Feature Review

CellPress

New tools in the box: An evolutionary synopsis of chromatin insulators

Peter Heger and Thomas Wiehe

Cologne Biocenter, Institute for Genetics, University of Cologne, Zülpicher Straße 47a, 50674 Köln, Germany

Despite progress in understanding genome organization and gene expression during the last decade, the evolutionary pathways that led to the intricate patterns of gene expression in different cells of an organism are still poorly understood. Important steps in this regulation take place at the level of chromatin, where the (epi)genomic environment of a gene determines its expression in time and space. Although the basic mechanisms of gene expression apply to all eukaryotes, multicellular organisms face the additional challenge of coordinating gene expression during development. In this review we summarize and put into evolutionary context current knowledge about chromatin insulators, an important class of regulatory factors mediating these tasks. Our interpretation of historical and recent findings points to a dynamic and ongoing evolution of insulator proteins characterized by multiple instances of convergent evolution, gene loss, and binding site changes in different organisms. The idea of two autonomously evolving insulator functions (as a barrier element and an enhancer blocker) further suggests that the evolution of metazoans and their enhancer-rich gene regulatory repertoire might be connected to the radiation of enhancer blocking insulators. Although speculative at the moment, such coevolution might create tools for complex gene regulation and therefore influence the evolutionary roadmaps of metazoans.

What are insulators?

Observations that led to the description of the first insulator protein (see Glossary) stretch back to 1932, when Calvin Bridges reported the first suppressor mutants in *Drosophila melanogaster* (Table 1) [1]. One of these mutations was at the *suppressor of Hairy-wing* [su(Hw)] locus and could revert the phenotype of certain spontaneous mutations at other loci. The spontaneous origin and random distribution of these mutations suggested the action of a mobile genetic element. It could not be shown before the late 1980s that a retroelement, *gypsy*, induced those mutations, that a specific factor, su(Hw), bound to the retroelement, and that this factor was able to inhibit enhancers upstream of the insertion site, thereby causing the mutant phenotype [2,3]. These experiments established a new class of regulatory elements with the ability to block enhancer action in a directional way, that is, only when placed between an enhancer and a promoter (Figure 1). At the same time it became clear from work with transgenic mice that vertebrates also possess regulatory sequences that insulate a transgene from the position effect of surrounding chromatin and that CTCF (CCCTC-binding factor), a highly versatile zinc finger (ZF) protein, is responsible for this effect [4–6]. Together with studies on position-effect variegation in *D. melanogaster*, these findings led to the conclusion that chromatin is divided into structurally and functionally discrete domains whose borders are maintained by specific boundary sequences. The identification of boundary elements and corresponding binding partners in yeast showed that a partition into active and repressed chromatin domains also applies to unicellular eukaryotes [7,8] and therefore is likely a universal eukaryotic feature.

These few examples illustrate the main characteristics of an insulator. (i) It typically consists of a short piece of genomic DNA, the insulator or insulator sequence, that is bound by a defined factor, the insulator protein, and this combination is necessary and sufficient for insulator activity, although sometimes in connection with cofactors. (ii) All insulators act as either barrier elements, maintaining the border between active and repressive chromatin environments, or enhancer blockers, delimiting the influence of positive regulatory elements (i.e., enhancers) in an orientation-dependent manner, or have both functions (Figure 1). (iii) Insulator function can be identified only experimentally. Assays typically require the time-consuming integration of a reporter construct into the chromatin of a cell line or transgenic animal. Importantly, these procedures often test insulator function in an artificial context that does not necessarily reflect *in vivo* function. (iv) Although a subset of insulator elements might be predictable on the basis of chromatin features, such as histone modifications [9], it is generally not possible to predict the factors they recruit or to find these factors intentionally. As a consequence, only a few insulator proteins have been verified experimentally. (v) Many of the known insulator proteins consist of multiple functional domains and interact with different partners. For some it is well established that they have additional functions beyond chromatin insulation, such as transcription factors (e.g., CTCF, TFIIIC). Insulator function may therefore reflect only one particular aspect of the repertoire of such a protein or the outcome of a multiprotein complex.

Corresponding author: Heger, P. (peter.heger@uni-koeln.de).

Keywords: genome organization; barrier element; enhancer blocker; DNase hypersensitive site; histone modification; coevolution.

0168-9525/

 http://dx.doi.org/10.1016/j.tig.2014.03.004

Glossary

CFTR: cystic fibrosis transmembrane conductance regulator, a 1.8-Mb region encompassing the CFTR locus, chosen for annotation of functional elements by the ENCODE consortium.

Chromatin boundary: Gr. χρῶμα = color, dye. Chromatin is the stainable content of a cell nucleus and is composed of DNA and proteins. A chromatin boundary is a transition between two distinct chromatin states defined by histone modifications, DNA methylation, or chromatin accessibility. It typically refers to the transition between a heterochromatin domain and an active gene locus.

Chromatin conformation capture (3C): restriction- and ligation-based assay for detection of long-range interactions. The 3C approach is used to analyze contacts between two defined regions, whereas 4C (circular 3C) is used to detect genome-wide contacts for a region of interest.

Chromatin remodeling: dynamic modification of nucleosome structure, composition, and positioning to facilitate gene expression. Histone-modifying enzymes and energy-dependent ATPases work together to move and restructure nucleosomes.

DNase I hypersensitive (DH) sites: short regions of genomic DNA (hundreds of base pairs) especially susceptible to digestion by nuclease. DH sites typically contain gene regulatory elements and correspond to 'open chromatin'. A subset of DH sites is bound by insulator proteins.

Enhancer: positively acting regulatory element that enhances transcription of a gene. Enhancers work position-independently and over large distances and are often responsible for cell-type-specific gene expression.

Fab-8: frontabdominal-8, one of several chromatin boundaries within the *D. melanogaster* bithorax Hox gene complex. Belongs to the large regulatory region of the Hox gene *Abd-B* and is necessary for specification of abdominal segments.

Heterochromatin: gr. ἕτερος = other. In contrast to actively transcribed euchromatin, this 'other' form of chromatin is transcriptionally silent, tightly packed, and enriched in repressive chromatin marks. Constitutive heterochromatin often has structural functions (e.g., at centromeres and telomeres), whereas facultative heterochromatin can lose its condensed status under the influence of developmental cues and become transcriptionally active.

Histone modifications: Histones, the constituents of nucleosomes, undergo a large number of post-translational modifications that are important for regulating gene expression and defining active and inactive chromatin compartments. Certain modifications, such as dimethylation on histone H3 lysine 4 (H3K4me2) and trimethylation of H3 lysine 36 (H3K36me3), are indicative of active chromatin, whereas others, such as H3K9me2 and H3K27me1, specify repressive chromatin.

iab-7, iab-8: infra-abdominal-7 and -8, enhancer regions within the *D. melanogaster* bithorax Hox complex that regulate Abd-B expression in a segment-specific manner.

Insulator protein: sequence-specific DNA-binding protein with the ability to interfere with enhancer–promoter communication in an orientation-dependent manner (enhancer blocker) and/or the ability to block the spreading of heterochromatin (barrier element) in functional assays.

Interband: less chromatin-rich regions between highly condensed, chromatin rich bands in polytene interphase chromosomes. The patterning into chromosomal bands and interbands can be distinguished by light microscopy (without staining).

Opisthokonta: gr. ὀπίσθιος = rear, posterior; κοντός = pole, rod. A monophyletic group of eukaryotes including choanoflagellates, animals, and fungi. Multicellular animals evolved in this group of eukaryotes whereas plants (land plants and green algae) originated in a different group.

Position effect: active and repressive chromatin can modulate the expression of randomly integrated transgenes. The level of transgene expression is thus dependent on the genomic environment, or position, of the integrated DNA and can vary accordingly (position-effect variegation). Often, a transgene is silenced by flanking repressive chromatin.

Puff boundary: to increase their transcription rate, some cells undergo endoreduplication (e.g., insect salivary gland cells) and produce a chromosome with highly amplified DNA content (polytene chromosome). At sites of active transcription, a microscopically visible, diffuse uncoiled region can form, known as the chromosomal puff or Balbiani ring.

TFIIIC: transcription factor IIIC, a heteromeric protein complex consisting of six subunits in yeast. Responsible for the transcription of tRNA and ribosomal RNA genes. Forms a complex of ~ 1.5-MDa with TFIIIB and RNA polymerase III (26 polypeptides) that is competent for transcription.

Upstream activation site (UAS): *cis*-regulatory sequence in yeast that increases the expression of a nearby gene. A UAS is often orientation- and distance-independent and therefore considered functionally analogous to metazoan enhancers [98], but is usually much closer to the transcription start site.

Recent years have seen increasing interest in insulators and their role in nuclear architecture and chromosome organization [10]. In many investigations, the authors centered on particular model organisms or molecular features without taking an evolutionary point of view. A first step towards understanding the origin of insulators and insulator proteins was the idea that they evolved as derivatives of promoters and transcription factors, a notion based on several conceptual and mechanistic similarities [11,12]. Extending the discussion, we focus here on evolutionary aspects of insulator proteins and the insights that can be gained from this perspective.

A catalogue of insulator proteins

Although examples of gene insulation have been reported from yeast to humans, the factors mediating this function differ substantially between organisms (Table 2).

Insulator proteins in D. melanogaster

Experimentally validated insulator proteins are restricted to a few model species in which functional assays are possible. The most abundant set of insulator proteins, eight at the time of writing, is known from *D. melanogaster* (Figure 2, Table 2). The discovery of chromatin boundaries at the 87A7 heat shock locus [13] and the observation of position effects in flies in the first half of the 20th century [14] have revealed that proteins involved in these phenomena can have insulator function. The first in this series was su(Hw), detected through its link to the transposable element *gypsy* [3]. Zeste-white 5 (Zw5) was originally identified in mutagenesis experiments as one of 12 complementation groups between the *zeste* and *white* loci on the *Drosophila* X chromosome [15]. It is synonymous with the earlier described mutation *deformed wing (dwg)* [16] and later studies showed that it conveys barrier activity of a specialized chromatin structure, scs, at one side of the *Drosophila* 87A7 heat shock locus [17]. In a search for factors binding to scs', the other boundary element of the locus, a novel boundary element-associated factor of 32 kDa (BEAF-32A) was identified that localizes to this region and many other interbands and puff boundaries [18]. Modifier of mdg4 [Mod(mdg4)] was discovered in a screen for mutations interfering with a *gypsy* induced phenotype [19]. It directly interacts with su(Hw) and contributes to the function of this insulator [20]. The *Drosophila* GAGA factor was initially described as a transcriptional activator of the homeotic gene *ultrabithorax* [21]. The GAGA factor caused *in vitro* disruption of nucleosome structure and opening of chromatin, which was at that time an important indication of how the repressive effect of nucleosomes on transcription could be overcome [22]. At the same time, mutations in the gene *Trithorax-like (Trl)* were known to act as dominant enhancers of position-effect variegation and to influence Hox gene expression [23]. The discovery that *Trl* encodes the *Drosophila* GAGA factor suggested that its chromatin remodeling function plays an important role in the expression of developmental genes [23]. However, the insulator function of the GAGA factor was discovered only later during analysis of the complex regulation of the *D. melanogaster even-skipped* gene [24]. Another insulator protein, CP190 (centrosomal protein of 190 kDa), was initially isolated as an integral part of the centrosome in an approach to identify centrosomal components [25]. Depending on the cell cycle, CP190 oscillates between centrosome and nuclear compartments, where it is associated with band/interband boundaries at a large