Seminars in Cell & Developmental Biology 57 (2016) 24–30

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

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

Towards a predictive model of chromatin 3D organization

Chenhuan Xu, Victor G. Corces*

Department of Biology, Emory University, 1510 Clifton Road NE, Atlanta, GA 30322, USA

ARTICLE INFO

Article history:
Received 13 September 2015
Received in revised form
24 November 2015
Accepted 26 November 2015
Available online 3 December 2015

Keywords:
Transcription
Nucleus
Architectural proteins
CTCF

ABSTRACT

Architectural proteins mediate interactions between distant regions in the genome to bring together different regulatory elements while establishing a specific three-dimensional organization of the genetic material. Depletion of specific architectural proteins leads to miss regulation of gene expression and alterations in nuclear organization. The specificity of interactions mediated by architectural proteins depends on the nature, number, and orientation of their binding site at individual genomic locations. Knowledge of the mechanisms and rules governing interactions among architectural proteins may provide a code to predict the 3D organization of the genome.

Contents

1. Introduction

The linear eukaryotic genome resides in the three dimensional nucleus in an organized manner [1]. Certain genomic regions are highly self-interactive, whereas interactions between these regions are much less frequent, thus organizing the genome into local chromatin interaction domains named topologically associating domains (TADs) [2–4]. Multiple lines of evidence suggest that this higher-order chromatin organization is linked to genome function. TADs contain genes with coordinated expression [3], they overlap with DNA replication timing domains [2,5], evolutionarily rearranged domains [6,7], and oncogenic translocation-induced hyperacetylation domains [8], suggesting that TADs are, at least in part, a physical representation of the functional partitioning of the genome.

Architectural proteins are enriched at TAD borders and at regulatory elements interacting with each other within TADs [2–4,9]. Binding sites for these proteins cluster at specific regions of the genome termed architectural proteins binding sites (APBSs) where they mediate chromatin interactions independent of their enhancer-blocking insulator function [10]. Here we first review the nature and genomic distribution of architectural proteins characterized in *Drosophila* and vertebrates. We then discuss results showing that loss of architectural protein function causes changes in chromatin interactions and alterations in transcription. Finally, we describe mechanistic models that aim to predict nuclear 3D organization from the linear information specified by the number, nature and binding site orientation of architectural proteins present at distinct sites in the genome.

* Corresponding author. Tel.: +1 404 727 4250; fax: +1 404 727 2880.
E-mail address: vcorces@emory.edu (V.G. Corces).

http://dx.doi.org/10.1016/j.semcdb.2015.11.013
1084-9521/

2. Architectural proteins

Contrary to vertebrates, in which CCCTC-binding Factor (CTCF) has been the main DNA-binding architectural protein studied in detail thus far, several DNA binding architectural proteins have been well characterized in *Drosophila*. These include CTCF, Boundary Element Associated Factor 32 (BEAF-32), Suppressor of Hairy-wing [Su(Hw)], Transcription Factor IIIC (TFIIIC), Z4 (also called Putzig), Insulator binding factor 1 and 2 (Ibf1 and Ibf2), Pita, and Zn-finger Protein Interacting with CP190 (ZIPIC) [10–12]. These proteins bind to specific sites in the genome and recruit a series of associated factors, including Centrosomal Protein 190 (CP190), Modifier of mdg4 [Mod(mdg4)], Rad21, CAP-H2, L(3)mbt, Fs(1)h-L, and Chromator (also called Chriz). Although each DNA binding protein has a preference to interact with a specific subgroup of accessory proteins, this preference is not strict, and it is possible to find combinations of any of the architectural proteins described above at some genomic location (Fig. 1). For example, TFIIIC, which is the main architectural protein found in yeast, is also present in *Drosophila* at tRNA genes together with Rad21 (a subunit of the cohesin complex) and CAP-H2 (a subunit of the condensin II complex) but it is also found at extra TFIIIC (ETC) sites with both DNA-binding and accessory architectural proteins, including CTCF, BEAF-32, Su(Hw), CP190, and Mod(mdg4) [10]. BEAF-32 and Z4 colocalize at many promoter-proximal sites in the genome together with Chromator and CP190, whereas Su(Hw) colocalizes preferentially with CP190 and the Mod(mdg4)2.2 isoform. These genomic sites containing individual DNA-binding architectural proteins and several associated factors are called APBSs. Importantly, all or most DNA-binding and associated accessory architectural proteins colocalize in different numbers and combinations at distinct sites called high occupancy APBSs, which are preferentially located at the borders between TADs [10] (Fig. 1). Additional candidate architectural proteins that possess canonical insulator function, including Early boundary activity (Elba), have been discovered in *Drosophila* but their genomic localization in the context of the ones described above has not been explored in detail [13].

CTCF and cohesin are the two main architectural proteins extensively characterized in vertebrates [14–17]. However, several other proteins have been shown to colocalize with CTCF at many genomic locations in mammals and to play a role in specific aspects of CTCF function, but their possible role in the establishment of 3D organization has not been explored in detail (Fig. 1). For example, Yin Yang 1 (YY1) functions with CTCF during X-chromosome inactivation and both proteins colocalize extensively at evolutionarily conserved CTCF sites located preferentially at promoter-proximal regions [18]. YY1 interacts with cohesin and condensin, and has been shown to contribute to the 3D organization of the *Igh* locus [19,20]. Furthermore, YY1 is enriched with CTCF at TAD borders [21]. As is the case in *Drosophila*, TFIIIC colocalizes with CTCF, cohesin, and the DNA-binding tumor suppressor protein Prdm5 at many locations of the mammalian genome [22–24]. The POZ-Zn finger transcription factor Kaiso interacts with CTCF but its genome-wide distribution or possible role in 3D organization has not been explored in detail [25]. Additional DNA binding proteins that colocalize extensively with CTCF include JunD, the Myc-associated zinc finger protein MAZ, and ZNF143 [26]. Finally, the nucleolar protein Nucleophosmin is required to recruit CTCF to the nucleolus in order to tether CTCF-mediated chromatin loops [27] (Fig. 1).

It is striking that most DNA-binding architectural proteins characterized so far in both *Drosophila* and mammals are zinc finger proteins but it is unclear whether this conservation reflects a requirement for specific aspects of architectural protein function. In the case of CTCF, it has been shown that different combinations of zinc fingers can recognize different sequence motifs, possibly exposing other zinc fingers for protein–protein interaction. This may confer greater specificity to both its DNA- and protein-interacting capacity, while preserving the flexibility to relocate and mediate new chromatin interactions when a cell changes its fate [28,29]. The degenerate consensus motif sequence of CTCF shows variable base content at many positions [30]. Indeed, three different types of CTCF motif sequences have been shown to be present at distinct genomic locations with respect to regulatory elements, different epigenetic features, and frequency of TSS-distal element interactions [31].

3. Architectural proteins mediate interactions between distant sequences

Hi–C and Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) experiments conducted in *Drosophila* and in multiple mouse and human cell lines have shown enrichment of CTCF and cohesin at TAD borders [2,3,9], and at anchor regions of chromatin interaction within TADs [32,33]. Contacts between cohesin-occupied anchor regions have significantly higher interaction frequency than those with cohesin at only one anchor or not mediated by cohesin [33], suggesting that cohesin facilitates the establishment or maintenance of contacts between the regulatory elements it occupies. However, multiple lines of evidence suggest CTCF and cohesin are not the only two architectural proteins involved in mediating contacts between distant sequences in mammals. In the human B-lymphoblastoid cell line GM12878, only 30% (2857 out of 9448) of all interactions are mediated by CTCF sites present at each of the anchor regions, whereas 54% (6991 out of 12,903) of all chromatin interactions have CTCF at only one of the anchors [32]. Similarly, only 41% of 14,701 RAD21-mediated interactions and 35% of 22,559 contacts mediated by the Smc1a cohesin complex subunit occur between two CTCF-binding anchors in the human leukemia cell line K562 and in mouse ESCs, respectively [33,34]. These results suggest that many chromatin interactions in mammals could be mediated by other architectural proteins or by the combination of cohesin with one or more of the possible candidate architectural proteins described above. For example, Znf143 and YY1 have been shown to be enriched at chromatin interaction anchors [21,32,33]. Although YY1 was found to extensively co-localize with CTCF in active genomic regions [18], TAD borders enriched with YY1 but not CTCF show higher enrichment for border-specific epigenomic features than borders enriched with both YY1 and CTCF [21], suggesting YY1 can function as an architectural protein independent of CTCF. Additionally, tRNA genes and SINE elements are also enriched at TAD borders [2], suggesting that TFIIIC, which binds to these sequences, may also be involved in mediating long range interactions [10,35]. Furthermore, tRNA-like Mammalian-wide Interspersed Repeat (MIR) elements were recently characterized as a new group of sequences in the human genome that possess canonical insulator activity, are close to TAD borders, and appear to be CTCF-independent, suggesting the existence of proteins that bind these sequences with an architectural function [36]. These results suggest that Znf143, YY1, and TFIIIC along with its interaction partner the condensin II complex, may function as new architectural proteins in the mammalian genome, but additional functional studies are required to verify their involvement in organizing chromatin contacts.

Although interactions that result in the formation of TADs are relatively stable during cell differentiation, contacts between architectural protein sites located within TADs are more variable [2,37,38]. It is possible that cell-type specific interactions result from the presence of architectural proteins in cell-type specific genomic locations. When the CTCF binding landscape was com-