertions using high-throughput sequencing as a readout for both. Such systems can be developed further to mutate or insert TF binding sites to gain information on whether TF recruitment is able to influence chromatin type.

Of particular interest will be the role of chromatin remodeling factors. It has been demonstrated that in glucocorticoid signaling in the mouse, glucocorticoid receptor can induce accessibility of some of its binding sites, while others are accessible even without hormone (John et al. 2008). SWI/SNF-type remodelers have been reported both to mediate TF binding during signaling through the JAK-STAT pathway (Ni et al. 2005) as well as to be recruited to target genes themselves upon stress signaling via the MAPK pathway (Proft and Struhl 2004).

As most studies along these lines highlight chromatin changes at only a few genes, it remains to be clarified if various remodeling complexes differ generally in their ability to serve as pioneer factors that reside constitutively at signaling target genes, or whether that differs per pathway.

Finally, it should increasingly be addressed if the behavior of populations of cells in terms of gene expression reflects that of single cells. In fact, evidence is accumulating that transcription occurs stochastically and in bursts such that a given gene in individual cells can undergo active transcription or be silent at a given time point (Raj and van Oudenaarden 2008).

This is because the molecular events leading to activation of a particular gene are not deterministic, but involve stochastic interactions of various factors. The same may be true for other important cellular events, such as modulation of nuclear architecture, and even cell fate decisions. It will therefore be of increasing importance to develop tools to monitor chromatin states in single cells.
References


An evaluation of 3C-based methods to capture DNA interactions *Nat Methods* **4**: 895-901.

Drosophila D1 overexpression induces ectopic pairing of polytene chromosomes and is deleterious to development. *Chromosoma* **119**: 287-309.


Site-specifically phosphorylated forms of H1.5 and H1.2 localized at distinct regions of the nucleus are related to different processes during the cell cycle. *Chromosoma* **118**: 693-709.

