Analysis of Genome Architecture Mapping Data with a Machine Learning and Polymer-Physics-Based Tool

Luca Fiorillo, Mattia Conte, Andrea Esposito, Francesco Musella, Francesco Flora, Andrea M. Chiariello, and Simona Bianco

Dipartimento di Fisica, Università di Napoli Federico II, and INFN Napoli, Complesso Universitario di Monte Sant’Angelo, 80126 Naples, Italy
{chiariello,simona.bianco}@na.infn.it

Abstract. Understanding the mechanisms driving the folding of chromosomes in nuclei is a major goal of modern Molecular Biology. Recent technological advances in microscopy (FISH, STORM) and sequencing approaches (Hi-C, GAM, SPRITE) enabled to collect quantitative data about chromatin 3D architecture, revealing a non-random and highly specific organization. To transform such tremendous amount of data into valuable insights on genome folding, heavy computational analyses are required. Here, we study the performances of PRISMR, a computational tool based on Machine Learning strategies and Polymer Physics principles, to explore genome 3D structure from Genome Architecture Mapping (GAM) data. Using such data, we show that PRISMR can successfully reconstruct the 3D structure of real genomic regions at various length scales, from mega-base sized loci to whole chromosomes. Importantly, the inferred structures are validated against independent Hi-C data. Finally, we show how PRISMR can be effectively employed to explore differences between experimental methods.

Keyword: Chromatin organization · Machine learning · GAM

1 Introduction

In cell nuclei, chromosomes are organized in a very complex architecture. In recent years, the three-dimensional (3D) structure of chromosomes has been investigated by novel, sophisticated technologies. These include sequencing methods – such as Hi-C [1], Genome Architecture Mapping (GAM) [2], SPRITE [3] - able to detect contacts between pairs of DNA sites genome-wide; microscopy techniques, measuring distances between loci in specific DNA regions [4, 5]. The huge amount of data produced by all these technologies has shown that the 3D organization of chromosome, far from being random, plays a key role for gene activity and transcriptional regulation [6–8]. For instance, it has been shown that chromosomes segregate in specific territories inside nuclei [9] and functional loops occur between promoters and enhancers [10]. Also, interactions are enriched within specific, mega-based sized regions (named topologically associated domains or, shortly, TADs [11, 12]). At larger scales, higher-order spatial patterns like
metaTADs [13] or A/B compartments [1] are found and span tens of mega-bases or entire chromosomes. All these structural features are strongly linked to genome activity [14–16]. However, many aspects of DNA 3D organization remain unclear and the key mechanisms leading to the formation of loops, TADs and so on are currently debated.

To explain the chromosome folding patterns in a coherent framework, Polymer Physics models have been proposed and provided insightful information on chromatin architecture and its folding mechanisms [17, 18]. Furthermore, their combination with the above-mentioned experimental data allowed to increase the accuracy of the description of real genomes. On the other hand, heavy computational efforts are needed to achieve reliable 3D reconstructions [19] and complex computational procedures based, e.g., on Monte Carlo or Molecular Dynamics (MD) simulations, have been developed [20–22].

In the present work, we show how the computational method PRISMR (polymer-based recursive statistical inference method) [21], based on Polymer Physics laws, can be applied on GAM data to infer the 3D structure of chromatin. We summarize the backbone features of the PRISMR algorithm and shortly review its performances when applied on Hi-C data. Since notable computational power is required to the procedure, as massive parallel MD simulations are performed, we describe the usage of High Performance Computing (HPC), involved as in typical other applications [19]. Then, we show how the method can be generally extended to work with different experimental technologies and focus on its application on GAM data [23]. To test the performances of the approach, we show the results from the model of a 6 Mb locus around the Sox9 gene in mouse embryonic stem cells (mESC, GAM data from [2]). Overall, the 3D structures derived by PRISMR successfully reproduce the input GAM data and match independent Hi-C experimental data [12]. Then, we describe the application of PRISMR on the entire chromosome 7 in mESC [23] and show how the inferred structure can be employed to study the relationship of genomic regions of interest with the surrounding environment. Finally, we briefly discuss how the 3D structures derived in-silico by PRISMR can be used to benchmark experimental technologies as GAM and Hi-C, highlighting the potential role that such computational approaches can have in helping the design of experiments. More generally, in this contribution we highlight how the combination of Theoretical Physics, Molecular Biology experimental data and powerful HPC resources allow investigating 3D genome architecture with an increasing level of accuracy [18].

2 The PRISMR Method and Its Extension for GAM Data

2.1 Overview of the PRISMR Method

The PRISMR method is a computational tool designed [21] to derive the 3D structure of a genomic locus starting from experimental data containing its structural features (Fig. 1a). Specifically, PRISMR employs experimental data detecting the pattern of contacts among DNA sites of the considered locus. To deconvolute the architectural information encoded in the data and produce physically meaningful structures, PRISMR is informed with the principles of a Polymer-Physics model of chromatin. Here we use the Strings&Binders Switch (SBS) model [24], but any other could be, in principle, employed. The SBS model is based on the biological scenario where diffusive molecules,
as transcription factors (TFs), bind to the DNA string and drive its folding. It describes the DNA filament as a self-avoiding walk polymeric chain where diffusive particles called *binders* can attach. Binders can only interact with specific beads of the polymeric chain, known as *binding sites*, and they can attach to more than one binding site simultaneously. In this way, binders allow for the formation of loops between distal polymer sites, driving the spatial conformation of the chain. In general, different types of binding sites can be used with homotypic interaction, i.e. binders can only anchor to the cognate binding sites. The different types of binding sites can be schematically visualized as different colors and all the same colored binding sites along the chain define a *binding domain*. The number of colors of an SBS polymer, the arrangement of binding sites along the chain and the concentration of their respective binders determine the folding properties of the polymer model, that is its possible equilibrium 3D configurations [19, 25, 26]. The equilibrium configurations can be employed as proxy for the real conformations of a genomic locus in nuclei.

**Fig. 1.** a) Outline of the PRISMR method [21]. Informed with a Polymer-Physics model of chromatin, PRISMR finds the best polymer to describe a given genomic region. This is done by a Machine Learning approach, i.e. training the polymer model over specific experimental contact data of the region, as Hi-C [1] or GAM [2]. b) To use PRISMR on GAM data, an algorithm simulating the GAM experimental method on 3D polymer structures was prepared. Briefly, cell nuclei are modelled as spheres, each one containing a 3D structure. A random plane is generated to simulate a real GAM nuclear profile (NP) for every sphere. Beads inside the simulated NP are counted and hence the segregation frequencies extracted (see text). Finally, the co-segregation frequencies whereby two beads were found in the same simulated NP are arranged in the *in-silico* co-segregation matrix. Adapted from [23].

Given a DNA locus of interest, PRISMR aims to find the best polymer model describing it, that is the best number and arrangements of colors for the SBS polymer chain. The input of PRISMR is the experimental contact data associated to the locus. Typically, they
are arranged in a matrix containing the contact frequencies between any pairs of DNA sites. Using a Simulated Annealing Monte-Carlo procedure, PRISMR scans through the huge space of all possible SBS polymer models minimizing a cost function $H$ that measures the difference between the input data and the model contact matrix.

Once the best polymer model is found, its equilibrium 3D structures are obtained by massive parallel Molecular Dynamics (MD) simulations. Typically, the polymer is prepared in a SAW state inside a simulation box and binders are randomly placed in the environment. The simulation is then performed until thermodynamic equilibrium is reached. This is repeated for a large number (hundreds) of independent polymers, so to eventually obtain an ensemble of equilibrium 3D structures. The parameters used (such as the profile of the homotypic bead-binder interaction, the concentration of binders and binding affinity) are described in classical polymer physics studies [27] and are widely used in the field [28, 29].

In order to deal with the large number of particles of an SBS system (typically $10^3 \div 10^4$ for a few mega-base sized locus) and to produce a reliable statistical sampling of the configurations space, High Performance Computing (HPC) resources are needed. To give a sense of the resources involved, a single run with a generic software optimized for parallel computing (as LAMMPS [30]) requires in general a number of processors ranging from 8 to 64, with a time limit of at least 24–48 h, necessary to achieve thermodynamic equilibrium. Additionally, as said, about hundred independent runs are performed to get the final equilibrium ensemble. So, the production of an ensemble of 3D structures through MD simulations is the most computational demanding step and makes the use of HPC one of the key tools for the overall strategy.

The PRISMR procedure has been proven successful in reproducing Hi-C data [26, 31, 32], in predicting the impact of mutations and structural variants [21, 33] and in explaining cell-to-cell structural variability as recently detected in microscopy [34]. We will review in the next sections how PRISMR can work successfully also on GAM data [23]. We recall that other methods have also been developed to reconstruct 3D genome structure, e.g. using polymer models informed with epigenetic data [35, 36] or optimization procedures of restraints given directly by the contact data [37, 38].

### 2.2 The in-silico GAM Algorithm

As explained above, the cost function of PRISMR compares the input experimental contact matrix with the matrix of the polymer model. Thus, it requires an algorithm capable to extract a contact matrix from the 3D structures of a given polymer model and, specifically, the contact matrix must be of the same kind of the input. Hence, for the application on GAM data, an algorithm simulating the generation of a GAM matrix from polymer structures was prepared.

In GAM experiments [2] a random nuclear profile (NP) is cut from each cell of a population. The genomic content of each NP is then sequenced and all the DNA loci are counted. A locus caught in a NP is defined as segregated. Hence, the segregation table is extracted, i.e. a table whereby the loci found for each NP are reported. From this, the co-segregation frequencies for pairs, triplets etc. of loci in the same NP can be derived. The pairwise co-segregation frequencies are usually arranged in the co-segregation matrix. To account for biases as different sequencing mappability, co-segregation matrices can be
normalized, e.g. with the \textit{linkage disequilibrium normalization} D’ [23]. These matrices can be used as input for PRISMR to find the best polymer structures of a given DNA region.

In order to adapt the PRISMR procedure on GAM data, we realized an \textit{in-silico} version of the GAM pipeline [23]. Precisely, we model cell nuclei as individual spheres containing a polymer 3D structure (Fig. 1b). Then, a plane with random orientation is generated in each sphere and all the beads of the polymer distant from the plane less than a threshold are counted as segregating. This step simulates the NPs extraction and sequencing. Then, the \textit{in-silico} segregation table is generated, the co-segregation frequencies are computed and arranged in a co-segregation \textit{in-silico} matrix, and eventually normalized. The threshold distance from the random plane is set according to the thickness of experimental NPs [2] and the sphere radius is fixed to match typical cell radius estimates taken from mESC cells [2]. In mESC, cell nuclei can be well approximated as spheres, while, for other cell lines, different nuclear geometries can be implemented.

3 \textbf{The Sox9 Locus Explored with GAM Data}

To test PRISMR on GAM data, we focused on the \textit{Sox9} locus in mESC (data from [2]), i.e. a 6 Mb long region centered around the \textit{Sox9} gene (chr11:109–115 Mb, mm9 genome assembly). Specifically, we applied PRISMR on both the co-segregation and D’ normalized GAM data of the \textit{Sox9} locus at 40kb resolution (see Fig. 2a, b, top matrices). The comparison between the PRISMR GAM matrix with the input experimental data is very good in both the co-segregation and D’ applications (Fig. 2a, b), as witnessed by the high values of the Pearson (r) and the Stratum Adjusted (scc) [39] correlation coefficients (\(r = 0.93\) and \(scc = 0.96\) for the co-segregation matrices, while \(r = 0.86\) and \(scc = 0.97\) for the D’ case). These values, together with the visual inspection of the matrices, show that PRISMR results effective on GAM co-segregation as well D’ normalized data. To test the robustness of the procedure, we compared the polymer models obtained from the two cases (co-segregation and D’). Importantly, the two best SBS polymers have the same number of colors and appear to have similar arrangements of the binding sites (Fig. 2a, b middle panels). Indeed, we computed the genomic overlap \(q\) between the binding domains in the two models, defined as the normalized integral along the locus of the product of the number of their binding sites. We find that the most overlapping domains (equally colored in Fig. 2a,b) have an average overlap \(q = 0.70\). As control, we did the same for bootstrapped binding domains, getting \(q = 0.47\), which is significantly lower (p-value = 3e−6, Mann-Whitney U test). This implies that the PRISMR procedure is robust, as it derives compatible polymer models when fed with raw or normalized GAM data. Also, this indicates that the SBS best polymer can work as a solid model for the 3D structures of a given genomic region.

We then generated a population with up to 5 * 10^2 equilibrium configurations for the polymer model trained on the co-segregation GAM data, using massive parallel MD simulations. Minute computational details of the simulations can be found in [23]. To test the accuracy of our GAM-derived polymer structures, we computed the contact matrix [26] from them, that is the matrix containing the frequencies whereby pairs of beads are in contact. The contact between two beads is called when their distance is less than a
As first application on GAM data [23], PRISMR was run over the data [2] of the Sox9 locus (chr11:109–115 Mb) of mESC at 40 kb. a) PRISMR is fed with the experimental GAM co-segregation matrix of the locus. In each of these cases, we obtained the best SBS polymer model (Fig. a, b, middle diagrams) and the corresponding in-silico GAM matrix (Fig. a, b, bottom matrices. On the top the experimental matrix, on the bottom the PRISMR derived matrix. Color bars indicate the percentiles of the matrices. In the middle, the binding domains found by PRISMR along the best SBS polymer of the locus are shown. The comparison between the two matrices is good (r = 0.93 and scc = 0.96), as proof of PRISMR effectiveness. b) Same as in panel a), but the input experimental matrix is the D’ normalized [2] GAM matrix of the Sox9 locus. Again, the comparison between experiment and model is nice (r = 0.86 and scc = 0.97). The binding domains found by PRISMR in this case (middle) and those found in the co-segregation case (panel a) have an average overlap q = 0.70 (see text). This is significantly higher than a random control case (p-value 3e—6, Mann-Whitney U test), meaning PRISMR inference is very robust to the input data. c) To further test the solidity of PRISMR, we considered the 3D structures derived from the co-segregation GAM matrix and extracted an in-silico Hi-C matrix from them (top), using an algorithm developed in [26]. We compare this matrix with independent Hi-C data of the same locus [12] (bottom). Strikingly, we get good similarity (r = 0.90 and scc = 0.51). Adapted from [23].
threshold [26]. Such in-silico contact matrix is compared with completely independent Hi-C data [12] for the same Sox9 genomic locus in mESC, at 40 kb resolution. The comparison reveals good agreement (Fig. 2c), with \( r = 0.90 \) and \( \text{scc} = 0.51 \). This result further supports the robustness of our computational approach when used on GAM data and, importantly, supports the accuracy in the description of the Sox9 locus from the produced 3D structures.

4 The Chromosome 7 Explored with GAM Data

Here, we show the application of PRISMR trained on GAM data for the entire chromosome 7, in mESC (data from [2]). We used D’ GAM data at 250 kb resolution. In line with the previously discussed results, the experimental D’ (Fig. 3a, top) and the PRISMR in-silico (Fig. 3a, bottom) matrices are similar, with \( r = 0.67 \) and \( \text{scc} = 0.96 \), showing that PRISMR achieves good results even for a system with higher complexity than a locus. As before, from the best SBS polymer (Fig. 3a, middle panel), a population of \( 5 \times 10^2 \) equilibrium 3D configurations was derived by use of MD simulations [23]. In Fig. 3b an example of 3D configuration of chromosome 7 is shown, colored according to the color scheme reported in Fig. 3a. From the population of 3D structures, relevant information about specific loci can be extracted, such as the relative location of biologically interesting regions with respect to the architecture of the overall chromosome. In this way, it is possible to reveal systematic tendencies to localize in peripheric or internal positions and, therefore, the possibility to form contacts with other chromosomes. Knowing the preferred contact pattern can be fundamental to understand the impact of mutations occurring on the studied region. We performed this analysis for the mouse orthologue of the human 16p.11.2 locus (chr7: 133.84:134.24 Mb, mm9 genome assembly), which is an interesting region for biomedicine since its mutations (as duplications or deletions) have been associated to the autism disorder [40, 41]. In Fig. 3b the 16p.11.2 locus is marked in red and appears to localize in a peripherical position. To quantitatively understand the preferential localization of the 16p.11.2 locus across all the polymer structures, we computed the distribution of the distance (\( r \)) between the locus center of mass and the center of mass of the whole chromosome. Specifically, working with a 250 kb resolution, we took the center of mass of a 2 Mb segment around the 16p.11.2 locus (chr7:133–135 Mb, mm9 genome assembly). Then we normalized dividing by the gyration radius of the chromosome (location ratio \( r/R_g \), Fig. 3c). A location ratio larger than 1 indicates a peripheral position, while a value lower than 1 would imply a more internal position inside the chromosome structure. In Fig. 3c we show the distribution of location ratios obtained from the population of structures. We compare against a control case computed for other 300 random loci, all 2Mb sized. The two distributions are not compatible (p-value = 0.01, Mann-Whitney \( U \) test), as the histogram of 16p.11.2 is shifted toward higher values. This suggests that the 16p.11.2 tends to be more peripheral than the control. Furthermore, in 10% cases the location ratio of the 16p.11.2 is greater than 1.5, i.e. located in the very periphery of the chromosome. This suggests such locus can give raise to important and functional contacts with other chromosomes nearby. Deeper studies as this, with PRISMR used on increased resolved data, could lead to the rigorous tracking of the contact network for the 16p.11.2. Finally,
Fig. 3. The PRISMR method was used on the D’ normalized GAM data [2] of the chromosome 7 in mESC [23] at 250 kb resolution. a) The input experimental matrix of the chromosome is compared against the PRISMR one. Visually and quantitatively they are similar to each other (\( r = 0.67 \) and \( \text{scc} = 0.96 \)). In the middle, the binding domains found along the polymer model of the chromosome are reported. b) An example of 3D structure inferred by PRISMR is shown. The polymer is colored according to the bar in panel a), under the top matrix. In red, a 2 Mb region (chr7:113–135Mb) containing the mouse orthologue of the human 16p.11.2 locus (chr7:133.85–134.24 Mb). c) An example of usage of the PRISMR derived 3D structures. The distribution (blue) of the location ratio \( r/R_g \) (see text) of the 16p.11.2 mouse orthologue locus across all PRISMR structures is plotted against a control distribution (gray), showing the location ratio for 300 randomly chosen loci of the same size. The distributions are not compatible (p-value = 0.01, Mann-Whitney U test), so that the 16p.11.2 location ratio results shifted toward higher values, indicating more likelihood to be peripheral in the chromosome architecture than the control. This suggests the 16p.11.2 mouse orthologue can establish significant contacts with adjacent chromosomes. Adapted from [23]. (Color figure online)

we note that for both the control case and the specific 16p.11.2 locus, the ratio between standard deviation and the average value is rather high (approximately 36%), hinting that there is a significant structural variability in the conformations. This is in agreement with recent experimental findings highlighting the high degree of structural variability of specific loci among different cells [4, 34, 42].
5 PRISMR-Derived Structures to Compare Different Technologies

In this last section, we compare the results of the PRISMR approach on data from different experimental technologies. To this aim, we consider the ensemble of 3D structures of the Sox9 locus inferred by PRISMR from GAM data [23] and the ensemble of 3D structures of the same locus derived from Hi-C data in another study [26]. The two in-silico populations of structures can be compared rigorously and could be used to understand the differences between the Hi-C and GAM technologies in detecting information about DNA 3D organization [43].

![Diagram](image)

**Fig. 4.** The structures derived by PRISMR can be interestingly employed to compare experimental technologies, as Hi-C [12] and GAM [2]. a) Examples of 3D structures for both the Hi-C and GAM derived cases are shown. Structures are colored according to bar in the bottom, which follows the TADs detected in [12]. These examples of configurations are visually comparable to each other. b) The distribution of the gyration radius across the 3D structures is computed. The GAM distribution (red) and the Hi-C distribution (blue) have the same shape, yet the former is shifted to the left, suggesting higher average compaction. Interestingly, such difference could be explored to understand if it reveals differences in how Hi-C and GAM «see» DNA 3D organization [43]. Adapted from [23]. (Color figure online)

For instance, given the 3D structures from Hi-C and GAM, we can compare their geometry by computing the average distance matrix, containing the average Euclidean distances between each possible pair of polymer beads. Comparison between the Hi-C and GAM derived distance matrices reveals good agreement, with \( r \sim 0.95 \) and \( sc \sim 0.60 \). Conversely, intrinsic differences between the technologies can lower the agreement between the inferred structures at deeper detail than the average distances and could be further explored [43]. In Fig. 4a we show examples of structures taken from the two ensembles and they look, overall, qualitatively similar. The distributions of the gyration
radius are quite similar to each other too (Fig. 4b), although GAM structures exhibit slightly lower values.

6 Conclusions

We have shown that PRISMR, a Machine Learning computational method that combines Polymer Physics and data from different experimental methods, can be effectively employed to infer the possible 3D structures of real genomic regions. Among the possible datasets today available, this can be achieved using Genome Architecture Mapping (GAM) data, a recent technique based on sectioning cell nuclear profiles and sequencing their genomic content [2]. We have described the application of PRISMR on GAM data for a 6 Mb locus in mouse embryonic stem cells [2] and showed that PRISMR successfully reproduced the input data. Furthermore, the GAM-derived 3D structures are then used to generate an in-silico contact matrix which results compatible with independent Hi-C data [12], supporting the generality and robustness of the computational approach. Next, we showed that the procedure can be extended to study systems with higher complexity as entire chromosomes. So, we generated a population of structures describing chromosome 7 using again GAM data in mESC [2]. Taking advantage of that, we studied the radial position of the mouse orthologue of the 16p.11.2 locus and found a general trend to localize in the peripheral region of the chromosome, with possible implications on its regulation. Finally, we tested the robustness of the 3D reconstruction by comparing the structures inferred from GAM [2] and Hi-C data [12] of the Sox9 locus and found that both models yield similar results. Differences in the polymer population likely reflect intrinsic differences between the two experimental technologies [43].

The perspective is that sophisticated computational tools combining Machine Learning, Polymer Physics and massively parallel Molecular Dynamics simulations, such as the PRISMR method, will become more and more effective to interpret the constantly increasing amount of data generated by experimental technology, shedding light on the mechanism regulating genome folding.

References

1. Lieberman-Aiden, E., et al.: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 80(326), 289–293 (2009)
2. Beagrie, R.A., et al.: Complex multi-enhancer contacts captured by genome architecture mapping. Nature 543, 519–524 (2017)
3. Quinodoz, S.A., et al.: Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. Cell 174, 744-757.e24 (2018)
4. Bintu, B., et al.: Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science 80(362), eaau1783 (2018)
5. Cardozo Gizzi, A.M., et al.: Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. Mol. Cell. 74, 212-222.e5 (2019)
6. Misteli, T.: Beyond the sequence: cellular organization of genome function. Cell 128(4), 787–800 (2007)
7. Bickmore, W.A.: The spatial organization of the human genome. Annu. Rev. Genomics Hum. Genet. 14, 67–84 (2013)
8. Finn, E.H., Misteli, T.: Molecular basis and biological function of variability in spatial genome organization. Science 365(6457), eaaw9498 (2019)
9. Cremer, T., Cremer, M.: Chromosome territories (2010)
10. 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)
11. Nora, E.P., et al.: Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385 (2012)
12. Dixon, J.R., et al.: Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380 (2012)
13. Fraser, J., et al.: Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol. Syst. Biol. 11, 852 (2015)
14. Dekker, J., Misteli, T.: Long-range chromatin interactions. Cold Spring Harb. Perspect. Biol. 7, a019356 (2015)
15. Dekker, J., Mirny, L.: The 3D genome as moderator of chromosomal communication. Cell 164, 1110–1121 (2016)
16. Spielmann, M., Lupiáñez, D.G., Mundlos, S.: Structural variation in the 3D genome. Nat. Rev. Genet. 19, 453–467 (2018)
17. Serra, F., Bäi, D., Goodstadt, M., Castillo, D., Filion, G., Marti-Renom, M.A.: 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)
18. Fiorillo, L., et al.: A modern challenge of polymer physics: novel ways to study, interpret, and reconstruct chromatin structure. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 10(4), e1454 (2019)
19. Conte, M., et al.: Hybrid machine learning and polymer physics approach to investigate chromatin 3D structure. Eur. Conf. Parallel Process. 572–582 (2019, in press)
20. Giorgetti, L., et al.: Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell 157, 950–963 (2014)
21. Bianco, S., et al.: Polymer physics predicts the effects of structural variants on chromatin architecture. Nat. Genet. 50, 662–667 (2018)
22. 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)
23. Fiorillo, L., et al.: Inference of chromosome 3D structures from GAM data by a physics computational approach. Methods S1046–2023(18), 30485–30487 (2019)
24. Nicodemi, M., Prisco, A.: Thermodynamic pathways to genome spatial organization in the cell nucleus. Biophys. J. 96, 2168–2177 (2009)
25. Nicodemi, M., Pombo, A.: Models of chromosome structure. Curr. Opin. Cell Biol. 28, 90–95 (2014)
26. Chiariello, A.M., Annunziatella, C., Bianco, S., Esposito, A., Nicodemi, M.: Polymer physics of chromosome large-scale 3D organisation. Sci. Rep. 6, 29775 (2016)
27. Kremer, K., Grest, G.S.: Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. J. Chem. Phys. 92, 5057–5086 (1990)
28. Rosa, A., Becker, N.B., Everaers, R.: Looping probabilities in model interphase chromosomes. Biophys. J. 98, 2410–2419 (2010)
29. 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. U. S. A. 110, E3605–E3611 (2013)
30. Plimpton, S.: Fast parallel algorithms for short-range molecular dynamics. J. Comput. Phys. 117, 1–9 (1995)
31. Barbieri, M., et al.: Complexity of chromatin folding is captured by the strings and binders switch model. Proc. Natl. Acad. Sci. U. S. A. 109, 16173–16178 (2012)
32. Bianco, S., et al.: Modeling single-molecule conformations of the HoxD region in mouse embryonic stem and cortical neuronal cells. Cell Rep. 28, 1574-1583.e4 (2019)
33. Kraggesteen, B.K., et al.: Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. 50, 1463–1473 (2018)
34. Conte, M., Fiorillo, L., Bianco, S., Chiariello, A.M., Esposito, A., Nicodemi, M.: Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. Nat. Commun. 11, 3289 (2020)
35. 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)
36. 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, 786-797.e11 (2018)
37. Hua, N., Tjong, H., Shin, H., Gong, K., Zhou, X.J., Alber, F.: Producing genome structure populations with the dynamic and automated PGS software. Nat. Protoc. 13, 915–926 (2018). https://doi.org/10.1038/nprot.2018.008
38. Stefano, M.D., Paulsen, J., Lien, T.G., Hovig, E., Micheletti, C.: Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. Sci. Rep. 6, 35985 (2016)
39. Yang, T., et al.: HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res. 27, 1939–1949 (2017)
40. Stein, J.L.: Copy number variation and brain structure: lessons learned from chromosome 16p11.2. Genome Med. 7, 13 (2015)
41. Loviglio, M.N., et al.: Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. Mol. Psychiatry. 22, 836–849 (2017)
42. Nagano, T., et al.: Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502, 59–64 (2013)
43. Fiorillo, L., et al.: Comparison of the Hi-C, GAM and SPRITE methods by use of polymer models of chromatin. bioRxiv. 2020.04.24.059915 (2020).