COMMENTARY

# Predictive model of 3D domain formation via CTCF-mediated extrusion

**Galip Gürkan Yardımcı[a] and William Stafford Noble[b,1]**

[a]*Department of Genome Sciences, University of Washington, Seattle, WA 98195; and*
[b]*Departments of Genome Sciences and Computer Science and Engineering, University of Washington, Seattle, WA 98195*

The extraordinary compaction of DNA in vivo, 2 m of DNA packed into a nucleus that is six orders of magnitude smaller, presents a conundrum: How can the cell maintain this highly dense chromatin structure while also carrying out exquisitely regulated processes like gene expression, DNA replication, and DNA repair? Over the past decade, we have begun to understand the fine structure of chromatin better and to unravel its previously underappreciated role in various functional processes. These insights have been driven largely by the development of high-throughput experimental methods, including Hi-C (1) and chromatin interaction analysis by paired-end tag sequencing (2), that measure the 3D proximity of pairs of DNA loci by cross-linking, ligating, and sequencing the DNA (Fig. 1*A*). Using such methods, researchers have profiled the 3D conformation of chromatin in vivo in a wide variety of species and human cell types. As experimental methods improve and sequencing costs drop, the effective resolution of this type of study has improved from 500 kb in 2009 down to 1 kb, achieved by Rao et al. in 2014 (3). In PNAS, Sanborn et al. (4), who are members of the same research team, describe an extensive reanalysis of the datasets of Rao et al. (3).

In the recently published study, Sanborn et al. (4) investigate the relationship between genomic distance and contact probability (i.e., the frequency with which two genomic loci are linked by paired-end reads in a Hi-C experiment). Inconsistencies between their observations and previously proposed chromatin models lead Sanborn et al. (4) to propose a model, the "extrusion model," that is consistent with known mechanisms of chromatin organization, can accurately simulate Hi-C measurements, and makes predictions that are validated by KO experiments.

Sanborn et al. (4) begin by calling into question the prevalence of the 30-nm fiber, a higher order solenoidal structure composed of DNA wrapped around nucleosomes. This structure has been observed in vitro, but its existence and prevalence in vivo are controversial (5). Sanborn et al. (4) measured the length distribution of cyclical Hi-C fragments that result from self-ligation of chromatin, observing a modal length of ∼1 kb. This length is much lower than the predicted modal length for the 30-nm fiber, for which cycles form at 30–60 kb. Similarly, the contact probability of a 30-nm fiber should reach a maximum around 30–60 kb, but Sanborn et al. (4) observe a decreasing trend in contact probability for all values >5 kb. Because the Hi-C protocol modifies native chromatin, we still cannot rule out the existence of the 30-nm fiber in vivo, but these findings imply that the 30-nm fiber is rare or nonexistent in the chromatin assayed by a Hi-C experiment.

The relationship between genomic distance and contact probability can also be used to describe more general properties of the overall conformation of chromatin in the cell. Generally, the contact probability decreases with increasing distance and is observed empirically to follow a power law distribution, where the probability of contact between two loci separated by $d$ bases is approximated by $d^{-\gamma}$. The observed value of the exponent $\gamma$ is critical, because theoretical analysis, supported by simulation studies, suggests that distinct types of conformations cause the exponent value to lie in specific numerical intervals (6). In the original Hi-C paper, Lieberman-Aiden et al. (1) pointed out that values between 1 and 1.5 correspond to a model called the "fractal globule," where sequential regions of the genome occupy distinct regions in space. This model agrees with observed contact probability distributions derived from Hi-C experiments at ranges from 500 kb to 7 Mb, and is consistent with a domain-like organization of chromatin.

In the recently published study, Sanborn et al. (4) point out that the fractal globule model breaks down when we zoom in on finer scale behavior using high-resolution data. Strikingly, they observed a markedly smaller exponent ($\gamma \approx 0.75$) for proximal pairs of loci, suggesting that the fractal globule model is inappropriate at this scale. The authors then

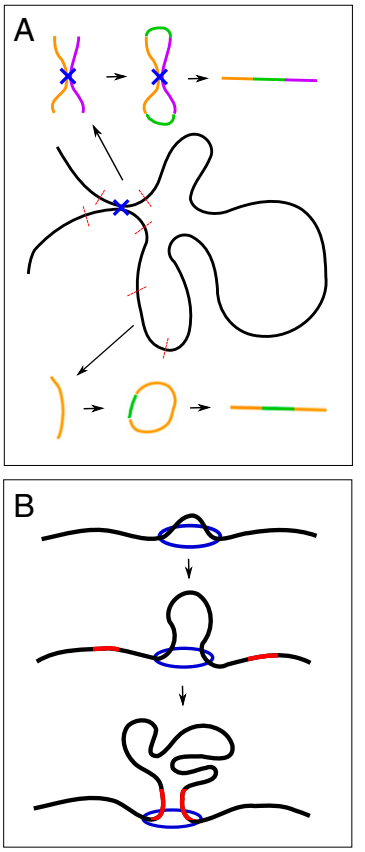


**Fig. 1.** (*A*) Hi-C protocol: A representative DNA loop with six restriction sites is shown as dashed lines, and a cross-linking event is shown with a blue cross. In a typical Hi-C experiment, two proximal regions of DNA are cross-linked, proximal DNA is cleaved, and ends are ligated to each other. The resulting circular fragment is sheared, sequenced from both ends, and mapped back to the genome. Sometimes, a single DNA fragment can fold back and ligate itself, resulting in cyclical Hi-C fragments. (*B*) Extrusion model: DNA loops are formed by an extrusion complex (blue circle) binding to a locus and beginning to extrude DNA. The complex moves in both directions, extruding DNA until it encounters two CTCF motifs (red). The extruded DNA forms a loop enriched for contacts.

Author contributions: G.G.Y. and W.S.N. wrote the paper.

The authors declare no conflict of interest.

See companion article on page E6456.

[1]To whom correspondence should be addressed. Email: william-noble@uw.edu.

leverage their previous observation that chromatin is organized into so-called "contact domains," contiguous regions of densely interacting DNA, to segregate pairs of loci into within-domain and between-domain pairs. This analysis shows that not only does the exponent γ change at small genomic distances but the "hinge" in the contact probability function occurs precisely at the boundary of the contact domain. Furthermore, the authors (4) note that the contact probability distributions are similar across different cell types from mouse and human and across the two previously described chromatin compartments (A and B) (1). These observations, together with a mathematical proof that the fractal globule model cannot produce the observed exponent of 0.75, suggest the need for two distinct chromatin models: the fractal globule for large-scale behavior and some other model for smaller scale, within-domain behavior.

Sanborn et al. (4) propose two potential models for within-domain behavior. The first model, called the "tension globule," can accurately recreate experimentally observed contact probability distributions but fails to explain various important features of contact domains, such as DNA loops. The tension globule is therefore superseded by the extrusion model, which is the primary contribution of the article.

The extrusion model draws upon our understanding of two widely studied protein complexes, CCCTC-binding factor (CTCF) and cohesin, that act jointly to organize chromatin (7). These factors are known to tether distal regions of the genome, thus forming DNA loops. Sanborn et al. (4) describe an extrusion complex, presumed to be a complex of CTCF and cohesin, that first binds to a locus and extrudes chromatin as the complex moves along the DNA in both directions (Fig. 1). When the complex overlaps with two CTCF-binding elements, CTCF proteins bind to the DNA and dimerize, thus creating a stable loop with CTCF-binding sites serving as anchors and CTCF dimers acting as a tether. Consistent with previous reports (3), this looping mechanism requires a convergent orientation of the two CTCF motifs.

Sanborn et al. (4) first demonstrate the utility of the extrusion model via simulation. Focusing on a 2.3-Mb region on chromosome 4, and using ChIP-sequencing data to identify CTCF sites, the simulation produces a contact map that strongly agrees with experimental data (Pearson's $r = 0.964$). Previously, predictive models have been built for the yeast genome, achieving similar levels of correlation between simulated and observed Hi-C data (8).

Next, Sanborn et al. (4) demonstrate the predictive power of their model using KO experiments. If the extrusion complex forms DNA loops via CTCF binding and dimerization, then elimination of CTCF activity should disrupt the formation of DNA loops. To test this hypothesis, Sanborn et al. (4) ran simulations on three loci, each with three DNA loops, and found that manipulation of CTCF elements away from the ideal configuration results in the loss of DNA loops. These results were confirmed by genome editing experiments, in which the CTCF sites were eliminated using the clustered regularly interspaced short palindromic repeat/Cas9 system. Capture Hi-C experiments that specifically target the loci of interest show that experimental data repeatedly matched the simulations. Interestingly, for certain configurations of CTCF-binding sites, experiments and simulations both show that contact domains can still form when a single CTCF anchor is deleted. The authors (4) propose that the activity of an intact extrusion complex can halt the activity of a neighboring extrusion complex that lacks an anchor. On the other hand, if the anchors for both extrusion complexes are removed, then the two separate contact domains merge. Overall, these observations support the extrusion model to explain the existence and formation of contact domains.

One of the grand challenges in the study of genome 3D architecture involves building a mechanistic model to explain our increasingly detailed picture of chromatin conformation at various scales. Beneath the large-scale structure defined by chromosome territories lie a confusing variety of domains of various types. Some types of domains, like isochores or heterochromatin and euchromatin, are primarily defined as contiguous genomic regions. More recently, driven by Hi-C and related assays, 3D domains have been proposed at various scales: A/B compartments, topologically associating domains, and contact domains. Largely lacking from these descriptions, however, has been a mechanistic model that explains the formation and maintenance of these various types of domains. The work of Sanborn et al. (4) provides a compelling model to fill this gap at the level of loop formation and contact domains.

A particularly striking feature of the model is the central role of CTCF. Initially considered a transcription factor, CTCF has been widely studied and implicated in a variety of biological processes, including gene regulation, alternative splicing, and the spreading of heterochromatin (7). It remains to be seen how these various CTCF functional roles relate to one another, and whether CTCF's central role lies in regulating chromosome conformation.

Moving forward, a key question is whether we can successfully link the multikilobase scale model of contact domains proposed here with the central gene regulatory model, which operates at the scale of nucleosomes, DNaseI hypersensitive sites, promoters, and transcription factor binding. Bridging this gap with an accurate, predictive model would help to establish the extent and mechanism relating genome 3D architecture to gene regulation.


**1** Lieberman-Aiden E, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289–293.

**2** Fullwood MJ, et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462(7269):58–64.

**3** Rao SS, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665–1680.

**4** Sanborn AL, 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.

**5** Maeshima K, Hihara S, Eltsov M (2010) Chromatin structure: Does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol* 22(3): 291–297.

**6** Mirny LA (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res* 19(1):37–51.

**7** Ong CT, Corces VG (2014) CTCF: An architectural protein bridging genome topology and function. *Nat Rev Genet* 15(4):234–246.

**8** Tjong H, Gong K, Chen L, Alber F (2012) Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res* 22(7):1295–1305.