**Giorgetti et al.**

**Supplementary Model Description**

**1. The polymer model**

We describe the chromatin fibre as an inextensible chain of beads (**Figure M1A**). The distance between consecutive beads is defined by the length *a*, whose numerical value (in nm) sets the spatial scale of the simulation. Beads are numbered from 1 to *N* from 5’ to 3’ along the genomic sequence; N is determined by the length of the simulated region and by the coarse graining level of its representation onto the chain of beads (see below).

System:Users:luca:Documents:ChrX:Draft:Supplementary_model:chromatin-1.pdf

**Figure M1.** A. Beads-on a string representation of the chromatin fibre. B. Spherical-well interaction potential between two beads as a function of their mutual distance.

Each bead *i* is allowed to interact with any other bead *j* along the chain through a spherical-well potential *Bij(dij)* that depends on their mutual 3D distance *dij* (**Figure M1B**). When the two beads are closer than the hard-core radius *rHC*, then *Bij=*+∞ (impenetrable hard core). If *rHC* *< dij < R*the two beads interact with energy *Bij,* which can be either negative or positive (corresponding to mutual attraction or repulsion, respectively). If *dij > R* the beads do not interact. Nearest-neighbour interactions between consecutive beads along the chain are not allowed (*Bi,i+1*=0*)*.

For any given conformation of the chain, the total energy of the fibre is given by the sum of the contributions of all pairs of beads, *E* = Σ*i<j Bij*. No explicit angular energy is used to increase the rigidity of the chain: allowing *Bij* to take positive values (especially between beads laying close along the chain, e.g. beads *i* and *i+2*) will ensure the emergence of a persistence length if this is needed to reproduce the experimental data.

From a biological point of view, this choice of potential corresponds to simulating an impenetrable chromatin fibre in which *short-range* interactions between distal genomic regions can take place (e.g. mediated by protein-protein interactions) when the corresponding parts of the fibre are sufficiently close to each other.

For the sake of comparing the simulation to the 5C contact maps, we further define two beads being *in contact* if *d < R*. In other words, two beads will be considered to be in contact (in the 5C sense) not necessarily when they touch each other (this happens at *d=rHC*), but when they are close enough to be able to interact. This is a crucial assumption as it allows comparing directly the simulations with 5C contact maps (cf. section 2 hereafter). Although there is no direct experimental justification for this, it is reasonable to assume that whenever two parts of the fibre are close enough in space to allow protein-protein interactions to take place, they will also be close enough to be crosslinked by paraformaldehyde and thus contribute to the observed 5C contacts.

System:Users:luca:Documents:ChrX:Draft:Supplementary_model:rf_size.pdf

**Figure M2.** Size distribution of restriction fragments in the 5C design of Nora et al, in the genomic region of TADs D (*Tsix*) and E (*Xist*).

**2. Mapping 5C data on the model polymer**

The 5C design of Nora et al. simultaneously interrogated nearly 250,000 possible chromosomal contacts around the X inactivation centre. Here we focus of the region of the topologically associating domains (TADs) harbouring the *Tsix* and *Xist* promoters (TAD D and E in Nora et al; mm9 chrX:100378306-101298738; see **Figure 2A** in the main text). In this genomic region, 124 forward (FOR) and 126 reverse (REV) primers anneal to alternate HindIII restriction fragments of variable size (**Figure M2**).

Since the average length of FOR and REV fragments in this region is 3078 bp, we chose to use 3000-bp beads, resulting in polymers of *N*=108 beads in the case of TAD D alone (chrX:100378306-100699670, total length 321’364 bp) and *N*=307 beads in the case of TAD D+E (total length 92’0432 bp). We believe that this represents an acceptable trade off between under- and oversampling 5C data: 35% of restriction fragments are longer then 3000 bp and will be mapped on multiple beads, while the remaining 65% will contribute together with their nearest neighbour to define the expected contacts of single beads.

To simulate 5C experiments, the uneven sampling of the genomic region provided by restriction fragments of different length must be mapped onto the even sampling provided by equally spaced beads. This is performed as follows (for an example of the alignment, see **Supplementary Figure 2B**):

1. The 5’ end of the first bead in the chain is positioned at the 5’ end of the first restriction fragment in the region of interest (in our case, REV\_489 in the design of Nora et al).
2. Each restriction fragment in the region is assigned two indices *i* and *j* (*i*,*j*=1…*N*) corresponding to the beads to which its 5’ and 3’ ends overlap.
3. 5C counts corresponding to each pair of restriction fragments are assigned to the corresponding pairs of indices (I,j) and (h,k), for example:



1. If two or more consecutive FOR or REV restriction fragments map to the same bead, their contributions are summed (<10% of all interactions).
2. 5C counts of two experimental replicates are averaged for each pair of restriction fragments and their standard deviation is taken as a measure of experimental uncertainty. A lower cut-off on experimental uncertainty was set at 1% of the average to avoid dispersing computational time in over-optimizing the agreement between simulated and experimental contacts (see Section 4 below).

**3. Thermodynamic interpretation of 5C data**

We assume that the 5C count associated to two segments spanning from *i* to *j* and from *h* to *k* along the fibre is proportional to the probability that the two segments are in contact throughout the cell population,

(1)

where *Z* is a normalization constant. In our notation, is the number of sequencing tags in the 5C experiment (normalized by the total number of reads), while is the actual probability of observing the corresponding contact (the definition of contact between two segments is given in Section 4 below). *Z* essentially defines the 5C count corresponding to contact probability equal to one, i.e. the value of above which two parts of the fibre are assumed to be in contact in 100% of cells.

Since *Z* is the key parameter controlling the conversion of 5C counts into contact probabilities, we set out to estimate its numerical value from the experimental data. To do this, we reasoned that adjacent restriction fragments (represented for example by pairs of beads (*i,j*) and (*h,k*) with *j*=*h*±1) can be considered to be in contact in 100% of cells, and thus that their average normalized 5C count (equal to 3097) could provide a reasonable estimate for the numerical value of *Z.* Further support for this choice came from the fact that the distribution of counts from non-adjacent fragments (expected to be in contact in <100% of cells, i.e. with counts smaller than *Z*) drops to zero around 3000 (**Figure M3**), in excellent agreement with our choice for the numerical value of *Z*.

Apart from the determination of the numerical value of *Z,* adjacent restriction fragments were not further used to constrain the model during structural reconstruction (see below).

System:Users:luca:Documents:ChrX:Draft:Supplementary_model:histo_counts.ps.pdf

**Figure M3.** Histogram of 5C counts from adjacent (i.e., *j=h+1,* *j=h-1*, *k=i+1* or *k=i-1*) and non-adjacent restriction fragments (in red).

**4. Optimization of interaction potentials**

For several values of hard core and interaction radii and , we have carried out an iterative Monte Carlo simulation to optimize the matrix of interaction energies *Bij*. Optimal values of and were selected *a posteriori* after comparison with DNA FISH (see **Supplementary Figure 2**).

In each simulation, the elements of the symmetric matrix of potentials *Bij* were determined by minimizing the *χ2* between the probabilities (obtained from the experimental 5C counts through Eq. 1) and the equilibrium contact probabilities *p*calculated from the model:

, (2)

where =1 if a 5C count is available for the associated quartet representing two segments along the chain (corresponding to a pair of restriction fragments), and zero otherwise [[1]](#footnote-1). is the experimental precision of *p(i,j;h,k)*, defined as the standard deviation between the counts in the duplicate 5C experiments, normalized by *Z*. Total numbers of parameters optimized in each Monte Carlo run is 8758 for TAD D alone and 39006 for TAD D+E.

Macintosh HD:Users:guido:Documents:my_articles:chromatin:flow.pdf

**Figure M4.** Scheme of the algorithm employed to optimize sets of interaction parameters *Bij.*

At the end of each Monte Carlo conformational sampling, the optimal values of *Bij* were determined with the help of an iterative Monte Carlo scheme (**Figure M4**). The starting point of the algorithm is the choice of an initial set of parameters *Bij*. This specific choice is not critical since they will be iteratively updated until convergence; trivially however, adequate initial choices will result in smaller numbers of iterations needed to reach convergence. A suitable choice resulted to be

, (3)

where *L=(j-k+1)(i-h+1)* is the number of pairs of monomers that can potentially interact across the two segments, *λ=h-j* is the length of the chain that separates the two segments and *B0* is the initial energy scale in units of *kBT*, *kB* being the Boltzmann’s constant. If *p5C(i,j; h,k)*=1 we set *Bij*=-2*B0*. We found *B0*=0.2 to be a good starting point. In the following we will express all energies in units of *kBT* and set *kBT*=1, for simplicity.

A Monte Carlo sampling of the conformation space is carried out to generate 5000 conformations that are representative of the system (i.e., they describe the thermodynamics of the system in the canonical ensemble).

Each simulation lasts for *NMC*=5⋅108 steps. Every *Nrec*=5⋅103 Monte Carlo steps a conformation of the chromatin fibre is recorded. The segments between *i* and *j* and between *k* and *l* are defined to be in contact if *at least one bead* in the range *i…j* is within distance *R* from a bead in the range *k…l* (**Figure M5**),where *R* is the contact radius that was defined in Section 1. This choice is motivated by the fact that whenever two restriction fragments are crosslinked and this interaction is detected in 5C, it is impossible to determine which parts of the two restriction fragments were actually in physical proximity.

System:Users:luca:Documents:ChrX:Draft:Supplementary_FIgures:contact_scheme.eps

**Figure M5.** Contact probability in the polymer model. Two restriction fragments represented by bead sequences *i…j* and *h…k* are considered to be in contact whenever two beads and

We define the function *Δ*s(*i,j; h,k*) on each pair of restriction fragments of each recorded structure *s*, equal to 1 if anyof the beads belonging to the two segments are in contact in that particular structure and zero otherwise:

(4)

where if and 0 otherwise. At the end of the Monte Carlo sampling, the contact probability of every pair of restriction fragments *p*(*i,j; h,k*) is calculated as

(5)

and all the contact probabilities along the chain are compared to the corresponding experimental values via the *χ2* calculated as in Eq. 2.

The recorded conformations are also used to evaluate which changes must be made on the potentials *Bij* to decrease the *χ2*. First, a random minimization is carried out where an element *Bij* is selected among those belong to segments whose 5C count is available, and updated summing a random number between -0.5 and 0.5 (in kBT units). The updated values of contact probabilities *p*(*i,j; h,k*) are then calculated by reweighting the recorded conformations as done in Norgaard *et al*. Biophys. J. 94, 182 (2008):

, [S5]

where

, [S6]

and *Esold* and *Esnew* are the energy of the chain before and after updating *Bij*, respectively. The new *χ2* is calculated and compared with its value before updating *Bij*. If *χ2* decreases, the updated values are maintained as the new *Bij*, otherwise they are discarded and another random update is attempted. This scheme is applied for *Nmin=103* iterations.

After a consistent change in the interaction matrix *Bij*, the set of recorded conformations cannot be regarded any longer as representative of the new system, and it is not possible to simply reweigh the conformations with the new potentials: a new Monte Carlo sampling of the conformation space is carried out to record a new set of *Nrec* conformations representative of the new potentials. This procedure is iterated until convergence of the *χ2* to a minimum (**Figure M6**). The final matrix *Bij* is then regarded as describing the set of interactions among segments of the chromatin fibre, and the associated set of conformations are regarded as representative of its equilibrium structural fluctuations.

System:Users:luca:Documents:ChrX:Draft:Supplementary_model:chi2_vs_iterations.pdf

**Figure M6.** Optimization of *χ2* through Monte Carlo sampling and minimization of interaction potentials *Bij*. Representative simulations with different values of *R* and *rHC* are shown; black and red curves are the replicate simulations with optimal values of *R* and *rHC* shown in Supplementary Figure 2D.

1. Given that 5C detects crosslinked FOR and REV fragments only, *p5C(i,j)* is available only when bead *i* and *j* belong to a FOR and a REV restriction fragment respectively, or vice versa. By constraining *Bij*=0 whenever this condition is not fulfilled we reduce the number of parameters in the optimization procedure. [↑](#footnote-ref-1)