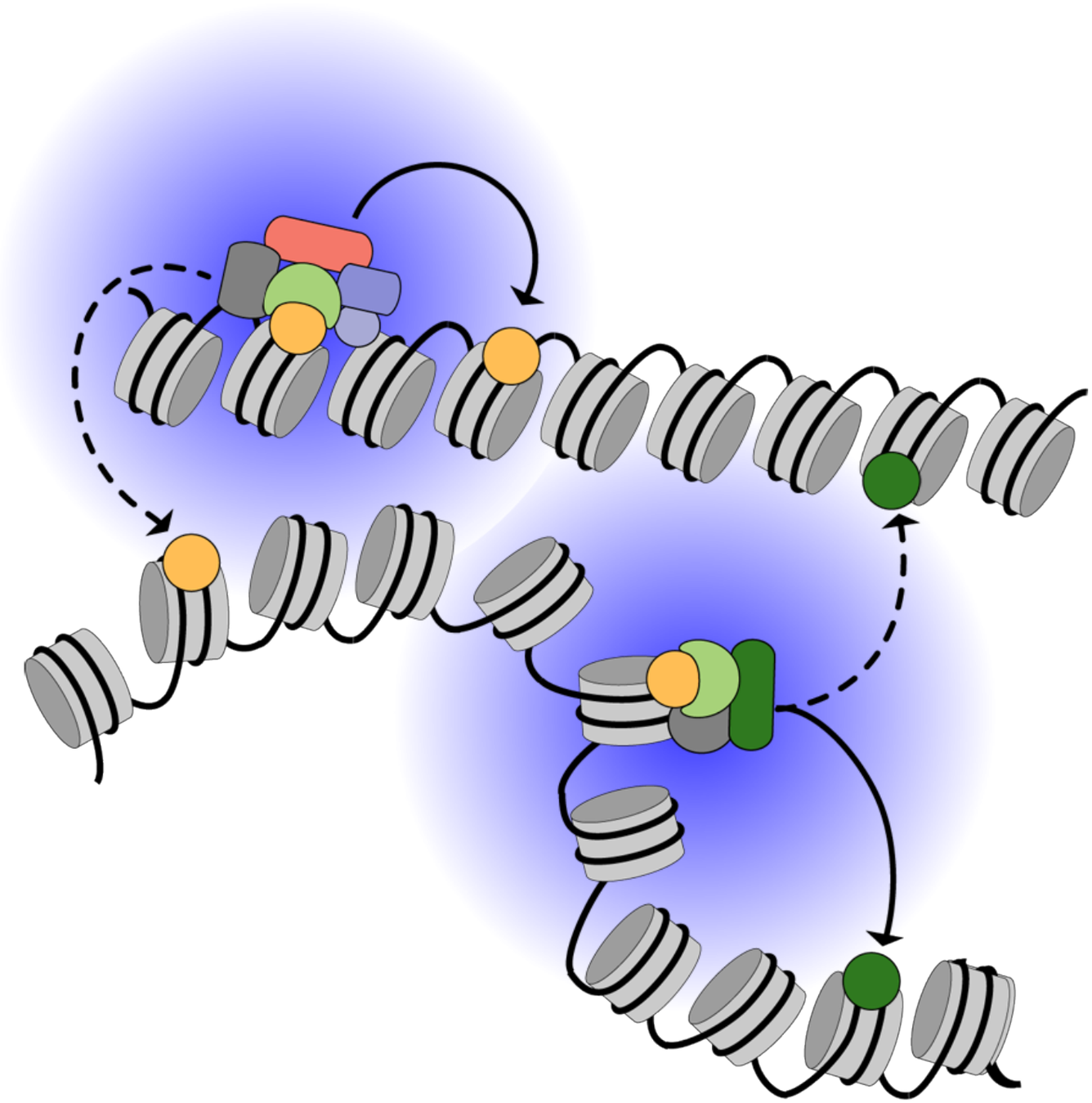
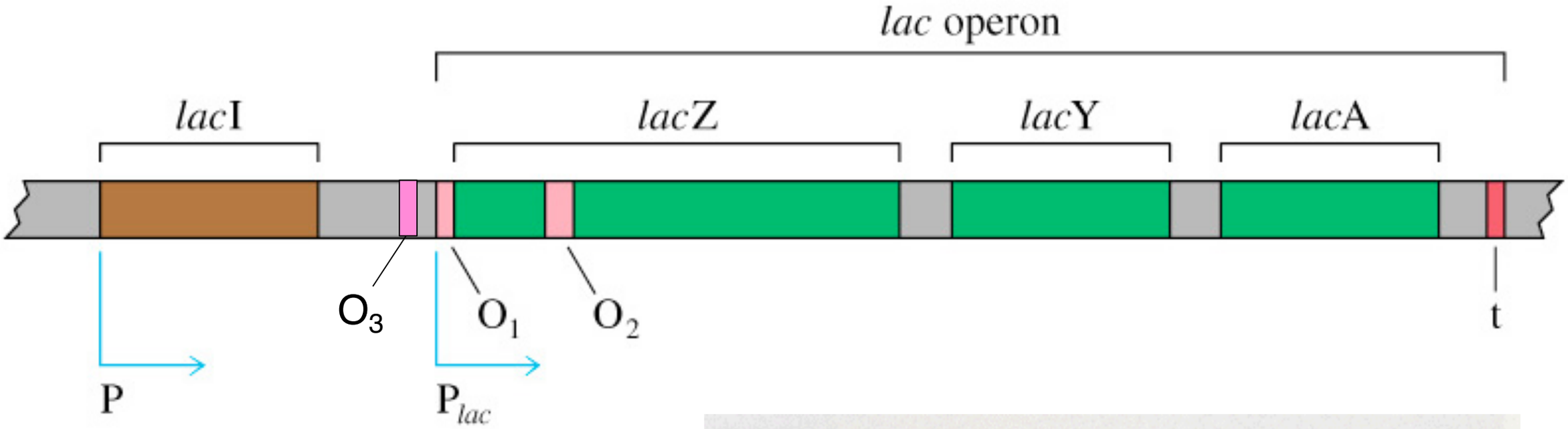
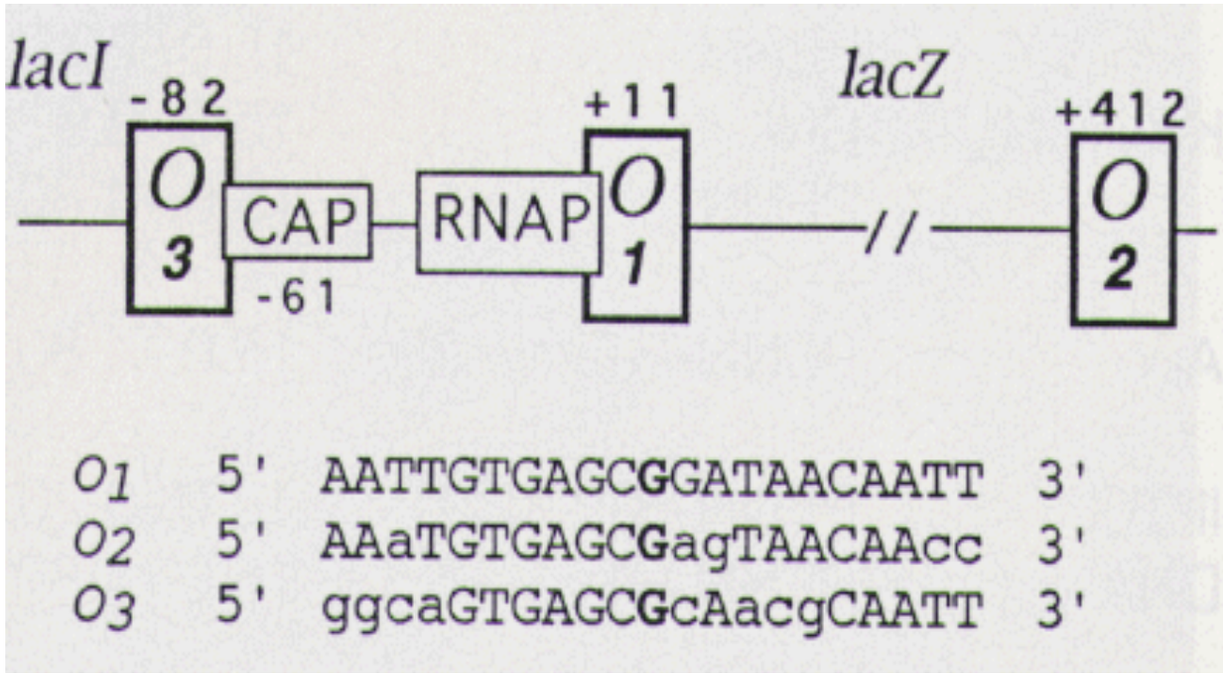
Interaction of DNA-bound proteins via DNA-looping

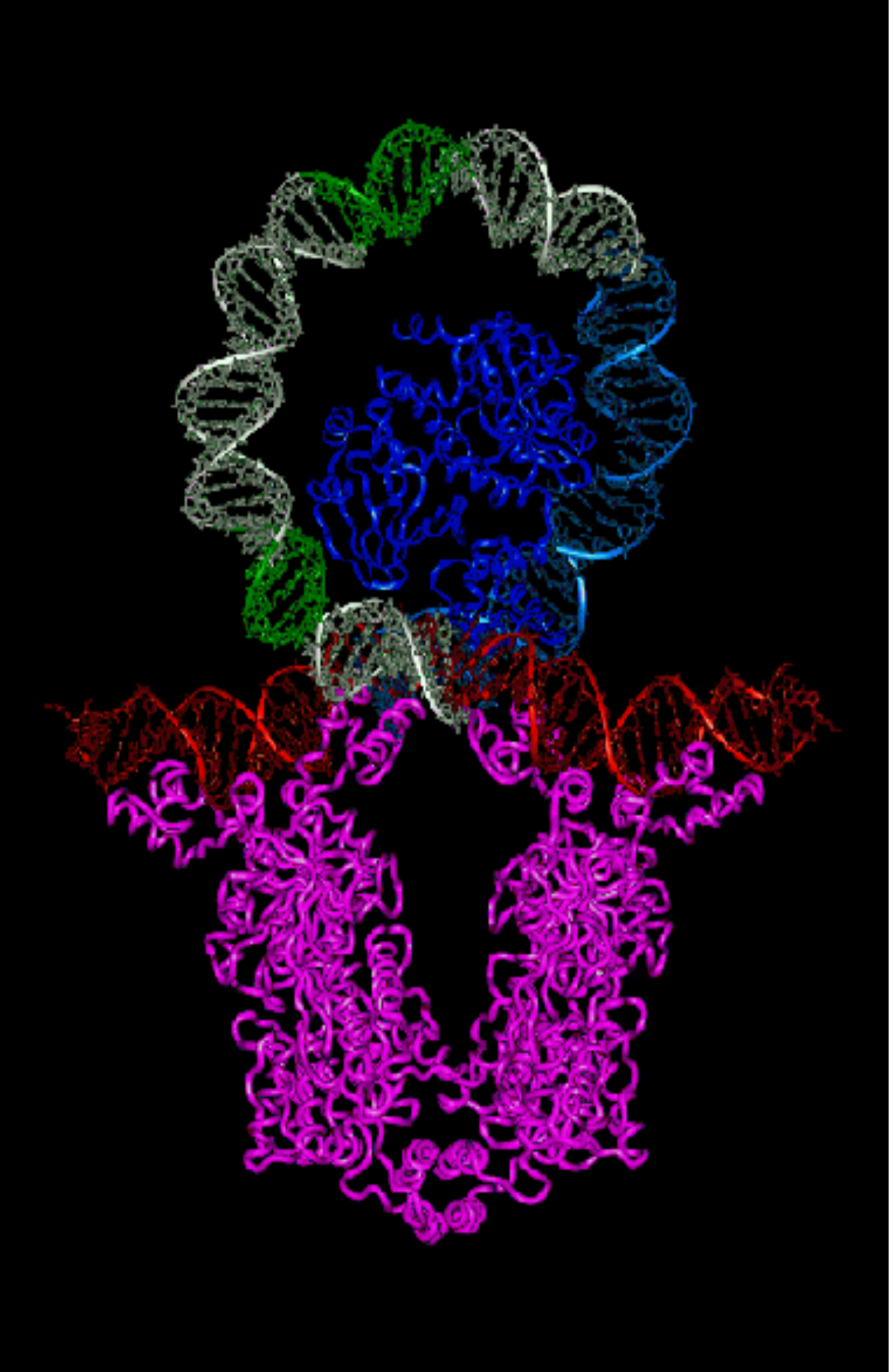


Organization of the genes regulated by Lac repressor, a transcription repressor protein in the bacterium E. coli



Lac repressor binds to the operators O_1, O_2 and O_3



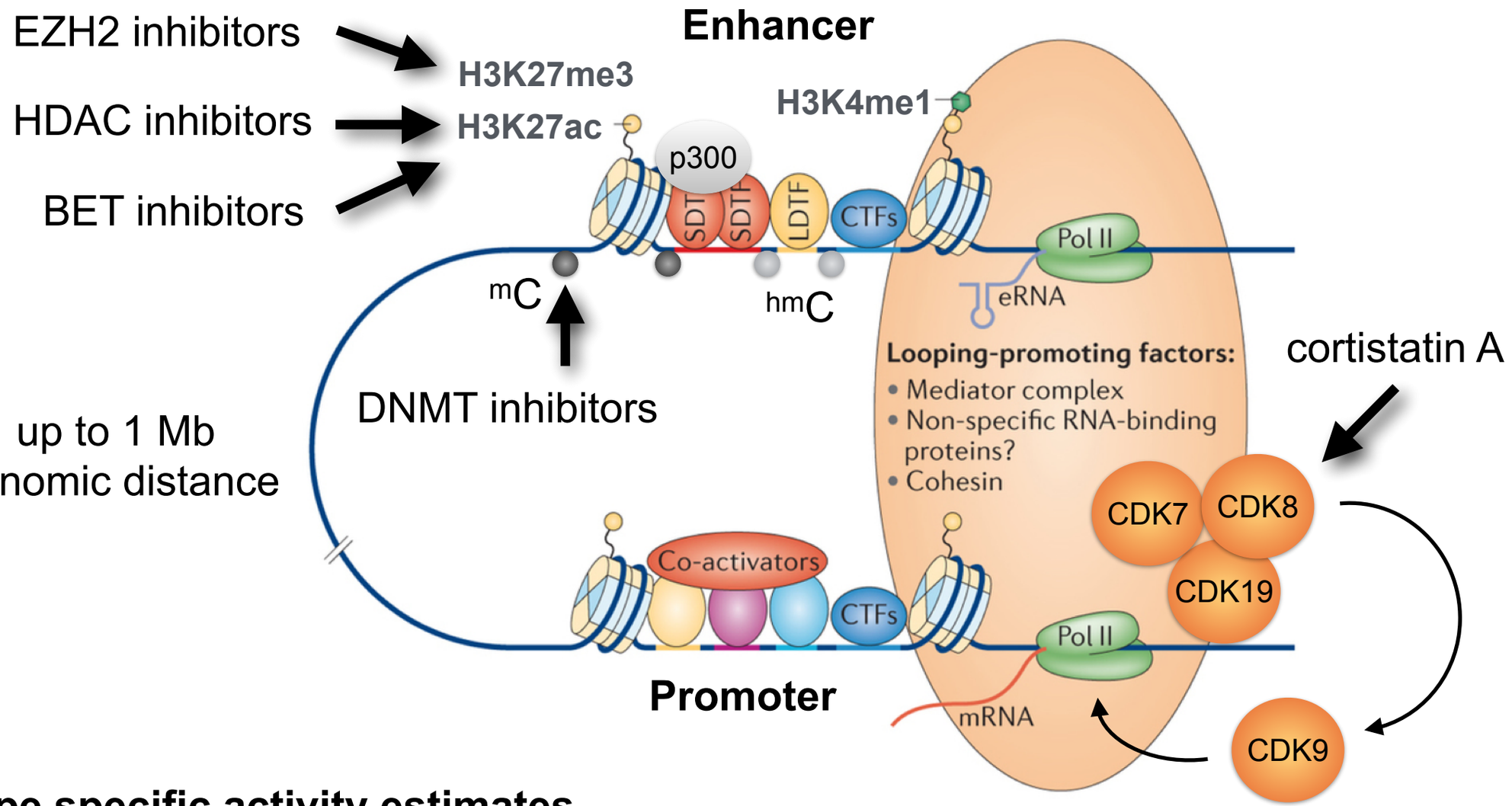


low glucose and low lactose
=> both CAP and LacI are bound => repression

⟵ CAP

⟵ Lac repressor bound to O1 and O3

More than **300 000** putative enhancers regulate **~54 000** annotated human genes (including lncRNAs)

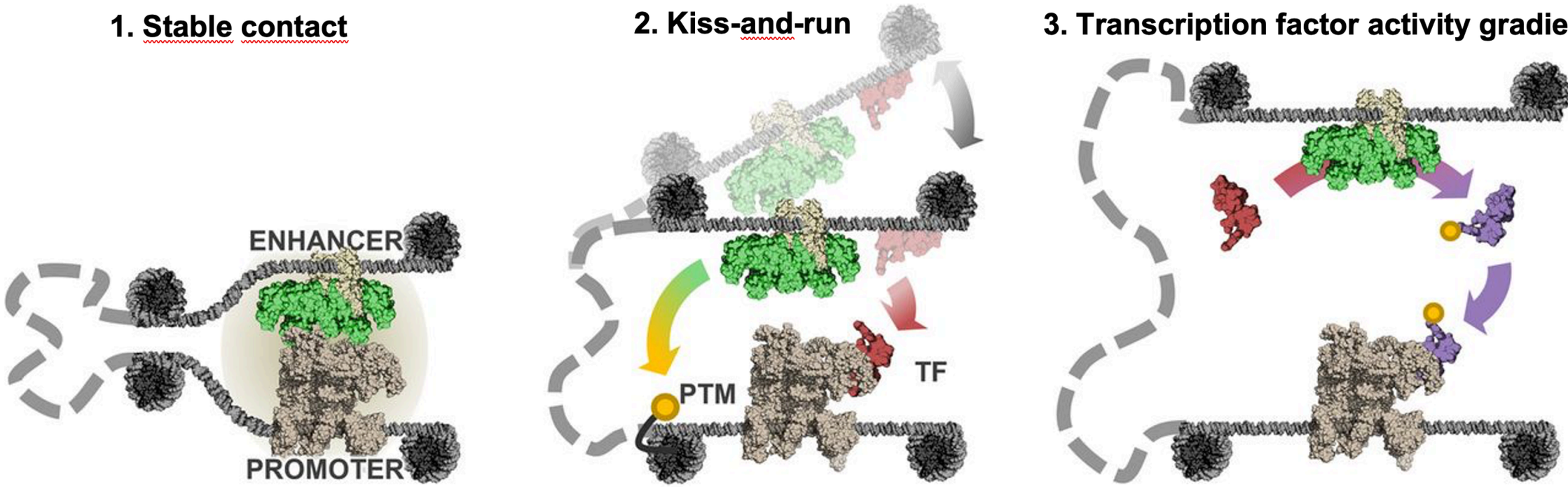


Cell type specific activity estimates

- ~20 000 active genes
- 80 000 - 240 000 active enhancers
- typical: 1-2 target promoter per enhancer
- ~10 different targets for some enhancers
- multiple enhancers for single promoter
- 300-500 super enhancers > 10 kb

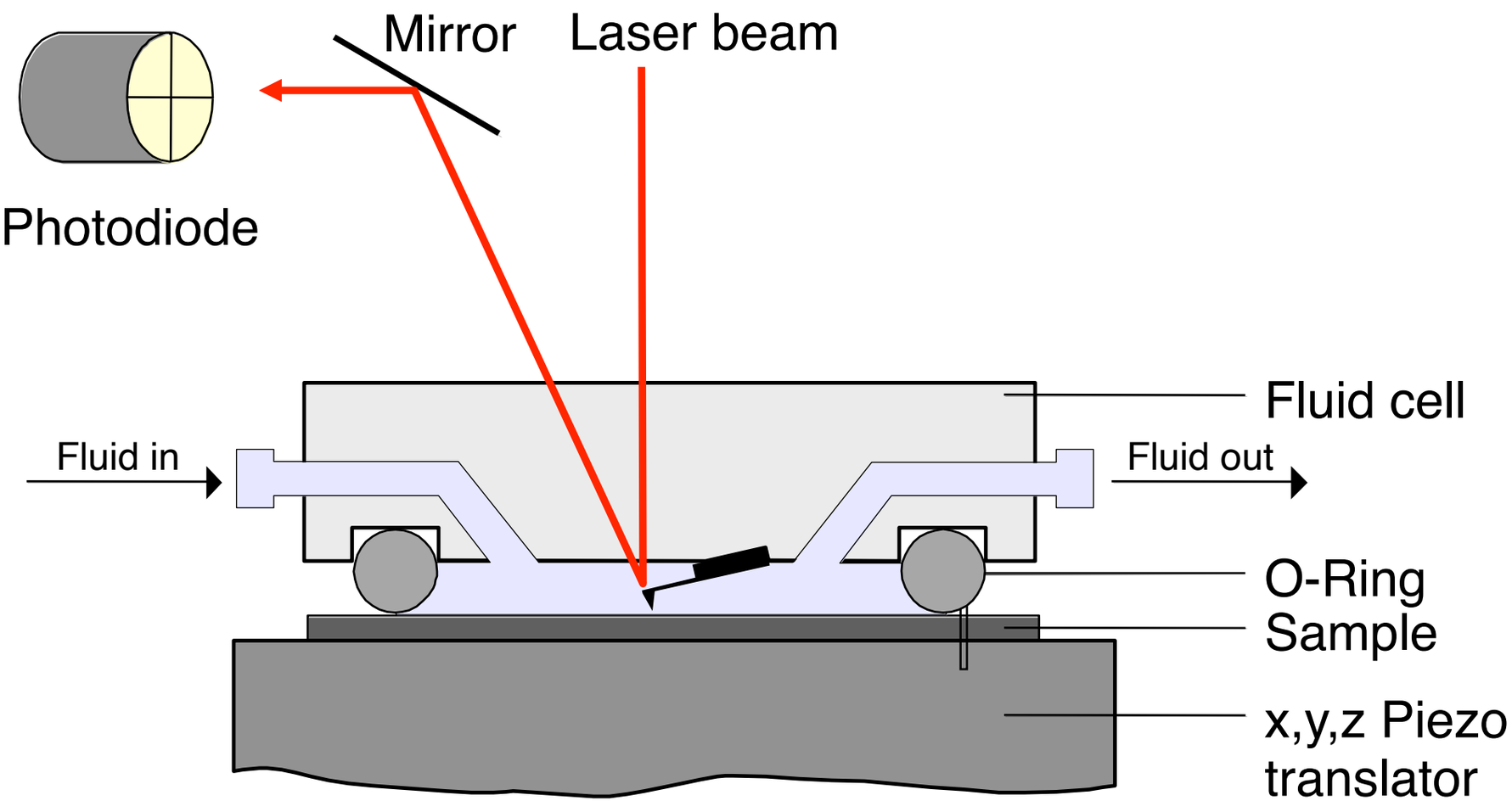
Heinz 2015, Nat Rev Mol Cell Biol
Roadmap Epigenomics Consortium 2015, Nature
FANTOM Consortium 2015, Nature

How do enhancers activate transcription?

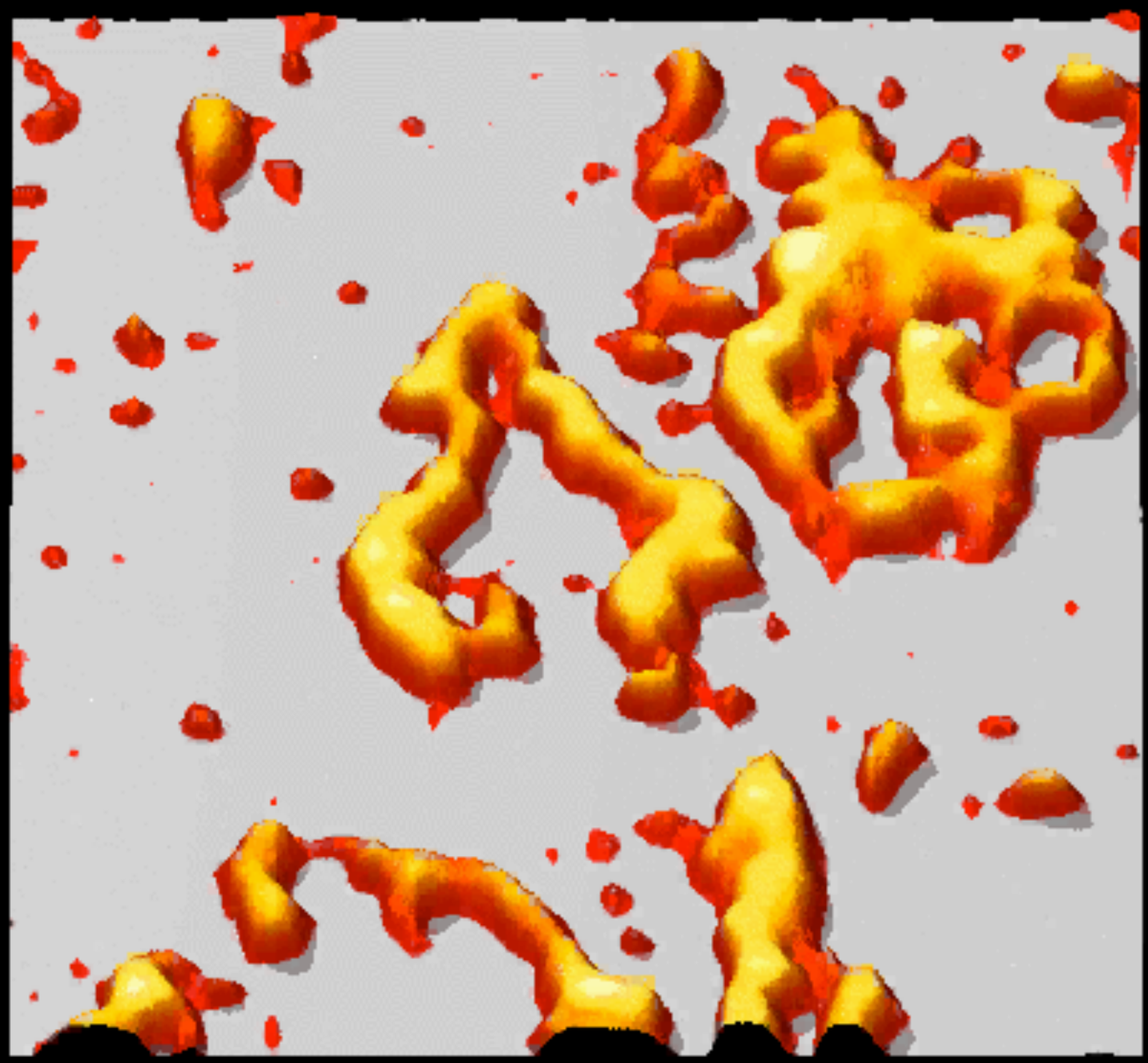


Karr et al. 2022, Genes & Dev 36, 7-16, https://doi.org/10.1101/gad.349160.121

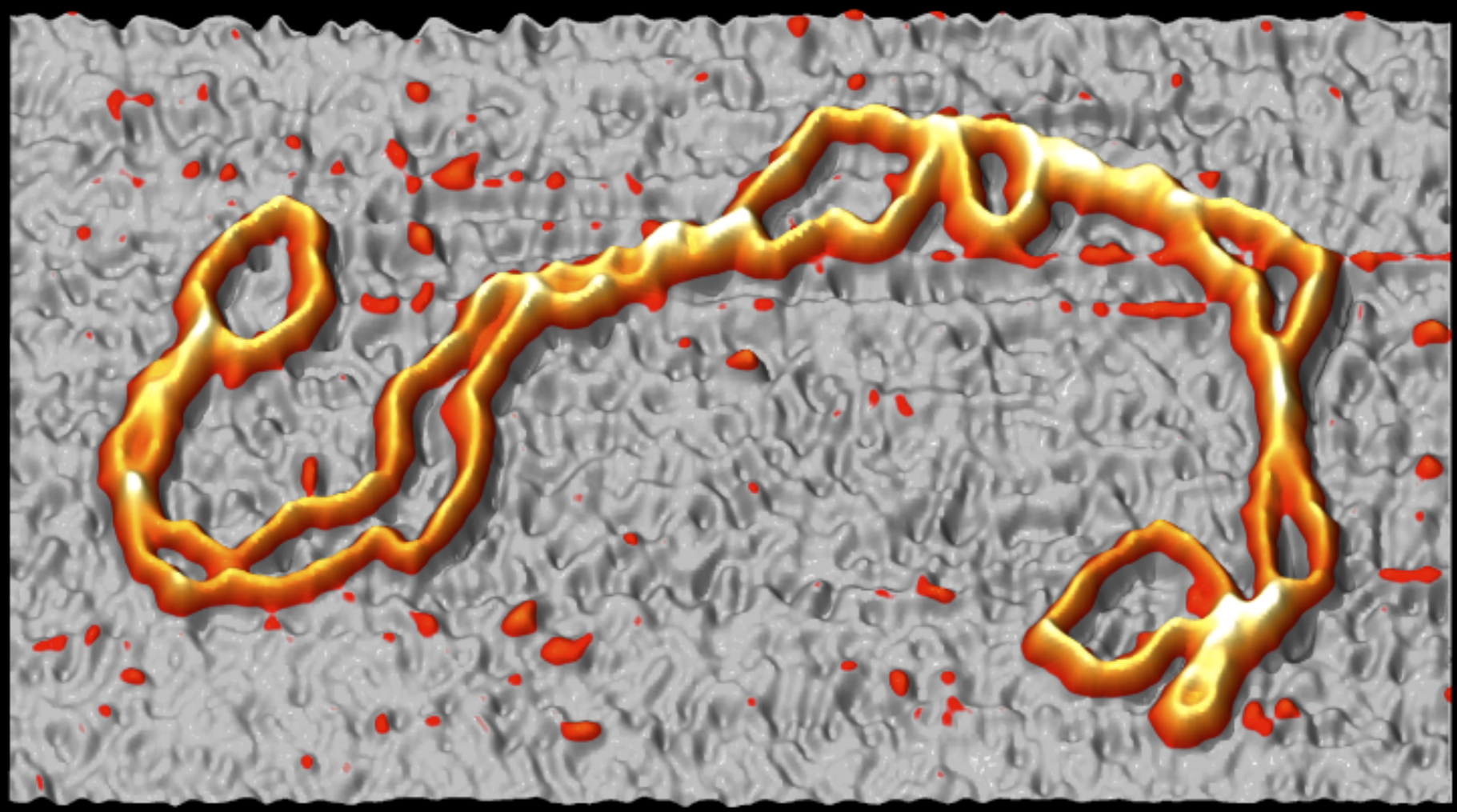
Scanning force microscopy (SFM) of DNA and protein-DNA complexes in air and in buffer solutions



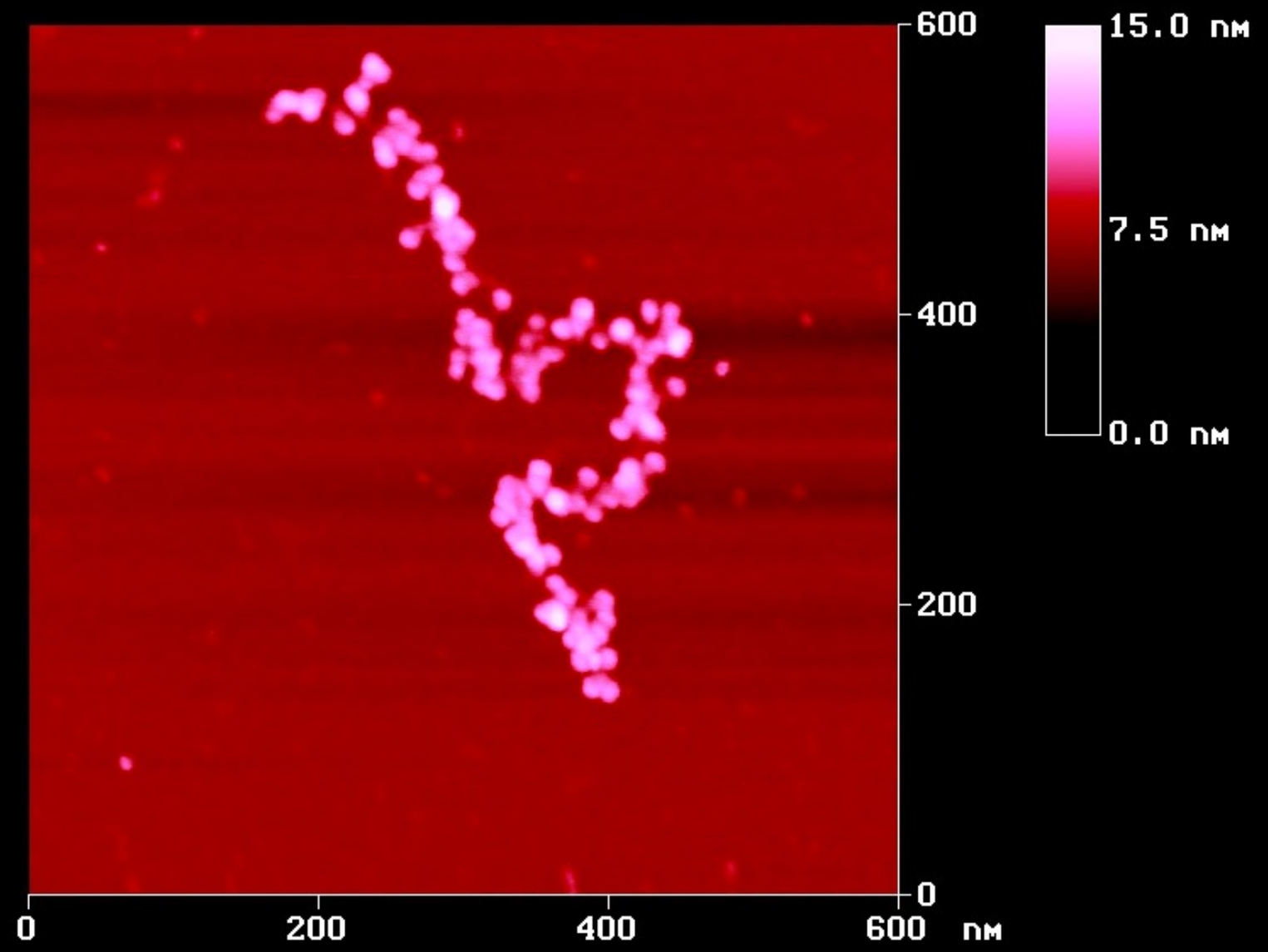
Movement of a DNA fragment on a mica surface visualized by scanning force microscopy (SFM/AFM)



SFM image of a 6.8 kb superhelical plasmid

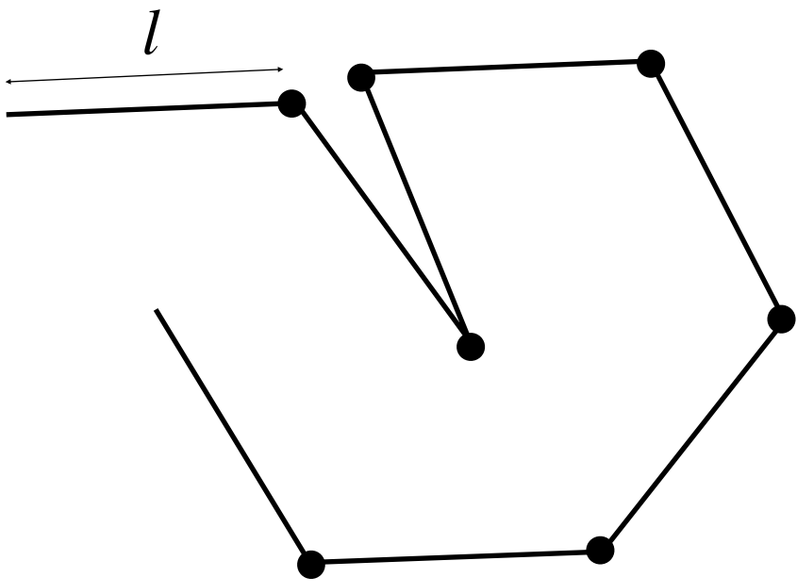


3. Looping of a chromatin fiber - the conformation of yeast chromosome III during interphase

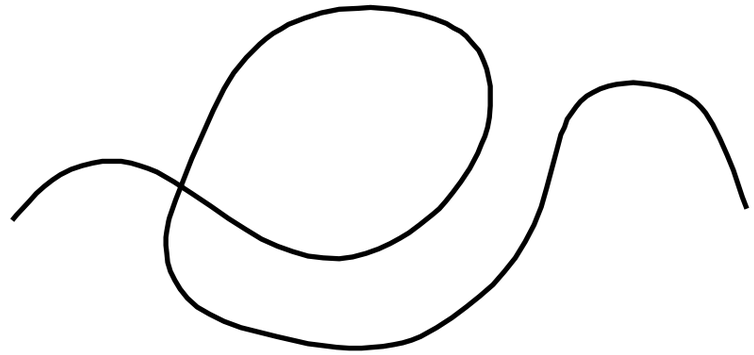


Theoretical description of DNA using the model of a freely jointed chain (FJC-chain) or the model of a elastic rod

statistical segment
or *Kuhn* length l



Freely Jointed Chain model (FJC) for
polymers ≥ 6 *Kuhn* segments



Kratky-Porod (KP) or worm-like chain
for polymers of all length

Numerical computer simulations
(Monte-Carlo, Brownian dynamics)

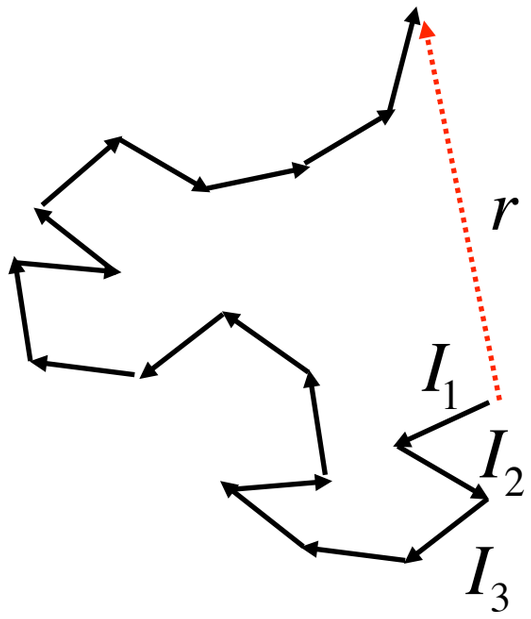
The freely joined chain model is similar to a random walk in three dimensions

The chain consists of *n* segments with length *l* and the end-to-end vector *r.*

$$\langle r^2 \rangle = r \cdot r = \left(\sum_{j=1}^{n} I_j \right) \cdot \left(\sum_{k=1}^{n} I_k \right) = \sum_{j=1}^{n} \sum_{k=1}^{n} I_j \cdot I_k$$

$$\langle r^2 \rangle = n l^2 + 2 \sum_{j>k} I_j \cdot I_k$$

$$\langle r^2 \rangle = n l^2$$



$$r = \sum_{i=1}^{n} I_i$$

average end-to-end distance:

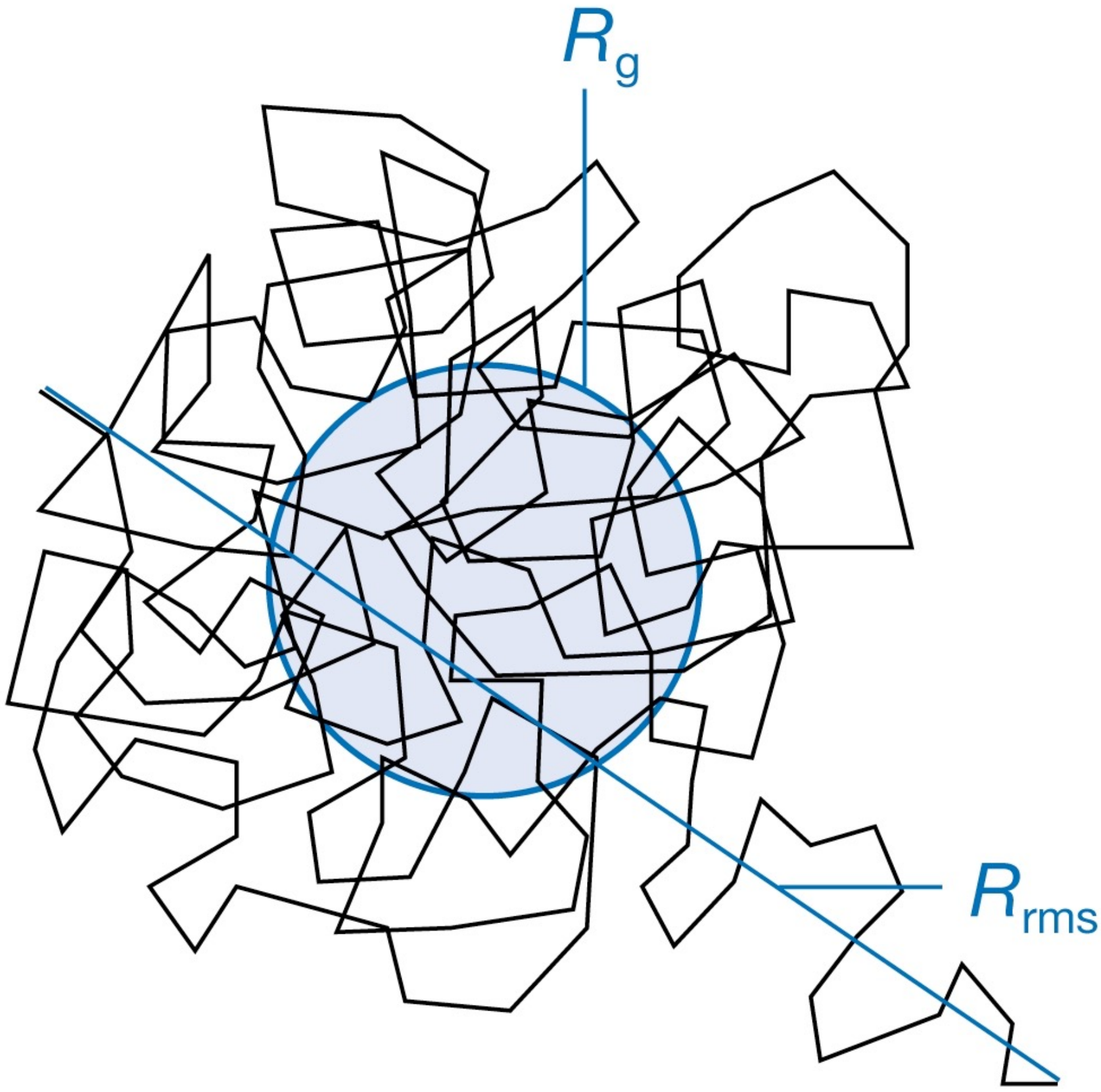
$$\sqrt{\langle r^2 \rangle} = \sqrt{n}\, l$$

radius of gyration:

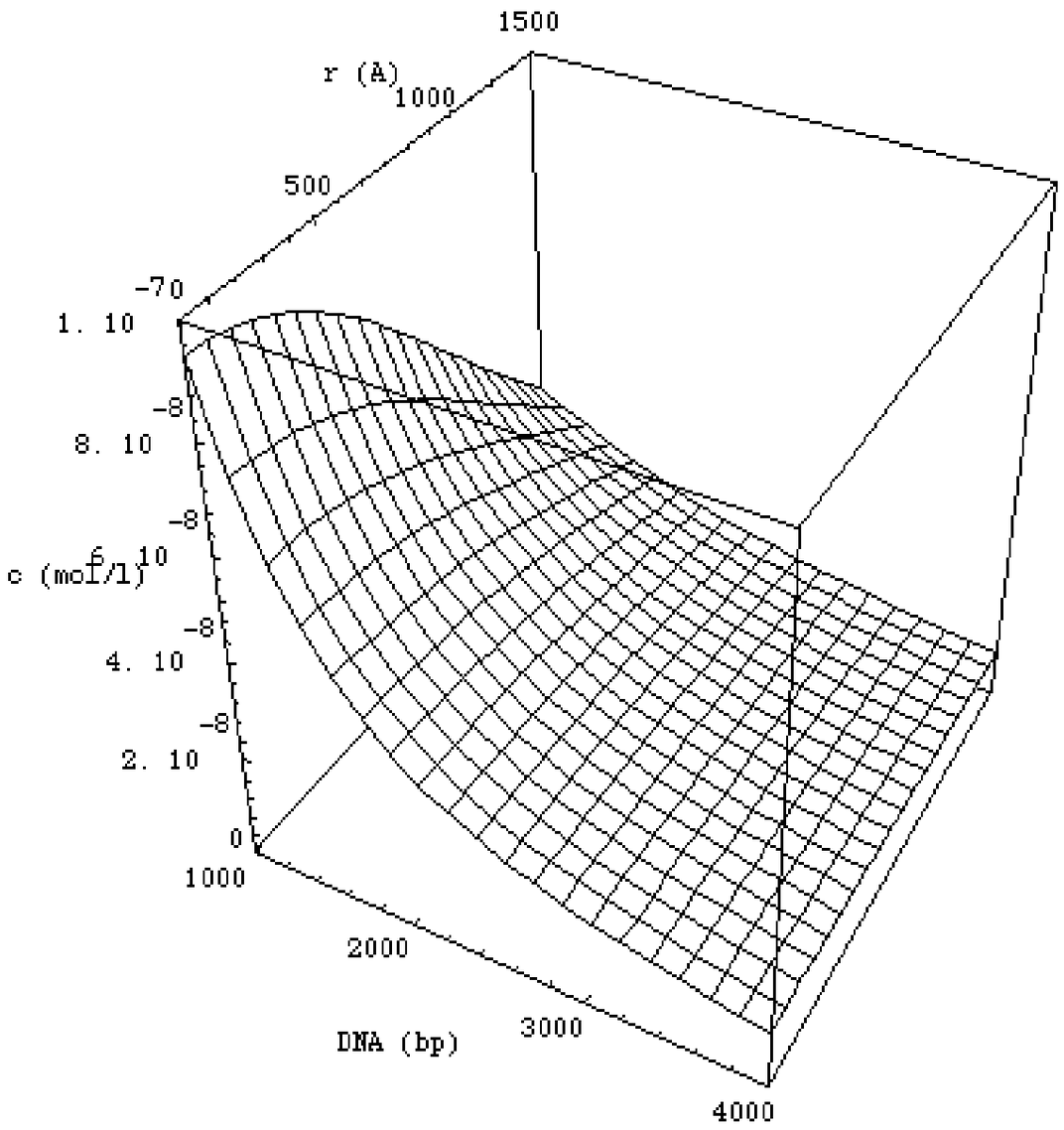
$$\sqrt{\langle R^2 \rangle} = \sqrt{\frac{n l^2}{6}}$$

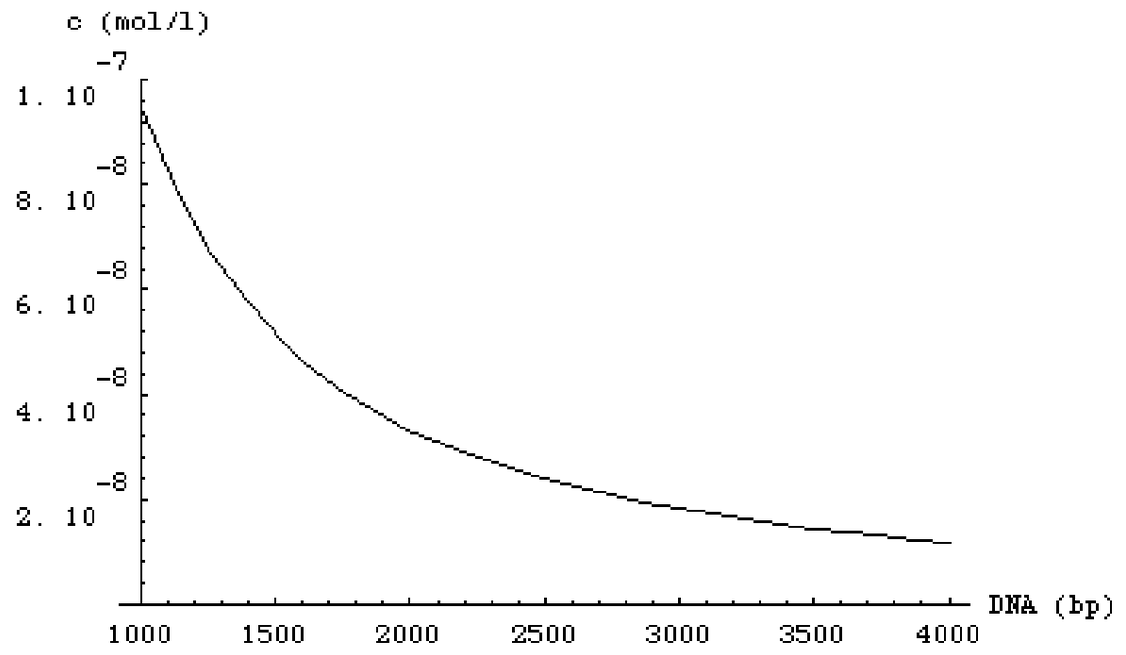
• The average end-to-end distance of a polymer chain as well as the radius of gyrations *R* is proportional to the square root of the chain length.

• The statistical segment length *l* reflects the stiffness of the polymers. For DNA at physiological salt concentrations it is *l* = 100 nm (300 base pairs).


R_g
R_{rms}

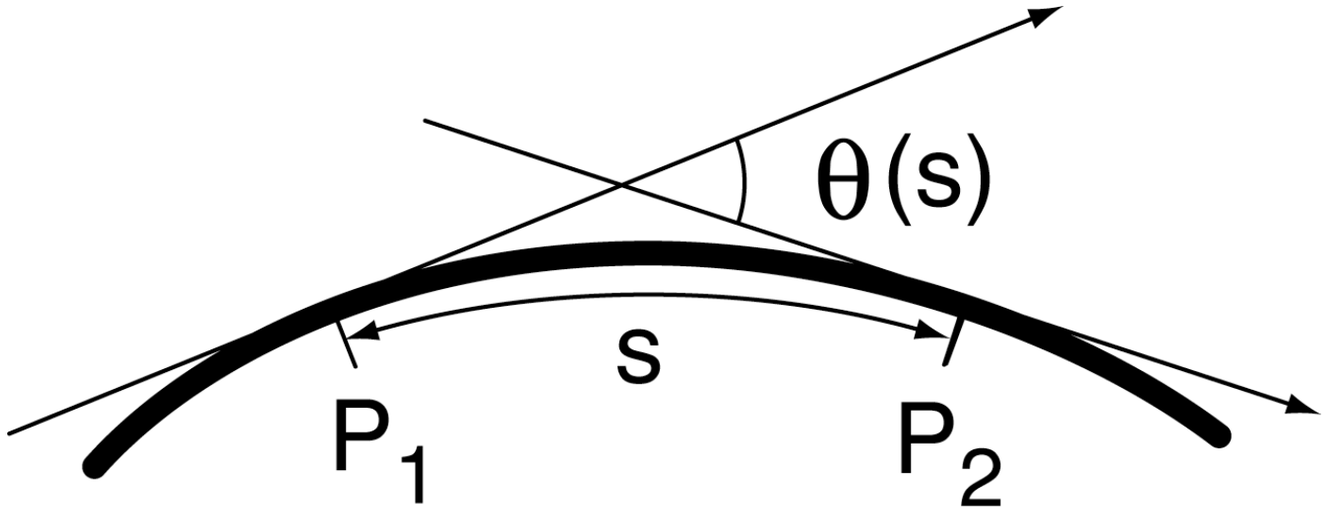
The local concentration function in dependence of r for the description of DNA acording to the freely jointed chain model





for small r

Theoretical description of a polymer using the model of an elastic rod or Kratky-Porod (KP) chain



$$\langle \cos\ \theta(s) \rangle = \exp\left(-\frac{s}{a}\right)$$

definition of persistence length ***a***

$$l = 2 \times a$$

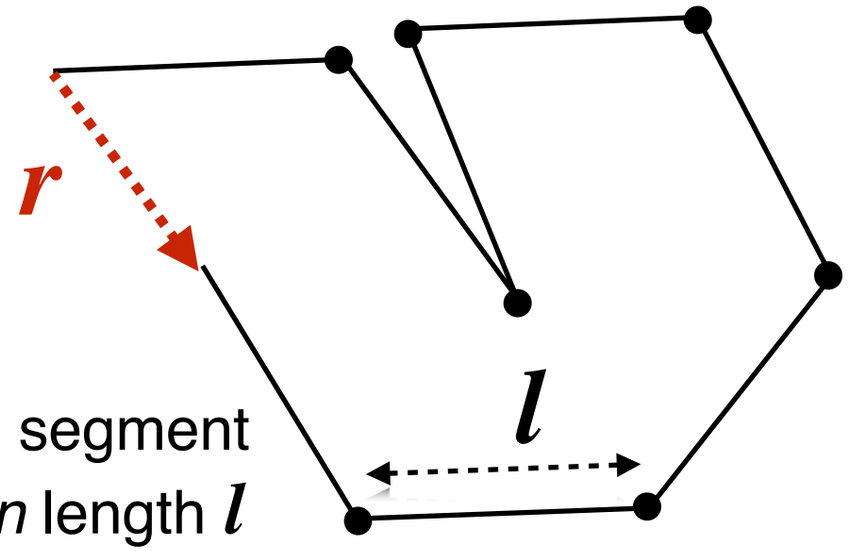
relation of *a* to *Kuhn* length l

$$\langle r^2 \rangle = 2a\ L_C \left[1 - \frac{a}{L_C}\left[1 - \exp\left(-\frac{L_C}{a}\right)\right]\right]$$

average squared end-to-end-distance $\langle r^2 \rangle$ for a KP-chain of contour length L_c

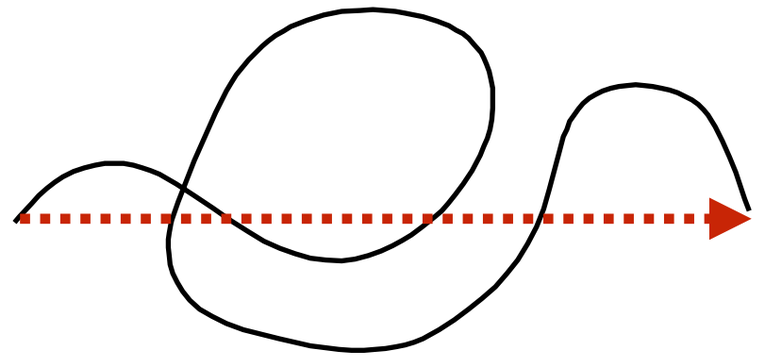
Theoretical description of DNA using the model of a freely jointed chain (FJC-chain) or the model of a elastic rod

Freely Jointed Chain model (FJC) for polymers with $n \geq 2$-3 segments



statistical segment with *Kuhn* length $\boldsymbol{l}$

Kratky-Porod (KP) or worm-like chain for polymers of contour length L_c



Persistence length $\boldsymbol{a}$

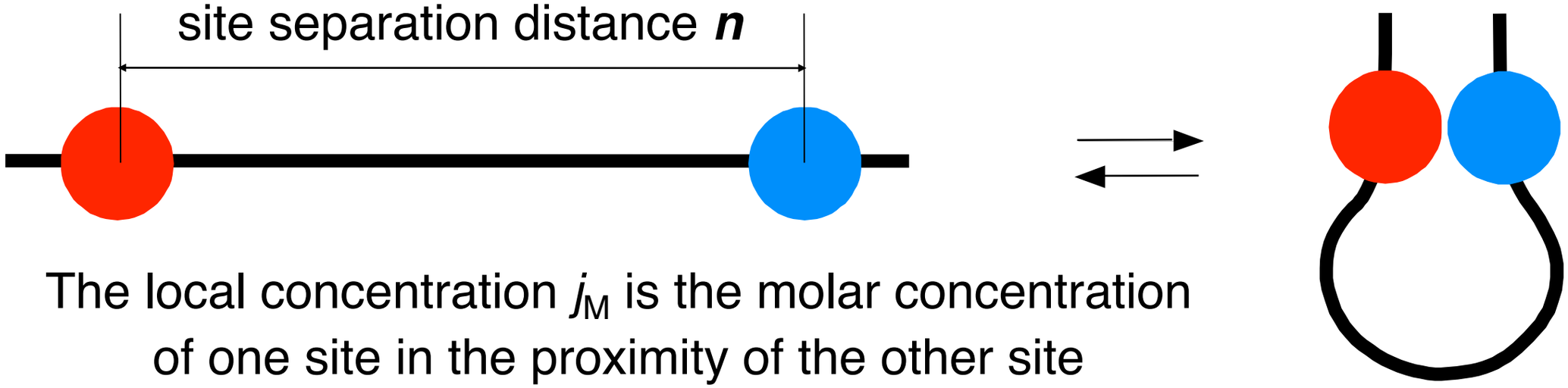
$$\langle r^2 \rangle = n l^2$$

$$\langle r^2 \rangle = 2a\, L_C \left[1 - \frac{a}{L_C} \left[1 - \exp\left(-\frac{L_C}{a} \right) \right] \right]$$

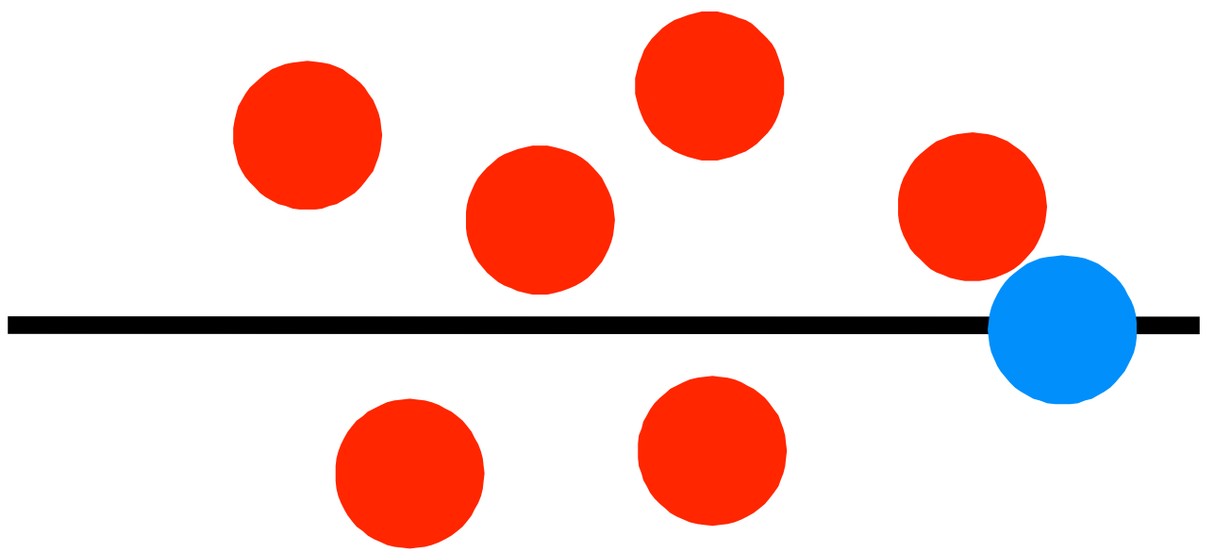
$$l = 2 \times a$$

relation of persistence length $\boldsymbol{a}$ to *Kuhn* length $\boldsymbol{l}$

The local molar concentration j_M of a protein in the proximity of another DNA bound protein

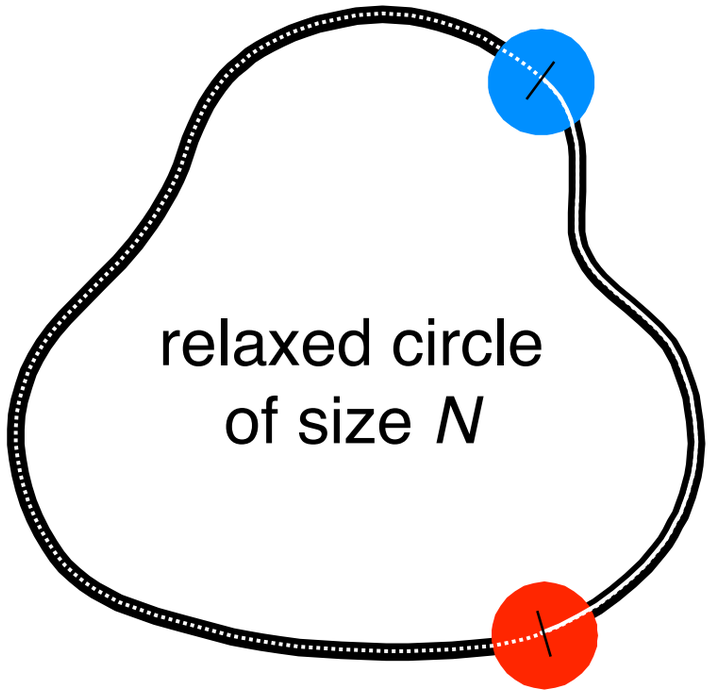


The local concentration j_M is the molar concentration of one site in the proximity of the other site

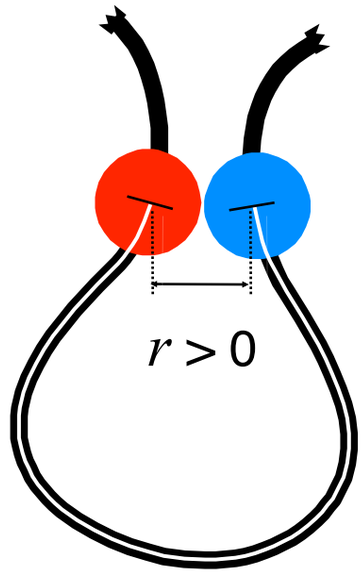


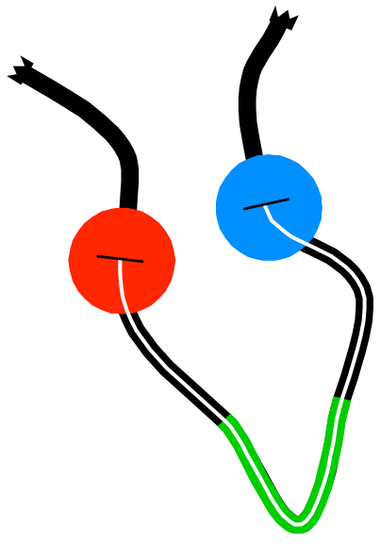
Interaction "*in trans*" without a nucleic acid linker

Interaction of two sites with a separation distance and a reaction distance ***r***

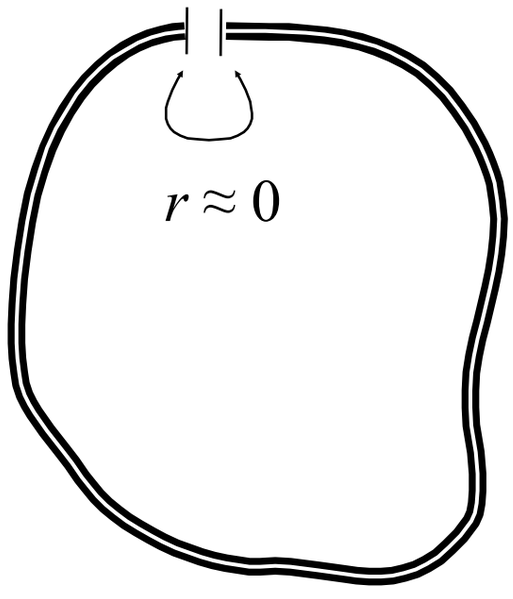


protein-protein contacts





intrinsically curved regions



cyclization of ends

The Gaussian distribution function for the freely jointed chain

$$W(r) = \frac{\left(\frac{3}{2\pi \cdot n \cdot l^2}\right)^{\frac{3}{2}}}{\exp\left(\frac{3r^2}{2\, n \cdot l^2}\right)} \quad \text{or} \quad W(r) = \frac{\left(\frac{3}{2\pi \cdot n}\right)^{\frac{3}{2}}}{\exp\left(\frac{3r^2}{2 l^2 n}\right) \cdot l^3} \quad \text{or} \quad W(r) = \frac{\left(\frac{3}{2\pi \left\langle r^2 \right\rangle}\right)^{\frac{3}{2}}}{\exp\left(\frac{3r^2}{2\left\langle r^2 \right\rangle}\right)} \quad \text{with} \quad \left\langle r^2 \right\rangle = n \cdot l^2$$

for small r the exponential term is ~ 1

$$W(r) = \left(\frac{3}{2\pi \cdot n \cdot l^2}\right)^{\frac{3}{2}} \quad \text{or} \quad W(r) = \left(\frac{3}{2\pi \cdot n}\right)^{\frac{3}{2}} \cdot l^{-3} \quad \text{or} \quad W(r) = \left(\frac{3}{2\pi \left\langle r^2 \right\rangle}\right)^{\frac{3}{2}}$$

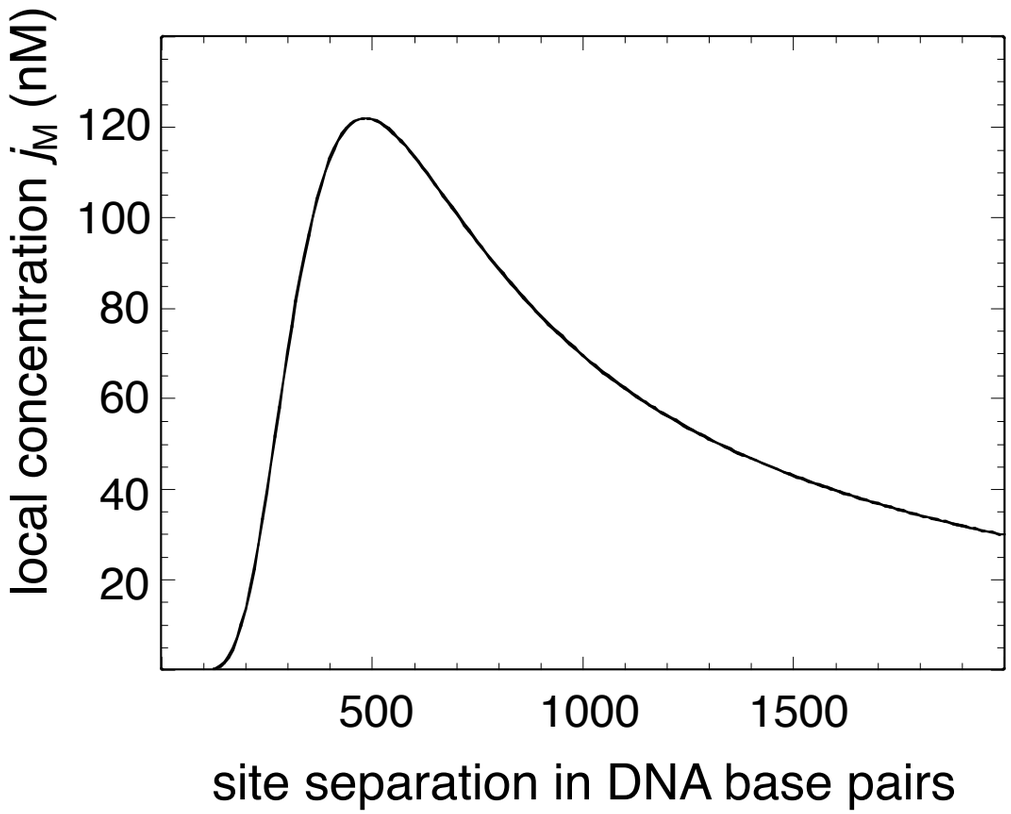
$$j_{\mathrm{M}}(r) = \frac{\left(\frac{3}{2\,\pi}\right)^{\frac{3}{2}}}{l^3\; n^{\frac{3}{2}}} \left[\frac{10^{27}\,\text{Å}^3\,\mathrm{M}}{6.022\text{e}23}\right]$$

and for l = 100 nm and 0.34 nm/bp

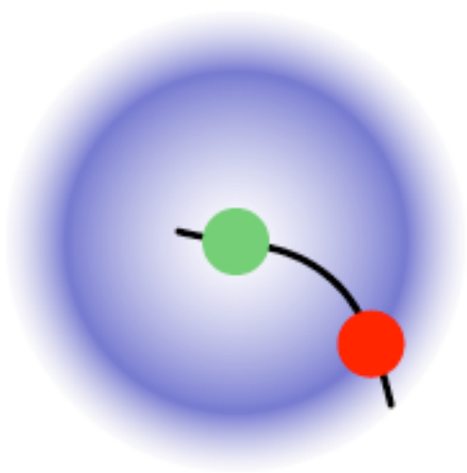
$$j_{\mathrm{M}}(bp) = 0.0028 \cdot bp^{-\frac{3}{2}} \left[\frac{\mathrm{mol}}{\mathrm{liter}}\right]$$

The local concentration j_M between two sites on double-stranded DNA depends on their separation distance

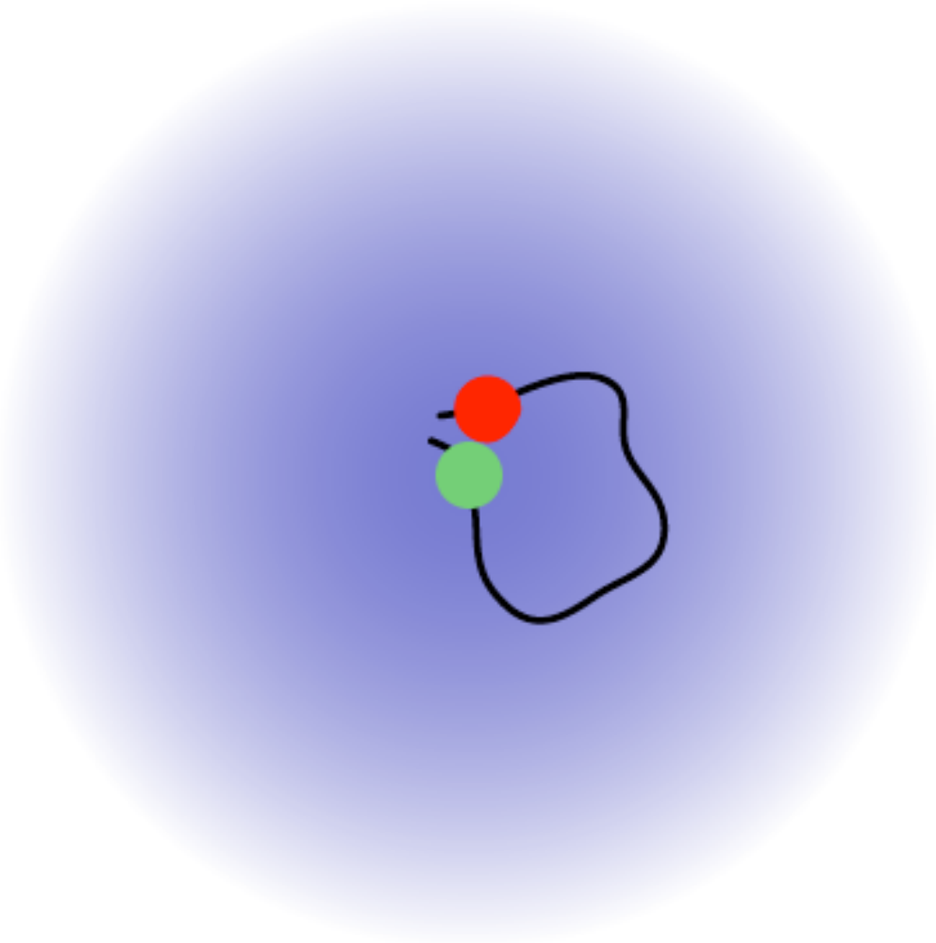
Dependence of local concentration on the site separation distance



150 base pairs distance



1500 base pairs distance



Values for j_M can be calculated from the contour length and flexibility of a specific nucleic acid chain

Nucleic acid chain	Length per monomer unit	Kuhn length (nm) (2x persistence length)
dsDNA	0.34 nm b^{-1}	100
dsRNA	0.27 nm b^{-1}	70-80
ssDNA	0.50-0.60 nm nt^{-1}	2-6
Single-stranded poly(rU)	0.65 nm nt^{-1}	4
Single chromatin fiber	8.6 nm kb^{-1}	60
Chromatin fiber	9.6 nm kb^{-1}	137-440
Metaphase chromosome	34 nm Mb^{-1}	300-5400

Abbreviations: b, base pair; nt, nucleotide; kb, kilobase pair; Mb, megabase pairs

Calculating the local concentration j_M of one site in the proximity of the other site for a linear polymer

approximation of FJC and KP-chain for linear polymer with $0.5 < n < 100$

$$j_{\mathrm{M}}(n) = 0.53 \times n^{-3/2} \times \exp\left(-\frac{2}{n^2}\right) \times l^{-3} \quad \frac{\mathrm{mol\ nm}^3}{\mathrm{liter}}$$

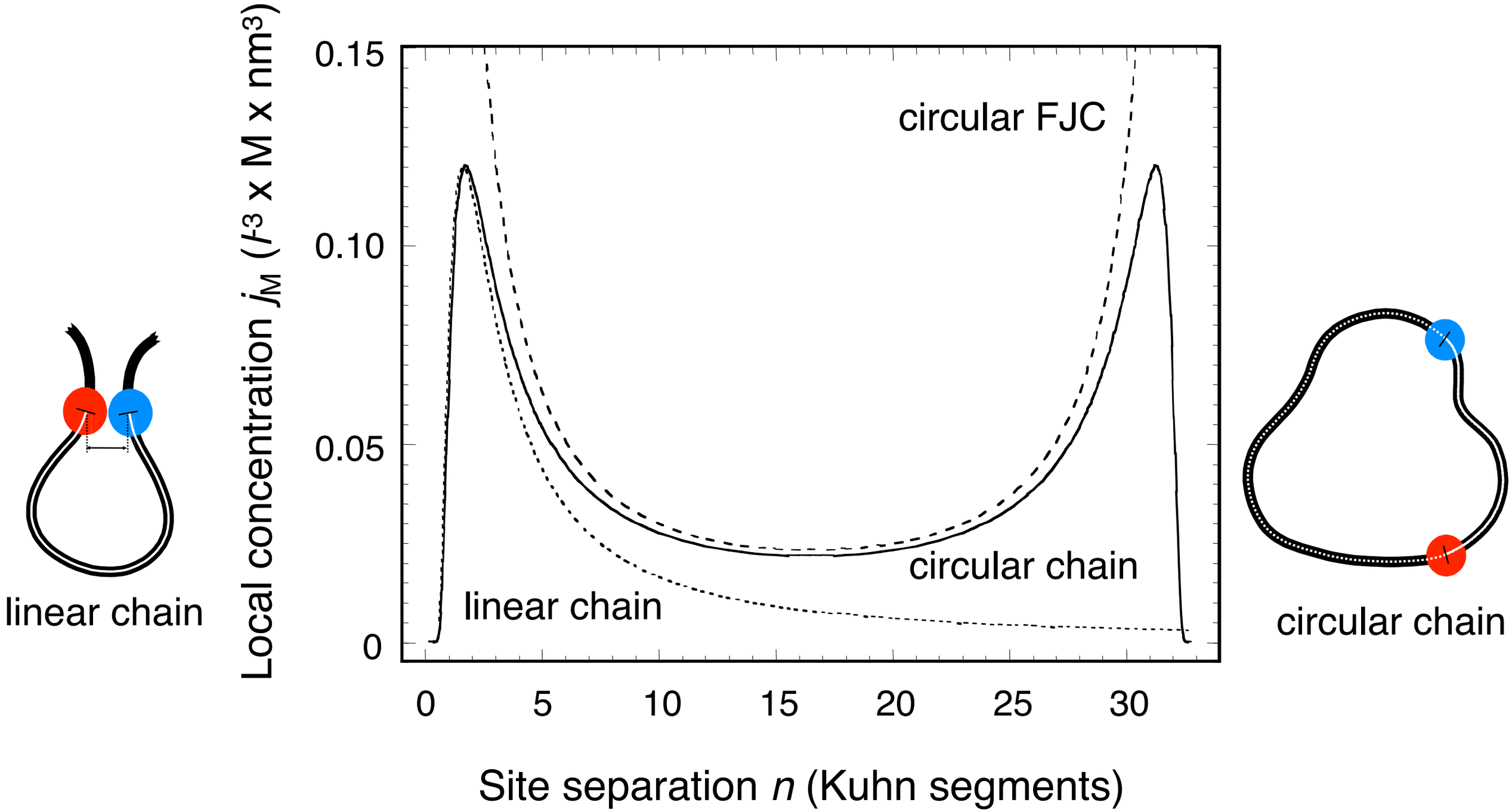
$$n = \frac{\text{number of monomers} \times \text{monomer length}}{\text{Kuhn length } l}$$

for a 30 nm chromatin fiber with a contour length of 11.1 nm/kb DNA and a Kuhn length l = 60 nm:

$$n_{\mathrm{fiber}} = \frac{s\,(\mathrm{kb}) \times 11.1\,(\mathrm{nm/kb})}{60\ \mathrm{nm}}$$

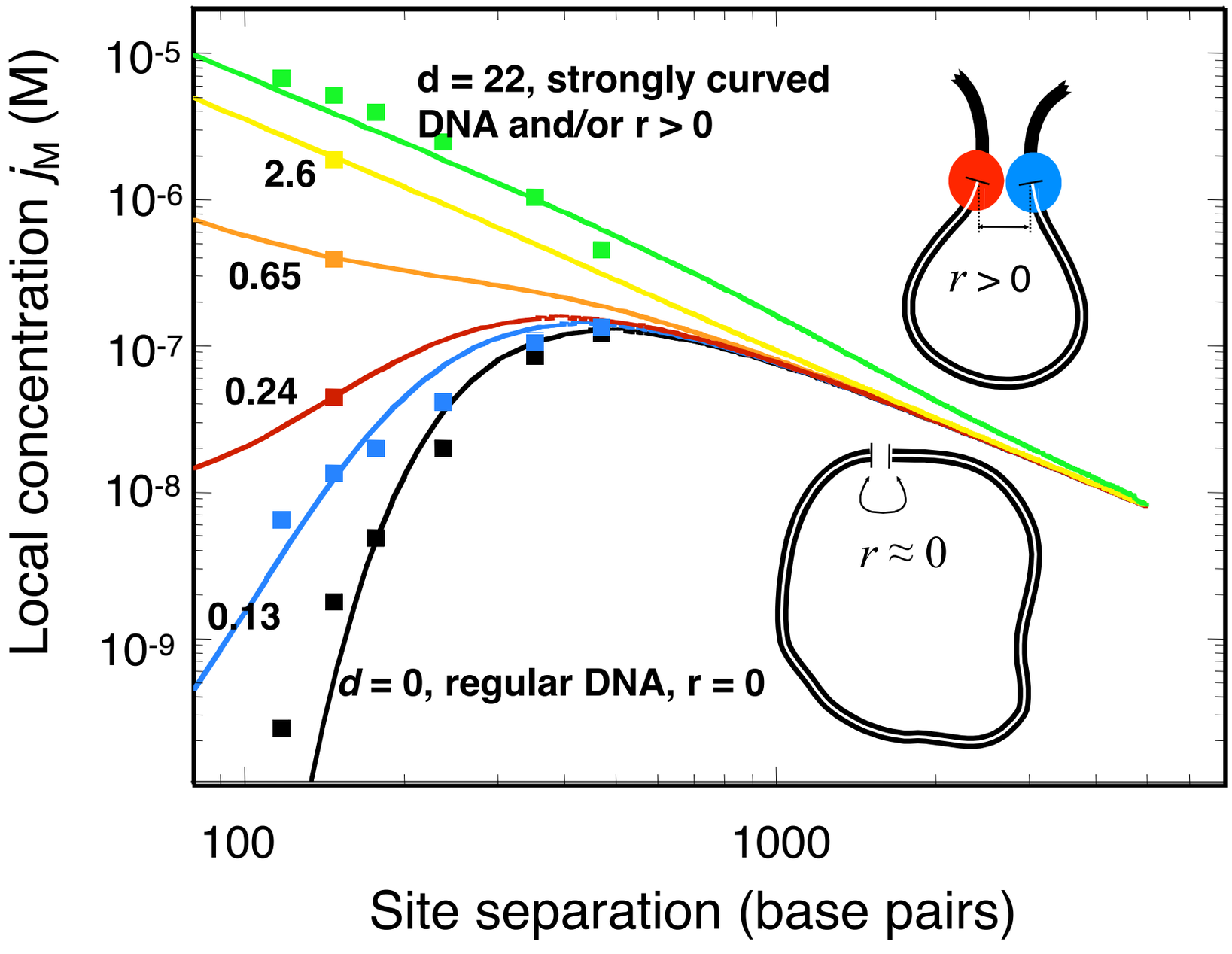
Rippe, K. (2001). Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* **26**, 733-740.

Dependence of the local concentration j_M on the site separation distance for circular and linear chains

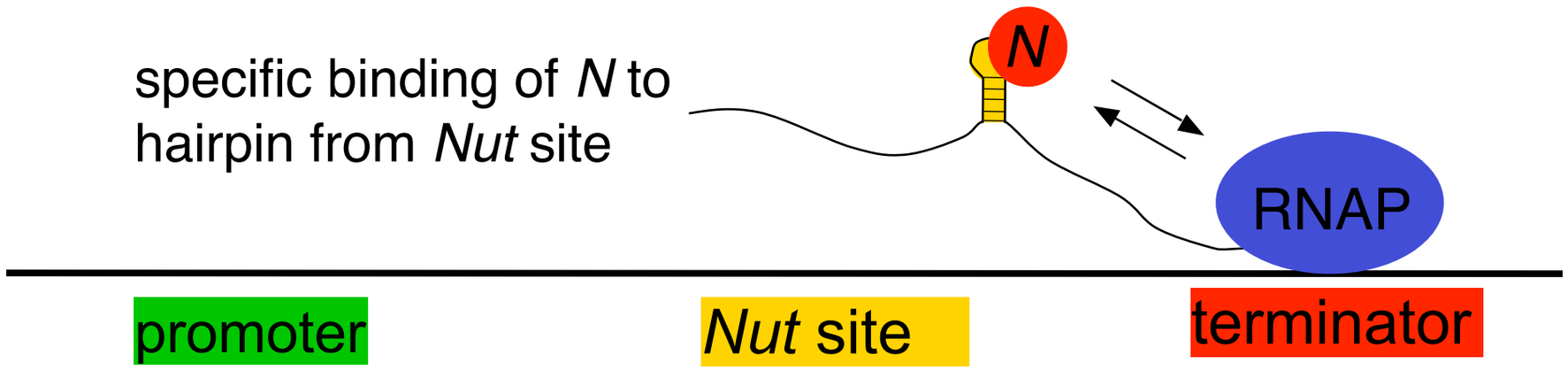


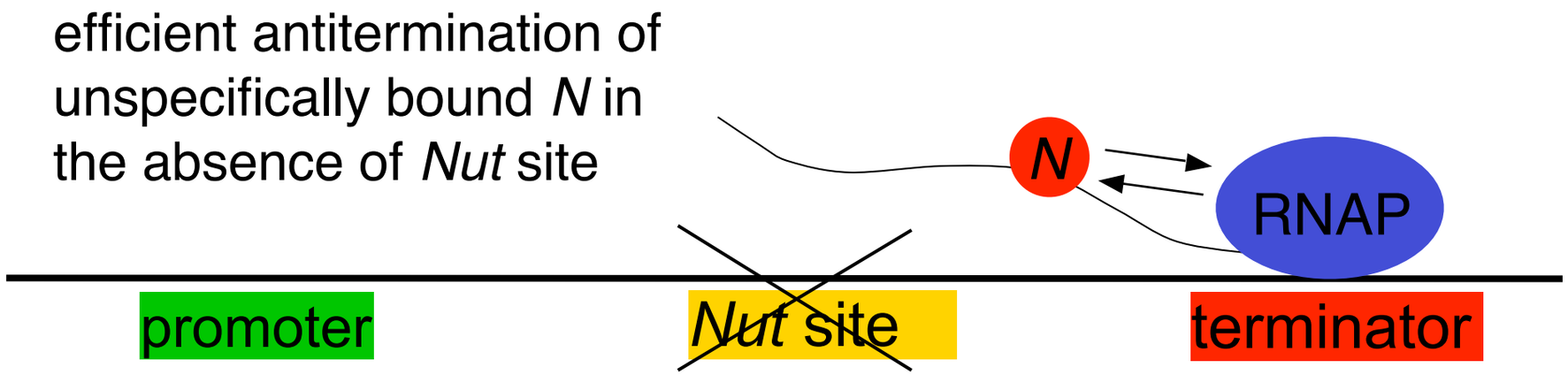
Dependence of local concentration j_M on the site separation distance b in base pairs for double-stranded DNA

$$j_{\mathrm{M}}(b) = 2.7\times10^{-3}\times b^{-\frac{3}{2}}\times\exp\left(\frac{d-2}{1.2\times10^{-5}\times b^{2}+d}\right)\ \frac{\mathrm{mol}}{\mathrm{liter}}$$



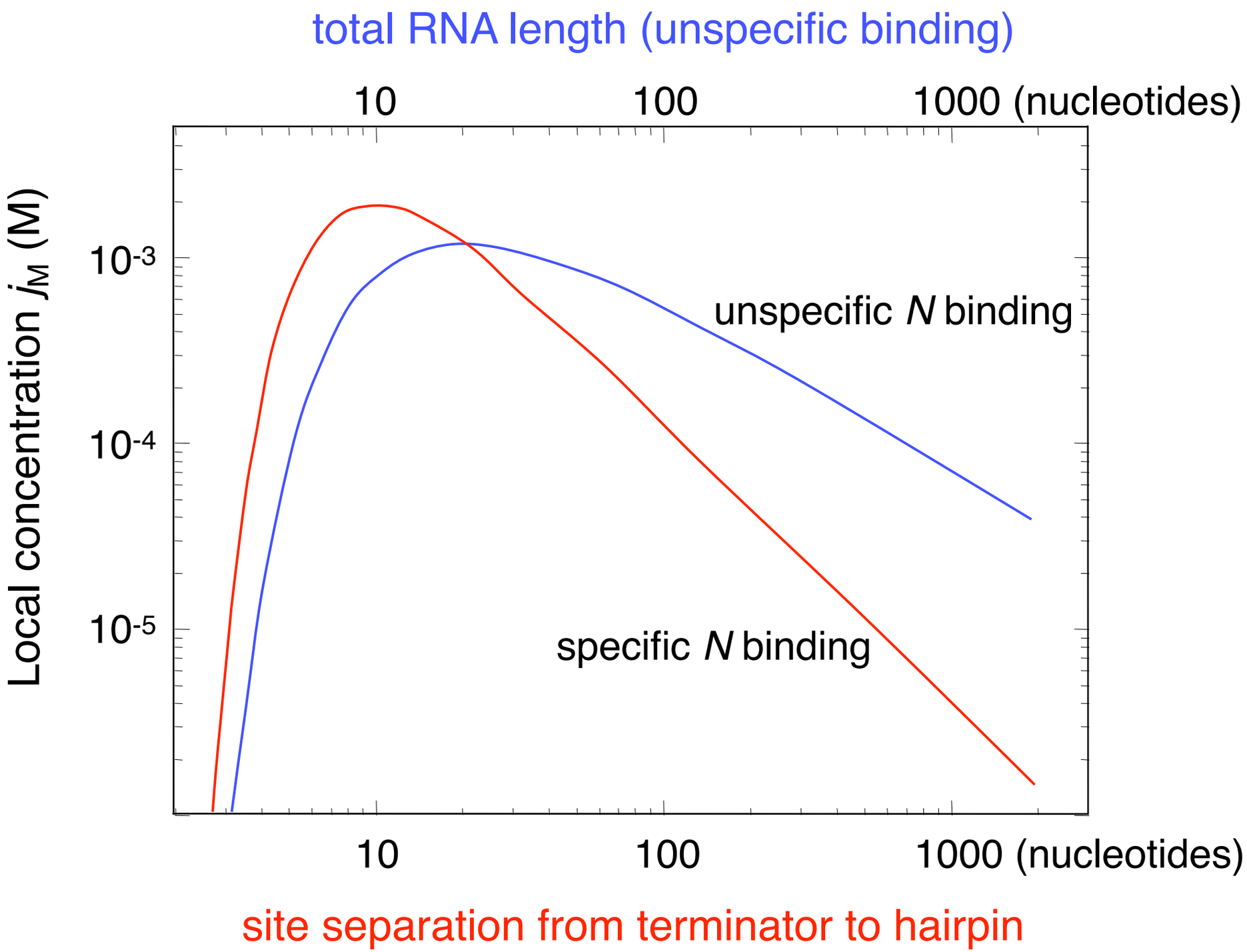
Looping of single-stranded RNA – Antitermination of RNA polymerase by phage lambda *N* protein





von Hippel, P. H., Rees, W. A. Rippe, K., Wilson, K. S. (1996) *Biophys. Chem.* **59**, 231-246.

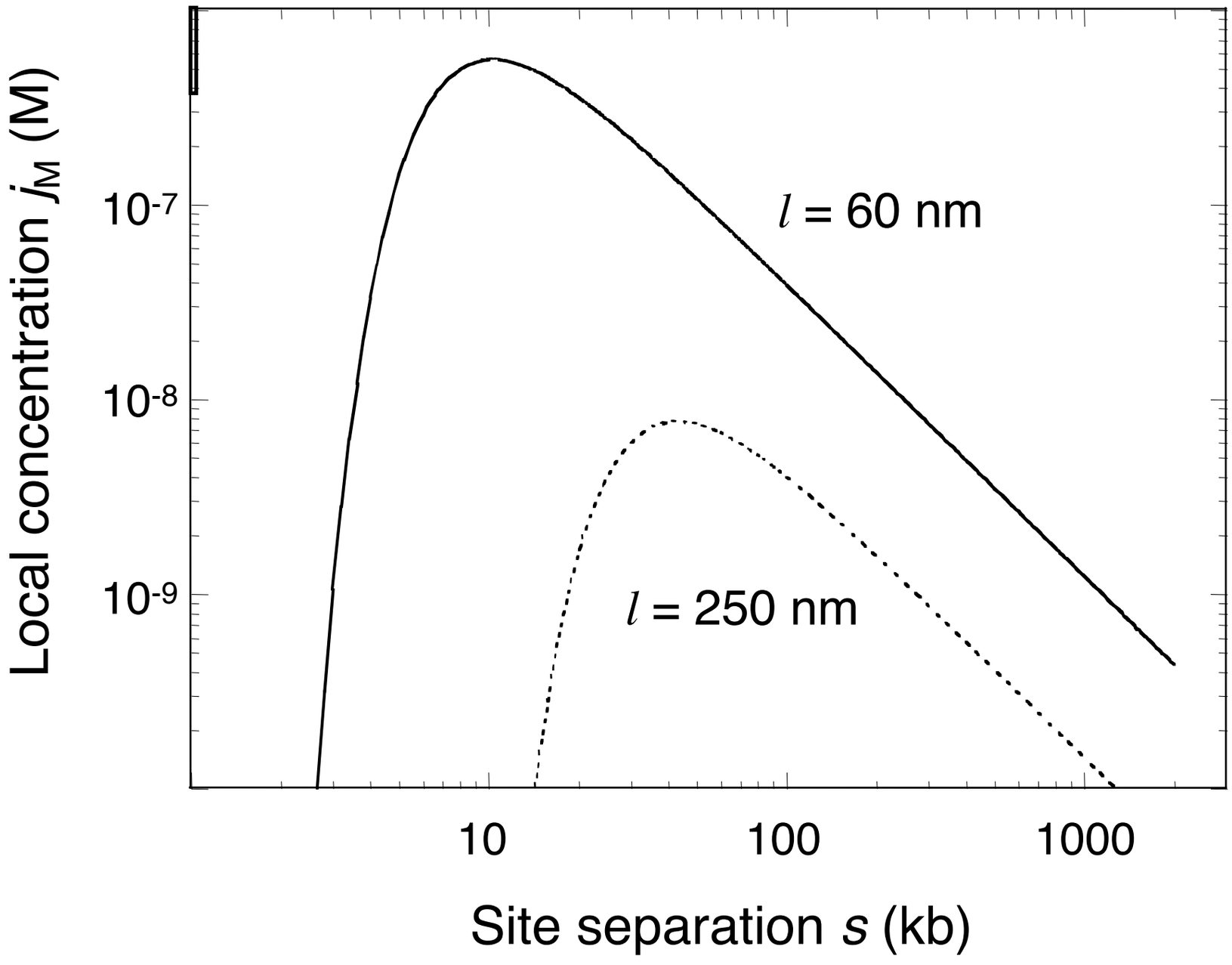
Local concentration of *N* protein in the proximity of the RNA polymerase for specifically and unspecifically bound *N*



Local concentration j_M for the 30 nm chromatin fiber

$$j_{\mathrm{M}}(s) = 3.9\times10^{-5}\times s^{-\frac{3}{2}}\times\exp\left(-\frac{80}{s^2}\right)\ \frac{\mathrm{mol}}{\mathrm{liter}}$$

$$j_{\mathrm{M}}(s) = 4.6\times10^{-6}\times s^{-\frac{3}{2}}\times\exp\left(-\frac{1400}{s^2}\right)\ \frac{\mathrm{mol}}{\mathrm{liter}}$$



Summary on polymer models for long range interactions

Homogenous polymer model with three parameters:

- **Stiffness:** statistical segment length or persistence length
- **Contour length**: polymer length in nm/base pair or nm/nucleotide
- **Contact distance**: r = 0 nm (DNA circle ligation) or r = 10 nm (protein-protein)

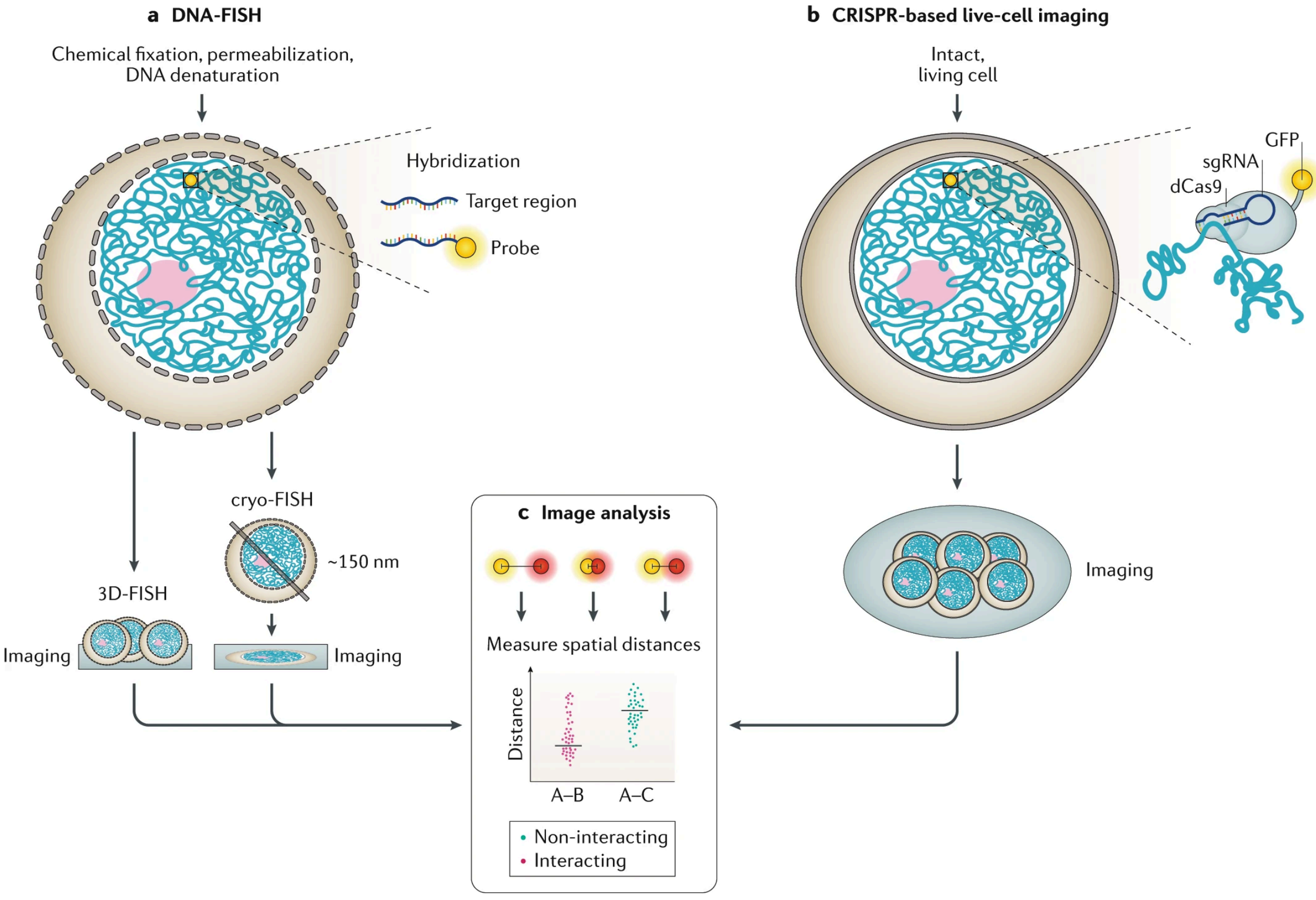
Local concentration j_M in mol/liter to express looping contact probability:

- Describes interaction probability of one sites on a polymer near another one
- j_M can be compared to the concentration without polymer tether (“in trans)
- j_M gives DNA concentration where circle and dimer formation have equal probability

Calculating j_M for two sites with a certain distance on a polymer:

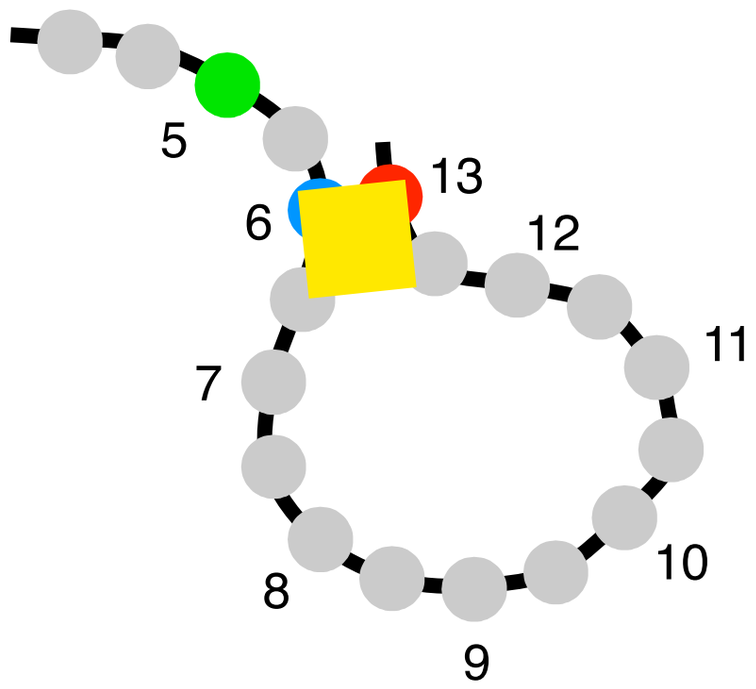
- Universal description of j_M vs separation distance in statistical segment length units
- j_M has a maximum at 1.7 x the statistical segment length
- Specific polymer: j_M as a function of stiffness, contour length and contact distance

Imaging-based approaches to visualize chromatin contacts



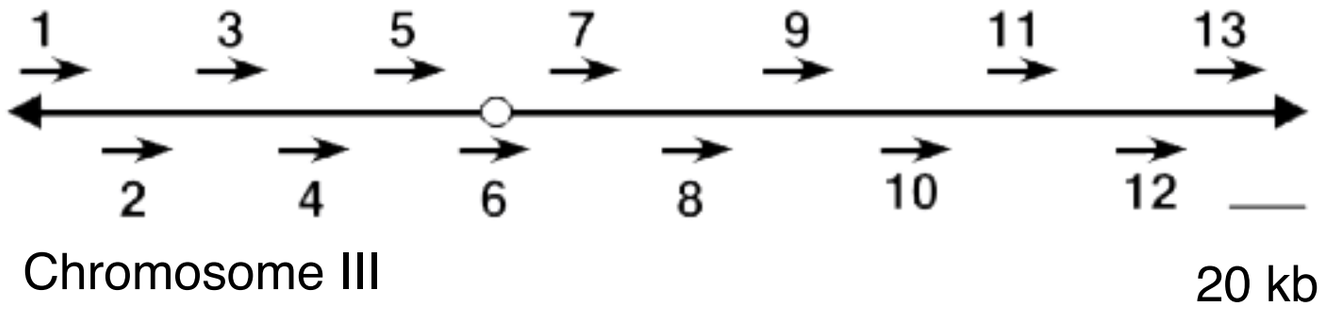
Szabo, Q., Bantignies, F., Cavalli, G. (2019). Science Advances 5(4), eaaw1668. https://dx.doi.org/10.1126/sciadv.aaw1668

Capturing **C**hromatin **C**onformation (3C assay) - mapping the 3D genome by cross-linging



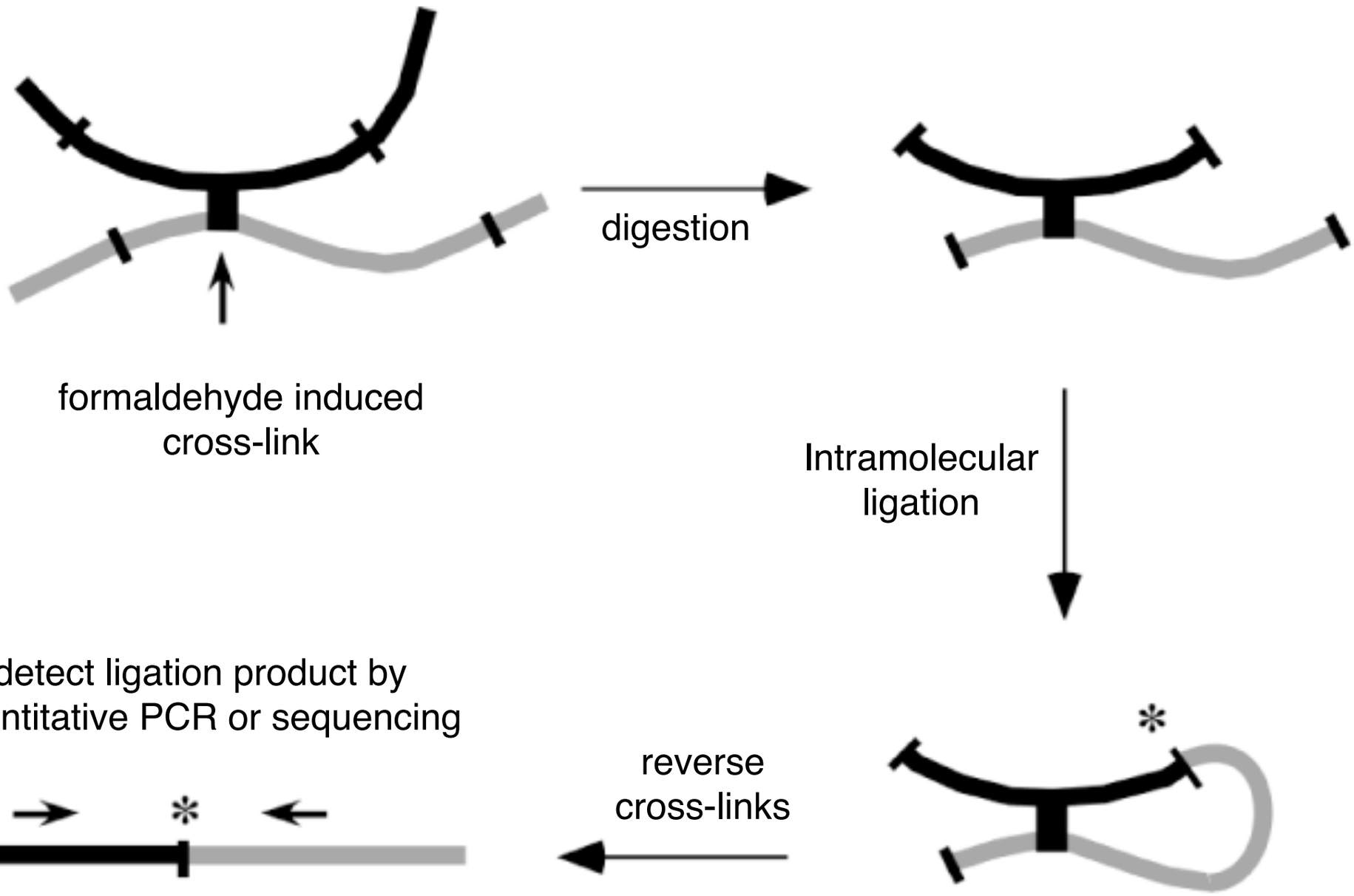
Chemical cross-link two sites on the chromosome

Quantitate cross-link efficiency and relate it to **genomic distance**

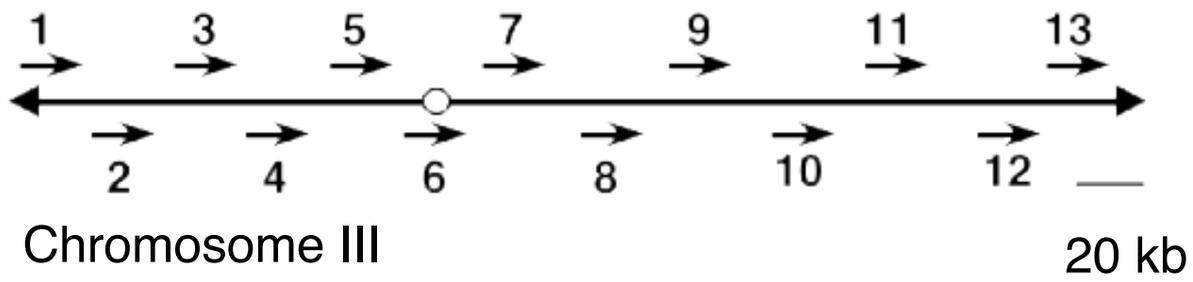


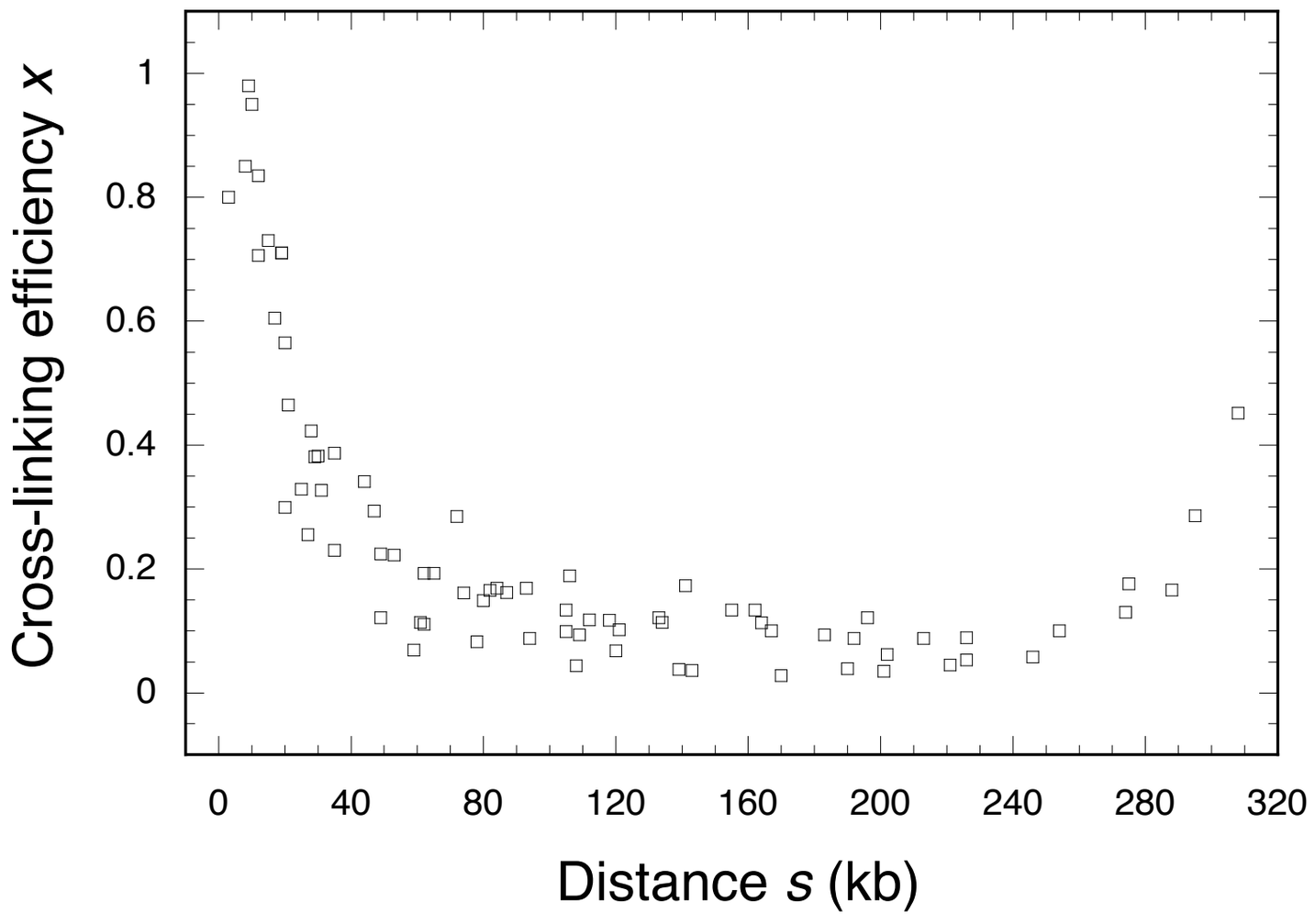
Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. (2002) *Science* **295**, 1306-1311.

Capturing **C**hromatin **C**onformation (3C assay) by *in vivo* cross-linking of whole cells or isolated nuclei



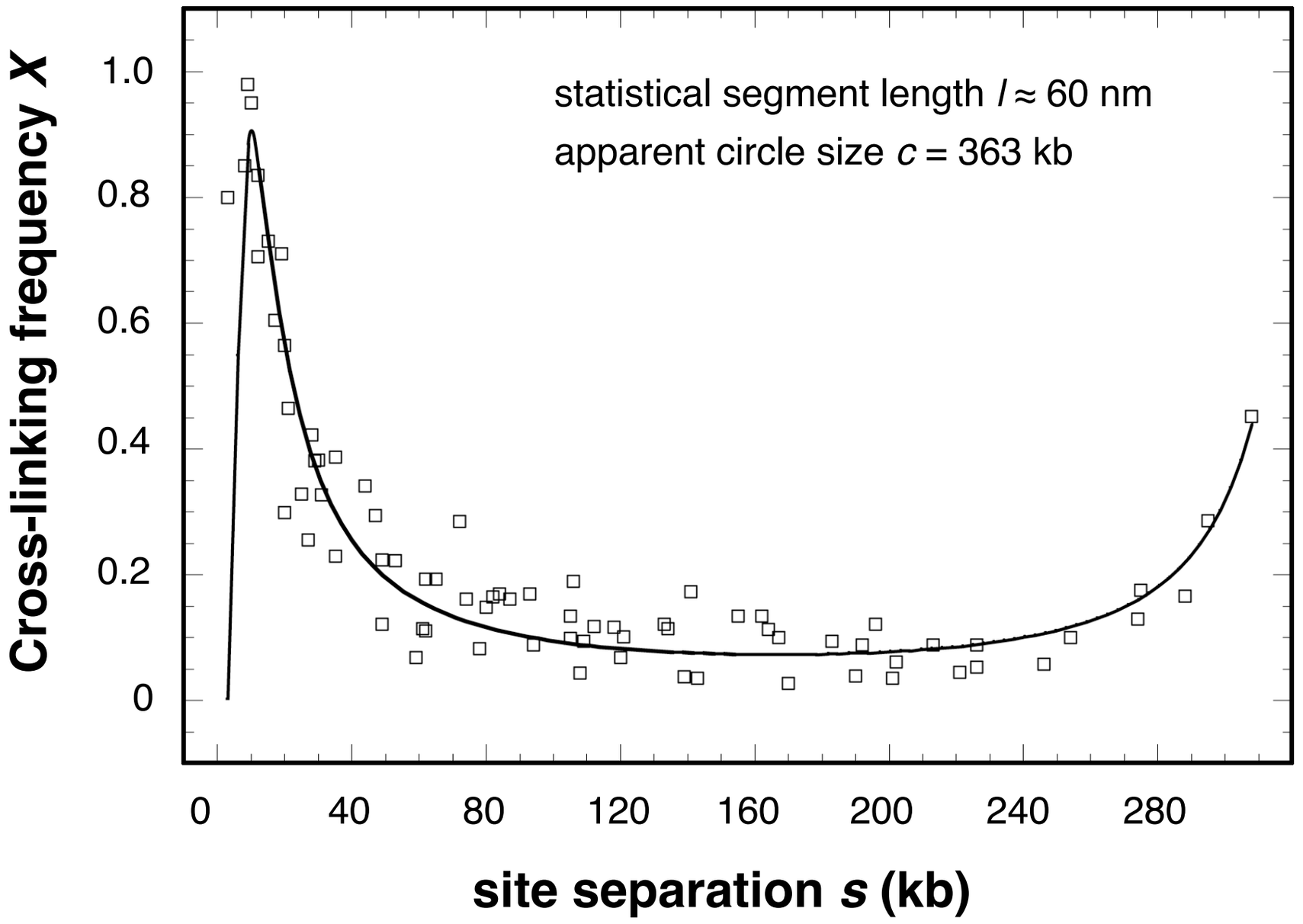
Cross-linking analysis of the *in vivo* interphase conformation of the complete yeast chromosome III (315 kb) in G1



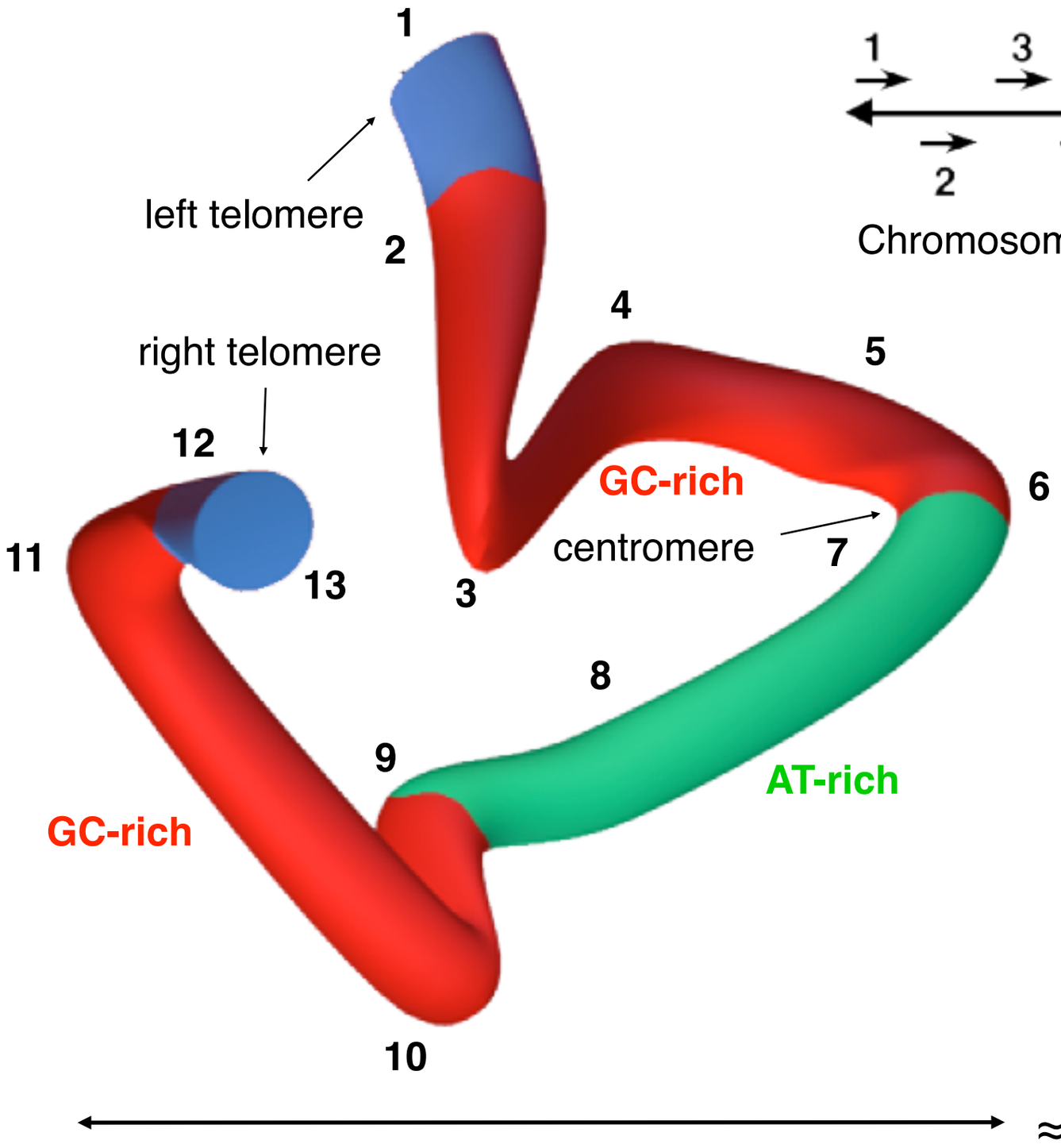


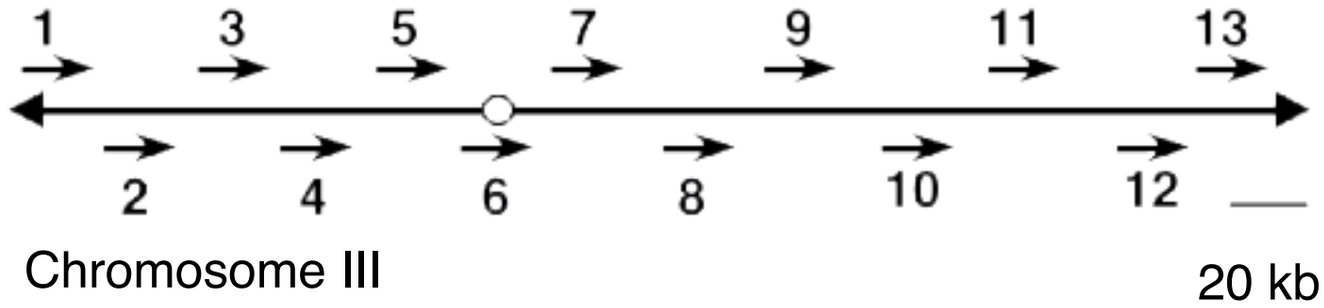
Cross-linking analysis of the *in vivo* interphase conformation of the complete yeast chromosome III (315 kb) in G1

$$X(s) = k \cdot j_{\mathrm{M}}(s) \qquad X(s) = k \times 0.53 \times \beta^{-\frac{3}{2}} \times \exp\left(-\frac{2}{\beta^2}\right) \times l^{-3}\,\frac{\mathrm{nm}^3\,\mathrm{mol}}{\mathrm{liter}} \qquad \text{with } \beta = 11.1\,\frac{\mathrm{nm}}{\mathrm{kb}} \times \frac{s}{l} \times \left(1 - \frac{s}{c}\right)$$



Average 3-D conformation of yeast chromosome III (315 kb) in interphase

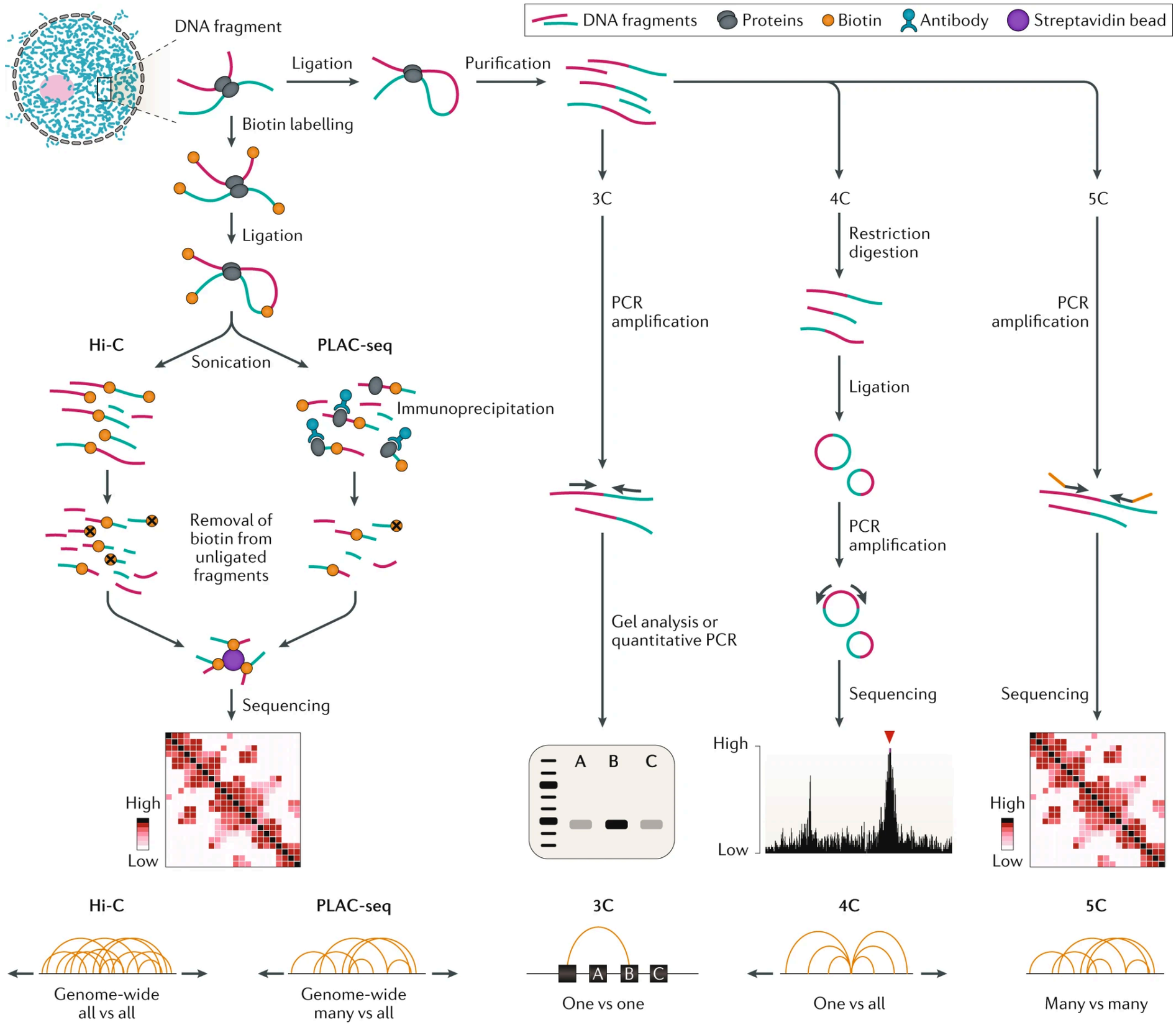




local concentration $j = \left(\frac{3}{2\pi \langle r^2 \rangle} \right)^{\frac{3}{2}}$

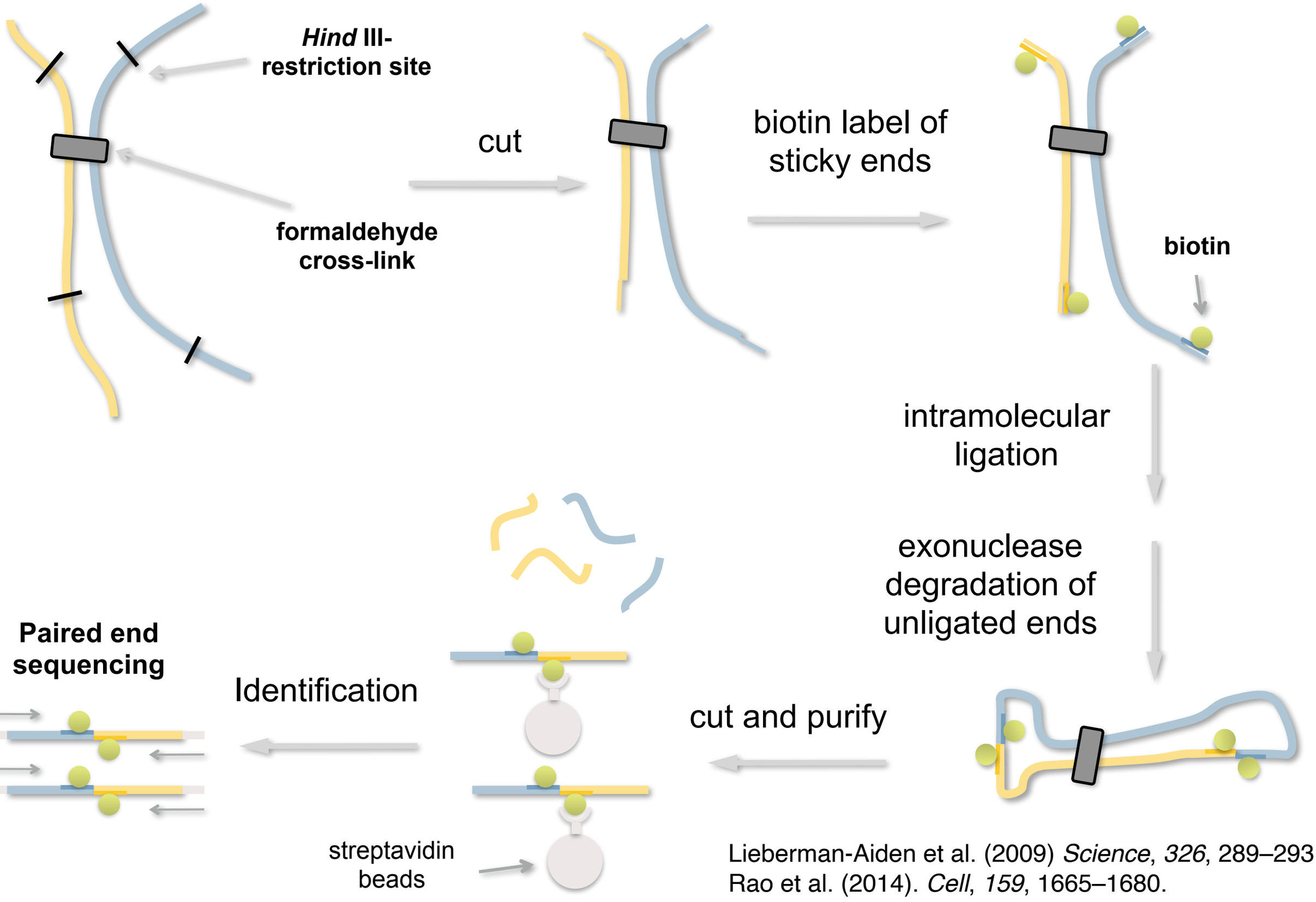
$$\sqrt{\langle r^2 \rangle} = 0.82 \times j_{\mathrm{M}}{}^{-\frac{1}{3}} \left(\frac{\mathrm{mol}}{\mathrm{liter}} \right)^{\frac{1}{3}} \mathrm{nm}$$

3C and its derivatives



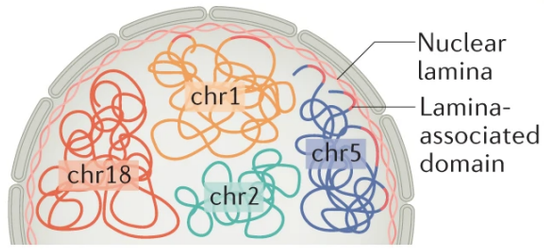
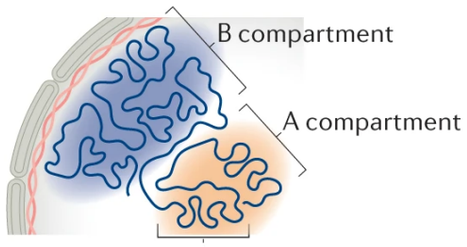
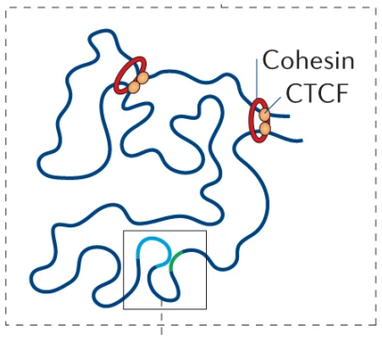
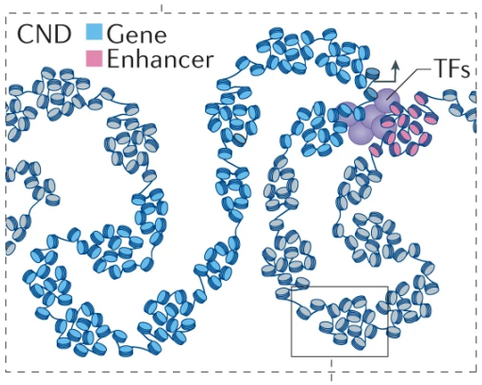
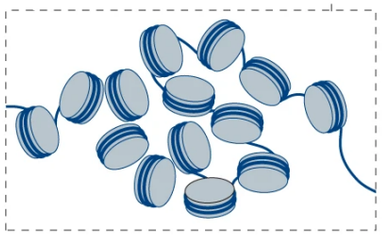
Kempfer, R., Pombo, A. (2019). **N**ature reviews. Genetics https://dx.doi.org/10.1038/s41576-019-0195-2

Hi-C – map all interactions

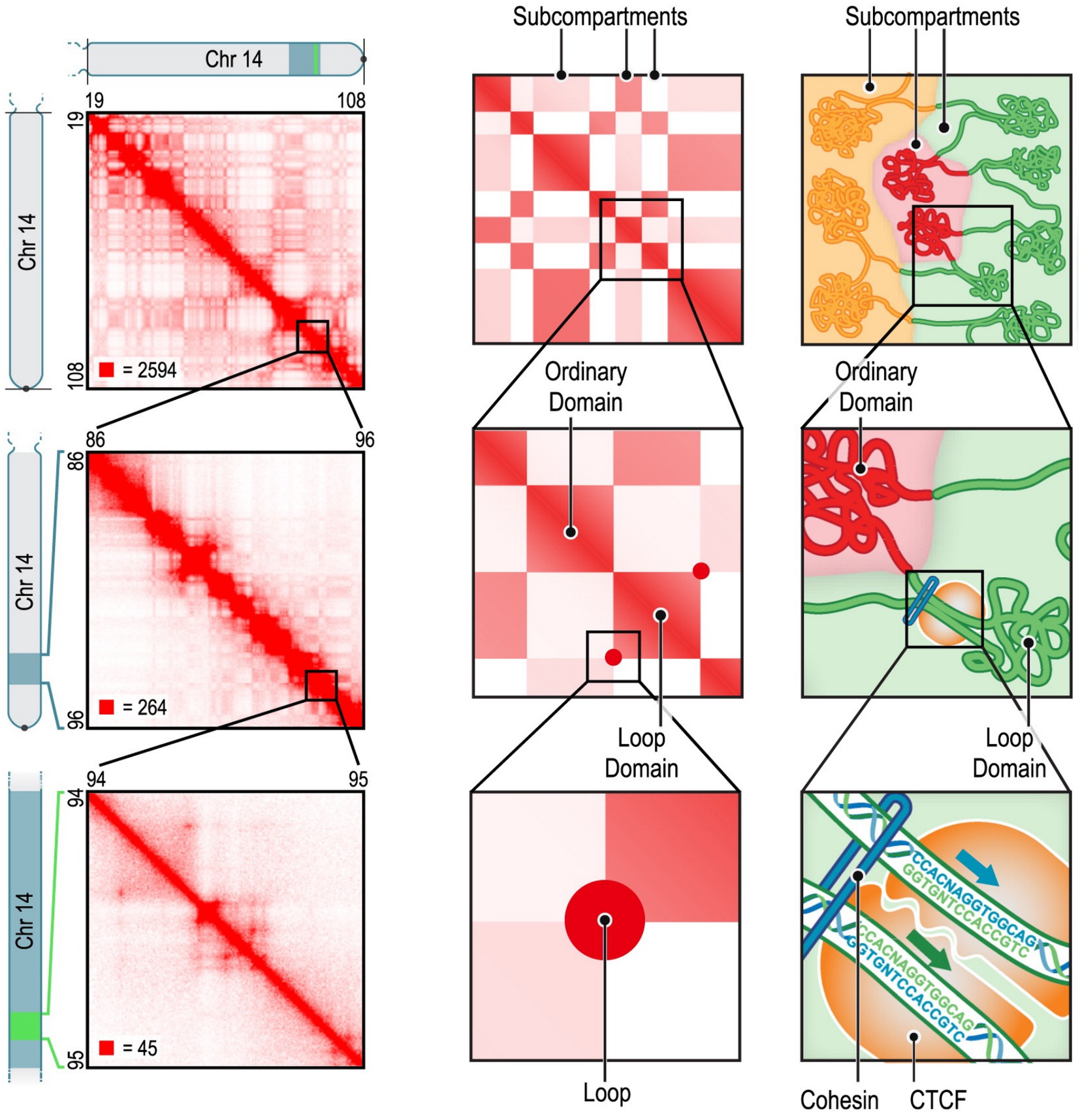


Lieberman-Aiden et al. (2009) *Science*, *326*, 289–293.
Rao et al. (2014). *Cell*, *159*, 1665–1680.

Hierarchical organization of the eukaryotic genome

Size	Organization level		Methods: Sequencing-based: Enrichment	Methods: Sequencing-based: Non-enrichment	Methods: Microscopy
Entire chromosomes	Chromosome territories			• GAM • GPSeq	3D FISH
Up to 100s of megabases	• Compartments • Hubs		• DamID • TSA-seq • ChIA-PET • ChIA-Drop	• GAM • Hi-C • Micro-C • GPSeq • SPRITE	• Super-resolution FISH • Electron microscopy
100 kb to few megabases	• Topologically associating domains • Chromatin loops		• Capture-C • cHi-C • 4C • ChIA-PET • ChIA-Drop • HiChIP • PLAC–seq	• GAM • SPRITE • Hi-C • Micro-C	Super-resolution FISH
10–100 kb	• Chromatin nanodomains • Functional loops (E–P contacts)		• Capture-C • cHi-C • 4C • ChIA-PET • ChIA-Drop • HiChIP • PLAC–seq	• GAM • Micro-C	Super-resolution FISH
~1–2 kb	Nucleosome clutches			Micro-C	• Super-resolution FISH • Electron microscopy

Jerkovic´ & Cavalli, Nat Rev Mol Cell Biol, 10.1038/s41580-021-00362-w


Chr 14
19
108
19
108
Chr 14
= 2594
86
96
86
96
Chr 14
= 264
94
95
94
95
Chr 14
= 45
Subcompartments
Ordinary
Domain
Loop
Domain
Loop
Subcompartments
Ordinary
Domain
Loop
Domain
CCACNAGGTGGCAG
GGTGNTCCACCGTC
Cohesin
CTCF

Summary on chromatin conformation capture analysis

- In situ cross-linking maps of sites that interact in the genome
- Either targeted to specific sites (3C, 4C) or genome-wide (HiC)
- Resolution depends on distribution of fragment size and sequencing depth
- Averaging over many cell can obscure the true conformation
- Regions of “random” coil polymer conformation but also specific loops
- Chromosome form “territories” and interactions on the same chromosome at short distances < 100 kb are most probable
- Identifies “topologically associated domains” (TADs) that are 1 Mb in size where interactions are facilitated
- TADs spatially organize promoter-enhancer interactions