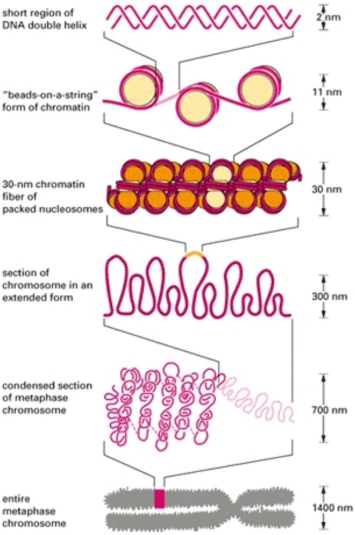
Origin of spatial organization of DNA-polymer in chromosomes.

Apratim Chatterji

IISER-Pune
apratim@iiserpune.ac.in

Bangalore Statphys-2017

DNA: Basic facts.



DNA is made up of: A,T,C,G.

The pair A-T or C-G: Base pair (BP).

Human DNA has 6.4×10^9 base pairs.

Length $2m$ packed in nucleus $\sim \mu^3$.

Some imp numbers:

Base pair : 3.3Å.

Helix pitch =3.4 nm OR 10.4 BP.
Chain width $\sim$ 2nm.

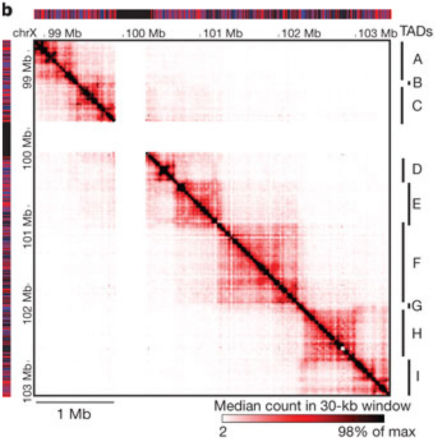
20 BP around histone.
Kuhn Segment: 300 BP or 100nm.

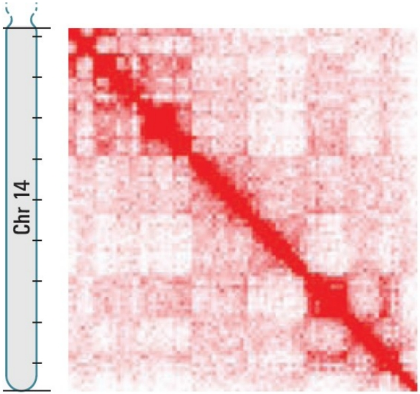
DNA packed in nucleus by organizing into hierarchy of Mesoscale structures.

Single Chromosome: Chromosmal Contact Maps.

A single chromosome is not organized Randomly:

Nora et. al., Nature (2012), Lieberman-Aiden et.al., Science, (2009)





Higher probabilities to find certain segments of chain in contact with specific other segments. How is the organization achieved?

What causes large scale organization of DNA?

Hypothesis:

- Cross-links at **specific** points along the chain leads to the organization of the DNA-polymer ?! - DNA binding proteins.

- **If the contacts are determined experimentally :**

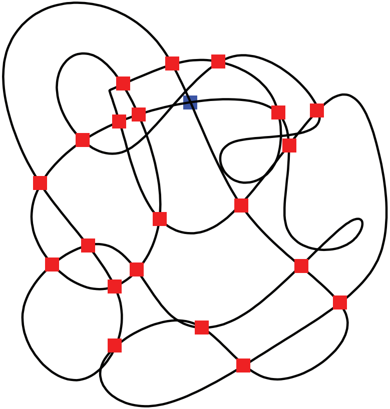
Can we test the hypothesis and determine *structure* ?

- Random cross-links: Nature of organization is different. More diffuse organization.

Bio-collaborators: Farhat Habib, G.P. Manjunath (IISER-Pune)
Ph.D. student: Tejal Agarwal.

Modelling-1: Choose Bacteria with single DNA.

1. Bacteria *E-coli*
4.642×10^6 BP = 4642 monomers.
2. 1 coarse grained monomer in model = 1000 BP.
3. Bead spring model of polymer.
4. Persistence length of DNA :
$\ell_p \equiv 150$ BP $\equiv 50$ nm
5. RING polymer.
6. Ex. Volume radius $= 0.2b$: Suitably truncated Lennard Jones interactions.



Cross-link between monomers A, B modelled by $V_{AB}(r) = \kappa_b(r_{AB} - b)^2$

Experimental Input To Simulations:

Experimental Input: High Frequency of monomer A and B to be proximity:

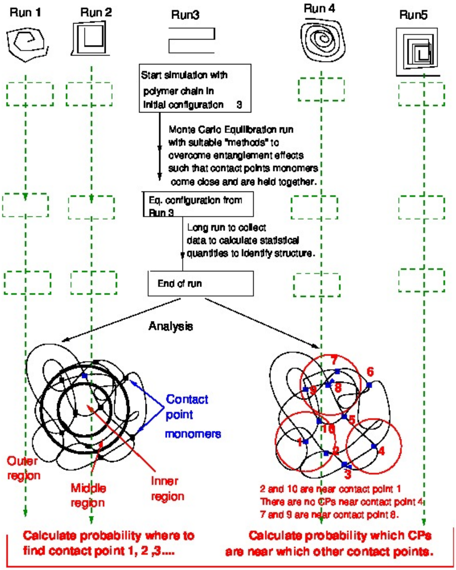
Frequency $> p_c \rightarrow$ Cross Link (CL).

CC	E.coli
60	82
26	27.

4017/4642 monomers in DNA-polymer

How to Determine structure of a FLOPPY chain ?

START: DIFFERENT INITIAL CONFIGURATIONS BUT IDENTICAL POLYMER CHAINS AND SET OF CONTACT POINTS (CPs).



Start simulation with polymer chain in Initial configuration 3

Monte Carlo Equilibration run with suitable "methods" to overcome entanglement effects such that contact points monomers come close and are held together.

Eq. configuration from Run 3

Long run to collect data to calculate statistical quantities to identify structure.

End of run

2 and 10 are near contact point 1
There are no CPs near contact point 4
7 and 9 are near contact point 8.

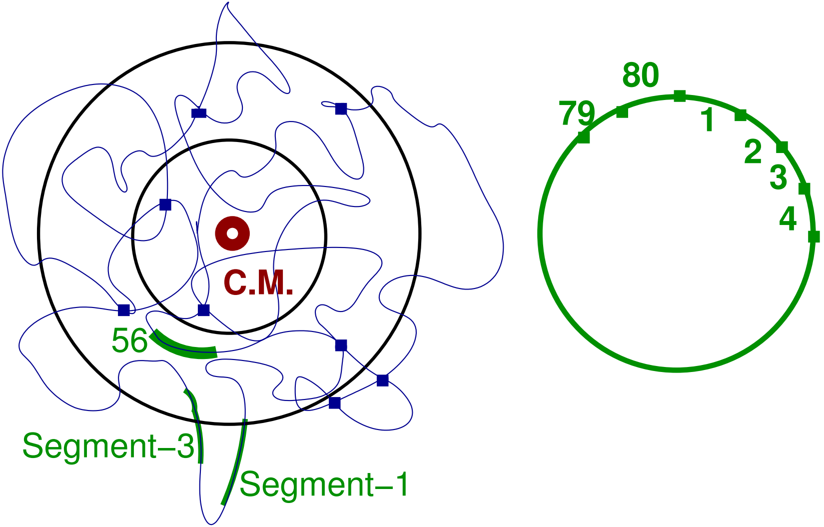
Calculate probability where to find contact point 1, 2 ,3....

Calculate probability which CPs are near which other contact points.

COMPARE STATISTICAL DATA (probabilities) from different runs.

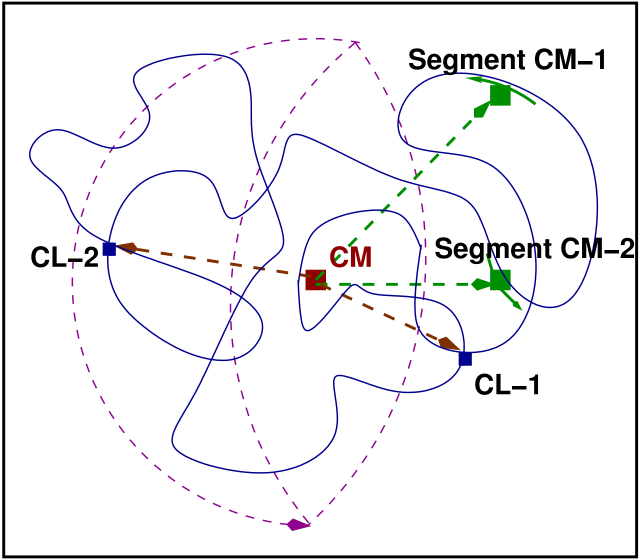
Quantities determining Structure ?? R_g..and..

CL: CROSS LINKs



Total ecoli length : 4642 monomers: 80 segments of 58 monomers each.

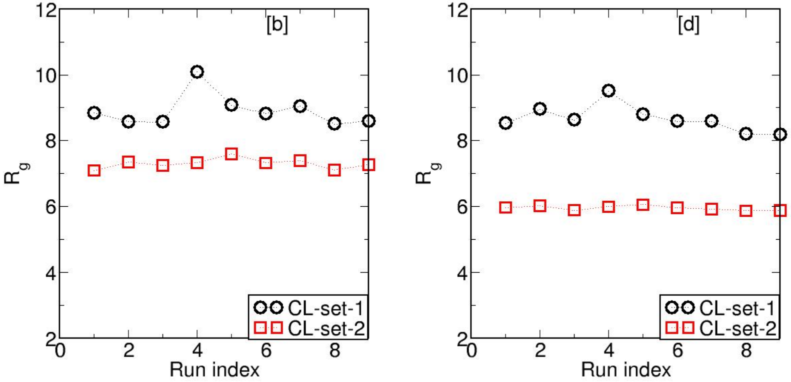
Segment-Segment Angular correlations:



Compare Radius of gyration R_g from different runs:

CL-set-1 : **27 effective** CLs.
CL-set-2 : **82 effective** CLs.

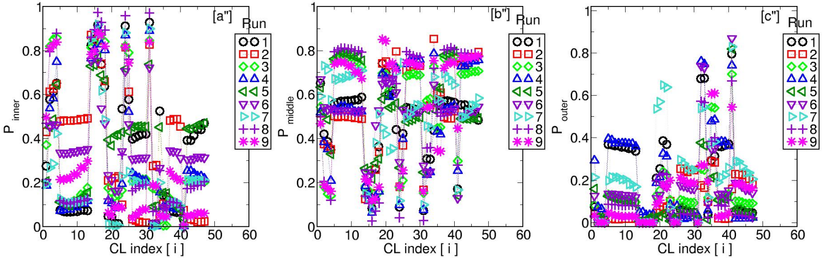
CL: Cross Links

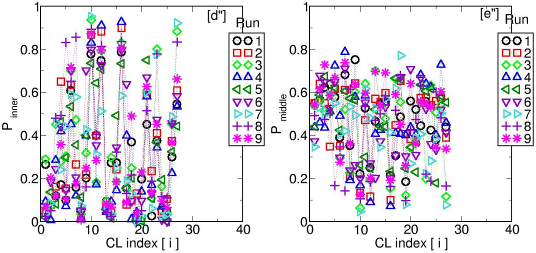


Ecoli Random .

Decrease in R_g more for Random CLs.

Spatial location of a CL: Inner/ Middle/ Outer?





47 Bio CLs,
27 *effective* CLs

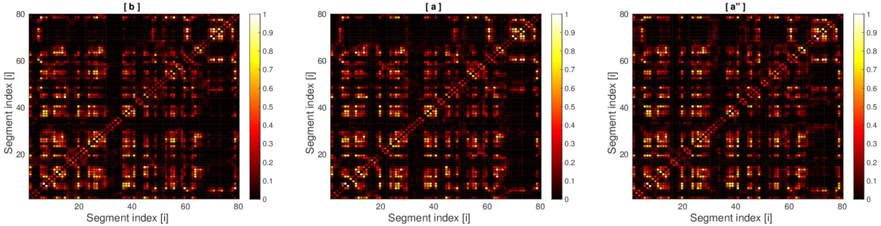
← 27 RANDOM CLs.

Inner Region: $r_{in} < 5b$;
Outer Region: $r_{outer} > 9b$.

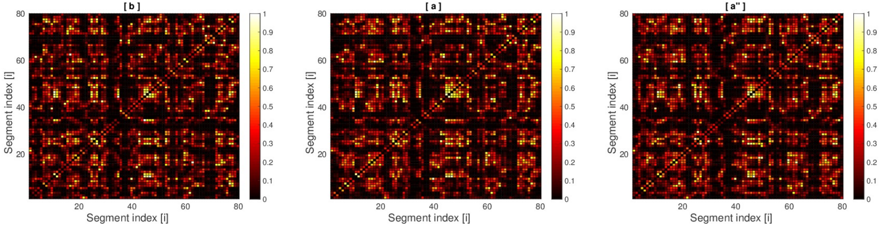
CLs: markers for segments of chain:
Well defined location of CLs/segments.

The neighbouring segments of a particular segment?

Ecoli $\equiv$ 82 CLs: Comparison from 3 independent runs:



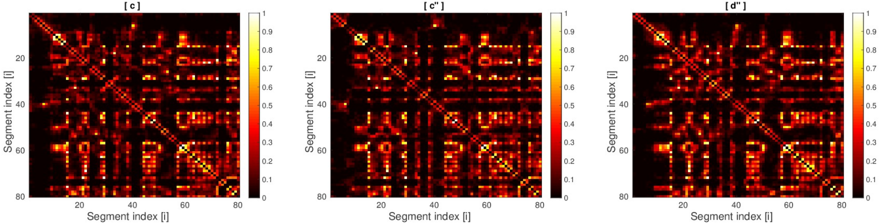
82 CLs at **Random** locations along chain: BUT



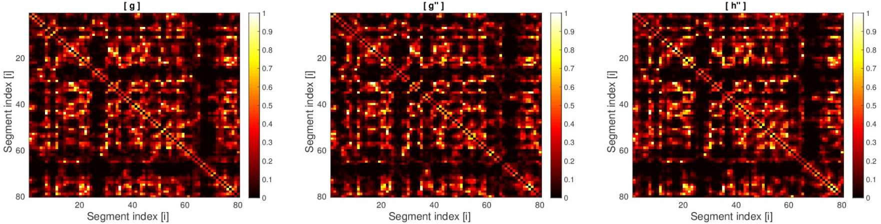
Trivial close points (by virtue of closedly spaced contacts) removed from data.

The neighbouring segments of a particular segment?

CC $\equiv$ 60 CLs: Comparison from 3 independent runs:

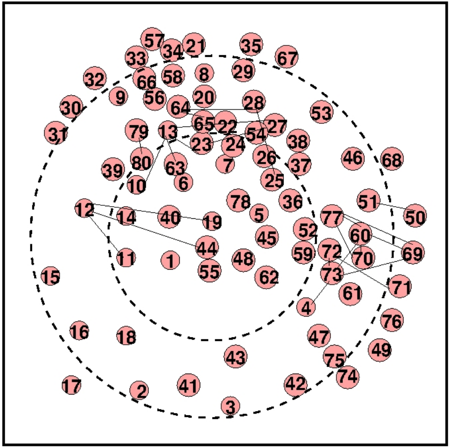


60 CLs at **Random** locations along chain: BUT

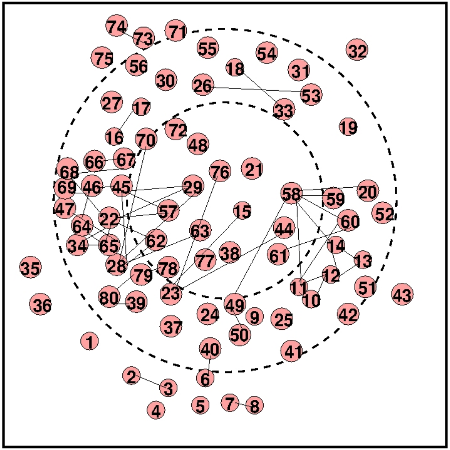


Trivial close points (by virtue of closedly spaced contacts) removed from data.

2-D map: Organization of 80 segments



E.Coli



Caulobacter Crecentus.

58 monomers in segment

50 monomers in a segment.

Theoretical prediction: Can be cross checked by Expts.

Conclusions.

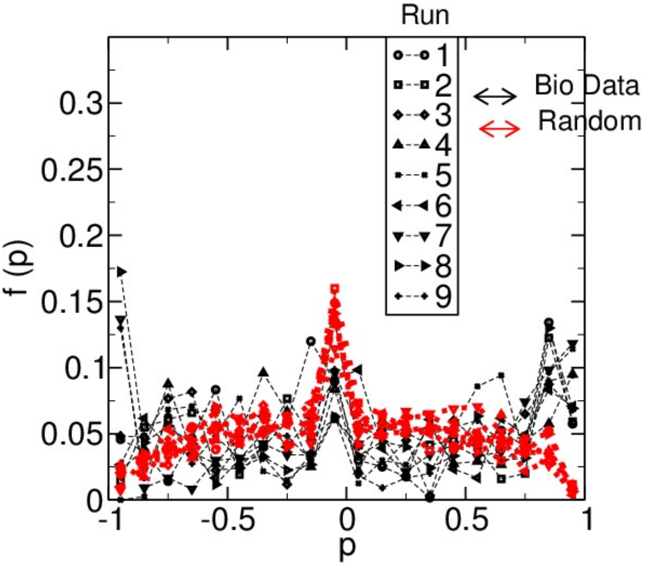
CROSS LINKS AT SPECIFIC POINTS ALONG CONTOUR DETERMINES

GLOBAL CHROMOSOMAL ORGANIZATION (μ length scale).

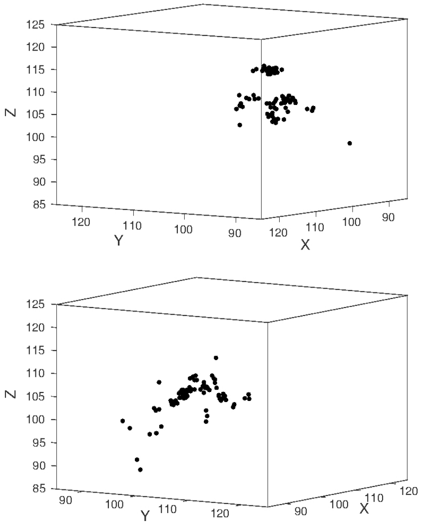
PhD Student : Tejal Agarwal
Bio-collaborators: G.P. Manjunath, Farhat Habib.

Thank you.

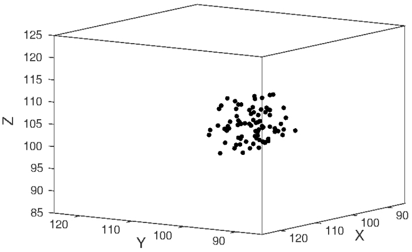
Comparision : Distribution of probabilities



Cross-Link Distribution



E.Coli & C.Crescentus.

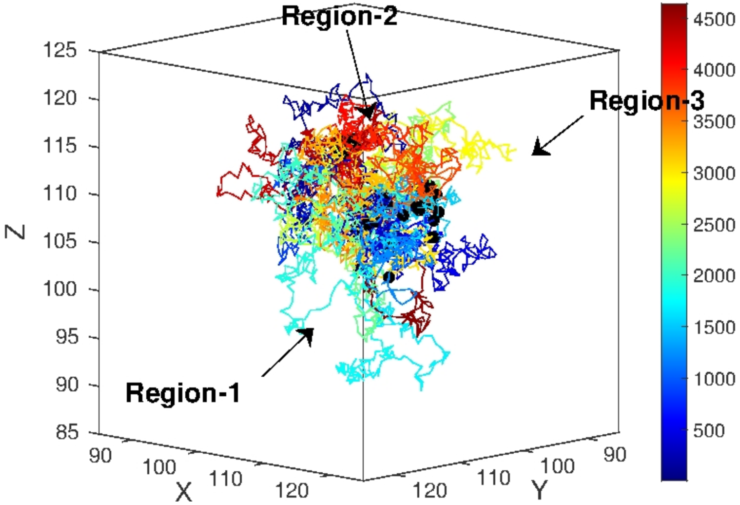


RANDOM

BIO-CLs are clustered spatially.

Random are more Scattered.

E.coli-Polymer Snapshot



Snapshot matches with Angular Correlation data.

Segment-Segment Angular correlations:

82 *effective* CLs. Last plot is with RANDOM CLs: