A Field-theoretical Approach to Chromosome Compaction

Benjamin Ruben*
Harvard Biophysics PhD Program
(Dated: May 21, 2021)

I. INTRODUCTION

The problem of understanding higher-order chromosome structure sits on the boundary between physics and biology [1]. Chromosomes are complicated polymers whose non-trivial mechanical and organizational features have implications for biological processes such as gene expression and cell division. Recent research combines high-throughput experimental techniques with computational modeling to decipher the mechanisms and consequences of chromosome folding. Here, we review the Hi-C conformation capture experiment, summarize previous approaches to modeling a chromosome's three-dimensional organization, and develop a new approach that employs the tools of statistical field theory.

I.1. The Hi-C Experiment

Hi-C conformation capture provides a wealth of information about chromosomestructure [2]. These methods work by treating a cell culture with a chemical cross-linking agent such as formaldehyde, creating covalent bonds between DNA segments in close spatial proximity. Then, the cells are treated with a restriction enzyme that cuts DNA at a specific, commonly occurring short sequence, leaving paired segments floating in solution. These bonded pairs are then ligated into loops and sequenced (see fig. 1A). By mapping paired sequences back to their original locations along the genome, a proximity map of the whole genome is constructed. The resulting contact map for one chromosome in DT40 (chicken) cells is shown in figure 1 at 50 kilobase resolution, for a population of chromosomes in interphase (1b) and during mitosis (1c), when chromosomes are compacted into their familiar rod-like structures. The observation of large-scale features in contact maps like these have provided hints about the mechanisms of genome folding, culminating in the discovery of LEFs (Loop Extrusion Factors): proteins that compact DNA by actively extruding loops, bringing loci far away from each other along the genome's contour length into close spatial proximity [3].

Figure 1D shows the contact scaling curve, which gives the mean probability of contact between loci as a function of their separation along the genome's contour length. The appearance of a non-monotonic "bump" in the scaling mitotic chromosome's scaling curve around 4000 kb is somewhat peculiar given that Condensin, the LEF that

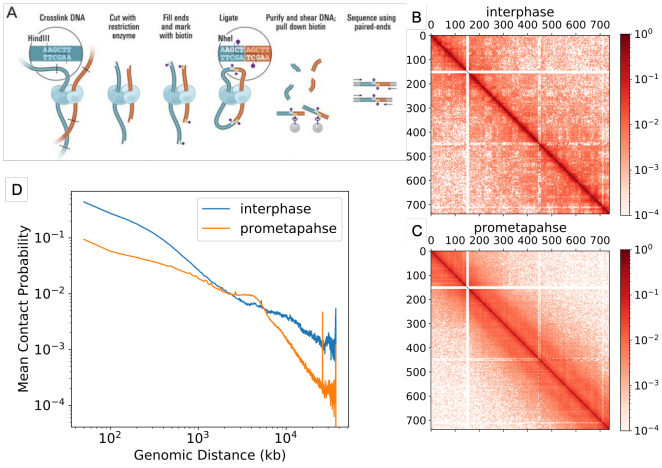


FIG. 1. The Hi-C experiment in interphase and prometaphase (mitotic) chromosomes. (A) Schematic explanation of the Hi-C experiment adapted from [2]. (B, C) Hi-C Contact Maps for DT40 (chicken) chromosome 7 in interphase and prometaphase. (D) Contact scaling curves calculated form interphase and prometaphase Hi-C Maps show contact probability as a function of genomic separation.

is thought to be responsible for compacting mitotic chromosomes, forms loops that are only 400 kb in size [4].

I.2. Existing Models for Chromosome Architecture

While Hi-C provides a mean contact map for populations of chromosomes, we would ultimately like to know the true three-dimensional structures that lead to these contact maps. Data-driven models seek to infer a three-dimensional structure, or ensemble of three-dimensional structures, directly from the Hi-C contact map. Usually this is accomplished through the use of a standard polymer model with additional interactions that can be tuned to reproduce experimental Hi-C contact maps. For example, [5] uses a square-well interaction to encode free energy losses associated with specific contacts and [[6]]) uses harmonic interactions between interacting loci. Both of these models are somewhat over-powered, in that they learn interactions directly from experimental data.

II. FIELD-THEORETICAL MODEL FORMULATION

Here, we will apply a Landau-Ginzburg approach analogous to that of the "crumpled paper" problem [[7]] to

* Ben Ruben: benruben@g.harvard.edu

model the chromosome. We begin by regarding the chromosome as an infinitely long, continuous polymer chain parameterized by a scalar variable $s \in \mathbb{R}$. The position of each point on the polymer is given by a position vector $\mathbf{r}(s) \in \mathbb{R}^3$. The chromosome's Hamiltonian must obey certain symmetries. Because there is no preferred direction or position along the polymer chain, it must be symmetric under the transformations in s: $s \mapsto -s$ and $s \mapsto s + d$. In addition, because a polymer can be translated or rotated in space without changing it's energy, the Hamiltonian must be invariant under the transformations: $\mathbf{r} \mapsto \mathbf{r} + \mathbf{d}$ and $\mathbf{r} \mapsto R\mathbf{r}$, where $\mathbf{d}$ is any constant vector and R is any unitary transformation.

These symmetries restrict the Hamiltonian to the following form:

$$\beta H_0 = \int_{-\infty}^{\infty} ds[\frac{t}{2}(\partial_s \mathbf{r})^2 + \frac{\kappa}{2}(\partial_s^2 \mathbf{r})^2 + \frac{v}{2}(\partial_s^3 \mathbf{r})^2 + ...] \quad (1)$$

where we have only included terms up to second order in $\mathbf{r}$. To more directly model the action of LEFs on the chromosome, we add a long-range interaction term of following form:

$$\begin{aligned} \beta H =& \beta H_0 + \beta H_{LR}, \text{ where} \\ \beta H_{LR} =& \int ds \int_{s'>s} ds' \frac{1}{2} W(s'-s)(\mathbf{r}(s') - \mathbf{r}(s))^2 \end{aligned} \quad (2)$$

Here, $W(s - s')$ encodes the strength of the looping interaction between loci as a function of their separation along the polymer chain. We note that this is a severe approximation: in reality, loops are known to form between specific loci, and in some cases to be actively extruded. However, in the interest of analytical tractibility, we hope to capture the effects of LEFs on the chromosome's contact scaling behavior through this potential.

In the language of Fourier modes, this Hamiltonian is represented as:

$$\beta H = \int \frac{dq}{2\pi} \frac{\Big((tq^2 + \kappa q^4 + vq^6 + ...) + [\tilde{W}(0) - \tilde{W}(q)]\Big)}{2} |\tilde{\mathbf{r}}(q)|^2 \quad (3)$$

where $\tilde{W}(q)$ is the Fourier transform of the interaction kernel, defined by:

$$\tilde{W}(q) = \int ds W(s) e^{iqs} \quad (4)$$

In order to achieve stability at this order, we insist that the coefficient of each Fourier mode is positive. From this expression, we see that the occupation of each of the Fourier modes are independent, and follow a Gaussian distribution:

$$\begin{aligned} P(\tilde{\mathbf{r}}(q)) \propto \exp\big[-\frac{1}{2}\Big((tq^2 + \kappa q^4 + vq^6 + ...) \\ + [\tilde{W}(0) - \tilde{W}(q)]\Big)|\tilde{\mathbf{r}}(q)|^2\big] \end{aligned} \quad (5)$$

$$\langle \tilde{r}_i(q)) \tilde{r}_j(q') \rangle = \frac{\delta_{ij}(2\pi)^d \delta^d(q + q')}{(tq^2 + \kappa q^4 + vq^6 + ...) + [\tilde{W}(0) - \tilde{W}(q)]} \quad (6)$$

Ultimately, we are interested in comparing this model's predictions to the contact scaling curves observed in experiments. As an intermediate step, we calculate the mean-square-distance between loci of a certain genomoic sepration:

$$\text{MSD}(d) = \left\langle \left(\mathbf{r}(s + \frac{d}{2}) - \mathbf{r}(s - \frac{d}{2}) \right)^2 \right\rangle \quad (7)$$

We make use of the following identity:

$$\mathbf{r}(s + \frac{d}{2}) - \mathbf{r}(s - \frac{d}{2}) = 2i \int \frac{dq}{2\pi} e^{iqs} \sin(\frac{qd}{2}) \tilde{\mathbf{r}}(q) \quad (8)$$

To calculate MSD(d) as the following:

$$\begin{aligned} MSD(d) = -4 \int \frac{dq dq'}{(2\pi)^2} e^{i(q+q')s} sin(\frac{qd}{2}) \\ sin(\frac{q'd}{2}) \langle \tilde{\mathbf{r}}(q) \cdot \tilde{\mathbf{r}}(q') \rangle \end{aligned} \quad (9)$$

$$= 12 \int \frac{dq}{2\pi} \frac{\sin^2(\frac{qd}{2})}{(tq^2 + \kappa q^4 + vq^6 + ...) + [\tilde{W}(0) - \tilde{W}(q)]}$$

This expression can readily evaluated numerically.

In [6], an expression which relates the mean square distance between two loci of a Gaussian chain to their contact probability is derived:

$$c_{ij} = \left(1 + \frac{< r_{ij}^2 >}{\xi^2} \right)^{-3/2} \quad (10)$$

Here, ξ is a characteristic length scale over which the probability of cross-linking decays as a function of the spatial distance between two loci.

III. RESULTS

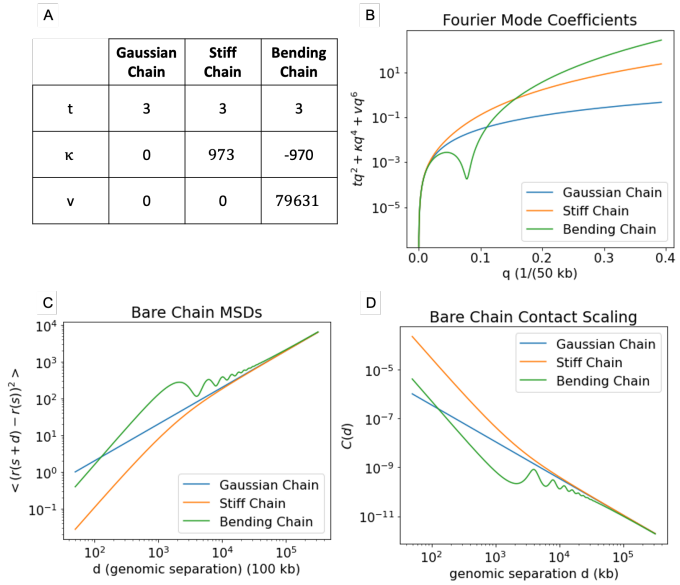
We test the contact scaling behavior of chromosomes described by many versions of the proposed Hamiltonian by varying both the Landau-Ginzburg parameters t, κ, and v, as well as the form of the interaction kernel $W(\Delta s)$.

III.1. Bare Chain Behavior

In this section, we investigate the behavior of a polymer chain with the standard quadratic Landau-Ginzburg interactions and no additional long-range interactions (i.e. $\beta H_{LR} = 0$).

In the case where $\kappa = v = 0$, we recover the standard Gaussian chain, and equation 9 can be solved analytically:

$$\text{MSD}(d) = 12 \int \frac{dq}{2\pi} \frac{\sin^2(\frac{qd}{2})}{tq^2} = \frac{3|d|}{t} \quad (11)$$



	Gaussian Chain	Stiff Chain	Bending Chain
t	3	3	3
κ	0	973	-970
v	0	0	79631

FIG. 2. Scaling Behavior of three Landau-Ginzburg Polymers: A standard Gaussian chain, a stiff chain and a chain with a propensity to bend. (A) Table of coefficients used in each polymer model. (B) Coefficients of Fourier modes in the Fourier-transformed Hamiltonian. (C) MSD (mean square distance) as a function of separation along the polymer contour. (D) Contact Probability as a function of separation along the polymer contour. Calculated using eq. 10 with $\xi = .01$

$$c(d) = (1 + \frac{3|d|}{t\xi})^{-\frac{3}{2}} \tag{12}$$

Here, $c(d)$ denotes the contact probability as a function of genomic separation d. As expected for a Gaussian polymer chain, the mean-squared distance grows linearly with genomic distance. A positive value for κ confers a stiffness to the polymer. And a negative value of κ can be used, so long as a positive value of v is included which prevents an instability in the quadratic Hamiltonian, by ensuring that $tq^2 + \kappa q^4 + vq^6 > 0$ for all q. A negative value of κ encodes a tendency for the polymer to bend. We show these three situations in fig. 2:

Intriguingly, the bending chain has a tendency to form regions of enhanced contact at long distances. This is a result of the "dip" visible for the bending chain in fig. 2B. It can be seen from equation 9 that the appearance of a dip in the coefficient of a Fourier mode $\tilde{\mathbf{r}}(q)$ leads to an enhancement in long-range contacts at the genomic separation reciprocal to q. The Landau-Ginzburg coefficients for this model were picked specifically to create an enhancement in contacts around 4000 kb.

III.2. Interaction Kernel Behavior

In this section, we investigate the behavior of a standard Gaussian chain with additional looping interactions (i.e. $\kappa = v = 0$, $\beta H_{LR} \neq 0$)

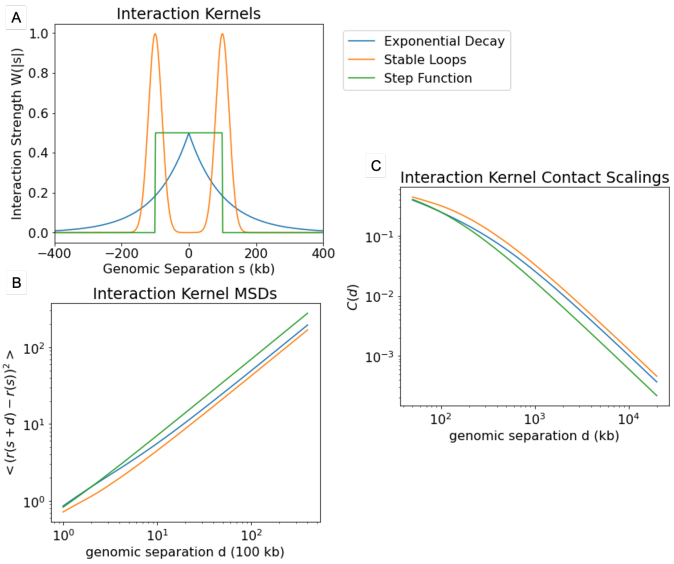


FIG. 3. Scaling Behavior of a Gaussian Polymer with Various Interaction Kernels. (A) Depiction of interaction kernels explored. $W(|s|)$ is the spring constant of a harmonic potential that all pairs of loci separated by $|s|$. (B) MSD and (C) Contact probability as a function of separation along the polymer contour.

We introduce a variety of biologically plausible forms for the interaction kernel $W(s)$. Previous work modeling mitotic chromosomes suggests that loop sizes are exponentially distributed [8].

This motivates an exponentially decaying interaction kernel:

$$W(s) = \frac{k_0}{\lambda} * e^{-\frac{|s|}{\lambda}} \tag{13}$$

Here, λ represents the average loop size. Other possibilities include looping around a specific distance:

$$W(s) = \frac{k_0}{\sqrt{2\pi\sigma^2}}(e^{\frac{(s-\lambda)^2}{2\sigma^2}} + e^{\frac{(s+\lambda)^2}{2\sigma^2}}) \tag{14}$$

and looping at all distances less than a cutoff length:

$$W(s) = \frac{k_0}{\lambda}\Theta(\lambda - |s|) \tag{15}$$

We calculate MSD and contact probability for each of these kernel forms as a function of genomic separation in fig. 3. To better understand which of these kernels offers an accurate depiction of the chromosome, we compare the calculated contact scaling curves for both the exponentially decaying kernel and the Gaussian kernel to experimental contact scaling curves from interphase and mitotic Hi-C experiments (see figure 4). Here, we sweep the parameter λ for average loop size to look for the best fit, and to extimate the mean size of loops that are formed by LEFs in interphase and prometaphase.

From these figures, it appears that an exponentially decaying interaction kernel best captures the scaling curve in both interphase and mitosis, with an average loop size of 50 kb and 275 kb respectively.

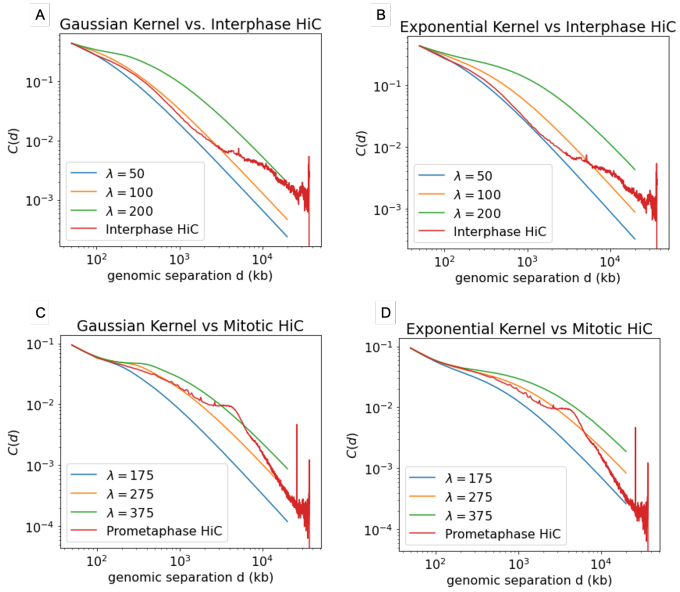


FIG. 4. Fitting λ to experimental data. Contact scaling behavior of standard Gaussian polymer with additional Gaussian (A,C) or Exponentially decaying (B, D) long-range interaction term. Calculated scaling curves are compared to experimental data from interphase (A,B) and mitotic (C,D) chromosomes. For each curve, the contact radius ξ was calculated so that the curve agreed with experiment at the initial contour length of 50 kb. In addition, for all Gaussian kernels, we use $\sigma = \frac{\lambda}{3}$

III.3. Competition Between Bending and Looping

When cells transition from interphase to mitosis, the average size of loops extruded by LEFs increases, and a peculiar long-range enhancement in contact probability (arounf 4000 kb) is observed. We have seen that a polymer with a propensity to bend can reproduce the peculiar long-range enhancement in contact frequency. We have also seen that a polymer with a Gaussian or exponentially decaying long-range compaction interaction can recreate the mitotic chromosome's short-range scaling behavior. In this section, we examine how these behaviors interact and propose a mechanism by which an enhancement of long-range interactions may occur when cells transition from interphase to prometaphase.

Fig. 5 shows the behavior of a polymer with a bending propensity as well as a looping interaction kernel. For the Gaussian kernel (A1-3), we see that long-range interactions appear when loop size decreases, opposite to what is observed in the interphase to mitosis transition. However, for an exponentially decreasing interaction kernel (B1-3), the trend is opposite: as loop size increases, the dip in the curve of figure B2 becomes more pronounced, and consequently so does the enhancement in the chromosome's contact frequency around 4000 kb! However, as is easily visible in the third row of fig. 5, the looping interaction fails to recreate the chromosome's short-range scaling behavior in this case.

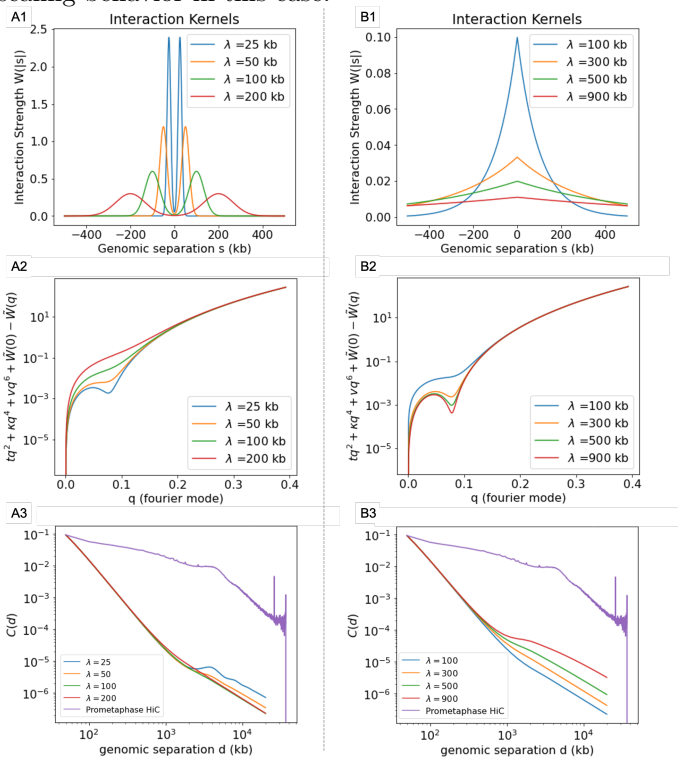


FIG. 5. Interplay between bending and looping. Contact Scaling Behavior for bending polymers with additional long-range interactions. (A1-3) Gaussian kernel. (B1-3) Exponentially decaying kernel. The top row shows the kernels used. As λ increases, normalization is held constant so that the area under all curves is the same. The second row shows the coefficients of the fourier modes in the fourier-transformed hamiltonian. The bottom row shows the resulting contact scaling behavior.

IV. CONCLUSION

In this paper, we put forth a dramatically simplified model for chromosome compaction by LEFs. Despte its many assumptions, this model is able to re-produce the short-range scaling behavior of both interphase and prometaphase chromosomes to a reasonable degree of accuracy. By studying parameter regimes where the polymer chain has a propensity to bend, a possible mechanism for the enhancement of long-range contacts upon entry to mitosis reveals itself: As loops get larger, the effect of the bending behavior becomes more pronouncd, leading to an enhancement of long-range contacts.

[1] G. Fudenberg and L. A. Mirny, Higher-order chromatin structure: bridging physics and biology, Current Opinion in Genetics & Development **22**, 115 (2012).

[2] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. 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, and J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, Science **326**, 289 (2009).

[3] M. Ganji, I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, and C. Dekker, Real-time imaging of dna loop extrusion by condensin, Science **360**, 102 (2018), https://science.sciencemag.org/content/360/6384/102.full.pdf.

[4] J. H. Gibcus, K. Samejima, A. Goloborodko, I. Samejima, N. Naumova, J. Nuebler, M. T. Kanemaki, L. Xie, J. R. Paulson, W. C. Earnshaw, L. A. Mirny, and J. Dekker, A pathway for mitotic chromosome formation, Science **359**, eaao6135 (2018).

[5] M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, and J. N. Onuchic, Transferable model for chromosome architecture, Proceedings of the National Academy of Sciences **113**, 12168 (2016), https://www.pnas.org/content/113/43/12168.full.pdf.

[6] G. Le Treut, F. KÃ©pÃšs, and H. Orland, A polymer model for the quantitative reconstruction of chromosome architecture from hic and gam data, Biophysical Journal **115**, 2286 (2018).

[7] M. Paczuski, M. Kardar, and D. R. Nelson, Landau theory of the crumpling transition, Phys. Rev. Lett. **60**, 2638 (1988).

[8] N. Naumova, M. Imakaev, G. Fudenberg, Y. Zhan, B. R. Lajoie, L. A. Mirny, and J. Dekker, Organization of the mitotic chromosome, Science **342**, 948 (2013).