Mechanistic modeling of chromatin folding to understand function

Mechanistic Modelling of Chromatin Folding to Understand Function

Chris A. Brackey[1], Davide Marenduzzo[1], and Nick Gilbert[2]

[1]SUPA School of Physics and Astronomy, University of Edinburgh, Edinburgh EH9 3FD, UK.

[2]MRC Human Genetics Unit, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK.

Correspondence to:

Nick.Gilbert@ed.ac.uk or Davide.Marenduzzo@ed.ac.uk

Abstract

Understanding 3-D chromatin organisation and function is a goal for molecular biology. Traditionally, questions were addressed using experiments but as datasets increase in size, reaching a mechanistic interpretation is challenging. Consequently, polymer simulations and mechanistic modelling are necessary to explain function. As these approaches are daunting we provide a guide for biologists to comprehend and appreciate what they are used for. When research questions can only be addressed by interdisciplinary teams understanding each other's languages and methods will facilitate breakthroughs.

Introduction

In the last couple of decades technological advances in both next-generation sequencing and microscopy have led to an explosion of new methods to probe genome structure and function. With this has come an increase in the size and complexity of experimental data sets, which are not always easy to understand and interpret. One example is the genome wide and high-throughput variant of chromosome-conformation-capture, known as Hi-C (for a review of this and other 3C methods see [1]). This method provides a population level read out of how likely two genomic loci are in spatial proximity. It has taught us that interactions within the genome are not random but are instead highly structured. At large scales, interactions between active chromatin regions and associations between inactive chromatin regions are enriched compared to active-inactive interactions—i.e. there is a segregation of the genome into compartments[2]. At shorter length scales, chromosomes are partitioned into contiguous regions which show enriched self-interactions, known as domains or TADs (topologically associating domains)[3]; many domains are associated with chromatin loops (between, for example, binding sites of the CCCTC-binding factor, CTCF)[4]. More recently, higher resolution data sets have revealed features such as "stripes" of enriched interaction emanating from super enhancers[5,6] (also called "frequently interacting regions" or FIREs). As experiments have become more quantitative and data sets larger, researchers are turning to methods common in other data-rich fields: for example, statistical modelling, machine learning, and – the subject of this perspective – mechanistic modelling. Mechanistic models, which seek to understand *why* a system behaves as it does, are common in the physical sciences; they are often based on simplified mathematical or computational representations of a system, fitting together individual components to study emergent behaviour.

Here we consider mechanistic modelling of chromatin structure and function, focussing on approaches which originate from soft matter, and polymer physics. Soft matter physics is the study of "soft" materials such as gels, colloids and polymers; materials often found in biology – in tissues, in the cytoplasm, or in the nucleus. The defining feature of soft matter systems is that their macroscopic behaviour originates from their mesoscopic properties: behaviour is governed by thermal fluctuations, self-organisation and entropy. Importantly the properties of these systems often do not depend on precise microscopic or molecular details, which enables simplified "coarse-grained" modelling approaches[7]. While molecular details are of course crucial for biological function (and active processes

can introduce phenomena, which transcend equilibrium thermodynamics), the mesoscale approach has often proved fruitful[8]. A reductionist approach and "search for universality" are common in physic; this may seem at odds with the view from biology, where frequently "the devil is in the detail". Nevertheless, there is a growing consensus that combining approaches, and a shared way of thinking, can provide understanding of molecular processes.

With mechanistic modelling come mathematical equations, numerical methods, and computer simulations – approaches less common in biological sciences. In this perspective we aim to demystify these concepts in the context of chromatin folding, chart progress to date, and give a perspective on what the future holds.

The approach to modelling: backwards or forwards.

When modelling the three-dimensional (3-D) structure of chromosomes there are two commonly used approaches: inverse or "fitting" based models and forward or "mechanistic" models. In the inverse case, you start with experimental data, such as chromosome interactions e.g. from 5C or Hi-C, and use that information to reconstruct the spatial arrangement of the chromosome in 3-D, via a fitting or iterative procedure. Early attempts used the data to infer a set of constraints on the separation of different genomic loci, using those as a basis for statistical or matrix methods to generate a single "average" chromosome conformation[9]. More recently, methods from polymer physics – where the chromatin fibre is represented as a connected chain of interacting units – have been applied[10,11]; the interaction parameters are then determined using an iterative procedure which improves the fit of the model. Usually, the polymer approaches generate a population of 3-D conformations, and in some cases once an optimal parameter set is found, the same model can be applied to different chromosomes or cell types[11].

Importantly, with inverse modelling one starts with an experimental data set and is interested in generating structures consistent with that data. Though the approach can successfully reproduce, for example, Hi-C interaction maps, it often reveals little about the microscopic mechanisms behind the observed structures. In mechanistic, or forward modelling the aim is different. Instead of reproducing experimental data, there is a desire to understand the mechanisms behind the observations, e.g.

addressing questions such as how does a chromosome segregate into an active and inactive compartment, or how do chromatin domains form? Mechanistic models usually start from first principles or hypotheses and ask whether different microscopic possibilities can generate the observed behaviours. A common approach (which is prevalent in the physical sciences) is to start with a simple description of the system (i.e. a minimal model), and then systematically add detail until the model reproduces the experimental observations. Such models can also incorporate experimental data, but the "output" will usually be a different kind of information from the "input". While in the inverse modelling case the input might be a Hi-C map, and the output will be a reproduced Hi-C map along with a 3-D structure, a mechanistic model might use protein binding data together with some hypothesised "rules" to generate 3-D structures and a simulated Hi-C map (but, Hi-C is *not* used as an input). In this way forward modelling is truly predictive, but the closeness of the prediction is unlikely to be as good as a fitting based model with many (or hundreds) of parameters; nevertheless, they have been hugely informative and provide an unique understanding of molecular mechanisms.

Inverse modelling can be a useful tool – especially when there is a lot of existing data. For example, early studies revealed the importance of loops[12], and that the observed interaction maps can be consistent with a population of chromosome structure which has a high degree of variability[10]. However, to understand what mechanisms are at play, forward modelling can often provide addition insights into the experimental system.

Dynamical Polymer Models: Key Concepts and Ideas

Most mechanistic modelling applied to chromatin has used simple models derived from polymer physics and molecular dynamics (MD) simulations[13]. MD is a scheme where the motion of atoms or molecules are simulated within the computer. The principle behind the method is simple: atoms move according to Newton's second law, which states that the force exerted on a body is equal to its mass multiplied by its acceleration (Figure 1A). This equation of motion is solved to simulate how the positions of all the atoms in a system change with time. After deciding where to initially place the atoms, the remaining problem is to calculate the forces which each atom will experience. In practice this can be challenging: each atom will experience forces due to interactions with all of the surrounding atoms (Coulomb

interactions due to charge, van der Waals interactions etc.); nevertheless, it has been possible to write down a set of interaction potentials (based on calculations from quantum chemistry, collectively described as a "force field") which results in realistic behaviour. Such "all-atom" MD simulations have been applied to proteins, as well as DNA and DNA-protein interactions. However, calculating the forces (including those between solvent atoms such as water and salt ions etc.) is computationally expensive, and typically this can only be done for a short DNA segment or a single protein subunit (for example it currently takes around a day to simulate the behaviour of a protein consisting of 142,000 atoms for 100 ns on 256 compute cores, or about 0.3 ns per day per core).

To simulate larger experimental systems, one must adapt the method – one way to do this is by removing the solvent from the simulation. Instead of explicitly treating all of the water molecules their effects can be included in an implicit way. Common methods include Brownian or Langevin Dynamics; the simplest method is to include only the viscous drag and "thermal jostling" which water provides, and done correctly this leads to the thermal diffusion properties of the system being accurately represented (Figure 1Bi and see Box 1 for details). More complicated approaches might include hydrodynamic interactions or to include the effect of salt ions one might need to modify the interaction force field, e.g. to account for electrostatic screening.

Even after removing the solvent, the size of systems which can be simulated with an all-atom treatment remains small. And anyway, keeping track of the position of every atom within a chromosome is unlikely to be informative. To circumvent these problems large molecular systems simulations are often coarse-grained (CG): collections of atoms or molecules are replaced by larger and simpler objects but the same method of solving Newton's law with an implicit solvent can be used (Figure 1Bii-iii). One must also have a scheme for determining the forces that the simplified objects will exert on each other – these interaction potentials are often phenomenological in nature (i.e. one uses a set of interactions which gives rise to the correct macroscopic behaviour, but which need not be microscopically realistic). Developing a CG model is no easy task; deciding on how much detail to include in a model in order to learn something meaningful is an art. This of course depends on the question (see Table 1): to investigate how a polymerase bends and twists DNA during transcription it might be important to model the double helix of the DNA[14,17]; on the other hand, to understand how a chromatin fibre is formed, perhaps the double helix structure is less important, but a description of the nucleosome needs to be included[18]. To

model a whole chromosome or chromosome region, most work has used a much lower level of detail where a chromatin fibre is represented by a connected chain of beads, with each bead representing several thousand base-pairs of DNA - it is these large-scale simulations which have been informative for understanding chromatin organisation.

While a chain of beads might sound like a crude model for chromatin, it has been remarkably successful in explaining many of the observations from experiments like Hi-C. This level of detail also allows concepts from polymer physics to provide novel insights. The key idea is that if a polymer is represented as a chain of connected units which are otherwise free to move in space, this then looks like a random walk (or a self-avoiding walk), which has well known statistics[7], that can be used to make predictions. Thus, starting from this knowledge, questions such as "*On average how much space will a polymer of given length take up?*" or "*What is the probability that the polymer will be forming a loop?*" can be addressed. Complexity can then be added to the description, for example by considering the flexibility of a polymer, or the solvent conditions. These statistical methods have been applied, for instance, to predict how the average level of Hi-C interactions decay with genomic separation[19], or how likely chromatin loops of certain length form. Often it is when an observed quantity *does not* fit the statistical description that we learn something – if the statistics of a free polymer predict one result, but something else is observed (either experimentally or computationally), there must be some mechanism we have not considered.

To understand how polymer simulations are performed, some practical details often give context. MD simulations and their variants are mature methods in the physics community and there are many software packages. The "bead-chain" simulations described below are often performed using very general and adaptable open-source MD codes developed with the programming savvy user in mind. The advantage of such software is that after many years of development they run efficiently on the latest computer hardware, are highly scalable when run on computer clusters, and have a broad user base who provide updates and new optimisations. In Box 2 some of the most popular software packages are listed, and outline how these are used and on what type of computer hardware.

Evolution and success of polymer models

Recent mechanistic models, which treat chromatin as a simple polymer, have improved the interpretation of results from Hi-C and other experiments.

Some of the first polymer simulations to study chromatin at a genomic scale considered the role of entropic effects in chromosome positioning, showing that chromatin fibre properties such as flexibility can control radial positioning of chromosomes or chromosome regions[20]. Other work used simulations and polymer theory to estimate time-scales for chromosome dynamics – revealing that the time required for two human chromosomes to become intermingled is of the order of 100s of years, putting into new light observations of chromosome territories, and how individual loci can be highly dynamic (e.g. structural changes after an inflammatory response can be observed within as little as 30 mins), while whole chromosomes appear relatively static (remaining in territories for the duration of the cell cycle[21]).

A more recent model which considered the formation of chromosome domains was the "strings-and-binders-switch" model[22]. This took a bead-chain polymer model and introduced generalized "binders": single beads which represented chromatin-binding protein complexes which form bridges between chromatin regions. That work revealed that bridging binders can lead to chromatin interactions with the right scaling of interactions as a function of genomic separation, and provides a mechanism for different levels of compaction in different chromatin regions. Around the same time, we also developed a bead-chain model with diffusing bridges (or "factors")[23]; by performing simulations of large chromosome regions (up to whole chromosomes) at a higher resolution (one to three kbps of DNA per bead) we uncovered a concept we called "bridging-induced attraction" – see Figure 2A. Mechanistically this describes a tendency for protein complexes which can form molecular bridges between multiple regions of DNA or chromatin (stabilising loops) to cluster together, even in the absence of interactions between the complexes. The attraction arises due to a positive feedback where the first loop to form creates a local increase in DNA density which promotes the binding of further proteins in that region, which further increases density, etc. The resulting cluster formation can be thought of as phase separation (or, more precisely, micro phase separation, as multiple clusters remain in steady state) – a mechanism now thought to be important for the formation of many nuclear structures[24]. Later work showed that using multiple species of bridges/factors and patterning the bead-chain with binding sites for these factors could drive the polymer into specific 3-D structures[25]. Using experimental data to position the binding sites it is possible to predict the Hi-C maps observed in experiments (Figure 2A top right; and since this

is a forward modelling scheme, it is truly a prediction as Hi-C data was not an input to the model). Rather than using specific known species of protein or protein complex, the factors are assumed to be generic chromatin binders, and then rather than identifying precise binding sites, data such as ChIP-seq for different histone modifications are used to identify broad regions of binding. With this "simplistic" model[25] it was found that with only two bridge species, an active factor (perhaps representing polymerase/transcription factor complexes), and a repressive factor (perhaps polycomb complexes, or HP1), remarkably it was possible to reproduce the interactions *vs* genomic separation scaling and predict the locations of 85% of chromatin domain boundaries.

As well as being able to generate interaction maps, these "transcription factor" (or "diffusing bridge") models give predictions on spatial organisation and dynamics of bridge complexes. The protein clusters which arise in the simulations look similar to some of the phase separated membrane-less organelles. For example, clusters of polycomb like proteins resemble polycomb bodies, and clusters of activating proteins can be thought of as transcription factories[30]. However, a more detailed inspection of the simulations showed that the dynamics of the protein clusters are not the same as nuclear bodies in an important respect: once formed the clusters are very stable, and proteins are not "turned over" (exchanged with a soluble pool). A refined model[31] adds a feature where the proteins stochastically switch from a binding to a non-binding state. This could represent post-translational modifications, active protein degradation, or programmed polymerase unbinding after transcription termination. This leads to more dynamic nuclear body-like clusters where proteins turn-over while the body retains its shape and size (Figure 2A, bottom right). Importantly this drives the system away from equilibrium (switching represents active chemical reactions which hydrolyse ATP) and provides a mechanism through which the cell can control protein clustering and concomitantly phase separation. The switching model is an example where a discrepancy between the original simulations with experimental observations led to model refinement and improved understanding.

Active protein unbinding is just one example of an out-of-equilibrium process which affects genome organisation. Biological systems are inherently away from equilibrium, as they take in energy from their surroundings to drive internal processes. Including out-of-equilibrium processes is therefore often important; a challenge in coarse-grained modelling is to recognise when this additional ingredient is necessary to explain the behaviour of the system. Active processes break "detailed balance" which can

dramatically effects macroscopic behaviour. At equilibrium, as each process in the system must be balanced by its reverse process, a movie of a simulation looks essentially the same when played forwards or backwards: this time-reversal symmetry is typically lost when models include active out-of-equilibrium processes. Another important consideration is that as the cell exits mitosis, the compacted mitotic chromosomes expand into their interphase configurations – they relax towards a new equilibrium state (but might not reach that state within biologically relevant time scales). This means that when running a simulation without an active process, one must consider carefully whether an equilibrium condition has been reached, whether such a state is relevant, and whether the outcome of the simulations will be affected by the initial configuration.

Other recent work[32] showed that to reproduce many of the features of a Hi-C map, it is not necessary to explicitly model the diffusing proteins – instead a direct attractive interaction between the polymer beads is sufficient. This could represent either bridging factors which are already bound to the chromatin, or direct chromatin-chromatin interactions mediated by charges on the nucleosome surface or histone tail interactions. While those models (usually known as "block copolymer models") can give good prediction of Hi-C maps (for example in *Drosophila*), obviously without an explicit representation of the bridging proteins, they cannot give any information about protein foci dynamics.

Another well-known model based on ideas developed through polymer physics is "loop extrusion"[26,27]. The loop extrusion model was first invoked to explain the puzzling observation that long-range interactions between binding sites of the CCCTC-binding factor (CTCF) – which has a binding motif with a specific direction on the DNA – tend to be found when the binding motifs have a specific "convergent" orientation[4]. Such a strong bias is difficult to reconcile with loops that form due to two sites diffusing into contact. Instead, the model proposes that a factor binds at some point between the CTCF sites and extrudes the loop outwards (Figure 2B). There is growing evidence that the SMC protein cohesin is involved in this extrusion process, but it is still unclear if a unidirectional "motor" is required to push the chromatin into loops[26,27], or if diffusive sliding of cohesin is sufficient[28]. Nevertheless, the loop extrusion concept has been successful, providing clear qualitative explanations for experimental observations. It explains changes in chromatin interactions resulting from "genome editing" experiments which manipulated CTCF binding sites[26], it also explains changes observed in Hi-C data when CTCF[33], cohesin[34,35], or factors involved in cohesin loading and unloading[36,37] are knocked out. It is also thought

that an extrusion mechanism, this time involving condensin, is involved in chromosome compaction during mitosis[29,38].

There is growing evidence that in reality, chromosome organisation is driven *both* by bridging protein complexes, and some form of extrusion, and combination models have now been developed[39]. While extrusion can explain CTCF looping, it does not provide a mechanism for compartment formation; conversely, a bridging mechanism readily explains compartments but cannot generate CTCF loops with a motif direction bias. Together, these two physics-based models explain many of the observations from Hi-C experiments, including compartments, domains and loops. Particularly, the fact that compartment patterns remain, but (loop-mediated) TADs are lost from Hi-C maps when cohesin (or its loading factor) is removed suggests that these features arise through different mechanisms[34,35].

Chromosome folding at the gene-scale

The polymer models described above have mainly been concerned with chromosome organisation at large scales, 100s of kilo-base to mega-base. In contrast, other efforts have focussed on an intermediate scale, studying the looping and folding of chromatin around specific genes and gene loci, which consider cis-regulatory promoter/enhancer interactions explicitly. Experimentally, higher resolution chromosome population level interaction data can be obtained using methods such as 4C or Capture-C[40], and single cell information can be collected through microscopy techniques such as fluorescent in-situ hybridisation (FISH). In simulations greater levels of detail can be probed using different coarse-graining, by using more "beads" to represent the same length of chromatin.

Our work simulating looping of mouse globin genes[41] revealed that to capture higher-resolution interactions it is necessary to use different input data for the polymer model. Although using histone modification data to infer protein binding can give good predictions of domains and compartments, more specific binding site placement is required to predict promoter-enhancer contacts. For example, ChIP data for transcription factors can be used, but we observed that Capture-C results could be predicted using only DNA accessibility information derived from DNase- or ATAC-seq experiments. Importantly, these protein-binding driven simulations could generate an ensemble of locus conformations which were largely consistent with both population (Capture-C) and single cell (FISH)

data. An important feature of these simulations is that the full details of each locus conformation within the population are retained, allowing the variability of the locus structure to be examined in a way that is still not possible experimentally. For the globin genes, analysis of the simulations implied that these loci tend to organise into one of a small number of possible structures, for example, forming a single compacted globule, two separated globules, or more extended shapes.

Applying a similar model (but now incorporating diffusing bridges and loop extruders) to the developmental gene *Pax6* in cell lines where the gene has different levels of transcriptional activity revealed different behaviour[42]. First, this model gave good predictions of interactions at the population level (simulating Capture-C data), but it failed to correctly predict the shape of the locus at the single cell level (simulating FISH data). Experimentally when *Pax6* was transcribed at a low to moderate level, the locus was more compact compared to the case where the gene was inactive (the separation between the *Pax6* promoter and its enhancers decreased). Surprisingly in a cell line where *Pax6* was highly active, the locus became more expanded (with separations between the promoter and one enhancer significantly increasing). This variation was not correctly predicted by the simulations; again, and importantly, the failure of the model led to new ideas for how the chromatin within the locus changes in these different cell types. The data showed that although large regions of the locus gained a histone acetylation mark associated with active enhancers (H3K27ac), this was not accompanied by an increase in looping between those regions and the promoters (as expected from the classic model of enhancer action). We reasoned that this mark was instead associated with some local change in the properties of chromatin which led us to a new "heteromorphic polymer" or HiP-HoP model[42] (Figure 3A). We hypothesised that the acetylation mark corresponded to regions of the chromatin with a less compact internal structure (an idea previously suggested by experiments such as RICC-seq[43]; indeed a simulation model which incorporates a fibre with varying linear compaction (different regions have a different amount of DNA per unit length) was able to reproduce all of the trends observed in the Capture-C and FISH data (Figure 3B-C). As the HiP-HoP model is predictive, it can now be used to analyse the structure of many other genes across different species (e.g. *SOX2*) to understand function. For example, the simulations of *Pax6* revealed a much larger variation of structure within a population of the same cell type compared to the case of the globin genes, suggesting that regulation of different genes involves alternate structural mechanisms.

An important point which arose from the *Pax6* study is that the classic model for enhancer-promoter interaction does not always seem to be at work. In comparing the inactive and low expression states, the distal *Pax6* enhancers gain histone acetylation marks, and physical interactions with the promoters are observed – as expected. Moving to the high expression cell line, for one enhancer the acetylated region broadens, but the frequency of interactions with the promoters *decreases* (and their separation *increases*); either this site does not have an enhancer action in these cells, or it functions by a mechanism other than physical contact with the promoter. At any rate, the action is different for different expression states. On a more technical level, we note that the simple nature of these models allows the large-scale behaviour of the chromatin to be simulated, without needing a detailed knowledge of the molecular details, so questions such as, is it a 30-nm or 10-nm fibre, a one-start or two-start helix, do not have to be addressed.

Outlook

There is a growing body of work using mechanistic polymer models to understand chromosome organisation and function. Importantly, it is clear that these methods are not only useful when applied to data after experiments have been completed, but that simulations can also be used to test new ideas, to uncover new mechanisms, and to drive new experimental studies. A good example of this is the recent work on loop extrusion: initially this was a theoretical endeavour – one of the first published studies was entirely computational[27] – but it rapidly prompted further experimental studies. In physics it is commonplace that computational work is conceptually ahead of what can be realised experimentally, and we believe that in the future this will become more prevalent in chromatin biology: in other words, we expect more and more often to see simulation work which does not follow, but instead provides new hypotheses that drive experiments.

What does the future hold for polymer simulations in chromatin biology? Our recent HiP-HoP model suggests that if we want to study gene loci in more detail, we need models that resolve some of the structural properties of chromatin (and how that varies at different locations). There are models which have detailed coarse-grained representations of nucleosomes[16,18]– this approach has recently been used to study the *HOXC* locus[44], which revealed that epigenetic factors play an important role in larger-scale

locus folding. But these detailed simulations are computationally expensive, and tend to be limited to small fibre sections – there is currently no model which resolves nucleosomes simply enough to allow simulation of large gene loci. If such a model were developed it could, for example, be used to study how specific patterns of nucleosome spacing around regulatory elements affects the 3-D fibre structure, or to better understand data from new experimental techniques such as ChromEMT[45] (where electron microscopy is used to image nucleosomes *in vivo*).

So, does progress equal developing models with more and more molecular detail? Not necessarily – the level of detail required in a model really depends on the questions, and the size of the system being investigated (see Table 1). For example, recent studies using less detailed models to examine compartmentalization and the global organisation of (hetero)chromatin within the nucleus have revealed that interactions between chromatin and the nuclear lamina play a key role, e.g. in the inverted structure observed in rod cells in nocturnal mammals[46], and in structural changes during senescence and ageing related diseases[47]. Similar low-resolution models might be appropriate to study how chromosomes are compacted during mitosis[38], or to study the kinetics of chromatid segregation during anaphase. Models with different scales and levels of detail might also be useful for gaining a more general understanding of mechanisms or phenomena such as DNA supercoiling, liquid-liquid phase separation of chromatin associated proteins, or active processes which occur within the nucleus.

Over recent decades polymer and coarse-grained or “mesoscale” simulations have proven useful for studying soft matter physics. They are starting to be used in biological sciences, and we expect such methods to become commonplace for chromatin biology. What remains challenging is developing models which incorporate pertinent molecular detail at a short length scale, while addressing large-scale behaviour and simulating physiologically relevant time scales. These exciting new applications for mesoscale modelling are therefore likely to drive new developments in multi-scale simulation methods, where systems are simultaneously modelled with different levels of detail, with results from one model feeding into another.

Acknowledgements

The authors would like to thank members of their groups for stimulating discussions. Research in the Marenduzzo group is supported by the European Research Council (CoG 648050, THREEDCELLPHYSICS); research in the Gilbert lab is funded by the UK Medical Research Council (MR/J00913X/1 and MC_UU_00007/13).

Author Contributions

All authors conceived and wrote the manuscript

Ethics Declaration

The authors declare no competing interests.

References

1. Han, J., Zhang, Z. & Wang, K. 3C and 3C-based techniques : the powerful tools for spatial genome organization deciphering. 1–10 (2018).

2. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-.).* **326**, 289–293 (2009).

3. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

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

5. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).

6. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **173**, 1165--1178.e20 (2018).

7. van der Maarel, J. R. C. *Introduction to Biopolymer Physics*. (World Scientific Publishing Company, 2007).

8. Ingólfsson, H. I. *et al.* The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 225–248 (2014).

9. Serra, F. *et al.* Restraint-based three-dimensional modeling of genomes and genomic domains. *FEBS Lett.* **589**, 2987–2995 (2015).

10. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).
This paper presents one of the first studies using polymer physics based models together in an iterative parameter finding procedure which produces a population of chromosome structures from 5C data.

11. Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. Transferable model for chromosome architecture. *Proc. Natl.*

Acad. Sci. USA **113**, 12168–12173 (2016).

12. Baù, D. *et al.* The three-dimensional folding of the $α$-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **18**, 107–114 (2011).

13. Frenkel, D. & Smit, B. *Understanding Molecular Simulation*. (Elsevier, 2001).

14. Fosado, Y. A. G. *et al.* Soft Matter A single nucleotide resolution model for large-scale simulations of double stranded DNA. *Soft Matter* **12**, 9458–9470 (2016).

15. Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138–141 (2005).

16. Schlick, T., Hayes, J. & Grigoryev, S. Toward convergence of experimental studies and theoretical modeling of the chromatin fiber. *J. Biol. Chem.* **287**, 5183–5191 (2012).

17. Ouldridge, T. E., Louis, A. A. & Doye, J. P. K. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **134**, 85101 (2011).
This paper gives the first presentation of the now widely-used 'oxDNA'. This coarse-grained model resolves DNA at the base-pair level with the full double helix structure and sequence specificity.

18. Arya, G. & Schlick, T. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. USA* **103**, 16236–16241 (2006).

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

20. Cook, P. R. & Marenduzzo, D. Entropic organization of interphase chromosomes. *J. Cell Biol.* **186**, 825–834 (2009).

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

22. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. USA* **109**, 16173–16178 (2012).

23. Brackley, C. A., Taylor, S., Papantonis, A., Cook, P. R. & Marenduzzo, D. Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. USA* **110**, E3605--11 (2013).
This study presents simulations which uncovered the 'bridging-induced attraction' mechanism which can drive formation of foci DNA-binding protein. This has become increasingly important in light of new data showing many protein foci have 'liquid droplet' like properties.

24. Erdel, F. & Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophys. J.* **114**, 2262–2270 (2018).

25. Brackley, C. A., Johnson, J., Kelly, S., Cook, P. R. & Marenduzzo, D. 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).

26. Sanborn, A. L. *et al.* 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).

27. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
This paper provides a detailed study of the popular loop extrusion model. This mechanism is now thought to play a major role in chromosome organization both during interphase and mitosis.

28. Brackley, C. A. *et al.* Extrusion without a motor: a new take on the loop extrusion model of genome organization. *Nucleus* **9**, 95–103 (2018).

29. Goloborodko, A., Marko, J. F. & Mirny, L. A. Chromosome Compaction by Active Loop Extrusion. *Biophys.J.* **110**, 2162–2168 (2016).

30. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (80-.).* **361**, 412–415 (2018).

31. Brackley, C. A. *et al.* Ephemeral Protein Binding to DNA Shapes Stable Nuclear Bodies and Chromatin Domains. *Biophys.J.* **112**, 1085–1093 (2017).

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

33. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930--944.e22 (2017).

34. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305--320.e24 (2017).

35. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).

36. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693--700.e14 (2017).

37. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).

38. Gibcus, J. H. *et al.* A pathway for mitotic chromosome formation. *Science (80-.).* **359**, (2018).

39. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. USA* **115**, E6697--E6706 (2018).

40. Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212 (2014).

41. Brackley, C. A. *et al.* Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.* **17**, 59 (2016).

42. Buckle, A., Brackley, C. A., Boyle, S., Marenduzzo, D. & Gilbert, N. Polymer Simulations of Heteromorphic Chromatin Predict the 3D Folding of Complex Genomic Loci. *Mol. Cell* **72**, (2018).
This paper presents a new model for predicting chromatin organization at the gene locus level. It combines mechanisms from several previous studies with a heteromorphic polymer description of chromatin.

43. Risca, V. I., Denny, S. K., Straight, A. F. & Greenleaf, W. J. Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature* **541**, 237–241 (2016).

44. Bascom, G. D., Sanbonmatsu, K. Y. & Schlick, T. Mesoscale Modeling Reveals Hierarchical Looping of Chromatin Fibers Near Gene Regulatory Elements. *J. Phys. Chem. B* **120**, 8642–8653 (2016).

45. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science (80-.).* **357**, (2017).

46. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).

47. Chiang, M. *et al.* Polymer Modeling Predicts Chromosome Reorganization in Senescence. *Cell Rep.* **28**, 3212--3223.e6 (2019).

48. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

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

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

51. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

52. Hinckley, D. M., Freeman, G. S., Whitmer, J. K. & de Pablo, J. J. An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139**, 144903 (2013).

53. Brackley, C. A., Morozov, A. N. & Marenduzzo, D. Models for twistable elastic polymers in Brownian dynamics, and their implementation for LAMMPS. *J. Chem. Phys.* **140**, 135103 (2014).

54. Bascom, G. D., Myers, C. G. & Schlick, T. Mesoscale modeling reveals formation of an epigenetically driven HOXC gene hub. **116**, 4955–4962 (2019).

55. Wocjan, T., Klenin, K. & Langowski, J. Brownian dynamics simulation of DNA unrolling from the nucleosome. *J. Phys. Chem. B* **113**, 2639–2646 (2009).

56. Brackley, C. A., Allan, J., Keszenman-Pereyra, D. & Marenduzzo, D. Topological constraints strongly affect chromatin reconstitution in silico. *Nucleic Acids Res.* **43**, 63–73 (2015).

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

58. Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 1–8 (2016).

59. Racko, D., Benedetti, F., Dorier, J. & Stasiak, A. Are TADs supercoiled? *Nucleic Acids Res.* **47**, 521–532 (2019).

60. Barbieri, M. *et al.* Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **24**, 1–13 (2017).

61. Serra, F. *et al.* 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).

62. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).

63. Bianco, S. *et al.* Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* **50**, (2018).

64. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).

Figure 1: Coarse-grained (CG) molecular dynamics (MD) simulations of chromatin. **A.** Schematic description of classic MD. All atoms in the system are represented. A set of interaction potentials are used to calculate the force exerted on each atom by others. Dynamics are then simulated by solving Newton's second law. **B.** A simplified description of the system reduces computational overhead. (i) Solvent atoms (red and grey in left diagram) are removed, and their effect provided implicitly by adding random force "kicks" to the solute atoms (represented by arrows in right diagram, also see Box 1). (ii) Coarse-graining is where collections of atoms are replaced by simpler objects. Left diagram shows an all-atom representation of a DNA molecule; on the right a CG model represents each nucleotide as a bead with a patch[14]. (iii) Chromatin can be coarse-grained at different levels depending on the question being asked (see Table 1). Left diagram shows an all-atom representation of two nucleosomes (based on the crystal structure from Schalch et al.[15]). Middle shows a chromatin fibre model where each nucleosome is represented by a solid body (modified from Schlick et al.[16]), whilst the right-hand diagram shows a bead-chain model where the internal structure of the fibre is not resolved.

Figure 2: Common models for understanding chromosome organisation. **A.** Left: Diffusing protein bridges stabilize loops and domains in chromatin or DNA, leading to bridging induced attraction (even in the absence of protein-protein interactions). Right top: specific patterns of binding sites on chromatin lead to domains and compartments (predicting Hi-C maps from protein binding data). Right bottom: ATP-driven chemical reactions altering bridging affinity lead to clusters with dynamics similar to nuclear bodies. **B.** The loop extrusion model[26,27] explains chromatin loop domains and the CTCF motif directionality bias. Extrusion could be an active "motor" effect, or diffusive[28] and a similar mechanism might drive chromosome compaction during mitosis[29].

Figure 3: A gene locus model: the highly predictive heteromorphic polymer model (HiP-HoP)[42]. **A.** Schematic showing the HiP-HoP model ingredients – a variable thickness fibre is combined with diffusing bridges and loop extrusion. **B-C.** This model was used to study the *Pax6* locus in three cell lines where *Pax6* is expressed at different levels; it predicts both population level chromatin interactions (Capture-C) and single cell microscopy (fluorescence in situ hybridization). Further insight can be gained by a more detailed study of the structures predicted by the simulations, or by editing the input data to perform *in silico* mechanistic experiments.

Box 1: Langevin dynamics for modelling molecular motion

As detailed in the text and in Fig. 1, running a molecular dynamics simulation essentially boils down to solving Newton's second law (the "equation of motion" or $F = ma$) for each atom (or coarse-grained (CG) object) in the system. To simplify the simulations the full details of the solvent can be neglected – a common scheme for doing this is Langevin dynamics, in which the effect of the solvent is approximated by adding two new forces to the equation of motion. Here we look at this in more detail. The equation of motion for atom i is

$$m\frac{d^2\boldsymbol{r}_i}{dt^2} = \boldsymbol{F}_i - \xi\frac{d\boldsymbol{r}_i}{dt} + \sqrt{6\xi k_B T}\boldsymbol{\eta}_i(t),$$

where $\boldsymbol{r}_i$ is the vector position in space of atom i. The left-hand side of the equation is mass times acceleration (the second derivative of position with respect to time). The three terms on the right are the forces experienced by atom i, and discussed below. To perform a simulation, this equation is solved numerically – by imagining that time evolves in discrete steps, we calculate the forces on the atom at one time point, and this equation tells us how those lead to a change in the velocity and position of the atom at the next time point. Each of the simulation software packages mentioned in the main text essentially solves this equation for all of the atoms (or CG objects) in the system.

The three force terms:

$\boldsymbol{F}_i$ This is the force experienced by "atom" (or bead) i due to interactions with all other atoms in the system. For an atomistic simulation this is found from a complex set of interaction potentials derived from quantum chemistry; for a CG simulation it may be a set of simplified phenomenological potentials. For a typical CG polymer model, interactions could include a potential to keep beads connected in a chain, a potential preventing beads from overlapping in space, and a potential giving rise to a polymer bending stiffness. This term could also include any external force applied on bead i.

$-\xi \frac{d\boldsymbol{r}_i}{dt}$ This is the first of two terms which approximate the effects of the solvent. It represents the viscous drag experienced by the bead as it moves, which is proportional (and in the opposite direction) to its velocity; ξ is a "friction" parameter related to the viscosity of the fluid.

$\sqrt{6\xi k_B T}\,\boldsymbol{\eta}_i(t)$ This second solvent term approximates thermal "jostling" due to solvent molecules. The symbol $\boldsymbol{\eta}_i(t)$ represents a random "kick" of force bead i receives at time t. There is a well-defined mathematical description of this "noise", but in a simulation context it amounts to generating a random number at each time step – which introduces stochasticity. The pre-factor ensures that the equation obeys the fluctuation dissipation theorem: in the context of Brownian motion there is a relationship between the viscous drag experienced by an object being pulled through a fluid (dissipation) and its diffusive motion (fluctuations).

This simple scheme neglects affects like hydrodynamics interactions (where fluid flows set up by motion of one object result in forces on another). Including these would require significant computational overhead, and commonly it is thought that they should not play a big role in the densely packed nuclear environment.

Box 2: Practicalities of Polymer Simulations

As detailed in the main text, most polymer-based molecular dynamics (MD) simulations of chromatin are performed using general-purpose codes. Common software packages include: LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator[48]) which is optimized for use on multicore computer clusters and is an adaptable and expandable code written in C++; HOOMD[49], which uses the Python scripting language and is optimised to run on GPUs; and ESPResSo (Extensible Simulation Package for Research on Soft Matter[50]), which also uses Python and was developed for soft matter physics. These packages come with extensive documentation, and tutorials are often a good starting point for new users (e.g., see cbrackley.github.io/simple_lammps_tutorial). The codes perform the simulations, and output "trajectories", i.e. details of the positions of atoms (or CG objects) as a function of time; since an implicit solvent simulation is stochastic, many such simulations can be performed to generate an ensemble of trajectories (representing e.g. a population of cells). Typically, users write further programs or scripts to take measurements from these trajectories which are compared to experimental measurements (e.g. one might measure the separation of CG objects, diffusion constants, or how often two objects are found together within a population of trajectories). Other software tools are used to visualise trajectories including Visual Molecular Dynamics[51] (VMD), a popular tool for generating images and animations.

All of the codes mentioned can be compiled to run on any standard Unix-based system (e.g. Mac or Linux). Though most software can run on any size of machine, from laptop or desktop up to large multi-core supercomputers, the size and scale of most studies necessitates use of multi-core machines (or alternative architectures like GPUs or other coprocessors). A typical simulation-based study might need thousands of CPU hours-worth of compute time and generate GBs of data. This requirement for specialist high-performance computing hardware means that these simulations should be viewed as "*in silico* experiments" in that – just as a wet-lab based project – specialist equipment, expertise, technicians and consumables (here compute time and data storage) are required.

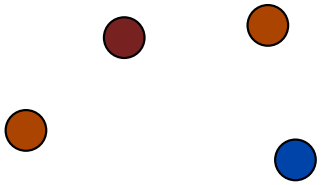
Table 1. Table showing how different coarse-grained polymer simulation models can be, or have been, used to study different biological questions. Different applications typically require models with different levels of detail, and this table provides a list of relevant models (although not exhaustive) for a set of specific biological questions. We separate "mechanistic" models, which are used to test hypothesis on underlying mechanisms, and "inverse" models, where data are used to, e.g., infer chromatin structures consistent with these.

Mechanistic (forward) models:		
Aim	*Examples*	
Understand microscopic DNA properties, in vitro experiments, or biotechnology applications.	These often involve CG models of DNA which resolve the DNA double strands and superhelical structure, e.g., the oxDNA model[17], or three-spheres-per-nucleotide (3SPN) model[52]. These can be used to study supercoiling[14], DNA melting, and DNA-protein interactions. Larger systems can be treated using simpler models which track twist deformations without resolving individual DNA strands[53].	
Understand chromatin structure at the nucleosome level.	Detailed CG models based on nucleosome crystal structures can be used to simulate small fibres[18,54]. Simpler models representing nucleosomes as disks or spheres have been used to simulate, e.g. in vitro nucleosome unwrapping[55], chromatin reconstitution[56], or micro-domains in yeast[57].	
Understand mechanisms of chromosome compartments and domains.	Simple bead-and-spring polymer models for chromatin (where nucleosomes are not resolved) can be used to investigate mechanisms for the formation of compartments and domains in, e.g. Drosophila[32], mouse[58], and humans[25].	
Understand mechanisms for chromosome loops and loop domains.	Simple coarse-grained bead-and-spring polymer models can be used to study these questions. On the basis of such studies, several different mechanisms have been put forward for the formation of cohesin and CTCF mediated chromatin loops, and loop domains. These include supercoiling[59], and loop extrusion[26–29].	
Predict the detailed structure of gene loci	As well as helping us to understand mechanisms, simple coarse-grained bead-and-spring polymer models can also be predictive, using some experimental data as an input. Unlike the inverse models detailed below, they do not involve fitting. The HiP-HoP model[41,42] is an example: data on DNA accessibility is an input, and simulated Hi-C/Capture-C is the output (see also ([60])).	
Understand chromosome organisation at the whole nucleus level	By reducing the level of detail (e.g. representing large chromosome regions as a single bead), whole chromosomes[20,21], or even whole nuclei can be simulated. This has been used, e.g. to study the effect of interactions between chromatin and the nuclear lamina[46,47].	
Inverse Models:		
Aim	*Examples*	*Input data*
Generate in silico configurations of a gene locus, or whole chromosome consistent with 3C-based data.	Two approaches to reconstruct of chromosome configurations from data are restraint-based and polymer-based models. Examples of the former include the TADbit software, which has been used to reconstruct gene loci in mammalian cells[12,61]; similar methods have been applied to reconstruct whole yeast nuclei[62]. In polymer-based methods, iterative parameter determination has been combined with both Monte Carlo[10] and	*Hi-C, 5C or single cell Hi-C*

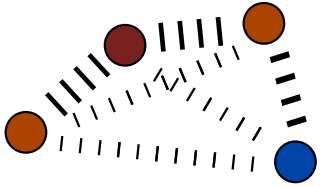
	Molecular Dynamics simulations[11,63] to generate structures from 5C and Hi-C data. By reducing model resolution, whole human chromosomes, or even whole nuclei can be modeled, e.g. to reconstruct configurations from single cell Hi-C data[64].	
Predict the effect of genome rearrangements	In some versions of the above models, once parameters are generated from one Hi-C data set, they can be used to make predictions about the effect of genome rearrangements[63], or cell differentiation[11].	*Hi-C of the "wild type"*

A. Molecular Dynamics

Positions of atoms tracked

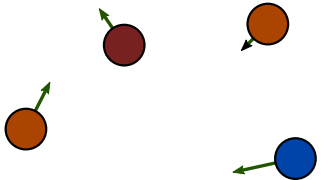


Atoms exert forces on each other



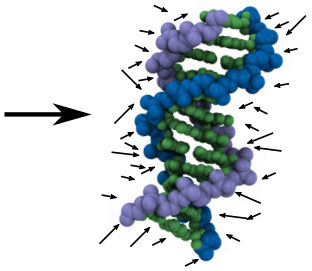
Motion simulated by solving Newton's Law

$$m\frac{d^2\mathbf{x}}{dt^2} = \mathbf{F}$$

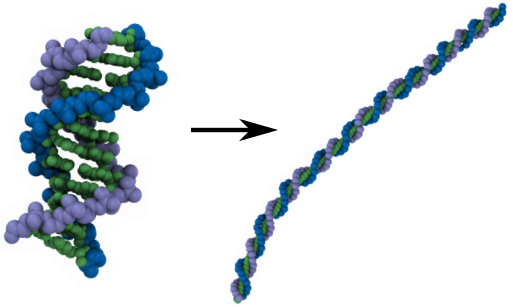


B. Coarse Graining

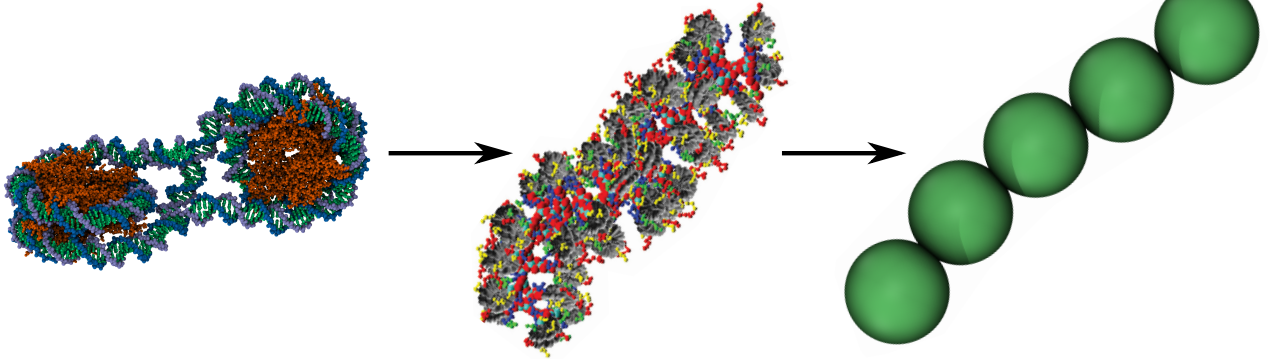
(i) implicit solvent

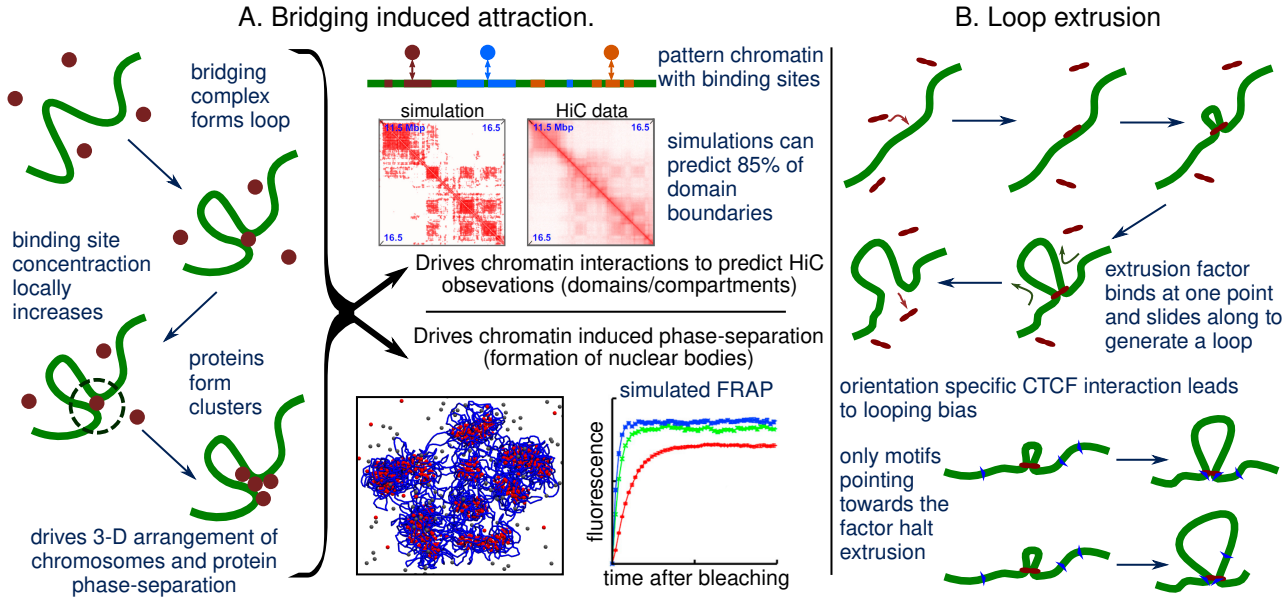


(ii) coarse-grained DNA



(iii) coarse-grained chromatin

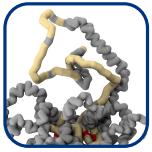


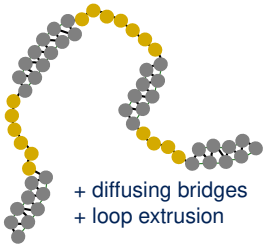

A. Bridging induced attraction.
B. Loop extrusion
bridging complex forms loop
binding site concentration locally increases
proteins form clusters
drives 3-D arrangement of chromosomes and protein phase-separation
pattern chromatin with binding sites
simulation
HiC data
11.5 Mbp
16.5
simulations can predict 85% of domain boundaries
Drives chromatin interactions to predict HiC obsevations (domains/compartments)
Drives chromatin induced phase-separation (formation of nuclear bodies)
simulated FRAP
fluorescence
time after bleaching
extrusion factor binds at one point and slides along to generate a loop
orientation specific CTCF interaction leads to looping bias
only motifs pointing towards the factor halt extrusion

A. HiP-HoP Model

HIghly **P**redictive **H**eter**O**morphic **P**olymer

variable fibre thickness

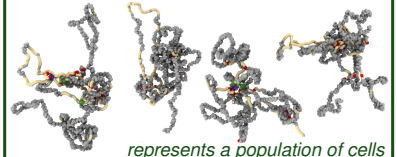




+ diffusing bridges
+ loop extrusion

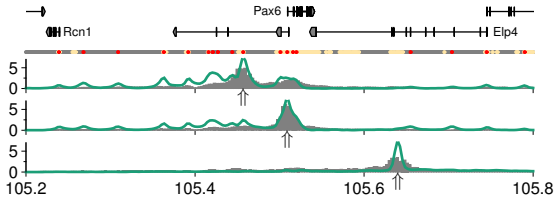
Input data: *DNA accessibility (e.g. ATAC-seq)*
CTCF / Cohesin ChIP-seq
H3K27ac ChIP-seq

Output: *a population of locus conformations*



represents a population of cells

B. 3C-style measurements at the Pax6 Locus

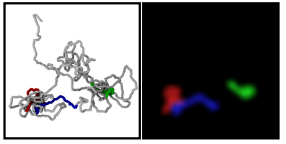


population level information can be compared to HiC / 4C / CaptureC etc.

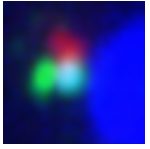
the full set of conformations can be studied, e.g. to look for variation within a population

C. FISH at the Pax6 Locus

simulation



compare with experiment



single cell-like information can be compared to microscopy