A Random Loop Model for Chromatin Organization

Manfred Bohn

Research Training Group "Simulational Methods in Physics"
Institute of Theoretical Physics
University of Heidelberg

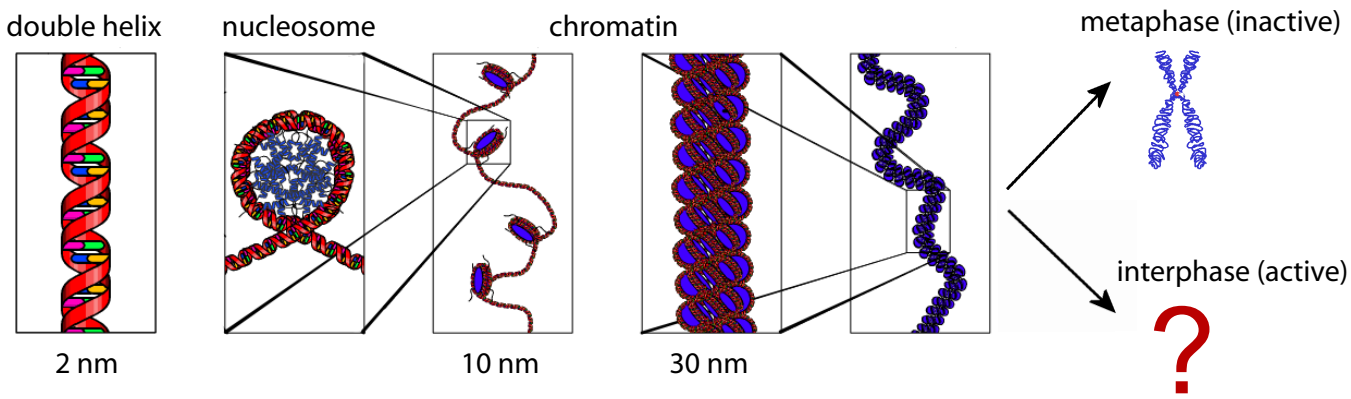
DELTA'07 meeting, 2007

M. Bohn and D. W. Heermann, *Phys. Rev. E* **76**, 2007

Outline

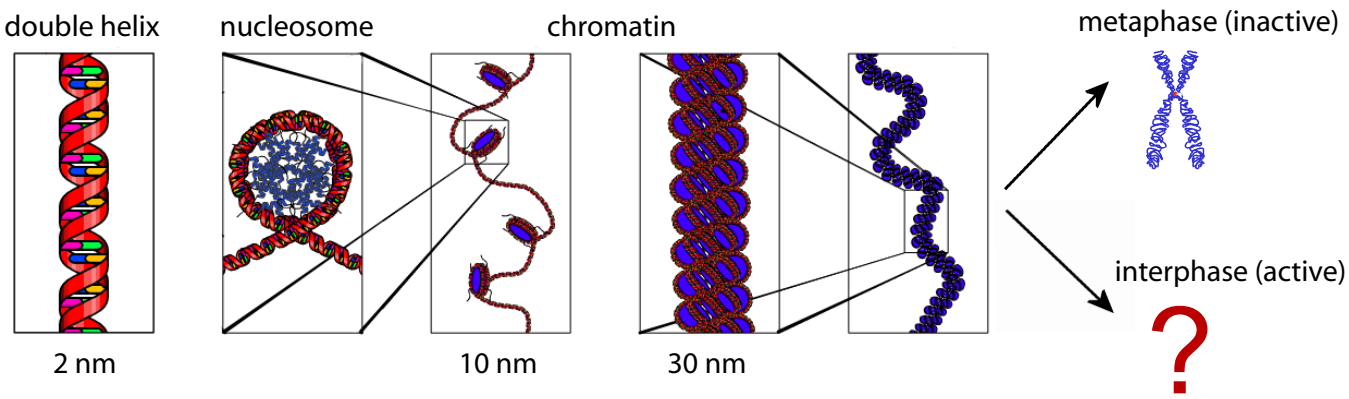
Problem:

- In human cells 2 m of ds-DNA have to be packed into a nucleus of 10 μm diameter.
- The first stages of compaction are well-known



Problem:

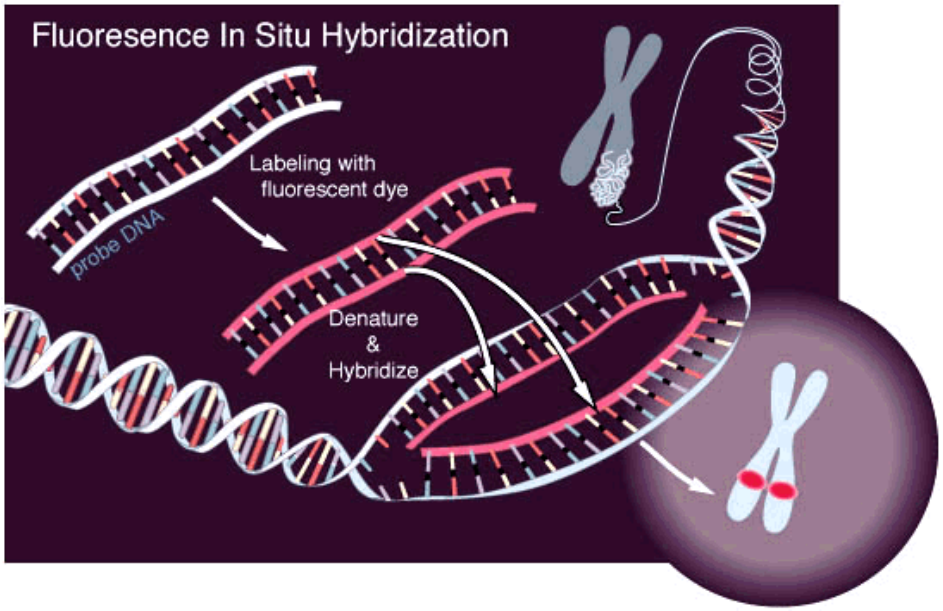
- In human cells 2 m of ds-DNA have to be packed into a nucleus of 10 μm diameter.
- The first stages of compaction are well-known

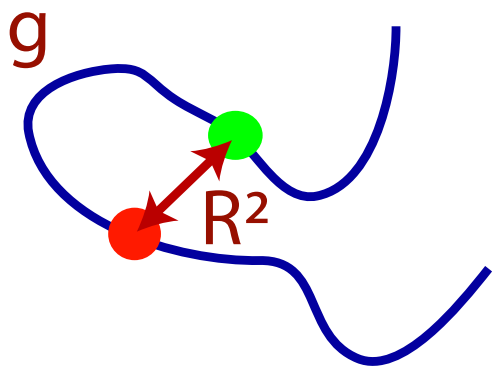


- The higher-order structure during interphase is widely unknown. Is the chromatin fibre just a random coil or is it somehow organized?
- How is the structure connected to the complex functionality of the nucleus?

Biological Background – Experimental approaches

- limitations of light microscopy $\rightarrow$ indirect approaches
- *FISH* (flourescence in situ hybridization)





- measurement of mean physical distance $\langle R^2 \rangle$ between two *FISH*-markers of known genomic distance g gives information about folding motifs and can be compared to polymer models

Polymer models

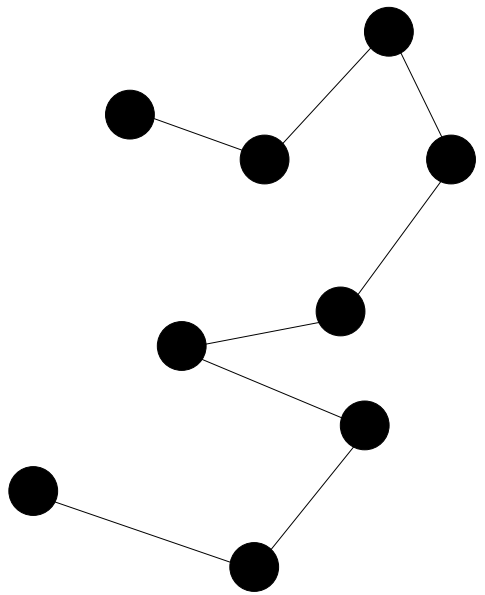
Long macromolecules often can be described by simple polymer models. On a scale much larger than a single molecule the structural details and stiffness can be neglected.

basic idea:

- divide the polymer into N equal subunits
- the subunits (beads) are connected by linkers, e.g.

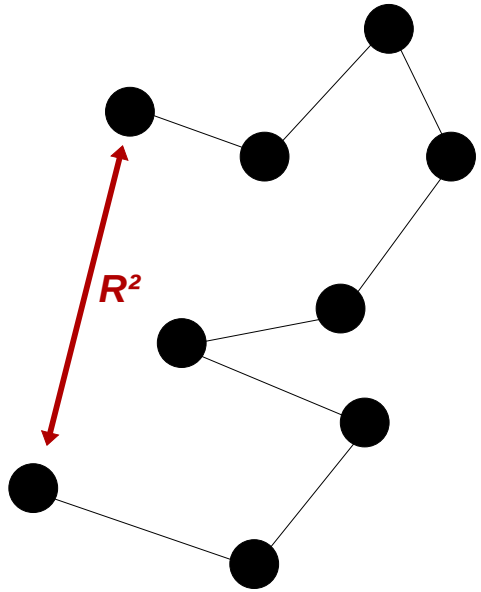
$$U = \sum_{i=1}^{N} \frac{1}{2} \kappa \parallel \mathbf{x}_i - \mathbf{x}_{i-1} \parallel^2$$

- two chain segments can rotate freely around each other



Polymer models – Gaussian chain model

- A characteristic feature of a polymer model is the mean squared end-to-end distance $\langle R^2 \rangle$



- For a Gaussian chain $U = \sum_{i=1}^{N} \frac{1}{2}\kappa \parallel \mathbf{x}_i - \mathbf{x}_{i-1} \parallel^2$, we have

$$\langle R^2 \rangle = b^2 N \qquad b^2 = \frac{3}{\kappa}$$

The proportionality to N is a universal feature and is also valid for other choices of U, e.g. stiff linkers

Polymer models – Effect of excluded volume

- In most applications, excluded volume interactions have to be taken into account, e.g. by an additional potential

$$U_{ex} = v k_B T \sum_{i<j} \delta(\mathbf{x}_i - \mathbf{x}_j)$$

- The relation for the mean square displacement changes to

$$\langle R^2 \rangle = b^2 N^{2\nu} \qquad \nu \approx 0.588$$

- But: Exact calculations become impossible

Polymer models – globular state

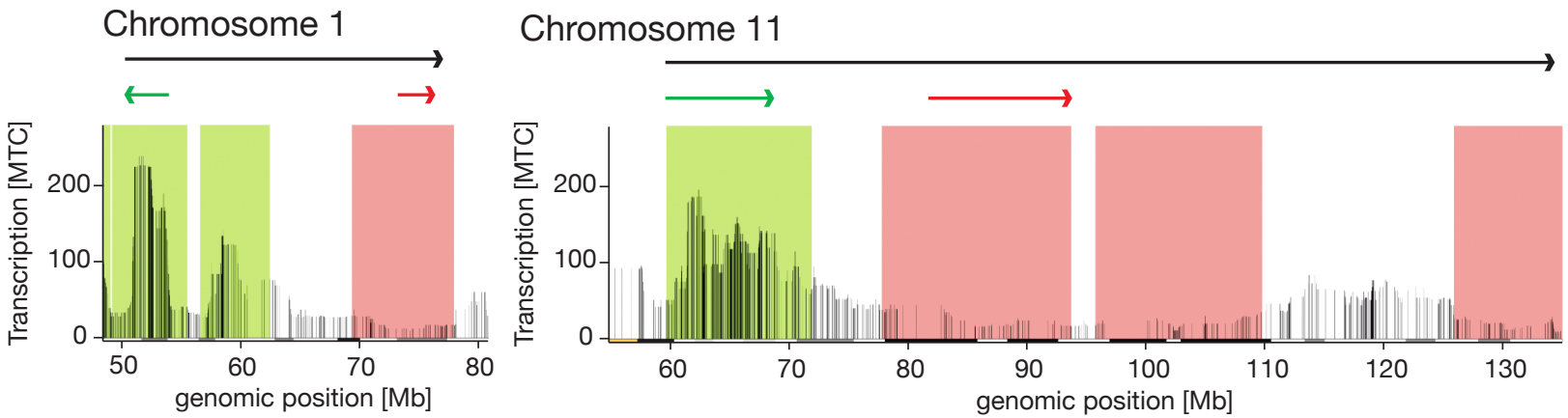
Third important polymer state: globular state

- further attractive interactions counteract excluded volume interactions
  $\longrightarrow$ collapse of the polymer
- e.g. proteins
- mean square displacement

$$\left\langle R^2 \right\rangle = b^2 N^{2\nu}, \qquad \nu = 1/3$$

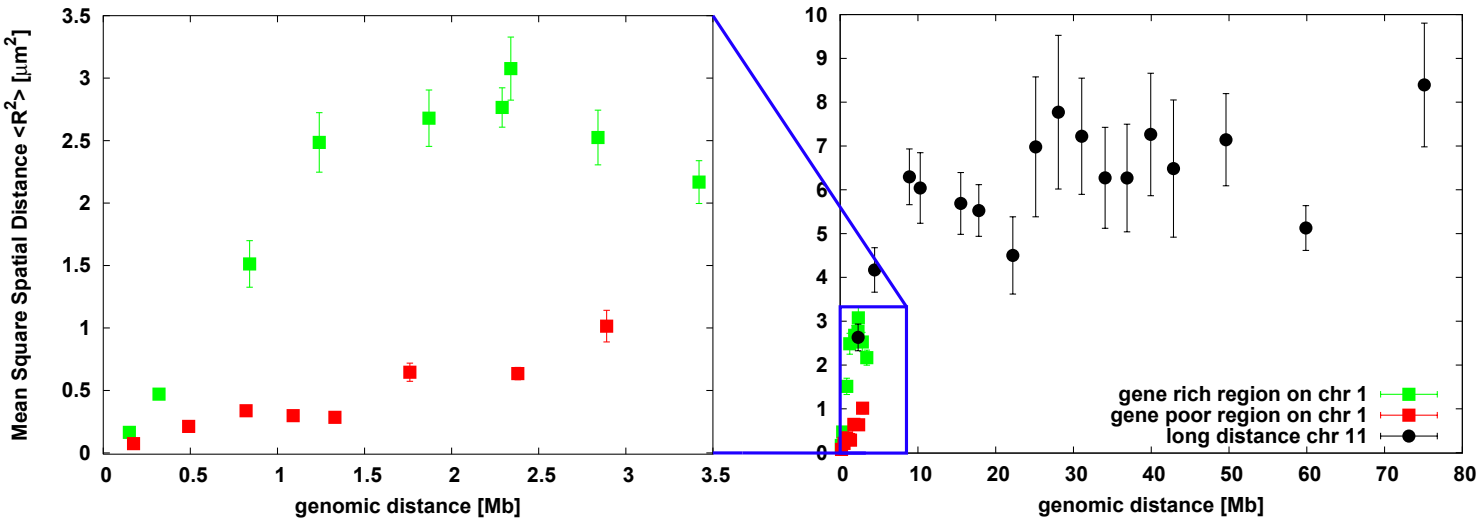
Experimental results

- Experiments (Roel van Driel, Amsterdam) measure mean square displacement in relation to genomic distance[1]
  - chromosome 1 and 11
  - gene-rich and gene-poor domains



[1] Mateos-Langerak *et al.*, 2007 (submitted to PNAS)

Experimental results



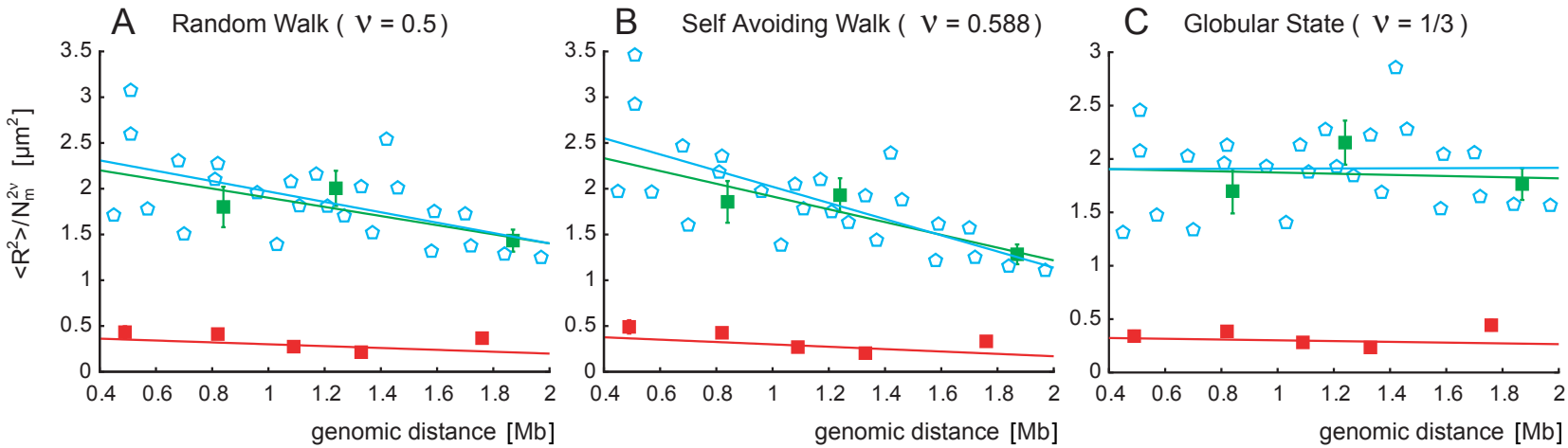
Results:

- Up to 2-3 Mb the mean square displacement (MSD) increases with genomic separation.
- Above this distance, the mean square displacement levels off and becomes independent of genomic distance

Experimental results

Which polymer model describes best the experimental results?
small genomic distances < 3 Mb

MSD increases like $\langle R^2 \rangle \sim N^{2\nu}$, but which ν (RW, SAW, GS)?

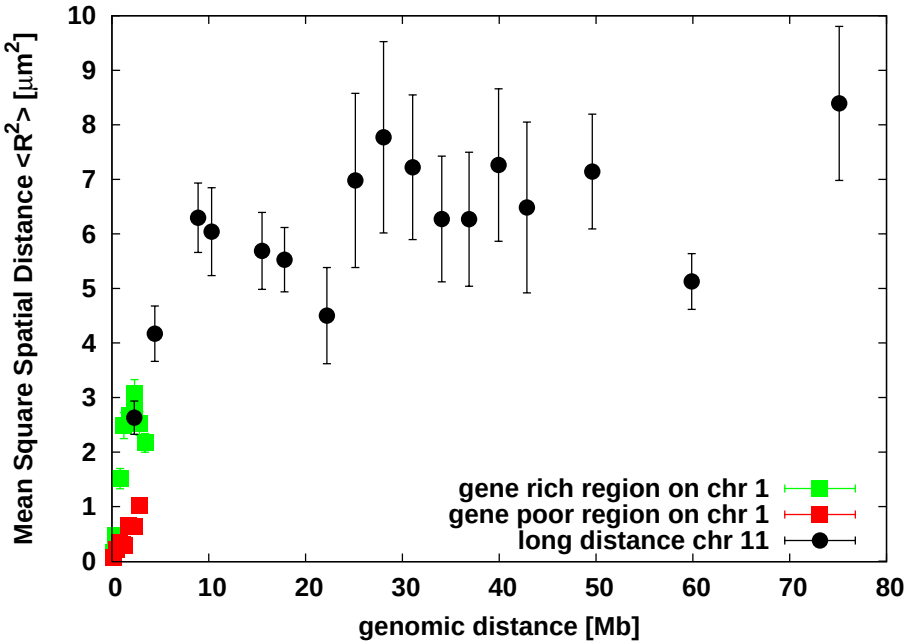


The globular state describes chromatin organization in the regime below ~ 2 Mb

additional data: Yokota *et al.*, 1995

Experimental results

Which polymer model describes best the experimental results?
large genomic distances

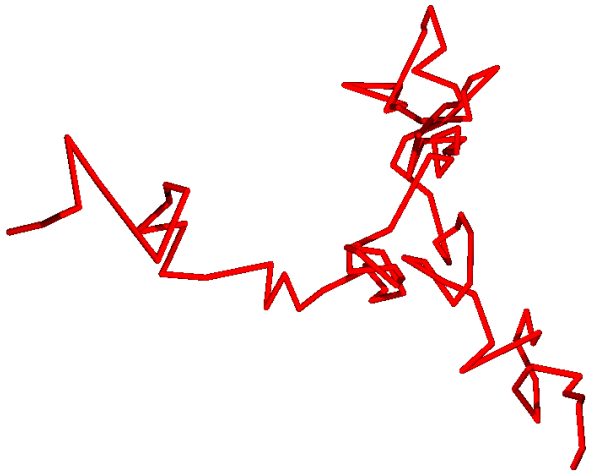


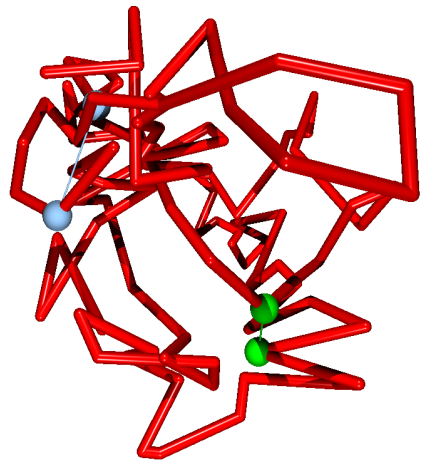
None of the polymer models with $\langle R^2 \rangle \sim N^{2\nu}$ can explain folding behaviour

A new model: the Random Loop Model

Idea

The formation of loops is necessary to obtain $\langle R^2 \rangle \sim \mathcal{O}(1)$





- The number of loops and their size are random and change in the course of time
- biological justification: there is evidence that different genes assemble to so-called “transcription factories”.

Mathematical model

- We consider $N+1$ monomers in a general potential

$$U = \frac{1}{2} \sum_{i<j} \kappa_{ij} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2$$

- connectivity of the chain: $\kappa_{ij} = \kappa, \quad |i-j| = 1$
- random loops: The κ_{ij} with $|i-j| > 1$ are Bernoulli distributed with probability $\mathcal{P}$

Note: We neglect the effect of excluded volume

Mathematical model

- We consider $N+1$ monomers in a general potential

$$U = \frac{1}{2} \sum_{i<j} \kappa_{ij} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2$$

- connectivity of the chain: $\kappa_{ij} = \kappa, \quad |i-j| = 1$
- random loops: The κ_{ij} with $|i-j| > 1$ are Bernoulli distributed with probability $\mathcal{P}$

Note: We neglect the effect of excluded volume

Question: What is the mean square physical distance between two arbitrary beads *I* and *J* of the chain?

Quenched vs. annealed average

We have two types of average in our model:

1. The thermal average over the positions $\mathbf{x}_0, \ldots, \mathbf{x}_N$, $\rightarrow$ partition sum $\mathcal{Z}$
2. The average over disorder of loops κ_{ij} (random variables)

Quenched vs. annealed average

We have two types of average in our model:

1. The thermal average over the positions $\mathbf{x}_0, \ldots, \mathbf{x}_N$, $\rightarrow$ partition sum $\mathcal{Z}$
2. The average over disorder of loops κ_{ij} (random variables)

Two ways of averaging in disordered systems:

Quenched average	**Annealed average**
$(\tau_{dis} \gg \tau_{eq})$	$(\tau_{eq} \gg \tau_{dis})$
$\langle F \rangle = k_B T \langle \log \mathcal{Z} \rangle_{\text{dis}}$	$\langle F \rangle = k_B T \log \langle \mathcal{Z} \rangle_{\text{dis}}$

We presume that the quenched average is the relevant one!

Calculation of the quenched average

- We are interested in the mean square displacement between two arbitrary beads I and J
- The average over the thermal ensemble for quenched disorder can be calculated to be

$$\left\langle r_{IJ}^2 \right\rangle_{\text{thermal}} = 3\left(\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}\right)$$

where

$$\sigma^{-1} = K = \begin{pmatrix} \sum_{\substack{j=0 \\ j\neq 1}}^{N} \kappa_{1j} & -\kappa_{12} & \cdots & -\kappa_{1N} \\ -\kappa_{21} & \sum_{\substack{j=0 \\ j\neq 2}}^{N} \kappa_{2j} & \cdots & -\kappa_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -\kappa_{N1} & -\kappa_{N2} & \cdots & \sum_{\substack{j=0 \\ j\neq N}}^{N} \kappa_{Nj} \end{pmatrix}$$

Calculation of the quenched average

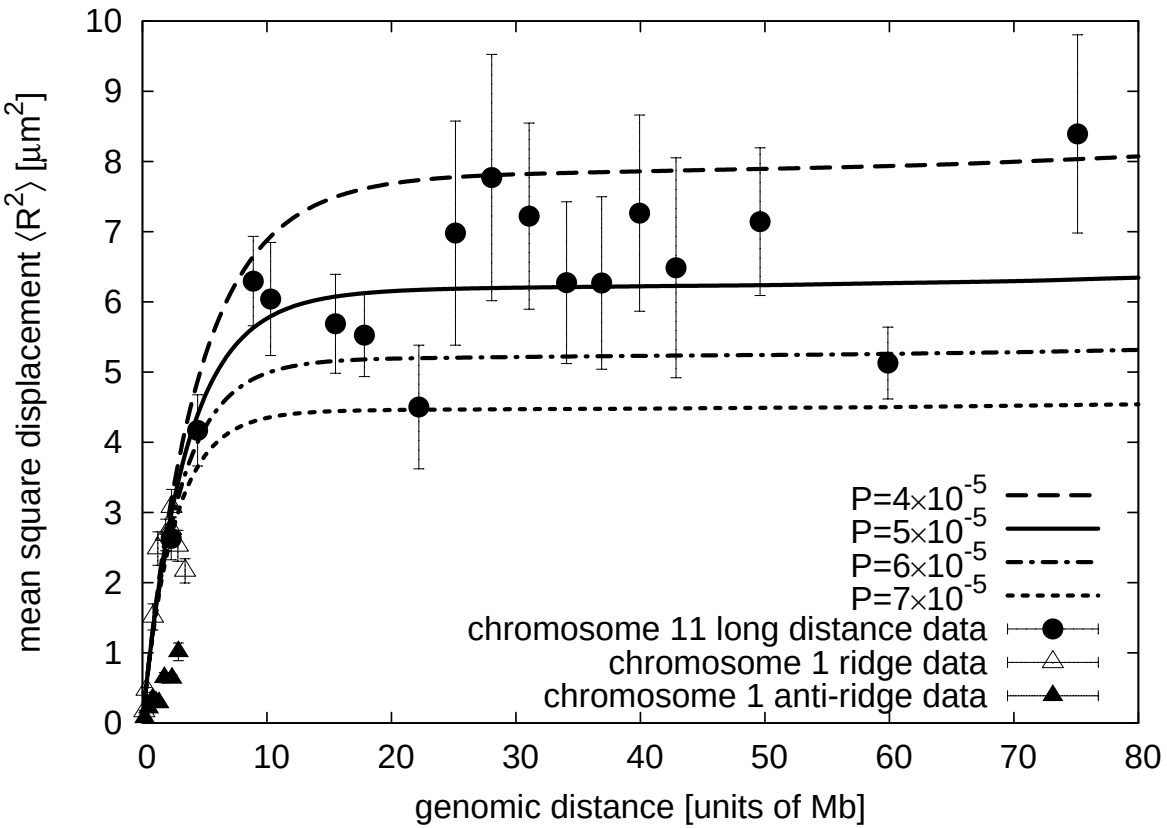
- Now we have to perform the average over the disorder of loops (the variables $\{\kappa_{ij}\}$).

$$\begin{aligned} \left\langle r_{IJ}^2 \right\rangle &= \left\langle \left\langle r_{IJ}^2 \right\rangle_{\text{thermal}} \right\rangle_{\text{loops}} \\ &= 3 \left(\left\langle \sigma_{JJ} \right\rangle_{\text{loops}} + \left\langle \sigma_{II} \right\rangle_{\text{loops}} - 2 \left\langle \sigma_{IJ} \right\rangle_{\text{loops}} \right) \end{aligned}$$

- This is equivalent to averaging over the ensemble of matrices K and can only be done numerically

Results

Mean square displacement



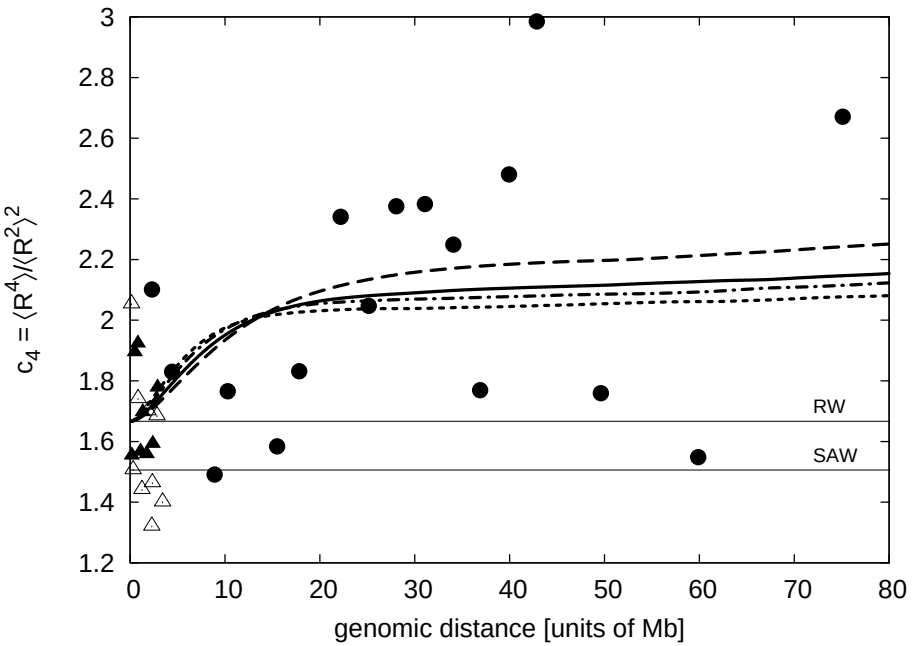
- RLM explains shows the leveling-off
- Only a few loops on all scales per 100 Mbp are necessary ($\sim$ 15)

Results

Relation between higher-order moments

$$c_4 = \left\langle R^4 \right\rangle / \left\langle R^2 \right\rangle^2$$

- Significant deviations from RW or SAW behaviour
- It is the average over the disorder in the RLM that makes c_4 larger



The effect of excluded volume

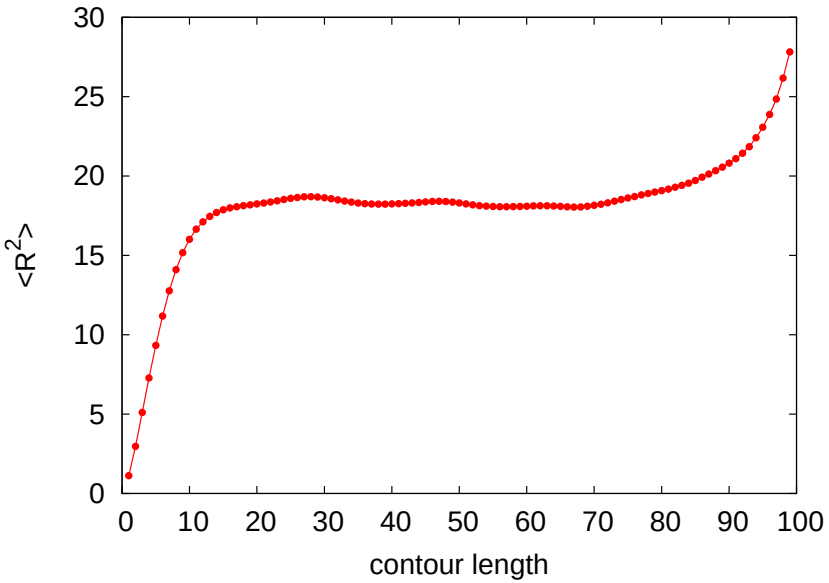
- We need to use molecular dynamics simulations (ESPResSo package)
- Excluded volume is modelled by a shifted Lennard-Jones potential with cutoff at minimum ($\rightarrow$ no attractive part).

$$U_{LJ}(r) = \begin{cases} 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) + \epsilon & \text{for } r < \sqrt[6]{2}\sigma \\ 0 & \text{otherwise} \end{cases}$$

The effect of excluded volume

- We need to use molecular dynamics simulations (ESPResSo package)
- Excluded volume is modelled by a shifted Lennard-Jones potential with cutoff at minimum ($\rightarrow$ no attractive part).

$$U_{LJ}(r) = \begin{cases} 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) + \epsilon & \text{for } r < \sqrt[6]{2}\sigma \\ 0 & \text{otherwise} \end{cases}$$



The same qualitative behaviour as without excluded volume interactions!

The annealed ensemble

- For the annealed ensemble we have to take the disorder average over the partition sum

$$\langle \mathcal{Z} \rangle_{\text{dis}} = \sum_{\{\kappa_{ij}\}} \mathcal{Z}\left(\{\mathbf{x}_k\}, \{\kappa_{ij}\}\right) \, p\left(\{\kappa_{ij}\}\right)$$

The annealed ensemble

- For the annealed ensemble we have to take the disorder average over the partition sum

$$\langle \mathcal{Z} \rangle_{\text{dis}} = \sum_{\{\kappa_{ij}\}} \mathcal{Z}\left(\{\mathbf{x}_k\}, \{\kappa_{ij}\}\right) p\left(\{\kappa_{ij}\}\right)$$

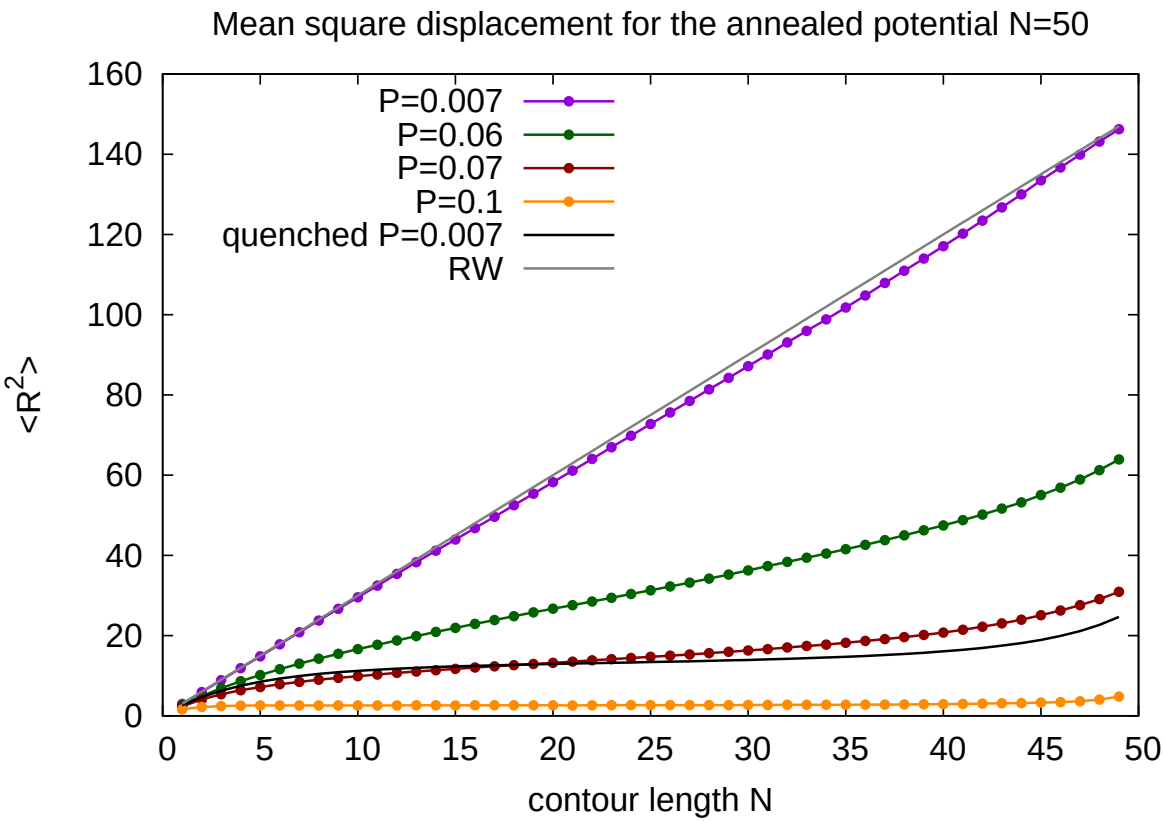
- The disorder average can be carried out, and the partition sum can be rewritten in terms of an effective potential

$$\langle \mathcal{Z} \rangle_{\text{dis}} = \int \int d\mathbf{x}_1 \ldots d\mathbf{x}_N \exp(-U_{\text{eff}})$$

$$U_{\text{eff}} = \frac{1}{2}\kappa \sum_{i=0}^{N-1} r_{i,i+1}^2 - k_B T \sum_{|i-j|>1} \log\left[1 + \mathcal{P}\left(e^{-\frac{1}{2}\kappa r_{ij}^2} - 1\right)\right]$$

The annealed ensemble

- MD simulations with annealed potential
- Leveling-off does occur only at much larger probabilities $\mathcal{P}$



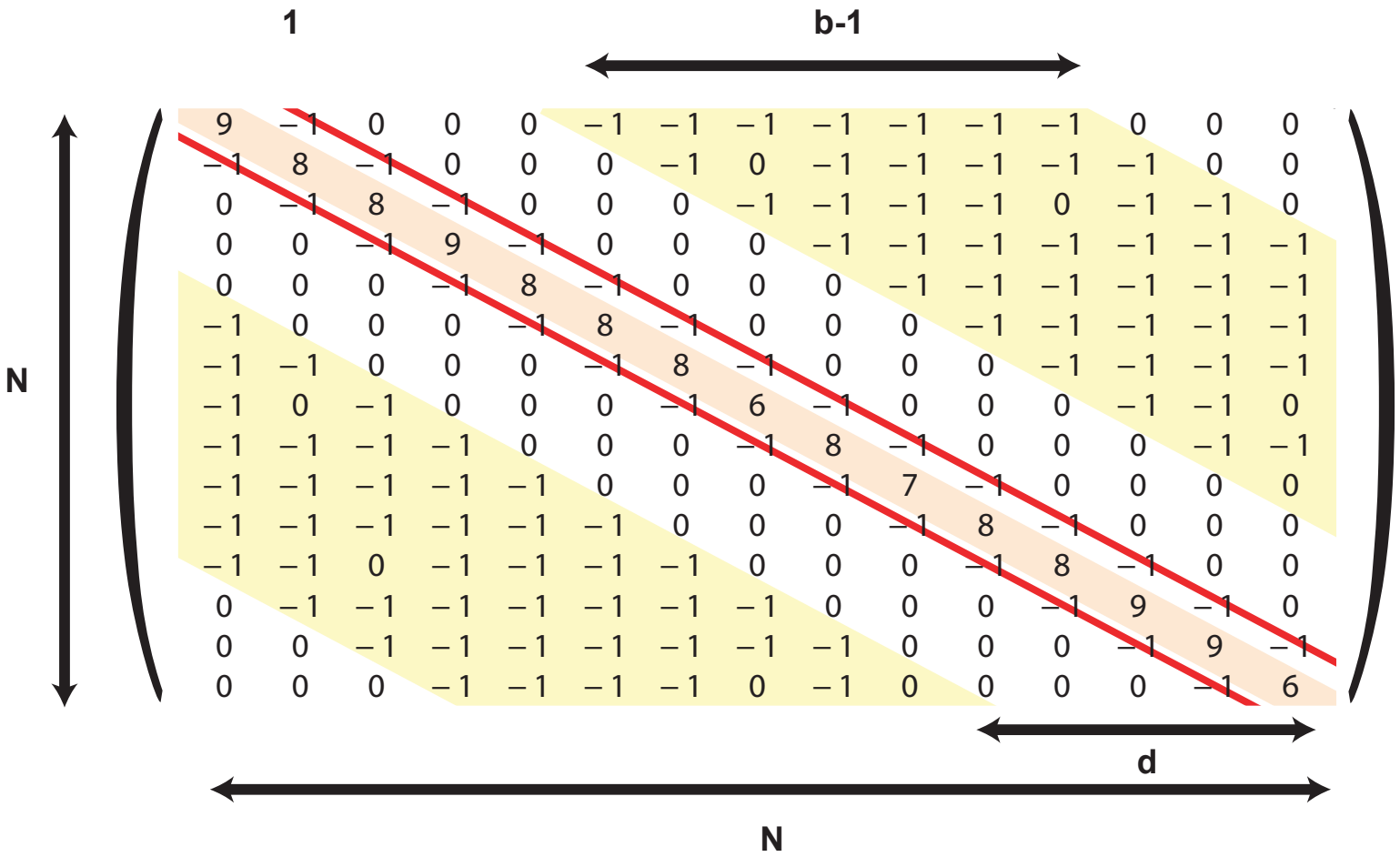
Random Matrices

For the quenched ensemble we had to average over random matrices.

$$\sigma^{-1} = K = \begin{pmatrix} \sum_{\substack{j=0 \\ j\neq 1}}^{N} \kappa_{1j} & -\kappa_{12} & \cdots & -\kappa_{1N} \\ -\kappa_{21} & \sum_{\substack{j=0 \\ j\neq 2}}^{N} \kappa_{2j} & \cdots & -\kappa_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -\kappa_{N1} & -\kappa_{N2} & \cdots & \sum_{\substack{j=0 \\ j\neq N}}^{N} \kappa_{Nj} \end{pmatrix}$$

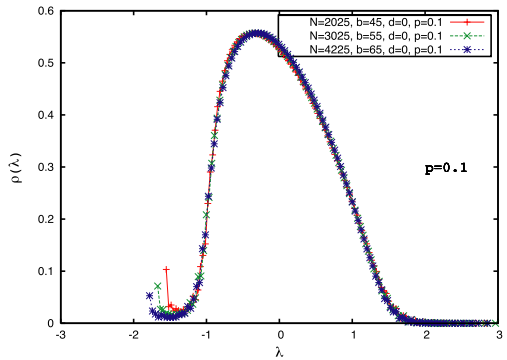
Random Matrices

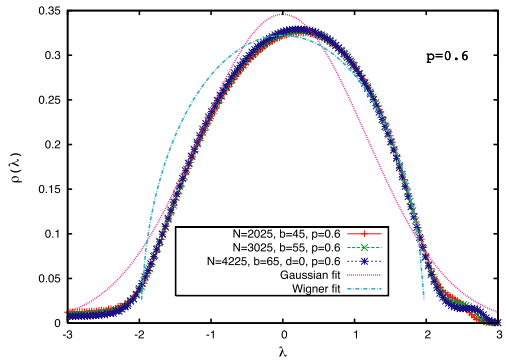
For the quenched ensemble we had to average over random matrices. We considered a slightly different ensemble of random matrices

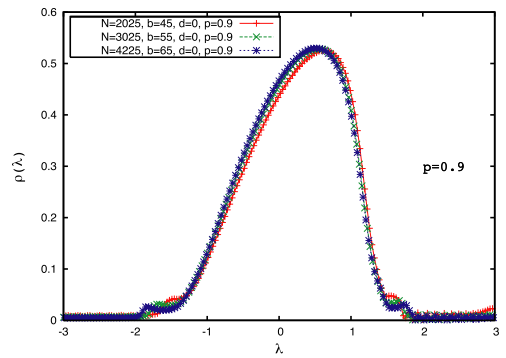


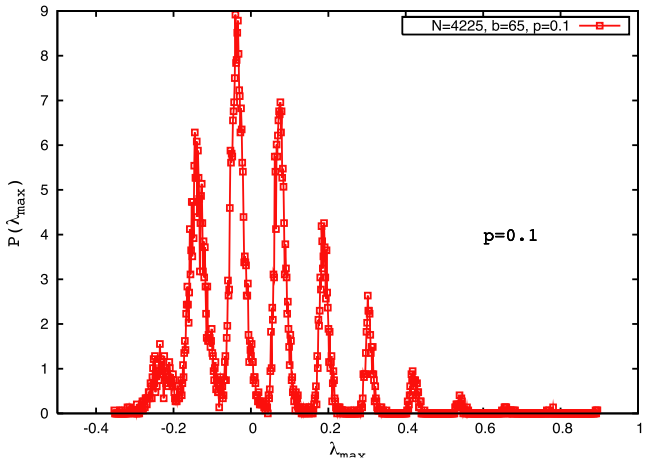
Random Matrices

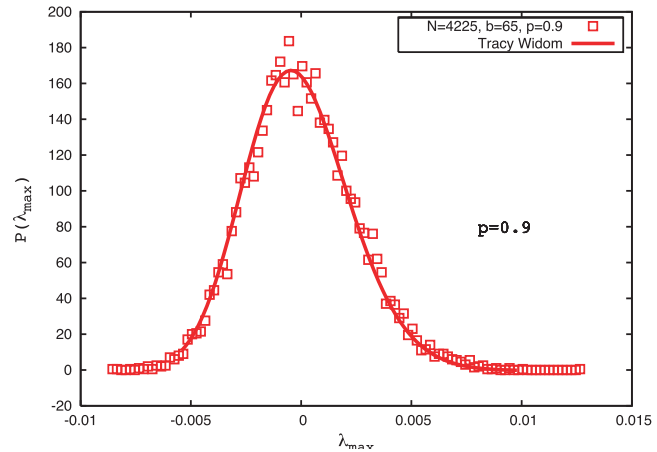
The eigenvalue distribution changes its skew at the percolation threshold. Above it shows a universal behaviour.











Conclusions

Results:

- There are two folding regimes:
  - On small genomic distances (< 2 Mb) the folding of chromatin can be described by a globular state
  - On large genomic distance ($> 5 - 10$ Mb) the folding motifs can be characterized by the formation of random loops

Open Questions:

- How can the crossover region ($2 - 10$ Mb) be described?
- Can one generalize this model to describe the common structure and intermingling of all chromosomes in the nucleus?

Conclusions

Results:

- There are two folding regimes:
  - On small genomic distances (< 2 Mb) the folding of chromatin can be described by a globular state
  - On large genomic distance ($> 5 - 10$ Mb) the folding motifs can be characterized by the formation of random loops

Open Questions:

- How can the crossover region ($2 - 10$ Mb) be described?
- Can one generalize this model to describe the common structure and intermingling of all chromosomes in the nucleus?

Thanks for your attention!