Integrating experiment, theory and simulation to determine the structure and dynamics of mammalian chromosomes

G. Tiana[1] and L. Giorgetti[2]
[1]Center for Complexity and Biosystems and Department of Physics, Università degli Studi di Milano and INFN, via Celoria 16, 20133 Milano, Italy
[2]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

Corresponding authors:
Guido Tiana, guido.tiana@unimi.it
Luca Giorgetti, luca.giorgetti@fmi.ch

Abstract

Eukaryotic chromosomes are complex polymers, which largely exceed in size most biomolecules that are usually modelled in computational studies and whose molecular interactions are to a large extent unknown. Since the folding of the chromatin fiber in the cell nucleus is tightly linked to biological function and gene expression in particular, characterizing the conformational and dynamical properties of chromosomes has become crucial in order to better understand how genes are regulated. In parallel with the development of experimental techniques allowing to measure physical contacts within chromosomes inside the cell nucleus, a large variety of physical models to study the structure and mechanisms of chromosome folding have recently emerged. Such models can be roughly divided into two classes, based on whether they adopt specific hypotheses on the interaction mechanism within chromosomes, or learn those interactions on the available experimental data using the principle of maximum entropy. All of them have played a key role in interpreting experimental data and advancing our understanding the folding principles of the chromatin fiber.

Highlights

- Chromatin is a large and complex molecular system whose interactions are to a large extent unknown.
- Coarse-grained physical models of chromosomes were developed based on different assumptions on internal interactions.
- Hypothesis-driven models allow to test predictions on the molecular mechanisms that govern chromosome folding
- Maximum-entropy based approaches allows developing models based on experimental data only, without introducing a priori hypotheses biases.
- Model predictions have proven useful to understand basic principles of chromosome folding and have important implications for gene regulation.

Introduction

Chromatin is a complex macromolecule, consisting of genomic DNA and histone proteins. Histones assemble into nucleosomes, around which DNA is wound in the eukaryotic cell nucleus. This contributes to compact meters of linear DNA to fit the μm-sized nucleus. In vitro, nucleosome-wound DNA folds into a fiber of ~30 nm diameter [1]. However, chromatin in vivo is far less characterized, and it is unclear if 30-nm fibers exist in living cells. It has been suggested indeed that in vivo, chromatin may occur as a dynamic assembly of thinner fibers [2] or even as a mixture of different fiber states [3,4]. Irrespective of the short-scale structure of the chromatin fiber itself however, in mammals the stretch of chromatin fiber corresponding to each chromosome (typically hundreds of million base pairs, or Mb) folds up into large 'chromosomal territories' that assume a globular conformation in interphase. Chromosomal territories occupy the large majority of the nuclear volume and remain only partially intertwined with each other [5].

Besides allowing DNA to fold compactly into the nucleus, the chromatin fiber has an important role in controlling core biological functions such as gene expression and DNA replication. It is well established that chemical modifications of DNA and histones, orchestrated by a variety of enzymatic complexes, provides a fundamental regulatory layer that contributes determining whether a gene is transcribed or not [6]. However, accumulating evidence suggests that an additional regulatory mechanism may act through the control of three-dimensional structure of the chromatin fiber inside chromosomal territories. Large-scale chromatin folding appears indeed to modulate the mutual distance between genes and specific regulatory genetic elements such as transcriptional enhancers, which can be placed hundreds of thousands of bases away from the genes that they control and whose spatial proximity is required for activating and sustaining gene expression [7]. Understanding the structure and the dynamics of chromatin over genomic scales ranging from tens of thousands to one million basepairs, where most interactions between enhancers and genes occur, has therefore become fundamental to understand transcriptional control.

Models of the chromatin fiber below the 10-kilobase scale have been developed based on biophysical and biochemical data, and successfully used to describe chromatin at atomic resolution or using multi-scale descriptions of DNA and of the nucleosomes (see Refs. [8,9] for recent reviews). However, the length scales that are relevant in studying transcriptional regulation by long-range enhancers are much larger and can be characterized using more coarse-grained polymer models based on experimental techniques derived from molecular biology, which are the focus of this review.

The main experimental tool that has been developed in the last two decades to investigate the three-dimensional arrangement of chromatin in vivo is Chromosome Conformation Capture (3C), a method in which digestion and successive re-ligation of crosslinked chromatin in cell nuclei allows the detection of spatial proximity between DNA sequences [10]. In recently developed methods derived from 3C, such as 4C, 5C and notably Hi-C [11,12], high-throughput sequencing and quantification of 3C ligation products enables the inference of interaction frequencies, i.e. the frequencies at which pairs of genomic loci are crosslinked in a population of cells because of their spatial proximity at the very moment of the crosslinking [13]. A typical Hi-C experiment results into a large two-dimensional map of crosslinking frequencies between any pair of genomic loci within and across chromosomes, with a resolution that can reach a few thousand base pairs in recent versions of the protocol [14,15] thus corresponding to a matrix of ~3 million bins for the human genome. With the exception of recently developed single-cell versions of the Hi-C protocol [16][17][18][19], whose

resolution is currently limited to hundreds of kb (hence too low to resolve contacts between genes and enhancers), contact probabilities returned by Hi-C are typically averaged over millions of cells (Figure 1).

Population-averaged Hi-C maps of mammalian chromosomes display well-defined patterns of interactions, which substantially deviate from what would be expected from polymers with uniform or random interactions. The most prominent feature is the presence of sets of nested interaction domains (Figure 1), which is indicative of a hierarchical folding of the fiber. Several levels of this hierarchy have been identified and investigated experimentally. At the 1-10 Mb scale, so-called 'compartments' were identified as large chromosomal regions that tend to preferentially associate and exclude each other, reflecting cell-type specific gene expression states [12]. At a lower level, in the 100kb-1Mb scale, topologically associating domains (TADs) appear as contiguous domains of preferential interactions of the chromatin fiber [20-22]. TADs are further partitioned into smaller domains, which largely overlap with loops bridging pairs of DNA binding sites for the transcription factor CTCF (CCCTC-binding factor) [14]. These levels belong to an even richer hierarchy wherein no level seems to be privileged from a structural point of view; strikingly however, several functional properties such as cell-type conservation of domain boundaries and gene co-regulation during cell differentiation are maximized at the scale of TADs [23]. TADs are indeed thought to act as regulatory microenvironments by restricting enhancer interactions to a limited set of genes (and vice versa), and TAD boundary deletions have been shown to result in ectopic gene expression leading to pathological states [24-26].

Despite our increasing ability in detecting chromosomal structures using 3C-derived techniques, we still ignore the mechanisms that give rise to these structures. The transcription factor CTCF plays a key role in the establishment of TADs [27,28], together with the cohesin complex [29,30]. However, whether it does so by inactively stabilizing loops between CTCF-bound sites [31], or rather by blocking ATP-dependent loop-extruding factors and thus effectively pulling two CTCF-bound sites together [32-35], is a matter of active debate. In addition, interactions between active genomic regions even in the absence of CTCF and/or cohesion could play an important role in mediating specific chromosomal interaction [27,29,30,36,37].

Several theoretical and computational approaches have been proposed to explain the complex patterns observed in Hi-C experiments. These models have two objectives: 1) reconstructing the actual three-dimensional shape of the chromatin fiber within chromosomes in the 10kb-1Mb scale, and 2) understanding the mechanisms that could give rise to such structures. Common to all modelling approaches is the necessity of employing a description of the chromatin fiber which is highly coarse-grained. Indeed, mammalian genomes are billions of base pairs long, which amounts to ~10^{12} atoms including histone proteins, even excluding including the surrounding water, ions and the proteins that are present in the dense nuclear environment. It is clear that no single computational approach can deal with these numbers and describe the whole genome within the nucleus, but rather multiple approaches must be used, each containing various levels of coarse graining [8,9]. The study of large-scale chromosome organization, in particular, benefits from highly reduced descriptions based on elementary units in polymer models spanning genomic regions of the order of several kb (or even higher depending on the size of the simulated region), which keeps no trace of the actual internal molecular organization of the chromatin fiber.

Hypothesis-driven models of chromosome architecture

One of the fundamental problems in modeling the conformational properties of chromatin is that the mechanisms mediating interactions between chromosomal loci (e.g. between DNA sites bound by CTCF) are unknown. Modeling higher-order chromatin architecture is thus very different from modelling the structure of proteins or RNA, where interactions between atomic constituents are known at least approximately. Several chromosome models have been proposed based on increasingly complex hypotheses on the forces that give rise to chromosomal structures (see Fig. 2). The simplest hypothesis is that chromatin is a self-repulsive polymer, where genomic sequences exclude each other by steric hindrance. This very simple assumption is sufficient to reproduce the segregation of chromosomes into chromosomal territories but, not unexpectedly, does not predict the specific patterns observed in Hi-C experiments [38]. Interestingly however, segregation into domains resembling TADs can be predicted by further including a torsion potential which simulates the effect of DNA supercoiling [39,40], pointing at a possible role for supercoiling in contributing to the establishment of specific chromosomal structures.

Assuming instead that direct, short-range interactions occur between specific sites looping out intervening DNA (which implicitly mimics protein-mediated interactions) allows polymer models to reproduce a subset of experimental observations. The presence of interaction domains for example can be reproduced by imposing attractive interactions between loci that share common histone modifications [41], by empirically imposing that co-regulated genes should interact [42], or also by quenching a random heteropolymer with attractive interactions [43,44]. Similarly, the scaling behavior of physical distances observed in DNA fluorescence in situ hybridization (DNA FISH) can be reproduced by a dynamic loop model, which describes the formation/disruption of bonds in the polymer as a stochastic process [45]. If these models show that short-range attractions between looped sites can explain some of the experimentally observed features, they however remain agnostic as to what would mediate short-range interactions at the molecular level.

A more sophisticated hypothesis is that interactions between chromosomal locations are mediated by soluble proteins, which diffuse in the cell nucleus, bind to specific locations on DNA and transiently interact with each other. This assumption is able to qualitatively reproduce the existence of interaction domains, quantitatively describe some of their statistical properties [46-48], and even predict detailed experimental contact maps of specific domains [49,50]. Although it is a plausible hypothesis, it should be noted that there is no direct evidence that direct protein-mediated looping is the main mechanism controlling chromatin folding. In fact, the stability of interaction domains predicted by this class of models crucially depend on protein concentration, whereas partial depletion of CTCF for example does not have a dramatic effect on TADs and CTCF-associated loops [27,51]. Moreover, in these models, interaction domains appear as a consequence of imposing as many interacting proteins as the number of domains, which does not seem to be compatible with experimental observations.

More recently, another mechanism has been suggested to be able to generate effective interactions between CTCF-bound loci, which is known as loop extrusion [33]. Under this hypothesis, ATP-driven motor proteins (potentially cohesin) would stochastically extrude loops along the chromatin fiber, until they are stopped by CTCF bound to DNA. This model has the merit of being able to reproduce several observations, such as the recent finding that the effective interaction between CTCF sites depends on their reciprocal orientation on DNA [14,52] (which cannot be accounted by a direct looping mechanism), the emergence of TADs as nested CTCF-associated ‘loops’, and the detailed position of strong contacts as a function of the position of CTCF along the fiber [35]. In a variant of the model [53], the extruding protein

moves by Brownian dynamics according to a ratchet mechanism, thus releasing the need of a motor protein. Despite the high predictive power of this model, is should be emphasized that a direct experimental proof that loop extrusion actually takes place on chromosomes is still lacking. In addition, the model remains agnostic as to why CTCF, but no other proteins should block loop extrusion and how.

Reconstructing the conformational ensemble of chromatin from experimental data

An alternative modelling strategy aims at reconstructing the actual folding of chromosomal regions rather than understanding their origin. It consists in making no a priori hypotheses on the interactions between genomic loci, but rather inferring them based on the available 3C-based experimental data.

A simple way to implement this strategy is using mean-field approaches, following a restrain minimization procedure similar to those used to obtain the structure of proteins from NMR data. This method leads to average structures of chromosomal regions [54,55], or even of the whole genome [56], where statistical fluctuations, which are likely to be relevant for transcriptional regulation [57], were neglected.

Another class of methods is based on the principle of maximum entropy [58]. Originally formulated to link statistical physics to information theory, it provides a scheme to generate the probability distribution for the states of a system based on some prior data, and has been largely applied to use experimental data for molecular modelling [59-62]. When used to model data derived from the population-averaged states of a system at equilibrium, the maximum entropy principle states that the model of choice should produce a probability distribution for its conformations displaying three features: 1) the average contacts calculated for each pair of loci from this distribution should agree with the experimental data, 2) the probability distribution should be as little well-defined as possible, namely it should maximize the associated Shannon entropy, and trivially 3) it should be normalized to 1. In particular, point 2 guarantees that the resulting model is minimally biased, avoiding introducing subjective information which is not supported by data. With simple calculations and under the basic assumption that the system is at thermodynamic equilibrium, it is possible to show that the above hypothesis are met choosing a potential whose functional has the same dependence on the conformation of the system as the experimental data [59]. The numerical parameters of the potential can be found from an iterative Monte Carlo scheme [57].

Experimental Hi-C data arise from the detection of close spatial proximity between two loci; therefore, the least-biased potential is given by the sum of short-range interaction terms in the form of contact functions, whose numerical parameters must be obtained from the experimental data. As a matter of fact, the interactions between different chromosomal regions are mediated by nanometer-sized proteins complexes, which can be reasonably approximated as contact interactions on the length scale of the model.

It is worth emphasizing that the equilibrium hypothesis is questionable. Although the dynamics of the chromatin fiber within living nuclei is only poorly characterized, mammalian chromosomes are extremely long polymers whose equilibration time could largely exceed the duration of one cell cycle [38]. In fact, the overall large-scale (>1Mb) statistical properties of Hi-C maps are compatible with crumpled globule model, which describes a polymer far from equilibrium [63]. However, the relaxation of the chromatin fiber at the scale of TADs is expected to be shorter than for an entire chromosome. Indeed, tracking of single chromosomal loci by time-lapse fluorescence microscopy suggests that within single domains

the motion is sub-diffusive and conformational changes might take place on the time scale of minutes, that is enough to guarantee equilibration at these length scales [64] (but likely not at the scale of the whole nucleus [65]).

An implementation of the maximum-entropy principle allowed to obtain a realistic reconstruction of chromatin conformation within TADs in mouse embryonic stem cells, along with its cell-to-cell variability [57] (see Fig. 3). In this case, Monte Carlo sampling [66] of a polymer with spherical short-range interaction potentials was used in combination with a perturbative resampling scheme to optimize computational time [67]. In support of the basic assumptions in this approach, model predictions could be validated with independent experiments where distances among genomic loci were measured using DNA FISH in wild-type cells, as well as in cells carrying the deletion of key genomic sequences [57]. Interestingly, analysis of simulations suggested that long-range contacts within TADs do not occur as stable loops as previously proposed [68], but rather as stochastic and potentially highly dynamic events [64]. Extension of this approach to all TADs genome-wide in mouse embryonic stem cells suggested that there is a strong relation between the local structural properties of the chromatin fiber within a TAD, and namely the average degree of isotropy of its three-dimensional structure, and the transcriptional activity of genes harbored within the TAD [69].
In another application of the maximum entropy principle, theoretic information landscapes were used to reconstruct the folding of human chromosomes starting from Hi-C data [70,71], which led to the prediction that TADs might prevent the formation of knotted topologies in the genome.

Conclusions

Chromatin is an extremely long and complex polymer whose three-dimensional structure has an important role in shaping biological activities in the nucleus. The formation of specific chromosomal structures across the nested hierarchy of interaction domains detected in Hi-C datasets is governed by mechanisms that are still poorly understood. Importantly, given the complexity and the size of chromosomes, it has not been possible to study these mechanisms in vitro so far. Physical models have therefore arisen as powerful tools to study the conformational properties of chromosomes. On the one hand, hypothesis-driven models have been fundamental in shaping our perception of the mechanisms that could give rise to specific chromosomal structures, and notably to interaction domains such as TADs and sub-TAD CTCF-associated loops. On the other hand, data-driven models based on the maximum entropy principle have provided realistic, unbiased reconstructions of chromatin conformation and have revealed its high cell-to-cell variability, notably at the scale of TADs. This in turn has important implications in understanding how genes are regulated by enhancers in the context of a variable and potentially dynamic chromosomal landscape.
The main challenge in simulating mammalian chromosomes *in vivo* at the Mb length scale is that different from proteins or bare nucleic acids, realistic force fields that can be used in simulations simply do not exist, and must therefore be built from scratch (and tested against experimental observations). In the case of hypothesis-driven models, such interactions obey to explicit 'risky' hypotheses on the actual molecular interactions that give rise to specific structures. In the case of maximum-entropy principle based models, explicit hypotheses are avoided by allowing the computation to infer ad hoc potentials that best describe the ensemble of experimental observations. Interactions obtained from the maximum-entropy principle might not coincide with those that would be obtained in an ideal world by

integrating the actual physical interactions defined at the atomic scale; however, comparison between model predictions and experimental data indicates that these models are realistic and predictive.

Acknowledgments

Research in LG's laboratory is supported by the Novartis Research Foundation.

References

1. Bian Q, Belmont AS: **Revisiting higher-order and large-scale chromatin organization**. *Current Opinion in Cell Biology* 2012, **24**:359–366.

2. Fussner E, Ching RW, Bazett-Jones DP: **Living without 30nm chromatin fibers**. *Trends Biochem Sci* 2011, **36**:1–6.

3. Belmont AS: **Large-scale chromatin organization: The good, the surprising, and the still perplexing**. *Current Opinion in Cell Biology* 2014, **26**:69–78.

4. Maeshima K, Hihara S, Eltsov M: **Chromatin structure: Does the 30-nm fibre exist in vivo?** *Current Opinion in Cell Biology* 2010, **22**:291–297.

5. Cremer T, Cremer M: **Chromosome territories.** *Cold Spring Harb Perspect Biol* 2010, **2**:a003889–a003889.

6. Peterson CL, Laniel M-A: **Histones and histone modifications**. *Current Biology* 2004, **14**:R546–R551.

7. Spitz F: **Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles**. *Seminars in Cell & Developmental Biology* 2016, **57**:57–67.

8. Ozer G, Luque A, Schlick T: **The chromatin fiber: multiscale problems and approaches.** *Current Opinion in Structural Biology* 2015, **31**:124–139.

9. Dans PD, Walther J, Gómez H, Orozco M: **Multiscale simulation of DNA.** *Current Opinion in Structural Biology* 2016, **37**:29–45.

10. ** Dekker J, Rippe K, Dekker M, Kleckner N: **Capturing chromosome conformation**. *Science* 2002, **295**:1306–1311.

The authors describe the main experimental method employed to obtain the contact map of chromosomes in a population of cells.

11. Denker A, de Laat W: **The second decade of 3C technologies: detailed insights into nuclear organization.** *Genes Dev* 2016, **30**:1357–1382.

12. 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.

13. Fudenberg G, Mirny LA: **Higher-order chromatin structure: bridging physics and biology**.

Curr. Opin. Genet. Dev. 2012, **22**:115–124.

14. * Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al.: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping**. *Cell* 2014, **159**:1665–1680.

15. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan W-L, et al.: **Formation of new chromatin domains determines pathogenicity of genomic duplications**. *Nature* 2016, **538**:265–269.

16. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, et al.: **3D structures of individual mammalian genomes studied by single-cell Hi-C**. *Nature* 2017, **544**:59–64.

17. Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J: **Massively multiplex single-cell Hi-C**. *Nat. Methods* 2017, **14**:263–266.

18. Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, Mendelson Cohen N, Wingett S, Fraser P, Tanay A: **Cell-cycle dynamics of chromosomal organization at single-cell resolution**. *Nature* 2017, **547**:61–67.

19. Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, Razin SV, Mirny LA, Tachibana-Konwalski K: **Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition**. *Nature* 2017, **544**:110–114.

20. 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.

21. 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.

22. 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.

23. Zhan Y, Mariani L, Barozzi I, Schulz EG, Blüthgen N, Stadler M, Tiana G, Giorgetti L: **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.* 2017, **27**:479–490.

24. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al.: **Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions**. *Cell* 2015, **161**:1012–1025.

25. Hnisz D, Day DS, Young RA: **Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control**. *Cell* 2016, **167**:1188–1200.

26. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE: **Insulator dysfunction and oncogene activation in IDH mutant gliomas**.

Nature 2016, **529**:110–114.

27. * Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG: **Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization**. *Cell* 2017, **169**:930–944.

28. Merkenschlager M, Nora EP: **CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation.** *Annu Rev Genomics Hum Genet* 2016, **17**:17–43.

29. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering C, Mirny L, et al.: **Two independent modes of chromatin organization revealed by cohesin removal**. *Nature* 2017, doi:10.1038/nature24281.

30. Rao S, Huang S-C, Hilaire BGS, Engreitz JM, Perez EM, Kieffer-Kwon K-R, Sanborn AL, Johnstone SE, Bochkov ID, Huang X, et al.: **Cohesin Loss Eliminates All Loop Domains, Leading To Links Among Superenhancers And Downregulation Of Nearby Genes**. *bioRxiv* 2017, doi:10.1101/139782.

31. Pant V, Kurukuti S, Pugacheva E, Shamsuddin S, Mariano P, Renkawitz R, Klenova E, Lobanenkov V, Ohlsson R: **Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance.** *Mol Cell Biol* 2004, **24**:3497–3504.

32. Alipour E, Marko JF: **Self-organization of domain structures by DNA-loop-extruding enzymes.** *Nucleic Acids Research* 2012, **40**:11202–11212.

33. * Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA: **Formation of Chromosomal Domains by Loop Extrusion**. *Cell Rep* 2016, **15**:2038–2049.

34. Nichols MH, Corces VG: **A CTCF Code for 3D Genome Architecture**. *Cell* 2015, **162**:703–705.

35. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al.: **Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes**. National Acad Sciences; 2015:E6456–E6465.

36. Andrey G, Schöpflin R, Jerković I, Heinrich V, Ibrahim DM, Paliou C, Hochradel M, Timmermann B, Haas S, Vingron M, et al.: **Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding.** *Genome Res.* 2017, **27**:223–233.

37. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG: **Evolutionarily Conserved Principles Predict 3D Chromatin Organization**. *Mol Cell* 2017, **67**:837–852.e7.

38. Rosa A, Everaers R: **Structure and dynamics of interphase chromosomes.** *PLoS Comput Biol* 2008, **4**:e1000153.

39. * Benedetti F, Dorier J, Burnier Y, Stasiak A: **Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes.** *Nucleic Acids Research* 2013, **42**:2848–2855.

40. Benedetti F, Racko D, Dorier J, Burnier Y, Stasiak A: **Transcription-induced supercoiling explains formation of self-interacting chromatin domains in S. pombe**. *Nucleic Acids Research* 2017, **45**:9850–9859.

41. * Jost D, Carrivain P, Cavalli G, Vaillant C: **Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains.** *Nucleic Acids Research* 2014, **42**:9553–9561.

42. Di Stefano M, Rosa A, Belcastro V, di Bernardo D, Micheletti C: **Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19.** *PLoS Comput Biol* 2013, **9**:e1003019.

43. Nazarov LI, Tamm MV, Avetisov VA, Nechaev SK: **A statistical model of intra-chromosome contact maps.** *Soft Matter* 2015, **11**:1019–1025.

44. Shukron O, Holcman D: **Transient chromatin properties revealed by polymer models and stochastic simulations constructed from Chromosomal Capture data**. *PLoS Comput Biol* 2017, **13**:e1005469.

45. Tark-Dame M, Jerabek H, Manders EMM, Heermann DW, van Driel R: **Depletion of the Chromatin Looping Proteins CTCF and Cohesin Causes Chromatin Compaction: Insight into Chromatin Folding by Polymer Modelling.** *PLoS Comput Biol* 2014, **10**:e1003877.

46. * Barbieri M, Chotalia M, Fraser J, Lavitas L-M, Dostie J, Pombo A, Nicodemi M: **Complexity of chromatin folding is captured by the strings and binders switch model.** *Proc Natl Acad Sci USA* 2012, **109**:16173–16178.

47. Johnson J, Brackley CA, Cook PR, Marenduzzo D: **A simple model for DNA bridging proteins and bacterial or human genomes: bridging-induced attraction and genome compaction.** *J Phys Condens Matter* 2015, **27**:064119.

48. Brackley CA, Johnson J, Kelly S, Cook PR, Marenduzzo D: **Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains.** *Nucleic Acids Research* 2016, **44**:3503–3512.

49. * Marenduzzo D: **Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models**. *Genome Biol.* 2016, **17**:1–16.

50. Barbieri M, Xie SQ, Torlai Triglia E, Chiariello AM, Bianco S, de Santiago I, Branco MR, Rueda D, Nicodemi M, Pombo A: **Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells**. *Nat Struct Mol Biol* 2017, **24**:515–524.

51. Zuin J, Dixon JR, van der Reijden MIJA, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPC, van de Werken HJG, Knoch TA, van IJcken WFJ, et al.: **Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells.** *Proc Natl Acad Sci USA* 2014, **111**:996–1001.

52. de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, de Laat W: **CTCF Binding Polarity Determines Chromatin**

Looping. *Mol Cell* 2015, **60**:676–684.

53. Brackley CA, Johnson J, Michieletto D, Morozov AN, Nicodemi M, Cook PR, Marenduzzo D: **Nonequilibrium Chromosome Looping via Molecular Slip Links**. *Phys Rev Lett* 2017, **119**:138101.

54. Baù D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, Dekker J, Marti-Renom MA: **The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules**. *Nat Struct Mol Biol* 2011, **18**:107–114.

55. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L: **Genome architectures revealed by tethered chromosome conformation capture and population-based modeling**. *Nature Biotechnology* 2011, **30**:90–98.

56. Di Stefano M, Paulsen J, Lien TG, Hovig E, Micheletti C: **Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization**. *Sci. Rep.* 2016, **6**:35985.

57. * Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, Tiana G, Heard E: **Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription.** *Cell* 2014, **157**:950–963.

58. Jaynes ET: **Information Theory and Statistical Mechanics**. *Physical Review* 1957, **106**:620–630.

59. Pitera JW, Chodera JD: **On the use of experimental observations to bias simulated ensembles**. *Journal of Chemical Theory and Computation* 2012, **8**:3445–3451.

60. Cavalli A, Camilloni C, Vendruscolo M: **Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle.** *J. Chem. Phys.* 2013, **138**:094112.

61. White AD, Voth GA: **Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data.** *Journal of Chemical Theory and Computation* 2014, **10**:3023–3030.

62. Cesari A, Gil-Ley A, Bussi G: **Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement**. *Journal of Chemical Theory and Computation* 2016, **12**:6192–6200.

63. Mirny LA: **The fractal globule as a model of chromatin architecture in the cell**. *Chromosome Research* 2011, **19**:37–51.

64. Tiana G, Amitai A, Pollex T, Piolot T, Holcman D, Heard E, Giorgetti L: **Structural Fluctuations of the Chromatin Fiber within Topologically Associating Domains**. *Biophys. J.* 2016, **110**:1234–1245.

65. **Confinement-Induced Glassy Dynamics in a Model for Chromosome Organization**. *Phys Rev Lett* 2015, **115**:198102.

66. Tiana G, Villa F, Zhan Y, Capelli R, Paissoni C, Sormanni P, Heard E, Giorgetti L, Meloni R: **MonteGrappa: An iterative Monte Carlo program to optimize biomolecular potentials in**

simplified models. *Computer Physics Communications* 2014, **186**:93–104.

67. ** Norgaard AB, Ferkinghoff-Borg J, Lindorff-Larsen K: **Experimental parameterization of an energy function for the simulation of unfolded proteins**. *Biophys J* 2008, **94**:182–192.

The authors develop a computational algorithm based on thermodynamic sampling and on re-weighting of the sampled conformations to optimize efficiently the numerical parameters of maximum-entropy modles.

68. Tolhuis B, Palstra R-J, Splinter E, Grosveld F, de Laat W: **Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus**. *Mol Cell* 2002, **10**:1453–1465.

69. Zhan Y, Giorgetti L, Tiana G: **Modelling genome-wide topological associating domains in mouse embryonic stem cells**. *Chromosome Research* 2017, **25**:5–14.

70. * Zhang B, Wolynes PG: **Topology, structures, and energy landscapes of human chromosomes.** *Proc Natl Acad Sci USA* 2015, **112**:6062–6067.

71. Di Pierro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JN: **Transferable model for chromosome architecture.** *Proc Natl Acad Sci USA* 2016, **113**:12168–12173.

Figure captions

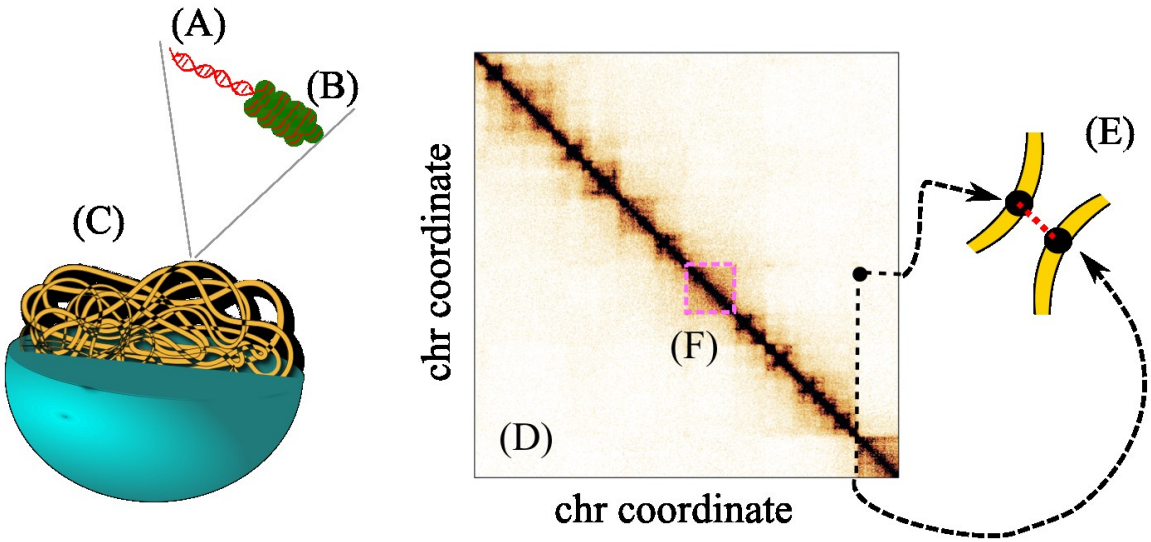

Figure 1: Chromatin is composed by the DNA double helix (A) wrapped around nucleosomes (B) into a fiber, which is folded into chromosomal territories in the nucleus of mammalian cells (C). The output of Chromosome Conformation Capture based experiments and notably Hi-C, which is shown here (D) in which every pixel encodes the probability that the two corresponding genomic loci are found in close proximity in the nuclear space (E), averaged over a population of millions of cells. Hi-C contact maps in vertebrate chromosomes display nested hierarchies of interactions domains F) corresponding to local preferential associations of the domains of the fiber

.

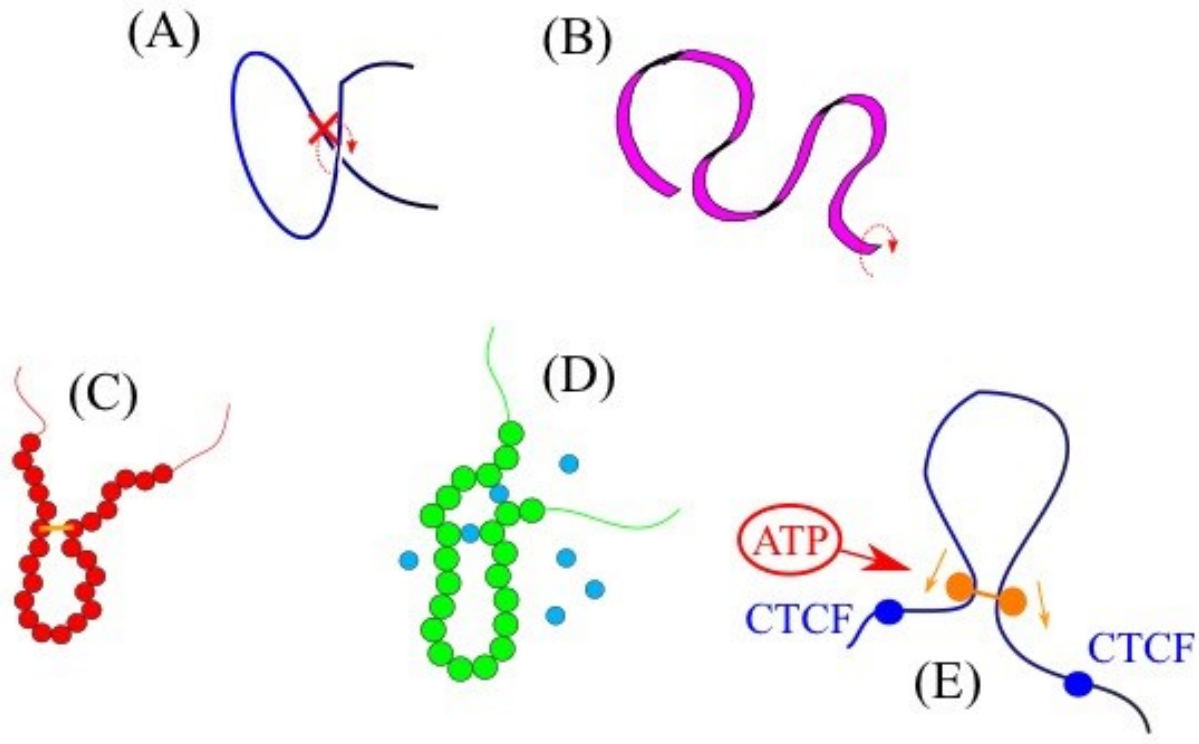

Figure 2: Hypotheses developed to describe the interactions governing the folding of the chromatin fiber include: (A) models based on topological constrains only (self-avoiding chains), (B) models based on supercoiling, (C) models based on direct short-range interactions between loci, (D) polymer models whose interaction are mediated by diffusing particles, (E) loop-extrusion models in which the effective interaction between two CTCF-bound loci is generated by ATP-driven molecular motors that extrude loops and are blocked by DNA-bound CTCF proteins.

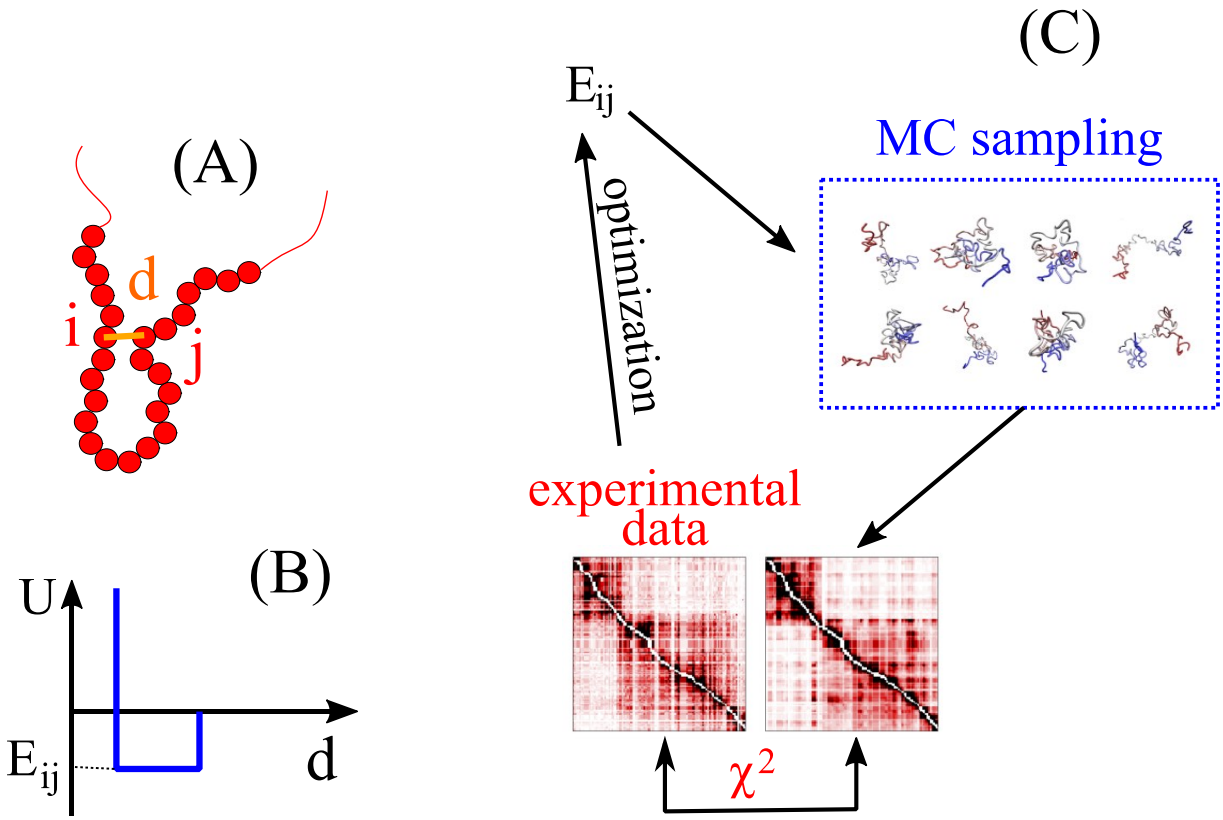

Figure 3: The implementation of the principle of maximum entropy for a coarse-grained model of chromatin requires a contact potential between each pair of beads (A-B), whose energy is obtained with an iterative Monte Carlo algorithm which minimizes the χ^2 between the predicted and the experimental data (C).