Biochimica et Biophysica Acta xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Surviving an identity crisis: A revised view of chromatin insulators in the genomics era☆

Leah H. Matzat, Elissa P. Lei *

Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Received 30 June 2013
Accepted 10 October 2013
Available online xxxx

Keywords:
Chromatin
Insulator
Nuclear organization
Drosophila

ABSTRACT

The control of complex, developmentally regulated loci and partitioning of the genome into active and silent domains is in part accomplished through the activity of DNA-protein complexes termed chromatin insulators. Together, the multiple, well-studied classes of insulators in *Drosophila melanogaster* appear to be generally functionally conserved. In this review, we discuss recent genomic-scale experiments and attempt to reconcile these newer findings in the context of previously defined insulator characteristics based on classical genetic analyses and transgenic approaches. Finally, we discuss the emerging understanding of mechanisms of chromatin insulator regulation. This article is part of a Special Issue entitled: Chromatin and epigenetic regulation of animal development.

Published by Elsevier B.V.

1. Introduction

A central question in biology centers on the ability for a single genome to produce a diverse array of transcriptional outputs, creating unique cell types, and ultimately, morphologically complex multicellular organisms. This developmental feat is achieved through spatially and temporally specified gene expression, which requires tremendous regulation. Studies of genome regulation and organization have long recognized the propensity for eukaryotic chromosomes to reside in distinct sub-nuclear territories, a feature conserved from *Saccharomyces cerevisiae* to humans (reviewed in [1]). Recently developed molecular techniques have allowed higher-resolution mapping of chromosomal domains, which confirmed a long-held hypothesis that units smaller than a single chromosome are non-randomly organized into functional domains. While the mechanisms underlying three-dimensional genome organization are not yet well understood, a key role for chromatin insulator proteins has emerged in defining chromatin domains both in a local chromosome environment as well as in long-range chromosomal interactions.

Chromatin insulator sequences, or boundary elements, were initially defined genetically as DNA elements that possess two key properties indicative of the capacity to define a chromatin domain. The first is termed enhancer blocking, the ability to interfere with enhancer–promoter communication only when placed between the two elements. The second feature is termed barrier activity, the ability to protect a flanked transgene from position-dependent silencing. For many years, insulator sequences, along with the specific effector proteins associated with these sequences, were predominantly studied at only a few model loci or within artificial contexts. These technical limitations permitted only a restricted view, leading to a certain set of predictions about where insulator complexes would be located throughout the genome as well as their functions within these contexts. With the advent of whole genome chromatin immunoprecipitation (ChIP) and chromosome conformation capture (3C) approaches, in addition to application of genome-wide transcriptome analyses, some of these predictions have been realized while others require re-evaluation. This review will examine the mechanisms and regulation of the main classes of chromatin insulator complexes present in *Drosophila* and attempt to reconcile their classically defined functional properties considering examples from other organisms as well as new insights from recent genome-wide studies.

Abbreviations: 3C, chromosome conformation capture; *Abd-B*, Abdominal-B; AGO2, Argonaute 2; ANT-C, Antennapedia complex; BEAF-32, boundary element-associated factor of 32 kDa; BTB, broad-complex, tramtrack, bric-a-brac; BX-C, bithorax complex; ChIP, chromatin immunoprecipitation; CNS, central nervous system; CP190, Centrosomal protein 190; CTCF, CCCTC-binding factor; DamID, DNA adenine methyltransferase identification; DREF, DNA replication-related element-binding factor; Elba, early boundary activity; E(y)2, enhancer of yellow 2; *fab*, frontabdominal; GAF, GAGA factor; *iab*, infra-abdominal; L(3)mbt, lethal (3) malignant brain tumor; LAD, Lamin-associated domain; MAR, matrix attachment region; Mcp, Miscadastral pigmentation; Mod(mdg4)2.2, Modifier of mdg4 2.2 isoform; ncRNA, noncoding RNA; ORC, origin recognition complex; piRNA, Piwi-interacting RNA; PcG, Polycomb Group; PRE, Polycomb Group Response Element; PTS, promoter targeting sequence; RNAi, RNA interference; Shep, Alan Shepard; *scs*, specialized chromatin structure; SRA, steroid receptor RNA activator; SUMO, small ubiquitin-like modifier; Su(Hw), Suppressor of Hairy wing; Topo II, topoisomerase II; TrxG, trithorax group; TSS, transcription start site; UTR, untranslated region; Zw5, Zeste-white 5.

☆ This article is part of a Special Issue entitled: Chromatin and epigenetic regulation of animal development.

* Corresponding author at: 9000 Rockville Pike, Bethesda, MD 20892, USA. Tel.: +1 301 435 8989; fax: +1 301 496 5239.
E-mail address: leielissa@niddk.nih.gov (E.P. Lei).

1874-9399/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.bbagrm.2013.10.007

2. Core components and mechanisms of chromatin insulator activity

2.1. Conservation of chromatin insulators between Drosophila and vertebrates

In *Drosophila*, multiple classes of insulator complexes have been identified, and these can be classified based on the specific DNA-binding component of the complex. Each category of insulator complexes displays enhancer blocking and/or barrier activities and also contains the common Centrosomal protein 190 (CP190). Only the CTCF insulator protein is known to be conserved in vertebrates, and CTCF is the only vertebrate insulator protein that has been identified thus far. Despite weak or no homology outside of the DNA-binding region, many functions of vertebrate CTCF mirror either that of its *Drosophila* counterpart or of other insulator proteins. One notable exception is the interaction of vertebrate CTCF with cohesin during interphase [2] (reviewed in this issue [Ball, Chen, and Yokomori]); this functional partnership does not exist in *Drosophila*, in which cohesin influences gene expression during interphase in an insulator-independent manner [3]. An important common feature of all insulator proteins is the capacity to mediate long-range interactions between distant genomic sites, which likely requires attachment to some type of scaffold within the nucleus.

2.2. Interpreting ChIP data with caution

Many of the studies discussed here analyze occupancy information derived from ChIP-chip or ChIP-seq experiments often performed in different laboratories. Methods used by different groups vary somewhat with respect to the particular antibodies and other technical parameters pertaining to ChIP, such as the extent of shearing. Earlier studies used microarray technology, and a variety of platforms were used in this phase. Later studies utilize high throughput sequencing, which yields higher resolution and greater dynamic range of signal, but computational methods for peak calling also vary widely across laboratories despite using the same sequencing platform. Another experimental consideration is that using formaldehyde crosslinking preserves both direct and indirect interactions, including looping-dependent interactions. Therefore, based on this method it is not possible to definitively determine whether a given ChIP signal corresponds to stable protein association with the DNA, either directly or as part of a protein complex. In one study, peak height was used as a main factor to discriminate true binding sites [4], while another used the presence of preferred binding motifs [5]. Although it is reasonable to classify subsets of ChIP peaks based on some criteria, either of these approaches could be argued to be somewhat arbitrary, and careful consideration should be applied when interpreting these results. Despite these caveats, which are not easily addressable, ChIP studies remain valuable tools for investigating genome-wide patterns.

2.2.1. The gypsy insulator

2.2.1.1. Nuclear organization and partitioning of the genome by the gypsy insulator. The *gypsy* insulator is the best defined of the three known classes of *Drosophila* insulator complexes. Its sequence specificity is dependent on the 12 zinc-finger DNA binding protein Suppressor of Hairy wing (Su(Hw)), which was first identified as binding an AT-rich 26 bp sequence element repeated twelve times in the 5′ UTR region of the *gypsy* retrotransposon [6,7]. Naturally occurring or endogenous Su(Hw) binding sites similar to those in the *gypsy* element are present as a single binding site or clusters of 2–6 repeats with variable spacing between them [8,9]. Su(Hw) is required for both enhancer blocking and barrier activity at *gypsy* as well as the handful of tested endogenous genomic binding sites [6–14]. Bound directly to Su(Hw) are the Modifier of mdg4 2.2 isoform (Mod(mdg4) 2.2) and CP190, which together form a tripartite complex that makes up the 'core' *gypsy* insulator complex required for *gypsy* enhancer blocking activity [10,15–19]. While neither are known to interact directly with DNA *in vivo*, both Mod(mdg4)2.2 and CP190 contain broad complex, tramtrack, bric-a-brac (BTB) dimerization domains that can interact with each other and possibly promote multimerization of insulator complexes [19–21]. This capacity for inter-complex interaction is consistent with the revealing discovery of the insulator bypass phenomenon, in which two tandem copies of the *gypsy* insulator cancel one another in an enhancer blocking assay, presumably by pairing and looping out of the intervening DNA [22,23]. Together with the finding that the AT-rich binding sites for Su(Hw) resemble nuclear matrix attachment regions (MARs) [8,24], these observations led to the hypothesis that *gypsy* insulator complexes could act as scaffolds to organize chromatin into higher order domains.

Although distributed throughout the nucleoplasm, *gypsy* insulator proteins coalesce at a small number of higher intensity foci in diploid nuclei, termed insulator bodies, which are found at both the nuclear periphery as well as the nuclear interior. These structures have been shown to colocalize with *gypsy* insulator sequences *in vivo* [20] and also tether to the nuclear matrix [25]. Disruption of Lamin (*Drosophila* Lamin B) results in diffusion and mislocalization of insulator bodies as well as loss of *gypsy* enhancer blocking activity [26]. However, association of insulator sites with the nuclear periphery is not a requisite for enhancer blocking activity [27]. Interaction between the *gypsy* insulator complex and Lamin appears to be mediated by the dTopors protein, which interacts directly with Mod(mdg4)2.2 (discussed in more detail later). Furthermore, a large number of mutations in *gypsy* insulator core components and regulatory factors that disrupt insulator body localization also display defects in *gypsy* enhancer blocking activity, indicating a tight correlation between the proper formation of these structures and robust insulator activity [19,20,26,28–38]. Nevertheless, proper localization of insulator bodies is not sufficient for *gypsy* insulator activity, as certain mutations in Mod(mdg4)2.2 disrupt enhancer blocking activity but do not perturb insulator body localization [34,39]. Recent work proposes the non-exclusive concept that insulator bodies could serve as storage sites for insulator proteins, protecting them from degradation or aiding in maturation of complexes [30]. Continued mechanistic studies as well as ultrastructural analyses of insulator bodies will provide additional insight into these intriguing nuclear structures.

Genome-wide profiling of Su(Hw) binding sites met many expectations for a protein believed to be involved in demarcation of transcriptional domains. Early cytological studies on highly replicated salivary gland polytene chromosomes showed that Su(Hw) and Mod(mdg4) 2.2 colocalize extensively at DAPI band–interband boundaries, indicating enrichment at borders between condensed and decondensed chromatin [17,29]. Similar to vertebrate CTCF [40,41], the majority of Su(Hw) binding sites correspond to intergenic regions and introns, while a smaller fraction of Su(Hw) sites are found at transcription start sites (TSSs) and within exonic sequences [31,42–44]. Consistent with interaction with Lamin and MARs, Su(Hw) binding occurs frequently at the borders of, as well as within, Lamin associated domains (LADs) [45], which tend to be gene poor and correspond to regions of low gene expression [46]. Furthermore, Su(Hw) binding is not correlated with any particular histone modifications profiled thus far; it is mainly associated with 'undefined' or 'black' chromatin in comprehensive genome-wide analyses of chromatin states [47,48]. Finally, Su(Hw) occupancy appears to be fairly static across various tissues and cell types [42,44], suggesting that its ability to interact with DNA is not widely regulated and also that Su(Hw) itself does not drive cell-type specific differences in gene expression.

2.2.1.2. Searching for endogenous functions of the gypsy insulator. Depletion and mutant studies have, unfortunately, yielded limited insight into endogenous Su(Hw) function. Supposing that the basic function

Please cite this article as: L.H. Matzat, E.P. Lei, Surviving an identity crisis: A revised view of chromatin insulators in the genomics era, Biochim. Biophys. Acta (2013), http://dx.doi.org/10.1016/j.bbagrm.2013.10.007