Clustered transposon insertion via formation of chromatin loops

Roshan Prizak^{1,a} and Lennart Hilbert^{1,a,b}

^aKarlsruhe Institute of Technology, Institute of Biological and Chemical Systems, Eggenstein-Leopoldshafen, Germany; ^bKarlsruhe Institute of Technology, Zoological Institute, Department of Systems Biology and Bioinformatics, Karlsruhe, Germany

Transposons, which are DNA sequences that can move to new positions in the genome, make up a large fraction of eukaryotic genomes and occur in clusters. The insertion of transposons into the genome is hindered by compact folding of chromatin, supposedly preventing aberrant or even pathogenic insertion. Chromatin can, however, be decompacted as a consequence of transposon insertion, leading to increased accessibility and, in consequence, further insertions. While these observations suggest a positive feedback between chromatin unfolding and transposon insertion, how such a feedback might contribute to clustered transposon insertion remains poorly understood. In this study, we analyze polymer models of a self-interacting chromatin domain that unfolds as increasing numbers of transposons are inserted and block the self-interaction. On the one hand, we find that, if additional transposons are inserted adjacently to already inserted transposons, the unfolding of the chromatin domain changes from a sharp globule-coil transition to a more gradual extension of loops from a core that remains folded. On the other hand, we find that adjacent transposon insertion emerges either when transposases are excluded from densely packed chromatin, or when transposon insertion proceeds very quickly in relation to the thermal equilibration of polymer configurations. We thus derive from our model physical conditions for clustered transposon insertion and the resulting spatial compartmentalization of chromatin. An according role was recently suggested for LINE-1 and Alu repeats, which occur in clusters and drive the mesoscopic compartmentalization of the mammalian genome.

Polymer simulations | Chromatin accessibility | Genome organization | Repetitive elements

Transposons, also called transposable elements, are sequences of DNA that are capable of inserting themselves into new locations in the genome. Transposon insertion often disrupts the function of existing DNA sequences, such as regulatory sequences and genes, and can compromise genome stability (1). These disruptions can lead to disease (2), including carcinogenesis and tumor progression (3–5). Accordingly, organisms have developed mechanisms to prevent transposition and remove or repress transposons (6, 7). However, due to high transposition rates, transposons also contribute to genotypic and phenotypic variation (8–10). As major drivers of genome evolution, they often get retained on the genome and acquire new functions. For instance, they contribute to gene regulation (11), replication (1) and development (12). Transposons make up a surprisingly large fraction of eukaryotic genomes – about 46% of the human genome, and an astonishing 85% of the maize genome, highlighting their ubiquity.

Bioinformatic analysis of reference genomes reveals that transposons are not uniformly distributed throughout the genome, but in clusters. This transposon distribution

is shaped by a combination of insertion processes and evolutionary changes in DNA sequence, but the relative contributions of these processes is still under debate (13). De novo insertions cluster on the 10 kb - 1 Mb scale, and often show a preference for regions containing active genes (13, 14). Transposon insertion is mediated by biological molecules like transposases and ribonucleoprotein complexes, which have to navigate the nuclear space, shaped in part by the three-dimensional (3D) organization of chromatin (15–19). The transposase-accessibility of chromatin can be assessed by the according ATAC method (20). Combining this method with sequencing (ATAC-seq) reveals differences in accessibility at the level of the chromatin fibre and nucleosome arrays (20). Combination with fluorescent tagging and imaging (ATAC-seq) suggests that also at scales of 100 nm and above, decompacted chromatin is more accessible to transposases (21). Differential accessibility for transposition might result from steric restriction (15, 17, 19, 22) or obstruction of the formation of relevant macromolecular assemblies by the surrounding chromatin mesh-work (23). Accordingly, differences in large-scale chromatin organization seem to play a role in the non-uniform insertion of transposons.

Recent studies have pointed out how transposon activity could, in turn, shape 3D chromatin organization. LINE-1 and Alu repeats, for example, have been implicated in the spatial separation of the genome into the A and the B compartment (24). Transposons can also serve as contact domain boundaries, as illustrated by the MERV family in the early mouse embryo (25), the HERV-H family in human pluripotent stem cells (26), and the MIR family in CD4⁺ T cells (27). Transposon insertion can also perturb chromatin organization, for instance, by disrupting sequence elements that set up chromatin organization, and lead to increase in chromatin accessibility (12, 25). Transposon activation and transcription can also induce chromatin unfolding and relocation of transcribed regions into the active nuclear compartment, away from inactive chromatin (28–36). Thus, both transposon insertion and transposon activity can trigger chromatin unfolding, potentially leading to chromatin configurations that favour transposase accessibility.

These observations suggest a possible feedback between transposition-mediated chromatin unfolding and further propensity for transposon insertion. While such a feedback can be expected to favor clustered insertion, it is still poorly understood under which conditions and by what exact process this feedback might take effect. In this study, we used polymer simulations to assess how the positive feedback between transposon insertion and chromatin unfolding affects transposon distribution. We find that unclustered insertion of

transposons leads to a sharp unfolding of chromatin, while clustered insertion leads to a gradual unfolding in the form of loops. Further, we find that larger transposases and faster transposition favor this unfolding in the form of loops, resulting in clustered transposon insertion.

Model overview

We model a 300-kb region of chromatin as a polymer chain with $N_{\text{tot}} = 100$ monomers. Monomers can be of two types - native or transposed. Only native monomers can cross-link with each other when they are in close proximity (Fig. 1a). Transposon insertion converts a native monomer to a transposed monomer, which cannot cross-link with other monomers. Each monomer represents a 3-kb chromatin region, and is assumed to have an effective diameter of 30 nm. We assume a persistence length of 3 monomers, or 90 nm, in line with experimental estimates of 50-200 nm (37), and consider harmonic bonds between adjacent monomers of the linear polymer chain. The cross-linking energy of native monomers is $1.5 k_B T$, in line with parameters used in previous chromatin polymer simulations (37, 38). All parameters are assigned effective values that include the influence of typical chromatin-associated proteins and RNAs.

Each simulation run starts from a compacted polymer configuration with only native monomers, and proceeds by alternating between transposon insertions and simulation segments during which polymer dynamics are evaluated (Fig. 1b). Transposon insertions are implemented differently in two different versions of the model. In the “basic model”, a placement adjacency parameter, $\alpha \in [0, 1]$, is fixed at the beginning of each simulation run. For each insertion, we choose an adjacent insertion position with probability α . At $\alpha = 0$, we generate polymers with insertion positions randomly chosen from all available positions, and at $\alpha = 1$, we generate polymers with insertion positions picked only from available positions that are adjacent to existing transposons (Fig. 1c). All simulation segments are of equal duration, containing

Fig. 1. Adjacent transposon insertion converts a sharp unfolding transition into gradual loop formation. (a) Schematic of the model chromatin polymer consisting of native monomers, which can form cross-links (yellow) with each other, and transposed monomers, which cannot form such cross-links. (b) The radius of gyration (R_g) of a chromatin polymer of fixed length ($N_{\text{tot}} = 100$) increases as transposons accumulate. Transposons accumulate one at a time, starting from a polymer with no transposons. The insertion history, referring to the sequence of positions on the linear polymer where transposons are inserted, is depicted in the inset. We fit a Hill curve to $R_g(N_{\text{tr}})$ to capture the sharpness of unfolding and the transition point. (c) A placement adjacency (α) of 0 leads to sharp unfolding, non-zero values of α lead to gradual unfolding. (d) Example snapshots (top row: $\alpha = 0$, bottom row: $\alpha = 1$) from the polymer simulations at different numbers of transposons (N_{tr}). Native monomers are shown as a white backbone. A semi-transparent white sphere with a radius equal to 40% of the bounding sphere for native monomers is shown to depict the polymer core. Transposed monomers are shown either in dark gray (if they are outside the semi-transparent sphere), or in pink (if they are inside the semi-transparent sphere).

enough timesteps to allow the polymer to equilibrate between consecutive insertions. For each simulation run and number of inserted transposons N_{tr} , we quantify the overall level of compaction by computing the polymer’s radius of gyration, R_g (Fig. 1b,c). In the “extended model”, the interaction of transposases with chromatin is modelled explicitly. A transposase molecule that is near a native monomer can convert it into a transposed monomer, marking a successful transposition. The transposase is then removed from the simulation, and a new transposase molecule is added at the boundary of the simulation volume to conserve the total number of transposases.

Results

The chromatin polymer unfolds as the number of transposons increases. How does an increasing number of transposons affect the chromatin polymer configuration? Chromatin polymers with few transposons adopt a compact configuration facilitated by cross-links between the native monomers. With further transposition, the polymer unfolds into an open configuration. As the polymer’s radius of gyration R_g represents the effective size of the polymer, this unfolding is quantified by the observed increase in R_g with the number of transposed monomers N_{tr} (Fig. 1b). The unfolding proceeds similar to the well-known coil-globule (CG) transition (SI Fig. 1) (39). Here,

Significance Statement

A large part of the genome is composed of repetitive sequences, so-called transposons. Transposons are involved in important processes, such as early embryonic development or control over which genes are used by the cell. Transposons frequently occur in clusters, where many similar sequence motifs are grouped together. Recent studies suggest that the insertion of transposons can result in local unfolding of the genome, favoring insertion of yet more transposons. Our work simulates a simplified region of the genome and transposases, which are the molecules that insert transposons into the genome. Surprisingly, large and fast-acting transposases favor the formation of distinct loops that contain most of the inserted transposons, providing a potential explanation for the clustered insertion of transposons.

Study design: RP, LH; investigation: RP; manuscript draft and editing: RP, LH.

The authors declare that they have no conflict of interests.

¹To whom correspondence should be addressed. E-mail: roshan.prizak@kit.edu, lennart.hilbert@kit.edu

the number of transposons at which the CG transition occurs (transition point) depends on the cross-linking energy, with weaker cross-linking leading to unfolding at a lower number of transposons. We fit a Hill curve to $R_g(N_{tr})$ to capture the sharpness of unfolding in terms of the Hill exponent n and the transition point in terms of the half-value K (Fig. 1b; see Appendix for details). A larger Hill exponent characterizes sharper unfolding, and a larger half-point characterizes a transition at a higher number of transposons.

Clustered insertion of transposons leads to gradual unfolding.

We can now use our simulations to assess how clustered transposon insertion might affect chromatin unfolding. To this end, we perform transposon insertion for different placement adjacency values α , and assess the resulting insertion histories and unfolding trajectories. By varying $\alpha \in [0, 1]$, we obtain insertion histories ranging from no placement adjacency ($\alpha = 0$) and hence no clustering, to maximal placement adjacency ($\alpha = 1$) resulting in complete clustering (insets, Fig. 1c). Placement adjacency α strongly affects the CG transition – while $\alpha = 0$ results in a sharp, switch-like unfolding similar to the canonical CG transition, transposon clustering at $\alpha = 1$ results in a gradual unfolding. In other words, when transposons are preferentially inserted in transposon-adjacent positions, transposon clusters emerge and the polymer unfolding becomes more gradual (Fig. 1c).

The influence of placement adjacency on polymer unfolding upon N_{tr} can be understood by visualizing the 3D polymer configurations (Fig. 1d). As the first transposons get inserted, they form a shell on the outer layer of the compact polymer core. For low α , as transposons accumulate further, the shell grows, but no unfolding occurs as long as the polymer core is intact (core-shell configuration). The core starts to open up around the transition point, leading to a sharp unfolding. On the other hand, clustered transposon insertion at large α leads to a transposon cluster looping away from the compact polymer core (loop configuration). This loop increases in size with further transposition, leading to a gradual increase in R_g . Taken together, we find that while unclustered transposon insertion leads to sharp unfolding via the core-shell polymer configuration, clustered transposon insertion leads to gradual unfolding via loop formation.

Larger transposases and rapid transposon insertion favor clustered insertion.

The observation that clustered transposon insertion leads to gradual unfolding begs the question of how such an insertion history might emerge. Transposons are preferentially inserted in less densely packed chromatin (19, 21), suggesting that clustered insertion might arise from chromatin packing differences. To further investigate this possible mechanism, we now analyze our extended model, in which we explicitly treat transposase molecules (Fig. 2a-b). In the basic model considered before, the position of each insertion, and consequently the insertion history, depend on an assumed value of placement adjacency α . In the current model, insertion history depends on the access of transposase to different parts of the polymer chain and emerges out of the transposase-chromatin interactions in 3D space. We assess the impact of two properties of the transposases. First, we consider size-based exclusion of transposase molecules from dense chromatin by varying transposase size, s_T

Fig. 2. Large transposases and fast insertion result in gradual unfolding and adjacent transposon insertion. (a) A simulation run is composed of multiple simulation segments interspersed by transposon insertion attempts. In each simulation segment, the chromatin polymer and other molecules in the system move according to the potentials they experience. In each insertion attempt, native monomers close to a transposase are converted to a transposed monomer. The transposon insertion rate is implemented by changing the number of iteration steps in the simulation segments in between consecutive insertion attempts. (b) Schematic of the polymer model with transposon insertion mediated by diffusing transposases. Differential accessibility of chromatin is considered via a size-based exclusion, which restricts access to compacted polymer regions. (c) The maximum-likelihood estimate α_{mle} for simulation runs for different values for the rate of insertion k_T and relative size of transposase s_T . (d-e) Hill exponent n and half-point K of the Hill curve fitted to $R_g(N_{tr})$ for different values of k_T and s_T .

(Fig. 2b). Exclusion of macromolecules from densely folded chromatin has been theoretically discussed (15, 17) and experimentally shown (16, 22). Effective pore sizes of 16-20 nm were determined for highly compacted chromatin (16, 40), matching well with the diameter of proteins ranging from 5-20 nm (17), implying the possibility of volume-based exclusion of large proteins from highly compacted chromatin. Second, we vary the rate of transposition, k_T (Fig. 2a). While the spontaneous rates of transposition are low (smaller than 1 per genome per generation), transposition bursts with high rates of transposition (5-10 transpositions per genome per generation) have been shown to occur in stress-related conditions in various species (10, 41, 42). We consider a wide-range of transposition rates, with low rates accounting for transpositions slower than, and high rates accounting for transpositions much faster than the chromatin polymer equilibration time.

Fig. 3. Clustered insertion is driven by exclusion of transposases from the polymer core. (a) Average insertion distance (first 30 insertions, relative to polymer center, normalized by R_g) for different transposase size s_T and insertion rate k_T . (b) Distance distribution native monomers, transposed monomers, and insertion events (first 30 insertions) for four different sets of s_T and k_T . (c) Distance distribution of transposases for four different sets of s_T and k_T .

To characterize the polymer unfolding profile, we fit Hill curves to $R_g(N_{tr})$ traces for different transposase sizes and insertion rates (Fig. 2d-e). For large transposases or fast insertion, the resulting Hill exponents are small, pointing to gradual unfolding as observed previously for large α . On the other hand, for small transposases or slow insertion, the resulting Hill exponents are larger, pointing to sharp unfolding as observed previously for small α . In the basic model, we saw that α and the resulting transposon clustering determined the unfolding profile. To check whether the differences in Hill exponents result from differences in transposon insertion bias also in the extended model, for each insertion history, we perform a maximum likelihood estimation of the underlying placement adjacency α_{mle} (Fig. 2c, see Methods and Appendix for details). We find that smaller transposases and slower insertion result in fewer adjacent insertions, consistent with the sharp unfolding and large Hill exponents. Thus, both larger transposases and quicker insertion result in more adjacent insertions, in line with gradual unfolding and small Hill exponents (Fig. 2d-e).

Large transposases and high insertion rates favor transposition in the polymer shell. Based on the observation of a core of native monomers in some of our simulations, we hypothesized that the ability of transposases to access this core is related to clustered transposon insertion. Indeed, for large transposases and high insertion rates, the mean normalized insertion distance was increased (Fig. 3a). A more detailed analysis confirms that fast insertion and large transposase size lead to more insertion within the shell and less insertion in the polymer core (Fig. 3b). When fast insertion and large transposase size are combined, insertion is almost fully restricted to the outer shell (Fig. 3b). This restriction of insertions towards the outer shell is reflected by an exclusion of transposases from the polymer core (Fig. 3c).

Clustered transposon insertion is associated with unfolding via loop formation. Taken together, our model analysis suggests

four scenarios how chromatin unfolding relates to progressive transposon invasion (Fig. 4a). Scenario I, which occurs for small transposases and slow insertion, is characterized by polymer unfolding via the core-shell mechanism (Fig. 4b). Scenarios II and III are characterized by unfolding via several small loops (Fig. 4b). In scenario II, large transposases cannot penetrate past the outer polymer shell, leading to clustered transposition. In scenario III, small transposases can, in principle, access the polymer core, but due to their fast transposition rate become exhausted before actually reaching the core. In scenario IV, in which transposons are large and insert transposons quickly, typically only one prominent loop emerges (Fig. 4b).

Discussion

In this study, we used polymer simulations of a model chromatin domain to study how clustered transposon insertion can arise from a feedback between transposon insertion and chromatin unfolding. We found that large transposons as well as fast transposon insertion favor clustered insertion. The crucial assumption for this effect is that transposon integration drives unfolding of chromatin, leading to the preferred integration of further transposons at loops that protrude from a globular polymer core formed from chromatin without inserted transposons. Indeed, in polymer models of retroviral insertion, insertion in decompacted is highly favored (19). These findings differ from the intuitive expectation that smaller molecules should access chromatin more easily, and faster insertion should favor transposon invasion. While our model is a strong simplification of the actual transposition process, it still presents an explanation of how changes in chromatin folding can shape the course of transposon invasion. Clustered transposon insertion, as predicted in our model, has been suggested as a major driver of the three-dimensional compartmentalization of chromatin regions with different epigenetic states. For example, transcriptional activation of non-coding LINE-1 repeat elements after fertilization regulates global chromatin accessibility in the early mouse embryo (12). Also, the non-coding repeats LINE-1 (clustered within heterochromatin regions) and Alu (clustered within euchromatin regions) exhibit homotypic interactions, leading to formation of the two spatially separated genomic A and B compartments (24, 43). Lastly, the relatively unfolded state of euchromatin has been associated with the SAF-A (also called U1 snRNP or HNRNPU) mediated retention of coding as well as non-coding RNAs (35, 44, 45). Many of the retained RNAs are transcribed from clustered non-coding elements (46), and our work provides a model for the emergence of such clustered placement of non-coding elements.

We assume chromatin compaction via cross-links and size-based transposase exclusion as key mechanisms underlying the clustered insertion of transposons. Other mechanisms have also been suggested as causes for clustered transposon insertion. First, considering chromatin compaction and unfolding, several relevant biophysical mechanisms have been identified (47, 48). For example, heterochromatin can be compacted by embedding in liquid-like or gel-like membrane-less organelles (49, 50). Another mechanism that was suggested for heterochromatin compaction is a polymer collapse transition, similar to the coil-globule transition seen in our model (51). A model that was recently applied to the formation of com-

Fig. 4. Different scenarios of transposon insertion and polymer unfolding. (a) Modulation of transposase size and insertion rate leads to four different scenarios of transposon-induced unfolding. In scenario I, small transposases and slow insertion lead to unclustered insertion and polymer configurations consisting of a compact core of native monomers and a shell of transposed monomers around it. Scenarios II and III exhibit unfolding via the formation of several small loops that occur as a result of clustered insertion - either because of size-based exclusion of large transposases or fast insertion on the outer shell. Scenario IV typically exhibits one prominent loop of clustered transposons. (b) Representative snapshots of polymer simulations at increasing numbers of transposons in the four scenarios (left to right, $N_{tr} \approx 10, 33, 50, 66$).

309 packed chromatin clusters by the Structural Maintenance of
310 Chromosomes (SMC) proteins is bridging-induced attraction
311 (38, 52). Unfolding of euchromatin, and specifically regions
312 harboring transcriptionally active genes, was explained via a
313 dispersing effect of nascent and chromatin-associated RNA
314 transcripts, leading to microphase separation from inactive
315 chromatin (35, 36). Second, considering the aspect of feed-
316 back between chromatin modifications to transposon insertion,
317 physical mechanisms different from chromatin decompaction
318 have also been investigated in terms of their effect on chro-
319 matin accessibility. For example, a screen of several hundred
320 transcription factors revealed that, whereas some of the DNA-
321 binding factors are excluded from compact chromatin, others
322 have a strong bias to colocalize with chromatin without being
323 affected by compaction (18). Another mechanism that might
324 limit DNA modification to distinct domains is the formation of
325 microphase-separated contact domains that restrict the reach
326 of chromatin-binding enzymes (53–55). Formation of such
327 contact domains was attributed to the microphase separation
328 of epigenetically distinct chromatin regions and was also de-
329 scribed by block co-polymer simulations (56–58). Another
330 relevant physical property of the chromatin fibre might be
331 the local stiffness, with previous work proposing that genomic
332 loci with lower bending stiffness might favor the integration
333 of transposons (59). Lastly, super-coiling of the chromatin
334 fibre was found to correlate with transcriptional activity and
335 chromatin decompaction (60). Polymer simulations show that
336 super-coiling can lead to clustered insertion via a positive
337 feedback loop similar to the positive feedback we found for
338 chromatin decompaction (61). From these different mecha-
339 nisms and explanations, a common theme emerges: clustered
340 insertion occurs via a positive feedback between a chromatin
341 modification resulting from transposon insertion and further,
342 localized insertion of transposons. Different processes can in-
343 stantiate this type of feedback, implying that clustered trans-
344 poson insertion can emerge on the basis of several biophysical
345 mechanisms.

346 In our model, transposases diffuse inward from the walls
347 of the volume containing the simulated 300-kb chromatin
348 polymer. This picture matches the “copy-and-paste” transpo-
349 sition mechanism. However, except for insertions occurring
350 within this local chromatin domain, our model would be the
351 same for “cut-and-paste” transposition. A polymer of 300
352 kb length also matches the scale at which clustered de novo
353 insertions of LINE-1 transposons was observed (13). Chro-

354 matin domains of this size hierarchically associate into larger
355 compartment domains (54, 62), raising the question of how
356 transposon insertion within a given 300-kb domain is related
357 to this overall genome organization. Similar to recent findings
358 on non-homologous double-strand break repair, the spreading
359 of transposon insertion might be limited by a CTCF-mediated,
360 relative reduction of 3D chromatin-chromatin contacts be-
361 tween neighboring domains (55). A better understanding of
362 how changes in chromatin accessibility might proliferate into
363 higher-order 3D interactions would require the deployment of
364 multi-scale polymer models, providing a relevant direction for
365 future work.

366 An uncontrolled sequence of transpositions, called transpo-
367 sion invasion, can severely disrupt genome integrity. Metazoans
368 and plants have developed mechanisms to protect against
369 transposon invasions, such as the piRNA pathway (63, 64) or
370 specific siRNAs (65). Our work suggests that certain modes
371 of transposon insertion that lead to gradual unfolding of chro-
372 matin could make the affected chromatin region vulnerable
373 to further invasion. Perhaps surprisingly, small transposases,
374 which lead to sharp unfolding, could in fact be less effective
375 in an invasion, as their unbiased insertion reduces chromatin
376 unfolding during the initial transposition events. Transpo-
377 sion insertion and removal processes span a wide range of
378 timescales, starting from the physiological timescale, on which
379 transposon insertion and transposon silencing occur, via the
380 inter-generational timescale of transposon silencing, up to evo-
381 lutionary timescales of transposon removal or proliferation
382 (66). To fully investigate the dynamics of transposon inva-
383 sion, including parameter regimes or situations which offer
384 protection against invasion, will require models that consider
385 transposon insertion as well as transposon silencing or removal
386 on various timescales.

387 Materials and Methods

388 **Model details.** Monomers in the chromatin polymer and trans-
389 posases (referred to as atoms henceforth) interact with phenom-
390 enological force fields (potentials) and follow Newton’s laws of motion.
391 We simulate the spatial dynamics of the system in the molecular
392 dynamics simulator LAMMPS (67). At the core of the simulation
393 is a standard Velocity-Verlet algorithm that solves the underlying
394 Langevin equations describing the system dynamics. The chromatin
395 polymer is held together by harmonic springs between adjacent
396 pairs of monomers, and the polymer’s bending rigidity is modeled
397 by a Kratky-Porod potential. To model pure (steric) repulsion be-
398 tween any pair of atoms, we consider the Weeks-Chandler Andersen
399 (WCA) potential, which is a shifted version of the Lennard-Jones

(LJ) potential with a cutoff at zero-crossing point. To model steric repulsion, this cut-off distance is typically set to be the mean diameter of the two interacting atoms. As all the monomers have a diameter $\sigma = 30$ nm, for all monomer pairs, the cutoff distance is set as $r_{\text{cut}} = 2^{1/6}\sigma$. Transposases have a size $s_T\sigma$, so the cut-off distance for the repulsive WCA potential between any pair of transposases is $2^{1/6}s_T\sigma$. For a monomer-transposase pair, we set the cutoff distance to be $2^{1/6}s_T\sigma$, a value smaller than the mean diameter of a monomer-transposase pair. This allows the transposase molecule to come closer to a monomer molecule than another monomer molecule can. Moreover, the cutoff distance is smaller for smaller transposases, allowing them to come closer to monomers than larger transposases. In effect, this models size-based exclusion of transposases from chromatin. Cross-linking interactions (cross-linking energy $E = 1.5 k_B T$) are modeled by using a shifted and truncated Lennard-Jones potential with a larger cut-off distance of 2.5σ . If a pair of native monomers are within a distance between $2^{1/6}\sigma$ and 2.5σ , they experience an attraction and form a cross-link.

Simulation run. Each simulation run in LAMMPS starts with an initial polymer with no transposons, and consists of an initial “soft” phase, an equilibration phase, and finally the main simulation block in which transposons are sequentially inserted into the chromatin polymer. In the basic model without transposases, the main simulation block contains $100000(N_{\text{tot}} + 1)$ time-steps, insertion is executed every 100000 time-steps. In the transposase model, insertion is attempted every $1000/k_T$ time-steps, leading to $100 - 10000$ time-steps per simulation segment. In between simulation segments, every native monomer that has a transposase within a distance of $\sigma(1 + s_T)/2$ is converted to a transposed monomer, and the transposase is moved to the simulation volume boundary.

Placement adjacency. If a native monomer has a transposon adjacent to it, depending on the value of the placement adjacency α , it can preferentially get the next transposon as compared to a native monomer without any adjacent transposon. At $\alpha = 0$, all native monomers have equal insertion probability, independent of being adjacent to transposons or not. At $\alpha = 1$, only native monomers adjacent to transposons receive further insertions. In the transposase model, we perform a maximum likelihood estimation of the underlying α as α_{mle} by using the first 50 insertions from multiple independent repeat runs.

ACKNOWLEDGMENTS. This work was supported by the Helmholtz Program “Natural, Artificial, and Cognitive Information Processing”. The study was started by discussions with Changjing Zhuge and Jinzhi Lei during an exchange visit supported by the Sino-German Center for Research Promotion. Davide Michieletto advised on the implementation of conversion reactions in LAMMPS simulations and commented on the manuscript. The authors declare that they have no conflict of interest.

1. G Bourque, et al., Ten things you should know about transposable elements. *Genome Biol.* **19**, 1–12 (2018).
2. LM Payer, KH Burns, Transposable elements in human genetic disease. *Nat. Rev. Genet.* **20**, 760–772 (2019).
3. TK Starr, et al., A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323**, 1747–1750 (2009).
4. E Lee, et al., Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967–971 (2012).
5. S Solyom, et al., Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* **22**, 2328–2338 (2012).
6. LN van de Lagemaat, L Gagnier, P Medstrand, DL Mager, Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res.* **15**, 1243–1249 (2005).
7. M Seczynska, S Bloor, SM Cuesta, PJ Lehner, Genome surveillance by HUSH-mediated silencing of intronless mobile elements. *Nature* **601**, 440–445 (2022).
8. D Schmidt, et al., Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348 (2012).
9. S Patoori, S Barnada, M Trizzino, Young transposable elements rewired gene regulatory networks in human and chimpanzee hippocampal intermediate progenitors. *bioRxiv* p. 2021.11.24.469877 (2021).
10. B Piegu, et al., Doubling genome size without polyploidization: dynamics of retrotransposon-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome research* **16**, 1262–1269 (2006).
11. RK Slotkin, R Martienssen, Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285 (2007).
12. JW Jachowicz, et al., LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).

13. T Sultana, et al., The landscape of L1 retrotransposons in the human genome is shaped by pre-insertion sequence biases and post-insertion selection. *Mol. Cell* **74**, 555–570.e7 (2019).
14. B Lucic, et al., Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. *Nat. Commun.* **10**, 4059 (2019).
15. SA Isaacson, DM McQueen, CS Peskin, The influence of volume exclusion by chromatin on the time required to find specific DNA binding sites by diffusion. *Proc. Natl. Acad. Sci. USA* **108**, 3815–3820 (2011).
16. SM Görisch, M Wachsmuth, KF Tóth, P Lichter, K Rippe, Histone acetylation increases chromatin accessibility. *J. Cell Sci.* **118**, 5825–5834 (2005).
17. K Maeshima, et al., The physical size of transcription factors is key to transcriptional regulation in chromatin domains. *J. Physics: Condens. Matter* **27**, 064116 (2015).
18. M Raccaud, et al., Mitotic chromosome binding predicts transcription factor properties in interphase. *Nat. Commun.* **10**, 487 (2019).
19. D Michieletto, M Lusic, D Marenduzzo, E Orlandini, Physical principles of retroviral integration in the human genome. *Nat. Commun.* **10**, 1–11 (2019).
20. JD Buenostro, PG Giresi, LC Zaba, HY Chang, WJ Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
21. X Chen, et al., ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
22. RM Martin, MC Cardoso, Chromatin condensation modulates access and binding of nuclear proteins. *The FASEB J.* **24**, 1066–1072 (2010).
23. Y Shin, et al., Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **175**, 1481–1491 (2018).
24. JY Lu, et al., Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* **31**, 613–630 (2021).
25. K Kruse, et al., Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. *bioRxiv* p. 523712 (2019).
26. Y Zhang, et al., Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388 (2019).
27. J Wang, et al., MIR retrotransposon sequences provide insulators to the human genome. *Proc. Natl. Acad. Sci. USA* **112**, E4428–E4437 (2015).
28. T Cremer, et al., The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett.* **589**, 2931–2943 (2015).
29. A Monneron, B Bernhardt, Fine structural organization of the interphase nucleus in some mammalian cells. *J. Ultrastruct. Res.* **27**, 266–288 (1969).
30. PJ Verschure, I Van Der Kraan, EMM Manders, R Van Driel, Spatial relationship between transcription sites and chromosome territories. *The J. Cell Biol.* **147**, 13–24 (1999).
31. S Chambeyron, WA Bickmore, Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
32. T Tumber, G Sudlow, AS Belmont, Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J. Cell Biol.* **145**, 1341–1354 (1999).
33. T Tsukamoto, et al., Visualization of gene activity in living cells. *Nat. Cell Biol.* **2**, 871–878 (2000).
34. EV Volpi, et al., Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113**, 1565–1576 (2000).
35. RS Nozawa, et al., SAF-A regulates interphase chromosome structure through oligomerisation with chromatin-associated RNAs. *Cell* **169**, 1214–1227 (2017).
36. L Hilbert, et al., Transcription organizes euchromatin via microphase separation. *Nat. Commun.* **12**, 1360 (2021).
37. A Buckle, CA Brackley, S Boyle, D Marenduzzo, N Gilbert, Polymer simulations of heteromorphic chromatin predict the 3D folding of complex genomic loci. *Mol. Cell* **72**, 786–797 (2018).
38. CA Brackley, S Taylor, A Papantonis, PR Cook, D Marenduzzo, Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. USA* **110**, E3605–E3611 (2013).
39. KA Dill, S Bromberg, D Stigter, *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*. (Garland Science), (2010).
40. HD Ou, et al., ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357** (2017).
41. C Vieira, D Lepetit, S Dumont, C Biémont, Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.* **16**, 1251–1255 (1999).
42. V Mérel, M Boulesteix, M Fablet, C Vieira, Transposable elements in *Drosophila*. *Mob. DNA* **11**, 1–20 (2020).
43. M Falk, et al., Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).
44. H Fan, et al., The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes. *Genome Res.* **28**, 192–202 (2018).
45. Y Yin, et al., U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* **580**, 147–150 (2020).
46. LL Hall, JB Lawrence, RNA as a fundamental component of interphase chromosomes: Could repeats prove key? *Curr. Opin. Genet. Dev.* **37**, 137–147 (2016).
47. EM Hildebrand, J Dekker, Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem. Sci.* **45**, 385–396 (2020).
48. F Erdel, K Rippe, Formation of Chromatin Subcompartments by Phase Separation. *Biophys. J.* **114**, 2262–2270 (2018).
49. AG Larson, et al., Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
50. AR Strom, et al., Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
51. F Erdel, et al., Mouse heterochromatin adopts digital compaction states without showing hallmarks of HP1-driven liquid-liquid phase separation. *Mol. Cell* **78**, 236–249.e7 (2020).
52. JK Ryu, et al., Bridging-induced phase separation induced by cohesin SMC protein complexes.

560 *Sci. Adv.* **7**, eabe5905 (2021).

561 53. E Lieberman-Aiden, et al., Comprehensive mapping of long-range interactions reveals folding
562 principles of the human genome. *Science* **326**, 289–294 (2009).

563 54. SSP Rao, et al., A 3D map of the human genome at kilobase resolution reveals principles of
564 chromatin looping. *Cell* **159**, 1665–1680 (2014).

565 55. F Ochs, et al., Stabilization of chromatin topology safeguards genome integrity. *Nature* **574**,
566 571–574 (2019).

567 56. M Barbieri, et al., Complexity of chromatin folding is captured by the strings and binders switch
568 model. *Proc. Natl. Acad. Sci. USA* **109**, 16173–16178 (2012).

569 57. D Michieletto, D Coli, D Marenduzzo, E Orlandini, Nonequilibrium theory of epigenomic
570 microphase separation in the cell nucleus. *Phys. Rev. Lett.* **123**, 228101 (2019).

571 58. SK Ghosh, D Jost, How epigenome drives chromatin folding and dynamics, insights from
572 efficient coarse-grained models of chromosomes. *PLoS Comput. Biol.* **14**, 1–26 (2018).

573 59. A Bousios, HW Nützmann, D Buck, D Michieletto, Integrating transposable elements in the 3D
574 genome. *Mob. DNA* **11**, 1–10 (2020).

575 60. MV Neguembor, et al., Transcription-mediated supercoiling regulates genome folding and loop
576 formation. *Mol. Cell* **81**, 3065–3081.e12 (2021).

577 61. G Forte, D Michieletto, D Marenduzzo, E Orlandini, Investigating site-selection mechanisms of
578 retroviral integration in supercoiled DNA braids. *J. Royal Soc. Interface* **18**, 20210229 (2021).

579 62. R Kempfer, A Pombo, Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.*
580 **21**, 207–226 (2020).

581 63. JS Khurana, et al., Adaptation to P element transposon invasion in *Drosophila melanogaster*.
582 *Cell* **147**, 1551–1563 (2011).

583 64. C Ernst, DT Odom, C Kutter, The emergence of piRNAs against transposon invasion to
584 preserve mammalian genome integrity. *Nat. Commun.* **8**, 1–10 (2017).

585 65. RK Slotkin, et al., Epigenetic reprogramming and small rna silencing of transposable elements
586 in pollen. *Cell* **136**, 461–472 (2009).

587 66. A Bousios, BS Gaut, Mechanistic and evolutionary questions about epigenetic conflicts
588 between transposable elements and their plant hosts. *Curr. Opin. Plant Biol.* **30**, 123–133
589 (2016).

590 67. AP Thompson, et al., LAMMPS - a flexible simulation tool for particle-based materials modeling
591 at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).