Predicting scale-dependent chromatin polymer properties from systematic coarse-graining

Sangram Kadam,[1,*] Kiran Kumari,[1] Vinoth M,[1] Shuvadip Dutta,[2] Mithun K. Mitra,[2] and Ranjith Padinhateeri[1,3,†]

[1] *Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai 400076, India*
[2] *Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India*
[3] *Sunita Sanghi Centre for Aging and Neurodegenerative Diseases, Indian Institute of Technology Bombay, Mumbai 400076, India*

Simulating chromatin is crucial for predicting genome organization and dynamics. Even though coarse-grained bead-spring polymer models are commonly used to describe chromatin, the relevant bead dimensions, elastic properties, and the nature of inter-bead potentials are unknown. Using publicly available nucleosome-resolution contact probability (Micro-C) data, we systematically coarse-grain chromatin and predict quantities essential for polymer representation of chromatin. We compute size distributions of chromatin beads for different coarse-graining scales, quantify fluctuations and distributions of bond lengths between neighboring regions, and derive effective spring constant values. Unlike the prevalent notion, our findings argue that coarse-grained chromatin beads must be considered as soft particles that can overlap, and we derive an effective inter-bead soft potential and quantify an overlap parameter. We also compute angle distributions between neighboring bonds giving insights into intrinsic folding and local bendability of chromatin. While the nucleosome-linker DNA bond angle naturally emerges from our work, we show that there are two populations of local structural states. The bead sizes, bond lengths, and bond angles show different mean behavior at Topologically Associating Domain (TAD) boundaries and TAD interiors. We integrate our findings into a coarse-grained polymer model and provide quantitative estimates of all model parameters, which can serve as a foundational basis for all future coarse-grained chromatin simulations.

PACS numbers:

Introduction

The eukaryotic genome is organized in the form of many long chromatin polymer chains, each essentially a string of nucleosomes—DNA wrapped around histone proteins—folded, looped, and condensed into domains of different compaction [1–4]. The spatial and temporal organization of these chains and their internal epigenetic states are crucial in deciding aspects ranging from cellular function to differentiation and development [5–8].

Chromosomes are typically simulated and studied as coarse-grained(CG) bead-spring polymer chains [9, 10]. A coarse-grained polymer picture is useful for many reasons: it is nearly impossible to simulate the huge polymer set (∼millions of nucleosomes) in its entirety. More importantly, the coarse-grained representation, with effective parameters, can be a powerful tool to understand chromatin organization and dynamics and make useful predictions [11–24]. However, since we still do not understand the chromatin structure and properties in detail, systematic coarse-graining has been a difficult task. We do not fully know the polymer properties/parameters relevant for simulating coarse-grained chromatin.

Owing to a large body of work over the past few decades, double-stranded DNA has a good coarse-grained description as a semi-flexible polymer [25–27]. We understand its coarse-graining size, bending stiffness, stretching elasticity, and other relevant parameters [26, 27]. However, chromatin is a more complex polymer having heterogeneous properties arising from different epigenetic states, amount of different proteins bound, and local folding [28, 29]. This complexity makes it difficult to accurately compute coarse-grained bead diameter, elastic constants, and other physical properties for a chromatin polymer.

Recent experimental advances have made it possible to understand chromatin structure using biochemical methods like Hi-C, Micro-C [30–41] and imaging methods like SAX, cryo-EM, and super-resolution imaging [42–49]. The studies so far show that chromatin is organized into different compartments and topologically associated domains (TADs) [2, 7, 31–33]. While histone modifications, transcription factors, and chromatin binding proteins greatly

*Electronic address: sangramkadam@iitb.ac.in
†Electronic address: ranjithp@iitb.ac.in

affect chromatin folding and make it a highly heterogeneous polymer, how the interplay between these factors decides the compaction and dynamics of chromatin is currently being investigated.

While different experimental methods provided us data to understand chromatin organization [30–33, 35–37, 42–44, 46–48, 50, 51], theoretical/computational studies have been pivotal in understanding and explaining chromatin characteristics [11–15, 18–22, 24, 52–67]. Models that simulated chromatin at nucleosome resolution primarily investigated how different molecular interactions influenced higher-order organization beyond the 10nm chromatin [53, 55, 56, 61, 66, 68–70]. Nearly all models that are employed to understand Hi-C/microscopy data represented chromatin as a bead-spring polymer chain with each bead representing chromatin of length in the range $\sim$ 1kb to 1Mb [12–17, 20–24]. However, physical dimensions and elastic properties of chromatin are not well understood. What is the diameter of a 1kb, 10kb, or 100kb chromatin bead? What is the magnitude of the spring constant that would represent the thermal fluctuation of chromatin at different length scales? Should chromatin be considered a flexible polymer or a semi-flexible polymer? Does it have an intrinsic curvature/bending stiffness? None of these questions have clear, definitive answers in the literature, currently. While it is well known that chromatin behavior is heterogeneous, depending on the epigenetic state, existing coarse-grained models of chromatin typically assume that the physical dimensions of the beads and elastic properties of the filament are uniform along the polymer, independent of the epigenetic state. How these properties—the size of the beads, stretching elasticity, bending elasticity, etc.—vary along the contour is also unknown. In the current models, the heterogeneity of chromatin is often incorporated into additional intra-chromatin interactions—interaction between two far-away beads [12, 24, 71]. Since one does not know the size of a coarse-grained chromatin bead, it is taken as a fitting parameter—a constant number across the filament—to achieve experimentally measured 3D distance values [12, 15, 22, 24, 72, 73]. Fig. S1 shows some of the reported values and how scattered they are. We do not understand this variability and what each number means.

One could not do systematic coarse-graining so far because the chromatin conformation capture data was available only with lower resolutions like 100kb, 10kb, and up to 1kb [30–33, 50]. Obtaining information smaller than the HiC resolution was not possible. Moreover, even at the smallest size scale ($\sim$ kb), the physical dimension of chromatin—chromatin bead diameter—was an unknown parameter. However, recent experiments have provided us chromatin conformation capture data at near-nucleosome resolution—200 bp resolution, which is essentially a nucleosome plus the linker DNA [34, 35, 39]. This data enables us to make a fine-grained chromatin model and systematically probe the properties of the coarse-grained chromatin. The advantage here is that the physical size of a 200bp chromatin is not a completely unknown free parameter; we do have a fair idea about the size of a 200bp chromatin bead. In this work, we take advantage of this recent fine-grained data, start from the 200bp Micro-C contact map, and construct chromatin polymer configurations that satisfy the map. From our work, without any arbitrary fitting parameter, the 3D distances and radius of gyration values emerge in a reasonable range comparable to known experiments. Using the 200bp-chromatin as a fine-grained polymer, we coarse-grain the chromatin systematically. This enables us to predict several quantities essential for anyone simulating a coarse-grained chromatin polymer. We predict (i) the physical sizes of coarse-grained beads of various chromatin length scales, (ii) the overlap between coarse-grained beads and an effective inter-bead soft potential energy, (iii) the value of the spring constant between neighboring beads dictating the fluctuation, and (iv) the distributions of bond angles and dihedral angles giving insights into the stiffness of chromatin. We show that some of the ideas we learned—e.g., soft inter-bead interactions that allow overlap—are crucial for obtaining sensible 3D distances/R_g when coarse-grained models are employed.

Model and Methods

Model-I. Fine-grained chromatin model with 200bp resolution chromatin: Our basic model is the fine-grained chromatin polymer model with 200bp resolution, constructed based on the publicly available Micro-C data for mouse embryonic stem cells (mESCs) [35, 36]. The polymer is made of N spherical beads, having the size of 200bp chromatin (diameter σ), with nearest-neighbor connectivity via harmonic springs and self-avoiding interaction via the repulsive part of the Lennard-Jones potential (see Fig. 1(a), Supplementary Information (SI) Sec. S1.A). Since each bead consists of a nucleosome and 50bp linker DNA, we call the bead a "nucleosome-linker" (NL) bead. To generate an ensemble of configurations consistent with Micro-C data, we connected (brought into proximity) bead pairs i and j with the experimentally observed contact probability P_{ij} in a two-step process. First, we defined a set of prominent (strong) contacts of the Micro-C contact map (see Sec. S1.A) [71]. Taking only the prominent contact probability values, we inserted harmonic springs between bead pairs i and j if $r_n < P_{ij}$, where r_n is a uniformly distributed random number between 0 and 1. Using this procedure, we generated 1000 independent polymer configurations and equilibrated them using Langevin simulations with LAMMPS [74]. We defined "prominent contacts" as follows [71]: Since the contact map depends only on $|i-j|$ for homogeneous polymers, we took the set of all P_{ij} values for a given $|i-j|$ and computed their mean and standard deviation. If P_{ij} was at least one standard deviation larger than the mean, we considered it as a prominent contact (see SI Sec. S1.A). Prominent contacts are defined for each $|i-j|$

line in the matrix (line parallel to the diagonal representing all equidistant bead pairs). Bonding prominent contacts ensured that all actively acquired far-away contacts (e.g., contacts via loop extrusion) were present.

In the second step, going beyond the prominent contacts, our aim is to insert contacts in the P_{ij} fraction of the configurations (out of the 1000 configurations) for each bead pair (i, j). To achieve this, we started with the ensemble of equilibrated configurations from step-1 and inserted harmonic springs between beads i and j in the P_{ij} fraction of configurations whose 3D distances (r_{ij}) are the smallest (see SI). This system was then equilibrated using Langevin simulations with LAMMPS. While the first step ensured that the strong contacts formed via events like loop extrusion were established, the second step ensured that all bonds closer in space would have priority in forming protein-mediated contacts.

We have used the minimal fine-grained model that accounts for polymer connectivity, self-avoidance, and contact probability. The assumption here is that all other properties of the fine-grained polymer (like inter-nucleosome interactions and stiffness) lead to the experimentally observed contact probability, which we have ensured. Our model generates all possible polymer configurations such that the experimentally known constraint of the contact map is satisfied.

Size of a 200bp chromatin bead (σ): Since a 200bp chromatin bead is bigger than a nucleosome, its size has to be greater than the size of the nucleosome (11nm) [3]. As geometrically shown in Fig. S2(a), since two neighboring nucleosomes are connected via a rigid 50bp linker DNA, the distance between them can be $\approx$ 28nm. However, two far-away nucleosomes can come as close as 11–12nm (with histone tails and other bound proteins). Hence, on average, one expects an effective size $\approx$ 20nm. In the Results section, we have shown that when $\sigma = 21$nm, the 3D distances and R_g values match well with experimental data. This is sensible considering the linker length and that a typical nucleosome in vivo will likely be covered by several enzymes/proteins like acetyl/methyl transferases, HMG, HP1, remodelers, etc. This is also consistent with the earlier observation that $\sigma = 25$nm for 250bp beads [72].

Since Model-I did not have explicit linker DNA, we also simulated short chromatin with nucleosomes, explicit linker DNA, and entry-exit angles between nucleosomal DNA. In this detailed Model-II, the chromatin polymer has two types of beads—linker DNA bead and nucleosome bead (see Sec. S1.B)[75]. In the Results section, we have compared the radius of gyration of chromatin segments from Model-II and the first fine-grained model. This also suggests that our $\sigma = 21$nm value is indeed reasonable.

Results

Constructing 3D chromatin configurations at near-nucleosome resolution consistent with Micro-C contact map and measured 3D distances

We simulated a fine-grained chromatin polymer made of "nucleosome-linker" (NL) beads, with each bead representing 200bp of chromatin (Fig. 1(a), Methods and SI), and generated an ensemble of steady-state chromatin configurations, taking the Micro-C contact probability data (P_{ij}) of mouse embryonic stem cells (mESCs) as input [35]. We simulated ten different genomic loci (see SI Table S1) having broad euchromatic or heterochromatic chromatin state characteristics. We computed the ensemble-averaged contact probability for each locus and compared them with the input contact map. The contact maps from the simulations appear visually similar to the Micro-C data (Fig. 1(b–d) and Fig. S3). The bottom panel shows representative snapshots from the simulations. For the Ppm1g locus, the beads are colored based on the domains in the contact map (color-strip at the top of Fig. 1(b-d)), while for the Gm29683 and Cbx8 loci, the far-away heterochromatic regions interacting with each other are shown in red and blue color (Fig. 1(c,d)). Even though these are representative snapshots, one can see signatures of domain separation (configurations below Fig. 1(b)) and interaction among far-away regions(configurations below Fig. 1(c-d)). The contact probability versus genomic distance plots from simulations and experiments are comparable (Fig. S4). We quantified the similarity between the experimental and simulation contact matrix by computing the stratum-adjusted correlation coefficient (SCC)[76] (Sec S1.D). The SCC values for most regions are above 0.9, suggesting that the simulations reproduce contact maps well (Table S1). Beyond the contact map, we also compared the mean 3D distance from our simulations for the alpha globin region with the available experimental data [12, 77] (see Fig. 1(e)). These results suggest that our simulations have generated an ensemble of configurations with the contact map and 3D distances comparable with experiments.

We now systematically coarse-grain all the above chromatin polymers. We chose n_b consecutive NL beads to form a coarse-grained CG bead (colored big bead in Fig. 2(a) top right). The coarse-grained polymer consists of N/n_b number of CG beads. We then measured various properties of the coarse-grained polymer such as the size of a CG bead R_g, bond length $l_{\rm cg}$, bond angle $\theta_{\rm cg}$ (Fig. 2(a) bottom), and dihedral angle $\phi_{\rm cg}$ (Fig. 4(a) bottom). We study how these properties depend on the coarse-graining size n_b and the genomic location. As a control, we have compared our chromatin results with the ideal chain (bead spring polymer with no self-avoiding interaction), the SAW (Self

Avoiding Walk; bead spring polymer with Weeks–Chandler–Anderson potential), and a highly packed globule (bead spring polymer with attractive Lennard-Jones potential with $\epsilon = 1k_B$T).

Predicting the size of coarse-grained chromatin beads and its variability along the genome

Since nearly all polymer simulations use coarse-grained beads of various genomic lengths like 1kb, 10kb, 100 kb, etc., it is important to understand the physical size (radius) of such a coarse-grained bead. Does bead size depend on the state of the chromatin (heterochromatin or euchromatin) and/or the genomic location of the bead (TAD interior, TAD boundaries)? To answer this, we first computed the radius of gyration (R_g) of the n_b consecutive beads that form a CG bead. Taking a sliding window of n_b beads, we plotted the average radius of gyration (R_g^i) as a function of genomic location (Fig. 2(b) and Fig. S5(a); real units in X_2 and Y_2 axes). For both euchromatic and heterochromatic segments, R_g values vary along the genomic length for the same coarse-graining, representing heterogeneities along the chromatin. In contrast, R_g curves for SAW and packed globule do not vary along the polymer contour (Fig. S5(b) and (c)). To understand how the locations of the peaks and troughs in the R_g correlate with the contact map, we compared them for a 1kb coarse-graining scale (Fig. 2(c, d) and Fig. S6). R_g values peak at the boundaries of TAD-like domains and are relatively small within the interior of the domain. That is, coarse-grained beads representing inter-TAD regions will have a larger physical dimension. This is consistent with what is observed in recent experiments [42].

How big is a typical 1kb, 5kb, or 10kb chromatin? We compared the mean R_g values (averaged over the entire region we considered) for several genomic segments – some of them are euchromatic, and others are heterochromatic regions in terms of dominating histone modification marks (Fig. 2(e)). Throughout this paper, the chromatin loci with broad euchromatic characteristics are plotted in shades of red, while the loci with broad heterochromatic marks are plotted using shades of blue. SAW polymer data is presented as a "control" having an expected $R_g \sim n_b^{0.6}$. Interestingly, even though there is variability among the gene region, they more or less fall in a narrow range. The range of average R_g values for 1 kb, 10 kb, and 80 kb chromatin regions are $27 - 28$nm, $85 - 97$nm, and $135 - 155$nm, respectively (see Fig. 2(e) inset). The curves for SAW and globule mark the extreme values possible. The R_g predictions for various regions from our simulations are of the same order as what is seen in experiments for the repressed chromatin domains in *Drosophila* cells [47] (orange data points Fig. 2(e)). Although the experimentally known R_g values are for a different cell type, they do indeed fall in the range that we are predicting, suggesting that R_g values from our simulation are reasonable. In Fig. S5(d–e), we also show the distribution of R_g values.

To independently check the order of magnitude of R_g values, we employed another realistic, detailed model of short chromatin with explicit nucleosomes having entry/exit angles and linker DNA explicitly (Fig. S2(b) and Sec. S1.B). We find that the size of a 1kb chromatin made of 5 nucleosomes in this model is comparable to what we predict using our basic fine-grained model (Fig. S2(c)), suggesting that the fine-grained model we use is reasonable.

Coarse-grained chromatin beads are not hard spheres; they overlap impacting bond length and stiffness

Another important quantity is the distance between the centers of two neighboring coarse-grained beads, defined as the bond length $l_{\rm cg}$, depicted in Fig. 2(a).

In Fig. 3, we present the bond length, its statistics, and the stretching elastic constants derived from it. First, mean $l_{\rm cg}$ values vary depending on the coarse-graining size and genomic location for all the genomic regions we studied. (Fig. 3(a, b) and Fig. S7 (a, b)). Similar to R_g, the bond length is also high at certain genomic locations like boundaries of TAD-like domains and low in the domain interiors (Fig. 2(c, d) and Fig. S6). This is essentially the same behavior found in recent experiments [42], which showed that the inter-TAD distances are larger than the intra-TAD distances, consistent with the spatial variation in coarse-grained bead sizes observed in our simulations. Given that a typical chromatin polymer will contain both euchromatic and heterochromatic regions of different compaction, it is instructive to compare extreme sizes of coarse-grained beads to understand the variability one can expect along the genome. For a 5kb segment ($n_{\rm b}$= 25), the mean $l_{\rm cg}$ values at different genomic locations vary in the range $\approx 95-120$nm (heterochromatin) to $\approx 105-150$nm (euchromatin)—compare $n_b = 25$ in Fig. 3(a) and (b). Interestingly, these values are comparable to the $l_{\rm cg}$measurements from microscopy experiments that "paint" 5kb chromatin segments [47]. This also implies that real chromatin will have highly heterogeneous bead dimensions, unlike the prevalent uniform bead size picture. The mean values of bond length, averaged along the polymer contour, are shown as a function of n_b (coarse-graining scale, genome size) for ten different chromatin loci (Fig. 3(c)). Equivalent coarse-grained bond lengths for SAW and globule are shown as control. Similar to R_g, there is some amount of variability in mean $l_{\rm cg}$ across different gene regions; however, they fall within the range of $58 - 67$nm for 1kb and $114 - 132$nm for 10kb.

How is $l_{\rm cg}$ related to R_g? Naively one would expect that $l_{cg} \approx 2R_g$. However, this is not the case; chromatin has $l_{cg} < 2R_g$ because the two nearby polymer segments can "mix"—CG beads can overlap—and have their center of

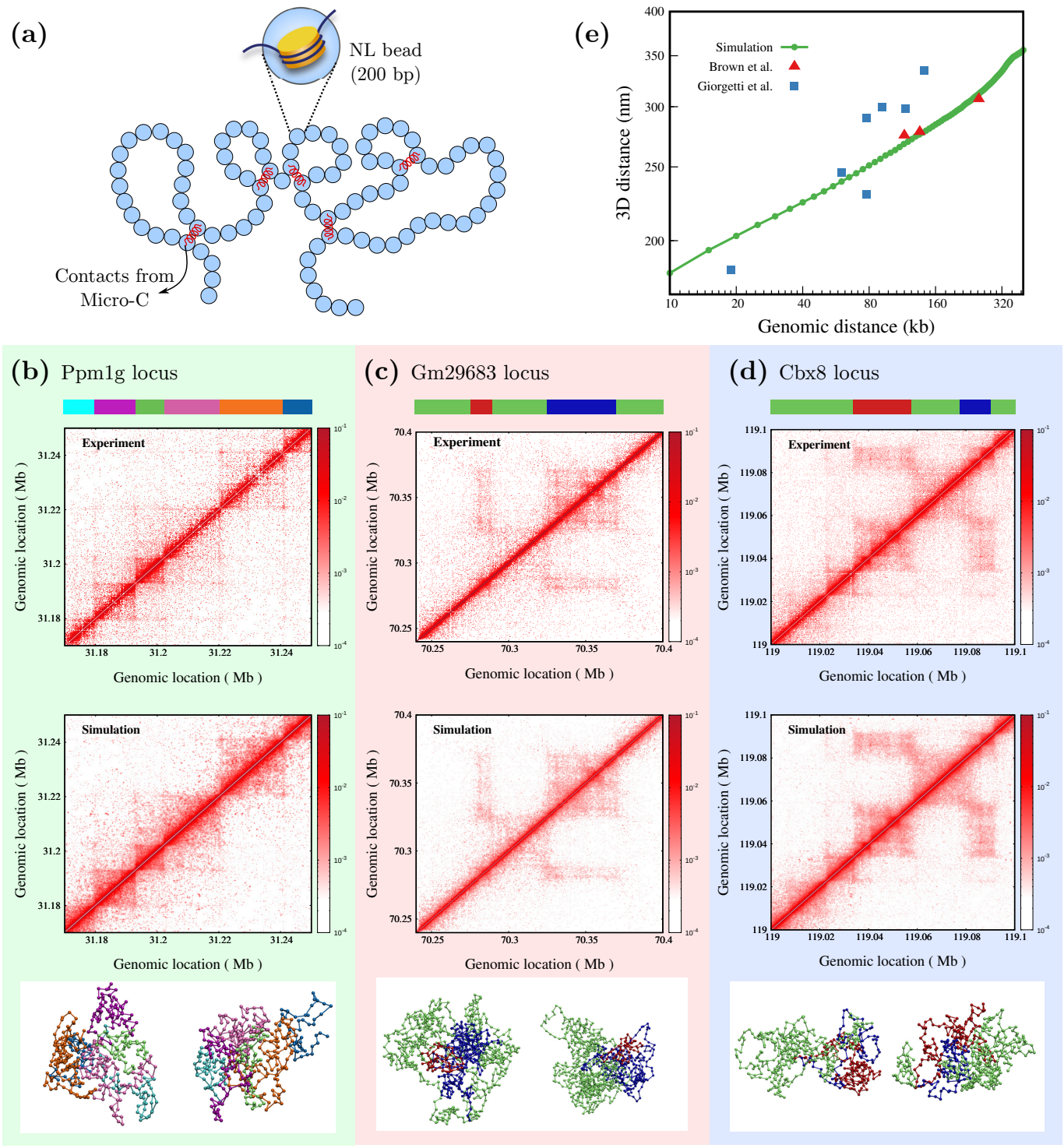

FIG. 1: Chromatin configurations are consistent with experiments: (a) Schematic of the fine-grained chromatin polymer with one bead representing 200bp nucleosome+linker (NL) chromatin. An ensemble of configurations is generated such that any pair of beads (i, j) is connected (red springs) with a probability p_{ij}, based on observed contacts in Micro-C experiments by Hsieh et al. [35] (see text). (b–d): Comparison of contact maps obtained from our simulation to experiments for a euchromatic region (b: Ppm1g locus) and two heterochromatic regions (c: Gm29683 locus, and d: Cbx8 locus). The bottom panel shows representative snapshots from the simulations, where the bead colors represent different domains, as shown in the color strip at the top. (e) 3D distance from our alpha globin simulation (green filled circles) compared with available experimental data for the same region (red filled triangles) taken from Brown et al. [77], and other regions (blue squares) of similar genomic length range taken from Giorgetti et al. [12].

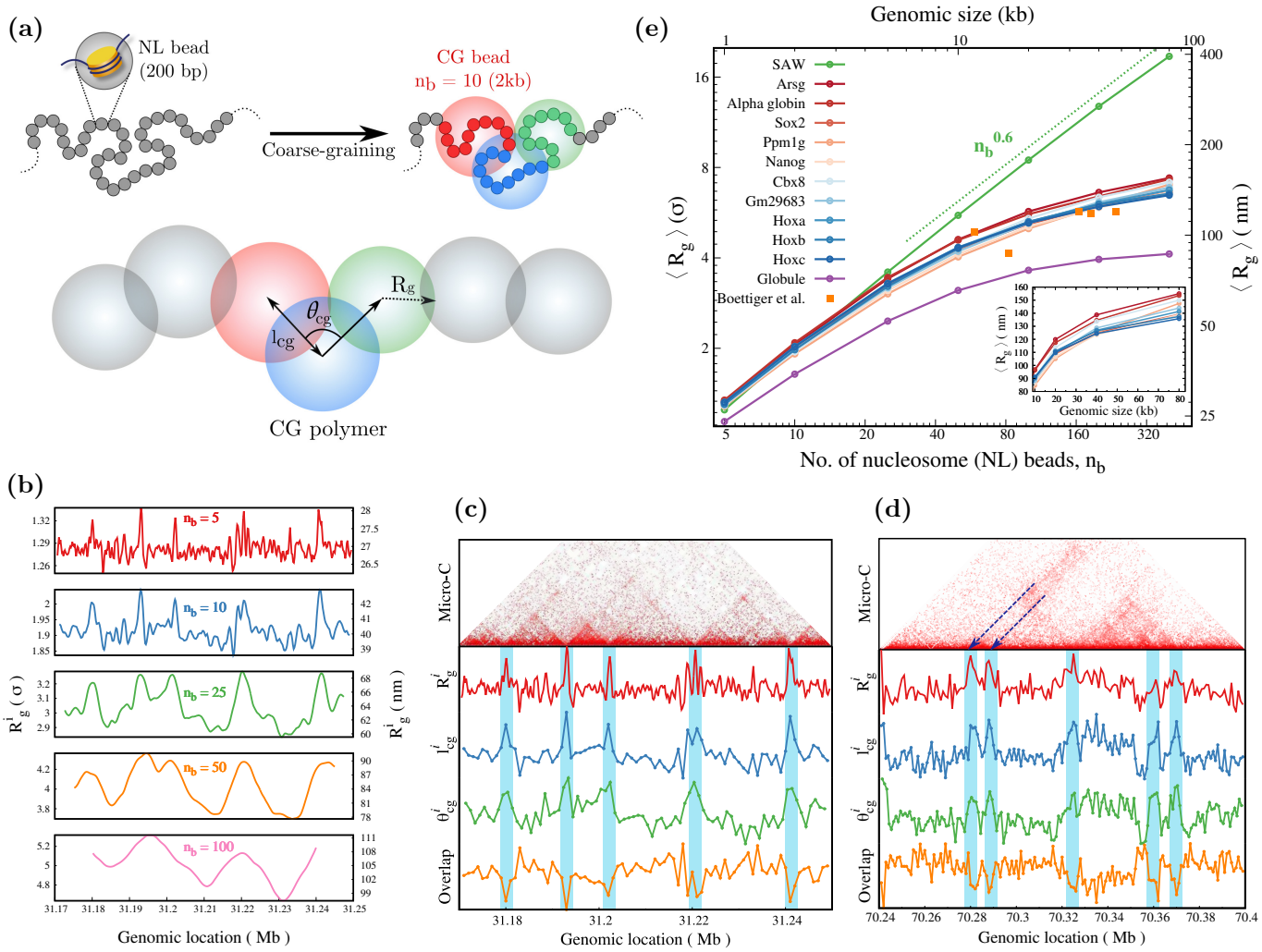

FIG. 2: (a) Schematic showing the coarse-graining procedure and quantities of interest. The fine-grained polymer of N beads (grey, small beads, representing 200bp of nucleosome+linker (NL)) is coarse-grained into N_{cg} big beads, with each big coarse-grained bead containing $n_b = N/N_{cg}$ beads. For illustration purposes, ten small colored beads ($n_b = 10$) are coarse-grained into one big CG bead. l_{cg} is the length of the bond connecting two successive coarse-grained beads, θ_{cg} is the angle between two successive bonds, and R_g is the radius of gyration of the coarse-grained polymer segment of size n_b. (b) Average radius of gyration at different genomic locations (R^i_g) for different n_b values for the Ppm1g locus. Variation of R^i_g and other quantities (l^i_{cg}, θ^i_{cg}, and overlap, see text) along the chromatin contour (n_b=5) for (c) a euchromatic Ppm1g locus and (d) a heterochromatic Gm29683 locus. Note different behavior at the domain boundaries. (e) Mean R_g for different coarse-graining sizes (n_b) showing the predicted range of values for different chromatin states. The chromatin loci with broad euchromatic characteristics are plotted in shades of red, while the loci with broad heterochromatic marks are plotted using shades of blue. Experimental data from Boettiger et al. [47] is added (orange squares; repressed chromatin domains in *Drosophila* cells) to demonstrate that the predicted R_g values are reasonable. The dotted line is shown as a guide to the eye. R_g values are presented in the units of NL bead size σ (Y_1 axis) and also in nm (Y_2 axis).

mass locations nearby (Fig. 3(d)). To quantify the overlap or mixing between the two adjacent coarse-grained beads, we define an overlap parameter $\mathcal{O} = (2R_g/l_{cg})$ (Fig. 3(e)). If the coarse-grained regions are perfectly spherical and non-overlapping(no mixing) $\mathcal{O} \leq 1$. Imagine two hard spheres of radius R_g connected by a spring. The thermal fluctuations will result in the average inter-bead distance (the equivalent of l_{cg} here) being slightly larger than $2R_g$ and $\mathcal{O} < 1$. As a control, one can see that for SAW, $\mathcal{O} < 1$ and is nearly independent of coarse-graining scale (n_b). For euchromatin and heterochromatin $\mathcal{O} > 1$ (i.e., $l_{cg} < 2R_g$), implying the mixing of adjacent polymer segments. We also find that $\mathcal{O}$ depends on the coarse-graining scale — overlap is high at larger n_b values. Since $l_{cg} \neq 2R_g$, should one consider l_{cg} as the size (diameter) of an effective coarse-grained bead, or $2R_g$? This is a relevant question

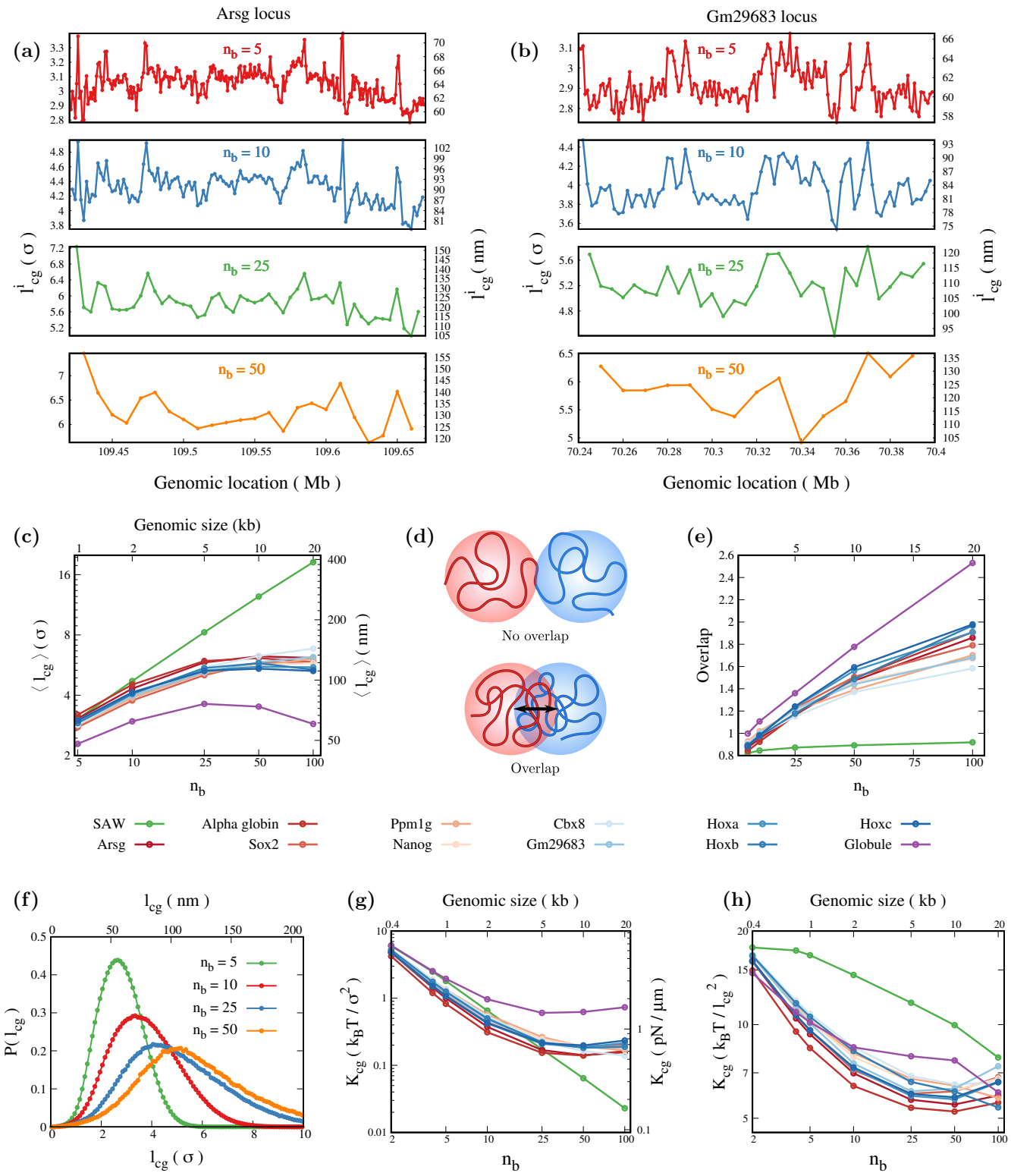

FIG. 3: Spatial variation of average bond length (l^i_{cg}) for (a) euchromatic and (b) heterochromatic regions is plotted for different coarse-graining sizes (n_b). (c) Average l_{cg} for different coarse-graining sizes showing the range of values possible from SAW to globule; presented in two different units in Y_1 and Y_2 axes. (d) Schematic showing two nearby long polymer segments "mixing" (or not mixing) in 3D space resulting in overlap (or no overlap) of coarse-grained (colored) beads. (e) A parameter that quantifies the extent of overlap is plotted for different n_b values. (f) Distribution of l_{cg} for euchromatin region. (g–h) The effective spring constant (K_{cg}) quantifies the fluctuation between neighboring coarse-grained beads for chromatin segments in different epigenetic states, presented in two different units. CG simulations would need units in (h).

for coarse-graining; given the fact that the beads can overlap, we propose that $l_{\rm cg}$ may be considered as the effective diameter of a coarse-grained bead since it is the effective bond length. We also computed $\mathcal{O}$ as a function of genomic location (Fig. S8(a, b). Comparing the overlap with the contact map shows that the overlap at the boundary of TAD-like domains is smaller than the domain interior (Fig. 2(c, d) and Fig. S6). While the above quantity measures the overlap between neighboring regions ("bonded" CG beads), any two chromatin regions residing far away along the polymer contour ("nonbonded" CG beads) can also overlap. To quantify this overlap, we first computed the probability distribution of 3D distance r_{ij} between any two CG beads, $P(r_{ij})$(Fig. S8(c)). The probability that $r_{ij}<$ $l_{\rm cg}$ is a measure of overlap among far away beads (see Fig. S8(d)).

Going beyond the average size, we computed the bond length distribution $P(l_{cg})$ that has all the information about the fluctuation and higher moments. As a control, for the ideal polymer chain, $P(l_{cg})$ from the simulation matches with the analytical relation proposed by Laso et al. [78](Fig. S7(c)). We then plot $P(l_{cg})$ for different chromatin loci for different coarse-graining sizes (Fig. 3(f) and Fig. S7(d, e)). From the distribution, we can derive an effective potential energy $V(l_{cg}) = -k_{\rm B}T \ln P(l_{cg})$ with which two neighboring beads interact. Even though the distribution is not perfectly Gaussian, a measure of the elastic constant of the interaction can be computed from the inverse of the standard deviation. Hence, we define an effective spring constant between two neighboring CG beads as $K_{cg} = \frac{k_B T}{\langle l_{cg}^2\rangle - \langle l_{cg}\rangle^2}$, where the angular brackets indicate the average computed using $P(l_{cg})$. In Fig. 3(g), we plot K_{cg} for different coarse-graining sizes and various gene loci.

The spring constant is scale dependent—it decreases as the coarse-graining size increases. For most gene regions, the spring constant values appear to saturate at a large coarse-graining scale, unlike the SAW polymer. The K_{cg} value for large n_b is in the range (0.1–1) $k_{\rm B}T/\sigma^2$, which is $\approx$ (1–10) pN/μm. This value is roughly comparable to some of the experimentally measured values from pulling long chromatin under certain in vitro conditions [79]. Note that, in contrast to pulling experiments where external forces can disrupt protein-mediated interactions, our estimate of the spring constant arises purely from thermal fluctuations and is thus expected to be a reliable signature of chromatin flexibility. Moreover, this is for relatively more dynamic MESc; hence the chromatin stretching stiffness obtained here will be less than that from the pulling experiments of the full-length mitotic chromosomes [80].

The spring constant above is presented in units of k_BT/σ^2 where σ is the size of a 200bp NL bead. However, in coarse-grained polymer simulations, one uses K_{cg} in units of k_BT/l_{cg}^2. Since $l_{\rm cg}$ also depends on CG size (n_b), the spring constant has a non-trivial behavior and is presented in Fig. 3(h). This gives a very useful range of numbers that can be used in all future coarse-grained simulations as $K_{cg} = 5$–$10\ k_B{\rm T}/l_{cg}^2$ for coarse-grained beads of size 1–20kb.

Predicting angle distribution and stiffness of coarse-grained chromatin segments

How flexible is a chromatin polymer segment? Do chromatin polymer segments have intrinsic curvature? While we understand the bendability of DNA reasonably well, we know very little about the bending elastic behavior of chromatin. From the large ensemble of structures that we have produced, consistent with nucleosome level Micro-C data, we computed the distribution of the bond angle ($\theta_{\rm cg}$)—angle between two neighboring bonds connecting three consecutive CG beads (see Fig. 4(a) top), and the dihedral angle ($\phi_{\rm cg}$)—angle between two neighboring planes formed by three consecutive bond vectors (Fig. 4(a) bottom).

The bond angle is defined as $\theta_{cg} = \pi - \alpha$, where $\alpha = \cos^{-1}(\hat{l}_{cg}^{i} \cdot \hat{l}_{cg}^{i+1})$, and it can take any value in the range $[0, \pi]$ (see SI Sec. S1.D). As a control, we computed the angle and its distribution for an ideal chain, and our results match well with the known analytical answer [78] for $P(\theta_{cg})$ for different coarse-graining sizes (Fig. S9(a)). Then we computed the distribution of angles $P(\theta_{cg})$ for chromatin segments in different epigenetic states. As shown in Fig. 4(b), for an ideal chain, even with no coarse-graining ($n_b = 1$), the distribution has a shape given by $P(\theta_{cg}) = \frac{1}{2}\sin(\theta_{cg})$ [78, 81]. This is due to the geometric measure, and it implies that when $\theta_{\rm cg}$ is near 90°, a large number of configurations are possible (having different azimuthal angles), while there is only one possible configuration for extreme cases if $\theta_{cg} = 0°$ and $\theta_{cg} = 180°$. Hence, to have a better understanding of the system, we also plot the corresponding probability density defined as $\tilde{P}(\theta_{cg}) = P(\theta_{cg})/\sin(\theta_{cg})$ in Fig. 4(c). For the ideal chain, with $n_b = 1$, $P(\theta_{cg})/\sin(\theta_{cg})$ is a flat curve (uniform distribution) reiterating the fact that the ideal chain is unbiased, and all configurations are equally likely. For SAW, the excluded volume would ensure that configurations with $\theta_{cg} \approx 0$ are not possible, and there is a natural bias towards extended configurations ($\theta_{cg} > 90°$).

The emergence of preferred inter-nucleosome angle from folded chromatin configurations: Our results here describe angle distribution for different of chromatin loci. For all the chromatin loci we simulated, at the nucleosomal (fine-grained, $n_b = 1$) resolution, a new peak emerges near $\theta_{cg} \approx 60°$ (Fig. 4(b)). The deviation from the ideal chain and SAW emerges due to intra-chromatin interactions. Since we do not impose any preferred angle in the fine-grained model, this population with angles near 60° emerges purely from the packaging, based on the contact probability map. To understand this better, we deconvoluted the $P(\theta_{cg})$ distribution and represented it as a sum

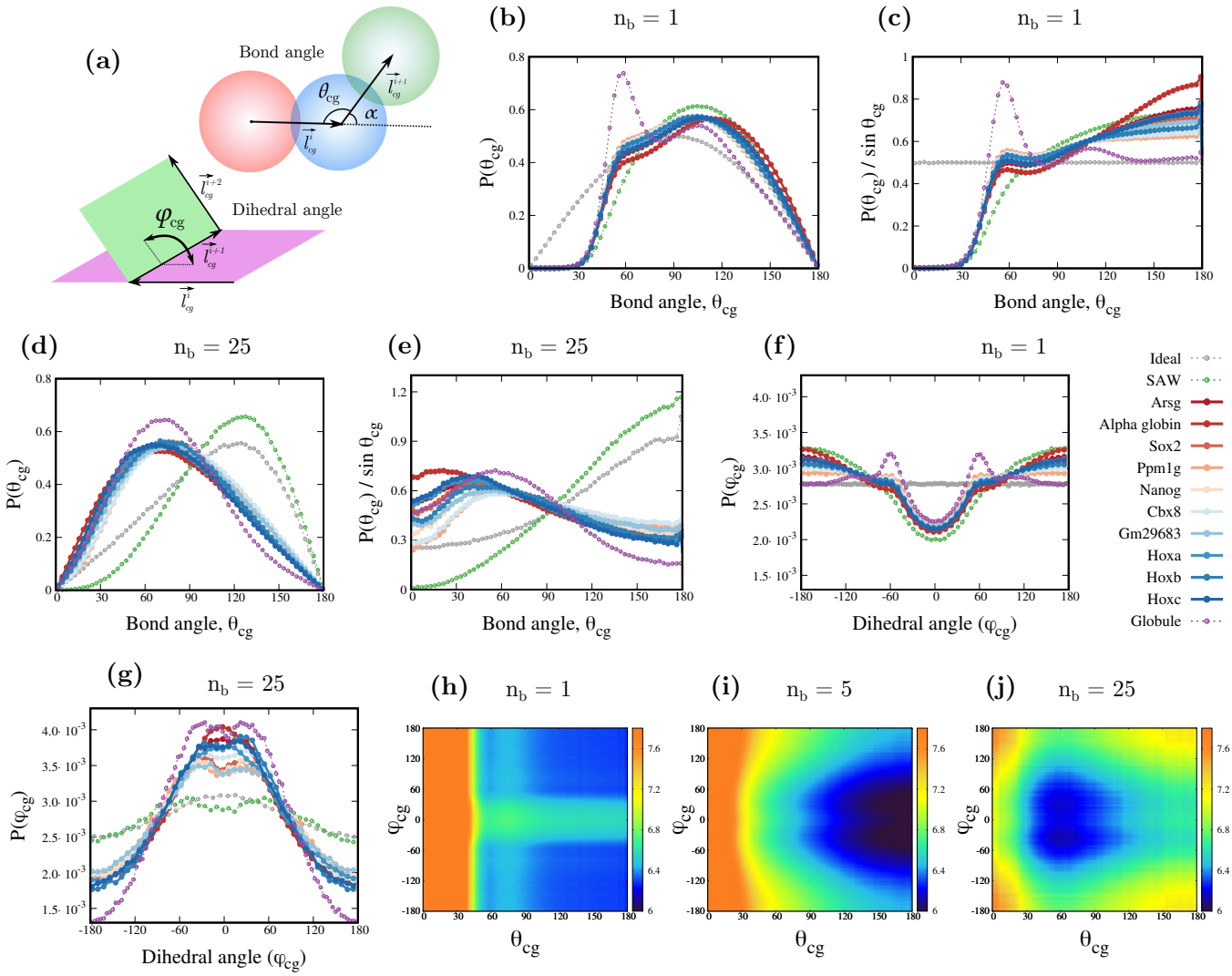

FIG. 4: (a) Schematic showing the bond angle θ_{cg} and dihedral angle ϕ_{cg} between coarse-grained beads. Distribution of angles for different chromatin loci, for the fine-grained model ($n_b = 1$) with nucleosome-linker (200bp) resolution (b) $P(\theta_{cg})$, and (c) the distribution with a different measure $P(\theta_{cg})/\sin(\theta_{cg})$ are shown. Similar angle distribution for the coarse-grained chromatin model with one bead representing 5kb ($n_b = 25$) of chromatin is shown in (d) $P(\theta_{cg})$ and (e)$P(\theta_{cg})/\sin(\theta_{cg})$. The distribution of dihedral angles for different chromatin loci for (f) the fine-grained model ($n_b = 1$) and (g) coarse-grained chromatin with one bead representing 5kb ($n_b = 25$) of chromatin. (θ_{cg}- ϕ_{cg}) energy plot for the euchromatic region for (h) the fine-grained model ($n_b = 1$), (i) CG chromatin with one bead representing 1kb chromatin ($n_b = 5$) and (j) CG chromatin with one bead representing 5kb chromatin ($n_b = 25$).

of two Gaussian distributions giving us two populations having mean values near $\approx 60°$ and $\approx 110°$ (Fig. S9(b)). Comparing the widths of the two populations suggests that the distribution with mean $\approx 60°$ is 2-3 times stiffer than the population with mean $\approx 110°$. For the highly folded globule, the peak at $\approx 60°$ is even more prominent, suggesting that tighter packaging could result in a population with $\approx 60°$ angles. In contrast to the prevalent notion of smaller angles around $\approx 60°$, our analysis shows that, at least in the case of mESC chromatin, there is a prominent signature of two sub-populations of angles, one highly folded and one extended, for $n_b = 1$ (Fig 4(b) and (c)).

Next, we examined the angle distribution of a coarse-grained chromatin polymer (Fig 4(d) and (e)). A lesser-discussed fact about polymers (even for the ideal chain) is that, when coarse-graining is performed ($n_b > 1$), the angle distribution gains a bias (or a shift), with a preference emerging for the larger θ_{cg} angles ($\theta_{cg} > 90°$) (See [78, 81] and Fig. S9(a)). The SAW polymer has an extra bias towards extended angles as smaller angles are disfavored due to excluded volume effects.

For coarse-gained chromatin, the angle distributions deviate a lot from the ideal chain and SAW, displaying a

preferred intrinsic angle around $\theta^0_{cg} \approx 60°$ for a coarse-graining scale of 5kb (Fig 4(d,e)). For chromatin loci, consistent with ideal chain and SAW, coarse-graining initially shifts the angle distribution towards larger angles (see $n_b = 5$ in Fig. S9(c-f)). However, for larger coarse-graining, long-range intra-chromatin interactions, such as TAD-forming loops, fold chromatin and shift the distribution towards smaller angles (see $n_b = 10, 50$ in Fig. S9(c-f)).

The effect of coarse-graining and deviation from ideal/SAW chain behavior is visible in $\tilde{P}(\theta_{cg})$ distribution as well (Fig. 4(e)). While the direct experimental readout of angles (e.g., via imaging) would yield $P(\theta_{cg})$, the scaled $\tilde{P}(\theta_{cg})$ is what would be useful for simulations; one can define an effective bond angle potential $V(\theta_{cg}) = -k_BT \log \tilde{P}(\theta_{cg})$ [81]. The $V(\theta_{cg})$ curves for chromatin segments have a well-defined minimum at preferred angles, which depends on the coarse-graining scale, n_b (Fig. S10). One can also compute the effective bending "stiffness" of chromatin segments by comparing the inverse of the standard deviation around the local maxima of $\tilde{P}(\theta_{cg})$ or by equating $V(\theta_{cg}) = \frac{k_b}{2}(1 + cos(\theta_{cg} - \theta^0_{cg}))$. We find that bending stiffness is of the order of thermal energy $k_b \approx k_BT$ for all coarse-graining scales. This suggests that the chromatin polymer is not highly stiff and explores a wide range of angles. This is consistent with the emerging notion that chromatin is highly dynamic [82–85], and has high cell-to-cell variability.

We examined how the angles vary along the genomic locations. Similar to R_g and $l_{\rm cg}$, angles too have heterogeneity along the genome (Fig. 2(c,d) and Fig. S9(g,h)). A comparison of the average angle for different genomic locations reveals that there are higher angles near TAD-like domain boundaries and lower angles in the interior of the domains. This spatial variation of chromatin properties could be important for understanding and reconstructing chromatin configurations.

Dihedral angle distribution: The distributions of the dihedral angles $P(\phi_{\rm cg})$ for different chromatin/polymer segments for fine-grained ($n_b = 1$) and coarse-grained ($n_b = 25$) level are shown in Fig. 4(f–g). For the fine-grained model ($n_b = 1$), the ideal chain (control) has a uniform angle distribution as expected; the SAW polymer has a dip near $\phi = 0$ indicating self-avoidance/steric hindrance (Fig 4(f)). Even for an ideal chain, the coarse-graining leads to non-trivial changes in the ϕ distribution where a preference for smaller dihedral angles arises. For the fine-grained chromatin and globule, due to high folding, the probability of obtaining smaller angles (ϕ near zero) increases, and larger angles become rarer compared with the SAW polymer. Folding also leads to a peak near $\phi \approx 60°$, which is prominent for the globule.

A preference for smaller ϕ values appears on coarse-graining, similar to the Ideal chain. At the same time, the folding via long-range intra-chromatin interactions results in the formation of peaks near $\phi \approx \pm 60°$. Both of these effects together define the coarse-grained ϕ distributions (Fig. 4(g)). Similar to the bending angle distribution, the dihedral angle distributions are very broad, implying a weak angle stiffness.

Similar to $V(\theta_{\rm cg})$, we define an effective dihedral potential $V(\phi_{cg}) = -k_BT \log P(\phi_{cg})$ (see Fig. S11). Assuming that the distributions of $\theta_{\rm cg}$ and $\phi_{\rm cg}$ are independent, we have plotted the heatmap of $V(\theta_{cg}, \phi_{cg}) = V(\theta_{cg}) + V(\phi_{cg})$ in the ($\theta_{\rm cg}$ – $\phi_{\rm cg}$) plane (see Fig 4(h–j)). Here the color-bar represents the energy $V(\theta_{cg}, \phi_{cg})$ values in $k_{\rm B}$T units. For the fine-grained model, low values of $\theta_{\rm cg}$ and $\phi_{\rm cg}$ are penalized due to self-avoidance (Fig. 4(h)). This effect reduces with coarse-graining. For lower coarse-graining, higher $\theta_{\rm cg}$ and intermediate $\phi_{\rm cg}$ values are preferred (Fig. 4(i)), while for higher coarse-graining $\theta_{\rm cg}$ in the range 50°–90° and $\phi_{\rm cg}$ values close to $\pm 60°$ are favored (Fig. 4(j)). This again shows that angle preferences for chromatin are scale-dependent—depending on the coarse-graining scale, the preferred values vary considerably.

Determining optimal soft inter-bead potential and simulating a coarse-grained chromatin

This work systematically estimates the size of coarse-grained beads, their fluctuations, the overlap among the beads due to the mixing of polymer segments, and the distribution of bond and dihedral angles. Here we integrate these quantities to simulate a coarse-grained chromatin polymer and predict 3D size or R_g measurable in microscopy experiments. While the spring constants and bead sizes ($l_{\rm cg}$) can be directly used from our results discussed so far, we lack the non-bonded interaction potential that would ensure appropriate compaction. Therefore, we perform an iterative Boltzmann inversion (IBI) to determine the form of an inter-bead soft potential that would achieve the 3D distance distribution consistent with our original fine-grained simulation (see Sec. S1.C and Fig. S12).

We implemented the iterative Boltzmann inversion method for the Arsg locus that we studied using the fine-grained model. Since the overlap (Fig. 3(e)) depends on the level of coarse-graining, it is expected that the potential energy would also depend on n_b. Hence, for each n_b, we simulated a coarse-grained bead-spring polymer with N/n_b beads connected by harmonic bonds with equilibrium bond length l_{cg} and spring constant K_{cg} taken from Fig. 3(c, h). Starting with a flat inter-bead potential energy $V^{nb}_{i=0} = 0$, at each step of iteration i, we simulated the polymer until

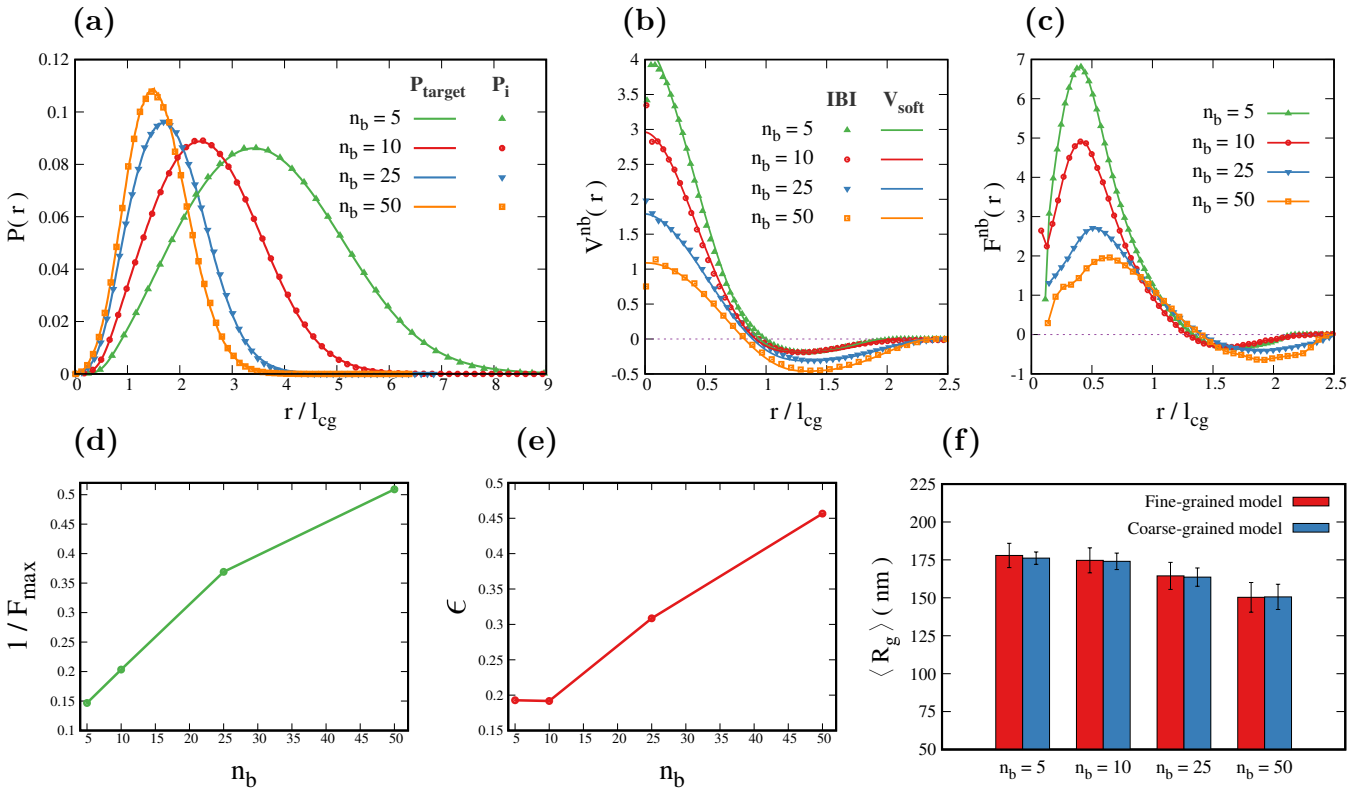

FIG. 5: (a) The distribution of distances between all non-bonded beads obtained from the iterative Boltzmann inversion (IBI) method after convergence $P_i(r)$ (points) matches well with the fine-grained chromatin target distribution $P_{\text{target}}(r)$ (solid lines) for appropriate levels of coarse-graining. (b) The non-bonded potential obtained from the IBI method (points) and the functional form for the soft potential $V_{\text{soft}}(r)$ (solid lines using Eq. 1; see parameters in Table S2) for different values of n_b. (c) The force derived from the potential (d) The inverse of the maximum value of the force as a function of n_b as a measure of softness. (e) The depth (ϵ) of the V^{nb} potential is plotted as a function of coarse-graining size. (f) The radius of gyration measured from the coarse-grained simulation with soft potential is compared with the R_g from the fine-grained model for corresponding levels of coarse-graining.

equilibrium and updated this potential using the relation:

$$V^{nb}_{i+1}(r) = V^{nb}_i(r) + \alpha(r)\ k_{\text{B}}T\ \ln\left(\frac{P_i(r)}{P_{\text{target}}(r)}\right).$$

Here $P_i(r)$ is the steady-state distribution of distances between all pairs of non-bonded beads. The distribution was compared with the known distance distribution, a target distribution $P_{\text{target}}(r)$ for the corresponding level of coarse-graining from our fine-grained model. $\alpha(r) = 0.2\ e^{-r^2/2}$ is taken as a decaying function to ensure that the resulting potential is short-range. We checked for the convergence of the algorithm by computing the Kullback-Leibler Divergence between the target and CG model distance distributions (see Sec. S1.C and Fig. S13(a)). In other words, for each n_b, we have computed a soft potential energy function between CG bead pairs and used it to perform CG polymer simulations that would reproduce 3D distance distribution exactly as we got from our fine-grained model (see Fig. 5(a)). The resulting potential energy functions V^{nb} for various n_b values are shown in Fig. 5(b). We also fit a functional form to this potential (solid line), such that the softness and depth of the potential can be tuned independently. We use the functional form

$$V_{\text{soft}}(r) = \begin{cases} V_0\left[1 - \left(\frac{r}{r_m}\right)^{\eta_1}\right]^{\eta_2} - \epsilon & r < r_m, \\ \frac{1}{2}\epsilon\left[\cos(\mu r^2 + \nu) - 1\right] & r_m \leqslant r < r_c, \\ 0 & r \geqslant r_c. \end{cases} \tag{1}$$

The first part of the equation ($r < r_m$) gives the repulsive part of the potential [86]. Here, V_0 controls the height of the potential at $r = 0$ (see Fig. S13(b)), r_m is the position of minima, and ϵ denotes the depth of the potential.

The parameters η_1 and η_2 can be tuned to get the desired softness. The second part of the equation ($r_m \leqslant r < r_c$) represents the attractive part of the potential [24, 87, 88]. This function has the advantage that it can ensure the continuity and differentiablity at $r = r_m$ and $r = r_c$ by tuning the values of μ and ν such that the value of potential is $V_{\text{soft}}(r = r_m) = -\epsilon$ and $V_{\text{soft}}(r = r_c) = 0$ (see SI Sec. S1.C and Table S2). The negative slope of the corresponding potential $F^{nb} = \frac{-dV^{nb}}{dr}$ is plotted in Fig. 5(c). The inverse of the maximum value of the force ($1/F_{max}$) can be used as a measure of the softness of CG beads (Fig. 5(d)).The depth of the potential captures the effective attractive interaction between a pair of beads (Fig. 5(e)). The important points to note are: (i) the potential is derived from fine-grained model that is consistent with the Micro-C experimental data. (ii) The potential is highly soft – softer than the typically used LJ potential (Fig. S13(c)). (iii) The potential energy and the two important physical parameters of the potential — softness and the attractive interaction strength — are scale-dependent. Different levels of coarse-graining have different softness and interaction strength. This is highly relevant for anyone wanting to simulate chromatin as a coarse-grained bead spring polymer.

Finally, we compare the radius of gyration of the chromatin polymer predicted by our CG simulations with our fine-grained model for various levels of coarse-graining (Fig. 5(f)). The radius of gyration values match with the fine-grained model. Note that this is equivalent to comparing a coarse-grained model simulation results with microscopy experiments that label DNA/chromatin (e.g., methods that "paint" chromatin [47]) with equivalent resolution. The radius of gyration of both fine-grained and coarse-grained polymers decreases slightly with increasing coarse-graining. This is because the distance of a CG bead from the center of mass of the polymer is smaller than the root mean square distance of the fine-grained beads it replaces (see Fig. S13(d)). This also predicts that the overall R_g value of a long chromatin region (made of many painted small segments) will marginally decrease as one increases the length of the labeled (painted) segment. This decrease is of course less than the size of the painted segment (l_{cg}). As mentioned elsewhere in this manuscript, we find that the most probable value that we predict for $l_{\text{cg}}$is comparable with the available data from chromatin microscopy experiments that paint 5kb segments.

Discussion and Conclusion

This paper addresses a fundamental question in modeling chromatin: what are the properties and parameters of a coarse-grained chromatin polymer, and how do they vary in a scale-dependent manner as we go from the $\sim$ 10 nm nucleosome scale to hundreds of nanometers gene scale, domain scale or micron-sized chromatin scale? Recent papers have given us a good understanding of the scaling laws, TAD formation, roles of phase separation, loop extrusion, and so on [72, 84, 89–93]. However, we do not understand the physical dimension of loci that we consider a "bead" in simulations, how stretchable chromatin loci are (spring constant), angle flexibility (bendability), how soft the inter-bead potentials are, and so on. We do not know how chromatin compaction (R_g), spring constant, bending angle, overlap, etc., depend on the local contact map (e.g., TAD) structures and epigenetic states.

To fill this gap, we used the recently published Micro-C contact map for mESCs and constructed an ensemble of chromatin configurations at 200bp resolution. These configurations simultaneously satisfy three constraints: (i) they comply with the Micro-C contact probability, (ii) the mean 3D distance values computed from the configurations are comparable with known experiments, and (iii) the size of the 200bp fine-grained bead (nucleosome + linker) is in a sensible range. We used this set of configurations and systematically coarse-grained them to predict physical properties and parameters relevant to a chromatin polymer bead-spring chain. We have determined the physical dimensions of chromatin loci (bead sizes of chromatin polymer) for ten different mESC gene regions having different epigenetic state characteristics. We have computed the distributions of the inter-bead distances, predicting how stretchable different chromatin loci are and quantifying their spring constants. We have also predicted the bending and dihedral angle fluctuations revealing how bendable chromatin loci are. Our work not only shows the similarity/variability among different loci but also reveals the effect of chromatin heterogeneity along the polymer contour, finding that TAD interior and TAD boundary have different properties and parameters — different CG bead dimensions, average angle values, overlap, etc. Contrary to the prevalent notion, our results show that CG chromatin beads should be modeled as soft particles that can overlap. We then compute the inter-bead soft potential and propose a functional form to quantify the softness. All our predictions reveal how chromatin properties and parameters change in a scale-dependent manner.

The chromatin polymer parameter values that we have predicted— bead sizes, spring constants, angle distributions, overlap/softness, etc. — are essential for anyone wanting to simulate chromatin polymer. To the best of our knowledge, this is the first comprehensive prediction of numerical values of all parameters starting with nucleosome resolution data. Moreover, our finding that chromatin polymer parameters depend on the scale one chooses to study is significant. The polymer parameters relevant for 1kb chromatin are not the same as that for 10kb or 100kb chromatin, which is essential to account for in future simulations. We also argue that many of these parameters (like overlap) are crucial for predicting 3D distance accurately. We have determined an effective inter-bead potential via an iterative Boltzmann

inversion method. We used all of these CG results to compute the R_g of a chromatin locus. In other words, our claim is: we have computed the relevant parameters for a polymer simulation of mESC chromatin at different scales. Anyone can use our parameters, simulate coarse-grained chromatin satisfying contact probability, and predict average 3D distances reasonably well within the region-to-region variability we show. Our work has biological significance for connecting chromatin structure to function. Many of the biological processes like recombination, DNA breakage/repair, enhancer-activation, and spreading of histone modifications occur at the scale of nucleosomes. The 3D structure we predict at nucleosome resolution is crucial for understanding these functional aspects. Our work connects the coarse-grained picture (100 nm to μm scale experiments having a few kb or Mb resolution) with a nucleosome-resolution picture and will enable Hi-C or Microscopy experiments to extrapolate and predict nucleosome-level structure. This is highly relevant for understanding the biological functions that occur at nucleosome resolution.

While building the fine-grained model, we made minimal assumptions. The primary assumption we made is that all the chromatin details (e.g., inter-nucleosome interaction potential, histone tails, etc.) result in deciding the contact probability; generating conformations that satisfy the contact map would implicitly account for the role of various local chemical and structural details. Since we use the Micro-C data with 200bp resolution, our model (model-I) cannot study details below this resolution. We also employed a model with linker DNA (model-II) and showed that our model-I results are sensible. Using model II, we also examined how the variability in nucleosome positioning would affect the overall size (R_g) of the folded chromatin. Apart from studying a fixed linker length of 50 bp, we have also performed simulations choosing linker lengths from a Gaussian distribution to incorporate variability. If there are 5 ± 1 nucleosomes in 1kb chromatin, the mean R_g is roughly the same order of magnitude as we reported (Fig. S2(d)). We have also reported these quantities for different mean linker length values. The difference in mean may represent different chromatin states.

For the ten gene loci we studied, heterochromatic and euchromatic regions have R_g, angles, and other quantities in a comparable range. This could be because (i) our study is for an embryonic stem cell where the chromatin could be more open. (ii) The underlying Micro-C data itself shows that heterochromatic and euchromatic regions have comparable contact probability as a function of genomic distance $P(s)$ (Fig. S4(e)). This can also be consistent with the irregular nature of chromatin organization as indicated by the power law decay of $P(s)$. Recent experiments have also indicated that heterochromatin can be diverse, and euchromatin can get highly folded due to multiple loops, resulting in similar compaction and other physical properties [48, 94, 95].

Our work predicts the mean values and variability in bead sizes and other physical properties like elasticity and bendability. It has been suggested that the variability in thickness and flexibilities could affect the chromatin properties below 100kb [96]. This implies that the variability we find may be relevant since many of the enhancers and promoters can be within 100kb [40, 97]. However, note that, apart from variability, we predict the average bead size, $l_{\rm cg}$, K_{cg}, etc.; the change in the average value would affect measurable quantities at all length scales.

One of the important results of our work is the inter-bead soft potential and the quantification of overlap. Very high-resolution models or models that used sub-beads to represent a larger CG bead would have some signatures of overlap [21, 72]. However, most of the current coarse-grained simulation studies use the Lennard-Jones potential for inter-bead interactions, and it quickly goes to infinity with negligible softness. Unlike earlier models [12, 13, 21, 72, 86], here we derive the functional form of the soft-potential starting with nucleosome-resolution contact map data and quantify the softness in a scale-dependent manner. One of the concerns regarding the soft potentials is that it may allow chain crossing leading to incorrect dynamics. However, recent experiments show that chain crossings are indeed present, and topoisomerase activity is required to remove these crossings and have entanglement-free interphase chromosomes [98]. This implies that more accurate dynamics would require the presence of enzymes like topoisomerase that actively regulate chromosome topology in terms of entanglements. This may be an essential feature necessary to study dynamics in coarse-grained models.

Experimental tests of our predictions: We simulate the fine-grained model (scale $\sim 10 - 20$ nm) and predict quantities at a much larger scale (~ 100 nm$-\mu$m) that can be measured in experiments. Our predictions of the radius of gyration, bond length, and 3D distances can be tested using microscopy experiments, and we have compared some of them in Fig. 1(e) and Fig. 2(e). Combining biochemistry and microscopy, recent studies have proposed methods to "paint" chromatin segments (size $\approx$ 5kb or higher) and trace the chromatin contour. This method allows one to test many polymer predictions, including coarse-grained inter-bead distances and angle fluctuations. Even though the experimental data is not available for the mESC segments that we simulated, we compared our predictions with the available data, and we found that the most probable value that we predict for $l_{\rm cg}$ is comparable with the measured data [47]. We also find that our prediction of the fluctuation of the angles — width of the angle distribution — is comparable to the experimentally measured values. Such experiments may be performed for the mESC gene regions we simulated to compare with our predictions. Future experiments could also test how these values change as one changes the segment size indicating how bead sizes and bendability would vary with the choice of the coarse-graining scale. Future experiments may also measure spring constants at different scales, either through measuring chromatin segment fluctuations or doing pulling experiments at various scales. All of our predictions can be tested

using microscopy, chromatin pulling, and other biophysical experiments.

It must be stated that the whole of our analysis is based on the Micro-C data for the mESCs from Hsieh et al. [35]. Hence the numbers emerging from this study would represent embryonic stem cell chromatin. In the future, analysis can be further extended to study various other cell types as new data emerge. The future direction is also to understand the role of nucleosome positioning heterogeneity and assembly/disassembly/sliding kinetics. It requires a much more detailed polymer model [99] and a model to understand how chromatin conformation capture contact maps are influenced by the heterogeneity of nucleosome organization.

Acknowledgment

RP Acknowledges useful discussions with Xavier Darzacq, Leonid Mirny, Daniel Jost, Geeta Narlikar, and Marc Marti-Renom. We acknowledge useful discussions with Vladimir Teif, Mayuri Rege, PB Sunil Kumar, Madan Rao, and Gaurav Bajpai. SK acknowledges fellowship support from the CSIR, India, and KK acknowledges iPDF support from IIT Bombay. We acknowledge funding from the Department of Biotechnology, India (Grant number: BT/HRD/NBA/39/12/2018-19). We also acknowledge the National Supercomputing Mission (NSM) for providing computing resources of 'PARAM Brahma' at IISER Pune, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India.

[1] B. Alberts, Molecular Biology of The Cell (Garland Science, Taylor and Francis Group, New York, 2014), 6th ed.
[2] W. A. Bickmore, Annu. Rev. Genomics Hum. Genet. **14**, 67 (2013).
[3] R. D. Kornberg and Y. Lorch, Cell **98**, 285 (1999).
[4] M. J. Rowley and V. G. Corces, Nature Reviews Genetics **19**, 1 (2018), ISSN 1471-0056, URL `http://www.nature.com/articles/s41576-018-0060-8`.
[5] C. B. Hug and J. M. Vaquerizas, Trends in Genetics **34**, 903 (2018).
[6] H. K. Long, S. L. Prescott, and J. Wysocka, Cell **167**, 1170 (2016).
[7] B. Bonev and G. Cavalli, Nature Reviews Genetics **17**, 661 (2016).
[8] A. D. Stephens, E. J. Banigan, and J. F. Marko, Current opinion in cell biology **58**, 76 (2019).
[9] O. Oluwadare, M. Highsmith, and J. Cheng, Biological procedures online **21**, 1 (2019).
[10] S. Bianco, A. M. Chiariello, M. Conte, A. Esposito, L. Fiorillo, F. Musella, and M. Nicodemi, Current Opinion in Cell Biology **64**, 10 (2020).
[11] M. Di Stefano, J. Paulsen, D. Jost, and M. A. Marti-Renom, Curr. Opin. Genet. Dev. **67**, 25 (2021).
[12] L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, Cell **157**, 950 (2014).
[13] M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, and J. N. Onuchic, PNAS **113**, 12168 (2016).
[14] Q. MacPherson, B. Beltran, and A. J. Spakowitz, PNAS **115**, 12739 (2018).
[15] C. A. Brackley, J. M. Brown, D. Waithe, C. Babbs, J. Davies, J. R. Hughes, V. J. Buckle, and D. Marenduzzo, Genome biology **17**, 1 (2016).
[16] C. A. Brackey, D. Marenduzzo, and N. Gilbert, Nature Methods **17**, 767 (2020).
[17] G. Bajpai, D. A. Pavlov, D. Lorber, T. Volk, and S. Safran, Elife **10**, e63976 (2021).
[18] M. Falk, Y. Feodorova, N. Naumova, M. Imakaev, B. R. Lajoie, H. Leonhardt, B. Joffe, J. Dekker, G. Fudenberg, I. Solovei, et al., Nature **570**, 395 (2019).
[19] A. Goloborodko, J. F. Marko, and L. A. Mirny, Biophys. J. **110**, 2162 (2016).
[20] S. K. Ghosh and D. Jost, PLoS computational biology **14**, e1006159 (2018).
[21] S. Bianco, D. G. Lupiáñez, A. M. Chiariello, C. Annunziatella, K. Kraft, R. Schöpflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, et al., Nat. Genet. **50**, 662 (2018).
[22] G. Shi, L. Liu, C. Hyeon, and D. Thirumalai, Nat. Commun. **9**, 1 (2018).
[23] N. Ganai, S. Sengupta, and G. I. Menon, Nucleic Acids Res. **42**, 4145 (2014).
[24] K. Kumari, B. Duenweg, R. Padinhateeri, and J. R. Prakash, Biophys. J. **118**, 2193 (2020).
[25] D. Chatenay, S. Cocco, R. Monasson, D. Thieffry, and J. Dalibard, Multiple Aspects of DNA and RNA: from Biophysics to Bioinformatics: Lecture Notes of the Les Houches Summer School 2004 (Elsevier, 2005).
[26] C. Bustamante, Z. Bryant, and S. B. Smith, Nature **421**, 423 (2003).
[27] J. F. Marko and S. Cocco, Physics World **16**, 37 (2003).
[28] M. Mir, W. Bickmore, E. E. Furlong, and G. Narlikar, Development **146**, dev182766 (2019).
[29] S. L. Klemm, Z. Shipony, and W. J. Greenleaf, Nature Reviews Genetics **20**, 207 (2019).
[30] 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, et al., Science **326**, 289 (2009).

[31] 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, et al., Nature **485**, 381 (2012).
[32] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, Nature **485**, 376 (2012).
[33] S. S. 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, et al., Cell **159**, 1665 (2014).
[34] T.-H. S. Hsieh, A. Weiner, B. Lajoie, J. Dekker, N. Friedman, and O. J. Rando, Cell **162**, 108 (2015).
[35] T.-H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, and X. Darzacq, Molecular cell **78**, 539 (2020).
[36] N. Krietenstein and O. J. Rando, Current Opinion in Genetics & Development **61**, 32 (2020).
[37] R. Jäger, G. Migliorini, M. Henrion, R. Kandaswamy, H. E. Speedy, A. Heindl, N. Whiffin, M. J. Carnicer, L. Broome, N. Dryden, et al., Nature communications **6**, 1 (2015).
[38] V. Ramani, D. A. Cusanovich, R. J. Hause, W. Ma, R. Qiu, X. Deng, C. A. Blau, C. M. Disteche, W. S. Noble, J. Shendure, et al., Nature protocols **11**, 2104 (2016).
[39] S. G. Swygert, S. Kim, X. Wu, T. Fu, T.-H. Hsieh, O. J. Rando, R. N. Eisenman, J. Shendure, J. N. McKnight, and T. Tsukiyama, Molecular Cell **73**, 533 (2019).
[40] Z. Islam, B. Saravanan, K. Walavalkar, U. Farooq, A. K. Singh, S. Radhakrishnan, J. Thakur, A. Pandit, S. Henikoff, and D. Notani, Genome Research **33**, 1 (2023).
[41] K. Guin, Y. Chen, R. Mishra, S. R. B. Muzaki, B. C. Thimmappa, C. E. O'Brien, G. Butler, A. Sanyal, and K. Sanyal, Elife **9**, e58556 (2020).
[42] Q. Szabo, A. Donjon, I. Jerković, G. L. Papadopoulos, T. Cheutin, B. Bonev, E. P. Nora, B. G. Bruneau, F. Bantignies, and G. Cavalli, Nat. Genet. **52**, 1151 (2020).
[43] R. Imai, T. Nozaki, T. Tani, K. Kaizu, K. Hibino, S. Ide, S. Tamura, K. Takahashi, M. Shribak, and K. Maeshima, Mol. Biol. Cell **28**, 3349 (2017).
[44] K. Maeshima, S. Ide, and M. Babokhov, Current Opinion in Cell Biology **58**, 95 (2019), ISSN 0955-0674, cell Nucleus, URL `http://www.sciencedirect.com/science/article/pii/S0955067418301881`.
[45] M. Eltsov, D. Grewe, N. Lemercier, A. Frangakis, F. Livolant, and A. Leforestier, Nucleic acids research **46**, 9189 (2018).
[46] H. D. Ou, S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'Shea, Science **357**, 1 (2017), ISSN 0036-8075, URL `http://science.sciencemag.org/content/357/6349/eaag0025/tab-pdf`.
[47] A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M. Imakaev, L. A. Mirny, C.-t. Wu, and X. Zhuang, Nature **529**, 418 (2016).
[48] H. A. Shaban, R. Barth, L. Recoules, and K. Bystricky, Genome biology **21**, 1 (2020).
[49] J. J. Parmar, M. Woringer, and C. Zimmer, Annual Review of Biophysics **48**, 231 (2019), ISSN 1936-122X.
[50] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, science **295**, 1306 (2002).
[51] Y. Nishino, M. Eltsov, Y. Joti, K. Ito, H. Takata, Y. Takahashi, S. Hihara, A. S. Frangakis, N. Imamoto, T. Ishikawa, et al., The EMBO Journal **31**, 1644 (2012), https://www.embopress.org/doi/pdf/10.1038/emboj.2012.35, URL `https://www.embopress.org/doi/abs/10.1038/emboj.2012.35`.
[52] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucleic Acids Res. **42**, 9553 (2014).
[53] J. Huertas, E. J. Woods, and R. Collepardo-Guevara, Current Opinion in Cell Biology **75**, 102067 (2022).
[54] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, Cell Rep. **15**, 2038 (2016).
[55] J. J. Parmar and R. Padinhateeri, Curr. Opin. Struct. Biol. **64**, 111 (2020).
[56] G. D. Bascom, C. G. Myers, and T. Schlick, PNAS **116**, 4955 (2019).
[57] A. Rosa and R. Everaers, PLoS Comput. Biol. **4**, e1000153 (2008).
[58] C. T. Clarkson, E. A. Deeks, R. Samarista, H. Mamayusupova, V. B. Zhurkin, and V. B. Teif, Nucleic Acids Res. **47**, 11181 (2019).
[59] M. Conte, L. Fiorillo, S. Bianco, A. M. Chiariello, A. Esposito, and M. Nicodemi, Nat. Commun. **11**, 3289 (2020).
[60] Y. Qi and B. Zhang, PLoS Comput. Biol. **15**, e1007024 (2019).
[61] G. Bajpai and R. Padinhateeri, Biophys. J. **118**, 207 (2020).
[62] H. Tjong, W. Li, R. Kalhor, C. Dai, S. Hao, K. Gong, Y. Zhou, H. Li, X. J. Zhou, M. A. Le Gros, et al., Proceedings of the National Academy of Sciences **113**, E1663 (2016).
[63] N. Hua, H. Tjong, H. Shin, K. Gong, X. J. Zhou, and F. Alber, Nature protocols **13**, 915 (2018).
[64] G. Tiana, A. Amitai, T. Pollex, T. Piolot, D. Holcman, E. Heard, and L. Giorgetti, Biophysical journal **110**, 1234 (2016).
[65] M. M. Tortora, H. Salari, and D. Jost, Current opinion in genetics & development **61**, 37 (2020).
[66] M. Ohno, T. Ando, D. G. Priest, V. Kumar, Y. Yoshida, and Y. Taniguchi, Cell **176**, 520 (2019).
[67] S. Guha and M. K. Mitra, Soft Matter **19**, 153 (2023).
[68] R. Collepardo-Guevara and T. Schlick, Proceedings of the National Academy of Sciences **111**, 8061 (2014), ISSN 0027-8424, https://www.pnas.org/content/111/22/8061.full.pdf, URL `https://www.pnas.org/content/111/22/8061`.
[69] S. E. Farr, E. J. Woods, J. A. Joseph, A. Garaizar, and R. Collepardo-Guevara, Nature communications **12**, 1 (2021).
[70] S. A. Grigoryev, Biophysical journal **114**, 2290 (2018).
[71] K. Kumari, J. R. Prakash, and R. Padinhateeri, Biophysical Journal (2022).
[72] A. M. Chiariello, S. Bianco, A. M. Oudelaar, A. Esposito, C. Annunziatella, L. Fiorillo, M. Conte, A. Corrado, A. Prisco, M. S. Larke, et al., Cell Reports **30**, 2125 (2020).
[73] G. Forte, A. Buckle, S. Boyle, D. Marenduzzo, N. Gilbert, and C. A. Brackley, bioRxiv (2021).
[74] S. Plimpton, Journal of computational physics **117**, 1 (1995).
[75] G. Bajpai, I. Jain, M. M. Inamdar, D. Das, and R. Padinhateeri, PLoS computational biology **13**, e1005365 (2017).

[76] T. Yang, F. Zhang, G. G. Yardımcı, F. Song, R. C. Hardison, W. S. Noble, F. Yue, and Q. Li, Genome research **27**, 1939 (2017).
[77] J. M. Brown, N. A. Roberts, B. Graham, D. Waithe, C. Lagerholm, J. M. Telenius, S. De Ornellas, A. M. Oudelaar, C. Scott, I. Szczerbal, et al., Nature communications **9**, 1 (2018).
[78] M. Laso, H. Öttinger, and U. Suter, The Journal of chemical physics **95**, 2178 (1991).
[79] Y. Cui and C. Bustamante, Proceedings of the National Academy of Sciences **97**, 127 (2000).
[80] M. G. Poirier and J. F. Marko, Proceedings of the National Academy of Sciences **99**, 15393 (2002).
[81] T. Vettorel, G. Besold, and K. Kremer, Soft Matter **6**, 2282 (2010).
[82] A. S. Hansen, C. Cattoglio, X. Darzacq, and R. Tjian, Nucleus **9**, 20 (2018).
[83] E. I. Prieto and K. Maeshima, Essays in biochemistry **63**, 133 (2019).
[84] M. Gabriele, H. B. Brandão, S. Grosse-Holz, A. Jha, G. M. Dailey, C. Cattoglio, T.-H. S. Hsieh, L. Mirny, C. Zechner, and A. S. Hansen, Science **376**, 496 (2022).
[85] A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, and X. Darzacq, elife **6**, e25776 (2017).
[86] S. Fujishiro and M. Sasai, Proceedings of the National Academy of Sciences **119**, e2109838119 (2022).
[87] T. Soddemann, B. Dünweg, and K. Kremer, Eur. Phys. J. E **6**, 409 (2001).
[88] A. Santra, K. Kumari, R. Padinhateeri, B. Dünweg, and J. R. Prakash, Soft Matter **15**, 7876 (2019).
[89] B. Bintu, L. J. Mateo, J.-H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, and X. Zhuang, Science **362** (2018).
[90] B. A. Gibson, L. K. Doolittle, M. W. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, and M. K. Rosen, Cell **179**, 470 (2019).
[91] M. Conte, E. Irani, A. M. Chiariello, A. Abraham, S. Bianco, A. Esposito, and M. Nicodemi, Nature Communications **13**, 4070 (2022).
[92] D. Racko, F. Benedetti, J. Dorier, and A. Stasiak, Nucleic Acids Research **47**, 521 (2019).
[93] Z. Tang, O. J. Luo, X. Li, M. Zheng, J. J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Ruszczycki, et al., Cell **163**, 1611 (2015).
[94] A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, and N. Gilbert, Molecular cell **72**, 786 (2018).
[95] G. Spracklin, N. Abdennur, M. Imakaev, N. Chowdhury, S. Pradhan, L. A. Mirny, and J. Dekker, Nature Structural & Molecular Biology pp. 1–14 (2022).
[96] A.-M. Florescu, P. Therizols, and A. Rosa, PLoS computational biology **12**, e1004987 (2016).
[97] S. Oh, J. Shao, J. Mitra, F. Xiong, M. D'Antonio, R. Wang, I. Garcia-Bassets, Q. Ma, X. Zhu, J.-H. Lee, et al., Nature **595**, 735 (2021).
[98] E. M. Hildebrand, K. Polovnikov, B. Dekker, Y. Liu, D. L. Lafontaine, A. N. Fox, Y. Li, S. V. Venev, L. A. Mirny, and J. Dekker, bioRxiv pp. 2022–10 (2022).
[99] W. Lu, J. N. Onuchic, and M. Di Pierro, PLOS Computational Biology **19**, e1011013 (2023).