Higher-Order Chromosomal Structures Mediate Genome Function
Ivana Jerković, Quentin Szabo, Frédéric Bantignies, Giacomo Cavalli

To cite this version:

HAL Id: hal-02990848
https://hal.archives-ouvertes.fr/hal-02990848
Submitted on 13 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Higher-Order Chromosomal Structures Mediate Genome Function

Ivana Jerković†, Quentin Szabo†, Frédéric Bantignies and Giacomo Cavalli

Institute of Human Genetics, CNRS and University of Montpellier, France

Correspondence to Frédéric Bantignies and Giacomo Cavalli: frederic.bantignies@igh.cnrs.fr, giacomo.cavalli@igh.cnrs.fr

https://doi.org/10.1016/j.jmb.2019.10.014

Edited by Romain Koszul

Abstract

How chromosomes are organized within the tridimensional space of the nucleus and how can this organization affect genome function have been long-standing questions on the path to understanding genome activity and its link to disease. In the last decade, high-throughput chromosome conformation capture techniques, such as Hi-C, have facilitated the discovery of new principles of genome folding. Chromosomes are folded in multiple high-order structures, with local contacts between enhancers and promoters, intermediate-level contacts forming Topologically Associating Domains (TADs) and higher-order chromatin structures sequestering chromatin into active and repressive compartments. However, despite the increasing evidence that genome organization can influence its function, we are still far from understanding the underlying mechanisms. Deciphering these mechanisms represents a major challenge for the future, which large, international initiatives, such as 4DN, HCA and LifeTime, aim to collaboratively tackle by using a conjunction of state-of-the-art population-based and single-cell approaches.

© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Chromatin is a complex, plastic, and multifaceted cluster of proteins, DNA, and RNA that regulate gene expression. It is built in a hierarchical manner starting with DNA forming a single base-pair up to whole genomes comprising of multibillion base pairs, stretching over ten orders of magnitude. Understanding this hierarchy requires the deployment of different but overlapping approaches, both theoretical and experimental, that are suited for the investigation at different scales. In the past decade, the application of high-throughput chromosome conformation capture techniques such as Hi-C has remarkably increased our knowledge of the chromatin fiber organization at an intermediate scale, between the multinucleosomal level and chromosomal territories. At this scale of tens to hundreds of kilobases, the genome is organized into distinct domains, called Topologically Associating Domains (TADs) [1–4], which exhibit higher preference for internal interactions compared to interactions with adjacent domains. The genome, therefore, appears as a succession of these higher-order structures, demarcated by boundaries. TADs are made of heterogeneous and extensive multipartite interactions [5–8] representing dynamic and preferential contacts [9,10] that can also be observed in individual cells as spatially defined domains [11,12]. This domain-based genome architecture is present in eukaryotes and similar domains have been reported in some prokaryote species as well [13–15], but are likely formed by different mechanisms, suggesting existence of different types of TADs. In mammals, most TADs are formed by the cohesin complex and its interaction with the insulator protein CCCTC-binding factor (CTCF) [16–19], probably driven by a dynamic “loop extrusion” process [20,21]. Notably, different chromatin states can be found within the same TAD, although these chromatin states often assemble sub-TADs of the same epigenetic content [22], which is possibly partially masked by cohesin loop extrusion activity [19,23]. In other species such as Drosophila, the TAD pattern is closely associated with partitioning of the chromatin fiber into distinct epigenetic domains [2,4,24,25]. In this case, TADs could arise from the differential folding of chromatin regions with their specific epigenetic states by a “self-assembly”
mechanism depending on the nature of chromatin marks associated with the domains [25]. Finally, other mechanisms such as transcription-induced supercoiling have been implicated and could also explain the formation of self-interacting domains [26]. Importantly, the genomic size of the TADs, from tens to hundreds of kilobases, represents a favored scale for gene regulation [27] where genes tend to be co-regulated [3,28] and in which functional enhancer-promoter contacts are privileged [29–31] (Fig. 1A). Moreover, TADs also correspond to timing-dependent replication domains, suggesting that they probably spatially segregate functional domains [32]. This role of TADs as regulatory units—mainly described in mammalian species—has been emphasized by the fact that disruption of TAD structures, for example, by disrupting TAD boundaries, can lead to neo functional landscapes associated with gene misexpression, as described in developmental diseases and cancer [33–36]. Understanding how functional contacts are controlled and established within TADs to ensure specific gene expression during development and cell differentiation has thus appeared as a crucial point. A striking example of this tight link between genome folding and function has been illustrated during mouse limb development, where Hoxd gene cluster is exposed to a dynamic switch in contacts between the two surrounding TADs defining proper and timely Hoxd expression [37].

Furthermore, although TAD borders are generally stable, TADs undergo substantial rewiring during cell fate change and reprogramming [29,38]. For example, predominantly confined within a TAD, cell-type specific interactions have been observed between promoters and their cis-regulatory elements [29]. Such enhancer-promoter contacts were also found in cell-type specific multiparticle interactions where they interact simultaneously within compartmentalized domains, suggesting that differentiation-related 3D chromatin rewiring is possibly functional [39,40]. In order to accurately assess TAD structure and function, several studies employed recently developed novel microscopy methods based on sequential imaging of chromosomal loci combined with RNA labeling, enabling both single-cell reconstruction of the chromatin fiber and the detection of transcriptional state [41,42]. These single-cell approaches revealed that the intrinsic TAD structure varies depending on its chromatin state and its transcriptional activity, highlighting the diversity of TAD folding in different cell types and their role in gene regulation. Moreover, super-resolution microscopy studies revealed dynamic and stochastic nature of TADs in single cells, with inconsistent TAD borders that preferentially locate at CTCF sites [10,11]. This dynamic TAD behavior is further corroborated by the loop-extrusion model, by the measurement of CTCF/cohesin-chromatin residence time and by CTCF/cohesin depletion studies in which TAD-scale chromatin contacts dissociate and later reappear

![Fig. 1. A. Schematic representation of TAD-mediated gene regulation. TADs represent higher-order chromosomal units that privilege functional and dynamic contacts between promoters and cis-regulatory elements. Most contact and intra-TAD changes occur concomitantly with specific transcription factor (TF) binding in a cell-type specific manner, although, in most cases, the causative relationship remains elusive. B. Schematic examples of extremely long-range interactions. Global folding of chromosomes includes homotypic long-range contacts between large chromatin regions of the same underlying epigenetic content, building A-active and B-repressive compartments. In addition, preferential cell-type specific extremely distant contacts occur both, in cis and in trans and usually involve gene(s), super-enhancers, and sometimes cell-type specific TFs. Together, these elements can assemble into functional hubs associated with transcription thereby regulating the activity of target genes. Likewise, repressive interaction networks between Polycomb group (PcG) proteins mediate long-range interactions between distal genes favoring transcriptional repression.](image-url)
upon reintroduction of the corresponding proteins [16–21,43,44]. However, it is still not clear with what kinetics do the TADs fold and dissipate in single cells and how does this affect gene regulation. Future studies integrating microscopy, biochemical, genetic, and computer modeling experiments are necessary in order to fully understand TAD dynamics in single cells and the role of TADs in the spatial regulation of functional contacts.

Above TAD organization, a higher-order chromatin folding also follows specific but distinct rules defining the overall chromosome conformation. In such fashion, chromosomes tend to segregate into regions of preferential long-range interactions (compartments) based on their underlying epigenetic content. This segregation occurs in a homotypic fashion, where active and inactive (repressive) chromatin intervals interact over tens to hundreds of megabases (Mb) building “A” and “B” compartments, respectively. Furthermore, both contiguous and noncontiguous local contacts and TADs assemble into compartments giving rise to higher-order structures [22,45]. When contiguous TADs interact, they can create a “meta-TAD,” while noncontiguous TADs form extremely long-range multipartite cis- and trans-interactions (Fig. 1B). The formation of such extremely long-range interactions relies either on underlying heterochromatin marks (i.e., H3K9me3) assembling into “TAD-cliques” on active marks, such as H3K27ac, building multipartite enhancer hubs [46–50], or on H3K36me3, correlating with contacts between coding regions of very active, intron-rich genes [29,31]. Moreover, assembly of higher-order interchromosomal multipartite interactions has been shown to center at specific nuclear bodies; nuclear speckles for gene rich, active hubs, and nucleolus and lamina for gene poor, inactive or heterochromatic regions indicating that spatial positioning of these (in)active hubs in the nucleus is not random [49–51]. This large-scale chromatin folding appears highly heterogeneous, indicating that chromosomes adopt various conformations in single cells [5–7,9,10,12,52]. Nevertheless, this cell-to-cell variability likely reflects the dynamic nature of chromatin folding rather than the lack of functional importance of this organization. Moreover, the association of higher-order interactions with nuclear bodies and lamina may act as a backbone that molds the overall 3D genome organization in the nucleus as well as helping to modulate its function [50,51]. However, understanding the functional significance of higher-order interactions is a key issue that is difficult to address comprehensively. Historically, one of the best-studied examples of long-range chromatin contacts involves H3K27me3 Polycomb-regulated regions at Hox loci, in cis and in trans, both in Drosophila and in mammals [4,29,53–57]. Here, Hox loci engage in multipartite interactions with other H3K27me3 domains and segregate into Polycomb foci, which dissolve upon Polycomb component removal and Hox genes become misexpressed, leading to homeotic transformations under sensitized conditions [53,56]. Nevertheless, despite these and other functional studies, it is yet not clear if all higher-order structures depend on the underlying histone marks and/or their writers, or whether other factors could play a role as well. Notably, although not unambiguously uncoupled from the histone marks, the active, but not inactive transcriptional start sites (TSSs) were shown to cluster together in a manner that depends on their expression level, exon and splicing event number, indicating that additional mechanisms could be governing and/or stabilizing long-range interactions [29]. Furthermore, similarly to the cell-type specific TFs instigating contacts at the TAD level, Oct4, Nanog and Klf4 orchestrate higher-order chromatin folding in mESC, indicating that possible alternative folding mechanisms likely exist [29,54,58,59]. A recent striking example of cell-type specific TF driven higher-order interactions comes from olfactory sensory neurons where 63 enhancers scattered over 18 different chromosomes form a hub in the high local presence of LH2X and LDB1 factors which stabilize the hub and facilitate the choice of the olfactory gene to be expressed in each neuron [60]. Therefore, although compartment-like homotypic domain interactions appear as a global feature of chromosomal folding, specific long-range chromatin interactions, such as the ones in olfactory receptors, are cell-type specific, suggesting that they could represent a key feature driving cell identity.

Together, these examples indicate that higher-order chromatin folding relies on several distinct rules; however, the actual mechanism is still elusive. While CTCF and cohesin are instrumental at short- and intermediate-level chromatin folding, it is clear that this mechanism is not employed at the compartment-level organization, as protosomal deletions of CTCF and cohesin almost entirely disrupt short- and intermediate-range chromatin folding but leave higher-order structures unaffected, or even exacerbated [16–19,44]. Yet, large-scale chromatin interactions also appear independent from one another, where the loss of Polycomb proteins specifically affects interaction among Polycomb target regions without altering the overall chromosome conformation [12,54]. Therefore, by their nature, higher-order chromatin interactions represent a heterogeneous but distinct form of architectural folding with an underlying functional component, proving that robust and precise gene regulation is rooted in several layers of chromatin folding as a prerequisite for life.

Chromosomal Structures Mediate Genome Function
Conclusions and Perspectives

The last decade has been a game changer for research on nuclear organization, higher-order chromatin structure and gene expression. A panoply of technologies has surged, allowing us to describe and modulate nuclear architecture with unprecedented versatility. One can describe nuclear organization by microscopy with unprecedented resolution [61] and multiplicity of molecular signatures, both in terms of number of chromatin loci to be studied and of their expression [9,11,52,62,63]. One can pinpoint the DNA counterparts of chromatin contacts at subkilobase resolution in bulk and in single cells [6,45,46,51,64,65] as well as provide indications on molecular distances between specific loci and nuclear landmarks genome-wide [66]. Finally, it is now possible to modify genome and epigenome function genome-wide or at specific loci thanks to genome technologies like CRISPR/Cas [67–72]. These technologies can be applied to a variety of organisms and cell types, with a decreasing number of limitations due to cell number requirement, cell type specificity, and thickness of the tissue being partially or totally lifted as technologies progress. Furthermore, data analysis and machine learning algorithms are improving at a proportionate speed allowing us to extract relations between data and to identify candidates for causality with much higher efficiency and accuracy. Therefore, it is gradually becoming possible to move from the study of cells in culture to more complex systems such as tissue becoming possible to move from the study of cells in culture to more complex systems such as tissue being organized (LifeTime https://lifetime-fetflagship.eu/) or being organized (LifeTime https://lifetime-fetflagship.eu/) to tackle these issues on a large scale, in a way that will allow us to gain much deeper understanding on these exciting questions and to apply the ensuing knowledge to improve human health.

Acknowledgments

I.J. was supported by an EMBO fellowship nr. ALTF 559-2018 and the Laboratory of Excellence EpiGenMed. Q.S. was supported by the French Ministry of Higher Education and Research, and La Ligue Nationale Contre le Cancer. F.B. was supported by CNRS. Research in the G.C. laboratory was supported by grants from the CNRS, the European Research Council No 788972 (Advanced Grant 3DEpi), the European Union’s Horizon 2020 research and innovation programme under grant agreement No 676556 (MuG), the Agence Nationale de la Recherche (ANR-15-CE12-0006 EpiDevoMath), the Fondation pour la Recherche Médicale (DEI20151234396) and the French National Cancer Institute (INCa).

Received 29 May 2019;
Received in revised form 2 October 2019;
Accepted 18 October 2019
Available online 02 November 2019

Keywords:
Chromatin structure;
Genome folding;
Transcriptional regulation;
Topologically Associating Domains;
Compartments

1Equal contribution.

References


Cell Atlas https://www.humancellatlas.org/) or Nucleome https://www.4dnucleome.org/, Human
[27] Y. Zhan, L. Mariani, I. Barozzi, E.G. Schulz, N. Bluthgen, M. Stadler, et al., Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes, Genome Res. 27 (2017) 479–490.


