FISH-ing for captured contacts: towards reconciling FISH and 3C

Geoffrey Fudenberg1,2 & Maxim Imakaev1

Chromosome conformation capture (3C) and fluorescence in situ hybridization (FISH) are two widely used technologies that provide distinct readouts of 3D chromosome organization. While both technologies can assay locus-specific organization, how to integrate views from 3C, or genome-wide Hi-C, and FISH is far from solved. Contact frequency, measured by Hi-C, and spatial distance, measured by FISH, are often assumed to quantify the same phenomena and used interchangeably. Here, however, we demonstrate that contact frequency is distinct from average spatial distance, both in polymer simulations and in experimental data. Performing a systematic analysis of the technologies, we show that this distinction can create a seemingly paradoxical relationship between 3C and FISH, both in minimal polymer models with dynamic looping interactions and in loop-extrusion simulations. Together, our results indicate that cross-validation of Hi-C and FISH should be carefully designed, and that jointly considering contact frequency and spatial distance is crucial for fully understanding chromosome organization.

While genomes are often considered as one-dimensional sequences, they are also physically organized in three dimensions inside the cell nucleus, with far-reaching consequences. One of the many important implications relates to gene regulation: regions of regulatory DNA are often very far away in the linear genomic sequence from the genes they regulate, yet their regulatory interactions are presumed to rely on direct encounters in three-dimensional space. Current methods provide indirect, yet complementary, readouts of the high-resolution chromosomal dynamics underlying these processes. Below we focus on two widely used techniques for assaying chromosomes, DNA-FISH and chromosome conformation capture (3C).

3C techniques have become popular for their high-throughput ability to connect spatial information to the genomic sequence. 3C-based approaches use crosslinking and ligation to capture the information that two genomic loci are spatially proximal. Moreover, 3C techniques are readily generalized to the genome-wide scale, and are then usually termed Hi-C; here we refer to the contact frequency obtained from Hi-C and 3C interchangeably. Importantly, while 3C records whether two loci were in contact in some fraction of cells in the population, it does not record where in the nucleus this contact occurred. Also, 3C is usually performed on large populations of cells. This is advantageous in that 3C can assay both very frequent and very rare events, as the large population allows a large dynamic range. However, information regarding cell-to-cell variability is not available in population-average maps of chromosomal contact frequencies.

FISH technologies are appreciated for their ability to specifically determine the spatial position of sets of chromosomal loci by imaging. FISH is based on optically labeled probes that hybridize to complementary regions of chromosomes. Importantly, as an imaging-based approach, FISH is intrinsically able to probe cell-to-cell variability and directly record spatial position inside the nucleus. Many different labeling approaches have been considered, including labeling pairs of loci (two-locus FISH) as well as labeling larger contiguous regions and even whole-chromosome painting. High-throughput and super-resolution FISH approaches are currently in development. Still, obtaining high-resolution pairwise distance distributions for all pairs of loci—that is, constructing a pairwise distance map similar to a genome-wide Hi-C contact map—currently remains out of reach. For further experimental background on the connection between FISH and Hi-C, see ref. 20.

In studies that rely primarily on 3C-based approaches, FISH is often performed on a subset of loci as a means of validation. Typically, for loci at increasing genomic separations, their average FISH spatial distance increases and 3C contact frequency decreases. Additionally, for a limited set of tested pairs of loci, it was found that loci in the same A/B compartment contact each other more frequently and are on average closer; similar findings are obtained for loci in the same topologically associating domain (TAD). Moreover, spatial distance and contact frequency are largely correlated at the ~300 kb–10 Mb scale. However, this does not seem to be strictly the case for all loci, and such assessments can even lead to seemingly paradoxical observations when comparing FISH and 3C.

Because of the stochastic and variable nature of chromosome folding in vivo, polymer models provide a useful framework...
for interpreting 3C or FISH data. Modeling efforts naturally started with homopolymer models, which assume the chemical equivalence of all monomers (i.e., that there is no sequence specificity for chromosomal interactions or folding). While chromosomes in vivo are generally better described by models that include locus-specific folding, homopolymer models are highly studied in the physics literature and are often more amenable to analytical understanding. In the majority of homopolymer models, the further apart two monomers are along the polymer chain, the further apart they are in space, and the less frequently they are in contact; this leads to an often-useful, but potentially misleading, heuristic that the two quantities are directly related (Supplementary Note). As measuring contact frequency was largely unachievable in polymer systems prior to the development of 3C for chromosomes, the concordance or discordance of contact frequency and spatial distance has received relatively little attention in the polymer physics literature.

Here we demonstrate how 3C and FISH generally probe different aspects of spatial chromosome organization. We first illustrate how spatial distance and contact frequency can display a seemingly paradoxical relationship in currently available experimental data, focusing on the simplest two-locus labeling approach for FISH, as it is most directly comparable to 3C. We then study the connection between contact frequency and average spatial distance in a simple polymer model; these simulations show that a minimal assumption, introduction of a single dynamic loop between two loci, can break the typical relationship between contact frequency and average spatial distance. We then consider polymer simulations with loop extrusion, and find that this process can also affect contact frequency and spatial distances to greatly different degrees. Together our results show how seemingly paradoxical relationships between contact frequency and spatial distance can easily emerge in experimental data for physical reasons, demonstrate that cross-validation of Hi-C and FISH must be very carefully considered, and argue that joint consideration of contact frequency and spatial distance will underlie further understanding of chromosome organization in vivo.

**RESULTS**

3C and FISH probe different aspects of spatial organization

To investigate the connection between 3C and FISH, we focused on the simplest case of both methods, in which each method probes the relationship between a pair of loci (Fig. 1a,b). A common design for FISH experiments involves labeling a pair of genomic loci to directly visualize their distances in a population of cells (Fig. 1a). This experimental design allows the measurement of the probability density function (PDF) of spatial distances between a pair of loci (Fig. 1c). Results from such experiments are often shown as cumulative distribution functions (CDFs, Fig. 1d) as these do not require binning or density-estimation steps to obtain relatively smooth curves for limited numbers of cells. In contrast with FISH, 3C experiments capture rare contacts that occur when the loci are closer than the capture radius imposed by crosslinking and ligation (Fig. 1b). Roughly, 3C measures the integral of the spatial distance PDF up to the capture radius (Fig. 1c), or, equivalently, the value of the CDF at the capture radius (Fig. 1d). Since such small distances are relatively rare, imaging many cells is certainly a requirement for directly comparing 3C contact probabilities with distances measured by FISH.

We further examined the connection between 3C and FISH by considering recent publicly available Hi-C data. The authors of this publication performed high-resolution Hi-C experiments and reported several thousands of CTCF-mediated peaks, termed loops, in the Hi-C data. As a validation of the loops by FISH, they report CDF FISH plots for four pairs of ‘loop’ and ‘control’ loci at matched genomic separations (for example, peak1–loop and peak1–control, re-plotted in Supplementary Fig. 1) for the same cell type. As part of the validation, the authors reported that for each of the loop–control pairs of loci, the median spatial distance changed concordantly with the Hi-C signal. While this holds, we also found that this was not always the case when we compared loops and controls from different pairs. Indeed, we found a seemingly paradoxical relationship between peak4–loop and peak3–control (Fig. 2): peak4–loop has higher contact frequency despite being further away on average than peak3–control. Nevertheless, the change in the value of the CDF at small distances actually was in agreement with measurements from Hi-C, suggesting that this short-range behavior of the CDF is more closely connected with contact frequency. A similar situation is observed for peak4–loop and peak2–control. In contrast, for all control–control pairs of loci, the median spatial distance changed concordantly with the Hi-C signal. We note that seemingly paradoxical pairs involved comparisons between a loop and a control.

Together, these observations suggest that locus-specific chromosome organization in vivo can be an important reason why average spatial distance and contact frequency could behave divergently. They additionally argue that to reconcile this divergence and cross-validate observations from 3C, it will be necessary to obtain the full spatial distance distribution from FISH, including very short distances. Given the currently limited availability of high-resolution matched experimental Hi-C and FISH data, we turned to polymer models to study the relationship between
Supplementary Fig. 4). However, the simulated ensemble of conformations.

Figure 2 | Experimental data demonstrate the complex relationship between Hi-C and FISH. (a) Comparison between a pair of typical loci shows that increased spatial distance as determined by FISH coincides with decreased Hi-C counts. (b) Comparison between a pair of loop loci and a pair of control loci shows a seemingly paradoxical increase in spatial distance as determined by FISH accompanied by increased Hi-C counts. FISH and Hi-C data re-plotted from ref. 23 for GM12878 cells (Supplementary Table 1). Horizontal gray line intersects the median spatial distance, and vertical gray line intersects the probability of an observation less than 300 nm, \( P(<300 \text{ nm}) \). Bar plots show median spatial distance, \( P(>300 \text{ nm}) \), and \( \log_{10}(\text{corrected Hi-C counts}) \).

spatial distance and contact frequency, where these two quantities can be unambiguously calculated for any desired pairs of loci from the same set of conformations.

**Simulations can reconcile contact frequency and spatial distances**

To understand the minimal set of assumptions that can decouple contact frequency from average spatial distance, we investigated both of these quantities in equilibrium polymer simulations of a single dynamic chromatin loop (Fig. 3). Following past work in which we investigated the effect of a fixed chromatin loop, we modeled chromatin as a semi-flexible polymer fiber with excluded volume interactions (Online Methods, Supplementary Note). We performed simulations using OpenMM\(^{20,31} \), and calculated simulated contact maps (Fig. 3a), spatial distance distributions (Fig. 3b,c), and average spatial distance maps (Supplementary Fig. 2b) from the simulated ensemble of conformations.

We imposed the dynamic looping interaction using a short-ranged attractive force. Monomers at the base of the 25-kb dynamic loop interacted with attractive energy (4\( kT \)) unless noted when they were closer than a distance of 2 monomer diameters; for other monomers the attractive part of the potential was set to be negligibly small (0.1\( kT \)). This pairwise interaction potential could arise from direct molecular interactions, and the two monomers involved in the dynamic looping interaction can be thought of as hard spheres that stick to some degree upon coming into contact, following a stochastic encounter in 3D (reviewed in refs. 8,32). In our simulations, the dynamic looping interaction is clearly visible in the contact frequency map, but is faint in a map of average spatial distances (Supplementary Fig. 2a,b). Interestingly, the PDF of spatial distances for monomers at the base of the dynamic loop and control monomers (Fig. 3b) appeared quite similar, apart from a sharp peak at short distances for the monomers at the loop base.

For typically considered polymer systems of indistinguishable monomers, mean spatial distance and contact probability are generally inversely related (Supplementary Note). However, our simulations demonstrate that even a minimal modification, introduction of a single dynamic loop, changes this typical behavior (Fig. 4). While a comparison between control loci of separation 15 kb or 25 kb displays the typical monotonic behavior over all genomic separations (Fig. 4b), an apparent paradox emerges when comparing the control loci separated by 15 kb with the 25-kb dynamic loop (Fig. 4c). While the 25-kb dynamic loop is further apart on average, it displays a higher contact frequency. This behavior can emerge because contacts are rare events, and therefore contact frequency can increase many-fold without large changes in the average distance (Supplementary Fig. 2d,f). Similar behavior emerges in simulations for a range of parameter values of chromatin stiffness, chromatin density, dynamic loop attraction strength, and loop size (Supplementary Fig. 3). Consideration of dynamic loop models as equilibrium ensembles of conformations provides further support for the widespread possibility of seemingly paradoxical pairs (Supplementary Note). Together, our simulations show how, even in a particularly simple case, seemingly paradoxical relationships can emerge between spatial distance and contact frequency, arguing for caution when designing comparisons between FISH and 3C.

**Simulations illustrate how experimental limitations could affect validation of a dynamic loop**

We next investigated how possible experimental limitations to either FISH or 3C can impact our ability to ascertain the presence of this simulated looping interaction.

In FISH experiments, assaying a finite number of cells both imposes uncertainty on the PDF and makes the probability of rare events difficult to estimate. Consistently, it is more difficult to reliably detect changes in contact frequency than median spatial distance in simulations when assaying a limited number of conformations (Supplementary Fig. 4). As contact frequencies are often quite low, this indicates that consistent validations of Hi-C by FISH would require assaying orders-of-magnitude-larger numbers of cells than is typical in FISH experiments.

For FISH, additional uncertainty can be imposed by factors including probe size, chromatin movement during denaturation and hybridization, background noise, and ambiguities arising...
from the presence of homologous chromosomes (reviewed in ref. 20). In simulations, we considered how the first two factors might affect spatial distance distributions of loop and control loci (Supplementary Fig. 5). To simulate the impact of probe size, we considered the pairwise distributions between centroids of chosen regions, rather than an exact pair of monomers. To simulate uncertainty or perturbation of relative distances due to chromatin movement during the FISH protocol, we simulated probe localization uncertainty by adding Gaussian noise to each set of simulated probe distances. We find that even a small uncertainty or imprecision in the spatial localization of probes during FISH makes the existence of a dynamic looping interaction much more difficult to ascertain, whereas larger probe size had a relatively smaller impact on the spatial distributions (Supplementary Fig. 5).

In a 3C experiment, whether two loci in close spatial proximity in a given cell are recorded as a contact depends on the effective capture radius. The capture radius can be influenced by a number of factors, including restriction efficiency, restriction frequency, and the details of crosslinking, which may depend on the particular complement of DNA-associated proteins at a given genomic locus. Our simulations show that a larger contact radius for simulated 3C can also obscure the existence of a dynamic looping interaction (Supplementary Fig. 5).

Additional measurement noise may also come from library complexity, sequencing depth, and ligations in solution, which could all obscure the detection of looping interactions in 3C-based methods (reviewed in ref. 36).

Together, these simulated perturbations to the idealized FISH and 3C protocols illustrate how considering many experimental details will be required to fully reconcile observations from FISH and Hi-C.

Loop-extrusion simulations can display divergent contact frequency and spatial distance

After considering spatial distance and contact frequency in this minimal model, we then considered their relationship in simulations of loop extrusion with locus-specific boundary elements, recently proposed by us and others as explaining key aspects of interphase chromosome organization. Loop extrusion is a potential mechanism for the formation of TADs and loops in mammalian Hi-C maps, making it a subject of recent interest (for review, see ref. 1). To consider genomic scales similar to those of the published loop-control pairs considered above, we simulated a genomic region containing several TADs of 210–870 kb, with parameters of the chromatin fiber as defined in our previous study of interphase loop extrusion. For each loop at the corner of every TAD, we considered a matched control at the same genomic separation, but offset by 100 kb.
The dynamics of loop extrusion are governed by key parameters: processivity, separation, and extrusion speed. In simulations, we found that certain combinations of these parameters lead to loop–control pairs with seemingly paradoxical relationships (Fig. 5). Such relationships could emerge because, much as in the minimal dynamic loop model considered above, certain regimes of loop extrusion can greatly increase contact frequency between subsequent boundary elements while minimally altering average spatial distance (Supplementary Fig. 6). In our simulations, the number of seemingly paradoxical pairs increased with larger separation, and also for slower loop extrusion (Supplementary Fig. 6). Interestingly, the best-fitting parameters from the sweep in our previous study did not produce seemingly paradoxical pairs for the considered TAD sizes. However, simply making loop extrusion slower was sufficient to create seemingly paradoxical pairs, while having little effect on the simulated contact map (Supplementary Fig. 7). This indicates that perturbing loop-extrusion dynamics could alter average FISH distances while having little effect on Hi-C contact maps, and thus using Hi-C alone to infer mechanisms of chromosomal folding may be insufficient.

**DISCUSSION**

Our results illustrate that while median spatial distance and contact frequency are often inversely proportional, they are far from equivalent. Indeed, our simulations show that a relatively minor perturbation—the existence of a dynamic looping interaction between two loci—clearly breaks the equivalence between these two quantities. In particular, we show that there is great freedom to make large shifts in contact frequency with small shifts to median spatial distance, since contacts between distal chromosomal loci are generally rare events. We then show that loop extrusion can similarly break the typical correspondence between median spatial distance and contact frequency. Together, our simulations demonstrate that our expectation in vivo should be a nontrivial relationship between contact frequency and spatial distance, and that Hi-C and FISH data together will be necessary to better understand chromosome organization.

Given these factors, 3C experiments cannot be simply validated (or invalidated) by FISH without carefully considering technical details of the two methods. Indeed, efforts to integrate results from these technologies will need to carefully address unknowns of the 3C capture radius and FISH localization uncertainty, in addition to assaying sufficiently large numbers of cells to populate the small-distance portion of the FISH distribution. While here we limit ourselves to considering the relatively simple comparison between 3C and two-locus FISH, many other comparisons would be valuable in future work, including how to best integrate information obtained from contiguously stained regions with Hi-C experiments.

In addition to the implications for validation, our results also caution against certain modeling approaches. In particular, our results show how a common strategy of simply transforming ensemble average 3C contact frequencies into spatial distances ultimately leads to inconsistent models of chromosomal organization (reviewed in ref. 8).

Finally, our results also demonstrate how polymer modeling can in principle be used to reconcile FISH spatial distances and 3C or Hi-C contact frequencies. This is because both quantities are readily calculable from an ensemble of simulated polymer conformations. One of the central goals of the recently formed 4DN consortium is to systematically compare Hi-C and high-throughput high-resolution imaging data and understand any potential discrepancies. As these matched data become available, systematically comparing polymer models to both Hi-C and imaging data will be an essential step toward understanding principles of chromosomal organization.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

The authors thank A. Goloborodko for thoughtful comments regarding the statistical mechanics analogy of re-weighting loop conformations, G. Nir for helpful discussions regarding imaging, and anonymous reviewers for thoughtful and detailed feedback. The authors also thank J. Dekker and other members of the UMass–MIT Center for 3D Structure and Physics of the Genome for helpful
feedback. Finally, the authors thank L. Mirny for comments on earlier drafts of this paper and for supporting their independent work. This work was supported by NSF 1504942 Physics of Chromosomes (PI: L. Mirny) and U54 DK107980 3D Structure and Physics of the Genome (PIs: J. Dekker and L. Mirny). During revisions, G.F. was supported by the San Simeon Fund (PI: K. Pollard).

AUTHOR CONTRIBUTIONS G.F. and M.I. jointly conceived of the study, performed analyses, and wrote the manuscript, and are listed in alphabetical order on the first page.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.
**ONLINE METHODS**

**Polymer simulation overview.** Polymer models were simulated with OpenMM\(^{30,31}\), a high-performance GPU-assisted molecular dynamics software (https://simtk.org/home/openmm). We used an in-house openmm-polymer library to implement simulations (publicly available http://bitbucket.org/mirnylab/openmm-polymer).

**Dynamic loop simulations.** Dynamic loop simulations were performed as in the work of Doyle et al., 2014\(^{29}\), albeit with a dynamically interacting loop rather than a static loop, implemented as described below. Following Doyle et al., we modeled chromatin as a semi-flexible polymer fiber with excluded volume interactions, where spherical monomers of 15-nm diameter represent ~500 bp, or approximately three nucleosomes. Adjacent monomers were connected by harmonic bonds with a potential \(U = 25^\star(r - 1)^2\) (here and below, energy is in units of \(kT\)). The stiffness of the fiber was modeled by a three point interaction term, with the potential \(U = k^\star(1 - \cos(\alpha))\), where \(\alpha\) is an angle between neighboring bonds and \(k\) is a parameter controlling stiffness, here set to 1 \(kT\).

To model the dynamic loop considered in the present work, we used a Lennard–Jones (LJ) potential \(U = 4\varepsilon_{ij}^\star(1/r^{12} - 1/r^{6})\) where \(\varepsilon_{ij}\) was set to 4 \(kT\) for the monomers at the base of the dynamic loop (\(ij\)), and was set to negligibly small otherwise (\(\varepsilon = 0.1\ kT\)). We initialized our simulations as a system of 8 compact rings (see ref. 47), and used periodic boundary conditions to achieve a density of 0.10 (in the middle of the estimated range for mammalian cells\(^{49}\)).

We then simulated 50 runs of this system using Langevin dynamics, for 10e8 time steps. For the fiber lengths considered here, polymer simulations reached equilibrium in less than 1e7 time steps; this was confirmed by observing that monomer displacement saturates after about 5e6 time steps. Conformations were saved every 1e5 time steps and an equilibrium ensemble of 900 conformations obtained after the initial equilibration was used for our analysis.

To obtain simulated contact maps, we first found all contacts within each polymer conformation, and then aggregated these contacts for all pairs of monomers. Following ref. 29, a contact was defined as two monomers being at a distance less than 3 monomer diameters. To obtain simulated FISH distributions, we calculated a list of spatial distances for a chosen set of loci, and built a histogram of distances starting at 0 in bins of 0.1 monomers. To display PDFs this histogram was then smoothed with a moving average window with a size of 0.7 monomers.

**Loop-extrusion simulations.** Loop-extrusion simulations were performed as in ref. 38 with slight modification to the sizes and number of TADs, and an upgraded loop-extruding factor (LEF) simulation engine.

To better span the genomic size range of TADs and loops probed by FISH in ref. 23, we considered simulations of a 6-Mb region, with 10,000 monomers, 12 TADs, and monomers representing 600 bp as in ref. 38. These TADs were separated by 11 boundary elements that stalled loop extrusion, positioned at monomers: 750, 1,550, 1,900, 3,000, 3,650, 4,300, 5,750, 6,550, 6,900, 7,750, 9,000. We used the same fiber stiffness and the same volume density as in the best-fitting model in ref. 38: density of 0.2 and stiffness of 2; where values of parameters are as defined as previously.

In the upgraded the loop-extruding factor (LEF) simulation engine, simulations do not need to be re-initialized between each subsequent step of loop extrusion. Instead, simulations are done in blocks of 100 loop-extrusion steps. At the start of each block, all LEF-mediated bonds that would occur in the next 100 loop-extrusion steps were initialized, and all but current (\(step = 0\)) bonds were given a strength of zero. Current bonds were given the same strengths as previously. Langevin dynamics (LD) was then advanced by a certain number of LD timesteps, reflecting the relative velocity of loop extrusion (1,000, 5,000, 20,000 used in this manuscript). After that, strengths of the bonds were adjusted such that only bonds that exist at step \(= 1\) of loop extrusion have nonzero strengths. This allowed us to avoid restarting simulations between subsequent extrusion steps, and do so only every 100 extrusion steps. The new engine lead to significant performance improvement, as it eliminated the necessity to frequently restart simulations. The new engine also allowed us to advance loop extrusion by only one step at a time (4 steps were used previously to decrease the number of restarts). This additionally allows higher-fidelity simulations of loop extrusion, with less abrupt motion of the polymer after updating the positions of bonds imposed by loop extrusion. The new loop-extrusion engine was added to the example folder of the openmmlib package (http://bitbucket.org/mirnylab/openmm-polymer).

We performed simulations for three different loop-extrusion speeds: 1,000, 5,000, and 20,000 LD steps per LEF step. Simulations with 1000 LD steps were run for 2,000,000 blocks of LD; simulations with 5,000 steps were run for 500,000 blocks of LD, and simulations with 20,000 steps were run for 250,000 blocks of LD. We obtained every 20th block for the simulation with 20 steps, and every 5th block for other simulations, yielding 100,000 total configuration for simulations with 1,000 and 5,000 LD steps per LEF step, and 50,000 configurations for 20,000 LD steps. Contact maps were built using capture radius of 10 monomers. Simulated FISH distributions were calculated as above.

**Experimental data.** FISH CDFs corresponding to ref. 23 were obtained from https://groups.google.com/forum/#!topic/3d-genomics/ from their UPDATED spreadsheet. Published publicly available Hi-C data\(^{23}\), GEO accession GSE63525, was re-processed, filtered, and iteratively corrected using hiclib, https://bitbucket.org/mirnylab/hiclib\(^{49}\).

**Data availability.** This manuscript used publicly available data as indicated above. Polymer simulation code relevant for this study is publicly available in the example folder of the openmmlib package (http://bitbucket.org/mirnylab/openmm-polymer).

---

47. Imakaev, M.V., Tchourine, K.M., Nechaev, S.K. & Mirny, L.A. Effects of topological constraints on globular polymers. Soft Matter 11, 665–671 (2015).

48. Halverson, J.D., Smrek, J., Kremer, K. & Grosberg, A.Y. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. Rep. Prog. Phys. 77, 022601 (2014).

49. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods 9, 999–1003 (2012).