Edinburgh Research Explorer

Predicting genome organisation and function with mechanistic modelling

Predicting genome organisation and function with mechanistic modelling

Michael Chiang,[1] Chris A. Brackley,[1] Davide Marenduzzo,[1] and Nick Gilbert[2,*]

[1]SUPA, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK

[2]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh, EH4 2XU, UK

*Correspondence: nick.gilbert@ed.ac.uk (N. Gilbert).

Fitting-free mechanistic models based on polymer simulations predict chromatin folding in 3D by focussing on the underlying biophysical mechanisms. This class of models has been increasingly used in conjunction with experiments to study the spatial organisation of eukaryotic chromosomes. Feedback from experiments to models leads to successive model refinement and has previously led to the discovery of new principles for genome organisation. Here, we review the basis of mechanistic polymer simulations, explain some of the more recent approaches and the contexts in which they have been useful to explain chromosome biology, and speculate on how they might be used in the future.

Keywords: genome organisation; chromatin modelling; mechanistic models; polymer physics.

Highlights

Mechanistic models provide testable hypotheses on principles of genome folding. It is common to start from a basic model and gradually introduce more ingredients to account for more experimental findings.

In the transcription factor (TF) model, multivalent chromatin-binding proteins cluster through positive feedback, resulting in phase separation. This 'bridging-induced attraction' explains the biogenesis of nuclear bodies and the formation of active and inactive chromosome compartments. In the loop extrusion (LE) model, structural maintenance of chromosomes (SMC) proteins drive the growth of chromatin loops. LE explains the formation of topologically associating domains and the bias favouring convergent CCCTC-binding factor (CTCF) loops.

The highly predictive heteromorphic polymer (HiP-HoP) model combines the TF and LE models and includes chromatin heteromorphicity. It can be used to predict 3D chromatin structure genome-wide.

Spatial organisation of eukaryotic genomes

The 3D organisation of eukaryotic genomes is complex. Individual chromosomes are folded in a way that enables them to fit within the nucleus while allowing essential tasks, such as transcription and replication, to be performed. A longstanding challenge in the field of genome biology is to uncover the principles governing the 3D folding of chromatin and, more importantly, to determine how folding is related to genome function and disease [1–5].

The spatial interactions between chromatin loci have been interrogated by both sequencing [e.g., **chromosome conformation capture** (3C; see glossary) and other ligation-free protocols] and microscopy-based techniques [e.g., **fluorescence *in situ* hybridisation** (FISH); see [6] for a detailed review of methods mapping 3D genome structure]. These methods show that the genome is organised in a hierarchical manner [1,6] and, on a large scale, individual chromosomes occupy distinct territories within the nucleus [7]. Early **Hi-C** experiments revealed that, within a territory, at the tens of megabase-pair (Mbp) level, the genome is segregated into two major **compartments**: one associated with transcriptionally active regions (A compartment) and the other with inactive regions (B compartment) [8]. Higher resolution Hi-C data showed that, within compartments, there are smaller chromatin regions, spanning hundreds of kilobase-pairs (kbp), called **topologically associating domains** (TADs), in which interactions are more favoured within than between domains [9,10]. Many TADs are associated with chromatin loops that are mediated by **CCCTC-binding factor** (CTCF) [11] and **promoter**–**enhancer** interactions [12].

The rapid advancement in experimental technologies has allowed us to gather a wealth of information on genome architecture, and theoretical modelling has been a powerful, complementary tool in enhancing our understanding. Modelling is important for visualising and interpreting the increasingly complex data generated from experiments; more potently, it can provide testable hypotheses on mechanisms regulating genome folding, which in turn can drive further experimental work. Additionally, modelling helps reconcile results from different experimental techniques. For instance, many 3C-based methods give population-averaged information, while super-resolution microscopy provides data on individual cells. Modelling can be used to generate an ensemble of structures to explain features found in both single cell- and population-based experiments.

In this review, we discuss the main concepts in theoretical modelling for chromatin folding. We focus specifically on mechanistic models, which predict structure without utilising 3D chromatin interaction information from experiments (e.g., Hi-C) [13]. We summarise recent progress made with this class of models in identifying key principles that shape nuclear structure. Finally, we highlight how a genome-wide modelling approach may be a promising direction to unravel the link between structure and function.

Polymer modelling of chromatin

To model the 3D folding of chromatin, it is useful to utilise a framework based on polymer physics. This is because chromatin is a long fibre, formed of hundreds of millions of DNA base pairs wrapped around histone complexes to form an array of nucleosomes. Here, we discuss some key considerations when constructing polymer models for chromatin.

Coarse-grained modelling of chromatin

A crucial part of setting up a polymer model is choosing an appropriate level of description for chromatin. This largely depends on the research question and the available experimental data. For instance, if one is interested in the effect of individual nucleotides or nucleosomal structure, an atomistic approach covering all base pairs, histones and other proteins, and solvent molecules may be necessary. Such high-resolution simulations can be precise, but they are computationally very expensive and only feasible for a small segment of chromatin over short timescales.

For investigating the large-scale arrangement of chromatin, it is more practical to remove some details, a process known as **coarse graining** (CG; Figure 1A) [14]. This approach is grounded in theories from polymer physics, which suggest that the macroscopic behaviour of a polymer (chromatin) does not depend on the specific details of its individual constituents (e.g., nucleotides) [15]. A common scheme is then to describe chromatin as a chain of beads, in which each bead represents several nucleosomes (see Box 1 for other CG schemes for modelling DNA and chromatin); consequently, the exact structure of chromatin within a bead is not resolved. The amount of chromatin covered by a single bead can be chosen depending on the size of the system to be modelled and the molecular resolution required. Simulations examining promoter–enhancer loops at specific loci may consider higher resolutions, whereas chromosome-wide simulations may choose lower resolutions for computational efficiency reasons.

Inverse modelling

Over the past decade, there have been two general directions in CG modelling of chromatin (see [16] for a comprehensive review). One direction, referred as the inverse (or 'top-down' or data-driven) approach, utilises 3C-based data (e.g., 5C or Hi-C) as an input to reconstruct the spatial structure of chromatin (Figure 1B). In some early models, constraints for the separation between chromatin loci were inferred from a 3C interaction matrix, and a single 'averaged' conformation that best satisfied the constraints was computed [17]. Recognising the ensemble nature of conventional Hi-C data, other work considered sampling multiple conformations based on the same set of constraints but starting from different initial conditions [18–20].

There were also studies that attempted to generate and iteratively improve a family of structures that, when measured collectively, reproduced the experimental data [21,22]. More recently, inverse models have been first trained against existing data in a well-studied condition, and then used to estimate the 3D conformations in an alternative condition for which there may be less information available, such as when the genome undergoes mutations or rearrangements [23]; however, this necessitates that assumptions present in the training set also exist in the alternative state.

Mechanistic modelling

The other direction in chromatin modelling, known as the mechanistic (or 'bottom-up' or first-principles) approach, focusses on understanding the physical and molecular mechanisms that give rise to prominent features of genome architecture [13]. Fundamentally, this approach starts from a minimal representation of chromatin (i.e., a bead-and-spring polymer) and explores possible microscopic mechanisms that can explain the folding features. In practice, the model input is a set of 'rules' based on known biophysical processes or hypotheses inspired from empirical observations (Figure 1C). For example, these could describe how chromatin loci interact with one another and with other nuclear landmarks; they could also model the effects of nuclear confinement. Experimental data, for example on histone modifications, may be incorporated, but, unlike the inverse approach, no fitting or training (e.g., to 3C-based data) is required. The result is then a set of 3D structures and a simulated interaction matrix, while some models additionally yield results on chromatin and protein dynamics. Models of this class are generally realised by performing molecular dynamics simulations (see [13] for more details) or Monte Carlo sampling.

Mechanistic models are powerful in several respects. First, because these models do not rely on structural data from experiments, they are truly predictive: the generated 3D conformations are purely based on the rules specified. Second, these models are parsimonious, and simulations can be completed relatively quickly. This enables an agile approach in assessing a range of potentially interesting hypotheses: they can first be studied by simulations, and only promising ones are then validated by experiments, which often take significantly longer to conduct. Third, these models typically contain only a handful of parameters. This allows more comprehensive sweeping of the parameter space, thus, a more precise mapping of parameter values to biological conditions.

The inverse and the mechanistic approaches are useful in different research contexts (Table 1). Inverse modelling is apt for situations in which data are abundant, whereas mechanistic modelling can offer predictions when experimental resources are limited or experiments are difficult to undertake (e.g., rare cell types). The two approaches are also not mutually exclusive. Indeed, models that combine elements from

both approaches have been developed, for example to examine structural variants at different genomic loci [23,24].

Successive refinement of mechanistic models

Within the bottom-up framework, it is common to start from a basic model and gradually introduce more ingredients to account for additional experimental findings (Figure 2, Key figure). Often, one gains new knowledge of mechanisms regulating genome architecture when model results deviate systematically from theory or experiment. Such situations signify that some information or rules are missing from the model; new hypotheses are then proposed and subsequently tested by further simulation and experimentation. Eventually, new model ingredients that account for the discrepancy can be incorporated into the model, and they become new principles learnt about chromosome folding. We recount here several examples in which the successive improvement of mechanistic models in this manner has been instrumental in advancing our knowledge of the biophysical processes underpinning the spatial arrangement within the nucleus.

Before adding complexity, we mention that, perhaps surprisingly, very simple polymer models can already provide fruitful insights into genome organisation. For instance, simulations of chromosome decompaction after mitosis reveal that the formation of territories may be attributed to the slow relaxation dynamics arising from polymeric topological constraints [25]. Entropic effects due to the flexibility of the chromatin fibre can partly account for the difference in radial positioning between chromosomes [26]. In addition, simple polymer simulations have suggested that the 3D conformation of chromatin is to some extent consistent with a fractal globule structure [27]. Nevertheless, generic polymer models can only capture limited aspects of nuclear arrangement. One needs additional details to recapitulate other phenomena discovered in recent higher resolution Hi-C and microscopy experiments. Here, we discuss two exciting developments in mechanistic modelling: one focussing on folding mediated by multivalent, chromatin-bridging proteins and the other on chromatin domains formed by the loop extrusion (LE) mechanism. We also review recent work on combining these two modelling frameworks.

Multivalent, chromatin-bridging proteins drive phase separation and compartmentalisation

Although the polymeric nature of chromatin is key to genome topology, theories based on polymer physics alone cannot fully account for the higher-order organisation of this fibre, such as the nonrandom partitioning into compartments and domains. A simple extension to the CG polymer model is the inclusion of diffusive chromatin-binding molecules (also called 'binders' or 'factors'), representing protein complexes, which associate with chromatin. This class of models has been known as the strings and binders switch (SBS)

model [28] or the **transcription factor (TF)** model [29]. The binders are typically represented as spherical beads for simplicity. In earlier versions of the model, they were allowed to bind nonspecifically to any sites along the chromatin fibre. Crucially, the binders are multivalent and, thus, they can form molecular bridges between two or more sites. Multivalency is a natural assumption, because each binder typically represents a complex of DNA-binding proteins: then, even if each of the single proteins is monovalent, the complex is multivalent. The model successfully reproduces Hi-C results on the decay in the contact probability between two chromatin segments as a function of their genomic separation [28].

An unexpected phenomenon that emerges from this model is that, when the interaction with chromatin is strong enough, the binders themselves, which are not directly attracted to one another, tend to cluster and form large aggregates. This effect was dubbed 'bridging-induced attraction' (BIA) [29], and it briefly works as follows (Figure 3A). First, diffusive multivalent binders associate with several chromatin loci, leading to a local increase in chromatin density. This effect, in turn, recruits more binders to that region, and they facilitate more bridging between chromatin segments, resulting in a further increase in density. This chain of events initiates a positive feedback loop, which ultimately results in **phase separation** [30], whereby a local high concentration of proteins is developed (see Box 2 for further discussion on BIA and phase separation). Recently, microscopy experiments examining DNA–protein interactions provided evidence of BIA in action, both *in vitro* and *in vivo* [31]. BIA also offers an appealing mechanism to explain the biogenesis of membraneless organelles seen within the nucleus or nuclear bodies [32].

When there is only a single type of nonspecific binder, protein clusters arising through BIA coarsen indefinitely, so that only one remains in steady state. This macroscopic phase separation behaviour is inconsistent with experimental data, which show nuclear bodies have well-defined sizes and do not typically grow beyond these [32]. To improve agreement with experiments, specific (e.g., sequence-dependent) protein-binding sites were incorporated into the chromatin fibre [33]. Specific binders create chromatin loops when they cluster, and the thermodynamic (entropic) cost associated with looping arrests coarsening and results in the formation of smaller, size-limiting clusters, qualitatively similar to nuclear bodies [34].

Protein clustering is accompanied by the formation of chromatin domains, regions of chromatin enriched in *cis*-interactions. As with clusters, typically there are multiple chromatin domains, which are enriched in different histone modifications. To account for this, multiple species of binders were added to the TF model (Figure 3B) [33,35], each binding to different types of chromatin (i.e., chromatin with different histone modifications). With this additional ingredient, multiple chromatin domains form spontaneously. Domains mediated by different types of binder strongly segregate in space, a phenomenon akin to compartmentalisation [30]. Surprisingly, by including only two types of binder, active (e.g., representing TFs

and polymerases) and inactive species (e.g., representing Polycomb complexes and heterochromatin protein 1), and by using experimental data [e.g., **chromatin immunoprecipitation with sequencing (ChIP-seq)** for histone modifications] to determine precise locations of the binding sites, this model accurately predicted 85% of the domain boundaries found in Hi-C data [33].

With multiple types of binder, BIA yields both clusters of active binders, which may be related to TFs [36], and inactive binders, which could be, for instance, Polycomb bodies [37]. It is also possible to use the model to study the dynamical properties of protein clusters or nuclear bodies. Experiments such as **fluorescence recovery after photobleaching (FRAP)** have shown that these protein foci are macroscopically stable but highly dynamic: there is a constant turnover of constituents within the foci with those from the soluble pool [34]. This feature was not captured by the original TF model, in which clusters were relatively static once established. As a result, another ingredient was added: binders that can 'switch' stochastically between a binding and a nonbinding state (Figure 3C) [34]. Switching mimics active (i.e., ATP-dependent) processes, such as post-translational modifications or protein degradation, driving the model out of equilibrium, a ubiquitous characteristic of biological systems. With switching, protein clusters can still nucleate via BIA, while their compositions become more dynamical.

Loop extrusion mediates the formation of TADs

Another cutting-edge development in the mechanistic modelling of chromatin is the LE model (see [38] for a detailed review). Extending the basic polymer framework, the model posits that there are protein complexes that can attach to chromatin or DNA at a single point and translocate outward to create loops (Figure 3D). LE was originally proposed as a mechanism for chromosome compaction during mitosis [39,40], but recently this concept has been revived thanks to the availability of higher resolution data on chromatin architecture. In the context of genome organisation during interphase, Hi-C results [11] revealed that many TADs are supported by chromatin loops, and domain boundaries are often colocalised with subunits of **structural maintenance of chromosomes (SMC)** complexes as well as CTCF, the binding motif of which on DNA is orientation specific. Surprisingly, most loop anchors are associated with convergent CTCF-binding sites, and this bias is incompatible with a model in which simple diffusive binders mediate the loops (i.e., the TF model). Therefore, new models were proposed: SMC proteins, such as cohesin and condensin, were postulated to act as extruding factors with motor activity that generate loops until reaching (occupied) convergent CTCF sites [41,42].

Theoretical modelling of LE has prompted more experiments on this topic. There has been growing evidence for extrusion mediated by SMC complexes and for its role in establishing topological domains with CTCF,

with genome-editing experiments showing that manipulating CTCF-binding sites can alter chromatin domain patterns [41,43]. Domains are also affected by degradation of CTCF [44], cohesin [45,46], and complexes involved in cohesin loading and unloading [47,48]. More recently, a plethora of single-molecule experiments provided direct evidence of the extrusion activity of cohesin and condensin on DNA *in vitro* [49–53]. Interestingly, these studies indicated that some SMC complexes may have a one-sided, asymmetric motor activity [50,52], inconsistent with earlier models, which assume a two-sided activity [41,42]. This finding led to more theoretical work examining the impact of one-sided versus two-sided extrusion on the resulting domain architecture [54,55]. In particular, it was shown that strictly one-sided extrusion cannot reproduce the 'dots' near TAD boundaries seen in Hi-C, suggesting that additional modifications, such as a mixture of one-sided and two-sided motors, are needed to capture all features of TADs (Figure 3E).

Combining mechanisms and modelling a heteromorphic chromatin fibre

The TF and LE models successfully capture many important aspects of chromatin architecture; however, these models on their own cannot account for all major findings from experiments. It has been suggested that phase separation (e.g., due to chromatin–protein–chromatin or protein–protein interactions) and LE are two independent competing mechanisms regulating genome topology [56,57]. A natural next step is to combine the TF and LE models [58]. This amalgamation results in a model that can give rise to compartments as seen in Hi-C interaction matrices, as well as TADs and the associated features that occur in these matrices due to LE.

Recently, this combined model was applied to look at the locus structure of the developmental gene *Pax6* in mice, and led to further refinement of the model and identification of an additional biophysical principle of chromatin folding related to its local compaction [59]. This study compared modelling results with those from **Capture-C** and FISH experiments to investigate how the 3D topology of the gene, particularly promoter–enhancer interactions, relates to its expression level. While model predictions and experimental data were in agreement when the gene was transcribed at a low or moderate level, they differed substantially when the gene was expressed at a high level. Here, FISH data indicated that the locus became more decompacted than in the moderate expression case, and there was an increase in separation between the gene promoter and one of its enhancers, an observation inconsistent with the classic picture of transcriptional activation associated with promoter–enhancer looping. Closer inspection revealed that there was a higher level of acetylation (i.e., histone 3 lysine 27 acetylation; H3K27ac) within the locus when gene transcription was upregulated. It was hypothesised that, locally, this modification may correlate with a more open or disrupted structure of the chromatin fibre, as suggested by previous experiments [60]. To incorporate these ideas, the

model was modified to describe chromatin as a heteromorphic rather than homogeneous fibre. While it is likely for chromatin to have several levels of local compaction depending on its surrounding environment, for simplicity only two levels were considered: a less compact fibre in regions with H3K27ac and a more compact one in those without; this improvement was already sufficient to recapitulate all experimental observations. This resulting model is known as the highly predictive heteromorphic polymer (HiP-HoP) model (Figure 4), and it has been used to study other loci, such as *Sox2* [59]. More recently, it has been used to examine the effects of genomic rearrangement on chromatin organisation in human cancer cell lines [61]. The HiP-HoP model allows one to simulate a population of structures for different gene loci. These structures can be compared with data from single cell Hi-C as well as those from more advanced FISH imaging techniques, such as DNA-based **multiplexed error-robust FISH (MERFISH)** and **seqFISH**.

It is remarkable to see that relatively simple polymer models, such as HiP-HoP, can reproduce many features of genome organisation. Yet, sometimes even simpler models are already sufficient to capture the biological phenomenon of interest. As an example, bead-and-spring models with direct polymer–polymer interactions (or block copolymer models), which could be a simplified representation of chromatin bridging mediated by proteins, can account for phenomena such as phase separation and compartmentalisation [56,62]. They can also explain how 3D folding has a role in regulating the epigenetic landscape along the chromatin fibre [63–65]. Moreover, these models can incorporate interactions with the nuclear lamina to examine large-scale genome reorganisation in rod cells of nocturnal mammals [66] and in cellular senescence [67,68].

Concluding remarks

Mechanistic polymer models are becoming more and more important in chromatin biology research because of the increasing need to interpret experimental results and identify mechanisms underlying a certain observation or behaviour. This type of modelling is predictive, because it requires minimal experimental data as input, and its output can be directly compared to existing and new data to validate or reject the hypotheses the model has been based on. The comparison can set up an iterative refinement process in which additional ingredients are introduced in a modular way to account for unexplained observations (Figure 2); when successful, this process can lead to the discovery of new fundamental principles of chromosome organisation.

We have reviewed here two biophysical principles that are central to most mechanistic polymer models for chromatin. The first is the bridging-induced attraction, which arises from the generic tendency of multivalent chromatin-binding proteins to aggregate and phase separate, thus providing a framework to explain the

biogenesis of nuclear bodies and the compartmentalisation of human chromosomes in active and inactive chromatin domains. The second is LE, which describes the process through which SMC complexes actively create growing loops of chromatin, and also explains the striking bias in favour of convergent CTCF loops found by Hi-C.

Models combining these two principles are promising candidates for a mechanistic model of chromatin folding, which can be used in a predictive way genome-wide and in many organisms, in view of the limited number of experimental data needed as input. As an example, we discussed here the HiP-HoP model, which also includes the heteromorphic nature of the chromatin fibre, in which active and inactive regions have different degrees of local compaction. While the HiP-HoP model has been used to explain the folding of specific genomic loci in mammalian cells, with sufficient computing resources it would also be possible to apply this model genome-wide, for instance to create a compendium of 3D structures for all human genes. Just as genome-wide association studies have been invaluable to discover new patterns in human genetics, we suggest that these genome-wide mechanistic models will provide an unprecedented resource with which to search for additional missing biophysical principles for chromatin folding. Since models such as HiP-HoP resolve both transcription-related proteins (e.g., RNA polymerases) and chromatin, we can attempt to link structure to transcriptional activity; this was demonstrated using the TF model in [69], which examined the relation between the spatial structure of active domains and gene regulatory networks. In the future, a more detailed treatment, such as the incorporation of LE and chromatin heteromorphicity, could be used to further elucidate the elusive biophysical link between 3D gene structure and transcription.

While polymer models have typically focussed on chromatin structure, they are also capable of predicting the intranuclear dynamics of chromatin in 3D. As higher resolution microscopy methods are developed in live cells, it would be of interest to see whether the predictions are quantitatively accurate. We also anticipate that mechanistic models may be used in the future to understand local changes in 3D chromatin structure following gene activation, or global changes due to differentiation and reprogramming [70]. To do so, it is likely that the models described here will need to be modified to account for a local dynamic chromatin state, such that the beads constituting the polymers can change properties as a region changes epigenetic mark or opens up following transcriptional activation. These innovations represent exciting challenges ahead (see Outstanding questions), which we hope can be addressed soon.

Outstanding questions

Do we know all of the main biophysical mechanisms underlying chromatin folding in 3D, or are there more to be discovered?

How well do existing mechanistic models of chromatin organisation predict 3D gene structure genome-wide in human and mammalian nuclei?

Current polymer models are focussed on Hi-C and structural properties of chromatin. Do they give an accurate prediction for its intranuclear dynamics?

Can we use mechanistic models to refine our understanding of the elusive link between 3D chromatin structure and transcriptional activity?

How does gene activation affect chromatin structure? Can we extend the current polymer models for chromatin to account for dynamic local structural changes due to, for instance, transcriptional activation?

What are the transcriptional and structural chromatin changes observed during cell differentiation and how are these linked? Can we predict them using mechanistic modelling?

Declaration of interests

None declared by authors.

References

[1]   T. Misteli. The self-organizing genome: principles of genome architecture and function. *Cell* 183, 28–45 (2020)

[2]   B. van Steensel and E. E. Furlong. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* 20, 327–337 (2019)

[3]   H. Zheng and W. Xie. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* 20, 535–550 (2019)

[4]   P. R. Cook and D. Marenduzzo. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res.* 46, 9895–9906 (2018)

[5]   D. G. Lupiáñez, M. Spielmann, S. Mundlos. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet.* 32, 225–237 (2016)

[6]   R. Kempfer and A. Pombo. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21, 207–226 (2020)

[7]   T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in

mammalian cells. *Nat. Rev. Genet.* 2, 292–301 (2001)

[8]   E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293 (2009)

[9]   E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, E. Heard. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385 (2012)

[10]   J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012)

[11]   S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680 (2014)

[12]   T. H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, X. Darzacq. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* 78, 539–553.e8 (2020)

[13]   C. A. Brackey, D. Marenduzzo, N. Gilbert. Mechanistic modeling of chromatin folding to understand function. *Nat. Methods* 17, 767–775 (2020)

[14]   H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole, S. J. Marrink. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 225–248 (2014)

[15]   J. R. C. van der Maarel. *Introduction to Biopolymer Physics*, World Scientific Publishing (2007)

[16]   G. Tiana and L. Giorgetti, eds (2019) *Modeling the 3D conformation of genomes*, CRC Press

[17]   F. Serra, M. Di Stefano, Y. G. Spill, Y. Cuartero, M. Goodstadt, D. Baù, M. A. Marti-Renom. Restraint-based three-dimensional modeling of genomes and genomic domains. *FEBS Lett.* 589, 2987–2995 (2015)

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

[19]   M. Di Stefano, J. Paulsen, T. G. Lien, E. Hovig, C. Micheletti. Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. *Sci. Rep.* 6, 35985 (2016)

[20]   F. Serra, D. Baù, M. Goodstadt, D. Castillo, G. Filion, M. A. Marti-Renom. Automatic analysis and 3D-

modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* 13, e1005665 (2017)

[21]   L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, E. Heard. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* 157, 950–963 (2014)

[22]   M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, J. N. Onuchic. Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. USA* 113, 12168–12173 (2016)

[23]   S. Bianco, D. G. Lupiáñez, A. M. Chiariello, C. Annunziatella, K. Kraft, R. Schöpflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, S. Mundlos, M. Nicodemi. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* 50, 662–667 (2018)

[24]   B. K. Kragesteen, M. Spielmann, C. Paliou, V. Heinrich, R. Schöpflin, A. Esposito, C. Annunziatella, S. Bianco, A. M. Chiariello, I. Jerković, I. Harabula, P. Guckelberger, M. Pechstein, L. Wittler, W. L. Chan, M. Franke, D. G. Lupiáñez, K. Kraft, B. Timmermann, M. Vingron, A. Visel, M. Nicodemi, S. Mundlos, G. Andrey. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat. Genet.* 50, 1463–1473 (2018)

[25]   A. Rosa and R. Everaers. Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* 4, e1000153 (2008)

[26]   P. R. Cook and D. Marenduzzo. Entropic organization of interphase chromosomes. *J. Cell Biol.* 186, 825–834 (2009)

[27]   L. A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* 19, 37–51 (2011)

[28]   M. Barbieri, M. Chotalia, J. Fraser, L. M. Lavitas, J. Dostie, A. Pombo, M. Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16173–16178 (2012)

[29]   C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, D. Marenduzzo. Nonspecific bridging- induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. U. S. A.* 110, E3605–E3611 (2013)

[30]   F. Erdel, K. Rippe. Formation of chromatin subcompartments by phase separation. *Biophys. J.* 114, 2262–2270 (2018)

[31]   J. K. Ryu, C. Bouchoux, H. W. Liu, E. Kim, M. Minamino, R. de Groot, A. J. Katan, A. Bonato, D. Marenduzzo, D. Michieletto, F. Uhlmann, C. Dekker. Bridging-induced phase separation induced by cohesin SMC protein complexes. *Sci. Adv.* 7, eabe5905 (2021)

[32]   Y. S. Mao, B. Zhang, D. L. Spector. Biogenesis and function of nuclear bodies. *Trends Genet.* 27, 295–306 (2011)

[33]   C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook, D. Marenduzzo. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* 44, 3503–3512 (2016)

[34]   C. A. Brackley, B. Liebchen, D. Michieletto, F. Mouvet, P. R. Cook, D. Marenduzzo. Ephemeral protein binding to DNA shapes stable nuclear bodies and chromatin domains. *Biophys. J.* 112, 1085–1093 (2017)

[35]   A. M. Chiariello, C. Annunziatella, S. Bianco, A. Esposito, M. Nicodemi. Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* 6, 29775 (2016)

[36]   W. K. Cho, J. H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, I. I. Cisse. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415 (2018)

[37]   R. Tatavosian, S. Kent, K. Brown, T. Yao, H. N. Duc, T. N. Huynh, C. Y. Zhen, B. Ma, H. Wang, X. Ren. Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J. Biol. Chem.* 294, 1451–1463 (2019)

[38]   E. J. Banigan and L. A. Mirny. Loop extrusion: theory meets single-molecule experiments. *Curr. Opin. Cell Biol.* 64, 124–138 (2020)

[39]   K. Kimura, V. V. Rybenkov, N. J. Crisona, T. Hirano, N. R. Cozzarelli. 13S condensin actively reconfigures DNA by introducing global positive writhe: implications for chromosome condensation. *Cell* 98, 239–248 (1999)

[40]   K. Nasmyth. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.* 35, 673–745 (2001)

[41]   A. L. Sanborn, S. S. P. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, E. L. Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* 112, E6456–E6465 (2015)

[42]   G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, L. A. Mirny. Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 2038–2049 (2016)

[43]   Y. Guo, Q. Xu, D. Canzio, J. Shou, J. Li, D. U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M. Q. Zhang, B. Ren, A. R. Krainer, T. Maniatis, Q. Wu. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 162, 900–910 (2015)

[44]   E. P. Nora, A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau. Targeted degradation of CTCF decouples local insulation of chromosome domains

from genomic compartmentalization. *Cell* 169, 930–944.e22 (2017)

[45] S. S. P. Rao, S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, E. L. Aiden. Cohesin loss eliminates all loop domains. *Cell* 171, 305–320.e24 (2017)

[46] W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. H. Haering, L. Mirny, F. Spitz. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51–56 (2017)

[47] J. H. I. Haarhuis, R. H. van der Weide, V. A. Blomen, J. O. Yáñez-Cuna, M. Amendola, M. S. van Ruiten, P. H. L. Krijger, H. Teunissen, R. H. Medema, B. van Steensel, T. R. Brummelkamp, E. de Wit, B. D. Rowland. The cohesion release factor WAPL restricts chromatin loop extension. *Cell* 169, 693–707.e14 (2017)

[48] G. Wutz, C. Várnai, K. Nagasaka, D. A. Cisneros, R. R. Stocsits, W. Tang, S. Schoenfelder, G. Jessberger, M. Muhar, M. J. Hossain, N Walther, B. Koch, M. Kueblbeck, J. Ellenberg, J. Zuber, P. Fraser, J.-M. Peters. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* 36, 3573–3599 (2017)

[49] T. Terakawa, S. Bisht, J. M. Eeftens, C. Dekker, C. H. Haering, E. C. Greene. The condensin complex is a mechanochemical motor that translocates along DNA. *Science* 358, 672–676 (2017)

[50] M. Ganji, I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, C. Dekker. Real-time imaging of DNA loop extrusion by condensin. *Science* 360, 102–105 (2018)

[51] I. F. Davidson, B. Bauer, D. Goetz, W. Tang, G. Wutz, J.-M. Peters. DNA loop extrusion by human cohesin. *Science* 366, 1338–1345 (2019)

[52] M. Kong, E. E. Cutts, D. Pan, F. Beuron, T. Kaliyappan, C. Xue, E. P. Morris, A. Musacchio, A. Vannini, E. C. Greene. Human condensin I and II drive extensive ATP-dependent compaction of nucleosome-bound DNA. *Mol. Cell* 79, 99–114.e9 (2020)

[53] S. Golfier, T. Quail, H. Kimura, J. Brugués. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *eLife* 9, e53885 (2020)

[54] E. J. Banigan, L. A. Mirny. Limits of chromosome compaction by loop-extruding motors. *Phys. Rev. X* 9, 031007 (2019)

[55] E. J. Banigan, A. A. van den Berg, H. B. Brandão, J. F. Marko, L. A. Mirny. Chromosome organization by one-sided and two-sided loop extrusion. *eLife* 9, e53558 (2020)

[56] J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, L. A. Mirny. Chromatin organization by an

interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.* 115, E6697–E6706 (2018)

[57]   L. A. Mirny, M. Imakaev, N. Abdennur. Two major mechanisms of chromosome organization. *Curr. Opin. Cell Biol.* 58, 142–152 (2019)

[58]   M. C. F. Pereira, C. A. Brackley, D. Michieletto, C. Annunziatella, S. Bianco, A. M. Chiariello, M. Nicodemi, D. Marenduzzo. Complementary chromosome folding by transcription factors and cohesin. *bioRxiv.* Published online April 23, 2018. https://doi.org/10.1101/305359 (2018)

[59]   A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, N. Gilbert. Polymer simulations of heteromorphic chromatin predict the 3D folding of complex genomic loci. *Mol. Cell* 72, 786–797.e11 (2018)

[60]   V. I. Risca, S. K. Denny, A. F. Straight, W. J. Greenleaf. Variable chromatin structure revealed by *in situ* spatially correlated DNA cleavage mapping. *Nature* 541, 237–241 (2017)

[61]   D. Rico, D. Kent, A. Mikulasova, N. Karataraki, B. A. Walker, B. M. Javierre, L. J. Russell, C. A. Brackley. High-resolution simulations of chromatin folding at genomic rearrangements in malignant B-cells provide mechanistic insights on proto-oncogene deregulation. *bioRxiv.* Published online July 26, 2021. http://dx.doi.org/10.1101/2021.03.12.434963 (2021)

[62]   D. Jost, P. Carrivain, G. Cavalli, C. Vaillant. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* 42, 9553–9561 (2014)

[63]   D. Michieletto, E. Orlandini, D. Marenduzzo. Polymer model with epigenetic recoloring reveals a pathway for the *de novo* establishment and 3D organization of chromatin domains. *Phys. Rev. X* 6, 041047 (2016)

[64]   D. Michieletto, M. Chiang, D. Colì, A. Papantonis, E. Orlandini, P. R. Cook, D. Marenduzzo. Shaping epigenetic memory via genomic bookmarking. *Nucleic Acids Res.* 46, 83–93 (2018)

[65]   D. Jost, C. Vaillant. Epigenomics in 3D: importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic Acids Res.* 46, 2252–2264 (2018)

[66]   M. Falk, Y. Feodorova, N. Naumova, M. Imakaev, B. R. Lajoie, H. Leonhardt, B. Joffe, J. Dekker, G. Fudenberg, I. Solovei, L. A. Mirny. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* 570, 395–399 (2019)

[67]   M. Chiang, D. Michieletto, C. A. Brackley, N. Rattanavirotkul, H. Mohammed, D. Marenduzzo, T. Chandra. Polymer modeling predicts chromosome reorganization in senescence. *Cell Rep.* 28, 3212–3223.e6 (2019)

[68]   S. Sati, B. Bonev, Q. Szabo, D. Jost, P. Bensadoun, F. Serra, V. Loubiere, G. L. Papadopoulos, J. C. Rivera-Mulia, L. Fritsch, P. Bouret, D. Castillo, J. L. Gelpi, M. Orozco, C. Vaillant, F. Pellestor, F. Bantignies,

M. A. Marti-Renom, D. M. Gilbert, J. M. Lemaitre, G. Cavalli. 4D genome rewiring during oncogene-induced and replicative senescence. *Mol. Cell* 78, 522–538.e9 (2020)

[69]   C. A. Brackley, N. Gilbert, D. Michieletto, A. Papantonis, M. C. F. Pereira, P. R. Cook, D. Marenduzzo. Complex small-world regulatory networks emerge from the 3D organisation of the human genome. *Nat. Commun.* 12, 5756 (2021)

[70]   M. Di Stefano, R. Stadhouders, I. Farabella, D. Castillo, F. Serra, T. Graf, M. A. Marti-Renom. Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. *Nat. Commun.* 11, 2564 (2020)

[71]   P. D. Dans, J. Walther, H. Gómez, M. Orozco. Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.* 37, 29–45 (2016)

[72]   T. A. Knotts, N. Rathore, D. C. Schwartz, J. J. De Pablo. A coarse grain model for DNA. *J. Chem. Phys.* 126, 084901 (2007)

[73]   D. M. Hinckley, G. S. Freeman, J. K. Whitmer, J. J. De Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* 139, 144903 (2013)

[74]   T. E. Ouldridge, A. A. Louis, J. P. Doye. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys. Rev. Lett.* 104, 178101 (2010)

[75]   T. E. Ouldridge, A. A. Louis, J. P. Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* 134, 085101 (2011)

[76]   O. Henrich, Y. A. Gutiérrez Fosado, T. Curk, T. E. Ouldridge. Coarse-grained simulation of DNA using LAMMPS: An implementation of the oxDNA model and its applications. *Euro. Phys. J. E* 41, 57 (2018)

[77]   J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, S. J. Marrink. Martini coarse-grained force field: extension to DNA. *J. Chem. Theory Comput.* 11, 3932–3945 (2015)

[78]   Y. A. Fosado, D. Michieletto, J. Allan, C. A. Brackley, O. Henrich, D. Marenduzzo. A single nucleotide resolution model for large-scale simulations of double stranded DNA. *Soft Matter* 12, 9458–9470 (2016)

[79]   C. Matek, T. E. Ouldridge, J. P. Doye, A. A. Louis. Plectoneme tip bubbles: coupled denaturation and writhing in supercoiled DNA. *Sci. Rep.* 5, 7655 (2015)

[80]   Y. A. Fosado, D. Michieletto, C. A. Brackley, D. Marenduzzo. Nonequilibrium dynamics and action at a distance in transcriptionally driven DNA supercoiling. *Proc. Natl. Acad. Sci. U. S. A.* 118, e1905215118 (2021)

[81]   E. Skoruppa, S. K. Nomidis, J. F. Marko, E. Carlon. Bend-induced twist waves and the structure of nucleosomal DNA. *Phys. Rev. Lett.* 121, 088101 (2018)

[82]   R. V. Reshetnikov, A. V. Stolyarova, A. O. Zalevsky, D. Y. Panteleev, G. V. Pavlova, D. V. Klinov, A. V. Golovin, A. D. Protopopova. A coarse-grained model for DNA origami. *Nucleic Acids Res.* 46, 1102–1112 (2018)

[83]   B. E. Snodin, J. S. Schreck, F. Romano, A. A. Louis, J. P. Doye. Coarse-grained modelling of the structural properties of DNA origami. *Nucleic Acids Res.* 47, 1585–1597 (2019)

[84]   Y. A. G. Fosado, Z. Xing, E. Eiser, M. Hudek, O. Henrich. (2019) A numerical study of three-armed DNA hydrogel structures. *arXiv* 2019, arXiv:1903.04186v2

[85]   Y. Fan, N. Korolev, A. P. Lyubartsev, L. Nordenskiöld. An advanced coarse-grained nucleosome core particle model for computer simulations of nucleosome-nucleosome interactions under varying ionic conditions. *PLoS ONE* 8, e54228 (2013)

[86]   O. Müller, N. Kepper, R. Schöpflin, R. Ettig, K. Rippe, G. Wedemann. Changing chromatin fibre conformation by nucleosome repositioning. *Biophys. J.* 107, 2141–2150 (2014)

[87]   T. M. Cheng, S. Heeger, R. A. Chaleil, N. Matthews, A. Stewart, J. Wright, C. Lim, P. A. Bates, F. Uhlmann. A simple biophysical model emulates budding yeast chromosome condensation. *eLife* 4, e05565 (2015)

[88]   O. Wiese, D. Marenduzzo, C. A. Brackley. Nucleosome positions alone can be used to predict domains in yeast chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 116, 17307–17315 (2019)

[89]   G. Ozer, A. Luque, T. Schlick. The chromatin fiber: multiscale problems and approaches. *Curr. Opin. Struct. Biol.* 31, 124–139 (2015)

[90]   G. Arya and T. Schlick. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J. Phys. Chem. A* 113, 4045–4059 (2009)

[91]   G. D. Bascom and T. Schlick. Chromatin fiber folding directed by cooperative histone tail acetylation and linker histone binding. *Biophys. J.* 114, 2376–2385 (2018)

[92]   C. P. Brangwynne, P. Tompa, R. V. Pappu. Polymer physics of intracellular phase transitions. *Nat. Phys.* 11, 899–904 (2015)

[93]   D. Michieletto, N. Gilbert. Role of nuclear RNA in regulating chromatin structure and transcription. *Curr. Opin. Cell Biol.* 58, 120–125 (2019)

[94]   J. Berry, S. C. Weber, N. Vaidya, M. Haataja, C. P. Brangwynne, D. A. Weitz. RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. U. S. A.* 112, E5237–E5245 (2015)

[95]   A. K. Shaytan, G. A. Armeev, A. Goncearenco, V. B. Zhurkin, D. Landsman, A. R. Panchenko. Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale:

atomistic insights into nucleosome functions. *J. Mol. Biol.* 428, 221–237 (2016)

[96]   S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 117, 1–19 (1995)

[97]   H. J. C. Berendesen, D. van der Spoel, R. van Drunen. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91, 43–56 (1995)

[98]   H-J. Limbach, A. Arnold, B. Mann, C. Holm. ESPResSo – an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.* 174, 704–727 (2006)

[99]   P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, V. S. Pande. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* 13, e1005659 (2017)

[100]   J. A. Anderson, C. D. Lorenz, A. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* 227, 5342–5359 (2008)

Box 1. Other coarse-grained schemes for DNA and chromatin modelling

The bead-and-spring polymer model is a common CG scheme for modelling chromatin and understanding its large-scale organisation. CG schemes have also been developed to examine the biophysical properties of DNA and chromatin at other length scales. At the nucleotide level, various models have been developed for single-stranded and double-stranded DNA [71]. These models forgo atomistic details of individual nucleotides. For example, some models represent each nucleotide as several effective beads, such as the 3SPN [72,73], oxDNA [74–76], and Martini models [77]. Others simplify the nucleotide as a single particle [78]. These models are useful for understanding how local mechanical properties of DNA influence its folding. They can be used to study topological effects, such as supercoiling [79,80] and twist-bend coupling [81], which are often neglected in lower resolution (more coarse-grained) models. They can also be used for investigating the properties of DNA-based materials in synthetic biology and nanotechnology, such as DNA origami [82,83] and hydrogels [84].

Beyond the base-pair level, there are CG models focussing on the nucleosomal-scale chromatin structure. Here, some models explicitly consider both the DNA fibre and the histone complexes [85,86], while others simply represent nucleosomes as single beads [87]. These models have been applied to examine nucleosome–nucleosome interactions [85] and how nucleosome positioning affects local chromatin compaction [86]. Recently, a polymer model describing nucleosomes as beads connected by linker DNA segments showed that nucleosome positioning alone can accurately capture chromatin domains as found in

Micro-C data for the budding yeast genome [88].

There are also multiscale CG models, which take advantage of simulation or experimental data from a higher resolution to inform model parameters at a lower resolution [89]. These models retain some information of microscopic effects, such as electrostatic interactions, while keeping the overall computational resource reasonable for studying chromatin arrangement mesoscopically. For instance, one model utilises crystal structure data to coarse-grain the nucleosome as a charged rigid body, while describing linker DNA and histone tails as simple bead chains [90]. This model was used recently to examine the interplay of histone tail acetylation and linker histone binding on the local compaction of the chromatin fibre [91].

Box 2. Bridging-induced attraction and phase separation

BIA provides a generic route to phase separation of chromatin and proteins within the nucleus (see Figure 3A–C in the main text) [34]. Given that this is not the phase separation mechanism most commonly discussed in the literature, here we explain it in more detail.

A classic route to intracellular phase separation is via weak multivalent protein–protein interactions [92], which are mainly electrostatic in nature. This type of protein-mediated phase separation can be observed *in vitro* in systems without any DNA or chromatin, at sufficiently large protein concentration (~µM) and compatible salt concentrations and temperature. Typically, protein-mediated phase separation results in the formation of large droplets, which coarsen: eventually, a single droplet remains in steady state, so that the system undergoes what is known as macroscopic phase separation. If proteins in the droplet are mobile, the droplet can further be classified as fluid (or liquid), and the emerging phenomenon is called liquid–liquid phase separation.

Alternatively, BIA can lead to phase separation via the feedback mechanism shown in Figure 3A in the main text. This feedback requires multivalent chromatin–protein interactions and does not necessitate any protein–protein interactions (the latter can be present but are not required). The emergent phase separation does not occur in the absence of DNA and is then best described as chromatin mediated; it can be achieved *in vitro* with protein concentrations as small as nM [31].

Chromatin-mediated phase separation can be macroscopic if the interaction between chromatin and protein is nonspecific, as in Figure 3A in the main text. However, if there are also sites of specific interactions between chromatin and proteins, then the end result is typically microphase separation, or arrested phase separation, with formation of clusters that do not coarsen indefinitely, but instead reach a self-limiting size. Microphase separation arises as the ensuing clusters bring together sites of specific interactions, which are linked by chromatin loops. Then, gathering several chromatin loops limits cluster growth both kinetically and

thermodynamically [34]. Kinetically, chromatin loops form a corona surrounding each cluster, which provides a barrier to cluster merging. Thermodynamically, the entropic cost of chromatin loops increases superlinearly with their number, so that the growth of large clusters is disfavoured. Protein switching (see Figure 3C in the main text) also leads to microphase separation.

Whether the clusters arising in chromatin-mediated phase separation are liquid, gel-like, glassy, or solid-like is an open question, and the answer likely depends on parameters. To what extent a cluster may be considered as a separate phase is also unclear, especially when the number of constituent molecules is ~10 or fewer. Protein-mediated and chromatin-mediated phase separation, while useful concepts, do not cover all possible types of phase separation occurring inside the cell. Another important class is RNA-mediated phase separation (e.g., [93,94]).

Table 1. Comparison of different approaches to modelling genome organisation

	Mechanistic modelling	**Inverse modelling**
Aim	• Predicts 3D genome structures using a set of biophysical rules that are based on hypotheses and/or experimental observations	• Reconstructs a single or an ensemble of 3D genome structures from 3C-based experimental data
When is it useful?	• Focussing on understanding the molecular mechanisms driving DNA or chromatin folding • Experimental resources are limited and/or experiments are difficult to undertake	• Focussing on extracting genome structural information from existing data sets • 3C-based data are abundant and can be obtained easily
Example models	• Atomistic model resolving the full structure of DNA • CG models for DNA structure at the base-pair level. Examples include 3SPN [72,73], oxDNA [74–76], and Martini models [77] • Bead-and-spring CG polymer models for chromatin structure beyond the nucleosome level. Examples include TF [29], LE [38], and HiP-HoP [59] models	• Bead-and-spring CG polymer models. Some examples include TADbit [20], MiChroM [22], and PRISMR [23,24]
Software tools	• Atomistic models are mostly implemented using molecular dynamics (MD) packages, such as LAMMPS [96] and GROMACS [97]	• Typically involves in-house code to perform the fitting, machine learning, or iterative Monte Carlo process

	• CG models involving MD can be implemented using packages such as LAMMPS, ESPResSo [98], OpenMM [99], and HOOMD [100] • CG models involving Monte Carlo sampling typically use in-house code	• Physics-based models that utilise MD to reconstruct structures can be implemented using packages similar to those for mechanistic modelling
Advantages	• Involves fewer parameters than inverse models, making it more feasible for extensive parameter sweeping • Allows rapid, systematic exploration and testing of different mechanisms regulating genome structure	• Provides 3D visualisation of contact maps from 3C-based experiments • Provides structures of individual gene loci that are highly consistent with experimental data
Limitations	• Devising new hypotheses and rules for these models is a nontrivial process requiring both biological and physical insights • Values of some parameters (e.g., binding and unbinding rates for loop extruding factors) are still not known from experiments • Genome-wide mechanistic modelling can be computationally intensive	• Often contains many parameters, making it difficult to interpret the biophysical meaning of each of them • Difficult to probe chromatin dynamics, such as those due to loop extrusion and protein binding, over short timescales • Predicted structures are strongly dependent on the quality of input data

Glossary

Capture-C: variant of the 3C technology, that uses specialised probes to selectively enrich ligased products for pairwise interactions that involve the targeted locus.

CCCTC-binding factor (CTCF): TF with zinc finger domains that is important for chromatin insulation and for regulating transcription and genome structure.

Chromatin immunoprecipitation with sequencing (ChIP-seq): technique that uses antibody binding and next-generation sequencing to measure genome-wide the enrichment of a specific protein or histone modification along the chromatin fibre.

Chromosome conformation capture (3C): molecular-based technology for measuring the frequency of interactions between pairs of chromatin loci using a protocol involving fixation, crosslinking, digestion, and ligation. Quantification was done originally by PCR and more recently by next-generation sequencing.

Coarse graining (CG): procedure whereby microscopic details of a physical system that are irrelevant to the

length scale of interest are removed from consideration.

Compartments: feature of chromatin organisation above the megabase level, with chromatin divided into two broad groups (A or B compartments). Segments within the same compartment interact more frequently than those in different compartments, and share similar physiological properties, such as transcriptional activity and replication timing.

Enhancer: stretch of DNA sequence recognised by TFs; helps regulate gene expression, typically by means of interacting with promoters.

Fluorescence *in situ* hybridisation (FISH): technique that uses complementary base pairing to bind fluorescence probes to targeted DNA sequences to visualise their location within the nucleus using microscopy.

Fluorescence recovery after photobleaching (FRAP): method for interrogating the dynamics of the internal constituents of a cellular or nuclear focus by inactivating fluorophores within the focus (bleaching) and analysing the rate of recovery of fluorescence signal afterwards.

Hi-C: high-throughput variant of the 3C technology with massively parallel sequencing to measure the frequency of interactions between all pairs of genomic loci from sampling a population of cells.

Micro-C: a variant of the 3C method that can probe chromatin interactions at the nucleosome level.

Multiplexed error-robust FISH (MERFISH): FISH technique that allows simultaneous detection of the spatial location of thousands of RNA or DNA loci by assigning unique binary barcodes to individual loci that are read sequentially via multiple rounds of hybridisation and imaging measurements. The barcode indicates whether the loci are fluorescently labelled in each round of hybridisation.

Phase separation: phenomenon in which different constituents of a system demix from a homogeneous mixture and form separated foci or phases, with each phase enriched in a specific type of constituent.

Promoter: stretch of DNA sequence associated with a gene that is located upstream and is responsible for controlling its transcription activity by mediating the binding of RNA polymerase to the region.
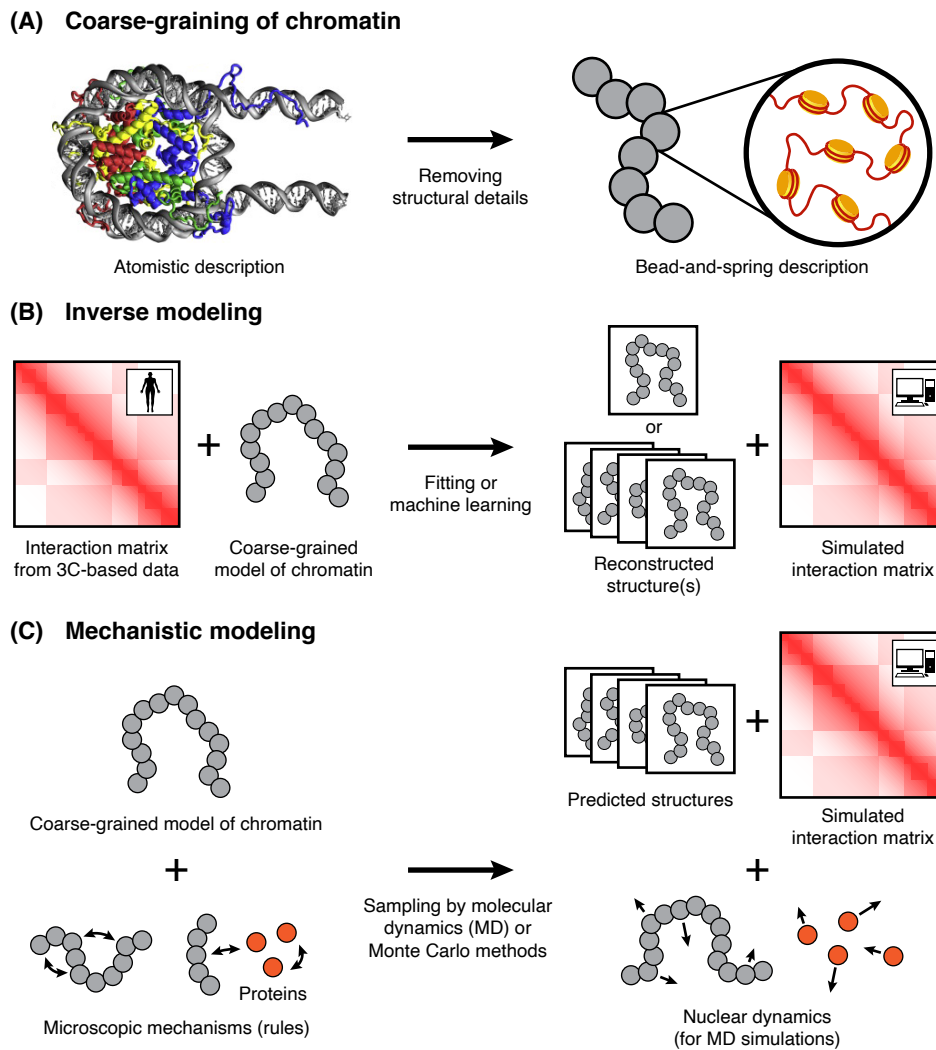
seqFISH: similar to MERFISH; another FISH technique that uses a barcoding system to simultaneously detect the location of thousands of RNA or DNA loci within a single cell. Here, the barcode is based on a combination of different fluorescent probe colours that are detected sequentially via multiple rounds of hybridisation.

Structural maintenance of chromosomes (SMC) complexes: class of architectural proteins that mediate the 3D compaction of chromosomes. Two well-known SMC complexes are cohesin and condensin.

Topologically associating domains (TADs): chromatin regions spanning hundreds of kbp with a higher frequency of intra- over inter-region interactions, as seen in Hi-C data.
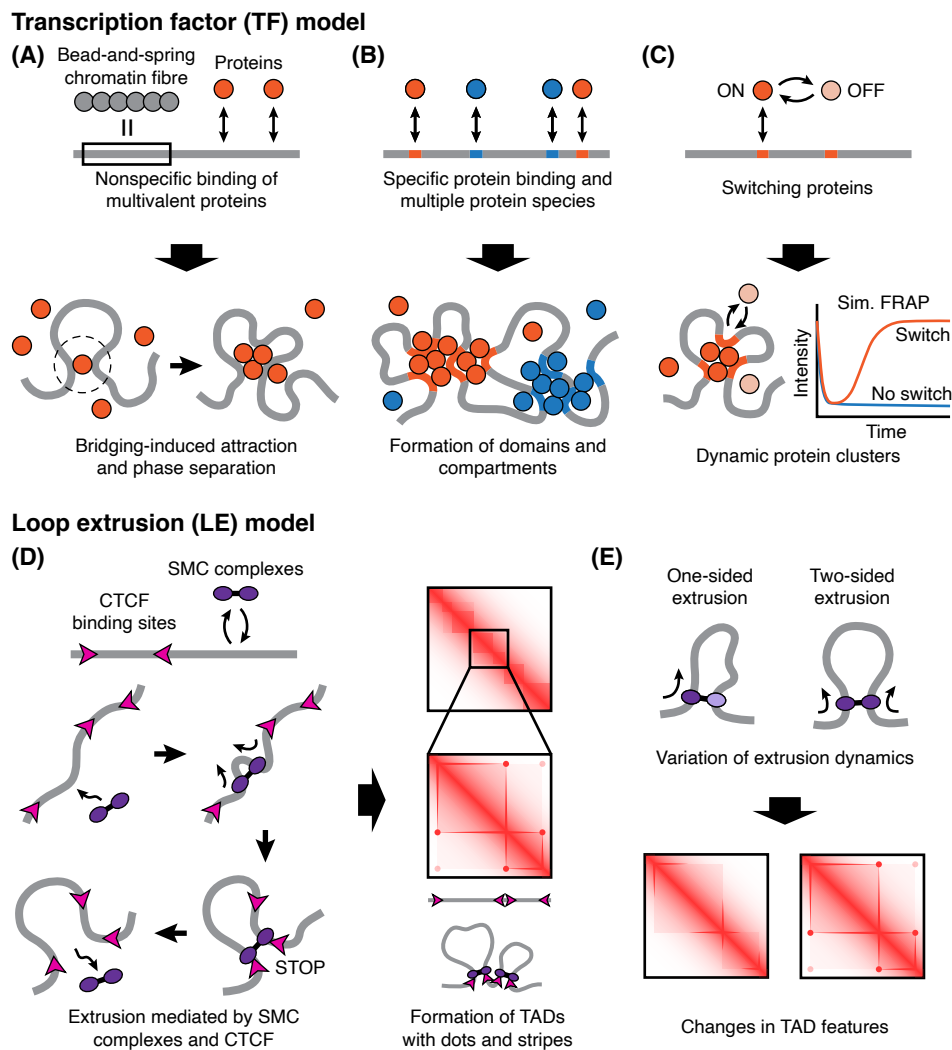
Transcription factors (TFs): protein complexes that bind to specific DNA sequences to regulate gene expression.

(A) Coarse-graining of chromatin

Atomistic description → Removing structural details → Bead-and-spring description

(B) Inverse modeling

Interaction matrix from 3C-based data + Coarse-grained model of chromatin → Fitting or machine learning → Reconstructed structure(s) or + Simulated interaction matrix

(C) Mechanistic modeling

Coarse-grained model of chromatin + Microscopic mechanisms (rules) / Proteins → Sampling by molecular dynamics (MD) or Monte Carlo methods → Predicted structures + Simulated interaction matrix + Nuclear dynamics (for MD simulations)

Figure 1. Various schemes in chromatin modelling. (A) Chromatin can be modelled atomistically with all components (DNA and histone complexes) accounted for. To simulate a larger region of chromatin, it is common to perform 'coarse graining' (CG) by removing some structural details. A popular CG scheme is the bead-and-spring polymer model, in which each bead represents a certain amount of chromatin (or number of nucleosomes). (B,C) Two prevalent approaches that use CG models to understand 3D genome folding are inverse modelling and mechanistic modelling. (B) In the inverse approach, 3C-based data are used as input to fit or train model parameters (e.g., interaction strength between chromatin loci). The output of the model is a population-averaged structure or an ensemble of structures, alongside a simulated interaction matrix. (C) In the mechanistic approach, the input is a set of microscopic mechanisms or 'rules' postulated to be important in regulating genome organisation. Molecular dynamics (MD) simulations or Monte Carlo algorithms implementing these rules are then used to sample possible conformations. As before, the output is a set of 3D structures and a simulated interaction matrix. Additionally, models implemented using MD simulations can also give information on nuclear dynamics. Modified from [95] (A).
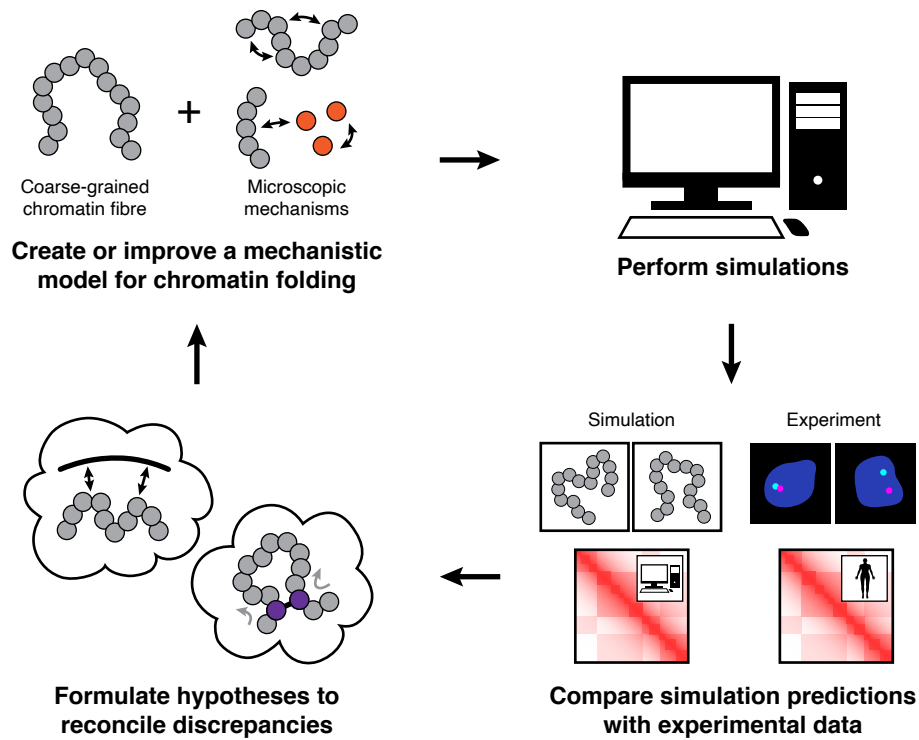
Figure 2. Key figure. Successive refinement of mechanistic models provides a systematic approach to uncover mechanisms regulating genome folding. This cyclic model development process is as follows: first, one starts with a simple mechanistic model with rules based on known mechanisms that shape genome architecture. Second, simulations are performed based on the defined model. Third, the predicted structures from the simulations are compared with available experimental data. Systematic discrepancies between two sets of results are noted, and new hypotheses or rules are then devised to reconcile the differences. The rules that successfully explain the differences between the simulation and experimental results can be incorporated into the model, generating new knowledge on the underlying mechanisms.

Transcription factor (TF) model

(A) Bead-and-spring chromatin fibre · Proteins

Nonspecific binding of multivalent proteins

Bridging-induced attraction and phase separation

(B)

Specific protein binding and multiple protein species

Formation of domains and compartments

(C) ON ⇄ OFF

Switching proteins

Dynamic protein clusters

Sim. FRAP · Intensity · Switch · No switch · Time

Loop extrusion (LE) model

(D) SMC complexes · CTCF binding sites · STOP

Extrusion mediated by SMC complexes and CTCF

Formation of TADs with dots and stripes

(E) One-sided extrusion · Two-sided extrusion

Variation of extrusion dynamics

Changes in TAD features

Figure 3. Two major developments in mechanistic modelling with successive refinements. (A–C) The transcription factor model. (A) Initially, the model considered a single species of multivalent protein, which binds nonspecifically anywhere along the chromatin fibre (modelled as a bead-and-spring chain). An emergent phenomenon is 'bridging-induced attraction', a positive feedback loop in which protein binding increases the local density of chromatin, leading to more proteins binding to the region (broken lines circle), which in turn drives further increase in density and so on, resulting in phase separation. (B) By introducing specific binding sites along the fibre and multiple protein species, the model yields chromatin domains and compartments. (C) With proteins switching stochastically between a binding and a nonbinding state, clusters become dynamic, and the model recapitulates trends in fluorescence recovery after photobleaching (FRAP) experiments for nuclear bodies. (D,E) The loop extrusion model. (D) This model postulates that there are active factors, such as structural maintenance of chromosomes (SMC) complexes, which bind to the DNA or chromatin fibre and translocate outward to form loops. These factors are thought to move along the fibre until reaching a CCCTC-binding factor (CTCF) site that is oriented opposite to its direction of travel. The model produces domains similar to topologically associating domains (TADs) in Hi-C experiments, with dots and

stripes near domain boundaries. (E) Motivated by recent single-molecule experiments studying the activity of SMC complexes on DNA [50,52], the model was modified with one-sided and two-sided extrusion considered [54,55]. These rules give rise to different domain features.
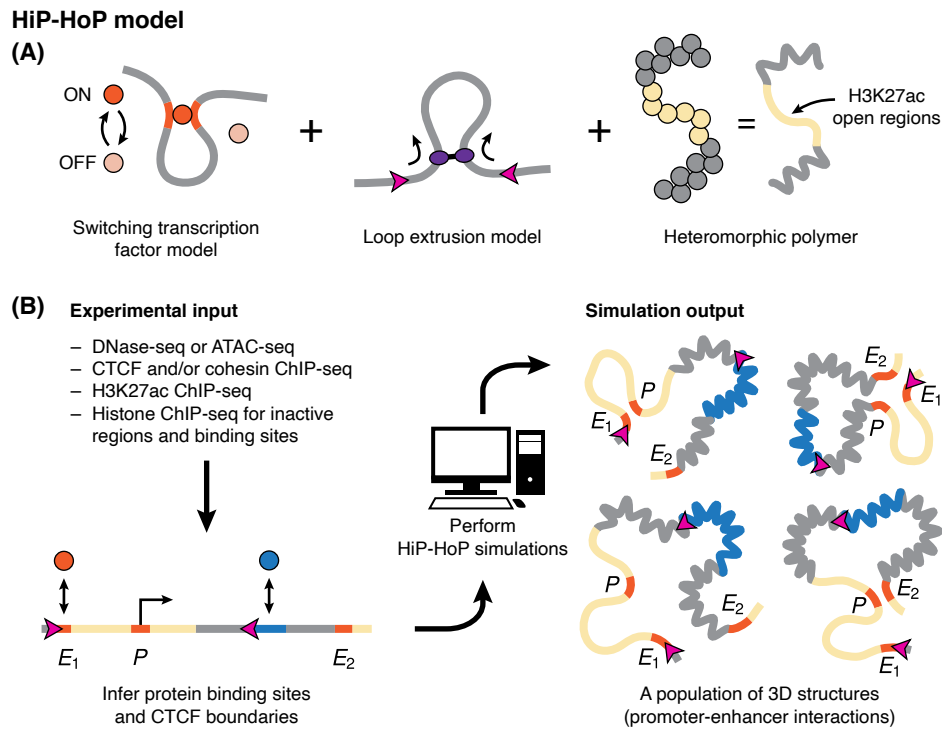
Figure 4. The highly predictive heteromorphic polymer (HiP-HoP) model. (A) This mechanistic model is an amalgamation of the transcription factor model with switching binders and the loop extrusion model with two-sided extrusion. The model additionally considers a heteromorphic chromatin fibre, with regions enriched in histone 3 lysine 27 acetylation (H3K27ac) postulated to adopt a more decompacted structure. (B) A typical workflow for running the HiP-HoP model. First, the model requires the following input data: chromatin accessibility (e.g., DNase-seq or ATAC-seq), ChIP-seq for CCCTC-binding factor (CTCF) and cohesin binding, and ChIP-seq for histone modifications (e.g., H3K27ac for active regions and H3K27me3 and H3K9me3 for inactive regions). These data are then used to infer protein-binding sites and CTCF boundaries. Afterwards, molecular dynamic (MD) simulations are performed based on the rules of the HiP-HoP model. By conducting many simulation runs, which can be viewed as different cells, a population of structures is generated. These structures allow one to investigate, for example, the interactions between promoters (P) and enhancers (E_1 and E_2), as done for the *Pax6* and *Sox2* loci in mice [59].