OPEN ACCESS

Repository of the Max Delbrück Center for Molecular Medicine (MDC) Berlin (Germany)

http://edoc.mdc-berlin.de/14021/

Models of chromosome structure

Nicodemi, M., Pombo, A.

Published in final edited form as: Current Opinion in Cell Biology. 2014 Jun ; 28: 90-95 | doi: 10.1016/j.ceb.2014.04.004 Elsevier ►

Title: Models of chromosome structure

Mario Nicodemi^{1,*} and Ana Pombo^{2,*}

¹Universita' di Napoli "Federico II", Dipartimento di Fisica, INFN Sezione di Napoli, CNR-SPIN, Complesso Universitario di Monte S. Angelo, Via Cintia, 80126 Napoli, Italy. *E-mail: mario.nicodemi@na.infn.it*

²Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Robert Rössle Strasse 10, 13125 Berlin-Buch, Germany. *E-mail: ana.pombo@mdc-berlin.de*

*To whom correspondence should be addressed:

Keywords: genome organization, long-range chromatin interactions, chromosome conformation capture (3C), microscopy, fluorescence in situ hybridization, polymer physics, computer simulations.

Abstract

Understanding the mechanisms that control chromosome folding in the nucleus of eukaryotes and their contribution to gene regulation is a key open issue in molecular biology. Microscopy and chromatin-capture techniques have shown that chromatin has a complex organization, which dynamically changes across organisms and cell types. The need to make sense of such a fascinating complexity has prompted the development of quantitative models from physics, to find the principles of chromosome folding, its origin and function. Here, we concisely review recent advances in chromosome modeling, focusing on a recently proposed framework, the Strings & Binders Switch (SBS) model, which recapitulates key features of chromosome organization in space and time.

Introduction

Chromosomes have a complex organization across spatial scales within the cell nucleus [1-3]. They occupy separate, yet interacting chromosomal territories [4, 1] where long-range chromatin interactions are functionally important. Each territory is partitioned in megabasepair-long domains enriched for internal contacts, known as topological associated domains (TADs) [1, 5, 6], while long stretches of chromatin interact with the nuclear lamina (the lamina-associated domains, LADs) [3], and with a variety of other functional compartments [1]. Genes come together, for instance, on specialized transcription factories, promoting proximity between different transcription units according to their expression state [7-10]. Splicing factors, the machinery that splices nascent transcripts into messenger RNA, accumulate in splicing speckles, which are often associated with active genes, while repressed chromatin associates with heterochromatic regions or Polycomb bodies [1]. Early replication origins are also clustered in replication factories, which stably reassemble in consecutive cell cycles [11].

The understanding of chromatin organization has inspired the development of models for many years [12-15]. Early models were often a visual summary of conformations identified by microscopy, aiming mostly to consolidate, in one image,

the many features observed (e.g. [1]). The increasing level of details exposed by experimental advances has triggered the need to further develop quantitative polymer models from physics. And here we focus especially on a few, more recent developments.

Polymer models of chromatin organization

The simplest model of a free polymer, i.e., a polymer experiencing only selfavoidance effects, is the Self-Avoiding Walk (SAW) model. A SAW polymer folds spontaneously in a random conformation (Fig. 1) and, thus, in a mixture, entropy induces full intermingling of polymers. Pioneer work by Kreth and Cremer (2004) introduced a model where polymers also experience a strong self-attraction force that can produce discrete chromosome territories [16].

Entropy might also play a role in favoring territorial separation by limiting the intermingling between different chromosomes: models of free polymers folded into clustered loops can hardly find the space to penetrate into each other, experiencing an effective repulsion of entropic origin partially preventing overlap [17-19, 14]. These models help bridge the original view of chromosome territories fully separated by channels devoid of DNA (the CT-IC view) [20, 1], and the discovery of their intermingling (the ICN view) [4, 21]. Entropy-based models, however, cannot explain the variety of specific, functional contacts occurring through chromatin looping (such as enhancer-promoter interactions), its domain structure (LADs, TADs) and, in particular, fail to describe the behavior of the contact probability between genomic loci reported by Hi-C experiments [22, 23].

The development of 'chromosome conformation capture' (**3C**) methods [24], and their genome-wide extensions such as Hi-C [25, 22, 26], has invigorated this field with the prospect of semi-quantitative measures of chromosome folding. Hi-C contact matrices have shown, for instance, that the average interaction probability between pairs of genomic loci decreases with their genomic separation, approximately with power law decay in the 0.5-7Mb range [22].

The contact probability and, in particular, its power law exponent, α , is different, though, in different systems [23]. For instance, human embryonic stem cells (ESCs) have $\alpha \sim 1.6$ [23], while $\alpha \sim 1.1$ in human interphase lymphoblastoid cells [22] and in metaphase HeLa cells $\alpha \sim 0.5$ [27]. The exponent α reported in different species also varies widely: in Drosophila $\alpha \sim 0.7$ -0.8 [28], in yeast $\alpha \sim 1.5$ [29], in mouse ESCs the *Xist* locus has $\alpha \sim 0.7$ -0.9 [6, 30]. Even in a given cell system different chromosomes can have different exponents [31, 23]; e.g. the gene-poor and gene-dense human chromosomes 18 and 19 have α around 1.08 and 1.3, respectively, in human lymphoblastoid cells [23]. These observations have contradicted the initial hope that a single universal architecture, originally envisaged in the Fractal Globule (FG) model [22] (where $\alpha = 1$), might describe chromosome folding.

The FG state [22] is a transient, highly unstable [32] conformation encountered, e.g., by a free polymer allowed to expand from an initial high-compaction, non-entangled state. For the FG to be observable, the activity of topoisomerases (important nuclear enzymes that resolve chromatin entanglement by cutting the DNA and passing it through) must be negligible. The FG is knot free, which might allow for quick unfolding events in the chromatin structure. However, the FG model also fails to explain the formation of TADs or LADs, and microscopy data of inter-locus distances, in particular the fact that the separation of distant regions on a chromosome saturates at the territory size [23].

An alternative scenario to explain chromosome folding was proposed in the Interchromatin Network (ICN) model [4], where folding derives from chromatin interactions with complexes that promote looping (e.g. contacts between coexpressed genes at transcription or replication factories) and associations with nuclear landmarks, such as the lamina. The Strings & Binders Switch (SBS) model [33, 34, 23] explored for the first time such a scenario on quantitative grounds using polymer physics (Fig. 2a).

The SBS model, in its simplest version, investigates a polymer and its interactions with diffusing binders that can crosslink the polymer specific binding sites. It can also be expanded to consider many polymers, different binder species, and interactions with nuclear landmarks [35, 36]. In the model, the position of the binding sites along the polymer is assigned; the concentration and affinity of binders can be changed [23], as resulting, for instance, from the up-regulation of a corresponding gene or from chemical modifications of a chromosome sequence.

With simple and generic settings, the SBS model has been used to illustrate several aspects of chromatin folding in quantitative terms: (a) three major folding classes exist (open, closed and Θ -point; Fig. 2b), corresponding to stable emergent phases of polymer physics (the Fractal Globule state is one of many possible transient states of the SBS model); (b) conformational changes can be sharply regulated (switchlike) by simple strategies, e.g., protein up-regulation or chromatin modifications, which may help understand the mechanisms employed, e.g., to transduce (*analog*) transcription factor levels into (*digital*) conformational switches; (c) that randomly diffusing binding molecules can establish and dynamically change, by thermodynamics mechanisms, architectural patterns, such as territory formation, TADs or LADs, and the looping out of large stretches of chromatin from territories (Fig. 2c); (d) that population and single cell microscopy and Hi-C data, such as contact probabilities (Fig. 3) and spatial distances, can be rationalized in a single framework [23]. The results of the SBS model are confirmed by similar findings in the Dynamic Loop (DL) model, which has been used, in particular, to explore the effects of entropy in mixtures of long, looped polymers [37, 14].

The picture that emerges from the SBS model is that chromatin exists inside nuclei as a complex mixture of differently folded regions according to local specific factors, which can self-organize across spatial scales by general physical mechanisms. Specificity of binding at different loci or domains can be achieved, for example through a combination of single molecular factors. In fact, more complex architectures, with different, nested layers of organization can be constructed

within the SBS model. Beyond binding factors, other molecular mechanisms are likely to contribute to chromatin folding into compact conformations, such as supercoiling [30] and plectoneme formation [38].

Technical limitations

Irrespective of promising developments, it is helpful to appreciate the limitations of the different methodologies. Imaging of nuclear architecture has improved to exciting extents, but it is not fraught from artifacts (e.g. milder fixation and structure loss) and currently suffers from low throughput options to screen broadly for chromatin interactions, with exciting exceptions [39]. 3C-based mapping of chromatin interactions has limitations too. Cells are also mildly fixed, broken up to extract nuclei, DNA is cut by restriction enzymes and DNA free ends ligated, before sequencing and alignment, all steps being fraught with unclear inefficiencies and biases [40, 41]. Efforts to improve all these technologies are ongoing.

Similarly, we have to overcome limitations of physics models, such as their hidden or oversimplifying assumptions. *Ad-hoc* prepared polymer structures have been discussed, which might reproduce architectures identified experimentally without adding a real new level of understanding of the underlying mechanisms. Due to computer speed limitations, it is currently only possible to model comparatively simple polymers, but as computer power grows soon whole chromosomes will be modelled at high resolution. Yet, the discovery that polymer properties scale with polymer size, awarded the Nobel Prize in Physics, has already demonstrated that general and important features of chromosome folding can be derived using relatively short and simple polymers [42].

Future perspectives

Many issues remain that require the development of more detailed models. On a general ground, for instance, the effects of crowding have been only partially considered as much as the constraints imposed by nuclear size and shape or by the associations with nuclear bodies. Nevertheless, polymer models can be employed to

study specific genomic loci (e.g., the *Xist* locus) [36] to discriminate different biological scenarios and thus helping to identify the determinants of folding.

Polymer models constrained by Hi-C data have been used to attempt reconstructions of chromosome 3D shapes [43, 44]. These efforts suffer from the limitation that population averaged Hi-C data return 'average' conformations depending on the specific type of constraints and algorithms employed. To circumvent such limitations, more realistic polymer models, such as the SBS, can be used to reconstruct spatial conformations, and have the potential to predict chromatin interaction sites that best explain Hi-C data.

From a biological point of view a number of key questions are still unresolved, such as the origin of **(a)** specificity of interactions (e.g. why there is a preference for pairing between specific chromosomes in specific cell types [45, 4] and why are chromosome homologues often separated from each other [46]), **(b)** chromosomal intermingling [4], or **(c)** the effects of other cellular processes, such as gene expression or DNA replication, on chromosome organization.

In conclusion, as the complex 4D organization of chromatin remains still largely mysterious, simple models of polymer physics, tested against real data, are trying to provide a first picture of the basic principles and molecular mechanisms of folding. In the near future, further developments in the models employed, exploiting technical progresses and more detailed biological information, could push even further our comprehension and could guide the design of targeted experiments to resolve some of the many crucial open questions, hopefully advancing also our understanding of related diseases.

Acknowledgements

A.P. thanks the Helmholtz Foundation for support.

References

- [1] Lanctot C, Cheutin T, Cremer M, Cavalli G, Cremer T: *Dynamic genome architecture in the nuclear space: Regulation of gene expression in three dimensions. Nature Rev Genet* 2007, 8:104-115.
- [2] Misteli T: *Beyond the sequence: Cellular organization of genome function. Cell* 2007, 128:787-800.
- [3] Bickmore WA, van Steensel B: *Genome architecture: Domain organization of interphase chromosomes. Cell* 2013, 152:1270-1284.
- [4] Branco MR, Pombo A: Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol 2006, 4:e138.
- [5] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012, 485:376-380.
- [6] Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, *et al.*: *Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature* 2012, 485:381-385.
- [7] Pombo A, Jackson DA, Hollinshead M, Wang Z, Roeder RG, Cook PR: *Regional specialization in human nuclei: Visualization of discrete sites of transcription by RNA polymerase III. EMBO J* 1999, 18:2241-2253.
- [8] Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, Corcoran AE, Fraser P: *Myc dynamically and preferentially relocates to a transcription factory occupied by igh. PLoS Biol* 2007, 5:e192.
- [9] Ferrai C, Xie SQ, Luraghi P, Munari D, Ramirez F, Branco MR, Pombo A, Crippa MP: *Poised transcription factories prime silent upa gene prior to activation. PLoS Biol* 2010, 8:e1000270.
- [10] Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS, *et al.*: *Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet* 2010, 42:53-61.
- [11] Jackson DA, Pombo A: *Replicon clusters are stable units of chromosome structure: Evidence that nuclear organization contributes to the efficient activation and propagation of s phase in human cells. J Cell Biol* 1998, 140:1285-1295.
- [12] Langowski J: *Polymer chain models of DNA and chromatin. Eur Phys J E Soft Matter* 2006, 19:241-249.
- [13] Emanuel M, Radja NH, Henriksson A, Schiessel H: *The physics behind the larger scale organization of DNA in eukaryotes. Phys Biol* 2009, 6:025008.
- [14] Tark-Dame M, van Driel R, Heermann DW: *Chromatin folding--from biology to polymer models and back. J Cell Sci* 2011, 124:839-845.
- * This review discusses multiple roles of entropy in shaping chromosome conformations.
- [15] Barbieri M, Scialdone A, Gamba A, Pombo A, Nicodemi M: *Polymer physics, scaling and heterogeneity in the spatial organisation of chromosomes in the cell nucleus. Soft Matter* 2013, 9:8631-8635.

- [16] Kreth G, Finsterle J, von Hase J, Cremer M, Cremer C: *Radial arrangement of chromosome territories in human cell nuclei: A computer model approach based on gene density indicates a probabilistic global positioning code. Biophys J* 2004, 86:2803-2812.
- [17] Marenduzzo D, Micheletti C, Cook PR: *Entropy-driven genome organization*. *Biophys J* 2006, 90:3712-3721.
- *A discussion of the effects of entropy-based interactions on chromatin organization.
- [18] Bohn M, Heermann DW, van Driel R: *Random loop model for long polymers. Physical review. E, Statistical, nonlinear, and soft matter physics* 2007, 76:051805.
- [19] Rosa A, Everaers R: *Structure and dynamics of interphase chromosomes. Plos Comp Biol* 2008, 4:e1000153.
- [20] Cremer T, Kreth G, Koester H, Fink RH, Heintzmann R, Cremer M, Solovei I, Zink D, Cremer C: *Chromosome territories, interchromatin domain compartment, and nuclear matrix: An integrated view of the functional nuclear architecture. Crit Rev Eukaryot Gene Expr* 2000, 10:179-212.
- [21] Branco MR, Pombo A: Chromosome organization: New facts, new models. Trends Cell Biol 2007, 17:127-134.
- [22] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al.*: *Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science* 2009, 326:289-293.
- **The Hi-C method was introduced, allowing genome-wide mapping of chromatin contacts, and measurement of the frequency of chromatin interactions. It also proposes the Fractal Globule model to explain properties of chromatin folding at 0.5-7Mb genomic scales.
- [23] Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, Nicodemi M: *Complexity of chromatin folding is captured by the strings and binders switch model. PNAS* 2012, 109:16173-16178.
- **The SBS model is used here for the first time to explore single cell FISH data and population based HiC data. It explains many features of available experimental data, and proposes a scenario where chromatin is a mixture of differently folded regions.
- [24] Dekker J, Rippe K, Dekker M, Kleckner N: *Capturing chromosome conformation. Science* 2002, 295:1306-1311.
- [25] Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W: *Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell* 2002, 10:1453-1465.
- [26] Rodley CD, Bertels F, Jones B, O'Sullivan JM: *Global identification of yeast chromosome interactions using genome conformation capture. Fungal Genet Biol* 2009, 46:879-886.
- [27] Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J: *Organization of the mitotic chromosome. Science* 2013, 342:948-953.
- *This paper shows that chromatin conformation depends on the cycle stage and that a compact folding state is observed at metaphase.

- [28] Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G: *Three-dimensional folding and functional organization principles of the Drosophila genome. Cell* 2012, 148:458-472.
- [29] Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS: *A three-dimensional model of the yeast genome. Nature* 2010, 465:363-367.
- [30] Benedetti F, Dorier J, Burnier Y, Stasiak A: *Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. Nucleic Acids Res* 2013.
- *This paper discusses the hypothesis that supercoiling can have an effect in the structure of Hi-C contact matrices.
- [31] Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L: *Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nature Biotech* 2011, 30:90-98.
- **Introduces an alternative 3C-based method, TCC, and the concept of modelling ensemble of chromosome conformations from genome-wide chromatin conformation capture. It incorporates the concept of population conformation in the interpretation of HiC data.
- [32] Schram RD, Barkema GT, Schiessel H: *On the stability of fractal globules. J Chem Phys* 2013, 138:224901.
- *An exploration of the unstable properties of the Fractal Globule state and the difficulties in generating it.
- [33] Nicodemi M, Panning B, Prisco A: *A thermodynamic switch for chromosome colocalization. Genetics* 2008, 179:717-721.
- **Here the idea that chromosome conformations result from chromatin interactions and are controlled by basic thermodynamic mechanisms of polymer physics was first introduced.
- [34] Nicodemi M, Prisco A: *Thermodynamic pathways to genome spatial organization in the cell nucleus. Biophys J* 2009, 96:2168-2177.
- [35] Scialdone A, Nicodemi M: *Diffusion-based DNA target colocalization by thermodynamic mechanisms. Development* 2010, 137:3877-3885.
- [36] Scialdone A, Cataudella I, Barbieri M, Prisco A, Nicodemi M: *Conformation regulation of the X chromosome inactivation center: A model. PLoS Comput Biol* 2011, 7:e1002229.
- *This paper introduces the idea that symmetry breaking and polymer physics mechanisms can contribute to produce different conformations on different alleles.
- [37] Bohn M, Heermann DW: *Diffusion-driven looping provides a consistent framework for chromatin organization. PloS One* 2010, 5:e12218.
- **This is an extensive study of the effects polymer interactions on chromatin folding.
- [38] Le TB, Imakaev MV, Mirny LA, Laub MT: *High-resolution mapping of the spatial organization of a bacterial chromosome. Science* 2013, 342:731-734.
- [39] Roukos V, Voss TC, Schmidt CK, Lee S, Wangsa D, Misteli T: *Spatial dynamics of chromosome translocations in living cells. Science* 2013, 341:660-664.
- [40] O'Sullivan JM, Hendy MD, Pichugina T, Wake GC, Langowski J: *The statisticalmechanics of chromosome conformation capture. Nucleus* 2013, 4:390-398.

- **Important study exploring limitations of chromosome conformation capture methods, in particular the biases introduced by ligation of each DNA fragment to at most two other fragments.
- [41] Belmont AS: Large-scale chromatin organization: The good, the surprising, and the still perplexing. Curr Opin Cell Biol 2014, 26C:69-78.
- [42] de Gennes PG, *Scaling concepts in polymer physics*. 1979: Cornell University Press.
- [43] Umbarger MA, Toro E, Wright MA, Porreca GJ, Bau D, Hong SH, Fero MJ, Zhu LJ, Marti-Renom MA, McAdams HH, *et al.*: *The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. Mol Cell* 2011, 44:252-264.
- [44] Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P: *Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature* 2013, 502:59-64.
- [45] Parada LA, McQueen PG, Misteli T: *Tissue-specific spatial organization of genomes. Genome Biol* 2004, 5:R44.
- [46] Khalil A, Grant JL, Caddle LB, Atzema E, Mills KD, Arneodo A: *Chromosome territories have a highly nonspherical morphology and nonrandom positioning. Chromosome Res* 2007, 15:899-916.

Figure Legends

Fig. 1. Polymer models and conformations

a) The self-avoiding walk (SAW) model describes a free polymer, i.e., a polymer having no interactions beyond excluded volume. A SAW polymer has a randomly folded conformation.

b) A mixture of polymers would show substantial levels of intermingling because of mixing entropy.

c) If each polymer experiences self-attraction (e.g. Kreth and Cremer model [16]), polymers can confine into clearly separated territories.

d) If polymers have looped conformations, 'rosette'-like, intermingling can be reduced because loops interpenetrate each other with more difficulty and thus experience an effective repulsion force (entropic force).

Fig. 2. The Strings & Binders Switch (SBS) model

a) In the SBS model [33, 23], chromatin is represented by a self-avoiding-walk (SAW) chain, which has binding sites for diffusing binders. The basic model parameters are the binder concentration, c_m, and their affinity for the polymer binding sites, E_x. The interactions of polymer with molecular binders can produce loops. The model stable emergent thermodynamic phases correspond to different conformational classes.

b) The SBS model identifies a switch-like behavior in chromatin folding in response to changes in binder affinity or concentration. The SBS phase diagram includes a phase where the polymer folds in a random open conformation (in the universality class of the free SAW) and a phase where it spontaneously folds into a compact closed conformation. At the phase transition point, there is the Θ -point state. Conformational changes are achieved by crossing the phase boundary, with no need of parameter fine-tuning.

c) The formation of chromatin domains and chromatin looping can be modeled within the SBS model, as result of the specialization of the polymer binding sites and their binding molecules. The corresponding contact matrices have general features

similar to those found by Hi-C methods.

Fig. 3. Contact probabilities estimated from HI-C data are explained by the SBS model.

a) The average contact probability, $P_c(s)$, measures the extent of chromatin interactions between pairs of loci separated by a given genomic distance.

b) In Hi-C data, the average contact probability, $P_c(s)$, is found to decrease with their genomic separation, *s*, with a power law decay at least within the 0.5-7Mb range, $P_c(s) \sim 1/s^{\alpha}$ [22]. The exponent, α , of the power law is different in different systems and in different chromosomes [23].

c) Within the SBS model, α is **2.1** in the open state, **1.5** in the Θ -point state, and **0** in the compact state. Thus, mixtures of modelled SBS polymers including, for example, a fraction, *f*, of open and a fraction, *1-f*, of compact polymer regions can easily explain the range of exponents found experimentally. This scenario simulates the fact that different genomic regions assume open or closed chromatin conformations, which need to be captured by polymer modelling approaches.

Nicodemi & Pombo, Fig I







