

Simulating topological domains in human chromosomes with a fitting-free model

C. A. Brackley,¹ D. Michieletto,¹ J. Johnson,¹ S. Kelly,² F. Mouvet,¹ P. R. Cook,³ and D. Marenduzzo¹

¹*SUPA, School of Physics & Astronomy, University of Edinburgh,
Mayfield Road, Edinburgh, EH9 3JZ, UK*

²*Department of Plant Sciences, University of Oxford,
South Parks Road, Oxford OX1 3RB, UK*

³*Sir William Dunn School of Pathology, University of Oxford,
South Parks Road, Oxford, OX1 3RE, UK*

Abstract

We discuss a simple polymer model for the 3D organization of human chromosomes. A chromosome is represented by a string of beads, with each bead being “colored” according to 1D bioinformatic data (i.e., chromatin state, GC content). Individual spheres (representing bi- and multi-valent transcription factors) can bind reversibly and selectively to beads with the appropriate color. During molecular dynamics simulations, the factors bind, and the string spontaneously folds into loops, rosettes, and topologically-associating domains (TADs) with the appropriate boundaries. Simulations also yield contact maps resembling those found experimentally using Hi-C. This organization occurs in the absence of any specified interactions between distant DNA segments, or between transcription factors. The model is “fitting-free” in the sense that it does not use Hi-C data as an input; consequently, one of its strengths is that it can – in principle – be used to predict the 3D organization of a chromosome in the absence of Hi-C data. We discuss how this simple model might be refined to include more transcription factors and binding sites, and to correctly predict contacts between convergent CTCF-binding sites.

Hi-C: contact maps, domains, and loops

The conformations adopted by human chromosomes in 3D nuclear space are key contributors to gene activity in health and disease [1], and understanding the principles driving genome folding is one primary goal of biophysicists studying DNA. An important recent experimental breakthrough has been the development of chromosome conformation capture (3C), and its high-throughput derivative – “Hi-C” – which allows contacts between different chromatin segments to be mapped genome-wide [2]. Contact maps obtained using Hi-C reflect some underlying chromosomal organization. For example, each chromosome folds into distinct “topologically-associating domains” (TADs) during interphase (but not mitosis when transcription ceases). Domain size is variable, with higher-resolution studies typically uncovering smaller TADs in the range between 0.1-2 Mbp. TADs are largely specified by the local chromatin environment, as the same 20-Mbp region in a chromosomal fragment or an intact chromosome yield similar contact maps [7]. This organization into TADs is conserved, as they are found in budding yeast [18] and *Caulobacter crescentus* (where they are called “chromosomal interaction domains” or CIDs) [19]. CIDs are also separated by strong promoters, and the bacterial ones are eliminated by inhibiting transcription.

Bioinformatic analysis suggests that mammalian TADs tend to be epigenetically determined; active and inactive regions typically form separate domains [2–7], with CTCF (the CCCTC-binding transcription factor) and active transcription units (or binding sites for RNA polymerase II) being enriched at inter-domain “boundaries” [3, 7]. These analyses also uncover chromosome loops apparently stabilized by transcription factors bound to promoters and enhancers [7, 9–14], and/or CTCF bound to its convergent cognate sites (presumably the latter loops are tethered by associated cohesin acting as a molecular “slip link”) [7].

While Hi-C data is normally obtained using cell populations, single-cell Hi-C experiments show that no two cells in the same population share exactly the same contacts; nevertheless, the organization is non-random as certain contacts are seen more often than others [17].

These observations point to central roles for transcription orchestrating the 3D organization, with transcription factors (e.g., CTCF) playing a central role in stabilizing the structure both locally and globally. Here we discuss results obtained using the simplest of

biophysical models – the binding of two transcription factors to cognate sites on DNA. As we will see, molecular dynamics simulations using this model yield contact maps remarkably similar to those obtained from Hi-C. We further discuss how this model can be extended to incorporate more transcription factors, and molecular “slip-links” like cohesin.

A toy model, and some basic principles

In this Section we introduce a toy model which is schematically described in Figure 1A: a chromatin fiber (a flexible bead-and-springs chain) interacts non-specifically with bi- or multi-valent spheres. These spheres represent transcription factors or complexes that can bind to two or more sites on the fiber; consequently they can form “molecular bridges” that stabilize loops. These factors stick to the chromatin fiber via a Lennard-Jones-like potential with depth E and range r_c . If the interaction strength is large enough to allow binding, and small enough to ensure the binding is reversible (so they behave like transcription factors), the bound proteins spontaneously cluster. This clustering is accompanied by the formation of rosettes, and “domains” in which intra-domain contacts are enriched. The (generic) principle underlying clustering – which occurs in the complete absence of any specified DNA-DNA or protein-protein interaction – has been called the “bridging-induced attraction” – as it does not occur with univalent proteins that cannot stabilize loops [22, 28?, 29].

The basic mechanism underlying this attraction is a simple (thermodynamic) positive feedback loop (Fig. 1B). First, proteins bind to chromatin, and – as they are bivalent – they can form a molecular bridge between two different DNA segments. This bridging brings distant parts of the chromosome together to increase the local chromatin concentration; this makes it more likely that additional proteins in the soluble pool will bind as they diffuse by. And once they have bound, the local high concentration of those proteins increases the chances that additional chromosomal segments will be trapped when they diffuse by. As this cycle repeats, protein clusters form and proteins or chromatin segments in these clusters rarely escape. Then, one or more such clusters organize a surrounding “TAD”. [We assume the protein concentration is

sufficiently low that proteins cannot uniformly cover the fiber even when all bind. If the concentration is large enough, bridging induces macroscopic collapse of the whole fiber [21, 24?].

In this simple case in which the transcription factors only bind non-specifically, the bridging-induced attraction yields clusters that continue to grow in size, ultimately giving one single cluster at the steady state [24]. However, most transcription factors also bind specifically, as well as non-specifically. A simple modification of the toy model includes a stronger specific binding (Fig. 1Ci). Clusters still form (Fig. 1Cii), but now they no longer grow indefinitely; instead, they reach a self-limiting size. This is because clustering of specifically-bound beads creates many chromatin loops, and bringing these together is entropically costly. Crucially, the entropic cost rises non-linearly with loop number, and this arrests cluster growth [22, 29, 40].

Another simple consequence of this generic organizing principle is that binding naturally creates “specialized” clusters. Imagine that two types of transcription factor (i.e., “red” and “green”) bind specifically to different beads on the fiber (i.e., pink and light green; Fig. 1D). As red and green beads are placed in different positions on the 1D chromosome map, the bridging-induced attraction then acts to increase the local concentrations of red factors plus pink beads in a different place in 3D space from the local concentrations of green factors and light-green beads. Consequently, the clusters that emerge tend to contain either red factors plus pink beads or green factors plus light-green beads. If red and green proteins represent complexes containing RNA polymerase II and III respectively, this naturally explains why distinct foci/“factories” are seen in human cells that contain one or other enzyme – but not both [30]. As discussed in the next Section, a similar mechanism probably underlies the organization of the “A/B” compartments uncovered in Hi-C experiments.

XXX Need here cartoon with (A) sketch with non-specific interaction; (B) clustering and coarsening; (C) sketch with non-specific and specific interaction; (D) clustering and self-limiting size XXX

A minimal, fitting-free, polymer model for chromosome folding

The toy model (Fig. 1) is easily extended to give a minimal model for genome organization. This is fitting-free because it is based solely on 1D information on the protein binding landscape. Thus, unlike more commonly-used approaches, it does not rely on contact information as an input, so predictive power is enhanced. In one simple version [29], the whole of chromosome 19 in GM12878 cells was modeled (Fig. 2A). In this case, each chromatin bead contained 3 kbp, and factors were of two types – “active” (modeling complexes of polymerases and transcription factors) or “inactive” (modeling heterochromatin-associated proteins like HP1 α , or even a simple linker histone like H1 – as both proteins are known to bind the genome in multiple places). Beads in the chromatin fiber are “colored” according to bioinformatic data to specify whether they bind the active or inactive proteins. Thus, active beads were colored using the “active” Broad ChromHMM tracks [34] on the hg19 assembly (i.e., using states 1,4,5 in the HMM track that signify an “Active Promoter” or “Strong Enhancer” to specify strong binding, and states 9 and 10 that signify “Transcriptional Transition” or “Transcriptional Elongation” to specify weak binding). Inactive beads were colored using either the appropriate HMM tracks or GC content – the latter is illustrated here as a low GC content is such a good predictor of an inactive (heterochromatic) nature.

Given the simplicity of this model, it is striking to see how well it allows correct prediction of the positions of TADs and their boundaries (Fig. 2C,D). For example, 85% boundaries are correctly identified to within 100 kbp; some inter-domain interactions are even correctly captured (see the off-diagonal blocks in the contact maps). While this agreement can certainly be improved by adding biological detail, we stress that it is especially remarkable as it appears in a fitting-free minimal model (the only relevant parameters are interaction strengths and cut-offs). The model can be applied, in principle, to any chromosome for which appropriate bioinformatic data is available (e.g., Broad ChromHMM track, histone modifications [45]); consequently, it can be used genome-wide in different cell lines and organisms. It can also be used to predict the contact map of any region of interest, and – of course – it can be applied at a higher resolution [45].

As in the toy model, active and inactive factors (and their cognate binding sites) cluster separately, and the model naturally yields the A (active) and B compartments seen in Hi-C contact maps. Moreover, the proteins cluster to give structures reminiscent of both nuclear “bodies” (e.g., Cajal, polycomb and promyelocytic leukemia bodies), and factories containing RNA polymerases II and III – all structures rich in distinct proteins binding to different DNA sequences [35–38]. The number of clusters is significantly smaller than that of chromatin domains: therefore our model predicts that a number of clusters will constitute a TAD.

As these simulations reproduce the overall Hi-C organization so well, it is of interest to ask what is special about beads at, or close to, boundaries between TADs. Figure 3 shows that the boundary beads *in silico* are depleted #-fold of inactive beads: this is consistent with bioinformatic analyses showing that boundaries are enriched #-fold in active marks and depleted of heterochromatic ones like HK39me3 and K3K27me3 [3]. An intriguing additional signal is that beads enriched at boundaries *in silico* are non-binding beads – which naturally form boundaries as they possess few contacts; this is consistent with ~25% Hi-C boundaries lacking any particular mark [3].

Fig. 3 analysis of simulated boundaries in chr19, Suppl. Fig. on NAR paper

Beyond the minimal model: adding colors and “slip-links”

The minimal models described thus far generally yield contact maps like those obtained from Hi-C data; however, exceptions do exist. In general, the percentage of TAD boundaries predicted accurately increases with transcriptional activity (the organization of chromosome 19 is predicted well, perhaps because it is the one containing the most active genes). In less-active regions, boundaries are predicted less accurately (e.g., Fig. 4A), and this makes it attractive to improve the “coloring” of inactive beads, and/or add more colors. Capture Hi-C results provide a way of adding more colors. Thus, Mifsud *et al.* [14] distinguished contacts between promoters on the basis of their histone marks. H3K27me3 is a classic inactive mark associated with facultative heterochromatin and binding of polycomb-group repressing complexes; it

marks “blue chromatin” in *Drosophila* [?]). Therefore, we stipulated that heterochromatic beads were classified according to histone modifications (instead of GC content), with two different colors for beads bearing the H3K9me3 or H3K27me3 mark [47]; we then also included in the simulations two proteins binding to these marks (modeling PcG-protein complexes, such as PRC1 binding to H3K27me3 marks [? ?]). Once different kinds of heterochromatic beads are specified, simulations predict TAD locations more accurately. We stress that this refined model is still fitting free as it does not rely on Hi-C data for input, but only assumes knowledge of the 1D protein-binding landscape, or histone-modification profiles.

Another (fitting-free) model similar in spirit to the one presented here is the “block-copolymer” model used to study folding of *Drosophila* chromosomes [23]. [For a non-fitting free version, see [30].] In this case, chromatin beads interact directly, so bridging proteins are implied but not explicitly modeled. This approach is equivalent to the one used in Figures 2-4 if bridging proteins are abundant enough to saturate binding sites; however, the two models differ in the regime where only some binding sites are occupied. The model used in Figures 2-4 also naturally explains the formation of nuclear bodies, and so can be used to study their biogenesis and kinetics (this is not possible with the block-copolymer model where bridges are assumed to be uniformly distributed at all times).

A recent study by Chiariello *et al.* [46] offers another avenue to improve simulation accuracy by using some information from Hi-C experiments (but then the model is no longer fitting free). In practice this is done through an iterative procedure which finds the minimal arrangement of binding sites and colors which best explain the Hi-C contact map; for example, simulations involving 16 colors gave contact maps for the Sox6 locus that were indistinguishable from those obtained by Hi-C.

An important unaddressed aspect concerns loops (or “loop domains”) stabilized by CTCF [7]. As discussed above, CTCF is more likely to bridge two cognate binding sites [7, 41] when sites are in a “convergent” orientation compared to a “divergent” one. Polymer models to explain this have been proposed [33?]; they involve loop-extrusion factors and “slip-links” that are simultaneously bound (linked) to beads on two different chromosomal segments and which can slide (slip) along the segments (in practice, these factors/slip-rings are cohesin and/or condensin). These models can account for the observed CTCF orientation bias, as they assume that the loop-extrusion factors

can stably stick only to one side of CTCF (which is true of cohesion). However, these models also require some as-yet undiscovered motor protein with a processivity sufficient to generate loops of hundreds of kb. Moreover, CTCF and its convergent sites cannot be the sole organizer of boundaries, as knock-outs of CTCF have only minor effects on domain organization in mammals [43, 44], and bacteria possess domains but no equivalent of CTCF. Nevertheless, it will be of interest to incorporate binding of CTCF to convergent sites into our model.

I. ACKNOWLEDGEMENTS

CAB, DMi and DMA acknowledge support from ERC CoG 648050 (THREEDCELL-PHYSICS).

- [1] Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290-299.
- [2] Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293.
- [3] Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M. Liu, J. S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380.
- [4] Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blthgen, N., Dekker, J., and Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385.
- [5] Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472.
- [6] Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. *Science* **342**, 948-953.
- [7] Rao, S.S.P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S. and Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1-16.
- [8] Sexton, T., and Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049-1059.
- [9] Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel,

B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains

- uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348-1354.
- [10] Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J. *et al.* (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98.
- [11] Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C. A., Schmitt, A. D., Espinoza, C. A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294.
- [12] Zhang, Y., Wong, C. H., Birnbaum, R. Y., Li, G. L., Favaro, R., Ngan, C .Y., Lim, J., Tai, E., Poh, H. M., Wong, E. *et al.* (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **503**, 290-294.
- [13] Heidari, N., Phanstiel, D. H., He, C., Grubert, F., Jahanbani, F., Kasowski, M., Zhang, M. Q., and Snyder, M. P. (2014). Genome-wide map of regulatory interactions in the human genome. *Genome Res.* **24**, 1905-1917.
- [14] Mifsud, B., Tavares-Cadete, F., Young, A. N, Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A. *et al.* (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Gen.* **47**, 598-606.
- [15] Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64.
- [16] Hsieh, T.-H., S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* **162**, 108-119.
- [17] Le, T.B., Imakaev, M.V., Mirny, L.A., and Laub, M.T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731-734.
- [18] Nicodemi, M., Panning, B., and Prisco, A. (2008). The colocalization transition of homologous chromosomes at meiosis. *Phys. Rev. E* **77**, 061913.
- [19] Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L.M., Dostie, J., Pombo, A., and Nicodemi, M. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. USA* **109**, 16173-16178.
- [20] Brackley, C. A, Taylor, S., Papantonis, A., Cook, P. R., and Marenduzzo, D. (2013). Non-specific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. USA* **110**, E3605-3611.

- [21] Jost, D., Carrivain, P., Cavalli, G., and Vaillant, C. (2014). Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* **42**, 9553-9561.
- [22] Johnson, J., Brackley, C.A., Cook, P.R., and Marenduzzo, D. (2015). A simple model for DNA bridging proteins and bacterial or human genomes: bridging-induced attraction and genome compaction. *J. Phys. Condens. Matter* **27**, 064119.
- [23] Le Treut, G., Kepes, F., Orland, H. (2016). Phase Behavior of DNA in the Presence of DNA-Binding Proteins *Biophys. J.* **110**, 51-62 (2016).
- [24] Brackley, C. A., Johnson, J., Kelly, S., Cook, P. R. and Marenduzzo, D. (2016). Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucl. Acids Res.* **44**, 3503-3512.
- [25] Giorgetti, L., Galupa, R., Nora, E.P., Pilot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950-963.
- [26] Kilic, S., Bachmann, A.L., Bryan, L.C., and Fierz, B. (2015). Multivalency governs HP1 α association dynamics with the silent chromatin state. *Nat. Commun.* **6**, 7313.
- [27] Sanborn, A. L., Rao, S. S. P., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochlov, I. D., Chinappan, D., Cutkosky, A., Li, J. *et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes, *Proc. Natl. Acad. Sci. USA* **112**, E6456-E6465.
- [28] Ernst, J., Kheradpour, P., Mikkelson, T. S., Shores, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M. *et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49.
- [29] Sleeman, J.E., and Trinkle-Mulcahy, L. (2014). Nuclear bodies: new insights into assembly/dynamics and disease relevance. *Curr. Opin. Cell Biol.* **28**, 76-83.
- [30] Pombo, A., Jackson, D.A., Hollinshead, M., Wang, Z., Roeder, R.G., and Cook, P.R. (1999). Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* **18**, 2241-2253.
- [31] Cook, P.R. (1999). The organization of replication and transcription. *Science* **284**, 1790-1795.
- [32] Papantonis, A., and Cook, P.R. (2013). Transcription factories; genome organization and gene regulation. *Chem. Rev.* **113**, 8683-8705.

- [33] Marenduzzo, D. and Orlandini, E. (2009). Topological and entropic repulsion in biopolymers. *J. Stat. Mech.*, L09002.
- [34] Wani, A. H., Boettiger, A.-N., Schorderet, P., Ergun, A., Munger, C., Sadreyev, R. I., Zhuang, X., Kingston, R. E., Francis, N. J. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Comm.* **7**, 10291.
- [35] Michieletto, D., Marenduzzo, D., Wani, A. H. (2016) Chromosome-wide simulations uncover folding pathway and 3D organization of interphase chromosomes. arXiv:1604.03041.
- [36] Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J. H., Gorkin, D. U., Jung, I., Wu, H. Y., Zhai, Y. N., Tang, Y. X. *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*. **162**, 900-910.
- [37] Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. (2016) Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**, 2038-2049.
- [38] Zuin, J., Dixon, J. R., van der Reijden, M. I. J. A., Ye, Z., Kolovos, P., Brouwer, R. W. W., van de Corput, M. P. C., van de Werken, H. J. G., Knoch, T. A., van IJcken, W. F. J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* **111**, 996-1001 (2014).
- [39] Hou, C., Dale, R., Dean, A. (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. USA* **107**, 3651-3656.
- [40] Brackley, C. A., Brown, J. M., Waithe, D., Babbs, C., Davies, J., Hughes, J. R., Buckle, V. J., Marenduzzo, D. (2016). Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Gen. Biol.* **17**, 59.
- [41] Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A., Nicodemi, M. (2016). Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 29775.
- [42] In practice, we used a threshold in histone modification tracks to color beads, but the exact value of the threshold played a minor role in the results.

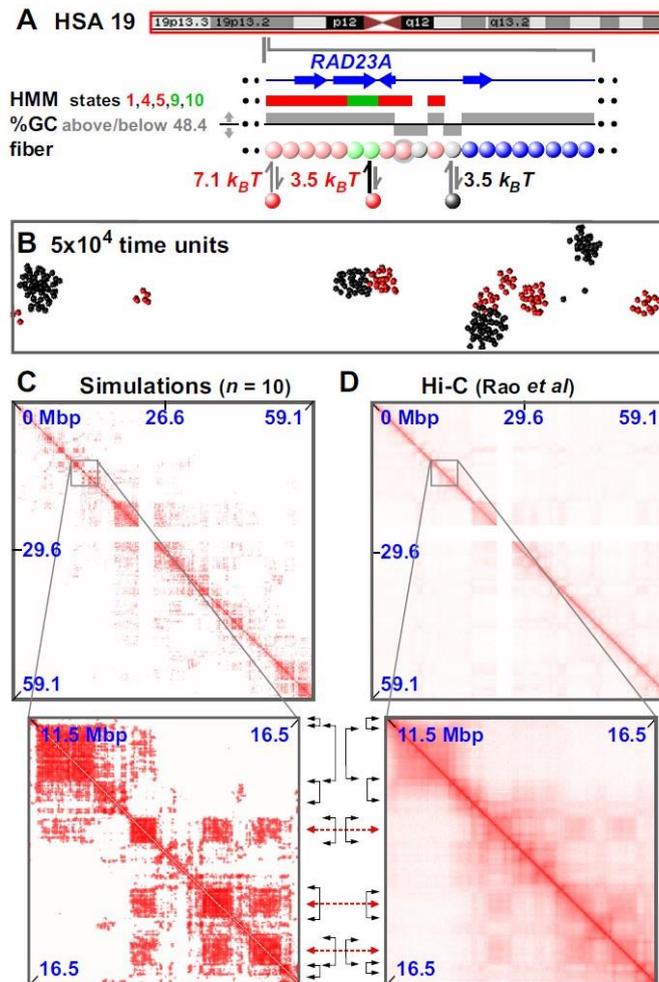


FIG. 1: **Fitting-free simulations of chromosome 19 in GM12878 cells.** (A) Overview. The ideogram (red box indicates whole chromosome simulated) and Broad HMM track (colored regions reflect chromatin states) are from the UCSC browser; the zoom illustrates an arbitrary region, around RAD23A, to show the details of the “coloring”. Beads (3 kbp) are colored according to HMM state and GC content: blue beads are non-binding; pink beads correspond to states 1+4+5 in the ChromHMM track; light-green to states 9+10. Grey beads correspond to beads which have <48.4% GC. Pink and light-green beads bind active factors (red in the figure); grey beads bind to inactive factors, linked to heterochromatinization (black in the figure). Note that the coloring rule is such that beads can have multiple colors: for instance, in the zoom two pink beads are also grey (represented by grey halos), so that such beads can bind both red and black factors. (B) Snapshot (without chromatin) of central region after 5×10^4 units; most clusters contain factors (or proteins) of one color. In other words, active and inactive proteins cluster separately. As discussed in the text, the formation of specialised clusters may underlie both the formation of A/B compartments (when looking at the chromatin pattern) and that of some nuclear bodies (when looking at the protein cluster patterns). (C,D) Comparison between contact map patterns found in simulations and experiments (see Ref. [29] for more details). Between zooms, black double-headed arrows mark boundaries of prominent domains (on the diagonal), and red double-headed ones the centers of off-diagonal blocks making many inter-domain contacts. Reproduced from Ref. [29], with permission.

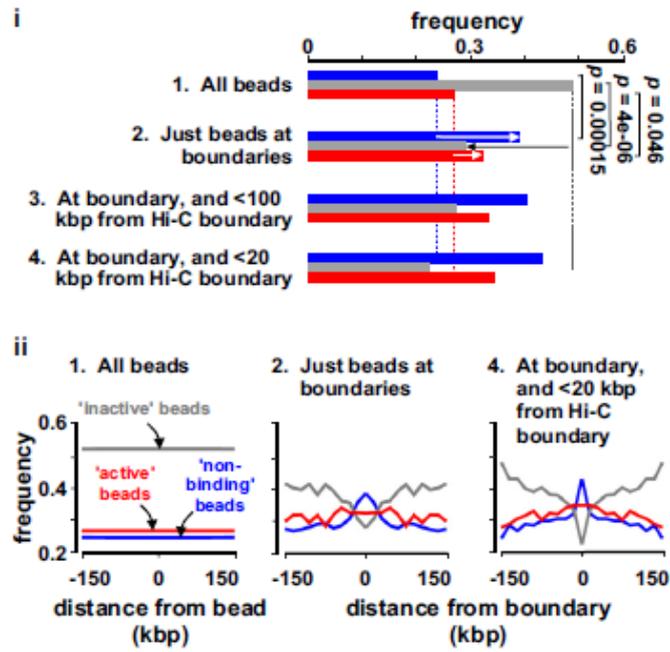


FIG. 2: **Characterization of TAD boundaries found *in silico*.** Adapted from Ref. [29].

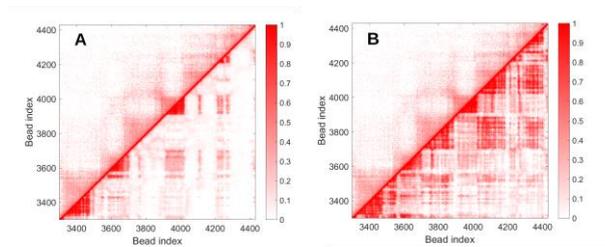


FIG. 3: **Adding “colors” leads to better agreement in some chromosome regions.** XXX chr14, add details of simulations and regions simulated (15 Mbp XXX). (A): no polycomb “coloring”. (B) with polycomb “coloring”.