



ARTICLE

DOI: 10.1038/s41467-018-04130-x

OPEN

Theoretical principles of transcription factor traffic on folded chromatin

Ruggero Cortini ^{1,2} & Guillaume J. Filion ^{1,2}

All organisms regulate transcription of their genes. To understand this process, a complete understanding of how transcription factors find their targets in cellular nuclei is essential. The DNA sequence and other variables are known to influence this binding, but the distribution of transcription factor binding patterns remains mostly unexplained in metazoan genomes. Here, we investigate the role of chromosome conformation in the trajectories of transcription factors. Using molecular dynamics simulations, we uncover the principles of their diffusion on chromatin. Chromosome contacts play a conflicting role: at low density they enhance transcription factor traffic, but at high density they lower it by volume exclusion. Consistently, we observe that in human cells, highly occupied targets, where protein binding is promiscuous, are found at sites engaged in chromosome loops within uncompact chromatin. In summary, we provide a framework for understanding the search trajectories of transcription factors, highlighting the key contribution of genome conformation.

¹Genome Architecture, Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain. ²Universidad Pompeu Fabra (UPF), 08003 Barcelona, Spain. Correspondence and requests for materials should be addressed to R.C. (email: ruggero.cortini@crg.eu) or to G.J.F. (email: guillaume.filion@gmail.com)

Transcription factors play a key role in the regulation of transcription¹. Upon binding to cognate regulatory sequences within the chromatin, the transcription factors trigger a cascade of molecular events culminating in the recruitment of the RNA polymerase and subsequent transcription². Understanding how transcription factors find their targets in the genome is thus the very first and pivotal step in understanding how transcription is controlled at the molecular level.

Transcription factors contain a DNA-binding domain with a non-specific affinity for DNA and a high specific affinity for some target sequence called the binding motif³. Affinity for their motifs predicts the binding of transcription factors *in vitro*⁴. However, recently acquired data using the ChIP-on-chip and ChIP-seq technologies showed a different picture in the nucleus of multi-cellular eukaryotes. Transcription factors occupy sites where their motif is absent^{5,6}, and leave unbound most sites where it is present^{7–10}, indicating that a simple protein–DNA interaction model is insufficient to describe the transcription factor dynamics *in vivo*.

The nucleus is a rich and heterogeneous environment, where higher-order interactions between molecules take place. Histones, the core component of chromatin itself, present a barrier to transcription factors binding. To overcome this constraint, the so-called pioneer transcription factors have been shown to bind their targets in the presence of a nucleosome, or even aid other transcription factors to gain access to their targets¹¹. However, as the binding of pioneer transcription factors is vastly different between cell types¹², the simple view that they bind their motif regardless of the chromatin context cannot hold true. Therefore, additional phenomena dictate and drive the search process of transcription factors within the nucleus.

One such mechanism is facilitated diffusion^{13,14}, which emphasizes the key role of non-specific affinity for the search kinetics. Transcription factors are first adsorbed onto DNA or chromatin through an electrostatic pull acting at a short distance¹⁵, and for a short period of time they diffuse on the polymer. This scanning or sliding mode is essential to discover the target sequence^{13,16}. However, both terms are somewhat inaccurate: transcription factors may merely detach and reattach to the chromatin immediately. At the points where the chromatin fibers meet, transcription factors can in theory fall off one fiber and reattach to the adjacent one, effectively jumping over large genomic distances while diffusing on the chromatin. Such effects have been observed *in vitro*, where the three-dimensional (3D) conformation of naked DNA was shown to impact the search kinetics¹⁷.

The importance of this mechanism of diffusion is well established in the eukaryotic nucleus¹⁸, and it is now clear that transcription factors diffuse intermittently in the nucleoplasm and on the chromatin fiber itself^{19,20}. However, little is known about the impact of the geometry of the chromatin fiber. Genomes have a characteristic 3D structure, revealed by chromosome capture methods such as Hi-C²¹. It is still unclear how genomes acquire a particular conformation, however, it is known that once preferential contacts are established, they can influence the trajectories of transcription factors diffusing along the chromatin.

Therefore, it is tempting to hypothesize that the knowledge of chromosome conformation could be used to obtain an insight into the search dynamics of transcription factors. However, the general principles are presently unknown for lack of a general theory. Previous work by ourselves and others suggested methods to infer the binding profiles of proteins on DNA from Hi-C matrices^{22–24}, but the validity of those approaches remains unclear as there is no guarantee that they correspond to realistic physical processes.

Here we establish the basic principles of transcription factor diffusion on folded chromatin. We use a molecular dynamics

simulation approach to investigate the role of chromosome conformation in the search process. Exploring configurations with the strings and binders model of Barbieri et al.²⁵, we find that geometry has a significant influence on the traffic of diffusing bodies with an affinity for the polymer. Strikingly, polymer loops increase traffic and occupancy in a wide range of conditions, but decrease them at high compaction. Consistently, Hi-C and ChIP-seq data show that massive protein binding accumulates at genomic loci engaged in long-distance contacts. Overall, this work suggests that the 3D conformation of the genome affects the discoverability of binding sites and contributes to the global distribution of transcription factors *in vivo*.

Results

Model and assumptions. We used molecular dynamics simulations to understand how the conformation of the genome can affect diffusion of transcription factors. This modeling strategy captures the behavior of simple objects according to realistic physical interactions. To develop a general and tractable model, we stripped down the specific features of chromatin and transcription factors to their bare essentials: chromatin was considered as a folded polymer and transcription factors as diffusible molecules with an affinity for the chromatin.

We simulated spherical particles, referred to as tracers, which were used as a model for transcription factors except for one important difference: tracers have no specific affinity for any particular site on the polymer, contrary to the specificity of transcription factors. This aspect of the model is essential to capture the dynamics of transcription factors in search mode, rather than in bound mode. Using these restrictions, tracers can be considered as either non-specific transcription factors, or transcription factors not bound to their target site. The strength of the non-specific affinity of the tracer on the polymer was labeled ϵ .

Folded polymers were simulated using a model originally developed by Barbieri et al.²⁵ and are summarized graphically in Fig. 1. Briefly, the model describes the large-scale structure of the polymer as an aggregate of stable loops formed between predefined anchor monomers. The loops are formed by special particles called binders that have a high affinity for anchor monomers and can thus bridge them together. The overall openness or compaction of the polymer depends on the number of loops, or more accurately on the fraction of these anchor monomers, labeled ϕ .

Varying ϵ and ϕ allowed us to explore a wide range of conditions. For each simulation, we computed the intra-polymer contact matrix where each entry was the number of times two given monomers were in contact during the simulation (meaning that their distance is less than a threshold t , see Methods). From this matrix, we computed the row sum, referred to as the polymer contacts, representing the total amount of contacts for each monomer. We also computed the total number of contacts between the tracers and the monomers, referred to as the traffic of the tracers. This profile represents the total amount of time tracers spent in contact with different regions of the polymer.

Most of the explored parameter space corresponds to physiologically relevant conditions. For example, recent estimates from electron microscopy suggest that within the nucleus of a typical human cell, the local chromatin density varies in the range of 12–50%²⁶, which corresponds to values of ϕ in the range of 0.1–0.4 (see Methods and Supplementary Fig. 2a). Other experiments suggest that for a typical human cell, a transcription factor spends approximately the same amount of time diffusing on the chromatin and in the nucleoplasm^{19,20}. This corresponds to $\epsilon \approx 2k_B T$ in the simulations, and for ϵ in the range of 1.0–2.5

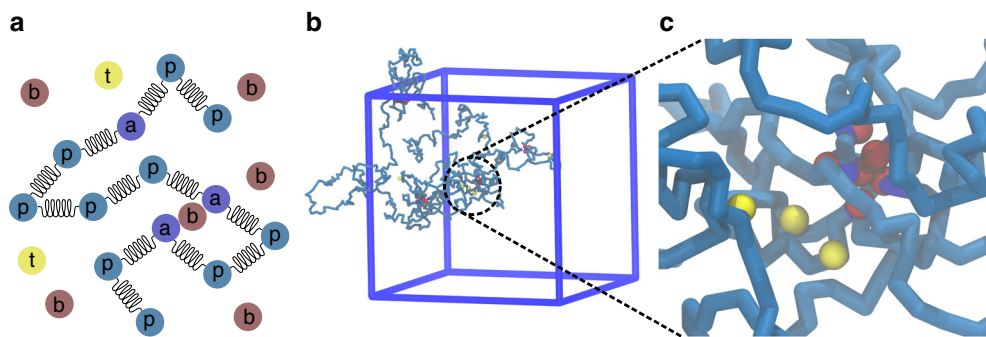


Fig. 1 Description of the simulation setup. **a** The polymer is made of p (polymer) and a (anchor) particles, connected by harmonic springs. Binders (b particles) are introduced to bridge a particles. Tracers (t particles) interact non-specifically with both p and a particles. Binders and tracers interact only by hard-core repulsion, i.e., by volume exclusion. **b** Snapshot of a simulation together with the simulation box. **c** Zoom-in of a loop formed by the binders and their binding sites, along with the tracers bound nearby (yellow)

$k_B T$, the time spent on the polymer varies from 5 to 95% approximately, which covers most possible cases (Supplementary Fig. 3c).

We carried out two independent sets of simulations, the first with 10 tracers and the second with 200 tracers. This allowed us to test the effect of the tracers themselves on the conformation of the polymer. The results of both simulations were similar. However, it is clear from the analysis of both simulations that when the tracer affinity exceeds a given threshold, the polymer changes conformation due to a local increase in the tracer concentration (Supplementary Note 1). The implications of this result are discussed in more detail within the Discussion section.

Importantly, the results clearly demonstrate the robustness of the model in regard to the details of the simulation. In particular, simulations with monovalent tracers (Supplementary Note 4), and in a crowded medium (Supplementary Note 6), gave the same qualitative results.

Polymer loops have two opposite effects on the tracer traffic. In what follows, the tracer traffic is defined as the total number of times that the tracers are detected in contact with each monomer. In other words, tracers visit more often a monomer with high traffic than a monomer with low traffic. In our simulation assay, polymer looping has two opposite effects on the traffic of the tracers, which is more clearly demonstrated when the values of the parameters are extreme (Fig. 2).

Of note is the correspondence between the monomer contact frequency and the traffic of the tracers. The correspondence is particularly clear in the cases where only a few loops are present in the polymer (low ϕ) and the binding of the tracers is strong (high ϵ).

We measured the correspondence between the polymer contacts and the tracer traffic using the Pearson correlation coefficient and the Kullback–Leibler (KL) divergence D_{KL} . Here, the KL divergence is computed as:

$$D_{\text{KL}}(C|R) = \sum_i C_i \log \frac{R_i}{C_i}, \quad (1)$$

where C_i and R_i represent the traffic on the i th monomer and the polymer contacts made by monomer i , respectively. The KL divergence and the Pearson correlation coefficient are both valid metrics although they represent different features of the data. When using the Pearson correlation coefficient, the two variables are interpreted as numeric random variables and the quantity of interest is their covariation. When using the KL divergence, the two variables are interpreted as probability distributions and the

quantity of interest lies in the information lost by using the polymer contacts array instead of the tracer traffic array.

If the polymer has few loops ($\phi = 0.02$) the high values of r and the low values of D_{KL} indicate that the total amount of contacts coincides with the traffic of the tracers (Fig. 2b). This is a direct result of the combination of the two following effects: Firstly, the on-rate of the association reaction is doubled at the contact points because tracers come from two distinct branches of the polymer. Secondly, the off-rate reduces as the local concentration of the polymer increases, thereby favouring binding (see Supplementary Movie 1). The latter effect is due to the attractive component of the interaction potential between the tracers and the monomers. The overall effect is an increase of the ratio between on-rate and off-rate, i.e., a net increase in traffic.

Highly compacted polymers with a multitude of looping sites ($\phi = 0.50$) define another category. High compaction combined with low tracer–polymer affinity (low ϵ) results in a strong anti-correlation between the contact amount and the occupancy profile of the tracers, i.e., the tracers have a tendency to visit the sites that make fewer contacts. The reason is that the polymer forms a globule, where the most visited monomers are at the surface, and their contacts with other monomers are less frequent. In comparison, when the tracer–polymer affinity is high (high ϵ) we observe a caging effect, whereby the tracers remain blocked in compacted inner structures (see below). In this case, both r and D_{KL} report that the tracer traffic and the polymer contacts are unrelated (see Supplementary Movie 2).

Figure 3 summarizes these results with the average values of the KL divergence and the Pearson correlation coefficient for all the parameters tested in our simulations. Regardless the number of tracers, a larger number of loops results in a poorer correspondence between the traffic of the tracers and the contacts of the polymer. In summary, the conformation of the polymer defines two regimes: one where the polymer is uncompact and contacts predict the traffic of the tracers, and another one where the polymer is compacted and contacts do not predict the traffic of the tracers.

High polymer compaction excludes tracers. Why does the conformation of the polymer cease to predict the traffic of the tracers when the number of loops increases? At least two non-mutually exclusive scenarios can be imagined. In the first, the polymer forms globular domains that are too dense for tracers to enter. In the second, tracers enter the domains but steric effects prevent them from binding to the anchors of the loops.

To address this question, we defined the coverage of the polymer as the percentage of monomers visited by a tracer at least

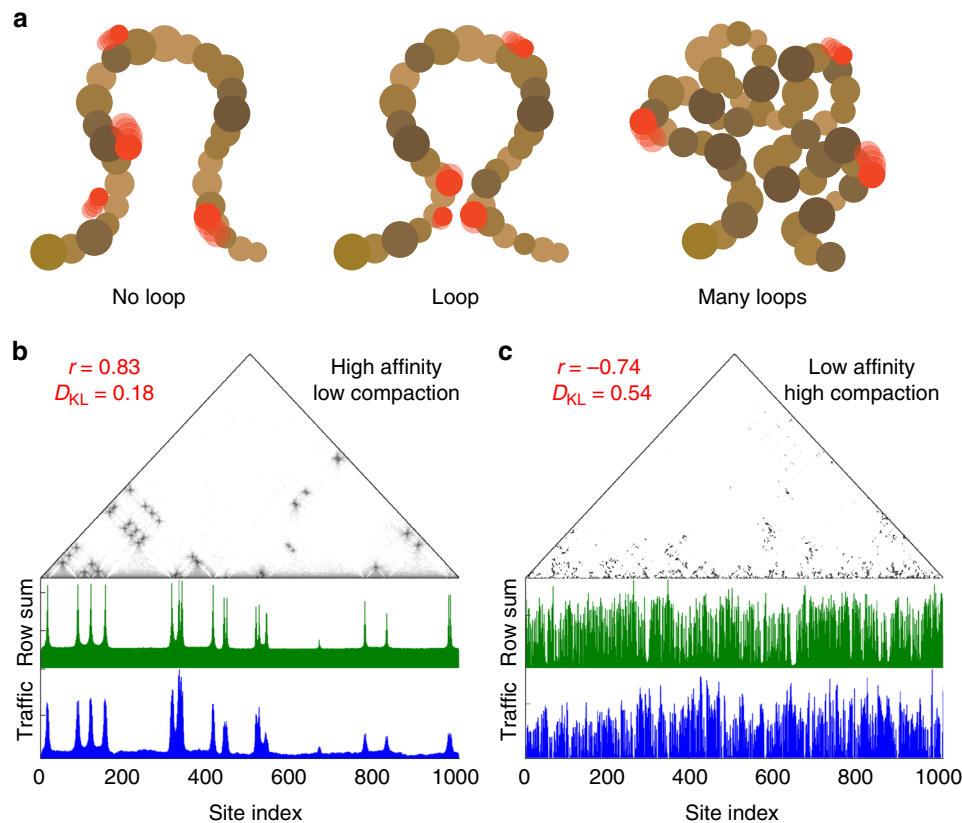


Fig. 2 The dual role of looping on transcription factor traffic. **a** Effect of chromosome conformation on the traffic of transcription factors. At low compaction (left), transcription factors (red) slide uniformly on the chromatin fiber (brown). In the presence of a loop (middle), transcription factors accumulate at the contact point. When the number of loops is high (right), transcription factors diffuse on the outer shell of the globule due to volume exclusion effects. **b, c** Examples of simulation results with the smallest and largest values of ϕ and ϵ in cross combinations. For each example, we show the contact matrix of the polymer in log scale, above the traffic of the tracers and the total amount of contacts for each monomer (Row sum). Each plot indicates the values of the Pearson correlation coefficient r between the two, along with their Kullback–Leibler divergence (see text). Panel **b** shows a simulation with low polymer compaction and high tracer affinity, resulting in a strong correlation between the polymer contacts and the tracer traffic. Panel **c** shows that the opposite happens with a low tracer affinity and high polymer compaction

once during the simulation. Figure 4 shows that the coverage decreases as the number of loops increases (high ϕ), consistent with the view that the tracers cannot access the core of the polymer. However, increasing the affinity of the tracers alleviates this effect, so the exclusion does not proceed by hard-core repulsion, otherwise even tracers with high affinity for the polymer could not penetrate the core. In fact, most of the monomers are accessible at some value of the non-specific affinity, so for the highest values of the compaction that we tested, the polymer rarely has a density as to be completely impermeable to tracers. We surmise that higher values of the compaction ϕ would result in a fully impenetrable polymer core.

We computed the average correlation between the occupancy of the tracers and the occupancy of the binders. The binders bridge the loop anchors and remain fixed on the same monomers for the duration of the simulation. As the number of loops increases (high ϕ), the correlation of the occupancy profiles becomes strongly negative, indicating that the binders exclude the tracers at looping sites. This can be partially alleviated by a higher non-specific affinity; however, the correlation remains negative at high number of loops.

These results suggest that in compact polymers, both effects apply: the core of the polymer domains becomes less accessible and the anchor sites become crowded. As a result, contacts within the polymer become poor predictors of the occupancy of the tracers.

To better understand the contribution of volume exclusion, we tested the behavior of tracers interacting with the polymer only through hard-core repulsion. We kept the polymer in a fixed state while simulating the motion of the tracers alone. In terms of traffic, a fixed polymer gives essentially the same results as a mobile polymer (since the simulated polymer structures are quite rigid) but substantially speeds up the simulations when modeling tracers without affinity for the polymer.

We ran the simulations at two extremes of the compaction spectrum (Fig. 5). At low polymer compaction ($\phi = 0.02$), tracers with and without affinity for the polymer have dissimilar traffic profiles ($r = 0.02$). This confirms the importance of non-specific affinity in shaping the traffic. In contrast, at high polymer compaction ($\phi = 0.50$), tracers with and without affinity for the polymer have comparable traffic profiles ($r = 0.58$). This shows that in this regime, the affinity between the polymer and the tracers has little impact on the occupancy. In conclusion, the pattern of tracer traffic in highly compacted polymers is mainly driven by volume exclusion effects.

Highly occupied targets in the human genome. An important question is whether the phenomena observed in our simulations are relevant to living cells. There is a general agreement that chromatin compaction tends to exclude transcription factors²⁷, but the idea that the geometry of the chromatin may guide transcription factors is still exploratory^{22,24,28,29}.

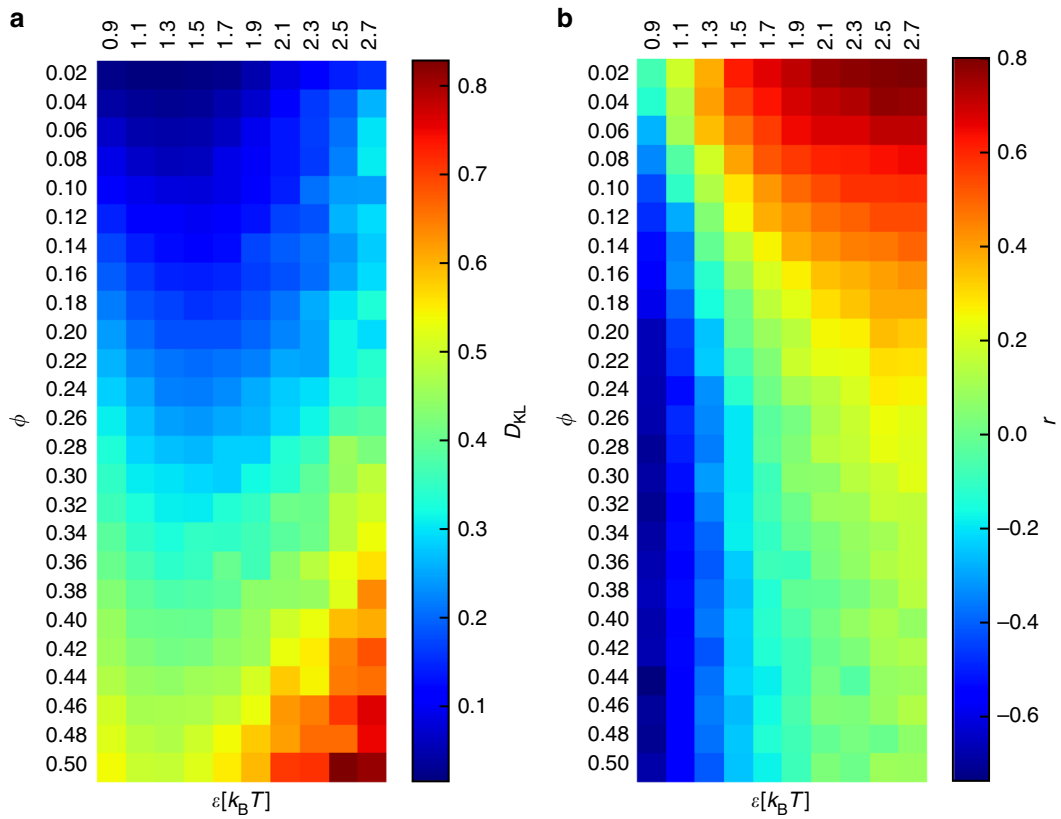


Fig. 3 Coincidence of the traffic of the tracers with the contacts of the polymers. Average values of the Kullback-Leibler divergence (a) and of the Pearson correlation coefficient (b) for different model parameters. Calculations were performed as described in Methods

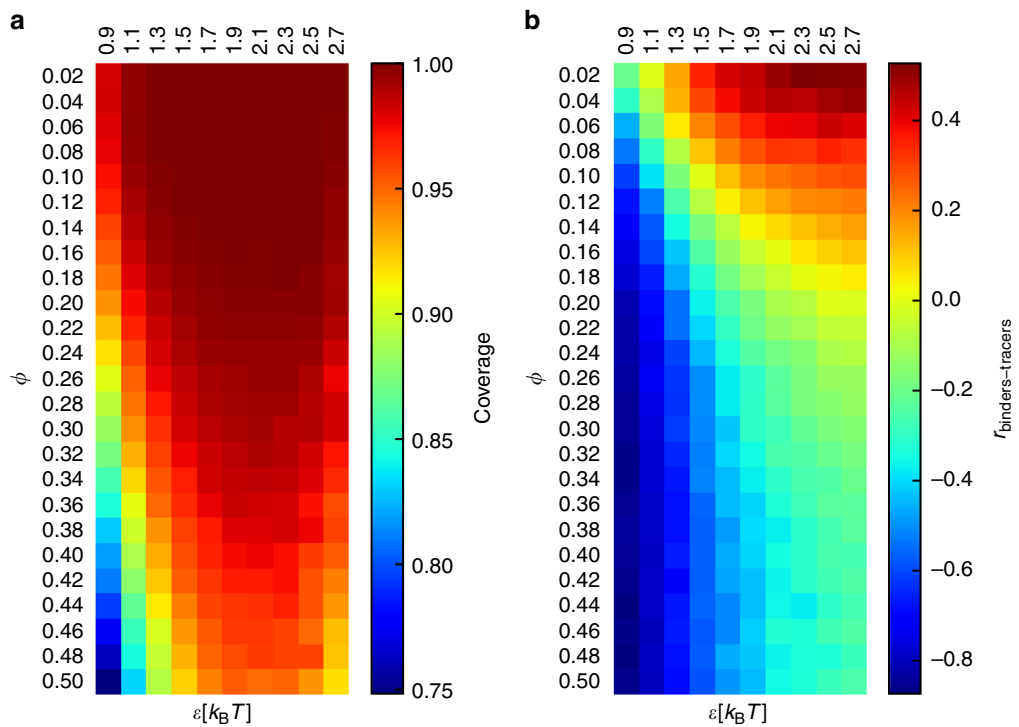


Fig. 4 Volume exclusion effects in the simulations. a Average percentage of sites of the polymer that are visited by the tracers during the simulations (coverage). b Average values of the Pearson correlation between the tracer and binder occupancy sets

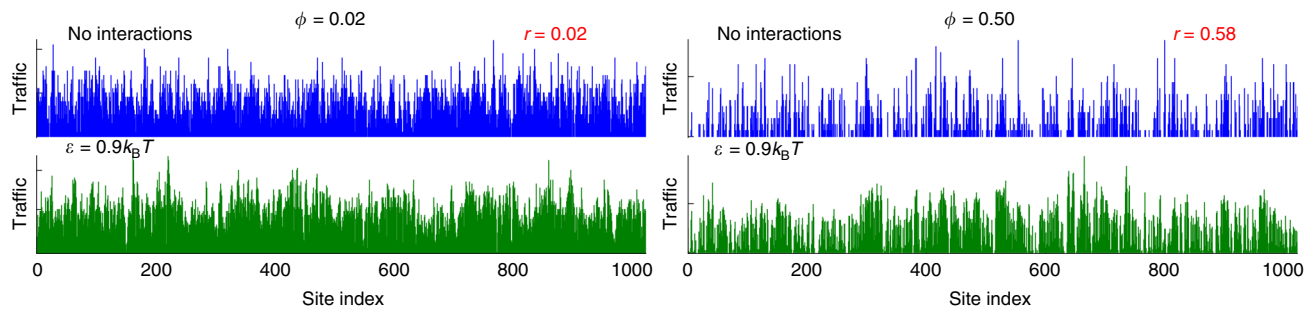


Fig. 5 Traffic with and without tracer-polymer interactions. In blue: traffic of tracers without tracer-polymer interactions, i.e., interacting only through hard-core repulsion. In green: traffic of tracers with tracer-polymer affinity equal to $0.9k_B T$. Left: low compaction with $\phi = 0.02$. Right: high compaction with $\phi = 0.50$. Values of the Pearson correlation coefficient between the polymer occupancy and the tracer traffic are reported in red

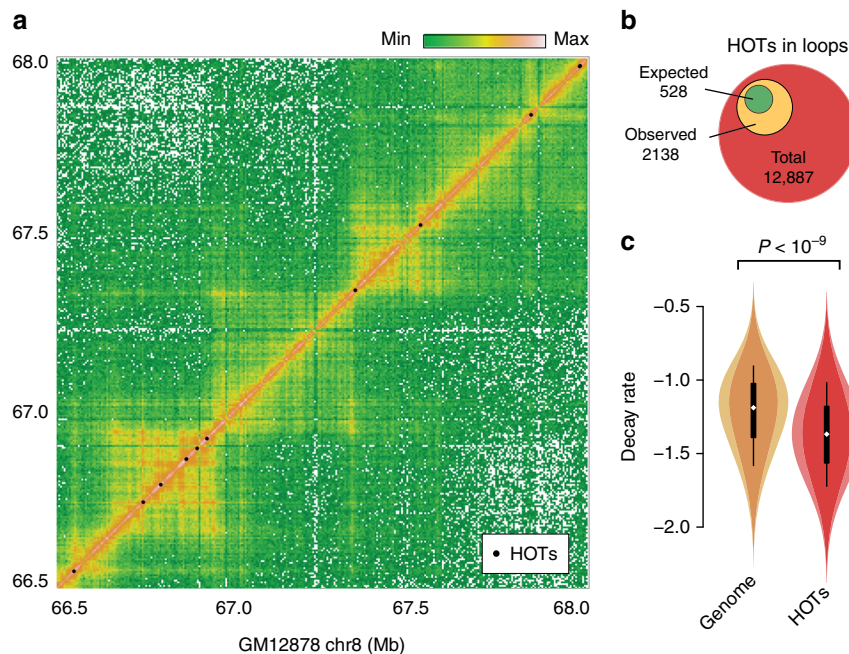


Fig. 6 Highly occupied targets in the human genome. **a** Hi-C map of the human cell line GM12878, along with the position of the highly occupied targets (HOTs) indicated by black dots. **b** HOTs are enriched at chromosomal loops in GM12878. The location of the HOTs from Foley and Sidow³⁰ was crossed with the location of the loops from Rao et al.^{30,31}. Out of 12,887 HOTs, 2138 were found in loops versus 528 expected (binomial distribution $P < 10^{-9}$). **c** HOTs are located in uncompacted chromatin. The violin plots show the distribution of contact decay in 50 kb windows in the whole genome versus each HOT. The mean values are -1.22 and -1.37 , respectively (95% confidence interval for the difference 0.146 – 0.157 , Student's t test with Welch-Satterthwaite correction and 18,080 degrees of freedom $P < 10^{-9}$). The white dot inside each violin represents the median of the distribution. The black block ranges from the first to the third quartile and the whiskers range from the first to the ninth decile

Our results indicate that at low chromatin compaction, long-range chromosomal contacts may increase the traffic of transcription factors. The simulations predict a 2–5-fold increase of non-specific binding (Fig. 2). The effect is modest, but it applies to every molecule with an affinity for chromatin and with a size comparable to that of the polymer (see Supplementary Note 5).

It is possible that loop-enhanced traffic may explain one of the least understood patterns of transcription factor occupancy, namely the existence of highly occupied targets^{5,6,30} (HOTs). HOTs, or binding hot spots, are small regions of the genome (typically less than 1 kb) bound by most proteins with an affinity for chromatin. The promiscuous binding of transcription factors in HOTs is independent of the presence of their binding motifs, suggesting that the process depends on non-specific affinity and protein–protein interactions. Could HOTs be a result of the formation of chromatin loops?

To test this hypothesis, we used high-resolution Hi-C data performed in the human cell line GM12878³¹, and HOTs locations obtained from the ChIP-seq profiles of 96 proteins mapped in the same cell line³⁰. It is visually apparent that HOTs tend to localize at regions engaged in long-distance chromosomal contacts (Fig. 6a). HOTs are strongly enriched at loop anchors, as defined by Rao et al.^{30,31}. Indeed, 2138 of the 12,887 HOTs lie at the basis of a loop, compared to 528 expected (Fig. 6b). Thus the trend is very robust, especially considering the imprecision in calling loops and HOTs.

The simulation results predict that HOTs should be present in the regions of the genome that are the least compact. The standard way to compare compaction levels is to estimate the local rate of contact decay, which is a measure of the polymer state²⁵. The probability of Hi-C contacts is locally proportional to $s^{-\alpha}$, where s is the linear separation between the loci and α is the decay rate. The most compact states of the polymer correspond to

the lowest local values of α and vice versa. The local contact decay surrounding HOTs in the dataset is significantly lower than the genome average (Fig. 6c), indicating that HOTs tend to occur within the least compacted regions of the genome, consistent with the simulation results.

Overall, these results indicate that massive protein binding tends to accumulate at loop anchors in open regions of the nucleus, consistent with our prediction that chromatin loops enhance the traffic of transcription factors in the nucleus.

Discussion

In this study, we used molecular dynamics simulations to establish the governing principles of transcription factor diffusion on folded chromatin (Fig. 2a). Despite the fact that our assumptions are reasonable, our model of chromosomes and transcription factors is simplistic, so we do not claim that it represents the real molecular detail at work in a cell nucleus. It only captures some general properties of the system.

The main feature of the polymer structure is the amount of contacts, acting in two opposite ways (Fig. 3): at low density, it increases the traffic of the tracers; at high density, the compaction of the polymer results in a volume exclusion effect that keeps the tracers away from loop anchors (Fig. 4). In the latter case, the tracers tend to interact with the monomers on the outside of globular domains, but they can also enter the structure, especially if their affinity for the polymer is high. These results are in line with experimental data obtained in live cells²⁷, where it was shown that chromatin compaction excludes diffusible factors, but no compartment is fully inaccessible.

Importantly, these effects describe the behavior of transcription factors in search mode only. Theoretical analyses³² show that chromatin loops may increase the on-rate by a factor of 2–5 (the number of fibers at the contact point, see Fig. 2), but the off-rate may vary by several orders of magnitude with the DNA sequence⁴. A 2-fold effect is large for the binding kinetics, but negligible for the average occupancy of high affinity binding sites. Consistently with the view that chromosome conformation has a mild impact on the occupancy of high affinity binding sites, it was recently shown that target sites of many transcription factors remain bound on the mitotic chromosome³³, where the conformation is radically different³⁴.

Nevertheless, transcription factors spend a large fraction of time in search mode^{19,20} and accumulate at HOTs, where most binding motifs are absent⁶. Our results suggest that this alternate binding mode depends on the conformation of the genome. The HOTs are found primarily at loop anchors in uncompacted regions (Fig. 6), as predicted by the simulations (Figs. 2 and 3). Importantly, these chromosome conformation effects apply to all molecules with an affinity for chromatin within the nucleus, provided the size of the particles is comparable to the pores of the chromosome folds (see Supplementary Note 5). This permits the appearance of emergent behaviors, induced by the interactions between the components.

Our model provides significant additional evidence to understand the formation of the 3D genome structure. In the molecular dynamics simulations, the structure of the polymer is given by the fixed distribution of the binders. This model recapitulates the essential features of Hi-C maps²⁵, but it does not correspond to the contentious issue of how the genome is actually folded³⁵. For example, the loop extrusion model proposes the existence of molecular machines capable of creating loops (i.e., cohesins^{36,37}). Other reports focused on the crucial role of epigenetic domains in shaping the 3D genome³⁸. Furthermore, architectural proteins such as CTCF also play a crucial role in shaping the 3D genome³⁹, and they too may act as the binders in our simulations.

Nevertheless, the molecular details underlying the formation of the chromatin loop is not relevant to the present work. All that matters is that the loops are stable enough for the effect described above to take place.

We showed, as mentioned in the Results section, that when the number of tracers and their affinity exceed a critical value, the tracers themselves create intra-polymer contacts (Supplementary Note 2). This effect was predicted by the strings and binders switch (SBS) model of Barbieri et al.²⁵ and the bridging-induced attraction mechanism^{40,41}, in which loops are created in a non-specific fashion. Here, we show that it is also valid when the polymer has specific loops in its structure, due to the fact that we separated the role of specific and non-specific bridging particles. This indicates that a high concentration of transcription factors may by itself induce the chromatin fiber to collapse locally, a result which suggests that the many transcription factors present in HOTs may stabilize long-range contacts. More work is necessary to clarify the effect of heterogeneous binding profiles of the transcription factors in mediating and stabilizing chromatin loops.

A strong assumption will need to be tested: transcription factors are close to saturating concentration in the nucleus, so that the weak effect of a chromatin loop is sufficient to nucleate an aggregate. It will also be important to elucidate whether such aggregates are stable or transient, and what other factors are necessary for their formation.

It was previously observed that cohesin binding is frequently associated with the presence of many transcription factors⁴², but the mechanism remained unclear. Since cohesin is necessary for the formation of the loops⁴³, we propose the following interpretation: cohesin binding induces the formation of a chromatin loop, which increases the local concentration of transcription factors. This nucleates the formation of an aggregate stabilized by protein–protein interactions. Finally, the high concentration of transcription factors compacts the chromatin the same way we observed for a large number of tracers, which further stabilizes the loop. Although we did not simulate cohesin loop formation explicitly, the binders in our simulations can be thought of as cohesin proteins. The time scale involved in the formation of cohesin loops is in fact very long⁴⁴, so that in our simulations we do not need to consider that the loops are dynamic. Once again, our model does not represent actual molecular mechanisms but only their essential properties regarding the diffusion of transcription factors.

This interpretation suggests that long-distance contacts may not be an accidental property of enhancers, but rather their essential mechanism of action. In this regard, the 3D organization of the genome could be a way to guide proteins of the same complex to the same locations in order to facilitate their assembly. This would also imply that the formation of liquid-like aggregates is dependent on the geometry of the chromatin polymer.

Our model is consistent with the experimental data regarding the distribution of HOTs in the nucleus. It also makes other predictions that can be experimentally tested. The first is that binding kinetics are 2–3 times faster when the target lies at the basis of a chromatin loop. Since more search trajectories lead to the target site when it lies within a loop, a rough estimate is that the search time is reduced by the number of chromatin fibers that meet at the cross point. The second prediction is that HOTs should disappear in mitosis. Indeed, the compaction of the chromosome and the disappearance of long-distance contacts and loops³⁴ are unfavorable conditions for the accumulation of transcription factors. Future experimental data will allow these predictions to be tested.

In summary, this study shows that the conformation of the genome should be taken into account to fully understand the

distribution of transcription factors in the nucleus of animal cells. More generally, the principles highlighted here pave the way for a general theory of transcription factor-facilitated diffusion on folded chromatin.

Methods

Simulation setup. Our simulation model is the same as the SBS model²⁵, and consists of the following elements (see Fig. 1):

- A bead–spring polymer, consisting of N total particles, which represent the individual monomers. The polymer is made of two types of particles: particles of type p (polymer), which have no special property, and type a, which are binding sites (anchors). The fraction of a particles in the polymer is the parameter ϕ .
- Binders, type b, which are free to diffuse in space, have strong attractive interactions with the a particles on the polymer. In our system, we keep the number of binders such that there are always two binders per binding site. Therefore, the number of binders is $n_b = 2\phi N$. The interaction strength is chosen in such a way that the binders are ordered⁴⁵.
- Tracers, labeled t, diffuse freely in space and have an attractive short-ranged interaction to the entire polymer, that is, with both p- and a-type particles. The number of tracers is n_t .

The polymer is described as a classical bead–spring system. Each particle in our system has a radius σ . We chose $\sigma = 15$ nm independently of the particle identity, and σ is set to be the length scale in the simulation units. The choice of the length scale is to match the approximate width of the chromatin fiber (30 nm fiber). Successive beads in the polymer are connected with harmonic springs of stiffness $k = 330k_B T/\sigma^2$, as illustrated in Fig. 1. The interaction between successive monomers is then given by

$$E_{\text{monomer}} = \frac{1}{2}k(|\mathbf{r}_{i+1} - \mathbf{r}_i| - r_0)^2, \quad (2)$$

where \mathbf{r}_i is the position vector of the i -th monomer, and $r_0 = 1.2\sigma$ is the rest position of the spring. Note that we set $k_B T = 1$ in the local simulation units.

All particles interact via the Lennard–Jones (LJ) potential with the other particles. The general interaction form is given by

$$E_{ij}(r) = 4\epsilon_{ij} \left[\left(\frac{\sigma}{r} \right)^{12} - \xi_{ij} \left(\frac{\sigma}{r} \right)^6 \right]. \quad (3)$$

Here, the coefficients ϵ_{ij} and ξ_{ij} depend only on the particle identity (p, a, b, or t); r is the center-to-center distance between the two particles. The parameter ξ_{ij} is set to zero for particles that interact only through hard-core repulsion, and it is set to 1 for particles that have an attractive component. As usual, we apply a cutoff distance for the non-bonded interactions, set to $r_{\text{cut}} = 3\sigma$.

The binders were assigned a fixed, strong interaction energy with the binding sites, $\epsilon_{ab} = 10k_B T$. The interaction between the tracers and the polymer is given by $\epsilon \equiv \epsilon_{tp} \equiv \epsilon_{ta}$. This is the other parameter that we vary in our simulations.

We used 25 different values of ϕ , from 2 to 50%, in steps of 2%, and 10 different values of ϵ , from $0.9k_B T$ to $2.7k_B T$, in steps of $0.2k_B T$. We simulated the system for all possible pairs of ϕ and ϵ . For each pair, ten independent simulations were carried out, each one with a different distribution of randomly placed binding sites (a sites) on the polymer.

The values of ϕ and ϵ that we chose are such that a large portion of the explored parameter space corresponds to physiologically relevant conditions. Measuring the radius of gyration r_{gyr} of the polymer (see Supplementary Note 2) allows us to probe the local density of the polymer by calculating

$$\rho_{\text{local}} = N \left(\frac{\sigma}{r_{\text{gyr}}} \right)^3. \quad (4)$$

Experiments by Ou et al.²⁶ on human cells suggest that the local chromatin density varies approximately 4-fold, and that the highest observed local density is of the order of 50%. In our simulations, this corresponds to $\phi \approx 10$ –40% (see Supplementary Fig. 2a). On the other hand, the experimental work by Normanno et al.²⁰ showed that the ectopically expressed TetR repressor spends about half of the time diffusing freely in the nucleoplasm and half of the time non-specifically bound to chromatin in human cells. As we show in Supplementary Fig. 3c, there is an excellent correspondence between the percentage of time that the tracers spend bound to the polymer and the value of ϵ . Fixing this value at 50% as in the case of TetR, we find $\epsilon \approx 2k_B T$, which is of the order of the weak hydrogen bonds formed by proteins transiently bound to chromatin. As ϵ varies from $0.9k_B T$ to $2.7k_B T$, the time spent bound to the polymer ranges from approximately 5 to 95%, covering most possible values.

Molecular Dynamics (MD) simulations were carried out using the HOOMD-blue software^{46,47}. To keep the temperature of the system constant at the value T ,

the system was simulated using Langevin dynamics. To each particle of the system was assigned a drag coefficient $\gamma = 1$ (in local simulation units), equal for each particle. For free particles, this gives rise to a bulk diffusion coefficient $D_0 \approx 1.2$ cm²/s in real units (see also Supplementary Note 3). We chose the simulation box to be cubic, with edge length $L = 50\sigma$. Periodic boundary conditions were applied in the simulations.

An equilibration of 2×10^7 time steps was run before taking samples of the system. After that, snapshots of the system were taken every 10^4 steps. The total length of the MD runs was 10^8 steps.

Polymer contact matrix and tracer traffic. To assess the relationship between the 3D structure of the polymer and the diffusing properties of the system, we evaluated the contact matrix of the polymer with itself. To obtain a result similar to chromosome capture-based interaction matrices, we evaluated a boolean contact matrix, which gives for each pair of particles a value of 1 if the two particles are in contact, and 0 otherwise. The criterion to establish whether the two particles are in contact was by assessing whether the center-to-center distance between the two particles is smaller than a given threshold, t . Note that distances must be calculated taking into account the periodic boundary conditions of the simulation box. All the distance matrices were computed using the Python package MDAnalysis⁴⁸. For each of the statistically independent snapshots of the system, we evaluated the contact matrix. The final matrix, H , is given by the sum of the contact matrices for each snapshot.

In a similar way to what was done for the polymer contact matrix, we evaluated the contacts between the tracers (t particles) and the polymer. For each snapshot, we evaluated the contact matrix between all the tracers and all the particles that compose the polymer. We then considered as a contact any mutual distance shorter than t . The choice of t is somewhat arbitrary. We set this value to $t = 2\sigma$, which seems a natural choice for it is the diameter of each particle, and also a distance at which the LJ interactions are practically zero. The tracer traffic at a monomer is then obtained by summing the contacts with all the tracers through all the snapshots of the simulation.

Volume exclusion effects. If there are many binding sites on the polymer (high ϕ), its configuration in three dimensions will be compact, with the binding sites buried deep inside the globular part of the polymer. The more this is true, the more we expect the tracers to be excluded from contacting the binding sites that are in the globular core. Instead, we expect the tracers to bind more to the outer shell of the globule, which is more enriched with non-binding sites (i.e., particles of type p). To test these intuitive ideas and quantify volume exclusion effects on the diffusion of the tracers, we used several metrics.

First, we calculated the tracer coverage, defined as the percentage of monomers that are visited by the tracers during the simulation time frames. If some monomers are never contacted by the tracers because of volume exclusion, the coverage will be less than 100%.

Next, we calculated the Pearson correlation coefficient between the tracer traffic and the binder occupancy (calculated the same way as the tracer traffic). If the tracers are excluded from the polymer because of the binders, we expect that the two profiles will be anti-correlated.

Biological data from GM12878. The Hi-C data were downloaded from GEO (Rao et al.^{30,31}, accession ID GSE63525). Loop positions were obtained from the file GSE63525_GM12878_primary+replicate_Arrowhead_domain_list.txt.gz. The locations of the HOTs produced by Foley and Sidow³⁰ were downloaded from <http://mendel.stanford.edu/sidowlab/downloads/hot/analysis/>. We used the file peaks_GM12878.fasta and obtained the positions of the domains from the headers. The overlap between the two was computed with custom R scripts using the package GenomicRanges⁴⁹.

To compute the rates of contact decay, the raw Hi-C reads were normalized in 50 kb windows using the TADbit pipeline⁵⁰ with default parameters. The contact decays were estimated using a simple linear regression between the logarithm of the linear separation between the windows and the logarithm of the normalized Hi-C signal. For each window, only the closest 20 windows on each side were used for the regression, as the signal in windows beyond the 20th was typically too noisy. Also, the self contacts in the window were discarded for being overly influential on the regression parameters. The slope of the regression line served as estimate for the rate of contact decay.

Code availability. The code that was used to obtain and analyze the simulation data is available at https://github.com/rcortini/sbs_tracers (Zenodo <https://doi.org/10.5281/zenodo.1201081>).

Data availability. Data available on request from the authors.

Received: 12 September 2017 Accepted: 5 April 2018

Published online: 30 April 2018

References

- Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 (2003).
- Saunders, A., Core, L. J. & Lis, J. T. Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.* **7**, 557–567 (2006).
- Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **16**, 16–23 (2000).
- Badis, G. et al. Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).
- Moorman, C. et al. Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 12027–12032 (2006).
- Kvon, E. Z., Stampfel, G., Yáñez-Cuna, J. O., Dickson, B. J. & Stark, A. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev.* **26**, 908–913 (2012).
- Farnham, P. J. Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* **10**, 605–616 (2009).
- Schmidt, D. et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
- Kaplan, T. et al. Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development. *PLoS Genet.* **7**, e1001290 (2011).
- Arvey, A., Agius, P., Noble, W. S. & Leslie, C. Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Res.* **22**, 1723–1734 (2012).
- Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
- Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).
- Berg, O. G., Winter, R. B. & Von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
- von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
- Dahirel, V., Paillusson, F., Jardat, M., Barbi, M. & Victor, J.-M. Nonspecific DNA-protein interaction: why proteins can diffuse along DNA. *Phys. Rev. Lett.* **102**, 228101 (2009).
- Mirny, L. et al. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A* **42**, 434013 (2009).
- van den Broek, B., Lomholt, M. A., Kalisch, S. M., Metzler, R. & Wuite, G. J. How DNA coiling enhances target localization by proteins. *Proc. Natl Acad. Sci. USA* **105**, 15738–15742 (2008).
- Tafvizi, A. et al. Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* **95**, L01–L03 (2008).
- Izeddin, I. et al. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife* **3**, e02230 (2014).
- Normanno, D. et al. Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat. Commun.* **6**, 7357 (2015).
- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Avcu, N. & Molina, N. Chromatin structure shapes the search process of transcription factors. Preprint at <https://www.biorxiv.org/content/early/2016/04/25/050146> (2016).
- Corrales, M. et al. Clustering of *Drosophila* housekeeping promoters facilitates their expression. *Genome Res.* **27**, 1153–1161 (2017).
- Wang, Y., Fan, C., Zheng, Y. & Li, C. Dynamic chromatin accessibility modeled by Markov process of randomly-moving molecules in the 3D genome. *Nucleic Acids Res.* **45**, e85–e85 (2017).
- Barbieri, M. et al. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl Acad. Sci. USA* **109**, 16173–16178 (2012).
- Ou, H. D. et al. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).
- Bancaud, A. et al. Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J.* **28**, 3785–3798 (2009).
- Bénichou, O., Chevalier, C., Meyer, B. & Voituriez, R. Facilitated diffusion of proteins on chromatin. *Phys. Rev. Lett.* **106**, 038102 (2011).
- Brackley, C. A., Cates, M. E. & Marenduzzo, D. Facilitated diffusion on mobile DNA: configurational traps and sequence heterogeneity. *Phys. Rev. Lett.* **109**, 168103 (2012).
- Foley, J. W. & Sidow, A. Transcription-factor occupancy at HOT regions quantitatively predicts RNA polymerase recruitment in five human cell lines. *BMC Genomics* **14**, 720 (2013).
- 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).
- Li, G.-W., Berg, O. G. & Elf, J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **5**, 294 (2009).
- Teves, S. S. et al. A dynamic mode of mitotic bookmarking by transcription factors. *eLife* **5**, e22280 (2016).
- Naumova, N. et al. Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).
- Cortini, R. et al. The physics of epigenetics. *Rev. Mod. Phys.* **88**, 025002 (2016).
- Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
- 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. USA* **112**, E6456–E6465 (2015).
- 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).
- Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
- Brackley, C. A., Taylor, S., Papantonis, A., Cook, P. R. & Marenduzzo, D. Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl Acad. Sci. USA* **110**, E3605–E3611 (2013).
- Brackley, C. A., Johnson, J., Kelly, S., Cook, P. R. & Marenduzzo, D. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* **44**, 3503–3512 (2016).
- Yan, J. et al. Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* **154**, 801–813 (2013).
- Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
- Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, e25776 (2017).
- Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 29775 (2016).
- Anderson, J. A., Lorenz, C. D. & Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* **227**, 5342–5359 (2008).
- Glaser, J. et al. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Comput. Phys. Commun.* **192**, 97–107 (2015).
- Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
- Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
- Serra, F. et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **13**, e1005665 (2017).

Acknowledgements

R.C. is grateful to Jean-Marc Victor, the members of the CNRS GDR 3536 (ADN), and Francesco Alessandro Massucci for useful discussions. The authors would like to thank Roni H.G. Wright for help with editing the manuscript, Egor Tiavlovsky for his help on an early version of the manuscript, and the CRG Scientific Information Technologies for helping with the simulation setup. We acknowledge the financial support of the Spanish Ministry of Economy and Competitiveness (Centro de Excelencia Severo Ochoa 2013–2017, SEV-2012–0208, Plan Nacional BFU2012-37168), of the CERCA Programme/Generalitat de Catalunya, of the European Research Council (ERC Synergy Grant 609989), and of the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement 608959.

Author contributions

R.C. set up, ran, and analyzed all simulations. G.F. analyzed the data from GEO. R.C. and G.F. designed the study and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-04130-x>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the

article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018