PNAS
www.pnas.org

Supplementary Information for

Anomalous Diffusion, Spatial Coherence, and Viscoelasticity from the Energy Landscape of Human Chromosomes

Michele Di Pierro[1,*,+], Davit A. Potoyan[2,*], Peter G. Wolynes[1,3,4,5], José N. Onuchic[1,3,4,5]

[1] Center for Theoretical Biological Physics, Rice University. Houston, Texas
[2] Department of Chemistry, Iowa State University, Ames, Iowa
[3] Department of Chemistry, Rice University, Houston, Texas
[4] Department of Physics & Astronomy, Rice University, Houston, Texas
[5] Department of Biosciences, Rice University, Houston, Texas
[*] Equal Contribution

[+] Correspondence to MicheleDiPierro@rice.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1806297115

Methods

We performed molecular dynamics simulations for a system composed of one copy of the interphase human chromosome 17 interacting with one copy of the interphase human chromosome 18. The chromosomes were modeled using the Minimal Chromatin Model described below. Each bead in the Minimal Chromatin Model represents a genomic segment spanning 50 Kb of DNA resulting in 1626 beads for chromosome 17 and 1564 for chromosome 18.

Hi-C data from Rao *et al.* (*1*) were used in (*2*) to train the empirical energy function of MiChroM in order to correctly characterize the structural features of human lymphoblastoid cells in interphase (cell line GM12878). Chromatin types annotations and loops locations, which are input to MiChroM, were also obtained for these cells from reference (*1*). The Gene Expression Omnibus (GEO) accession number for the data sets used for the training is GSE63525.

Minimal Chromatin Model (MiChroM)

All simulations methods in this manuscript are the same as those used in references (*2, 3*).

Minimal Chromatin Model Energy Function

The MiChroM energy function is:

$$U_{MiChroM}(\vec{r}) = U_{HP}(\vec{r}) + \sum_{\substack{k \geq l \\ k,l \in \text{Types}}} \alpha_{kl} \sum_{\substack{i \in \{\text{Loci of Type } k\} \\ j \in \{\text{Loci of Type } l\}}} f(r_{ij}) + \chi \cdot \sum_{(i,j) \in \{\text{Loops Sites}\}} f(r_{ij}) + \sum_{d=3}^{500} \gamma(d) \sum_{i} f(r_{i,\,i+d})$$

with the contact function:

$$f(r_{ij}) = \frac{1}{2}\left(1 + \tanh\left[\mu\left(r_c - r_{ij}\right)\right]\right)$$

and with parameters $\mu = 3.22$ and $r_c = 1.78$.

Parameter Set

The parameters α's governing the type-to-type interactions are:

	A1	A2	B1	B2	B3	NA
A1	-0.268028	-0.274604	-0.262513	-0.258880	-0.266760	-0.225646
A2	-0.274604	-0.299261	-0.286952	-0.281154	-0.301320	-0.245080
B1	-0.262513	-0.286952	-0.342020	-0.321726	-0.336630	-0.209919
B2	-0.258880	-0.281154	-0.321726	-0.330443	-0.329350	-0.282536
B3	-0.266760	-0.301320	-0.336630	-0.329350	-0.341230	-0.349490
NA	-0.225646	-0.245080	-0.209919	-0.282536	-0.349490	-0.255994

The parameter χ governing the loop interactions is equal to -1.612990.

Ideal Chromosome Term

The Ideal Chromosome Potential is:

$$\gamma(d) = \frac{\gamma_1}{\log(d)} + \frac{\gamma_2}{d} + \frac{\gamma_3}{d^2}$$

with parameters $\gamma_1 = -0.030$, $\gamma_2 = -0.351$, $\gamma_3 = -3.727$.

Homopolymer Model

The homo-polymer potential $U_{HP}(\vec{r})$ models a generic polymer and consists of the following five terms, U_{FENE}, U_{Angle}, U_{hc}, U_{sc} and U_c.

$$U_{HP}(\vec{r}) = \sum_{i \in \{\text{Loci}\}} U_{FENE}(r_{i,i+1}) + \sum_{i \in \{\text{Loci}\}} U_{hc}(r_{i,i+1}) + \sum_{i \in \{Angles\}} U_{Angle}(\theta_i) + \sum_{\substack{i,j \in \{\text{Loci}\} \\ j > i+2}} U_{sc}(r_{i,j}) + \sum_{i \in \{\text{Loci}\}} U_c(\vec{r}_i)$$

U_{FENE} (Finite Extensible Nonlinear Elastic potential) is the bonding potential applied between two consecutive monomers:

$$U_{FENE}(r_{i,j}) = \begin{cases} -\frac{1}{2} k_b R_0^2 \ln\left[1 - \left(\frac{r_{i,j}}{R_0}\right)^2\right] & \text{for } r_{i,j} \le R_0 \\ 0 & \text{for } r_{i,j} > R_0 \end{cases}$$

A hard-core repulsive potential

$$U_{hc}\left(r_{i,j}\right)=\begin{cases} 4\varepsilon\left[\left(\dfrac{\sigma}{r_{i,j}}\right)^{12}-\left(\dfrac{\sigma}{r_{i,j}}\right)^{6}+\dfrac{1}{4}\right] & \text{for } r_{i,j}\le\sigma 2^{\frac{1}{6}} \\ 0 & \text{for } r_{i,j}>\sigma 2^{\frac{1}{6}} \end{cases}$$

is added between bonded monomers to avoid overlap.
A three-body term is applied to three consecutive monomers in the following form

$$U_{Angle}\left(\theta_i\right)=k_a\left[1-\cos\left(\theta_i-\theta_0\right)\right]$$

where θ_i is the angle defined by the two vectors $\vec{r}_{i,i+1}$ and $\vec{r}_{i,i-1}$.
All non-bonded pairs interacts through a soft-core repulsive interaction

$$U_{sc}\left(r_{i,j}\right)=\begin{cases} \dfrac{1}{2}E_{cut}\left[1+\tanh\left(\dfrac{2U_{LJ}\left(r_{i,j}\right)}{E_{cut}}-1\right)\right] & r_{i,j}<r_0 \\ U_{LJ}\left(r_{i,j}\right) & r_0\le r_{i,j}\le\sigma 2^{\frac{1}{6}} \\ 0 & r_{i,j}>\sigma 2^{\frac{1}{6}} \end{cases}$$

The Lennard-Jones potential $U_{LJ}\left(r_{i,j}\right)=4\varepsilon\left[\left(\dfrac{\sigma}{r_{i,j}}\right)^{12}-\left(\dfrac{\sigma}{r_{i,j}}\right)^{6}+\dfrac{1}{4}\right]$ is capped off at a finite upper magnitude, allowing for chain crossing at finite energetic cost. r_0 is chosen as the distance at which $U_{LJ}\left(r_{i,j}\right)=\dfrac{1}{2}E_{cut}$. Allowing occasional chain crossing is essential to model the action of topoisomerases.
The potential U_c restricts the chromosome in a spherical region. The spherical wall is included to mimic a similar confinement experienced by chromosomes inside the cell. Each monomer *i* of the chromosome interacts with its nearest point on the wall $\vec{r}_{np}$ through the potential $U_{hc}\left(r_{i,np}\right)$.

Molecular Dynamics Simulations

First, we condense the polymer from an extended configuration initialized as a straight line. To condense the polymers, we perform $2\text{x}10^4$ step MD simulation under the potential energy function

$$U_{Eq}(\vec{r}) = \sum_{i\in\{\text{Loci}\}} U_{FENE}(r_{i,i+1}) + \sum_{i\in\{\text{Loci}\}} U_{hc}(r_{i,i+1}) + \sum_{i\in\{Angles\}} U_{Angle}(\theta_i) + \sum_{\substack{i,j\in\{\text{Loci}\} \\ j>i+2}} U_{sc}(r_{i,j}) + \frac{1}{2}K_{Eq}\left(R_g - R_g^0\right)^2$$

which is the homopolymer potential with an additional harmonic bias on the radius of gyration R_g. We set $K_{Eq} = 200\varepsilon/\sigma^2$ and $R_g^0 = 1$. There is no spherical confinement in this phase of the simulation. Then, from these condensed polymer configurations, we perform 20 million steps of equilibration with the potential energy function $U_{HP}(\vec{r})$, which now also includes the confinement potential. The radius of the confinement potential was set to reproduce a volume ratio of 0.1.

All chromosome simulations were performed using the molecular dynamics package LAMMPS (*4*). In reduced units the simulation parameters that were used are:

$$k_a = 2\varepsilon \qquad k_b = \frac{30\varepsilon}{\sigma^2} \qquad E_{cut} = 4\varepsilon \qquad \varepsilon = K_B T$$
$$R_0 = 1.5\sigma \qquad \sigma = 1 \qquad \theta_0 = \pi$$

Simulations were maintained at a constant temperature $T = 1.0$ via Langevin dynamics (*5*) with a damping time of 10.0τ, where τ is the unit of time. Using Einstein's relation this damping time results in a diffusion coefficient of 10 σ^2/τ for the free bead.

A time step $\Delta t = 0.01\tau$ was used for the simulation. The system was simulated for approximately 25 x 10^6 time steps. Configurations were recorded each 10 time steps (i.e. each $0.1\,\tau$) leading to a total of 2.5 x 10^6 configurations sampled. The total angular momentum was restrained to be zero to eliminate rigid rotations of the system. An equilibration of 10^6 time steps at high temperature (T=10) was performed before starting sampling chromosome conformations.

Physical Units

The unit of length in simulations can be calibrated by using available FISH data of human lymphoblastoid cells as was done in (*3*). Using this information we found that the unit of length σ in our model corresponds to 0.165μm, meaning that one bead has a radius of about 825Å. With this calibration, the chromatin density in our simulations is 0.002

bp/nm^3, which in is excellent agreement with previously reported estimates (*6*), especially when considering the variability in size between cell types and even within a homogeneous cell population.

Using the Einstein-Stokes relation we calculate the diffusion coefficient of a bead $D = K_B T / 6\pi\eta r$ as that of a sphere of radius $r = 0.0825 \mu m$ immersed in water at 298.15K. Using the viscosity of water $\eta = 8.9$ x 10^{-4} Pa·s we obtain the diffusion coefficient of 2.97 μm^2/s. Comparing this latter estimate for the diffusion coefficient of the free bead with the previous estimate of the same quantity in reduced units ($D = 10\sigma^2 / \tau$) we obtain that the unit of time τ in our simulations corresponds to approximately 0.1 seconds. The time step used in the simulations then corresponds to $\Delta t \sim 1 ms$.

Rouse mode analysis of chromosomal chains

This section is devoted to the quantitative determination of Rouse modes from the molecular dynamics simulation data. To this end we start from the known exact solution of the Rouse model for simple Gaussian chains. The effective Hamiltonian for a Gaussian chain is:

$$H(r^N) = \frac{3k_B T}{2\sigma^2} \sum_{i=1}^{N-1} (\vec{r}_{i+1} - \vec{r}_i)^2$$

Where $\vec{r}_i$ are positions of monomers connected by springs with equilibrium bond length σ. The entire chain is embedded in thermal bath with temperature T, which results in the following equation for each monomer:

$$\frac{d}{dt}\vec{r}_i = -\zeta \frac{\partial H}{\partial r_i} + \eta_i(t)$$

Where η_i are stochastic displacements, which are related to the temperature of the heat bath via fluctuation dissipation theorem. Fourier transformation of the monomer coordinates yields the Rouse modes:

$$\vec{X}_p(t) = \left(\frac{2}{N}\right)^{1/2} \sum_{i=1}^{N-1} \vec{r}_i \cos\left[\frac{p\pi}{N}(i - 1/2)\right], \quad p = 0,1,2, \ldots N-1$$

Using this definition of Rouse modes, one can compute the same quantity for the MiChroM chains using simulation data. For the chains obeying Rouse dynamics the following exact result holds:

$$\langle \vec{X}_p(t) \vec{X}_p(0) \rangle = \langle \overrightarrow{X^2}_p \rangle \exp\left(-\frac{t}{\tau_p}\right)$$

Where the amplitude is given by $\langle \overrightarrow{X^2}_p \rangle = A \frac{N}{p^2}$ and relaxation times $\tau_p = B \left(\frac{N}{p}\right)^2$ with A and B being constants. The angular brackets denote ensemble average. Both amplitude and relaxation time show inverse quadratic dependence on mode number. The rouse modes extracted from MiChroM simulation data fit Rouse model imperfectly by showing significant deviations from theoretical predictions for the long wavelength modes N/p>>1. In particular the autocorrelation function of Rouse modes for the MiChroM chains is better fitted by stretched exponential $\langle \vec{X}_p(t)\vec{X}_p(0) \rangle = C \exp\left(-\frac{t}{\tau_p}\right)^{\beta_p}$ with varying exponents β_p. To make connections with rouse mode relaxation times we compute effective relaxation times of MiChrom chain $\tau_p^{eff} = \int_0^\infty \exp\left(-\frac{t}{\tau_p}\right)^{\beta_p} = \frac{\tau_p}{\beta_p}\Gamma(\beta_p^{-1})$ where $\Gamma(x)$ stands for gamma function. The effective relaxation times exhibit two different power law scaling: ~$p^{-0.5}$ for first few modes (p=1-15) and ~$p^{-1.7}$ for the rest. Using p=15 we can estimate the size and time scale associated with the power law crossover. Length scale corresponding to crossover is L=(N/p)σ = 18 μm and time scale $\tau_{cross} = \tau_{15}^{eff} = 0.3$ s.

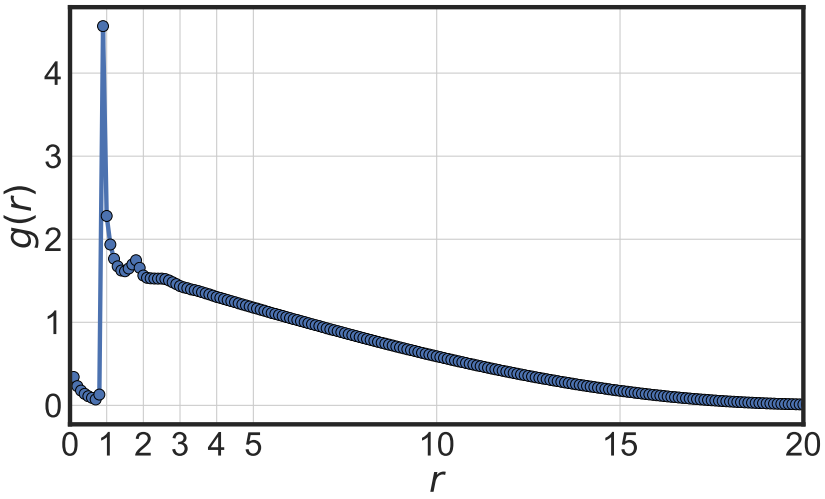

Figure S1: Radial distribution function of chromosomal loci.

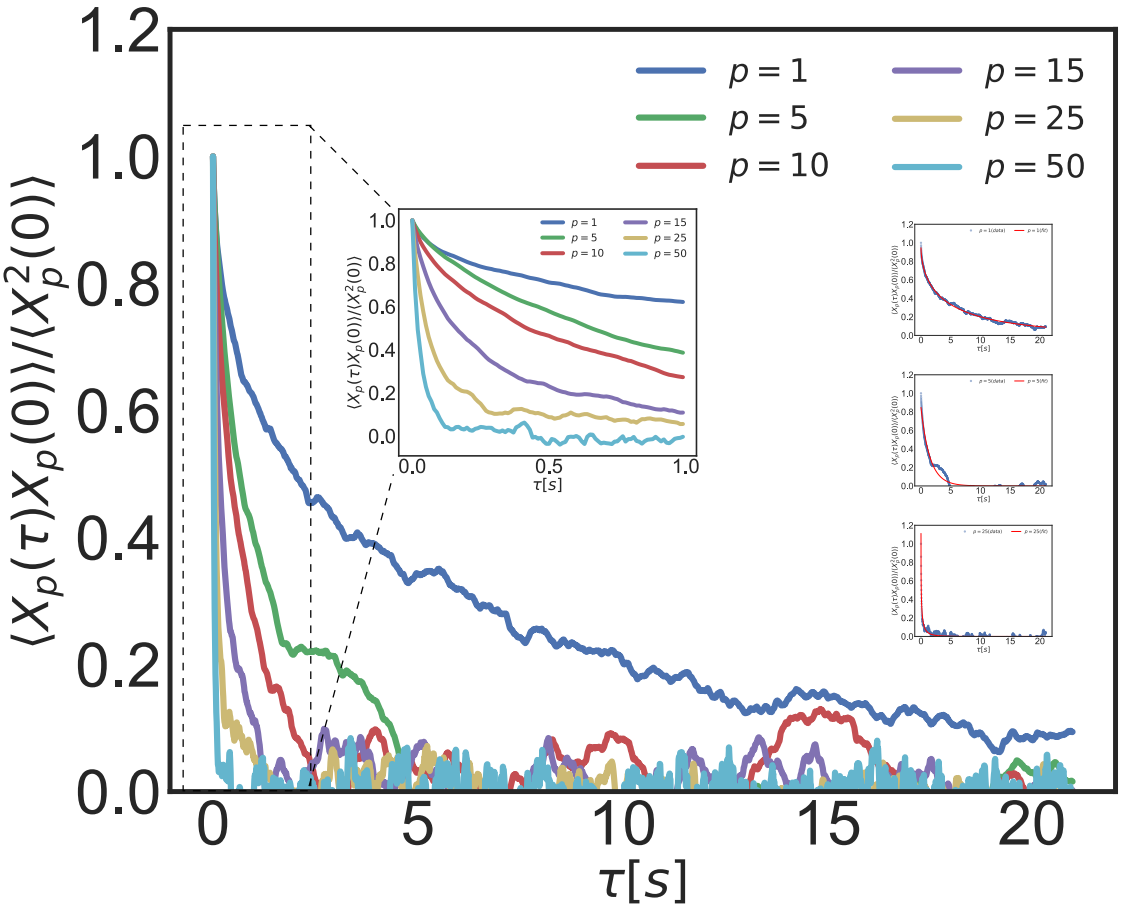

Figure S2: Autocorrelation function of Rouse modes. Inset on the left shows the short time behavior of Rouse mode autocorrelation functions. Inset on the right shows the stretched exponential fits to the Rouse mode autocorrelation functions for p=1, p=5 and p=25.

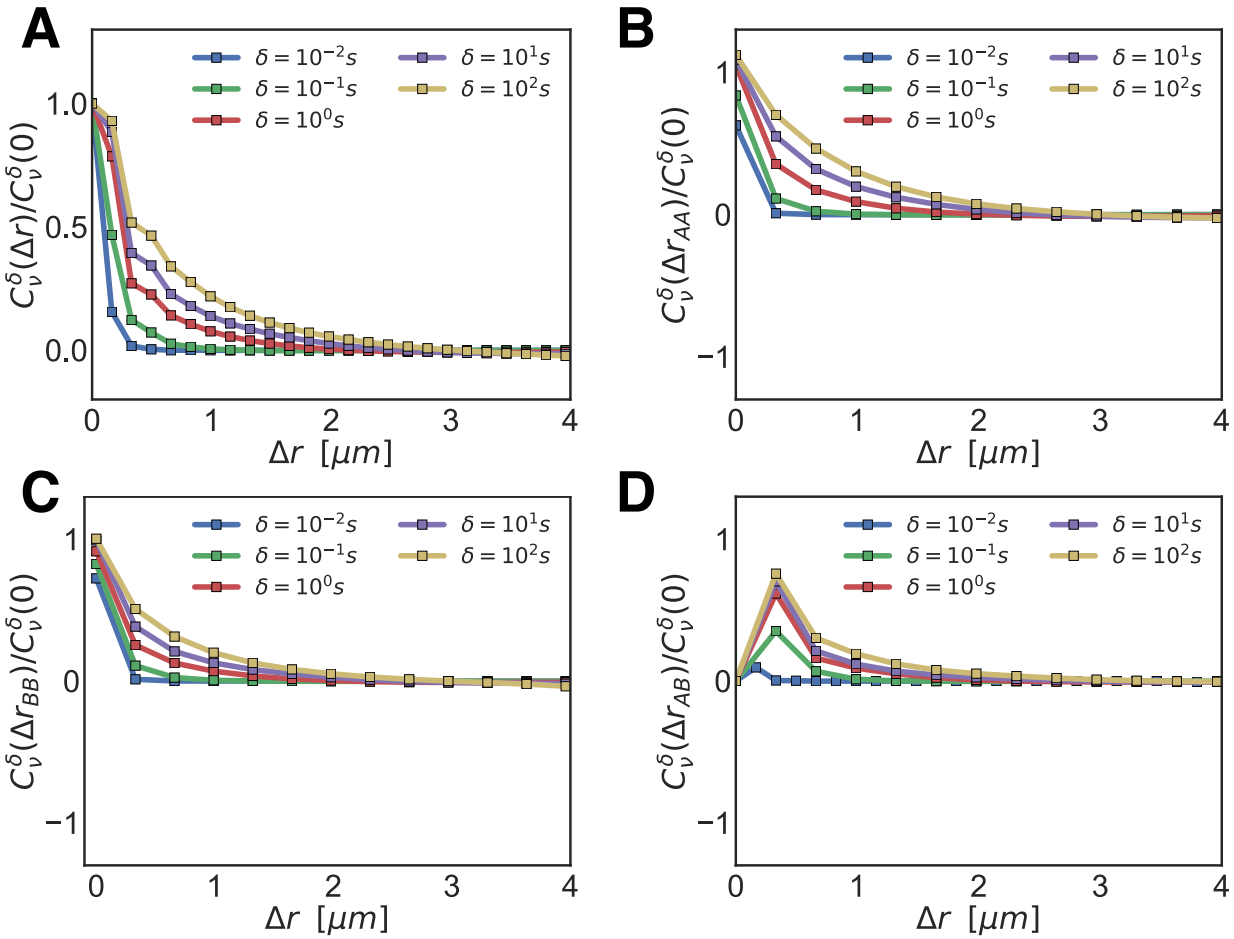

Figure S3: Spatial autocorrelation functions as a function of distance in space computed for beads of (A) All types (B) Only between A and A types (C) Only between B and B types (D) only between A and B types.

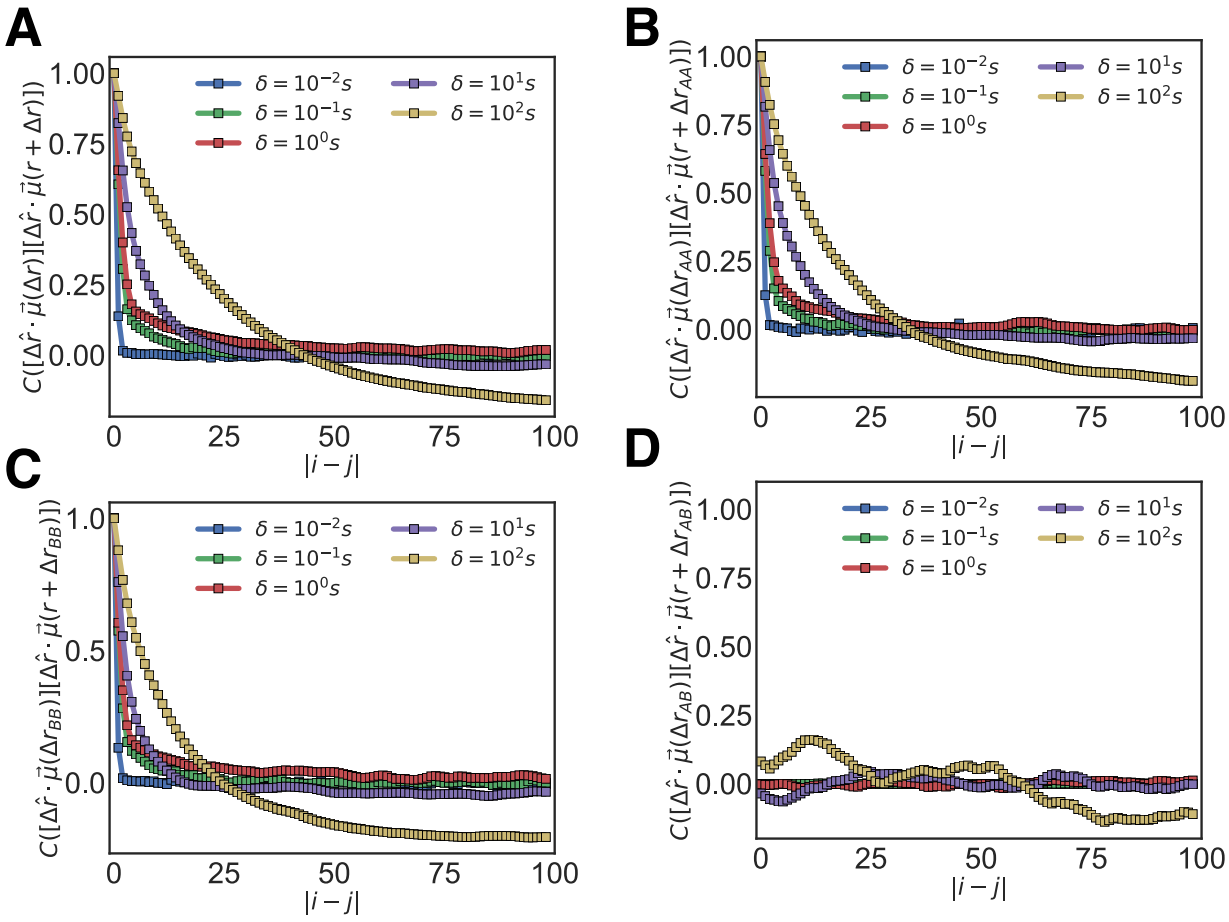

Figure S4: Flow autocorrelation functions as a function of genomic distance computed for beads of (A) All types (B) Only between A and A types (C) Only between B and B types (D) only between A and B types.

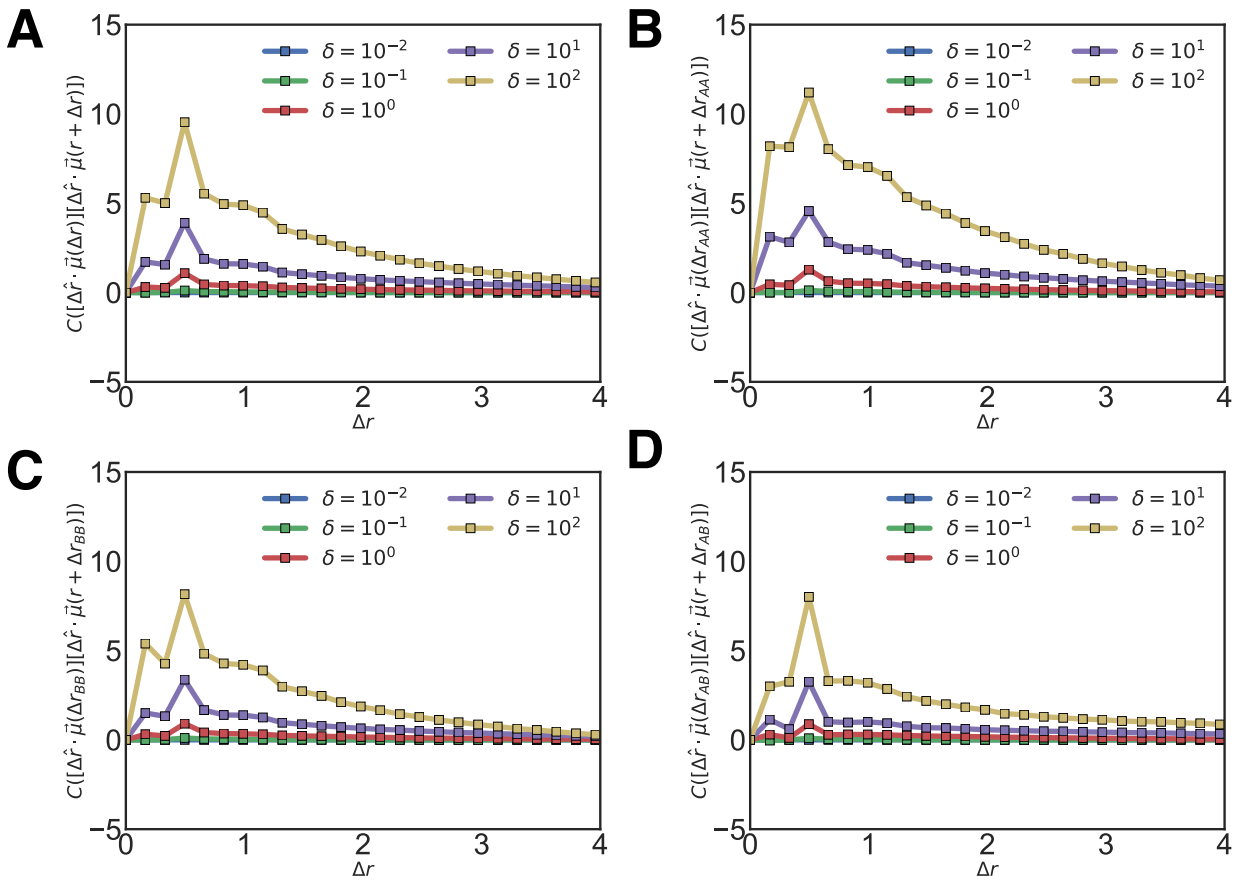

Figure S5: Flow autocorrelation functions as a function of distance in space computed for beads of (A) All types (B) Only between A and A types (C) Only between B and B types (D) only between A and B types.

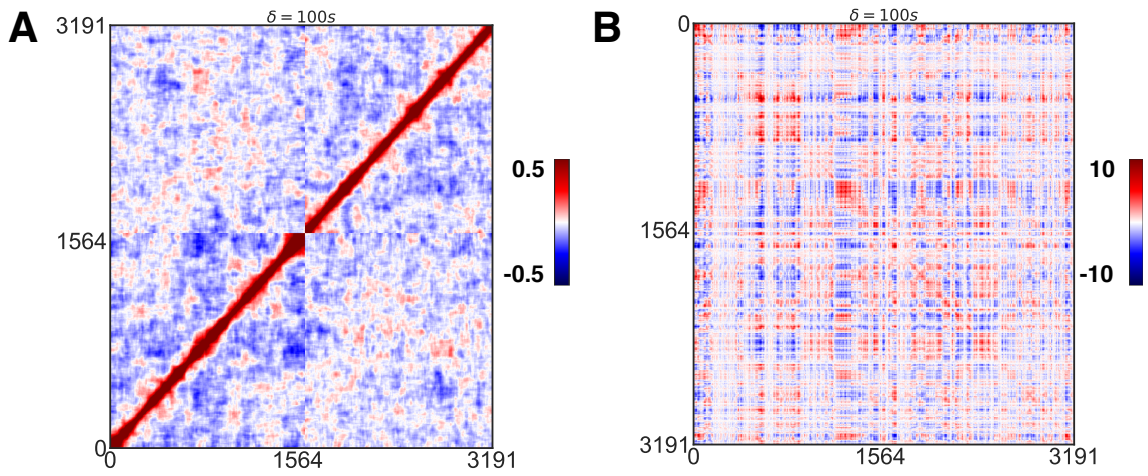

Figure S6: (A) Ensemble averaged displacement map for 100s time displacement (B) Single snapshot displacement map for 100s time displacement

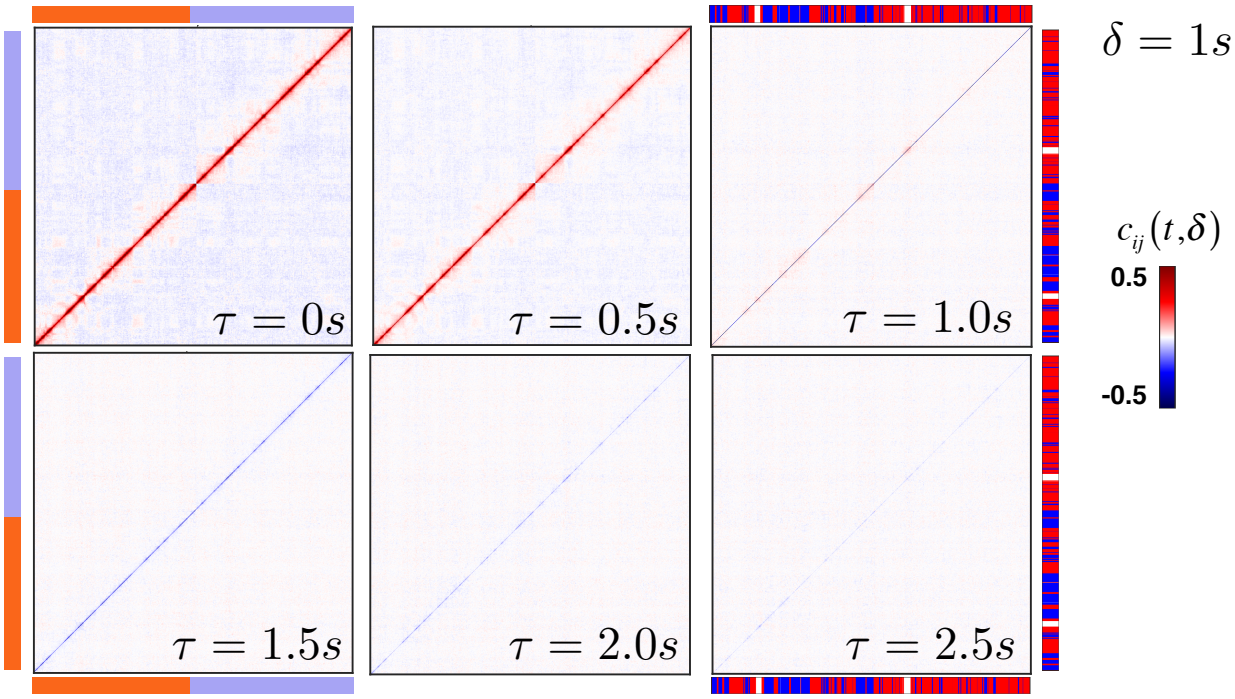

Figure S7: Average velocity-velocity autocorrelation maps as a function of time lag.

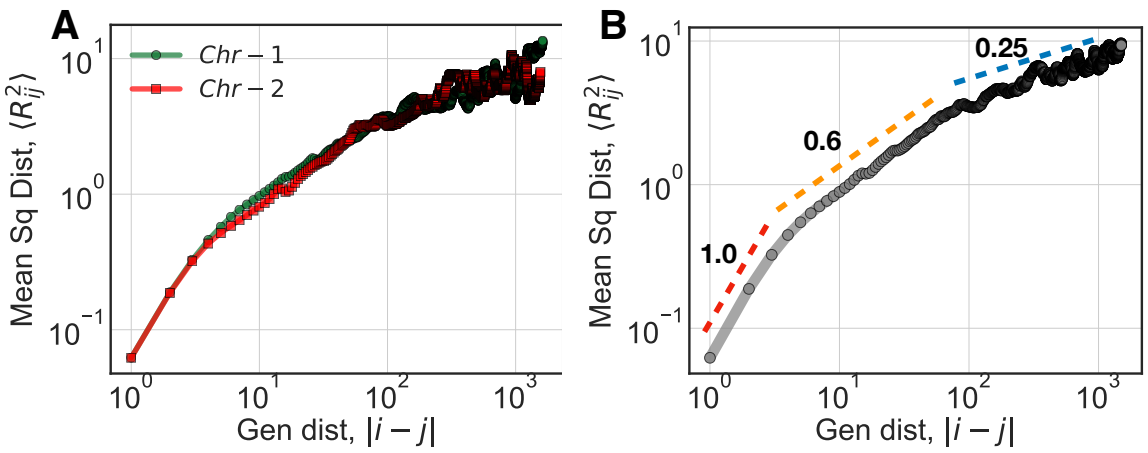

Figure S8: Behavior of the mean-square displacement as a function of the genomic distance. (A) Mean square distance between loci *i* and *j* of individual chromosomes (shown in green and red) as a function of their genomic distance. (B) Aggregated data from both chromosomes are used to calculate the apparent scaling exponent for the genomic distances 1 to 10 (exponent 1.0), 10 to 100 (exponent 0.6), and 100 to end of chains (exponent 0.25). At long length scales, heterogeneity is particularly evident; however, despite the noise introduced by heterogeneity, it is possible to observe a change in the apparent scaling exponent around genomic distances of about 100 beads, or about the size of individual DADs. This last finding supports the analysis of the p-modes scaling reported in the main text.

References

1. S. S. P. Rao *et al.*, A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665-1680 (2014).
2. M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 12168-12173 (2016).
3. M. Di Pierro, R. R. Cheng, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proceedings of the National Academy of Sciences*, 201714980 (2017).
4. S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *Journal of Computational Physics* **117**, 1-19 (1995).
5. T. Schneider, E. Stoll, Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **17**, 1302-1322 (1978).
6. A. Rosa, R. Everaers, Structure and Dynamics of Interphase Chromosomes. *Plos Computational Biology* **4**, (2008).