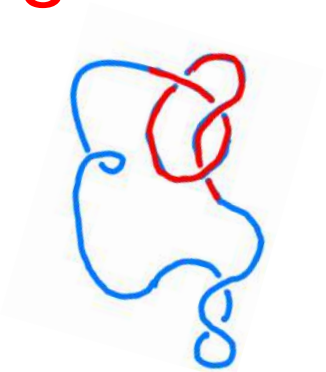
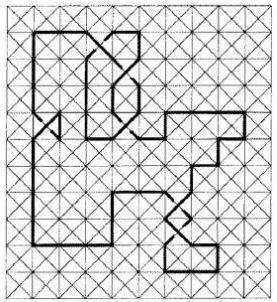
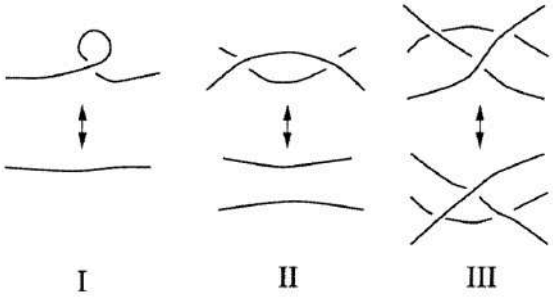
Working with Emmanuel

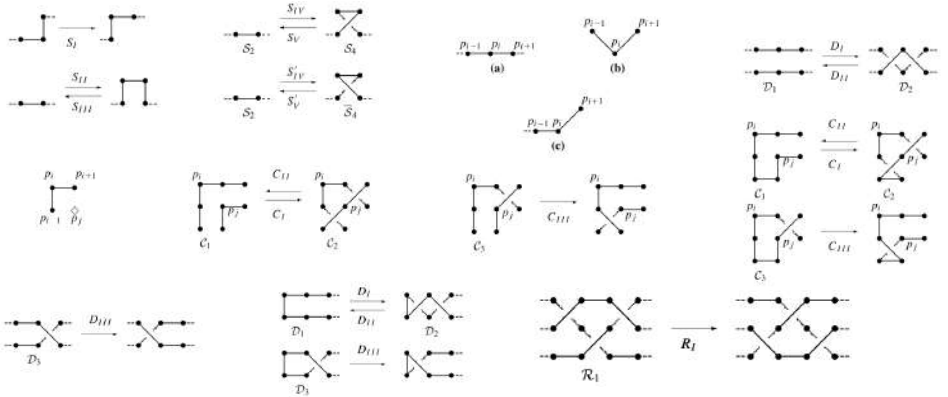
How the presence of a knot in a polymer ring can affect its metric and entropic properties ?

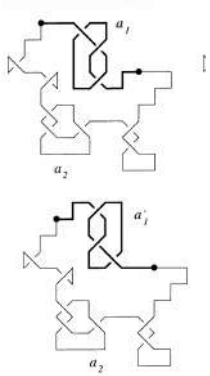


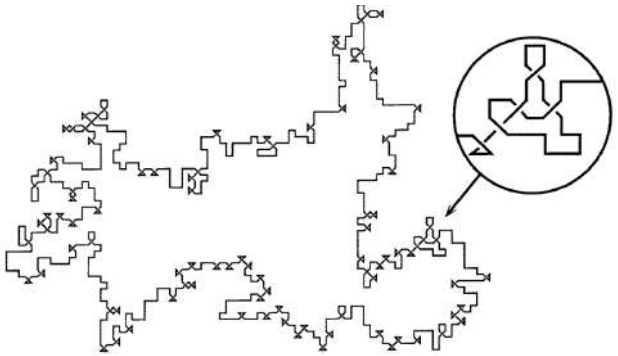


Redeimeister moves



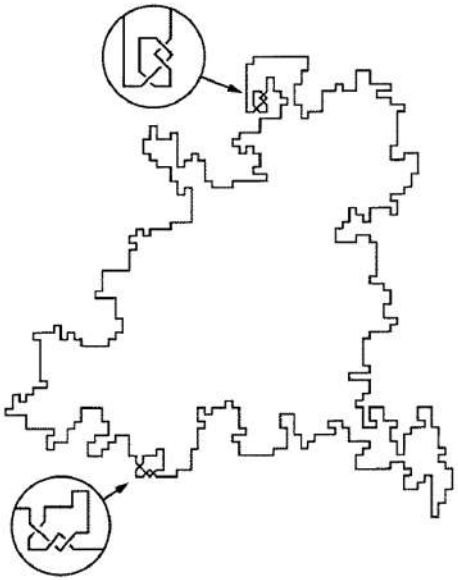






$$p_n(\tau) \sim A(\tau)(\mu(\tau))^n n^{\alpha(\tau)-3}.$$

$\mu(\emptyset) = 3.254 \pm 0.01$	$\alpha(\emptyset) = 0.58 \pm 0.07$
$\mu(3_1) = 3.250 \pm 0.06$	$\alpha(3_1) = 2.10 \pm 0.08$
$\mu(4_1) = 3.255 \pm 0.07$	$\alpha(4_1) = 2.13 \pm 0.10$
$\mu(3_1\#3_1) = 3.249 \pm 0.13$	$\alpha(3_1\#3_1) = 3.91 \pm 0.24$
$\mu(3_1\#4_1) = 3.261 \pm 0.15.$	$\alpha(3_1\#4_1) = 4.05 \pm 0.25.$



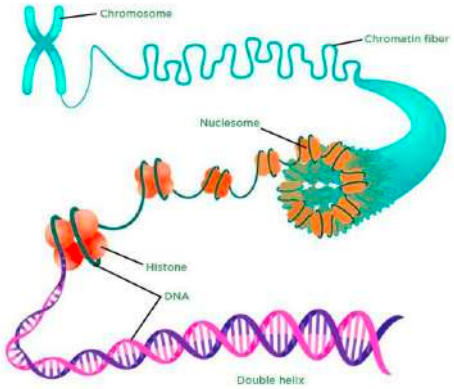
Voici a monstrous configuration sampled by Emmanuel's Monte Carlo method and now sitting in the IphT coffe room

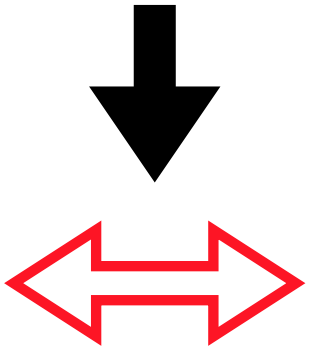
Topological spectra and entropy of chromatin loop networks

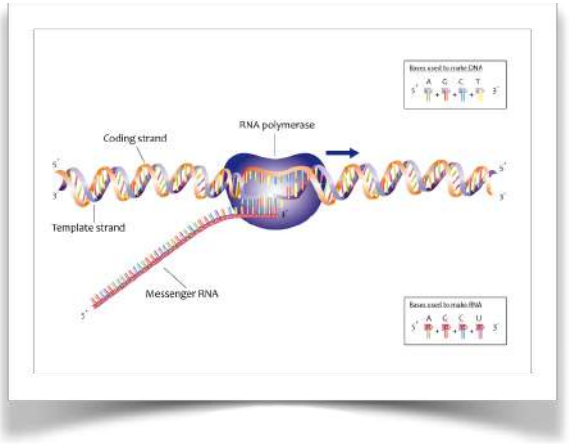
E. Orlandini (Università di Padova)

A.Bonato (Strathclyde), M. Chiang (Edimburg) D. Corbett (Edimburgh), S. Kitaev (Strathclyde), D. Marenduzzo (Edimburg), Al. Morozov (Edimburg)

Polymer models + Combinatorial arguments





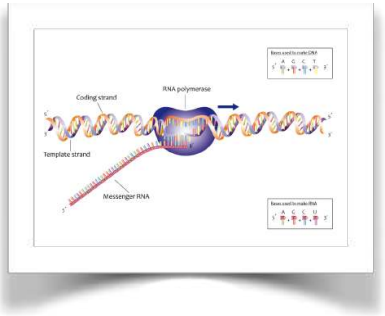


3D-folding of Chromatin

Transcription activity

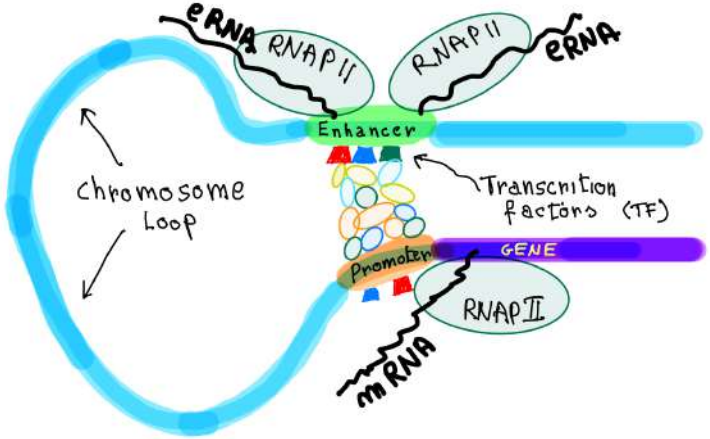
L'ESPRIT DES CARTES, : une conférence en l'honneur d'Emmanuel Guitter
CEA, Saclay 15-16 May 2025

Transcription and chromatin organisation

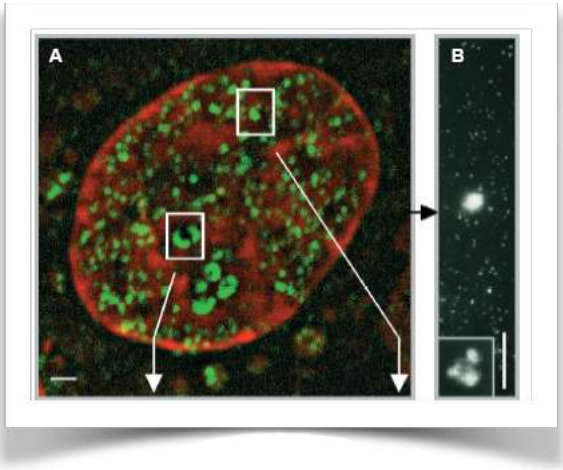


Transcription in 1D: process through which an RNA polymerase "reads" DNA and produces an RNA transcript

Role of the 3D chromatin structure on transcription: often promoters and enhancers are close in space (via looping) to trigger transcription.



Transcription is not organised randomly inside the nucleous



Polymerases and related proteins cluster in the so called "transcription factories" (0.1 -1μ)

P. Cook Science 1999, Image by A. Pombo,

Polymer models of chromatin folding as valuable investigation tools to link transcription activity with DNA mammalian organisation

Simple model of 3D transcription within chromatin

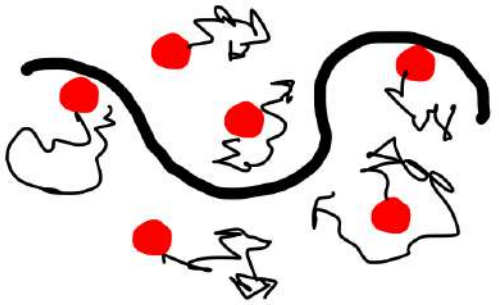
C. A. Brackley et al., PNAS (2013); see also M. Barbieri et al., PNAS (2012)

Chromatin: semi flexible polymer

interacting with chromatin-binding proteins

polymerase, Transcription factors, etc

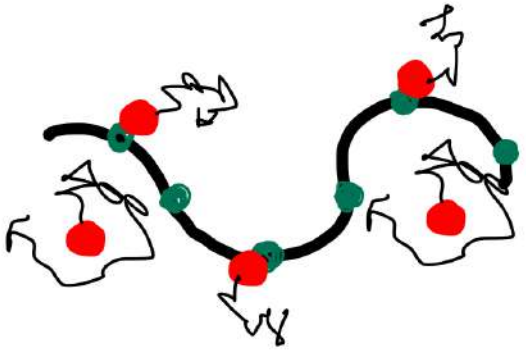
model: diffusing spherical particles



Polymerases interact with most sites on the DNA weakly

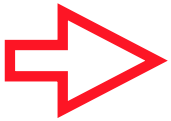
but strongly with some of them

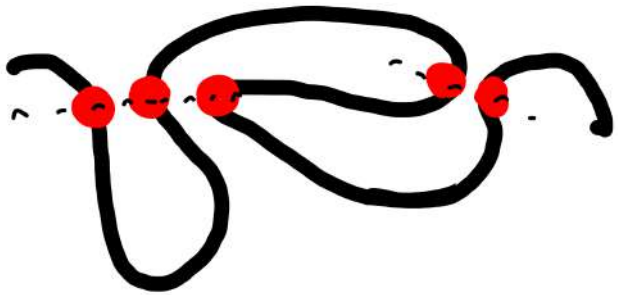
(promoters/enhancers of genes, TU)



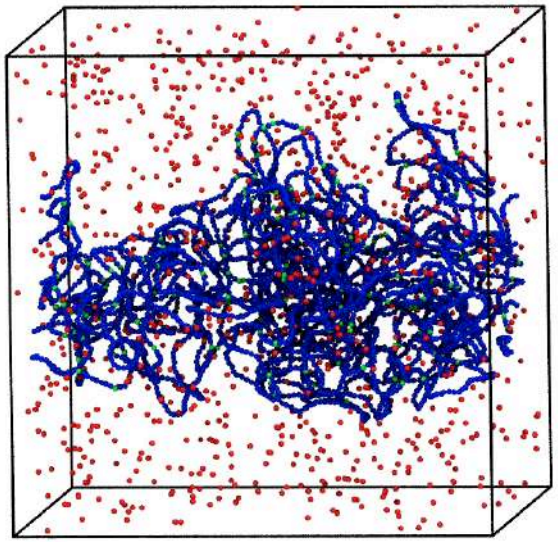
Important point:

protein complexes can bind multivalently ==> Molecular bridges



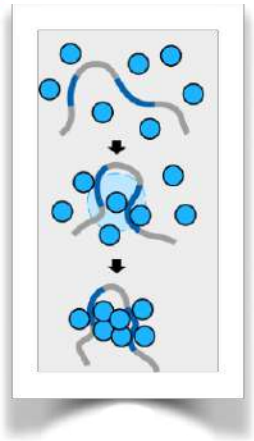


What are the resulting chromatin structures ?



Spontaneous formation of several finite-size clusters of proteins and binding sites decorated by loops and separated by ties

Microphase separation can be explained via a positive feedback loop known as bridging-induced attraction + entropy loos



The recruitment of chromatin-binding proteins increases bridging and loop formation

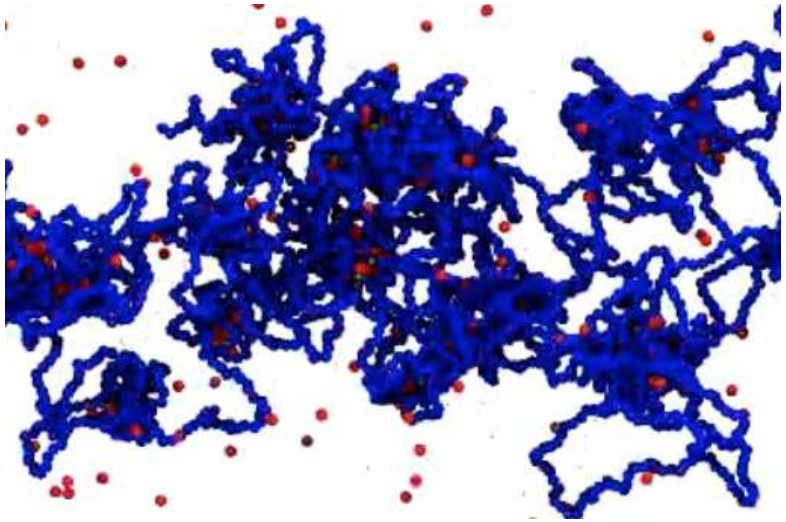
The presence of many loops arrest the phase separation:

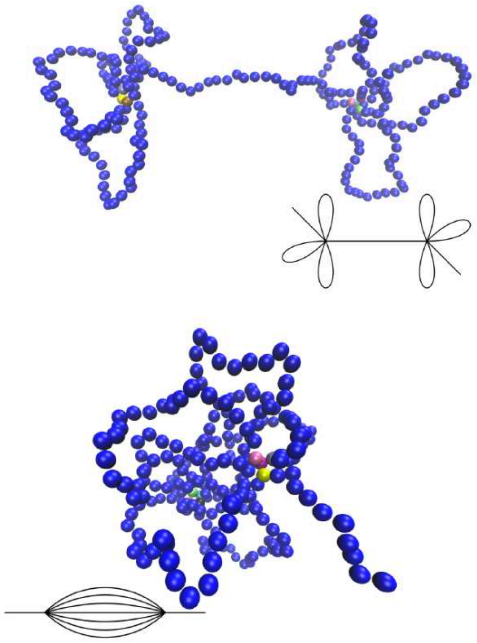
(i) corona of loops hinders kinetically the targeting of other binding proteins

(ii) the entropic cost to form many loops increases super linearly with their number

B. Duplantier: J.Stat.Phys. (1989) D. Marenduzzo, E.O , J.Stat.Mech. (2009)

Competition between gain in binding energy and entropy loos
=> clusters do not coarsen past a typical size





Typical 3D motifs of loops and clusters whose topology can be described in terms of simple graphs

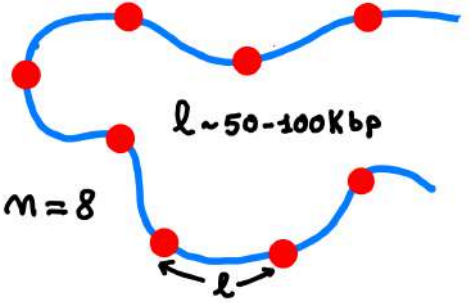
Claim: The collection of these motifs determine the 3D structure of gene and their transcriptional activity

⇩ Aim ⇩

(i) Classify and enumerate these 3D motifs in terms of the topology of their underlying graphs;

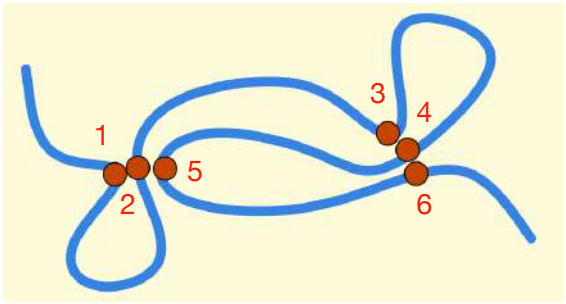
(ii) Predict their probability of occurrence along the genome by computing their multiplicity and configurational entropy.

Focus on a stretch of chromatin with n=8
highly sticky sites (TUs)
(most gene loci have #TUs ~5-10)
Chiang et al bioRxiv 2022

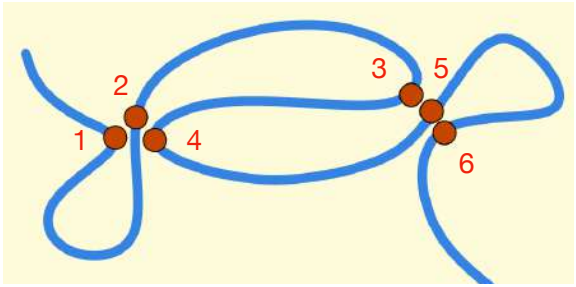


If one looks at the relative position of the TUs (transcriptional activity of the promoter)

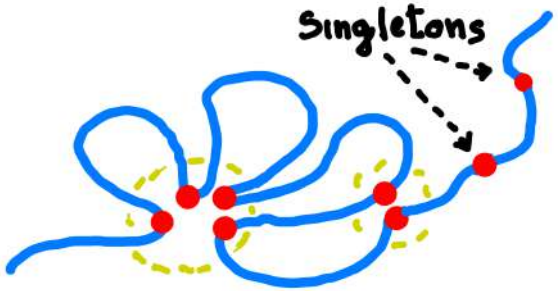
Distinguishable vertices (Labelled graphs)



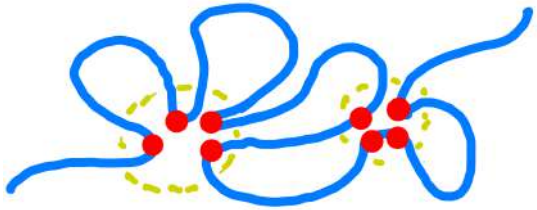
≠



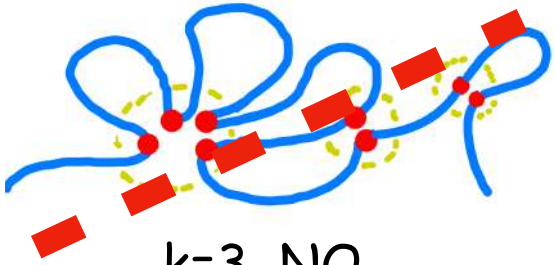
No singletons



Case with 2 clusters only (k=2)



k=2 OK



k=3 NO

# of inequivalent labelled topologies

Stirling number of second kind

Singletons

$$N(n,2) = S(n,2) - n = 2^{n-1} - n - 1$$

Partition of 8 letters into 2 groups

For a model with n TUs and k clusters, the number of inequivalent topologies (without singletons) obeys the recursion

$$N(n,k) = kN(n-1,k) + (n-1)N(n-2,k-1).$$

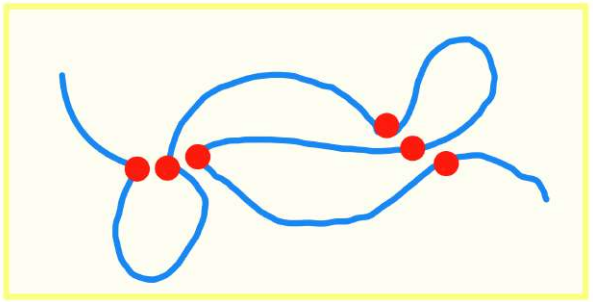
From this relation we can compute the generating function

$$G(t,x) = \sum_{n\geq 0} N(n,k)\frac{t^n}{n!}x^k$$

$$\boxed{G(t,x) = e^{x(e^t-1-t)}}$$

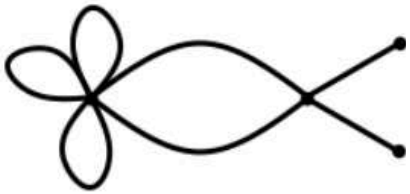
If interested to the **statistical weights** of different topologies the labelling of the TUs can be omitted

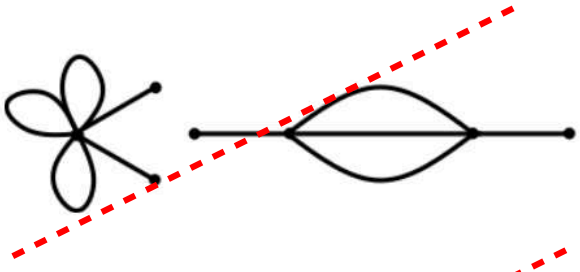
Focus on graphs with indistinguishable vertices
(Unlabelled graphs)



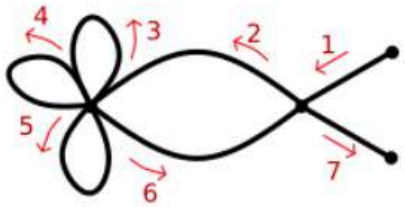
Constraints to satisfy

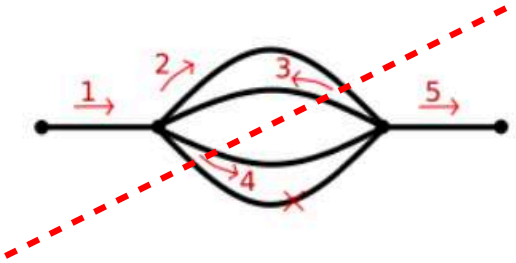
Connectivity





Traversability



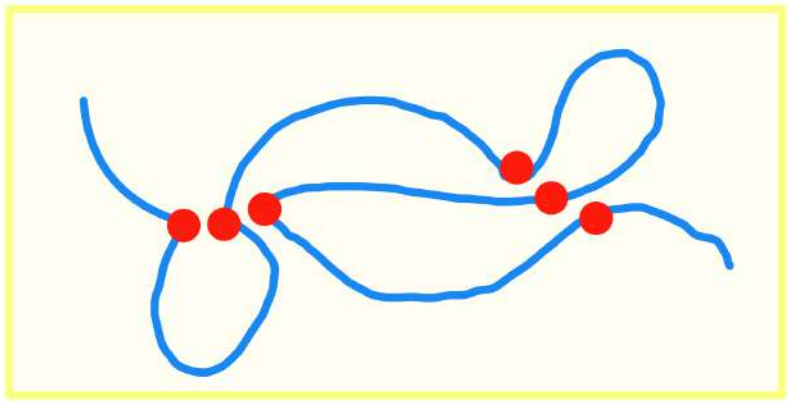


Very challenging to compute $N_u(n,k)$ for k generic

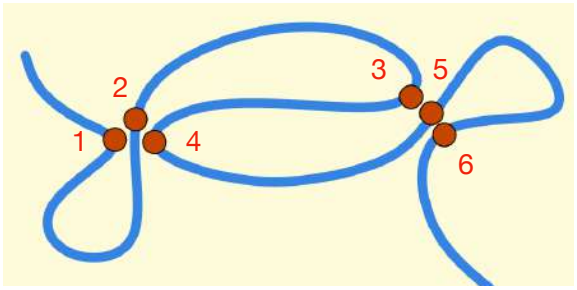
but if k=2

$$N_u(n,2) = \frac{n(n-1)}{2} - 2 - \frac{\lfloor \frac{n-1}{2} \rfloor \lfloor \frac{n+1}{2} \rfloor}{2} \sim n^2$$

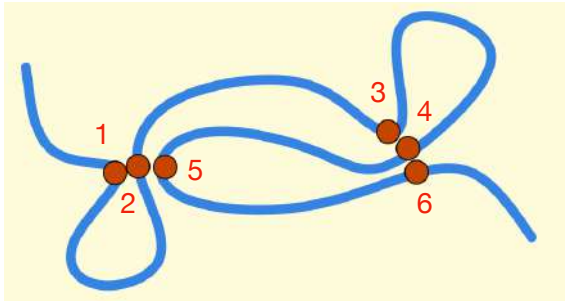
Our specific case $N_u(n=8,2) = 20$



Unlabeled network topology



Multiple labeled graphs



Combinatorial weight or multiplicity Ω_i

Number of different labelled networks corresponding to the i-esim unlabelled one.

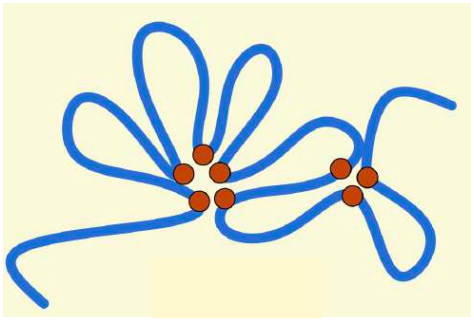
N.B. This is different for different topologies.

$N_u(n=8,2)=20$

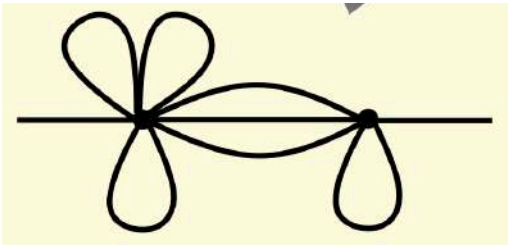
L_1, L_2: node degree

n_l number of loops;

n_t number of ties



$L_1=10, L_2=6$



$n_l=4, n_t=3$

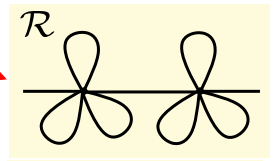
Diagram	n_t	n_l	L_1	L_2	Ω
	1	6	8	8	1
	2	5	8	8	3
	3	4	8	8	9
	4	3	8	8	9
	5	2	8	8	9
	6	1	8	8	3
	7	0	8	8	1
	1	6	6	10	2
	2	5	10	6	4
	2	5	6	10	2
	3	4	6	10	16
	4	3	10	6	12
	4	3	6	10	4
	5	2	6	10	12
	6	1	10	6	4
	1	6	4	12	2
	2	5	12	4	5
	2	5	4	12	1
	3	4	4	12	10
	4	3	12	4	10

$L_1=8$
$L_2=8$

$L_1=6$
$L_2=10$

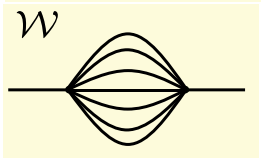
$L_1=4$
$L_2=12$

Extreme cases: rosettes and watermelon

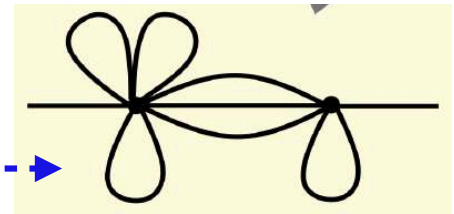


$n_t=1, n_l=6$

$\Omega_i=1$



$n_t=7, n_l=0$



$n_t=3, n_l=4$

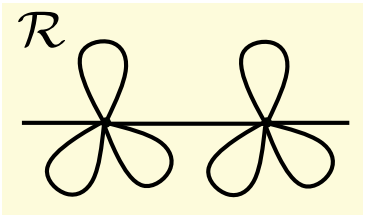
$\Omega_i=16$

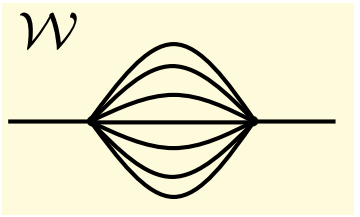
N.B. Multiplicity tends to be larger for hybrid networks (intermediates between R and W)

Do they also occur more often ?

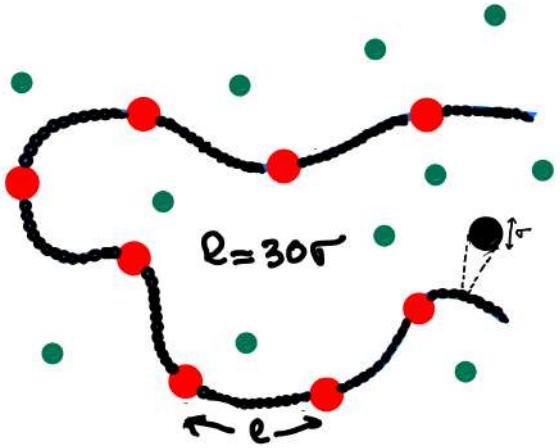
This should depend in the competition between combinatorics multiplicity and entropic cost to form loops or ties

In other words …

Is it more frequent 

or 

?

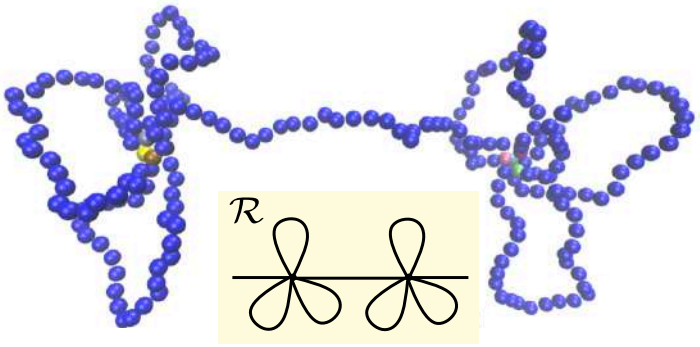


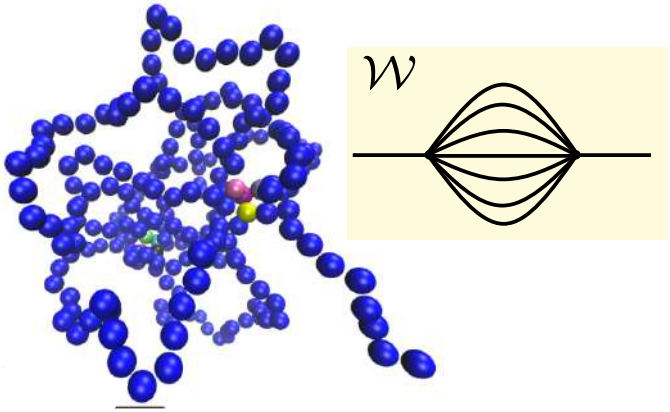
Entropy estimate by MD simulations

Diffusing beads are binding to specific sites (red beads)

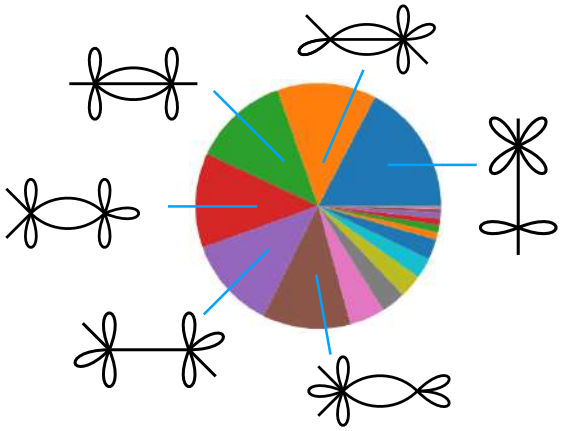
Non-specific (weak) interactions are neglected

Several topologies occur but we focus on those with k=2





How about the statistical weight of each network (entropy) ?



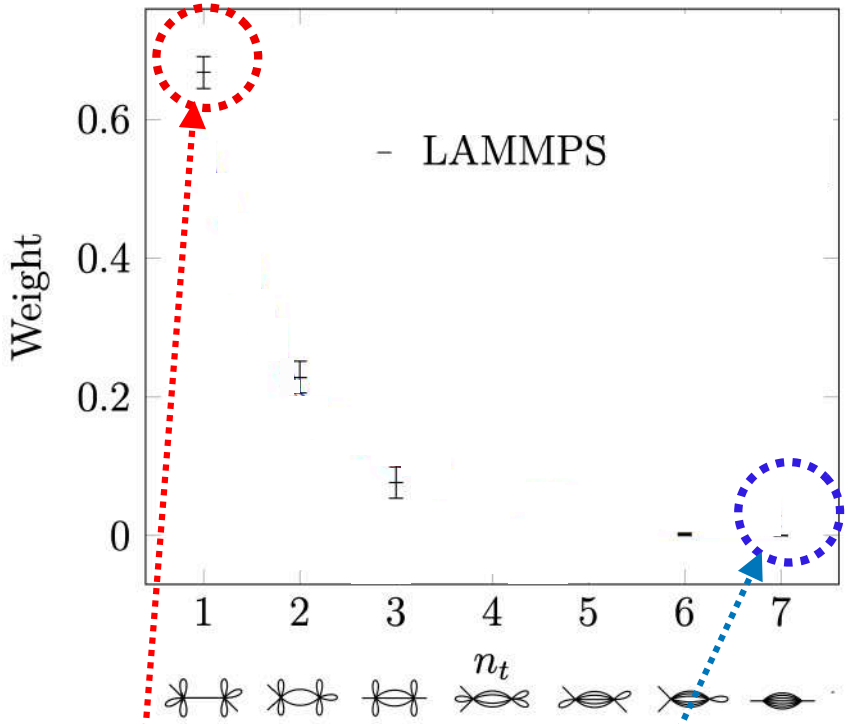
Very few topologies are highly populated

top six topologies account for ~ 80% of the total structures

Diagram	n_t	n_l	L_1	L_2	Ω
	1	6	8	8	1
	2	5	8	8	3
	3	4	8	8	9
	4	3	8	8	9
	5	2	8	8	9
	6	1	8	8	3
	7	0	8	8	1

$L_1 = 8$

$L_2 = 8$

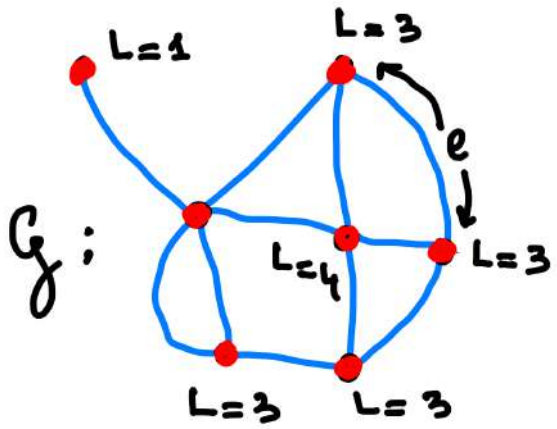


Networks with large nt (W-like) are much less populated than R-like ones

Stat weight of R is over two-order of magnitude larger than that of W

Duplantier' theory of self-avoiding networks

(B. Duplantier: J.Stat.Phys. 1989)

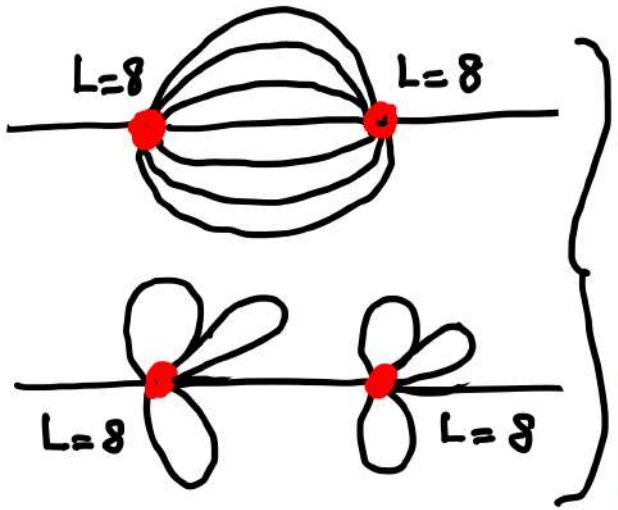


$$n_1 = 1$$
$$n_3 = 4$$
$$n_4 = 1$$
$$n_5 = 1$$

$$\mathcal{L} = 1 + \frac{1}{2}\sum_L n_L (L-2)$$

$$Z_g \overset{\sigma \ll \ell}{\sim} A_g \mu^{N_g} \ell^{\gamma_g - 1}$$

$$\gamma_g - 1 = -\nu d \mathcal{L} + \sum_L n_L \sigma_L$$



$n_g = 2$

$\mathcal{L}_w = \mathcal{L}_R$

$\gamma_w = \gamma_R$

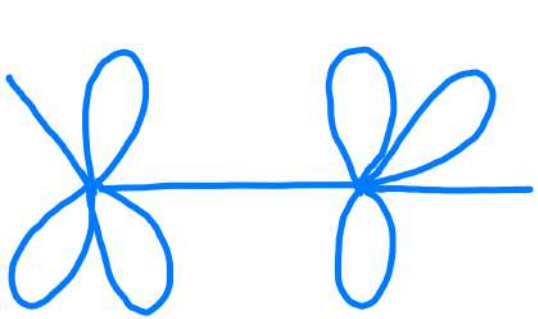
ℓ Cannot be arbitrarily large !!

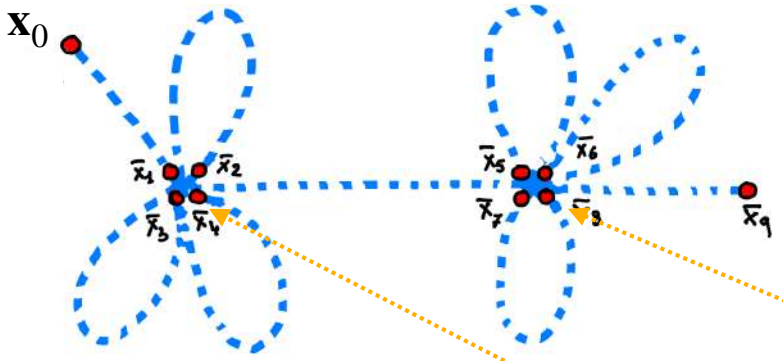
$\ell \sim (50 - 100)\,\mathrm{kbp}$ Human chromatin

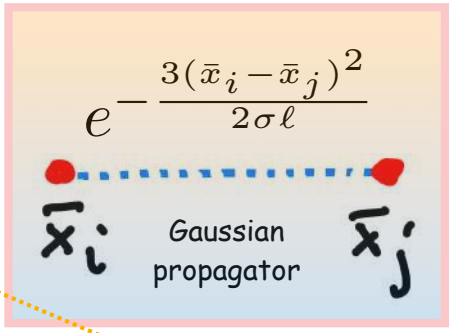
P. Cook "Principles of nuclear structure and function" Wiley (2001)
C. Brackley et al Nat Comm 2021

Amplitude A_G becomes relevant !

Computing amplitudes (Gaussian approx)







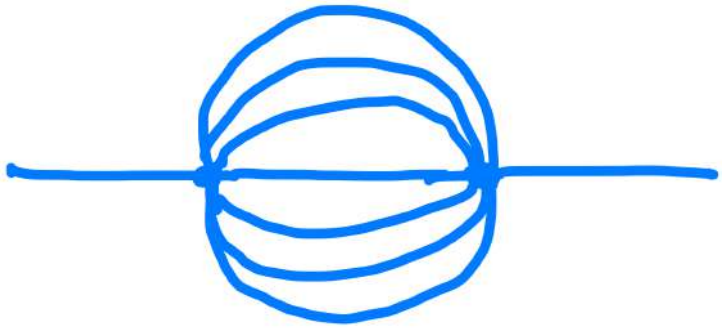
$$Z_{\mathcal{G}} = \int d\mathbf{x}_0 \dots d\mathbf{x}_{n+1}\, \delta(\mathcal{G}) \prod_{i=0}^{n} e^{-\frac{3(\mathbf{x}_{i+1}-\mathbf{x}_i)^2}{2l\sigma}} \qquad \delta(\mathcal{R}) = \prod_{i=2,3,4} \delta(\mathbf{x}_1 - \mathbf{x}_i) \prod_{j=6,7,8} \delta(\mathbf{x}_5 - \mathbf{x}_j)$$

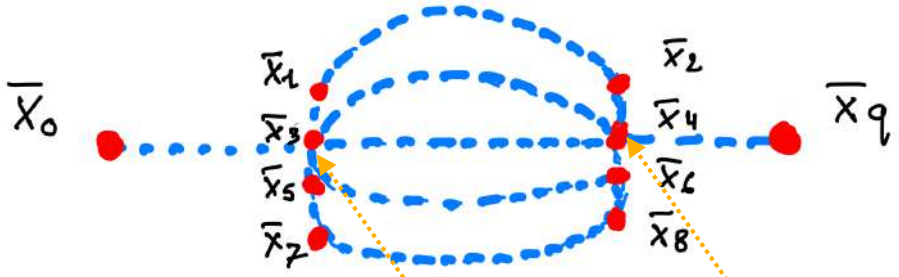
$$\Rightarrow \quad Z_{\mathcal{R}} = \int d\mathbf{x}_0 \dots d\mathbf{x}_9 \left[\prod_{i=0}^{8} e^{-\frac{3(\mathbf{x}_{i+1}-\mathbf{x}_i)^2}{2l\sigma}} \right] \prod_{i=2,3,4} \delta(\mathbf{x}_1 - \mathbf{x}_i) \prod_{j=6,7,8} \delta(\mathbf{x}_5 - \mathbf{x}_j)$$

Note: $$\int d\mathbf{x}_0\, e^{-\frac{3(\mathbf{x}_1-\mathbf{x}_0)^2}{2l\sigma}} = \int d\mathbf{x}_0\, e^{-\frac{3\mathbf{x}_0^2}{2l\sigma}} = W_0^{-1}, \qquad W_0 \equiv \left(\frac{3}{2\pi l\sigma}\right)^{3/2}$$

$$Z_{\mathcal{R}} = W_0^{-2} \int d\mathbf{x}_1 d\mathbf{x}_5\, e^{-\frac{3(\mathbf{x}_1-\mathbf{x}_5)^2}{2l\sigma}} = W_0^{-3}\, V,$$

Computing amplitudes (Gaussian approx)





$$\delta(\mathcal{W}) = \prod_{i=3,5,7} \delta(\mathbf{x}_1 - \mathbf{x}_i) \prod_{j=4,6,8} \delta(\mathbf{x}_2 - \mathbf{x}_j).$$

$$Z_{\mathcal{W}} = W_0^{-2} \int d\mathbf{x}_1 d\mathbf{x}_2 \, e^{-7 \cdot \frac{3(\mathbf{x}_2 - \mathbf{x}_1)^2}{2l\sigma}} = \frac{W_0^{-3} V}{7^{3/2}}$$

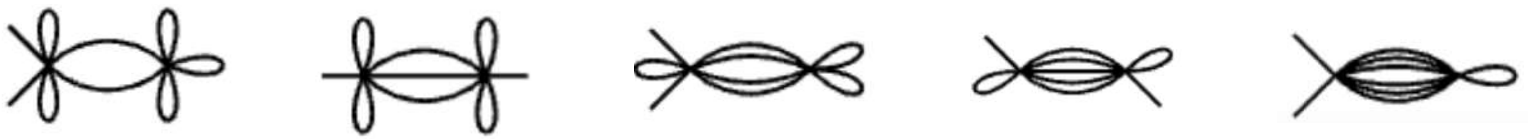
$$Z_{\mathcal{W}} = \frac{Z_{\mathcal{R}}}{7^{3/2}}$$

The topological weight of the watermelon is much smaller than that of rosette.

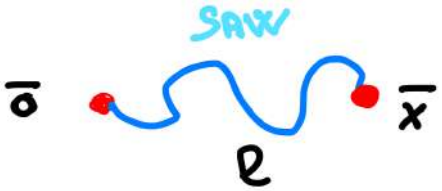
For generic hybrid rosette-watermelon configs with k=2 clusters and n_t ties

$$Z_{\mathcal{G}} = \frac{Z_{\mathcal{R}}}{n_t^{3/2}}$$

It explains why 2-cluster topologies with n_t ties are less frequent in simulations



Including (partially) self-avoidance



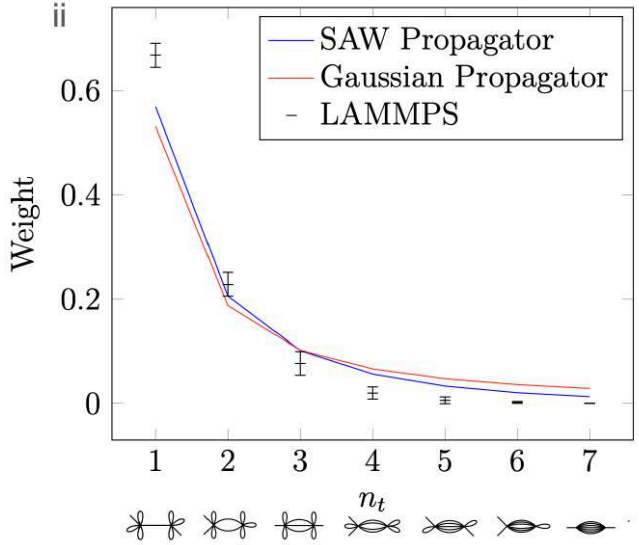
$$p(x \equiv |\mathbf{x}|; l) \sim \left(\frac{x}{R_F}\right)^{n_t g} e^{-n_t \left(\frac{x}{R_F}\right)^{\delta}}$$

$$R_F = R_F(\ell)$$
$$g = \frac{\gamma - 1}{\nu}$$
$$\delta = \frac{1}{1-\nu}$$

$$\frac{Z^{\mathrm{SA}}_{n_t}}{Z^{\mathrm{SA}}_{\mathcal{R}}} = \frac{\Gamma\left(\frac{n_t g+3}{\delta}\right)}{n_t^{\frac{n_t g+3}{\delta}} \Gamma\left(\frac{g+3}{\delta}\right)} \sim \frac{e^{-\frac{n_t g}{\delta}}}{\sqrt{n_t}}$$

$$\gamma \simeq 7/6 \qquad \nu \simeq 3/5$$

$$g \simeq 5/18 \qquad \delta \simeq 5/2$$

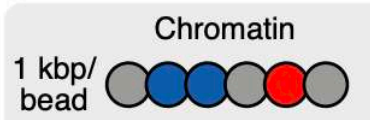


Excluded volume interactions sharpen the difference between the topological weights, rendering predictions closer to the values observed in simulations.

Q: Are these predictions confirmed by more realistic models of chromatin organisation ?

Highly Predictive Heteromorphic Polymer (HiP-HoP) model

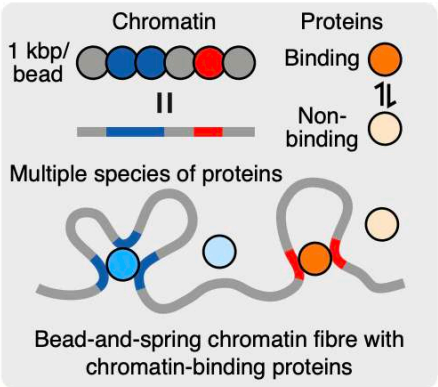
(Buckle et al Molecular Cell 72, 2018; Chiang et al bioRxiv 2022)



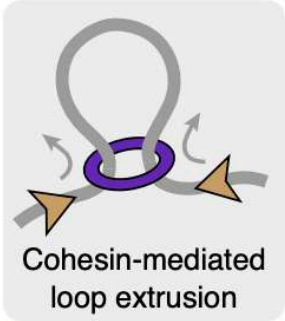
Chromosome region: bead-and-spring polymer.
Each bead represent a 1kbP of chromatin.

Three different mechanisms at work

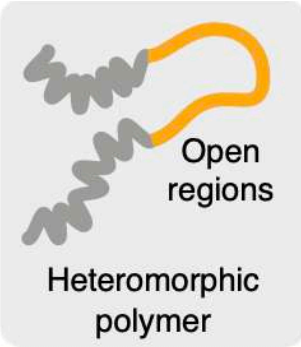
(i) Proteins (RNA polymerises, TFs) diffuse and interact multivalently with chromatin at specific binding sites (epigenetic marks)



(ii) Loop extrusion by cohesin & CTCF



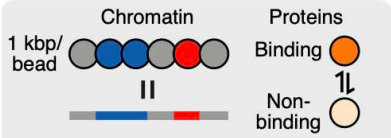
(iii) Heteromorphic polymer local variation in compaction

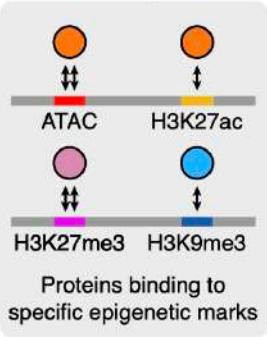


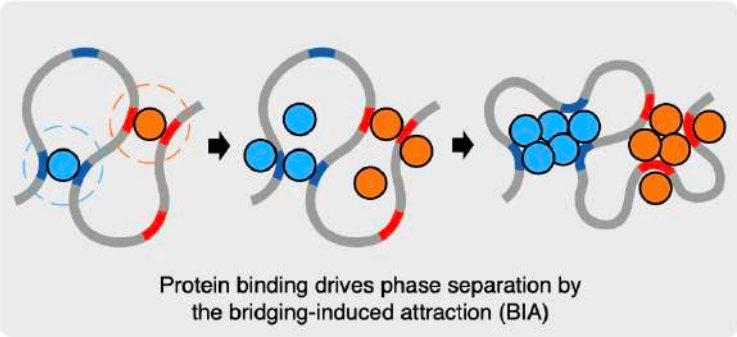
Highly Predictive Heteromorphic Polymer (HiP-HoP) model

(Buckle et al Molecular Cell 72, 2018; Chiang et al bioRxiv 2022)

(i) Multiple species of proteins diffusing and binding

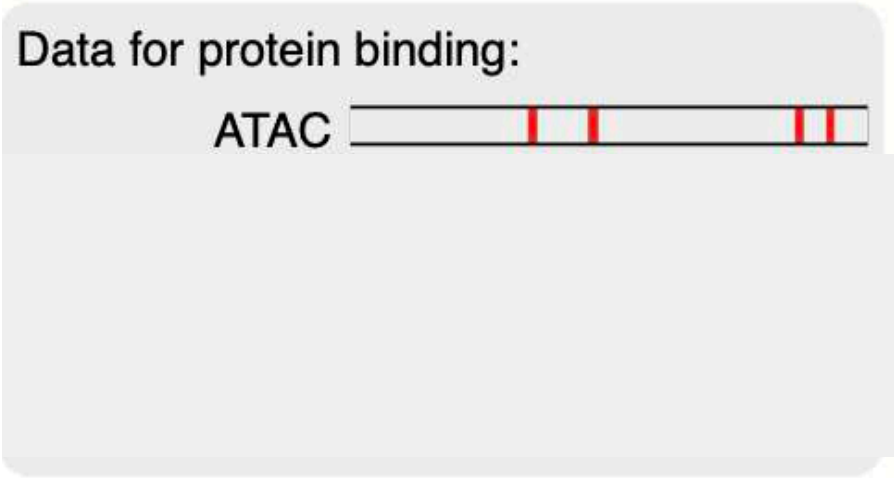






Multivalent binding:
Bridging induced attraction and microphase separation

Input data:



ATAC: **A**ssay for **T**ransposable **A**ccessible **C**hromatin
(determine chromatin accessibility). Based on the binding of TN5 transposes

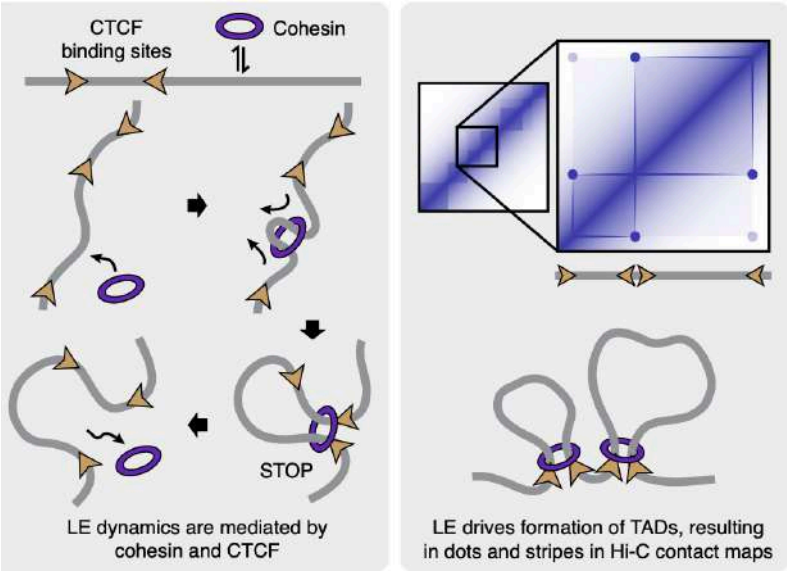
ATAC-seq peaks identifies accessible DNA regions (promoter and enhancers)

ChIP-seq: method to analyze protein-DNA interactions. Histone ChIP-seq pipeline (ENCODE)
H3K27ac (open or euchromatin)
H3K27me3, H3K9me3 (heterochromatic, transcriptional non-active regions)

Highly Predictive Heteromorphic Polymer (HiP-HoP) model

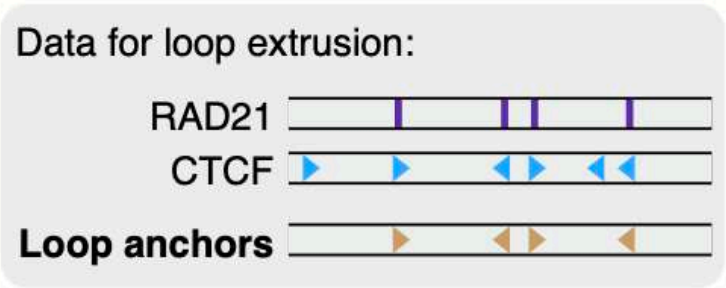
(Buckle et al Molecular Cell 72, 2018; Chiang et al bioRxiv 2022)

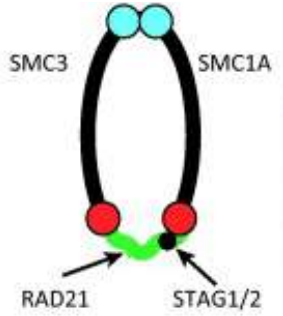
(ii) Loop extrusion mechanism mediated by cohesion complex and CTCF



Input data for cohesion and CTCF

Chip-seq analysis



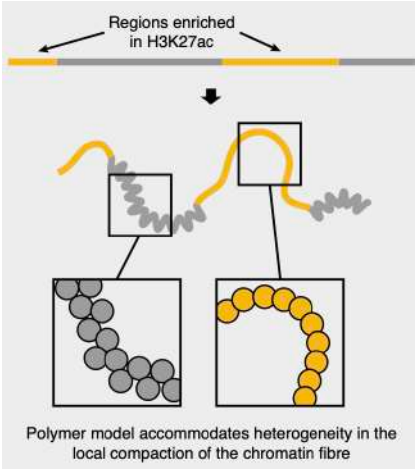


RAD21: sub-component of the **cohesin complex)**

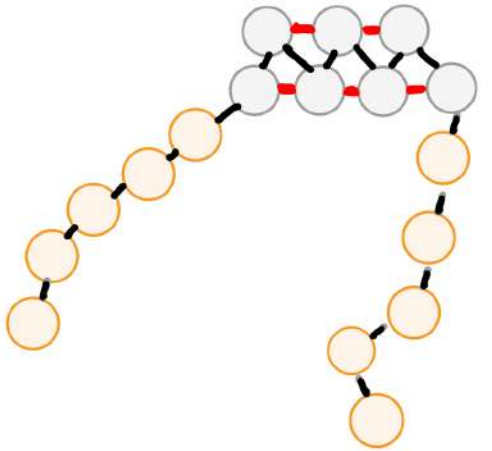
Highly Predictive Heteromorphic Polymer (HiP-HoP) model

(Buckle et al Molecular Cell 72, 2018; Chiang et al bioRxiv 2022)

(iii) Heteromorphic polymer It takes into account local variation in degree of compaction, active euchromatin vs inactive heterochromatin regions

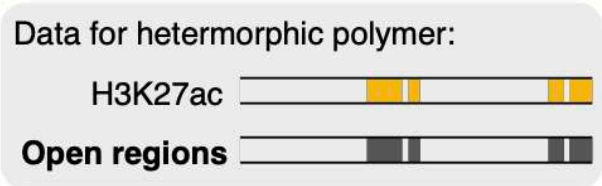


Model: add next-to-nearest harmonic spring



Input data:

ChIP-seq

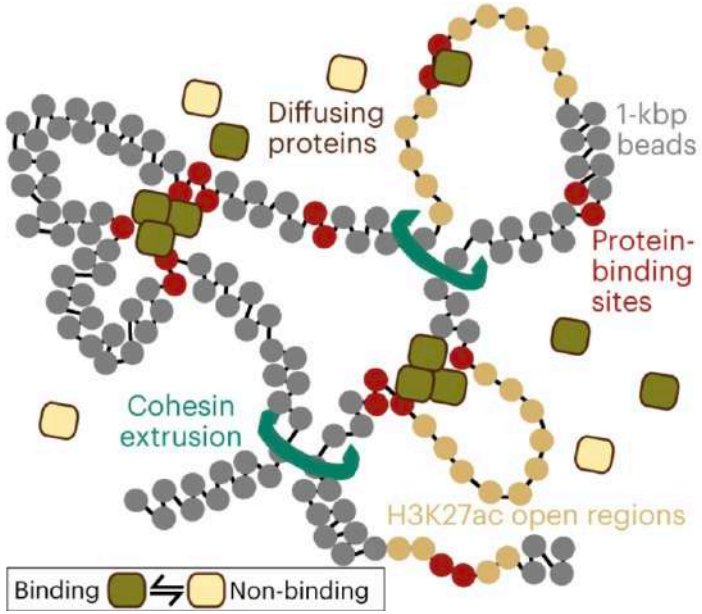


H3K27ac mark: associated with transcriptional active chromatin that typically adopts a more open and disrupted structure

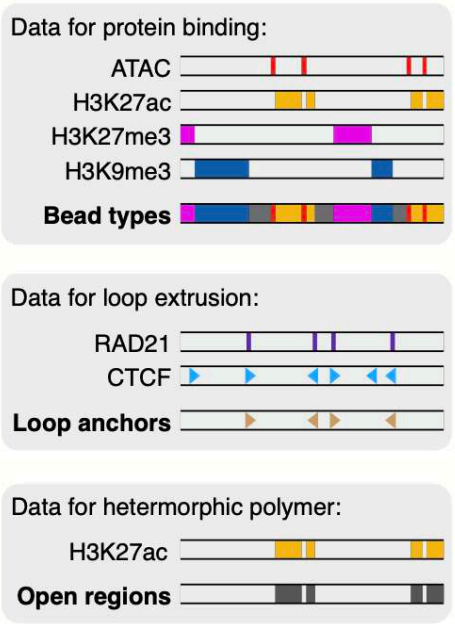
Highly Predictive Heteromorphic Polymer (HiP-HoP) model

(Buckle et al Molecular Cell 72, 2018; Chiang et al bioRxiv 2022)

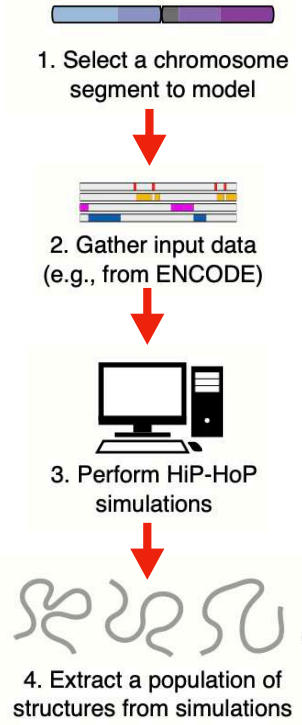
Model



Input data



Simulation workflow



Analysed two gene loci in lymphoblast cells:

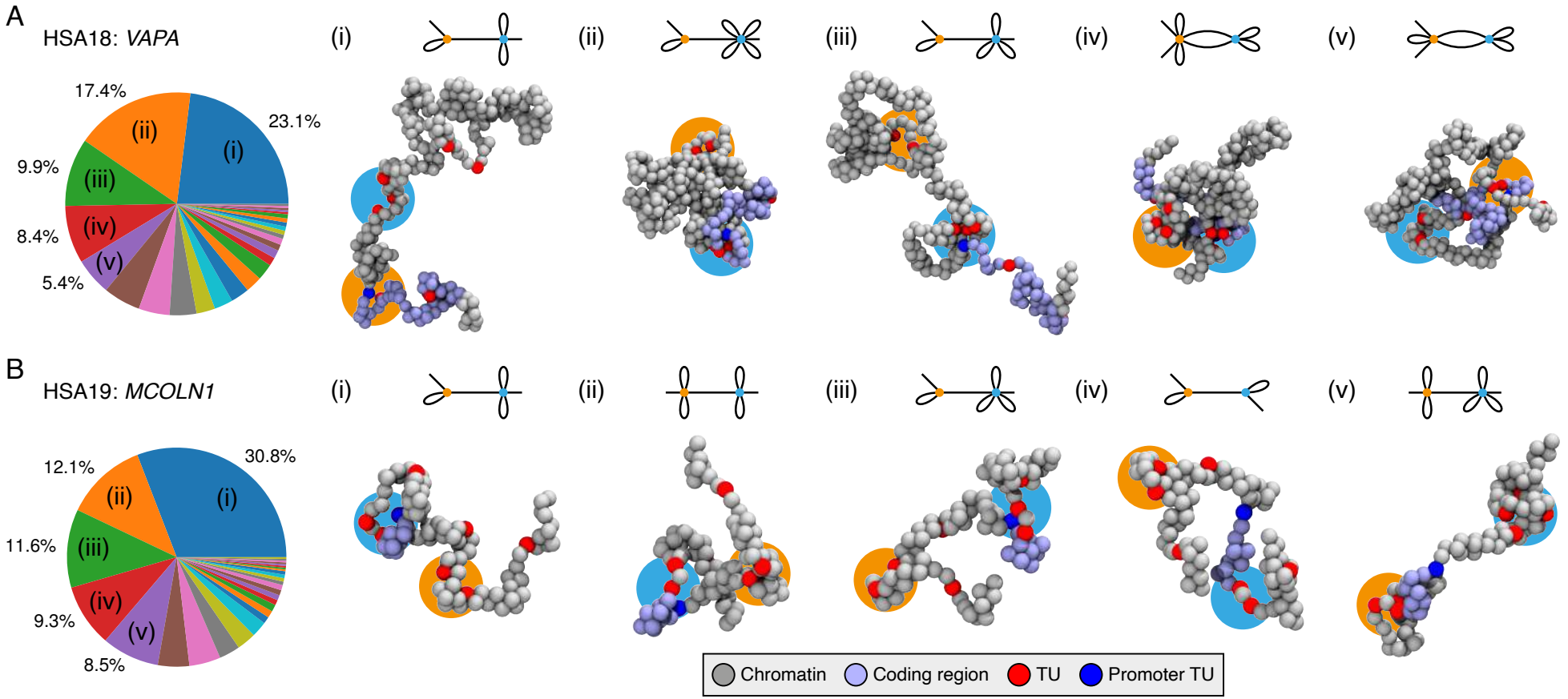
VAPA (in HSA 18, gene poor)

MCOLN1 (in HSA 19,gene rich)

Similar size and both with 8 highly interacting TUs

300 independent simulations per gene loci

TU partners are those that >10% interacting with the chosen TU



Rosette-like structures with local loops are dominant genome-wide

Conclusions

Studied topological spectrum of chromatin loop networks that emerge from 3D folding of polymer models of chromatin

Theory of partition provides a general way to enumerate the topologies of labelled networks. Counting unlabelled ones is more challenging (k=2)

Out of all possible topologies only a small fraction is largely frequent

Prediction:

gene loci are overwhelmingly organised in rosette-like structures

Dominance of rosettes is also found in more realistic models of human chromatin (HiP-HoP) where ATP-regulated loop extrusion, heteromorphic structures and epigenetic marks are considered.

Joyeux anniversaire

Emmanuel !!