A study of bacterial Chromosome organization using bead-spring polymer model

Tejal Agarwal
Reg. Id: 20142035

Supervisor: Dr. Apratim Chatterji

Department of Physics

Indian Institute of Science Education and Research, IISER Pune, India

This dissertation is submitted for the degree of
Doctor of Philosophy

I would like to dedicate this thesis to my loving parents and sister . . .

Certificate

It is certified that the work incorporated in the thesis entitled ("The study of bacterial Chromosome organization using bead-spring polymer model") submitted by Ms. Tejal Agarwal was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Date: 2nd December 2019
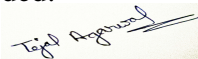
Apratim Chatterji

(Supervisor)

20 July, 2020

Declaration

I declare that this written submission represents my research work in my own words and wherever other's ideas or works have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsif ed any idea/data/fact/source in my submission. I understand that violation of the above will cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Tejal Agarwal

Reg. Id: 20142035

June 2020

Abstract

The chromosome, a long polymer ($\sim$ 2m in the human cell and $\sim$ 1mm in a bacterial cell), consists of all the genetic information and is packed $\sim$ 1000 fold inside the nucleus (in a human cell) or in the bacterial cell. The chromosome is involved in various cellular processes, e.g., DNA replication, transcription and is not like a random-walk polymer but shows a unique spatial organization at μ length-scales, as is evident from the recent experimental studies. The contact map generated from the experimental technique Hi-C gives the frequency of two chromosome segments of 1 kilo base-pair- 1 mega base-pair) each, to be in proximity to each other. The Hi-C contact maps which give essential insights about the chromosome organization but do not provide the overall 3D structure of the chromosome. Also, despite several theoretical and experimental efforts, the mechanism leading to the μ length-scale organization of the chromosome remains elusive. Currently, the primary challenge in the field is to build a physical model of the chromosome that can predict the 3D organization and shed light on the mechanism in the chromosome organization obtained in the Hi-C contact map. Developing such a model requires statistically significant high-throughput data from individual cells, with close dialogue between modeling and experiment.

Here in this thesis, we use bead-spring model of the ring polymer to understand the role of different physical mechanisms, e.g., DNA binding proteins, the crowding environment due to various proteins, enzymes inside the cell, release of topological constraints and the confinement of the cell wall in the chromosome organization of bacteria *E. coli* and *C. crescentus*. We also predict the overall 3D organization of the chromosome and validate our prediction with the available experimental data. In particular, we use the data from the Hi-C contact map of *E. coli* and *C. crescentus* to introduce the cross-links at particular positions in the bead-spring polymer where one coarse-grained bead represents 1000 base pairs (BP) of the DNA. The cross-links at specific positions mimic DNA-binding proteins. Via suitable Monte Carlo simulations, we show that the presence of $< 2\%$ cross-links leads to a particular organization of the chain at large (micron) length scales. We also investigate the structure of a ring polymer with an equal number of cross-links at random positions along the chain. We find that though the polymer does get organized at the large length scales, the nature of the organization is quite different from the organization observed with cross-links at specific

biologically determined positions. We next individually study the effect of chain crossing, crowding-environment of the cell, or the confinement due to the cell wall on the organization of the cross-linked polymer and then combined all the effects and predict the 3D organization of the chromosomes of bacteria *E. coli* and *C. crescentus*.

List of Publications

- **Tejal Agarwal**, G P Manjunath, Farhat Habib, Pavana Lakshmi Vaddavalli, and Apratim Chatterji. Role of special cross-links in structure formation of bacterial dna polymer. Journal of Physics: Condensed Matter, 30(3):034003, 2018.

- **Tejal Agarwal**, G. P. Manjunath, Farhat Habib, and Apratim Chatterji. Origin of spatial organization of dna-polymer in bacterial chromosomes. EPL (Europhysics Letters), 121(1):18004, 2018.

- **Tejal Agarwal**, G. P. Manjunath, Farhat Habib, and Apratim Chatterji. Bacterial chromosome organization. i. crucial role of release of topological constraints and molecular crowders. The Journal of Chemical Physics, 150(14):144908, April 2019.

- **Tejal Agarwal**, G. P. Manjunath, Farhat Habib, and Apratim Chatterji. Bacterial chromosome organization. II. few special cross-links, cell confinement, and molecular crowders play the pivotal roles. The Journal of Chemical Physics, 150(14):144909, April 2019.

Acknowledgements

The last four and half years of work for this thesis would not have been possible without the constant support and motivation of several people, and I would like to acknowledge them here.

Firstly, I would like to thank my thesis supervisor Dr. Apratim Chatterji for giving me the freedom to pursue my research. His constant advice has helped me to improve my writing and presentation skills. I express my gratitude to our collaborators Dr. G. P. Manjunath and Dr. Farhat Habib, for introducing such an interesting problem of chromosome organization to us and for all the help to learn essential biological terminology. I also would like to thank my RAC members Dr. Ranjith Padinhateeri and Dr. Vijay Chikkadi, for the fruitful advice during the meetings. I am thankful to Dr. M.S. Madhusudan, Dr. Arijith Bhattacharyay, Dr. Surajit Singh, and Dr. Prasad Subramanian for all the help and advice.

I am thankful to my parents and sister, who were with me in my good and bad situations as parents cum friends. Without their constant support, motivation, and believe I couldn't have completed my Ph.D. thesis.

I want to thank my present and past labmates, Devanshu, Alex, Neha, Debarshi. Devanshu is one of the most helpful friends I have got in IISER, who motivated me a lot throughout my stay in IISER. Neha a very kind and gentle person who always brought home-cooked food for me and saved me from the mess-food. Alex, a great friend, and lab-mate with whom I have shared many good memories, which I can never forget. I can always remember the long discussions about politics, college life with Debarshi, who was there at the time of my thesis writing.

I am thankful to my friend Ankita who was always there to listen to my talks and with whom I have shared long morning walks, cycle rides, and endless activities. I am thankful to Mayur, my chess, badminton, and table tennis partner, and a person who always makes me smile. I am grateful to Deepak, one of my first friends at IISER, and who was always there to help me. I want to thank my friend, Sayali, my first friend and roommate at IISER, a very matured and kind person I have met. I am also thankful to Yashwant, the tech-person of our batch, without whom I couldn't have got a great laptop. Also, thanks to being there with me in my bad moods and cheering me up. I also thank all batchmates from the 2014 i-Ph.D.

batch, specifically my Tamil-mess partner, Puneeta, Deepak Kumar, Kriti, Dipti, Shubham, Prashant, Soham, Vikash.

I am thankful to my other friends outside IISER, specifically Garima, Surabhi, Piyush, Ashwani, Rashmi. Garima is the best friend I have met during my undergraduate, but our friendship has become deeper even after those many years. She is always there to listen to my long-talks over the phone, and we never miss a chance to meet whenever I go to Delhi. Surabhi, my school friend and a very kind person who is always happy with all my success and achievements.

I want to acknowledge the support from my other family members, specifically my aunt, uncle, and grandparents, who always believed in me and supported me for all my decisions. I also would like to thank my cousins Ojas, Shreya, Prafull, Khushi, Rachit, Sakshi, Snehi, Achal, Rahul, and Piyush, who were always there to cheer me up and with whom I have shared so many good memories since childhood.

I extend my thanks to Prof. Suckjoon Jun for allowing me to do the experimental work in his lab at the University of California, San Diego, for six months. I thank all the lab-members of Jun Lab, specifically, Guillaume, Sarah, Dongyang, JT, Liyana, Fangwei. Guillaume a perfect and caring person I have met in the USA and who made my stay in San Diego pleasant. Sarah taught me all the experiments and helped me throughout to get familiar with American culture. I also would like to thank my other friends in the USA, Hamda, Can Cao, Srushti, Irene, Bhagyashree, who helped me and cherished me throughout my six months stay.

I would also like to thank the academic staff at IISER, specifically Tushar sir, Anjali ma'am, Prabhakar sir, and Dhanashree ma'am, who helped me a lot in getting signed all the paperwork in less time. Lastly, I am thankful to IISER Pune, DST, and IUSSTF for funding support for fellowship, funding support to attend Lindau Nobel laureate meeting, and visitation program through WISTEMM internship. The experience I have got through these is life-long and invaluable.

Table of contents

Chapter 1

Introduction

1.1 Introduction to DNA

DNA, the fundamental unit of the cell, is a long biopolymer which consists of two nucleotide chains (strands). Each nucleotide chain is made up of one of the four nitrogen bases A (adenine), T (thymine), G (Guanine), and C (cytosine), sugars, and acidic phosphate group. The two nitrogen bases from the two DNA strands are bonded by the hydrogen bonds to form a double-helical structure of the DNA firstly identified by Watson and Crick by the analysis of the X-ray diffraction experiments acquired by the Rosalind Franklin's group in 1953 [1]. The hydrogen bonding between the nitrogen bases of the two strands is not random but follows a rule that only A and T (base pairs (BP)) or G and C can form a hydrogen bond resulting double-helical structure of the DNA. The sequence of these base pairs is unique to one particular individual and contains the whole genetic information. The sequence of the base pairs is not random but a specific message for the mRNA to code the sequence of the amino acids for the synthesis of a specific protein, ribonucleic acids (RNA). These specific base-pair sequences, which are the code for the synthesis of the proteins, RNA are called the genes. The DNA in human cells have approximately 10^9 base pairs, and the bacteria *E. coli* has 4.6×10^6 BPs. Out of the billions and millions of BPs, the number of protein-coding genes in a human cell is 20000, while in the bacteria *E. coli*, the number of protein-coding genes is 4000 per cell [2].

The typical length of the bacterial DNA e.g., *E. coli*, *C. crescentus*, *B. subtilis* etc., if stretched straight is ≈ 1 mm while the total length of the DNA in a human cell is ≈ 2 m [2, 3]. In higher eukaryotes, e.g., human cells, the DNA is packed inside nucleus while prokaryotes (e.g., bacteria or archaea) do not have a nucleus wall and the DNA is found in the cytoplasm of the cell [2]. In bacteria, the region in which the total genetic material is found is called nucleoid. In *E. coli*, the nucleoid consist of the $15 - 25\%$ of the cell

volume [4]. The cell size of the *E. coli* is $\approx 1 \mu m^3$. Thus, the DNA in *E. coli* cell is compacted ≈ 1000 times inside the cell. This 1000 fold packaging of the DNA inside the cell is not like a random polymer because DNA has to perform many biological processes like transcription, translation, replication, etc. During transcription and translation, RNA polymerase attaches with the specific part of the DNA, copy the DNA sequence into the mRNA, and then ribosomes synthesize the proteins from the mRNA [2]. The process of transcription and translation are together called the gene expression. Gene expression is considered to be one of the most important processes in genetics, as the changes in the genetics of the cells have the observable traits (phenotype). In other words, the synthesis of specific proteins or enzymes through gene expression affects the pathways of the metabolism and the development of the cell. During replication, DNA in the cell makes another copy of itself (replicates) so that after the cell division, each mother and the daughter cell get one copy of the DNA.

In the nucleus/nucleoid of the cell (Eukaryotic/bacterial), the thread-like DNA molecule is packed tightly and dressed with various proteins. This resulting DNA macromolecule is called a chromosome. Chromosomes in higher organisms are generally not visible under a microscope except in the metaphase of the cell cycle (the stage just before the cell division), where it is found in the condensed form as an X-shaped structure. In higher organisms, this high level of condensation is achieved by the hierarchy of the structures [2, 5], see Fig4.1. First, the double helix DNA (2nm) is wrapped around histone proteins, resulting in 10nm wide beads on the spring-like structure, which is called nucleosomes. At this stage, ≈ 150 BP of DNA is wrapped around histone proteins, and two histone core particle wrapped with DNA is connected with linker DNA which consists of $10 - 80$ BPs. These nucleosome units further fold into a solenoid like structure resulting in 30nm fiber (chromatin), which gives the packing ratio of ≈ 50 [6]. This chromatin fiber further folds into a fiber of diameter $200 - 300$ nm, where each domain contains $30 - 100$ kBPs of DNA. This fiber further compacts in the metaphase stage of the cell cycle resulting in the X-shaped chromosome, which can be seen under microscopes. But, the mechanisms which lead to the organization of the DNA polymer above 30nm length-scales are poorly understood [3, 7], and the study of this thesis. From now on, we will use the term large-length scales to refer to the length scales above 50nm throughout the thesis.

Unlike higher organisms, bacteria do not have histone proteins, and the folding at small length scales is not achieved [7]. Thus, the chromosome organization in bacteria cannot be compared with the organization of the chromosomes in eukaryotes. However, the bacterial chromosome shows the chromosome organization at higher length scales, and DNA forms macrodomains, where each domain represents > 10 kBPs of DNA [7, 8]. Yet, bacteria have
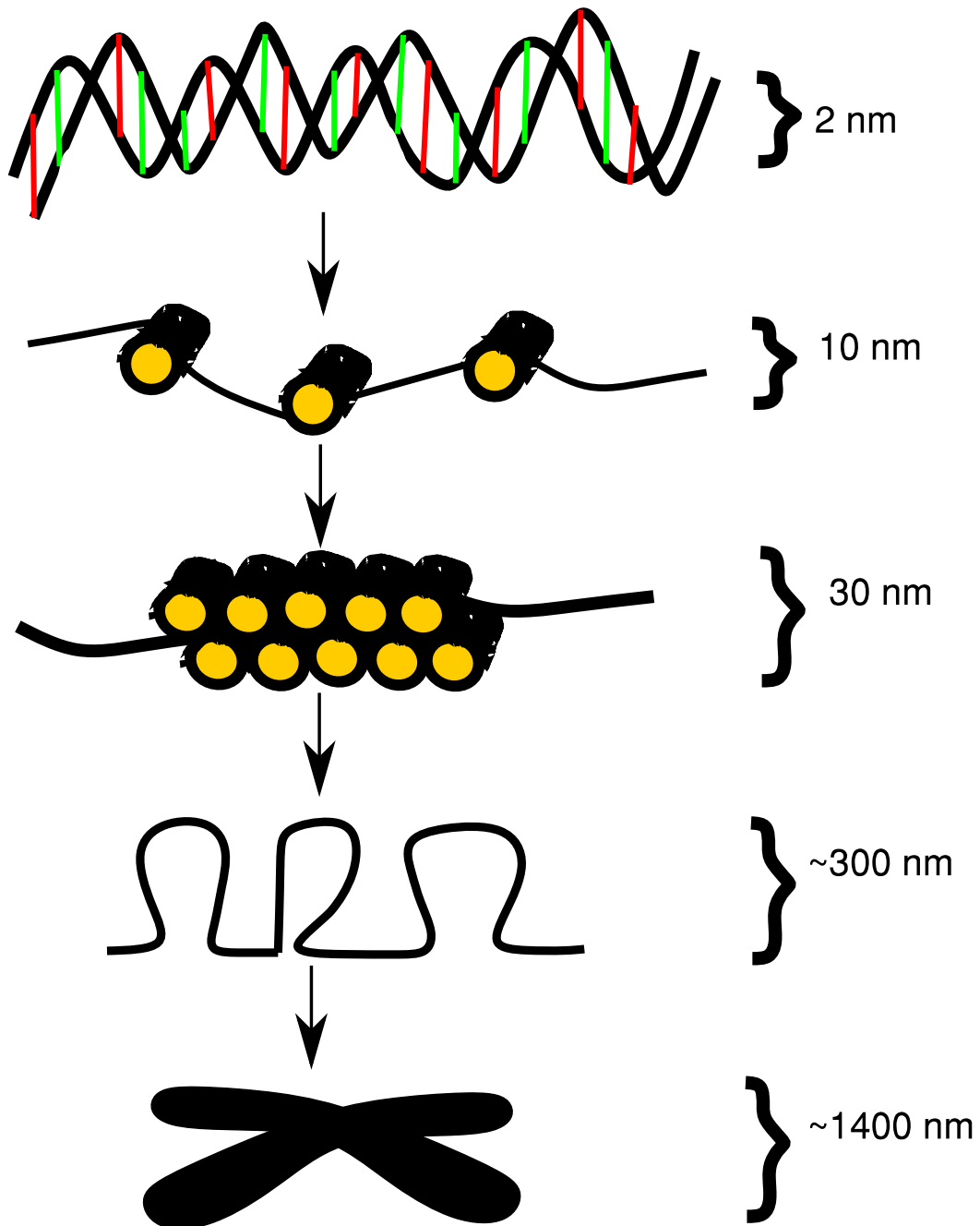
Fig. 1.1 The schematic figure shows the heirarchy of chromosome organization in higher eukaryotes.
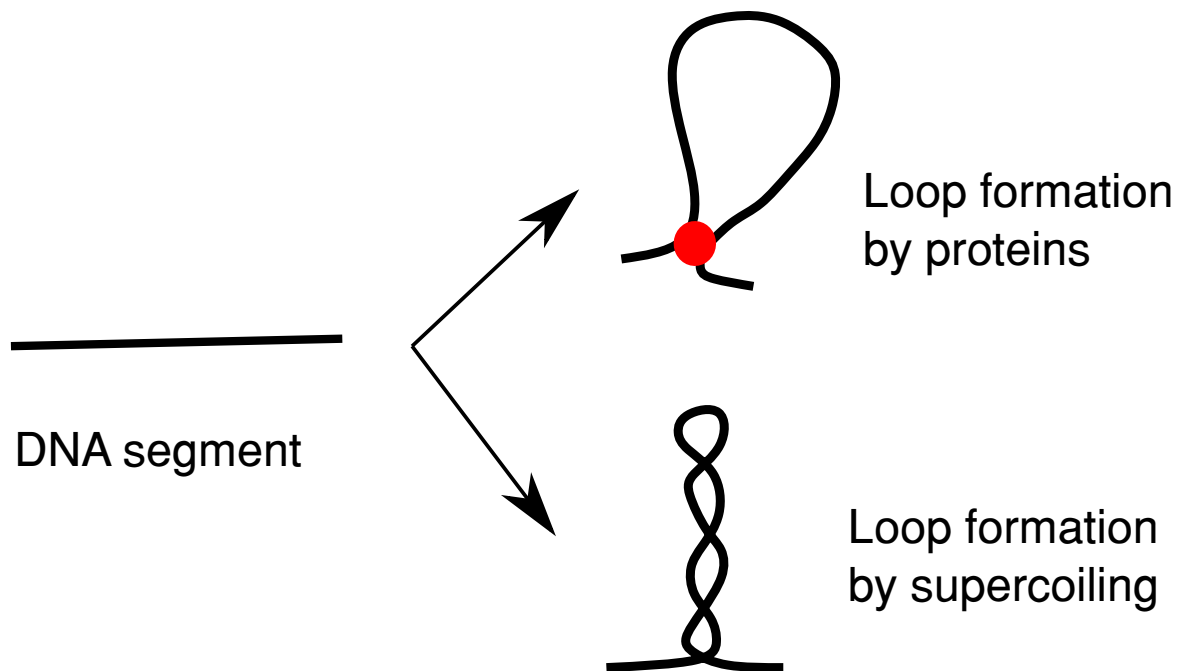
Fig. 1.2 The figure shows the two possible folding mechanisms of a DNA segment by the nucleoid associated proteins and by the supercoiling of the DNA.

nucleoid-associated proteins, e.g., H-NS, FIS, etc. Furthermore, most bacterial cells have a single circular chromosome, and the DNA is negatively supercoiled. Presumably, the supercoiling and the other nucleoid-associated proteins may help the chromosome structure at higher length scales by the formation of loops [7–16]. The loop is formed if the two distant segments (along the chain contour) of the chromosome are bound by a DNA binding protein or come closer to each other by the supercoiling (Fig.4.2).

As mentioned before, in human cells, the 23 pairs of chromosomes are packed inside a spherical membrane called nucleus, which ensures the separation of the genetic material (DNA) from the other organelles of the cell. In contrast, bacteria lack the nuclear membrane, and the single chromosome is found inside the cylindrical cell in the region called nucleoid. Moreover, the chromosome is not isolated in the cell but is in the crowding environment of the various proteins, RNA, enzymes, and cytosol, etc. Presumably, the organization and dynamics of the chromosome are also affected by the presence of this crowding environment and the confinement of the cylindrical cell wall [4, 17, 18]. In addition, the bacteria also have enzymes topoisomerase, which can help to release the topological constraints by cutting and rejoining the chain, unlike the conventional polymer where the chain crossing is strictly prohibited [19]. This release of topological constraints might help in a unique organization of the chromosome.

In this thesis, we will study the role of different mechanisms, e.g., nucleoid associated proteins, the effect of molecuar crowders, chain crossing, and the confinement of the cell wall in the organization of the bacterial chromosomes of *E. coli* and *C. crescentus*. I will start with briefly reviewing the experimental methods and the key literature on chromosome organization at large length-scales.

1.2   Experiments to determine chromosome structures at large-length scales

In the past two-three decades, several attempts have been made to study the large-length scale structure of the chromosome by various experimental techniques e.g., fluorescence in-situ hybridization (FISH), fluorescent repressor operator system (FROS) including fluorescently tagged DNA binding proteins or genome-wide methods e.g., chromatin immunoprecipitation (CHIP), chromosome conformation capture methods (3C, 5C and Hi-C) [8, 20–29, 7, 30–38]. In the following paragraphs, I will briefly review the various experimental techniques to obtain the chromosome organization.

1.2.1   FISH

In FISH, fluorescently labeled DNA probes are attached to the specific locus along the chromosome in the fixed cells (cells in which all the chemical processes are stopped). Fluorescent microscopy techniques are used to determine the positions of the fluorescent tags in the cell. This technique gives accurate and highly specific positons of the tagged loci, but the fixation of the cells can alter the chromosome structure. Furthermore, there are technical challenges to attach the probe at specific loci as the labeling yield is less than 100%.

1.2.2   FROS

This is another method to label the different chromosome segments fluorescently. In FROS, the chromosome is labeled using the fluorescent proteins in contrast to the DNA probes used in FISH. In other words, the fluorescent proteins are fused either in the repressor protein or in the ParB sites to track the location of different chromosome segments. Using this method, we can track the chromosome loci in the live cell in contrast to the FISH, where fixation of the cells was mandatory. Thus, using this method, the dynamics of the loci in a single live cell can be studied.

1.2.3   CHIP

This is a genome-wide method to identify the protein-binding sites in the chromosome. In this method, the DNA and the associated proteins are cross-linked and then fragmented using sonication. Then, deep sequencing on the fragments is used to identify the protein binding sites in the chromosome.

1.2.4   Chromosome conformation capture (CCC)

CCC is a relatively new molecular biology method to identify the segments which are in spatial proximity in the 3D space with each other but can be very far along the contour of the chain. These interactions may occur because of the random polymer looping, supercoiling, or looping by the DNA binding proteins. The basic experimental procedure involves cross-linking the whole chromosome by chemicals and then cut the cross-linked chromosome in smaller segments using the enzymes. Then, using high-throughput end to end sequencing of these DNA fragments, the population average probability matrix is generated, which gives the information about the interaction frequency of different segments of the chromosome with each other at that particular time. There are different variants of genome-wide conformation capture techniques, namely 3C, 4C, and Hi-C. In 3C (CCC), the interaction frequency between the two DNA segments is identified. This technique further developed to 4C, where the interaction frequency between one to many other DNA segments can be identified. Finally, in the Hi-C technique, the contact frequency between all the DNA segments with all the other DNA segments is obtained. Each segment represents 1 kBPs to 1 MBPs of the DNA [8, 20, 21]. The Hi-C experiments can be performed at different stages of the cell cycle to obtain the temporal organization of the chromosome. The contact maps obtained from the Hi-C experiments is shown in Fig. 1.3.

The two complementary methods, i.e., fluorescent imaging of the chromosomal loci and Hi-C, give critical insights about the chromosome organization. Specifically, from the Hi-C contact maps, we can identify two kinds of patterns by visual inspection: one is the plaid pattern, and the other is the self-interacting DNA segments, which appear as the blocks along the diagonal. These self-interacting patterns are called Topologically Associated Domains (TADs). The mechanism of the formation of these patterns is not fully known and is an active topic of research. The resolution in the Hi-C method is $\approx 1 - 10$ kBPs, which is relatively better compared to the FISH or FROS techniques, which has a resolution ≈ 100kBPs limited by the diffraction limit of the confocal microscopes [22, 24, 25, 28]. With the advent of super-resolution fluorescent microscopy, it is now possible to track the tagged loci at the nanometer scale (≈ 30 kBPs) [39, 40]. Another experimental approach
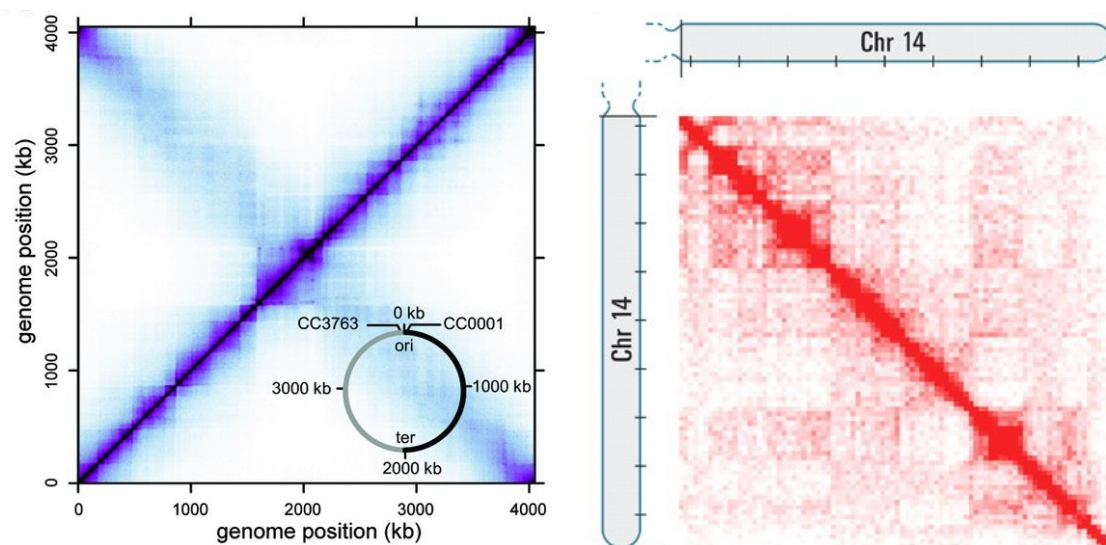
Fig. 1.3 The figures show the Hi-C contact map of the chromosome of bacteria *E. coli* and the human chromosome number 14. The x-axis and y-axis show the DNA segments at the resolution of 10kBP and 1MBP for *E. coli* and human chromosome number 14, repectively. The color show the probability of finding two DNA segments to be in contact with each other. The figures have been reprrduced from papers Le et.al. Science, **342** (6159):731–734, 2013 and Aiden et. al. Science, **326** (5950):289–293, 2009 with the permission from the publishers.

is to tag the chromosome loci using two-color quantum dots [41] and track the positions of these loci using electron microscopes. It is very recently that technological advances have made it possible to use these super-resolution techniques to obtain a large amount of quantitative data at the single-cell level. The single-cell data obtained from the experiments combined with theoretical modeling can help in understanding the chromosome organization, segregation, dynamics, and the mechanism of gene expression [22–28, 42]. We will also use the probability contact map data obtained from the Hi-C techniques in our bead-spring polymer model to study the chromosome organization in chapter 2, 3 and 4 of the thesis. While in the Appendix A of the thesis, we will present the data of the experiments based on fluorescent imaging techniques.

1.3 Bacterial growth law

The number of bacterial cells, if grown in suitable conditions, gets doubled after a fixed amount of time. For example, suppose, the *E. coli* cells are inoculated in a flask filled with transparent fresh growth media. With good shaking of the flask at 37°C, we can observe that the media becomes turbid after some time because of the cellular growth in the media.

Using a photo spectrometer, we can measure the change in the turbidity of the media by the quantity called optical density (OD), which in turn is the measure of total cell mass in the media [43]. If we plot the OD of the media v/s time in a semilog plot, we obtain the growth curve consisting of three phases. First is the lag phase, in which cells get adjusted to the new growth media and does not get doubled in constant time. The lag phase is followed by the exponential phase in which the cells get doubled after a constant time. The exponential phase is seen as the straight line in the semilog plot of OD v/s time. The exponential phase is succeeded by the stationary phase [44, 45]. In the stationary phase, the cells have exhausted all the nutrients from the media, and the total cell mass remains constant over time. During the exponential phase, the number of cells $N(t)$ at time t follow the following law [46]:

$$N(t) = N_0 exp(\lambda t) \tag{1.1}$$

Here, N_0 is the number of cells at time $t = 0$ and λ is growth rate. The growth rate $\lambda = \ln 2/\tau$; where τ is the generation time or the doubling time. The mathematical relation in equation 1.1 is called the growth law. Depending on the nutrient conditions in the media, the generation time τ can change. In other words, cells can grow faster in the nutrient-rich media compared to the minimal media.

Most of the experiments are done in the exponential phase of the cell culture. This is because after several generations in exponential phase, all the extensive properties of the cell culture, e.g., the number of proteins, enzymes, RNA etc increase exponentially with time, while all the intensive properties, e.g., the average size of the cell, average DNA content remain constant with time, thus reaches the steady-state. In steady-state, many physiological parameters of the cells, e.g., length, width, DNA, RNA, protein content per cell, can be measured using the microscopy and biochemistry experiments [45, 47]. However, the age distribution of the cells cannot be measured directly from the cell culture in the flask but can be analytically calculated from the length distribution of the cell culture or the length distribution can be calculated from the age distribution. For a given cell culture in the exponential growth, it can be shown that the length $\rho(l)$ and age $\phi(t)$ distribution are given as:

$$\rho(l) = \frac{l_d}{l^2} \tag{1.2}$$

$$\phi(t) = \frac{2ln2}{\tau} 2^{-t/\tau}, \qquad 0 \leq t < \tau \tag{1.3}$$

Refer [48] for the derivation of $\phi(t)$, the length distribution $\rho(l)$ can be calculated from $\phi(t)$ using change of variables method and assuming the exponential elongation of the cell

length with age, i.e. $l(t) = l_0 exp(ln2 * t/\tau)$. In the above equations, l_d is the length of the cell at the division, and τ is the generation time. For the study in this thesis, we will study the cells in the exponential phase.

1.4 Bacterial cell cycle

The cell cycle is the series of events in which cells double each organelle inside the cell, such that the daughter cells after the cell division get one copy of all the cellular organelles. The most important part of the cell cycle is the replication of the DNA, which is followed by the division of the cell. In eukaryotes, the cell cycle consists of two stages, namely, interphase and mitosis. In interphase, the cell replicates its DNA and accumulates all the nutrients required for the cell division, while in the mitosis stage, the cell divides into two daughter cells [2, 5].

In contrast to the eukaryotic cell cycle, the bacterial cell cycle is divided into three parts, namely B, C, D period. The B period is time duration between cell division and the initiation of the DNA replication, C period is time taken for the DNA to get duplicated, and the D period is time between the end of DNA replication to cell division (see fig.4.4). The combined time period $C + D$ is termed as the cell cycle time τ_{cyc} [2, 5, 45]. Another important quantity is the doubling time, which is time duration between the cell birth and cell division. In slow growth conditions, the doubling time is equal to $\tau = B + C + D$ and in faster growth conditions the cell exhibits multifork replication (defined in the next section) and the B period does not exist.
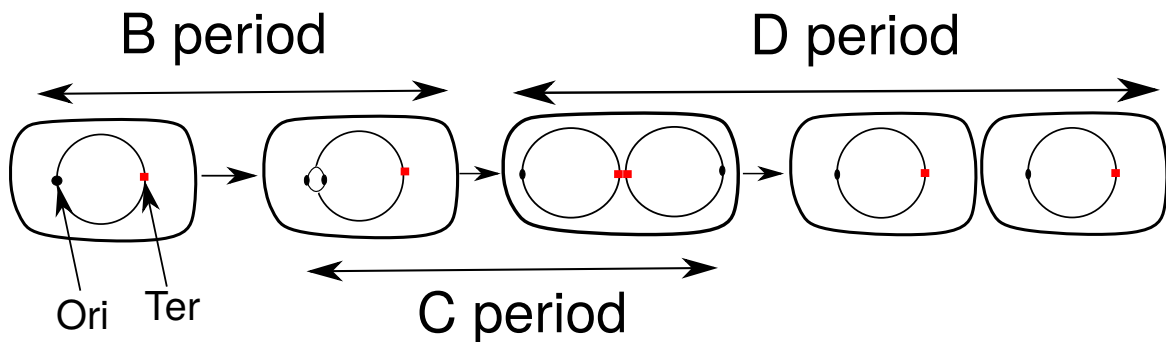
Fig. 1.4 The figure represents the cell cycle of a bacterial cell.

In bacteria, DNA replication starts at a specific position in the chromosome termed as Ori. The replication proceeds bidirectionally along the left and right arms of the chromosome from Ori and stops at the location called Ter. The average speed of the replication across two arms is the same, i.e., ≈ 1000 BP/s [45]. For bacteria *E. coli* in fast-growth conditions

(doubling time < 60 minutes), the C period is experimentally measured to be ≈ 40 minutes and the measured D period is ≈ 20 minutes. Thus, the combined time period $\tau_{cyc} \approx 60$ minutes [49, 50].

1.5   Multifork replication

If the bacteria *E. coli* and *B. subtilis* are grown in nutrient-rich media, the cells double in $\approx 20 - 30$ minutes [49–51]. However, as mentioned before $\tau_{cyc} = C + D = 60$ minutes. In other words, if it takes 40 minutes for the cells to duplicate its chromosome, then how the cells are dividing in 20 minutes with each daughter cell getting at least one copy of the chromosome?

This puzzle was solved by Helmstetter and Cooper in the 1970s, where they show that in fast-growth conditions, the cells exhibit multifork replication [52]. In multifork replication, the multiple cell cycles overlap such that the ongoing round of replication in the current generation initiates in the mother or grandmother generation, and the cell is born with a partially replicated chromosome (see fig. 1.5).

The idea of Helmstetter and the Cooper model was to set the cell division as a reference point and then trace back the initiation of replication from this reference [45]. For example, suppose the doubling time of *E. coli* cell is 25 minutes. Now, take the cell division as a reference point, trace the initiation of the DNA replication, which was initiated 60 minutes before. If we consider doubling time to be 25 minutes, the cell in the current generation was born 25 minutes before, the mother was born 50 minutes before, and the grandmother was born 75 minutes before. This implies that the DNA replication in the current generation has been initiated in the grandmother cell, and the three cell cycles overlap. If we know the C, D period and the generation time, we can trace how many overlapping cell cycles exist for the cells in a given growth media.

Assuming the age distribution of eq. 1.3, it can be shown that the average number of ori and ter in multifork replication is given as [53]:

$$\langle N_{ori} \rangle = 2^{(C+D)/\tau} \tag{1.4}$$

$$\langle N_{ter} \rangle = 2^{D/\tau} \tag{1.5}$$

Unlike *E. coli* and *B. subtilis*, the bacteria *C. crescentus* does not show the multifork replication, and its cell cycle is more similar to the eukaryotic cell cycle where all the stages of the cell cycles are separated in time [54, 55].
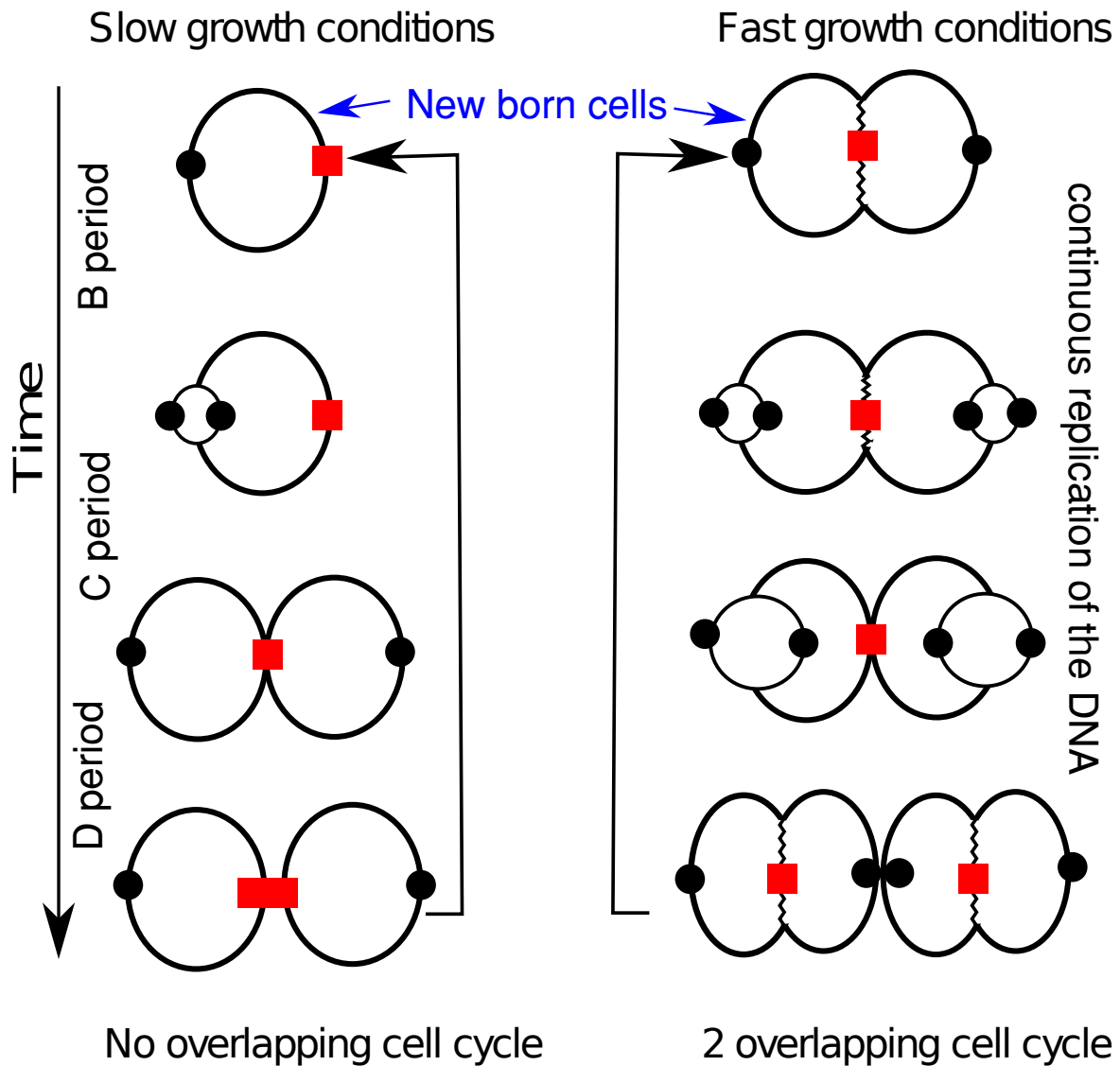
Fig. 1.5 The schematic diagram show the cell cycle in the no overlapping cell cycle and multifork replication (two overlapping cell cycle) cases, respectively. The figure is adopted from the paper Youngren et. al., Genes and Development, **28(1), 2014**

1.6   Spatiotemporal organization of the chromosome in different growth conditions

It is known from several experimental studies that the average size of the bacterial cell changes in different growth media [45, 49, 50, 56–58]. In particular, in bacteria *E. coli* length and width of the cell increases in the rich nutrient conditions in contrast to the cell of bacteria *B. subtilis* where only the length of the cell increases with improved nutrient conditions [45, 49]. In other words, the *E. coli* cells become longer and fatter in the rich

media. Consequently, for *E. coli* the aspect ratio (*length/width*) remains constant over different growth conditions while in *B. subtilis*, the cell becomes longer, and width remains constant for various growth conditions. Furthermore, the bacteria *E. coli* and *B. subtilis* increase the replication complexity as the nutrient conditions are improved. Most of the structural and dynamical studies of the chromosome in *E. coli* (the most studied bacteria) have been limited to the slowest growth conditions because of the simplicity of the cell cycle in slow growth conditions [24, 25, 28]. Recent studies have shown that the chromosome organization in the slowest growing cell and faster-growing cells is qualitatively different [22, 24, 25, 28, 59]. For example, in the slowest growing *E. coli* cells, the left and right arms of the chromosome occupy each cell half along the *longitudinal axis*. While during mulifork-replication, the left and right arms are arranged in each cell halves along the *radial axis* of the cylindrical cell (see Fig. 1.6).

Some of the longstanding questions in the literature are, why the chromosome organization changes in different growth conditions and which mechanisms drive these changes? What will be the chromosome organization in case of multiple overlapping cell cycle?

1.7 Role of polymer modeling in the chromosome organization

Modern experimental techniques have given essential insights about the chromosome organization and dynamics at different length scales and time in the cell cycle[8, 20–22, 24, 25, 28, 31, 33, 60–69]. But, the physical mechanisms leading to the large-scale organization of the chromosome remains elusive despite several efforts. Furthermore, experimentally exploring a large parameter space is neither cost nor time effective because of the complexity in the biological systems. Thus, there is a need for simple theoretical models that can explain the existing experimental data and make predictions to reduce the size of the experimental parameter space. But, because of the intricacy of the biological system, modeling the entire system with today's computational facilities is challenging. For example, the simplest bacterial cell is abundant with millions of proteins, enzymes, millions of base pairs of DNA, etc., hence all atom modeling and studying the dynamics of billions of particles will be quite challenging even with the help of supercomputers.

In this regard, coarse-grained approaches and statistical physics principles have been used widely in the past and were proven successful to give the important physical insights about the problem [59, 70–104]. In the coarse-grained approach, the DNA polymer is modeled as a bead-spring polymer where each bead represents 10 BPs to 1 MBPs depending on the
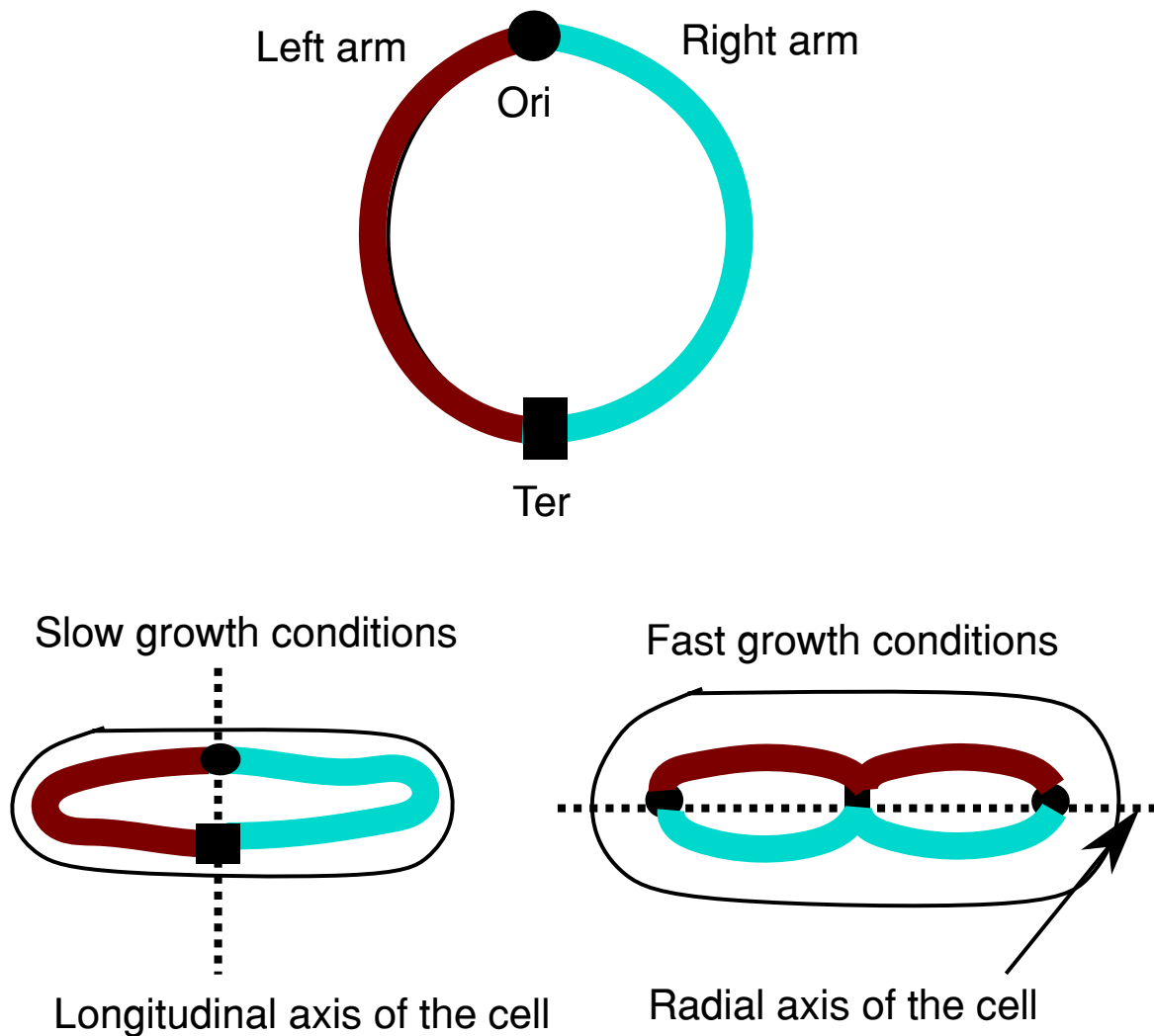
Fig. 1.6 The schematic diagram show the organization of the chromosome in the slow and fast growing conditions.

length scale of the study. In particular, it was shown that the chromosomes in the human cells are not like an equilibrium polymer globule but exists in a non-equilibrium, unentangled state called crumpled/fractal globule [20, 70, 77, 78], which will be explained later in this chapter. Furthermore, all the 23 pairs of chromosomes in the nucleus of the human cell are not in the mixed state but occupy well-defined space with respect to each other in the nucleus called chromosomal territory. Polymer physics and statistical physics have shed light on the physical mechanism of the formation of this non-equilibrium state (fractal globule) of the chromosome organization. Simulation techniques like Monte-Carlo, Brownian dynamics, molecular dynamics are widely used to study the biological systems in the past [8, 105–112]. However, these techniques rely on the simplified assumptions as the interaction potentials

used in these kinds of simulations are the simplified approximations and do not represent the actual atomic interactions between the particles of the system. In the next section, we will briefly review standard polymer physics terminology and the models of the chromosome. For the detailed review of the polymer physics, we will ask the interested readers to look at the classical literature of the field in [113–118].

The polymer is a long chain consisting of many repeating structural units called monomers. In reality, the polymer can be highly flexible, where at very large length scales, it appears as a random walk in 3D space (spaghetti like structure) or it can be very stiff like a rod or the case in between the two extremes (highly flexible or rigid) exists as well which is called the semiflexible chain. The length scales at which the polymers can be considered flexible, semifleixible or rigid can be defined by the quantity called persistence length.

1.7.1 Persistence length

Persistence length is the measure of polymer stiffness. In other words, persistence length is the measure of the length over which the correlation of the tangential vectors in the polymer is lost.

Let r_0 and r_j are the two tangent vector of a polymer segment at position 0 and j along the contour. Suppose θ is the angle between these two vectors, r_0 and r_j. Then, it can be shown that the average of the cosine of the angle decreases exponentially along the length of the polymer.

$$\langle cos(\theta) \rangle = exp(-j/P)$$

Here, the variable P is called the persistence length.

Generally, the polymer with length much less than the persistence length can be thought of as a rigid or rod-like chain, while the polymer with a length above persistence length can be modeled statistically as a random walk in the 3D space. The persistence length of the double-helical DNA is $\approx$ 150BPs or 50nm.

1.7.2 Kuhn length

Kuhn length is the theoretical concept to describe a polymer chain in which the chain is composed of N Kuhn segments of size a. These Kuhn segments can be thought of as freely joined with each other. The Kuhn segments can take any position and orientation in the space like a random walk in 3 dimensions. The length of the fully stretched polymer can be written as $L = Na$, where N is the number of Kuhn segment in the chain. Typically, Kuhn's length is

double the persistence length for the worm-like chain and we can consider the polymer to be flexible above kuhn length.

For double-helical DNA, the Kuhn length is ≈ 300BPs or 100nm.

1.7.3 Mean square end-end distance and radius of gyration

In polymer physics, a polymer consisting of many elementary repeating units called monomers can be thought as made up of $N+1$ coarse-grained units with position co-ordinates $\vec{r}_i = \vec{r}_1, \vec{r}_2....\vec{r}_{N+1}$ and blob diameter b. In the context of DNA polymer, each coarse-grained unit represent few base-pairs to 1MBPs of DNA depending on the length-scales of the study.
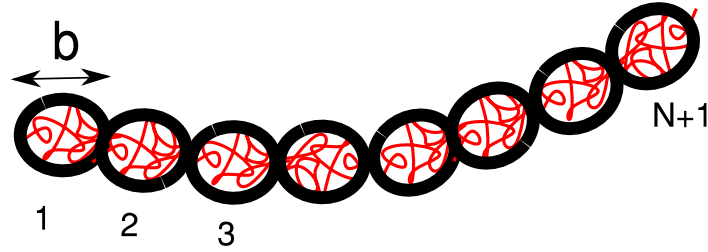
Fig. 1.7 The schematic show the coarse-grained polymer consisting $N+1$ coarse-grained unit. The diameter of each blob is b.

The end to end vector for a polymer consiting of $N+1$ monomer units is defined as

$$\vec{R}_{ee} = \sum_{i=1}^{N+1} \vec{u}_i$$

The $\vec{u}_i = \vec{r}_{i+1} - r_i$ is the bond vector between two adjacent monomers along the chain. For a random walk polymer, $\langle R_{ee} \rangle = 0$ and it does not give the information about the extent of the polymer chain. Thus, the quantity of importance is mean square end to end distance which is given as

$$R_{msee} = \langle \vec{R}_{ee}^2 \rangle$$

The mean square end to end distance defines the size of the linear polymer very well, but for the ring polymer or branched polymer, this quantity is not useful. Hence, the more vital quantity to define the extent of the polymer of any topology (including linear chains) is the radius of gyration R_g, which is defined as

$$R_g^2 = \frac{1}{N+1} \sum_{i=1}^{N+1} (\vec{R}_i - \vec{R}_{cm})^2$$

Here, $\vec{R}_{cm}$ is the position vector of the center of mass of the polymer chain. The radius of gyration R_g and mean square end to end distance R_{msee} are related as follows:

$$R_{msee} = 6R_g^2$$

1.7.4   Ideal chains

The simplest model to describe the flexible polymer is an ideal chain (gaussian chain), model. In this model, the polymer is assumed to perform the random walk in the 3D space, and any other kind of interactions are neglected. The monomers/blobs only interact only with the adjacent monomers along the chain contour, and two or more monomers can occupy the same positions in space.

For an ideal polymer chain, we assume that all the bond vectors are independent and identically distributed random variables. Thus, using the central limit theorem, it can be shown for $N >> 1$ the end to end distance R_{ee} follows the gaussian distribution with zero mean and unit variance. Hence, the polymer can be thought of like the string of monomers connected with each other by a harmonic spring.

The mean square end to end distance for the freely jointed chain is.

$$R_{msee} \approx Na^2 \tag{1.6}$$

Here, a and N represents the size and number of Kuhn monomer in the chain, respectively. Thus, root-mean-square end to end distance R scales as $R \sim N^{1/2}a$ for an ideal chain.

In theory and simulations, the interaction between two adjacent monomers of the polymer chain can be modeled by the harmonic spring potential of the form

$$v_{spring} = \kappa(r_i - a)^2 \tag{1.7}$$

The parameters a and κ represent the bond length and the spring constant, respectively.

Another interaction potential, which has been used widely in the literature to model the polymers is finitely extensible non-linear elastic (FENE) potential of the form

$$V_{FENE} = -\frac{3\kappa_e r_0^2}{2a^2} \sum_{i=1}^{N+1} ln\left(1 - \frac{u_i^2}{r_0^2}\right)$$

The parameter r_0 is the distance above which the non-linear effects start coming into the picture, and κ_e is the spring constant of the non-linear spring.

However, for the work in this thesis, we will use V_{spring} to model the interaction between adjacent monomers of the chain.

1.7.5 Real chains

The polymers, in reality, are not ideal chains in which the interactions between the distant monomers along the chain is completely ignored. But, these interactions can affect the conformations and dynamics of the chains, significantly. Moreover, in real chains, two monomers cannot overlap each other, and the space occupied by one monomer is not available for another monomer. This type of interaction is called excluded volume interaction. Therefore, the real chains are modeled as self-avoiding 3D random walk where the positions taken by one monomer is not available for other monomers of the chain.

In self-avoiding, random walk, the scaling behavior of R_g of a random walker is changed from $R_g \propto \sqrt{N}$ to $R_g \propto N^{0.586 \pm 0.004}$ in the limit $N \to \infty$. This scaling law had been obtained in the experiments on a dilute polymer solution of linear polymers [119]. From the mean-field Flory theory, on the dilute polymer solution, it can be shown that $R_g \propto N^{0.6}$. However, more mathematically involved calculations (renormalization group) gives the scaling for a self-avoiding random walk as $R_g \propto N^{0.588 \pm 0.0015}$ [120]. For self-avoiding ring polymers R_g scales as $R_g \sim N^{0.65}$ [121]. Which suggests that the scaling is independent of the chemistry, and the simplest self-avoiding random walk model can define the atomic interactions between the monomers and the properties of the polymer.

Furthermore, the structure and dynamics of the polymer also depend on the quality of the solvent. Depending on the interaction between monomers and solvent molecules, three kinds of solvents are defined in the polymer literature.

Good solvent

In a good solvent, the interaction between the monomers and the solvent is energetically favorable. In other words, the polymer likes the solvent molecules and becomes swollen. This can be thought of as the repulsive interactions between the monomers of the polymer. In this case, the radius of gyration scales as $R_g \propto N^{0.6}$.

Poor solvent

In poor solvent, the interaction between the monomers and solvent molecule is energetically unfavorable, and as a result, the monomer-monomer attractive interactions are enhanced. In poor solvent, the polymer contracts, and it forms a dense droplet conformation called a polymer globule. In this case, the radius of gyration scales as $R_g \propto N^{1/3}$.

Theta solvent

In theta solvent, all the monomer-monomer interactions are screened, and the polymer behaves like an ideal chain with R_g scales as $R_g \propto N^{1/2}$. In this case, the solvent is poor enough, such that it cancels the effect of excluded volume interactions.

Furthermore, the quality of the solvent can be changed by varying the temperature. At higher temperatures, the excluded volume effect plays a major role, and polymer expands. At very low temperatures, the interaction between the monomers dominates over thermal energy $k_B T$, and the polymer shrinks and forms a globule. At intermediate temperature, the excluded volume (repulsion between monomers) and monomer-monomer attraction cancel each other, and the polymer behaves as an ideal chain. The temperature at which the polymer behaves as an ideal chain is called the theta temperature. Therefore, changing the temperature results in the phase transition of the polymer from the coil (swollen chain) to a globule state.

In coarse-grained simulations, the effect of excluded volume between two non-bonded monomers is modeled by the repulsive part of the lennard-jones potential called Weeks Chandler Anderson (WCA) potential of the form:

$$V_{ev} = \begin{cases} 4\varepsilon \left[\left(\dfrac{\sigma}{r}\right)^{12} - \left(\dfrac{\sigma}{r}\right)^{6} \right] + V_{ev}(r = 2^{1/6}\sigma), & \text{if } r \leq 2^{1/6} \\ 0, & \text{otherwise} \end{cases} \tag{1.8}$$

Here, ε is the strength of the potential in the units of thermal energy $k_B T$. The parameter σ represents the diameter of the spherical monomer, and r is the distance between two monomers.

1.7.6   Charged polymers

DNA and the proteins are charged molecules and interact with themselves and each other electrostatically. But, because of the presence of ions in the cytoplasm of the cell, these electrostatic interactions are screened. From the Debye-Hückel theory, it can be shown that the screened electrostatic interaction decays exponentially with distance [113]. The characteristic length scale r_{DH} above which the interactions are fully screened is ≈ 1nm for the salt concentration $0.1M$ at $25°$ celsius under physiological conditions [4]. Therefore, at the length scales of the study in this thesis, i.e., > 30nm, we can neglect the effect of the electrostatic interactions.

1.7.7 Worm-like chains

Very stiff polymers can be modeled as worm-like chains. In these models, bond-angles are constraints to take small values, and large bond-angles are energetically unfavorable. For example, the double helix DNA is a very stiff polymer (persistence length $\sim$ 150BPs or 50nm) and is modeled as a worm-like chain. But at the coarse-grained scales ($>$ 150BPs), the DNA polymer can be modeled as a flexible polymer.

For the coarse-graining in this study, we will model the DNA polymer as a flexible chain.

1.7.8 Suggested models of chromosome architecture

For a polymer globule, the end to end distance of the whole polymer chain scales as $R \sim N^{1/3}$, as mentioned previously. In the literature, there are two kinds of globules suggested, namely equilibrium and fractal, to explain the chromosome architecture for higher organisms [74, 77, 94].

Equilibrium globule

In equilibrium globule, the end to end distance of the sub-chain scales as $R(s) \sim s^{1/2}$. Note that the exponent differs from the end to end distance of the whole chain, which has the exponent $\mu = 1/3$. In other words, the subchain can be thought of as behaving like a random walker. It was argued that the end to end distance of the subchain in case of equilibrium globule scales as:

$$R_{eq}(s) \sim \begin{cases} s^{1/2}, & \text{if } s \leq N^{2/3} \\ constant, & \text{otherwise} \end{cases} \tag{1.9}$$

Also for equilibrium globule, the contact probability of subchain scales as

$$P_C(s) \sim \begin{cases} s^{-3/2}, & \text{if } s \leq N^{2/3} \\ constant, & \text{otherwise} \end{cases} \tag{1.10}$$

In addition, it is found in several theoretical and simulation studies that the equilibrium globule is an entangled state of the polymer which is abundant with polymer knots [70, 77, 78]. It is also mentioned that the formation of equilibrium globule is a very slow process (equilibrium time N^3).

Fractal globule

Fractal globule is another suggested globular form of the polymer introduced by Grosberg in 1988. Fractal globule is a long-lived, unentangled, and non-equilibrium state and emerges because of the topological constraints.

In this case, the end to end vector of the subchain scales with the length of the subchain as:

$$R(s) \sim s^{1/3}$$

and the contact probability of the subchain scales with the length of the subchain as:

$$R(s) \sim s^{-1}$$

It is argued in [20] that the scaling laws obtained in the Hi-C experiments of human chromosome are consistent with the fractal globule model. But another study in [122] shows that the same scaling laws can also be obtained for the knotted and entangled polymers. The knotted and entangled polymer model is also consistent with the topological complexity in the human chromosome. Moreover, fractal globule architecture has not been found for bacterial chromosomes.

1.7.9   Other polymer models

To explain the organization obtained in the Hi-C contact map, more detailed polymer models have been proposed in the literature. Some of the models are data driven [65, 109, 123] which use Hi-C contact frequency as an input to the model and other models try to explain the patterns in the Hi-C contact map from the first principles [81, 87, 88, 91, 96].

In one of the first principle model namely switch and binder model [87], the chromosome is modeled as a bead-spring co-polymer consisting of different kind of monomers and the proteins are modeled as diffusive particles. The diffusive particles can bind or unbind with different monomers of the polymer with different binding affinities. The size of the particles, number of the diffusing particles, the binding affinities of the particles etc. are taken as parameters. By optimizing multiple parameters researchers reproduce the Hi-C contact map.

Another successful first principle model to explain the Topologically Associated Domains (TADs) in the contact map is loop extrusion model [88, 91]. To reiterate, TADs are the genomic regions in the chromosome which interact with each other more, compared to the other regions of the chromosome and appear as blocks along the diagonal in the contact map. In the loop extrusion model, there are boundary elements (BE monomers) at specific sections along the chains. A pair of special monomers (LE-monomers), probabilistically bind with

each other to extrude loops of variable lengths by diffusing/translocating along the length of the chain, but LE-monomers are constrained to remain bounded between the BE-monomers. Again, a search through a large parameter space leads to optimal TADs with a quantitative match with experimental data.

In data driven models of the chromosome, the contact frequency data from the Hi-C contact map is taken to generate the equilibrium configuration of the polymer. It is assumed that the probability of contact (p_{ij}) of two segments i and j in the Hi-C contact map is inversely proportional to the distance (d_{ij}) between them in the 3D space [123]. With this assumption, all the segment pairs i, j are bound by harmonic potential where the spring constant or equilibrium distance of the bond is chosen according to the contact probability of segments i, j in the Hi-C contact map. In another approach [109], all the segments are bound by the harmonic spring potential and researchers try to find the spring constant matrix κ_{ij} between all the segment pairs such that the contact probability in the simulated system matches with the contact probability in the Hi-C conatct matrix. But, one important caveat in the interpretation of Hi-C data is that the contact matrix does not represent the contact of chromosome in one cell but represents the average contact frequency in the population of cells. Thus, all the contact probabilities in the Hi-C matrix may not represent the true contact between the polymer segment in the 3D space.

It is also in the previous theoretical and experimental studies that the DNA polymer compacts and collapses in the presence of the crowding environment (DNA condensation) [124, 125]. In the presence of the charged molecular crowders, the "effective" interactions between DNA-DNA segments increase which leads to the segregative phase separation [4, 124, 126]. The segregative phase separation may happen if two crowders (e.g., protein, RNAs, etc.) or a crowder and DNA segments of the same charge repel each other more strongly than two different segments of the DNA [4]. Another simulation study of ring polymers with uncharged molecular crowders and cylindrical confinement shows that increasing the density of the mutually repulsive crowding molecules (modeled by repulsive Weeks-Chandler potential) leads to the decrease in the radius of gyration R_g of the ring polymer due to the depletion interaction between the polymer and the molecular crowder [124]. In particular it was shown that, the component of R_g along the longitudinal direction of the cylinder decreases $\approx 8\%$ when the volume fraction ϕ of the crowders is increased from $\phi = 0$ to $\phi = 0.3$, while the component of R_g along the radial axis of the cylinder does not change significantly. These and other studies which incorporate dissimilar crowders established that the crowding environment primarily leads to the compaction of the DNA molecule in the nucleoid region inside the cell [127]. Thus, the effect of molecular crowders can be modeled as the weak attractive attraction between the monomers. These studies give important insights

about the role of molecular crowders in the organization of the DNA polymer but completely neglect the role of either DNA-binding proteins or chain crossing. Researchers model the activity of enzyme topoisomerase explicitly by taking a suitable potential to allow chain crossing [8]. But, we introduce different extents of ease of release of topological constraints in our model by systematically changing the diameter of the monomer bead. It varies from the value where chain crossing can occur freely to the value where chain crossing is disallowed.

Furthermore, it is argued in several studies that the confinement plays a crucial role in the organization of the chromosomes in bacteria as well as in higher organisms [8, 18, 36, 71, 93, 128–135]. In higher organisms, the chromosomes are confined by the spherically shaped nucleus membrane, and in bacteria, chromosomes are confined within the capsule shaped cell membrane. In a recent study of polymers under spherical confinement, it has been shown that the confinement slows down the dynamics of the polymers (glassy-dynamics) and the glassy dynamics helps in the segregation of the chromosomes in the nucleus for the human chromosomes [130]. Other studies on bacterial chromosome organization took the cylindrical confinement of the cell into account and model the chromosome as a bottle-brush polymer with several loops emanating from the central backbone, and the backbone is arranged parallel to the length of the cylinder [8, 131]. In another separate study of bottle-brush polymers (backbone and side-loops), it is reported that the backbone attains the helical structure spontaneously when confined to the small cylindrical volume and this structure is maintained for the different aspect ratio of the cylinder for high packing fractions [71]. Several researchers have also reported the dynamics of the polymer under cylindrical confinement for various aspect ratios [93, 124, 132]. In these papers, it is pointed out that in the presence of cylindrical confinement, the segregation of the two ring polymers is entropically driven.

1.8 Scope of the work in the thesis

Taking the limitations of other polymer models (optimization of multi-parameters in the first-principle models or taking all the data of the contact map in data-driven models) into consideration, we cross-link only a few specific segments which are in spatial proximity with each other with very high probability in the Hi-C contact map. Next, we try to find the minimum number of cross-links, which gives the organization to the DNA polymer consistent with the experimental contact map. We do not aim to explain the patterns in the contact map like other theoretical studies but address a different question that if we are given the contact map, can we predict the overall 3D organization of the DNA polymer? Note that the contact map only gives only the information about two segments to be in proximity and does

not give the overall 3D organization. We will also study the effect of molecular crowders, chain crossing and the confinement effects of cylindrical wall in the 3D organization of the polymer with a few specific cross-links.

1.9 Aim and organization of the thesis

This thesis aims to obtain the 3D organization of two bacterial chromosomes *E. coli* and *C. crescentus* at μ length scales using the bead-spring model of the polymer and using the data from the Hi-C contact maps. Moreover, we focus on the effect of different mechanisms (specific cross-links, chain crossing, molecular crowders, confinement of the cell wall) in the large-length scale structure of the bacterial chromosomes during the period of the cell cycle when the DNA is not replicating. We hypothesize that, if two specific segments of the chromosome are cross-linked (bound together) in the bead-spring polymer, can give rise to the specific organization to the polymer. These cross-links in the DNA polymer can represent the effect of DNA binding proteins or random looping of the polymer. We determine the positions of the specific cross-links from the Hi-C contact maps of the bacteria *E. coli* and *C. crescentus*. In other words, if two segments are in contact with each other with high probability in the contact map, then we cross-link them in our bead-spring polymer model. By setting different frequency cut-off in the contact map, we can have a different number of cross-links. Thus, we aim to find the minimum number of cross-links that are required for the polymer to get organized into a particular structure. We then address the question: What will happen if the positions of the cross-links are chosen randomly rather than from the experimental contact maps? We use the Monte-Carlo simulations to generate the different microstates of the system. We first study the role of different mechanisms, e.g., chain crossing and the effect of molecular crowder (modeled by the weak attractive LJ potential) and the confinement of the cell wall in the organization of the polymer with cross-links, separately. Next, we combined all the effects and then predict the overall 3D organization of the DNA polymer. As a complementary study, we also look into the organization of the chromosome of bacteria *E. coli* during multifork replication throughout the cell cycle. The thesis is organized as follows.

In chapter 2 of the thesis, we present the work published in [136, 137] where we study the organization of the DNA polymer of bacteria *C. crescentus* and *E. coli* at large length scales in the presence of different number of cross-links whose positions are chosen either from the Hi-C contact map or randomly. Here, we do not incorporate the effect of cell wall confinement. We also find the minimal number of cross-links required for the polymer to get organized in a unique structure. We then predict the overall 2D organization of the

chromosome of bacteria *E. coli* and *C. crescentus* from our simulation and validate our predictions with the available experimental data.

In chapter 3 of the thesis, we study the crucial role played by the crowding environment of the cell in the organization of the chromosome. We also study the role of the release of topological constraints in the structure and architecture of the bacterial DNA polymer. We obtain the suitable parameters from this study and use these parameters for the study in chapter 4. The work of this chapter has been published in [138].

In chapter 4, we first study the effect of confinement of the cell wall in the organization of the ring polymer. We then introduce the cross-links and the effect of molecular crowders and study the combined effect of cross-links, molecular crowders, and chain crossing in the structure of the bacterial chromosomes of *C. crescentus* and *E. coli*. We take the suitable parameters from the study of chapter 3 and introduce the effect of chain crossing and molecular crowders in the polymer model. The work in this chapter is published in [139].

Note that the studies in chapters 2, 3, 4 correspond to the period of the cell cycle when the DNA is not replicating, and the cell can be assumed in the state of local equilibrium. Thus, we can apply the principles of equilibrium statistical physics to study the structure of the bacterial chromosomes.

In Appendix A of the thesis, we present the experimental studies to study the organization of the chromosome of bacteria *E. coli* during multifork replication over the period of the cell cycle where two or more cell cycles overlap.

The study in chapter 2, 3 and 4 of the thesis is carried out in IISER Pune, India while the experimental work of Appendix A has been performed in the lab of Prof. Suckjoon Jun in University of California, San Diego, USA.

Chapter 2

Role of special cross-links in the polymer organization

2.1   Introduction

In this chapter, we discuss the organization of the polymer in the presence of the different number of cross-links, where the positions of the cross-links are either chosen from the Hi-C contact map of the two bacteria *E. coli* and *C. crescentus* (Biological cross-links) or chosen randomly (random cross-links). We use the bead-spring model of the polymer physics to study the organization of the bacterial chromosome, where each monomer in the coarse-grained polymer represents 1000 BPs of DNA. Furthermore, we study the chromosome organization in the stage of the cell cycle when the DNA is not replicating, and we assume the cell to be in the local equilibrium. We aim to find the minimal number of cross-links that are required for the polymer to get organized into a particular structure. We next compare the polymer organization in the presence of biologically chosen and randomly chosen cross-links cases to address the question about the uniqueness of the biological cross-links. We also give the 2D prediction of the overall organization of the two bacterial chromosomes and validate our predictions with the available experimental data. We start by describing our model and then show our results and conclusions.

2.2   Model

The circular bacterial chromosome of *C. crescentus* and *E. coli* consist of 4017 and 4642 thousands of base-pairs (BPs), respectively. We model these as a flexible ring polymer having 4017 and 4642 monomers. Each monomer bead in our coarse-grained polymer represents

1000 of BPs of DNA, which is much greater than the persistence length of the naked DNA (150 BP); thus, we can model the chromosome as a flexible polymer. The coarse-graining of one monomer in our model matches the resolution of the Hi-C contact map, which is 1000 BPs. The nearest neighboring monomers along the polymer contour interact by the harmonic potential of the form

$$V_s = \frac{1}{2}\kappa(r-a)^2$$

Here, a is the equilibrium bond length, and we chose the value of $a = 1$, which also sets the length scales in our system, κ is the spring constant, which we take as $\kappa = 200k_BT/a^2$ high compared to the thermal energy k_BT. We chose the value of $k_BT = 1$, which is also the unit of energy in our simulation.

We model the excluded volume interaction between the monomers (excluding the adjacent monomers along the chain) using the suitably shifted and truncated Lennard Jones potential, which is also called the Weeks, Chandler, and Anderson potential (WCA) potential.

$$V_{LJ} = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right] + \frac{1}{4}, & \text{if } r \leq 2^{1/6}\sigma \\ 0, & \text{otherwise} \end{cases}$$

The parameter ε is chosen to be $1k_BT$ and we chose the bead diameter $\sigma = 0.2a$. The small value of the bead diameter σ is chosen such that the chain can cross itself. We allow the chain crossing in our system to model the activity of enzyme topoisomerase, which helps the chain to release the topological constraints by suitably cutting and rejoining the chain. In chapter 3 of the thesis, we also vary the parameter σ to study the effect of chain crossing in the organization of the polymer.

As mentioned earlier, we also use the data from the Hi-C contact map of the bacterial chromosomes of *C. crescentus* and *E. coli* as an input to our model of the chromosome. The Hi-C contact maps have been generated by the analysis of the raw experimental data obtained from [8, 21]. The procedure of the analysis is given in Appendix A of the thesis.

We reiterate that the Hi-C contact map gives the frequency of the different DNA segments i, numbered $i = 1, 2..., N$ of 1 kilo-BPs each to be in spatial proximity with the other DNA segments j, numbered $j = 1, 2...., N$. Therefore, we take monomer pairs with monomer index i and j, which are in the spatial proximity in the experimental contact map with very high frequency and cross-link them. The number of monomer pairs that we cross-link depends on the threshold frequency we set in the Hi-C contact map. For two different values of cutoff frequency, we have 49/47 and 153/159 number of cross-links (CLs) for the bacteria *C. crescentus*/*E. coli*. We term these sets of monomer pairs, which we take from the

experimental contact map as biological cross-link (BC-1 and BC-2) sets of CLs throughout our study. Note that the BC-1 CL set is the subset of the BC-2 CL set, i.e., BC-2 CL set contains all the CLs which are present in the BC-1 CL set. We model the cross-linking between specific monomer pair by harmonic spring potential:

$$V_{CL} = \frac{1}{2} \kappa_c (r_{cc} - a)^2$$

Here, r_{cc} is the distance between the specific monomer pair we take from the BC-1 and BC-2 CL sets, and κ_c is the spring constant. We take the value of $\kappa_c = 200 k_B T / a^2$, similar to the spring constant between the adjacent monomers along the polymer contour. We also observe that all the CLs in the biological cross-link sets are not independent viz. suppose monomer index i and j constitute a cross-link and is present in the biological CL-set, and if the monomer index i+1 and j+1 which form a cross-link is also present in the biological CL sets then we should not consider these as independent cross-links since they are effectively one cross-link. To get the effective number of CLs, we set the cutoff distance equal to $5a$ along the length of the contour. In other words, if the monomers constituting two different CLs are at a distance $\leq 5a$ (arbitrarily chosen) from each other along the polymer contour, we consider them as only one independent CL and randomly chose one out of two cross-links for our simulation. Thus, we have 26/27 and 60/82 effective CLs in place of 49/47 and 153/159 number of CLs in the BC-1 and BC-2 CLs sets for the bacteria *C. crescentus/ E. coli*.

We also study the systems in which the positions of the monomers to be cross-linked are chosen randomly, and the number of CLs are the same as the effective number of CLs in the biological CL-sets. The CL sets corresponding to the randomly chosen positions of monomers with 26/27, and 60/82 number of CLs are termed as random cross-link sets RC-1 and RC-2, respectively. We study the organization of the polymer with 10 different RC-1 and RC-2 sets and compare it with the organization of the polymer with BC-1 and BC-2 CL-sets corresponding to the bacteria *C. crescentus* and *E. coli*. To see the list of monomers which are cross-linked, corresponding to the BC-1, RC-1, BC-2, and RC-2 CL sets, refer the tables C.1 and D.1 in the Appendix B of the thesis; where we have only enumerated the monomers which constitute the CLs corresponding to 2 random realizations of the cross-links.

For the study in this chapter, we only study the role of cross-links in the organization of the polymer and do not take the confinement of the cell walls into account. Therefore, we keep the polymer in a huge simulation box of size $200a \times 200a \times 200a$. We chose the huge simulation box to avoid the interaction between the different polymer segments due to the periodic boundary conditions.

To check whether the large length-scale organization of the chain is determined primarily by the presence of CLs, we start our simulations from nine independent initial conditions and generate different conformations of the polymer using Monte-Carlo simulations. Instead of taking random coil initial configurations of the polymer, we design different initial conditions of positions such that the adjacent monomers along the chain contour are at a distance $a = 1$, but the specific monomers which we cross-link are not placed at a distance of $a = 1$ with each other in the 9 different initial conditions. For example, in one of the initial conditions, the monomers of the ring polymer are arranged along a circle of radius $30.73a$ such that one circle has 193 monomers. The circles of monomers are stacked up to form a cylinder. Note that this will lead to monomer numbered one and the last monomer $N = 4017/4642$ to be at a distance much larger than a, but for a ring polymer the first and the last monomer should be connected. Also, the other specific monomers which constitute the CLs can be at distances much larger than a as they are arranged along the cylinder. But these will come closer due to the presence of harmonic spring potentials acting between monomer-pairs as the polymer is allowed to relax during the Monte-Carlo run. For the second and third initial conditions, we arrange the monomers in circles of radius $40.92a$ and $36.94a$, respectively. The monomers are arranged in squares of side $90a$, $80a$, and $70a$ for the fourth, fifth, and sixth initial conditions, and these squares are then stacked up. In the last three initial conditions, we arrange the monomers in equilateral triangles of sides $40a$, $50a$, $60a$, and stack them to form a vertical column.

We reiterate that in the 9 independent initial conditions, the pair of monomers that constitute the CLs are not at a distance of a, but they interact by a spring potential. If we take the initial value of spring constant to be very high, i.e., $k_c = 200k_BT/a^2$, the system will not be able to relax due to the very high energy contribution from the cross-linked monomers. Thus, to obtain the equilibrium conformation of the polymer system, we take a very small value of the spring constant $k_c = 0.2k_BT/a^2$ between the CL monomers at the start of the simulation. We then increase the value of k_c by $0.02k_BT/a^2$ after every 1000 Monte-Carlo steps (MCS) till it reaches a value of $k_c = 200k_BT/a^2$. We also attempt the large trial displacements with $\delta = 1.2\sigma$ in every 100 MCS, which may help to release the topological constraints.

To ensure that the system has relaxed from its initial condition, we monitor the potential energy for the chains with BC-1 and BC-2 sets of CLs corresponding to the model chromosomes of bacteria *E. coli* and *C. crescentus*. The value of the potential energy should be approximately same across different initial condition, if the system has relaxed, Indeed, the value of energy relaxes to the same value at the end of 10^5 MCS from the 9 different runs starting from independent initial conditions, see figure 2.1(a), (b), (c) and (d). It gives us the
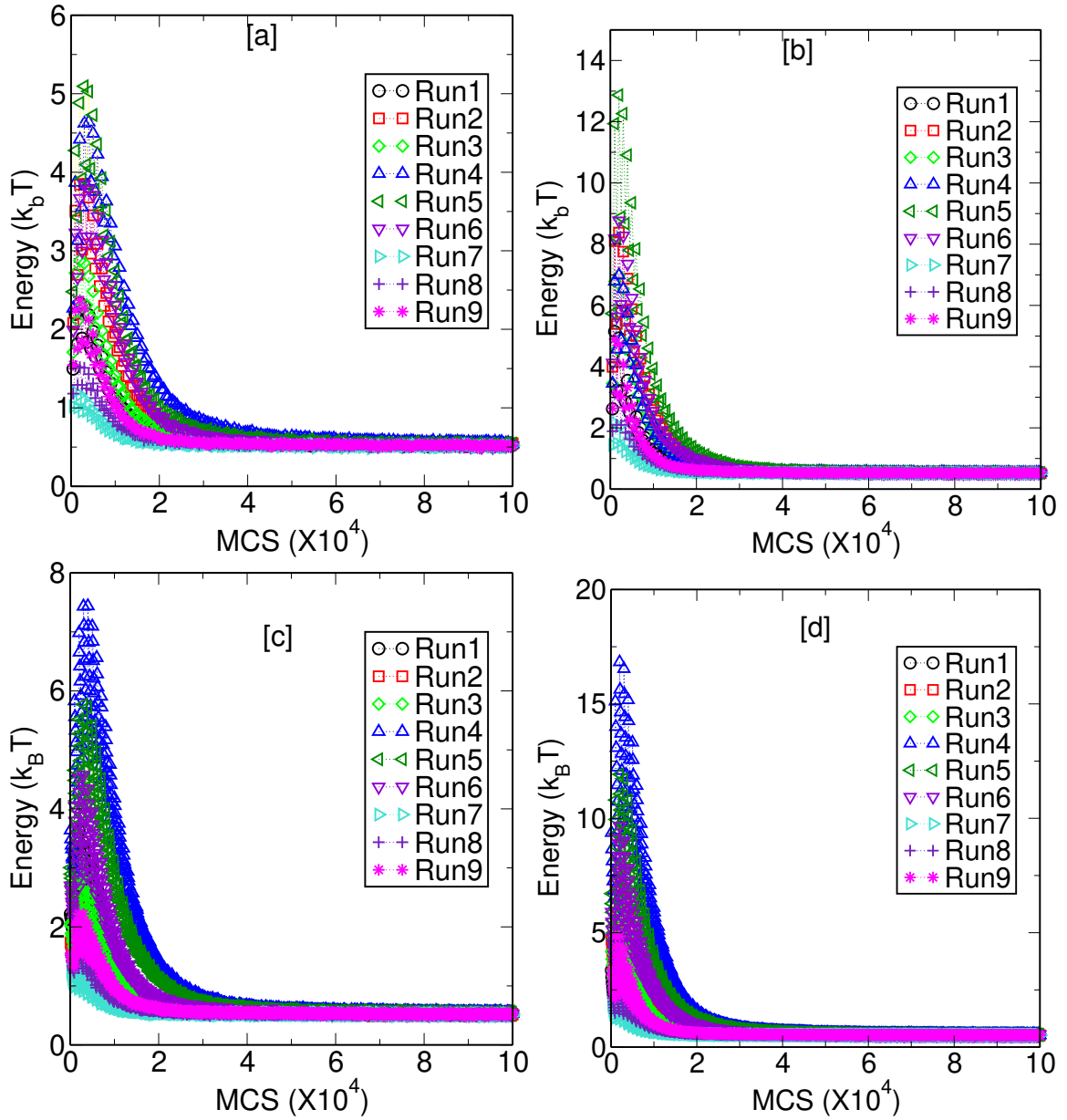
Fig. 2.1 The plot of the energy versus Monte Carlo steps (MCS) as the systems evolve to a relaxed state from the 9 different initial conditions. The energies all converge and fluctuate around the same value establishing that the chains are unlikely to be stuck in some meta-stable states because chains can cross each other due to the relatively small excluded volume of beads considered. Plots (a) and (b) correspond to the polymer with BC-1 and BC-2 sets of CLs taken from the Hi-C contact map of bacteria *E. coli* and the plots (c) and (d) are for the model polymer of bacteria *C. crescentus* with BC-1 and BC-2 sets of CLs.

confidence that the chain conformations are not stuck in meta-stable energy minimum. We then allow the system to evolve for another 1.2×10^7 MCS.

After the initial equilibration steps, we start calculating the average statistical quantities and compare the organization of the polymer with cross-links across 9 independent initial conditions. If we get the same organization of the polymer for the runs starting from independent initial conditions, then we can claim that the cross-links give rise to a non-trivial, unique organization to the polymer because a ring polymer without any cross-links can take any conformations and we do not expect the polymer to organize in a particular structure.

2.3   Results

Now, we discuss the statistical quantities which we use to investigate the organization of the ring polymer. We will check if the statistical quantities from different runs starting from 9 independent initial conditions with the same set of CLs give similar results? This is to infer that the polymer has a similar organization across different runs. We further compare data from 10 different RC-1 and RC-2 CL sets of CLs with data from *E. coli* and *C. crescentus* CL sets BC-1 and BC-2.

The first quantity we estimate is the size and extent of the polymer with CLs. To that end, we calculate the moment of inertia tensor **I** with respect to the center of mass (CM) of the polymer coil and diagonalize the matrix to get its principal moments for each micro-state. We then calculate the average principal moments I_1, I_2, I_3. Where I_1 is the largest eigenvalue, and I_3 is the smallest eigenvalue of the moment of inertia matrix.

In Fig. 2.2(a) we show the values of I_1/I_3 for distinct random CL-sets having the same number of CLs as RC-1 and RC-2 CL-sets corresponding to the bacteria *E. coli*. For each random CL-set, the average is taken over 9 independent initial configurations. In subplot (b), we show I_1/I_3 for Biologically determined CLs: BC-1, BC-2, for nine independent initial conditions. In plot (c) $R_g = \sqrt{(I_1 + I_2 + I_3)/2M}$ for different random CL sets having same number of CLs as RC-1 and RC-2 is shown and in subplot (d) we show R_g for 9 independent initial conditions for BC-1 and BC-2 respectively. Here M is the sum of masses of the individual monomers $M = \sum m_i$, $m_i = 1$ is the mass of each monomer. The value of I_1/I_3 is the ratio of major and minor axes and gives a measure of shape asymmetry of the coil. Comparing the value of I_1/I_3 in 2.2(a), (b) we see that I_1/I_3 has a lower value for all ten RC-2 sets compared to BC-2 set. A plausible explanation for this difference is given in the following paragraphs and will be confirmed by the end of this chapter. Subplots Fig 2.2 (b) and (d) show the values of R_g obtained from randomly determined CLs and Biologically determined CLs. The calculated value of R_g for ring polymer without CLs and the average value is $\approx 12a$. The value of R_g decreases as we increase the number of CLs from BC-1/RC-1 set to BC-2/RC-2 sets; this decrease in the value with an increase in the number of *effective*
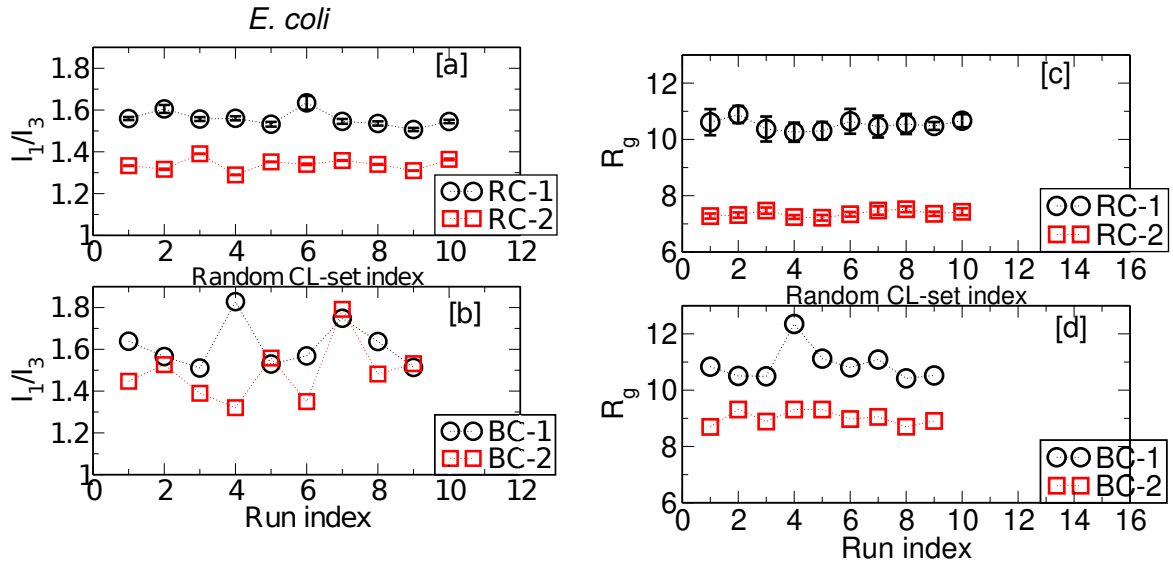
Fig. 2.2 (a) The value of I_1/I_3, the ratio of the largest and lowest eigenvalues of the diagonalized moment of inertia matrix for distinct random CL sets (no. of CLs for each RC-set is equal to that for RC-1 and RC-2, for bacteria *E. coli*). For each random CL-set index, the plot shows the average and s.d. of I_1/I_3 over 9 independent initial conditions; the s.d. is smaller than the size of the symbol. (b) The value of I_1/I_3 versus run index for runs starting from 9 independent initial conditions for biologically determined CLs: BC-1 and BC-2 for the bacteria *E. coli*. (c) The plot of Radius of gyration $R_g = \sqrt{(I_1 + I_2 + I_3)/2M)}$ versus the random CL-set index. (d) The plot of R_g and run index from 9 independent initial conditions for biologically determined CLs: BC-1 and BC-2 for *E. coli* bacteria.

constraints is expected. But interestingly, the change in the value of R_g as we go from BC-1 to BC-2 is distinctly less than the decrease in R_g as we go from RC-1 to RC-2. We interpret the difference between the two cases as follows: the *effective* CLs in BC-1 are already at critical positions along the contour, which give partial organization to the DNA polymer. On increasing the number of CLs further (i.e., BC-2 CLs), the organization of the polymer improves along with the already established framework. On the other hand, an increase in the number of random CLs leads to overall shrinkage in the size of the coil and not necessarily to accentuate a preferred set of conformations. The lower values of I_1/I_3 for all 10 sets of RC-2 compared to that in BC-2 also point towards such an understanding. This idea will get further substantiated in the rest of the chapter. The similar conclusions can be drawn for the biological and random CL-sets for the bacteria *C. crescentus* from the figures in Fig.2.3.

Next to get the idea of how the monomers of the polymer are distributed in space and if there is any difference in the radial arrangement of the bio-CLs and random CLs, we investigate the radial distribution of monomer number densities and the normalized CL
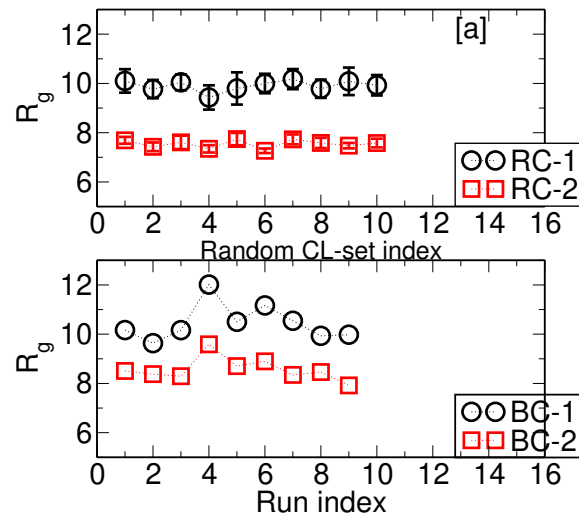
Fig. 2.3 (a) The figure shows the value of the radius of gyration (R_g) v/s random CL-set index. The random CL set index represents ten different realizations of randomly chosen CLs, and in each RC-set, the number of CLs is equal to the RC-1 and RC-2 sets for bacteria *C. crescentus*. The error bar shows the standard deviation across 9 independent initial conditions. (b) The figure shows the radius of gyration for biological cross-links set versus the run index. The run index goes from 1 to 9 for the runs starting from 9 independent initial conditions.

Fig. 2.4 The monomer number density $n_M(r)$, is plotted versus r, where r is the distance of the position of the monomers from the center of mass of the DNA-polymer coil for the model bacterial chromosome of *E. coli*. Plot (a) is for the polymer with BC-1, RC-1 CL sets, whereas plot (b) is for the polymer with BC-2,RC-2 CLs. The value of $n_M(r)$ is averaged over 9 independent initial conditions and error the bar shows the s.d. from the average.

number density with the distance r from the center of mass (CM) of the polymer coil. The quantities $n_M(r)$ and $n_{CL}(r)$ are calculated by calculating the average number of monomers and CLs in radial shells of width $2a$ from the CM of the coil, divided by the volume of each shell. The CL-density is further normalized by the total number of CLs for a particular case under consideration to obtain $n^*_{CL}(r)$. Data for $n_M(r)$ and $n^*_{CL}(r)$ from 9 independent runs are plotted for each of set of CLs: BC-1, BC-2, and one set of RC-1, RC-2 in Figs.2.4 and 2.5, respectively corresponding to the bacteria *E. coli*. Small standard deviation from average for monomer number densities and the normalized CL number density is an indication that the arrangement of monomers and CLs have relaxed to similar distributions, and is independent of starting configurations of monomers.

Comparing subplots (a),(b) of Figs. 2.4 and 2.5 for BC-1 and BC-2 establish that coils with a higher number of CLs lead to more number density of monomers and CLs at the center of the coil. As the coil gets into a more compact coil structure with an increased number of CLs in set RC-2 as compared to RC-1, we again see an increase in the number density of monomers, CLs (suitably normalized) at the center. Comparing data for BC-1 and RC-1, respectively, from Figs. 2.4 and 2.5, we observe that the distribution of monomers and CLs are similar at different r for the 2 cases. In contrast, the normalized density of CLs at the central region is more for BC-2 compared to that for RC-2. Moreover, monomer density $n_M(r)$ for BC-2 is lower than that for RC-2 at the center of the coil, whereas there are more monomers present at the periphery (for $r > 12s$) for BC-2 as compared to RC-2. Since the number of monomers in each shell is divided by the volume of the shell, the difference in the number of monomers at the periphery is just discernable from the number density $n_M(r)$ plots. Lastly, the number density of monomers/CLs drops down significantly beyond a distance of $8a$ from the coil's center. The other nine sets of randomly chosen CLs also is in tune with the above observations (data not shown).

To gain some more insight into the global structural organization of the DNA-coil, the simplest question to ask is whether a particular CL is always found near the center of the coil or near the periphery of the coil. To this end, we compute the probability of each of the CLs to be found in the *inner, middle* and *outer* regions of the DNA-coil. We use the cutoff radii $R_{inner} = 5a$, $R_{middle} = 9a$ (chosen from the knowledge of the value of $R_g \approx 8a$) and calculate the probability $P_{inner}, P_{middle}, P_{outer}$ of finding the $i-$th CL within distance $r < R_{inner}$ (inner region), $R_{inner} < r < R_{middle}$ (middle region) and $r > R_{middle}$ (outer region), respectively, from the coil's center of mass. If the values of $P_{inner}, P_{middle}, P_{outer}$ for each CL has small deviation from the average value in each of the 9 independent runs, it would indicate that the presence of CLs leads to a similar organization of the DNA across independent runs. Also,

we compare the probability distribution of CLs for runs with bio-CLs and random-CLs to investigate if bio-CLs lead to organization distinct from that obtained with random-CLs.
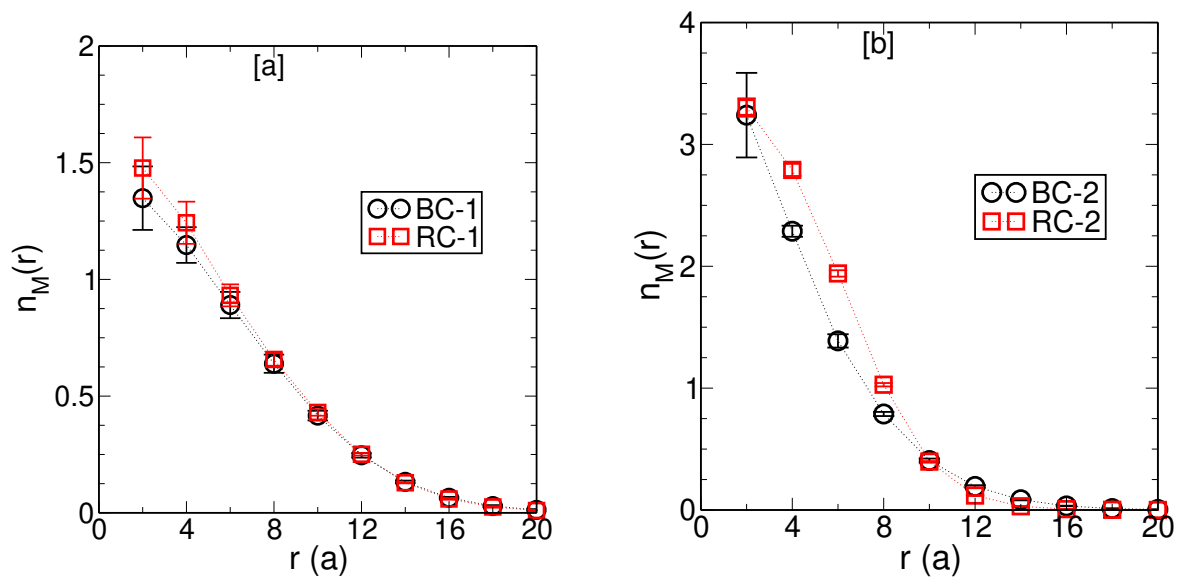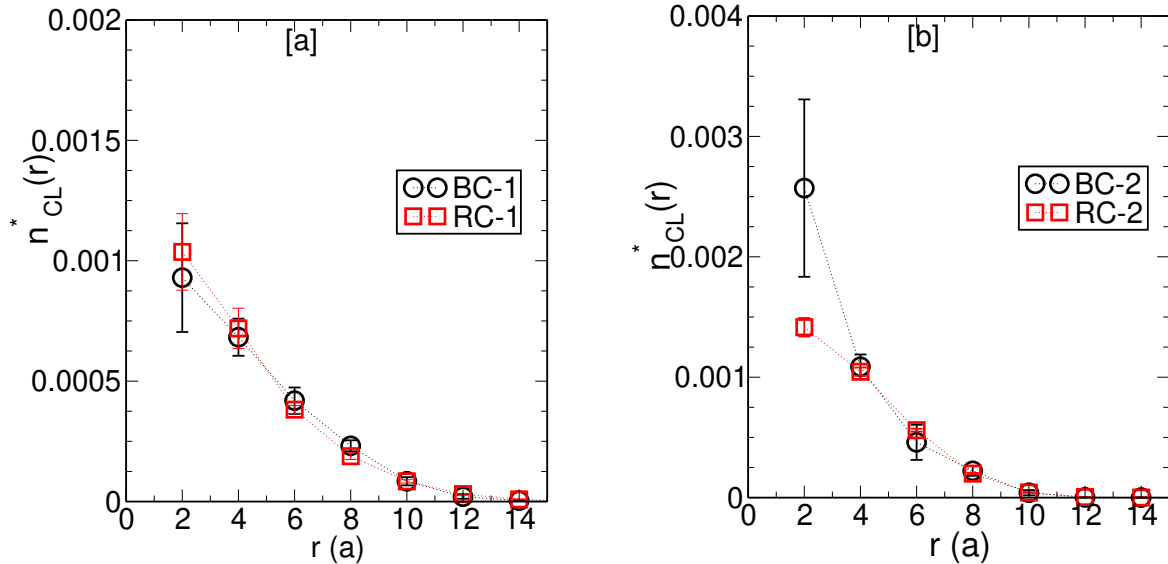

Fig. 2.5 The normalized CL number density $n_{CL}^*(r)$ is plotted versus r, where r is the distance of the position of the CLs from the center of mass of the DNA-polymer coil. Subplot (a) is for the polymer with BC-1 and RC-1 CLs, whereas subplot (b) is for the polymer with BC-2 and RC-2 sets of CLs. Further $n_{CL}^*(r)$ is averaged over nine independent runs starting from different initial conditions, and error bars show the standard deviation from the average value. The number of CLs in both graphs are equal to the CL sets of the bacteria *E. coli*.

We carry out the same exercise for different segments of the polymer chain. The *E. coli* chain with 4642 monomers is divided into 80 segments with 58 monomers in each segment, and the segments are labeled from $i = 1, 2 ... N_s$ as we move along the contour. We can then calculate the location of the CMs of each segment and find out the probability of finding the CMs in the central, middle, and outer regions. The segments in a random-walk polymer model (without CLs) can take any conformation, and there is no reason to believe that certain segments will preferably be found in the inner or outer regions of the coil. If the segments were completely delocalized, we would expect the polymer in different microstates to contribute to all the $P_{inner}, P_{middle}, P_{outer}$ quantities for each segment. The question is to what extent will this basic behavior of polymer coils get modified by the presence of bio-CLs and random-CLs?

Fig. 2.6 Subplots (a), (b) and (c) (in the left column) shows the probabilities of the center of mass (CM) of 80 DNA-polymer segments to be found in the inner, middle and outer region of *E. Coli* DNA coil with BC-2 set of CLs. The x-axis is segment index. In each case average, values of P_{inner}, P_{middle} and P_{outer} are taken over nine runs starting from independent initial conditions, deviation from the average is shown as error bars. Small error bars indicate that the probability of finding CM of segment i in a particular region is nearly the same across different runs. Data on subplots (d), (e), (f) are for the random choice of cross-link position (set RC-2) with 82 CLs in a chain with 4642 monomers. Each segment has 58 monomers.
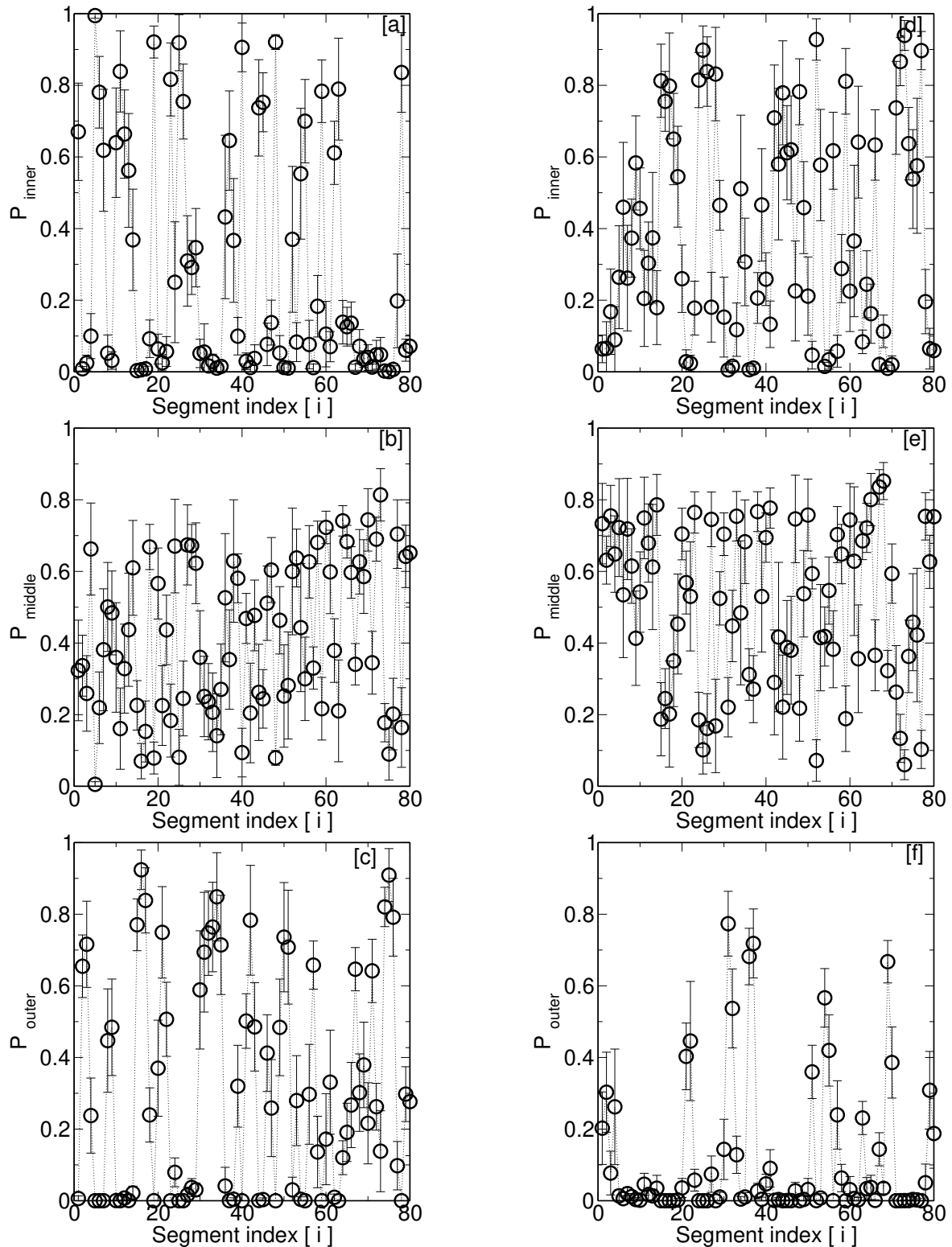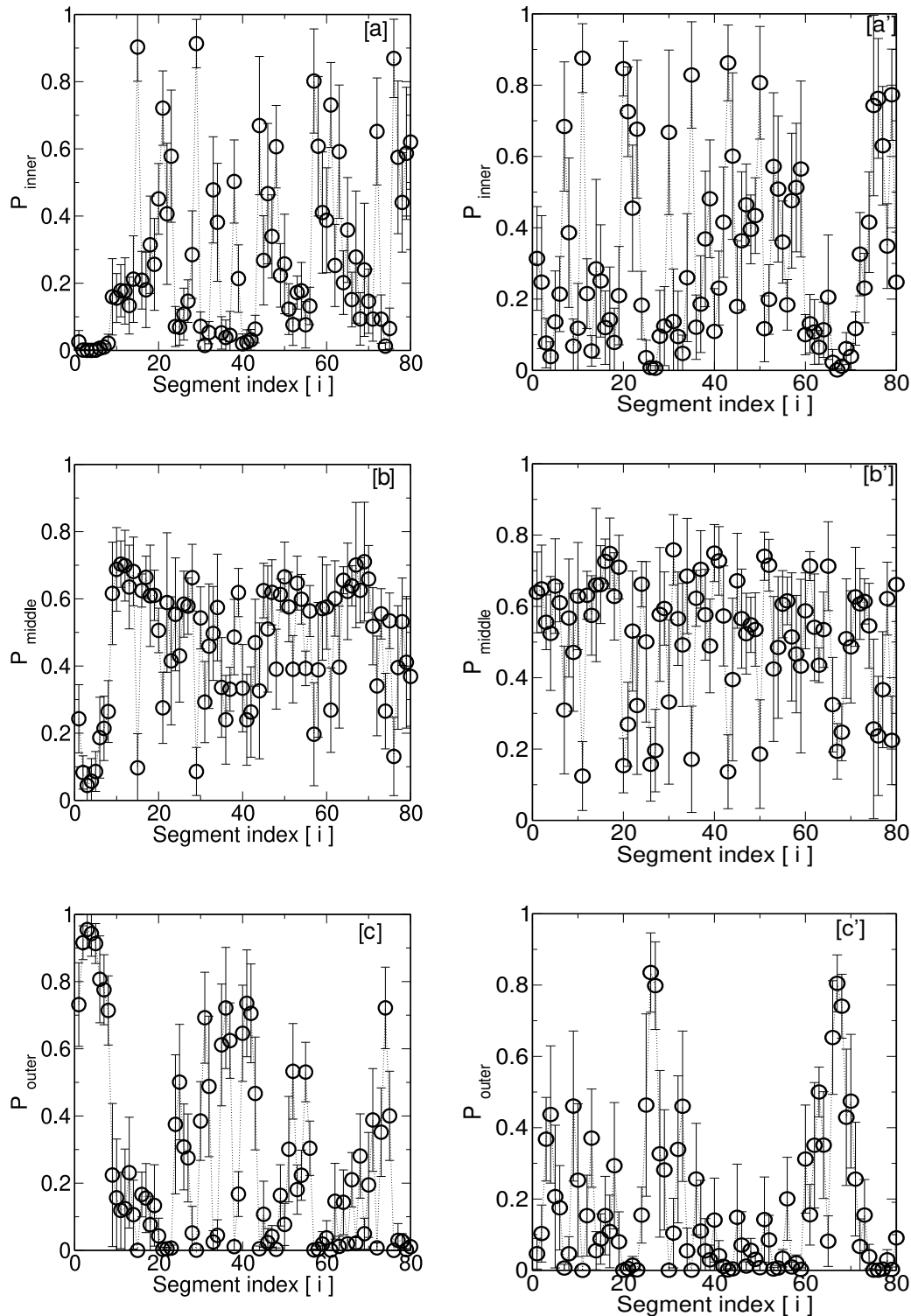
Fig. 2.7 Subplots (a), (b) and (c) (in the left column) shows the probabilities of the center of mass (CM) of 80 DNA-polymer segments to be found in the inner, middle and outer region of *C. crescentus* DNA coil with BC-2 set of CLs. The x-axis is segment index. In each case average, values of P_{inner}, P_{middle} and P_{outer} are taken over nine runs starting from independent initial conditions, deviation from the average is shown as error bars. Small error bars indicate that the probability of finding CM of segment i in a particular region is nearly the same across different runs. Data on subplots (a'), (b'), (c') are for the random choice of cross-link position (set RC-2) with 60 CLs in a chain with 4642 monomers. Each segment has 50 monomers.

Probability data about the location of CLs and segments for BC-2 and RC-2 is given in B.2 and 2.6, respectively. Data for BC-1 and RC-1 is shown in B.1 and B.3. Furthermore, from Figs B.2 and 2.6(a),(b) and (c) we see that some CLs (e.g., the CL with index 60) has the nearly equal probability of being in the inner or middle region of the coil, but very low probability to be found in the outer region. For BC-2, most CLs are found in the inner and middle regions of the coil, whereas for RC-2 CL set, there are some CLs at the periphery; refer Fig.B.2. On the other hand, from Fig.2.6 we see a larger number of segments have a finite probability of being in the outer regions for BC-2 as compared to data for RC-2. The data consistently show that the position of CL, as well as segments, are localized in space across different runs. The probabilities of the 80 segments to be found in the inner, middle, and outer region for the DNA coil of bacteria *C. crescentus* with BC-2 CL-set is shown in the Fig.2.7 (a), (b), (c), respectively. The graphs (a'), (b'), (c') in Fig.2.7 represent the probability of finding the 80 segments in the inner, middle, and outer region of the DNA coil with RC-2 CL set corresponding to the bacteria *C. crescentus*. The polymer coil is divided into 80 segments in a 4017 monomer chain. Thus each segment consists of 50 monomers.

Having established that the CLs and segments of the DNA-polymer coil have some degree of radial organization, we try to extract more detailed structural information about the position of segments relative to each other within the coil. We calculate the probability of each CL (alternatively, each segment) to be in proximity to other CLs (alternatively, other segments). If there are no particular well-defined relative positions of CLs/segments within the chain-coil, there is no reason to expect CM of some segments (or independent CLs) to be found spatially close to each other with high probability, especially when the segments/CLs are separated along the chain contour. We define two CLs/segment's CM to be spatially close to each other if the distance r between the CLs/segment-CMs is $< 5a$, which is just more than $0.5R_g$. We emphasize that we have cross-linked monomers, these constraints are at the monomer (1000 BP) length scale, whereas we are investigating the organization of polymer segments at a much larger length-scale. The position of CLs, position of CM of segments are just 2 different markers of different sections of the chain and we use relative position of both to identify spatial correlations between different sections of the chain.

In Fig. 2.8 we present the colormaps showing the average probability $p(i,j)$ of finding each pair CLs i,j at distances of $< 5a$ for BC-2, RC-2 for four independent runs starting from different initial conditions. As the Monte Carlo simulation evolves, at each microstate if the distance d between a pair of CLs is such that $d < 5a$, a counter $c(i,j)$ for pair i,j is incremented. The probability $p(i,j)$ at the end of the MC-run is the value of $p(i,j) = c(i,j)/N_{micro}$, where N_{micro} is the number of microstates over which data is calculated for calculation. The x-axis and the y-axis represent CL indices i,j, and the colored pixel indicates
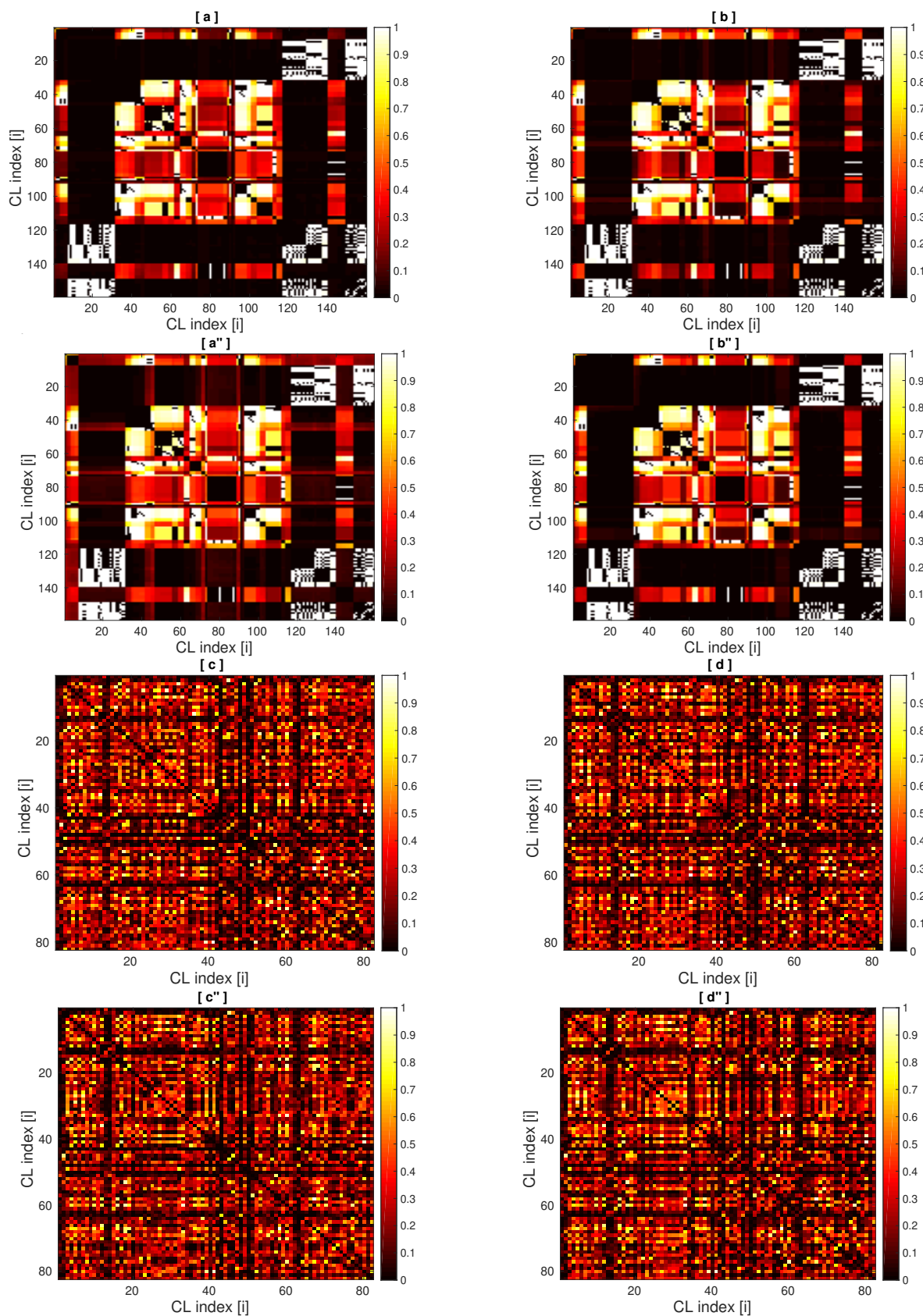
Fig. 2.8 Colormaps to represent probability $p(i, j)$ to find CLs i spatially close CLs j. The top 4 figures are data from 4 independent runs with BC-2 with 159 CLs ($i, j = 1...159$) and the bottom two subplots are from 4 independent runs with RC-2 with $i, j = 1, ...82$.

the value of $p(i, j)$. The top four colormaps of Fig.2.8) represent data obtained for BC-2, and the bottom four colormaps Fig.2.8 show corresponding data from two independent runs with RC-2 set of CLs. A pair of CLs that are near each other along the contour of the chain will have the distance $d < 5a$ between them by default and will show up as high probabilities in the colormap. We set these $p(i, j) = 0$ in the calculation if the monomers constituting a pair of nearby CLs are separated by less than 6 monomers along the contour. That is the reason that the diagonal $p(i, i)$ is dark in all colormaps. We do this because we want to see only non-trivial correlations between different CLs. Following Fig. 2.9 and 2.11, the positional correlation colormaps show the probability of finding a pair of segment-CMs within a distance of $5a$ for BC-1/RC-1 and BC-2/RC-2, respectively, for the bacterial DNA polymer of *E. coli*. Note that these probability colormaps give much more detailed information than a pair correlation function $g(r)$, which would just give the average distance between CLs or segment-CMs.

Thus we find further evidence of our hypothesis that the set of CLs decides the large scale structure of the polymer.

Secondly, the number of the bright pixels are much more in colormaps obtained using CL sets BC-2 and RC-2 (Figs.2.11) as compared to colormaps for BC-1 (Fig. 2.9). It is not surprising as more constraints due to the presence of a higher number of CLs lead to a relatively more compact well-defined structure and a large number of CLs (or segments) near one another. With the few bright patches for BC-1, RC-1 CL set with 27 *effective* CLs, one cannot clearly define the mesoscale conformation of the whole chain, though there are indications of the emergence of structure. However, a set of 82 effective CLs for BC-2, RC-2 might be enough to deduce and define the large-scale organization of DNA-polymer as we now know which segments are neighbors of a particular segment.

Thirdly, comparison of colormaps for BC-2 and RC-2, especially in Figs. 2.11 shows a different nature of the organization of the DNA polymer. For BC-2 CL-set, adjacent segments show a higher propensity to be together, which can be deduced by observing that there are clusters of adjacent bright pixels. Comparatively, bright pixels are scattered more randomly in the colormaps for RC-2. From the colormaps, we can clearly, observe that there is a difference in the nature of patterns for BC-2 and RC-2. Similar conclusions can be drawn from the positional correlation colormaps of Fig.2.10 and 2.12 for the bacteria *C. crescentus* with BC-1/RC-1 and BC-2/RC-2 CL-sets, respectively.

Fourthly and importantly, the reasons for the formation of clusters of bright pixels seen in the top two colormaps of Figs. 2.8 (for CLs) is not the same as that of Fig.2.11 (for segment-CMs). To understand the bright patches of Fig. 2.8, we remind the reader that the CLs are often found adjacent to each other along the chain contour for BC-1 and BC-2.

Fig. 2.9 Colormaps to represent probability $p(i, j)$ to find CM of segment i spatially close to CMs of other chain segments j. There are 80 segments in the *E. coli* polymer with 58 monomers per segment. The top 4 figures are runs with BC-1 and the bottom four for RC-1.
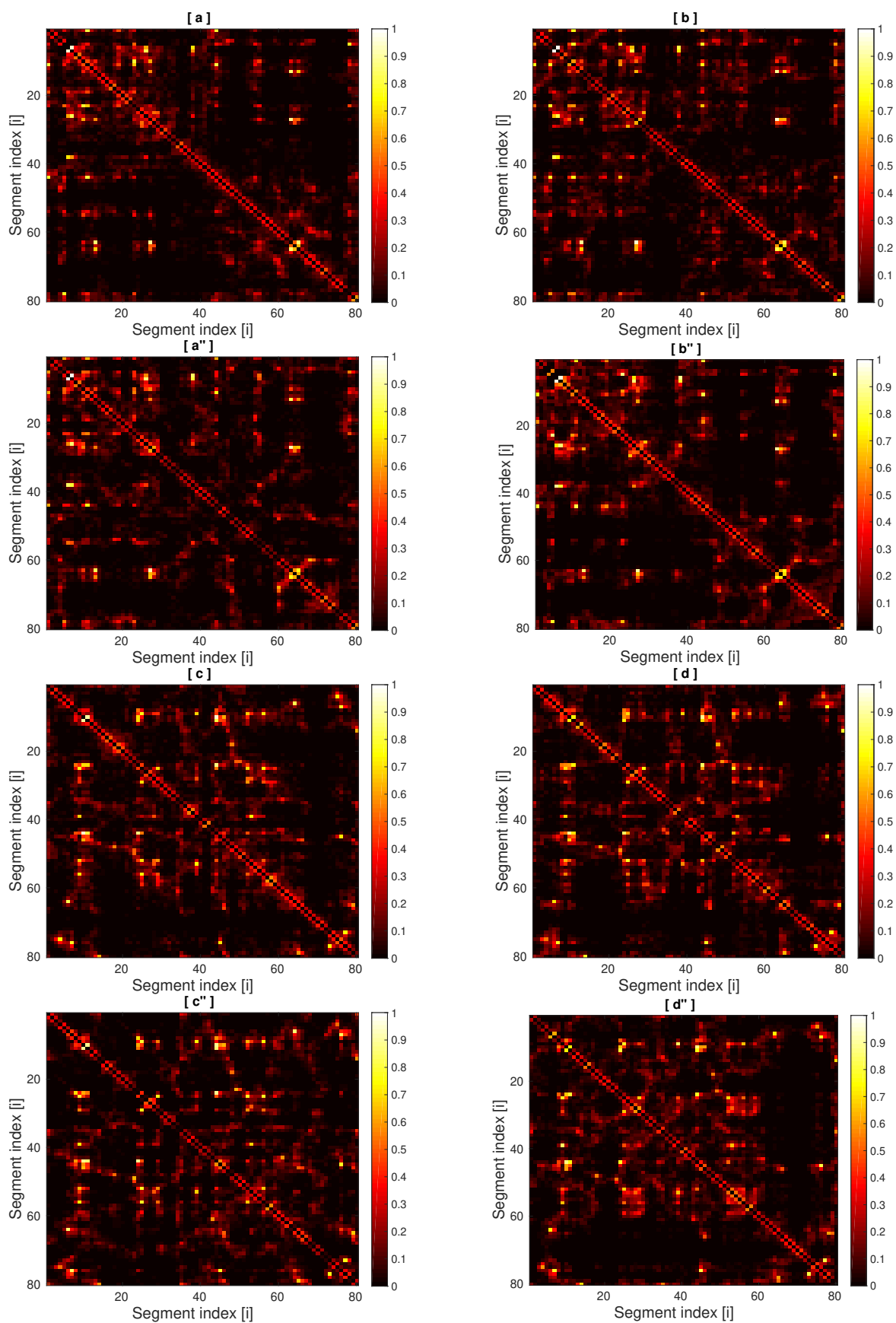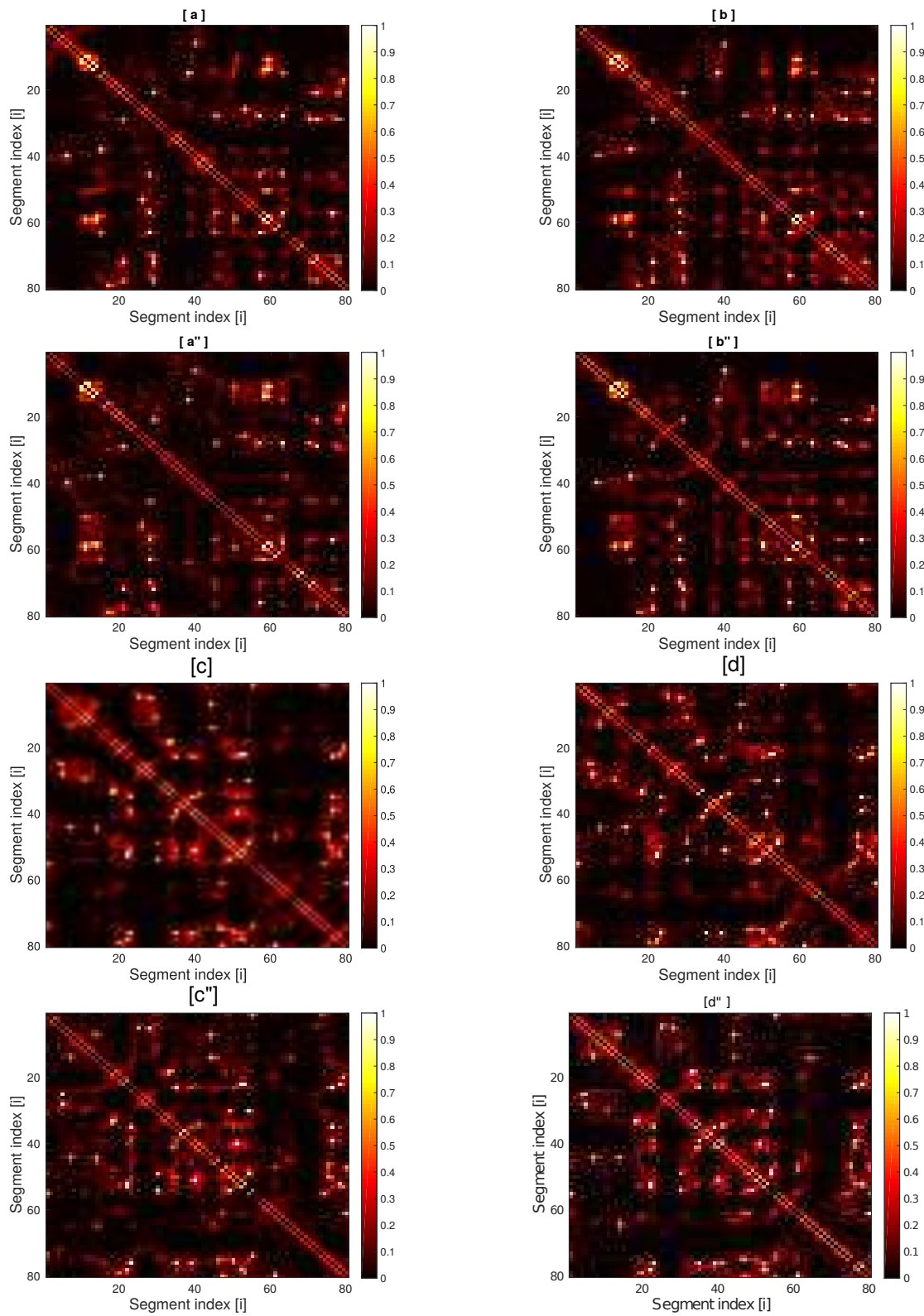
Fig. 2.10 Colormaps to represent probability $p(i, j)$ to find CM of segment i spatially close to CMs of other chain segments j. There are 80 segments in the *C. crescentus* DNA polymer with 50 monomers per segment. The top 4 figures are runs with BC-1 and the bottom four for RC-1.

Suppose, CL-i, CL-j and CL-k are next to each other along the chain. Note that then $p(i,j)$, $p(i,k)$, $p(j,k)$ has been explicitly put to zero. But if CL-m, which is far from i and j along the contour, comes within a distance of $5a$ from CL-i, then CL-m is also automatically close to CL-j,k and three adjacent pixels will appear in the colormap, viz., $p(i,m)$, $p(j,m)$, $p(k,m)$. Thus, the bigger bright patches for BC-2 in Fig. 2.8 should not be interpreted as evidence for a more organized polymer. A similar arrangement of bright/dark pixels across runs is just evidence of similar organizations across different runs. We do not get big bright patches in the colormaps for RC-1 and RC-2 in Fig. 2.8, because the calculation are done using only the *effective* number of CLs. Thereby there are no redundant CLs in the RC-1, RC-2 set of CLs. We have also checked that the organization is maintained for BC-2 if we To quantify the differences in the colormaps of BC-2 and RC-2 in Fig. 2.11 and 2.12, we calculate the number of segments, $n_{seg}(i)$, which are near (i.e. within distance $d < 5a$) to the CM of the i-th segment with probability $p(i,j) > 0.05$. That is we count the number of non-black pixels in the colormaps of fig.2.11 and 2.12) for a particular segment with index i. Then we divide $n_{seg}(i)$ by the total number of segments to get $f(i)$ to get an estimate of the fraction of a total number of segments that approach segment i with any finite probability. It is shown in the Fig.2.13 and 2.14 for RC-2 and BC-2 correseponding to the model chromosome of *E. coli* and *C. crescentus*. A cutoff of 0.05 for the value of $p(i,j)$ is appropriate as anyways most of the colormap is black and deep red going upto yellow for very few pixels. From the figure, we observe that the value of $f(i)$ is relatively high for RC-2 set of CLs as compared to $f(i)$ for bio BC-2, this suggests for random CL-set many more segments can approach a particular segment for RC-2 compared to that for BC-2. We interpret this as a more spatially organized structure with BC-2 cross-links, as it has fewer but well-defined neighbors, as can also be checked from the colormap of Fig.2.11 and 2.12. As an example, segments with indices 70-78 for BC-2 are only close to their adjacent segments (a bright diagonal patch in the colormap) giving relatively very low value of $f(i)$ in Fig.2.13(a).

We have also obtained colormaps for the 10 different sets of random CLs, and for each CL-set we can calculate $f(i)$ for each segment index i. Moreover, we can calculate f_{av}, that is the average value of $f(i)$ summed over all the segment indices, i.e. $f_{av} = (\sum_i f(i))/80$. Furthermore, we can calculate the mean of f_{av} over 9 independent runs for each CL set, and thereby obtain $\langle f \rangle$. In Fig.2.13(b) and 2.14(b), we plot $\langle f \rangle$ versus the random CL-set index each set has the same number of CLs as in RC-2. We compare this data with the $\langle f \rangle$ for the one set of biologically obtained CLs: BC-2. We clearly see that for each random CL-sets, the quantity $\langle f \rangle$ has a relatively higher value than $\langle f \rangle$ for BC-2. Observing the differences in colormaps for BC-2 and RC-2 for both the bacteria *E. coli* and *C. crescentus*, we claim that the positions of CLs along the chain for DNA are not completely random. An equivalent

number of CLs in random positions also give an organized structure in that the colormaps from 9 independent runs look similar, but the nature of the organization is very different from the case where biological positions of CLs are chosen.

To extract further insight into the structural organization of the DNA-polymer, we would next probe whether the segments are at geometrically fixed positions with respect to each other, of course accounting for thermal fluctuations. Thereby, we next calculate the angular correlations between CLs and equivalently between the segment's CMs.

To that end, we calculate the dot product of the radial vectors from the CM of the polymer coil to the respective positions of a pair of CLs (i, j) and check if the value of $\cos(\theta_{ij}) > 0$ or < 0, where θ_{ij} is the angle between the two vectors. If the value of $\cos(\theta_{ij}) > 0$, we can say that the two CLs are on the same side/hemisphere of the coil, and increment counter $c^{opp}(i, j)$ by 1. If $\cos(\theta_{ij}) < 0$ we decrement $c^{opp}(i, j)$ by 1. For all possible pairs of CLs, we calculate the average value of $\langle c^{opp}(i, j) \rangle$ suitably normalized by the number of snapshots used to calculate the average. The value of $\langle c^{opp}(i, j) \rangle \approx -1$ would indicate that the pair of CLs i, j are always on two opposite hemispheres. A value of $\langle c^{opp}(i, j) \rangle \approx 1$ means that the two CLs remain in the same hemisphere. We should not interpret $\langle c^{opp}(i, j) \rangle \approx 0$ as we cannot claim that the average angle between the radial vectors is nearly a right angle. The reason is that if the CLs are closer to the center of the DNA-coil, small positional displacements could cause the quantity $c^{opp}(i, j)$ to fluctuate between 1 and -1 and cause $\langle c^{opp}(i, j) \rangle$ to average out to zero. The $\langle c^{opp}(i, j) \rangle$ data for all pairs of CLs are given in Fig.B.4 for BC-2/RC-2 respectively, the corresponding data for relative angular positions for the segment's CMs are given in Figs.B.5 & 2.15 for BC-1/RC-1 and BC-2/RC-2. As before, the top four colormaps in are from two independent initial conditions with BC-1/BC-2 and the bottom four colormaps are for four independent runs with RC-1/RC-2.

In the colormaps of Figs. B.4, B.5 and 2.15 we see there are patches of bright and dark pixels, the size of patches are larger for BC-2 compared to RC-2. As mentioned before, if $\langle c^{opp}(i, j) \rangle = 0$ which is represented by orange/deep yellow color in the colormap we cannot predict the angular positions of the CLs/segment's CMs because of the reason explained above. We can clearly see that the colormaps from independent runs starting from different initial conditions look similar.

In figure B.5, comparing segment-CM colormaps in (a),(b) (for BC-1) with (c),(d) (for RC-1) we do not find any difference in the nature of the distribution of patches. But as the number of CLs increase as we go from BC-1 to BC-2 and RC-1 to RC-2 in Fig.2.15, we find differences in the pattern of colormaps on comparing (a),(b) with (c),(d) corresponding to BC-2 and RC-2 CL sets, respectively. In contrast, for color maps (a),(b) of Fig.2.15 we observe large patches of bright pixels as compared to the patches in (c),(d). Large patches of

bright/dark pixels for BC-2 suggest adjacent segments along the chain contour are on the same/opposite hemispheres with respect to the CM of the coil. The small patches of bright and dark pixels in (c),(d) for RC-2 suggest more of random distribution of different segments. The polymer is organized in both BC-2 and RC-2 CL sets as colormaps from independent runs look similar, but the nature of the organization is different. The reasons for large bright patches in the colormaps for CL-angular positions as shown in Fig. B.4 is not the same as for the colormaps in Fig.2.15. The reasons for the difference has been explained previously for positional correlation colormaps.

The same conclusions also remain valid for the angular correlations of different segments for the model chromosome of bacteria *C. crescentus*. The corresponding colormaps for DNA polymer of bacteria *C. crescentus* is given in the figures B.6, 2.16. The colormaps in Fig.B.6 and 2.16 represent the angular correlations of different segments for BC-1/RC-1 and BC-2/RC-2 CL-sets, respectively.

Finally, we show a representative snapshots of the DNA-polymer of *C. crescentus* and *E. coli* in Fig.2.17 (a) and (a'). The polymer is colored from blue to red as we go from monomer index 1 to 4642/4017 along the contour. This snapshot confirms what we have deduced from the previous figures of positional and angular correlations. Large sections of the chain are localized together in space. The snapshot confirms the kind of conformations expected from the colormaps of angular correlation shown in Fig.2.15, 2.16 (a) and (b). For example, in snapshot 2.17(a) the section marked Region-1 representing monomers around 1750 (segment index 30) is diametrically opposite Region-3 with monomer index 2990 (segment index 50). In Fig.2.15 (a) we see the pixel corresponding to segment indices (30,50) are black. The Region-2 represents monomer numbered around 4100, segment index 71. We can see the pixels corresponding to segment indices (30,71) are yellow, whereas pixels for (50,71) is white. The bottom figure shows the CL distribution in space: only one of the monomers out of the pair, which constitutes a CL, has been plotted.

It is interesting to observe in Fig.2.17 (b) and (b') that the CLs are clumped together in space in about four aggregates. We believe that this helps in the mesoscale organization of the chain as multiple segments of the chain are pulled towards the coil's center with multiple loops on the periphery of the coil. The peripheral loops can lead to relatively large fluctuations in the values of I_1/I_3, as seen in Fig.2.2. This is further validated by the Fig.2.6, 2.7 (c), where we see a large number of segments are to be found in the outer region with significant probabilities. Thus BC-2 set of cross-links leads to the reorganization of the CLs in space such that they form clusters in space with the possibility of polymer loops emanating from the CL-clusters in a rosette-like structure. We interpret that loops from a particular CL cluster would-be neighbors of specific other polymer segments due to the nature of the

arrangement, as opposed to spatial proximity to many segments as seen for RC-2 in Fig.2.13, 2.14 while comparing colormaps for BC-2 and RC-2.

We also calculate the distribution of length of segments between two adjacent CLs in the bio and eight representative random sets. It is given in the Fig.2.18. The distribution of lengths is a fixed quantity once one chooses a particular CL set as an input to the simulation. The number of segments of length l between two adjacent CLs along the chain contour has been shown on x-axis (using a bin size $d\ell$ of 10 monomers) and the frequency density function (FDF) $P(l)$ is plotted on the y-axis, where $P(l) = N_S(l)/(N_S * d\ell)$. We denote the number of segments of length ℓ by $N_S(l)$, and N_S is the total number of segments between CLs. Thus as an example for *E. coli* CL set, $N_S = 82 * 2 - 1$ for RC-2 as each CL is constituted of a pair of monomers, and l essentially is a count of the number of monomers between two CL-monomers along the contour. For the bio-CL, one particular monomer is attached with many other monomers (see supplementary material Table-1); hence, there is a peak at segment length value 0 to 10. We also observe in the randomly chosen CL sets after segment length ≈ 150 FDF is almost zero while in the biologically obtained CL set, there are a few segments for segment length > 250. This shows in the biological CL set there are several long segments that can form bigger loops as compared to CLs chosen in a random manner.

Finally, using the aggregate of all the structural quantities presented in this chapter, we are able to piece together the large scale organization of DNA-polymers in a 2D map for both *C. crescentus* and *E. Coli*(Fig.2.19). To construct the 2-D map, we assumed a spherical globular structure of the DNA-coils, and then we use a 3-step procedure in conjunction. Firstly, we know which segments are to be found primarily in the inner, middle, and peripheral regions of a globule from Figs.2.6 and 2.7. Secondly and thirdly, we use the information about which segments are to be found on the radially opposite hemispheres from Figs.2.15 and 2.16, in tandem with the data about positional correlation given in 2.11 and 2.12. We use this collective quantitative information to distribute the segments within the schematic diagram of the globule.

In lieu of experimental confirmation of our predictions, we use the information content of experimental contact maps to validate our prediction of DNA-organization. To compare the positional colormaps from our simulations to experimental contact maps, we condense contact map data into an 80×80 matrix by suitable coarse-graining (averaging over neighboring bins) of the data of experimental contact maps. We then compare the coarse-grained contact map (frequency of larger DNA-sections) with the highest probabilities (> 0.5) of the simulation color-maps, refer Fig. 2.20: these show a good match. This is by no means an obvious match: we take very few (but significant) points from the experimental contact maps as cross-linked monomers in our simulations and are able to predict from our simulations the positions of

highest contact frequencies in the coarse-grained experimental contact map. The simulation and the experimental contact maps for the bacteria *E. coli* match well with each other, as can be seen in Fig.2.20(b), (d). The difference between Figs. 2.20*a* and *c* for *C. crescentus* is the absence of clear high-probability diagonals in simulation colormap. The plausible reason for this difference could be the presence of plectonemes, as proposed in [8]. The effect of plectonemes are not accounted for in our model as we have taken only 153 CLs (only 60 effective CLs). But it is enough to match our simulation positional correlation color maps with experimental data even for segments far separated along the contour. Furthermore, the cutoff distance $R^C = 5a$ we choose while generating colormaps is comparable to the $R_g \approx 4a$ of 50 segments. Thereby, only a few segment CMs can occupy positions within distance R^C.

Furthermore, the list of 47 most expressed genes of *E. coli* out of ≈ 4000 genes in [140]. Moreover, the position of the gene along the chain contour is also known [141]. We find that 37 out of the 47 most expressed genes are always found in the segments which remain in the peripheral region of the coil as per our simulations. The list of these 47 genes, their locations along the DNA-contour as well as their corresponding monomer and segment in appendixD.1. This is an indirect validation of our results, as biologically, it is expected that more transcriptionally active genes are found in the peripheral regions for easy access by transcription factors [142]. We did not find the list of most expressed genes for *C.crescentus*.

We can also identify the number of genes in a particular segment of 58 kilo-BP and 50 kilo-BP from [141] for *E. coli* and *C. crescentus*. In Fig.2.19 the segments are colored according to the number of genes in that segment; dark color represents more genes and bright color represents fewer genes. For *E. coli* the most gene dense regions are found in the peripheral regions in simulation as well as on the same side of the globule, but for *C. crescentus* the gene dense regions are spread over the extent of the polymer. This seems to be an interesting observation whose physiological relevance might be identified in the future.

Fig. 2.11 Colormaps to represent probability $p(i, j)$ to find CM of segment i spatially close to CMs of other chain segments j. There are 80 segments in the *E. Coli* polymer with 58 monomers per segment. The top 4 figures are runs with BC-2 and the bottom four subplots RC-2.
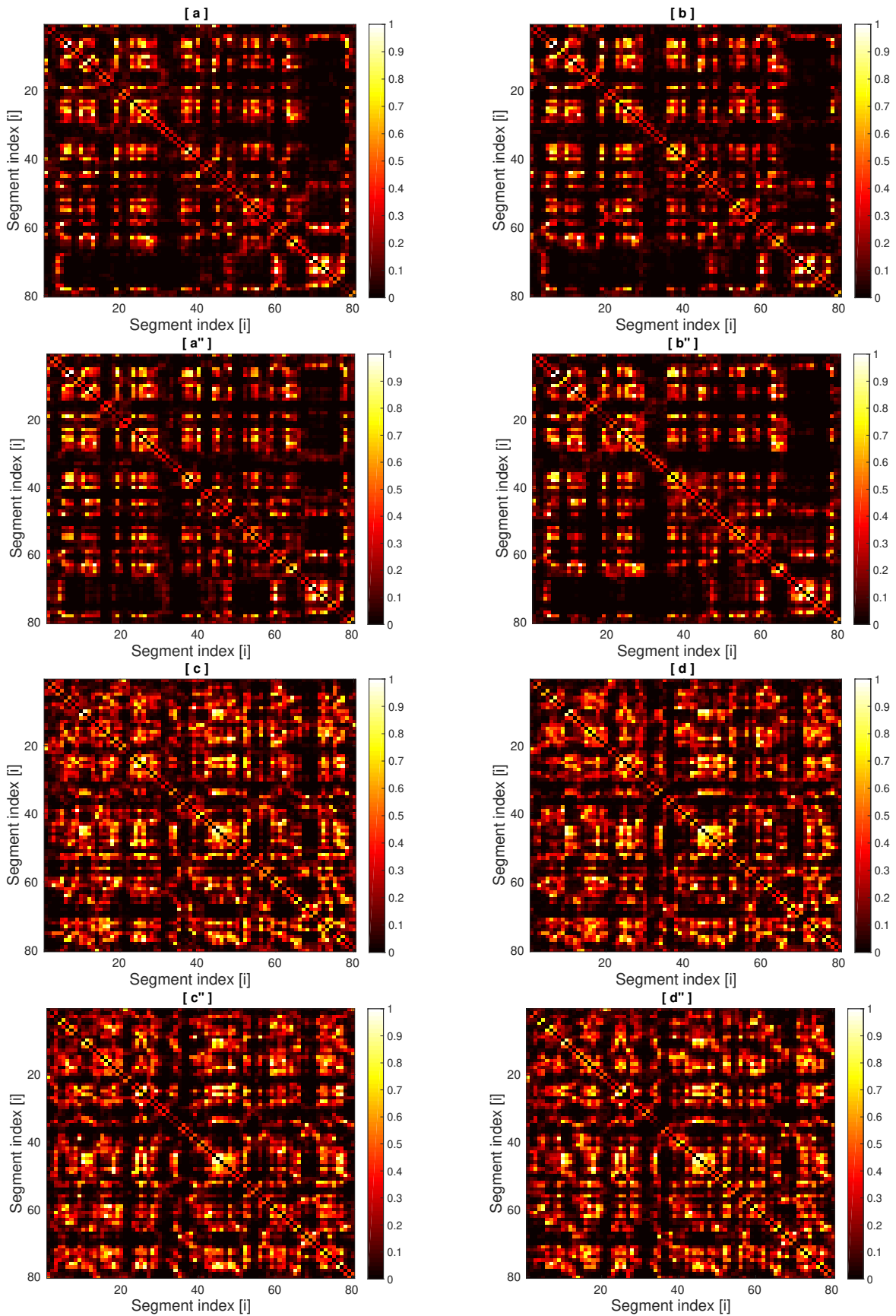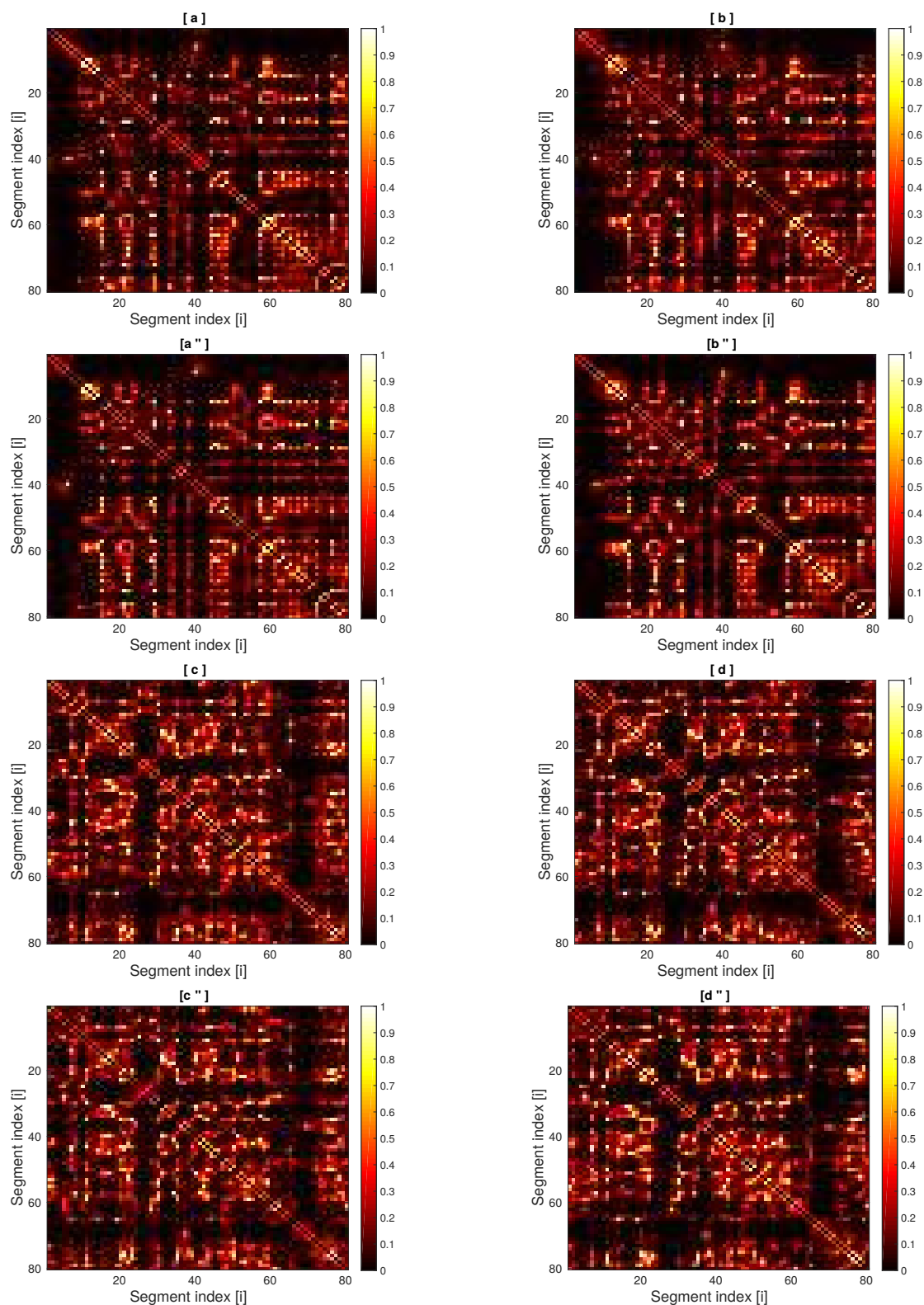
Fig. 2.12 The olormaps represent probability $p(i, j)$ to find CM of segment i spatially close to CMs of other chain segments j. There are 80 segments in the *C. crescentus* polymer with 50 monomers per segment. The top 4 figures are runs with BC-2 and the bottom four subplots RC-2.
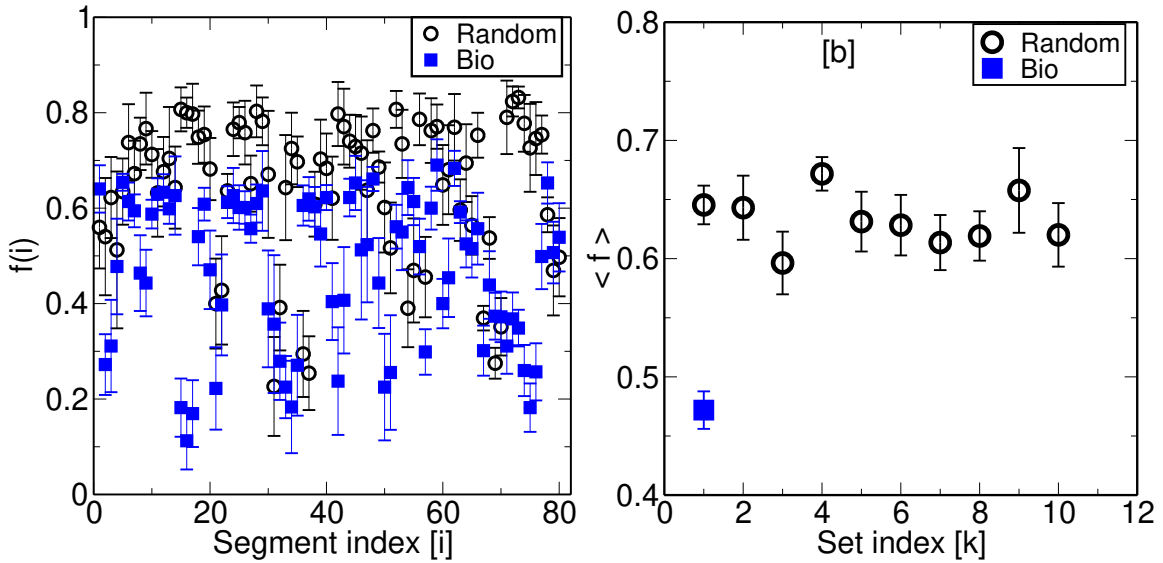
Fig. 2.13 Subplot (a) shows number of pixels $f(i)$ with probability $p(i,j) > 0.05$ for a particular segment i in the Fig. 2.11, normalized by the total number of segments versus the segment index. Subplot (b) shows $< f >$ for 10 distinct random CL sets, labelled $K = 1..10$ and one biologically determined CL set. The error bars show the standard deviation in f_{av} (see text) calculated for the 9 independent runs for each CL-set. Eighty-two pairs of monomers have been chosen randomly and then cross-linked for each CL set.
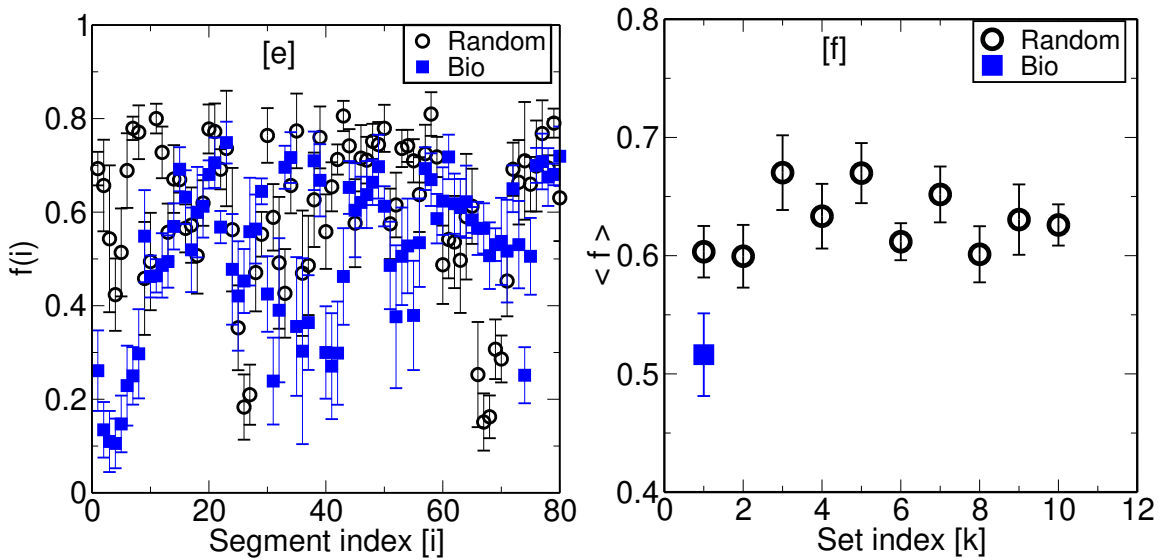
Fig. 2.14 Subplot (a) shows number of pixels $f(i)$ with probability $p(i,j) > 0.05$ for a particular segment i in the Fig. 2.12, normalized by the total number of segments versus the segment index. Subplot (b) shows $< f >$ for 10 distinct random CL sets, labelled $K = 1..10$ and one biologically determined CL set. The error bars show the standard deviation in f_{av} (see text) calculated for the 9 independent runs for each CL-set. Eighty-two pairs of monomers have been chosen randomly and then cross-linked for each CL set.

Fig. 2.15 Colormaps to investigate the angular location of different DNA-polymer segments with respect to each other for bacteria *E. coli*. Subplots (a),(b),(a"),(b") are for BC-2 and (c),(d),(c"),(d") for RC-2 with different initial conditions, respectively.
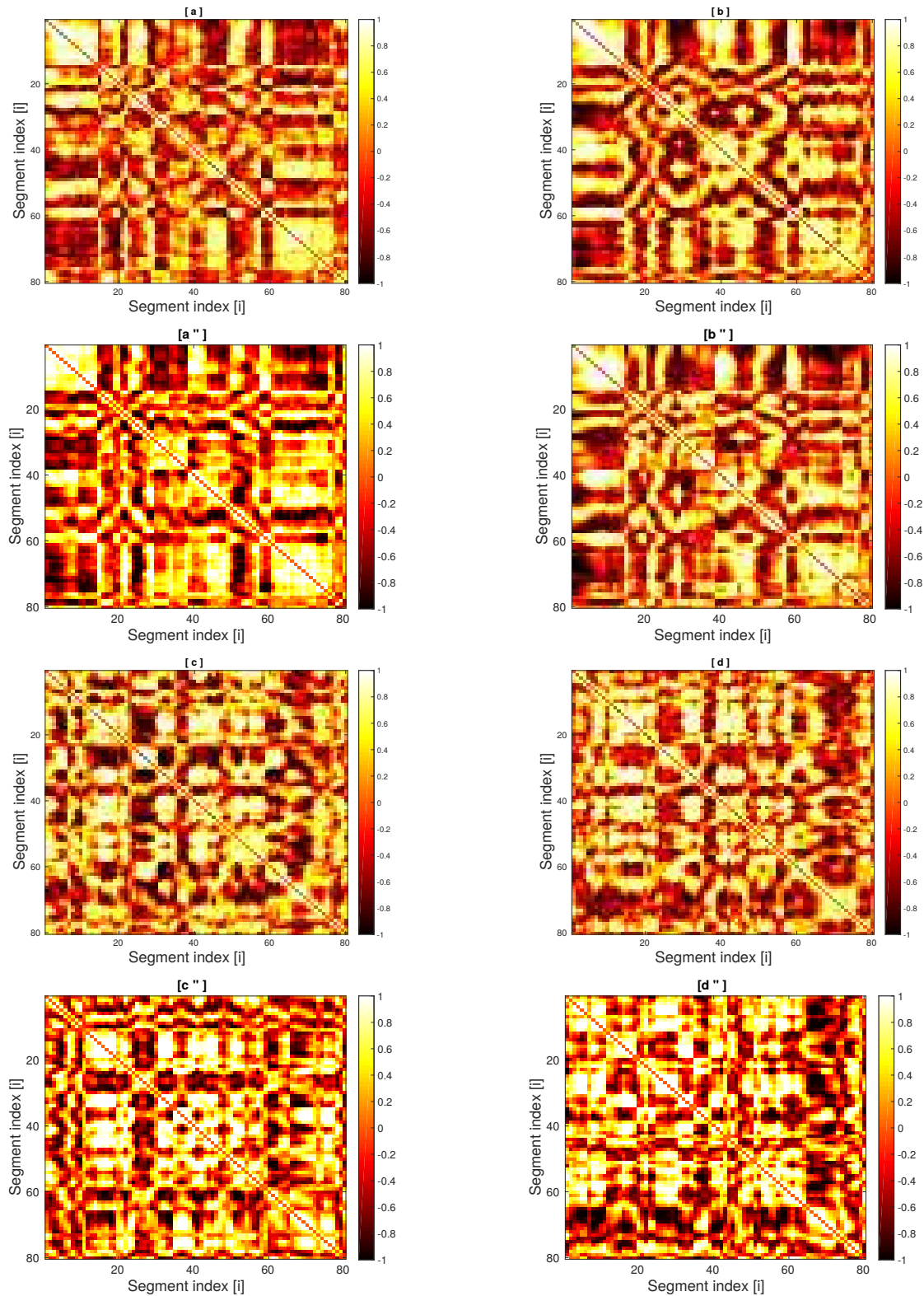
Fig. 2.16 Colormaps to investigate the angular location of different DNA-polymer segments with respect to each other for bacteria *C. crescentus*. Subplots (a),(b),(a"),(b") are for BC-2 and (c),(d),(c"),(d") for RC-2 with different initial conditions, respectively.

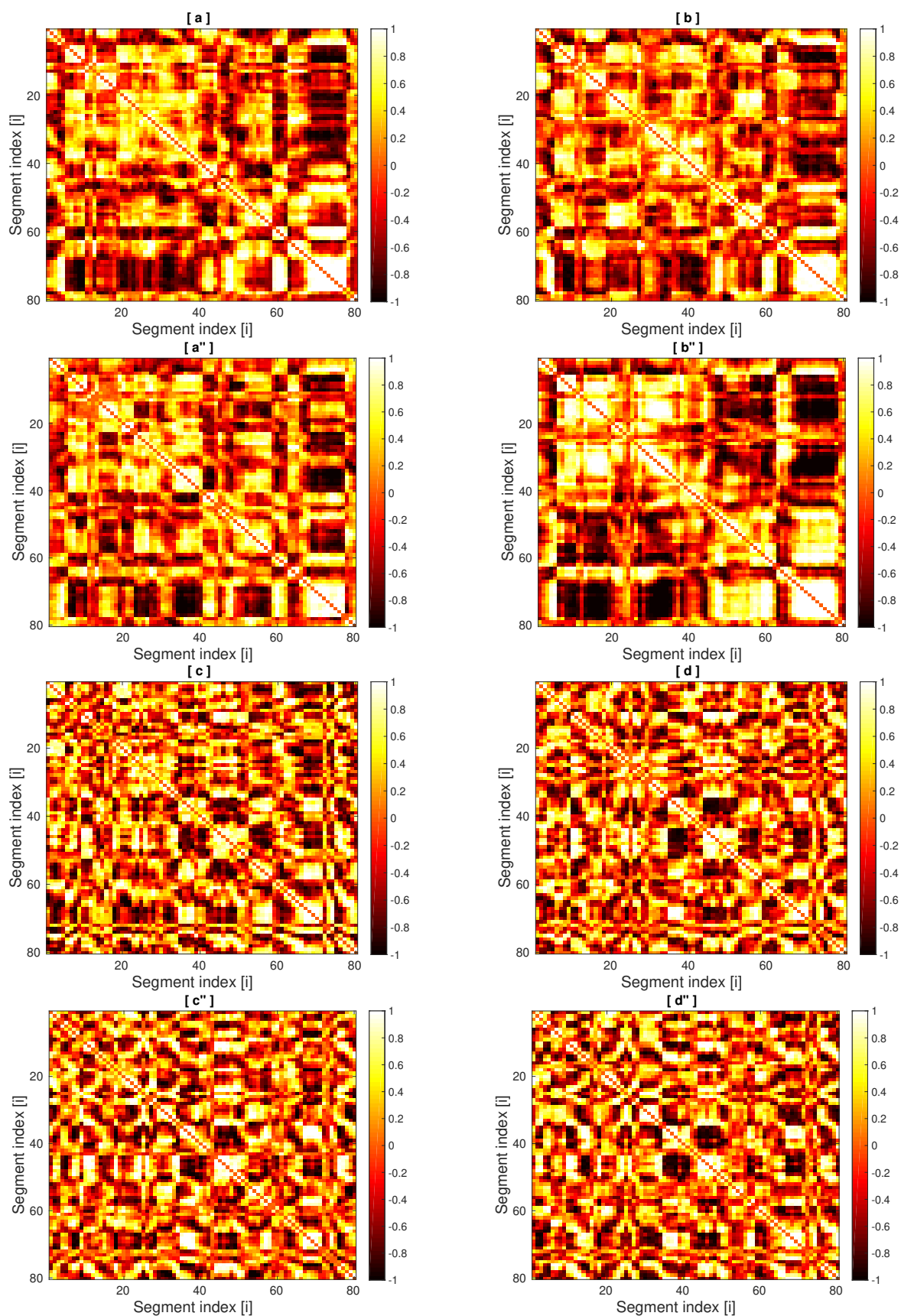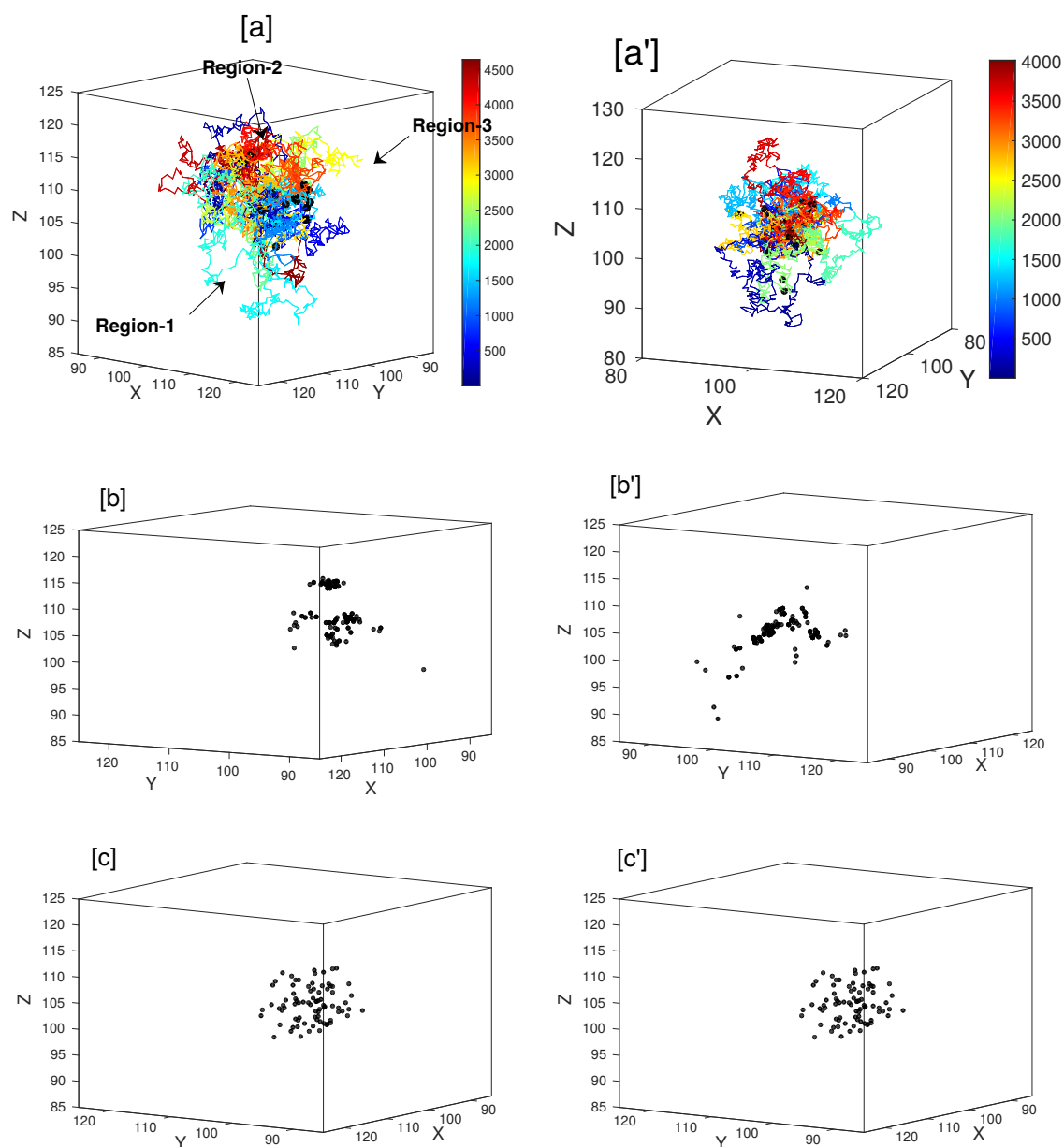Fig. 2.17 The Subfigures(a) and (a') are the representative snapshot from our simulation of DNA-polymer of bacteria *E. coli* and *C. crescentus* with BC-2 CL-sets. The color bar on the right shows the color, which is used to represent the monomers numbered from 1 to 4642/4017. The black circles show the positions of CLs. The figures (b) and (b') show the position of CLs in space for *E. coli* and *C. crescentus* where we have removed the other monomers for better visualization. The subfigures (c) and (c') represent the distribution of CLs for the RC-2 set of CLs in the polymer chain of 4642 and 4017 monomers. In figures (b) and (b'), we see there are approximately 4 clusters of CLs in space for BC-2 while the CLs are uniformly distributed in space for RC-2, as can be seen in snapshots (c) and (c').
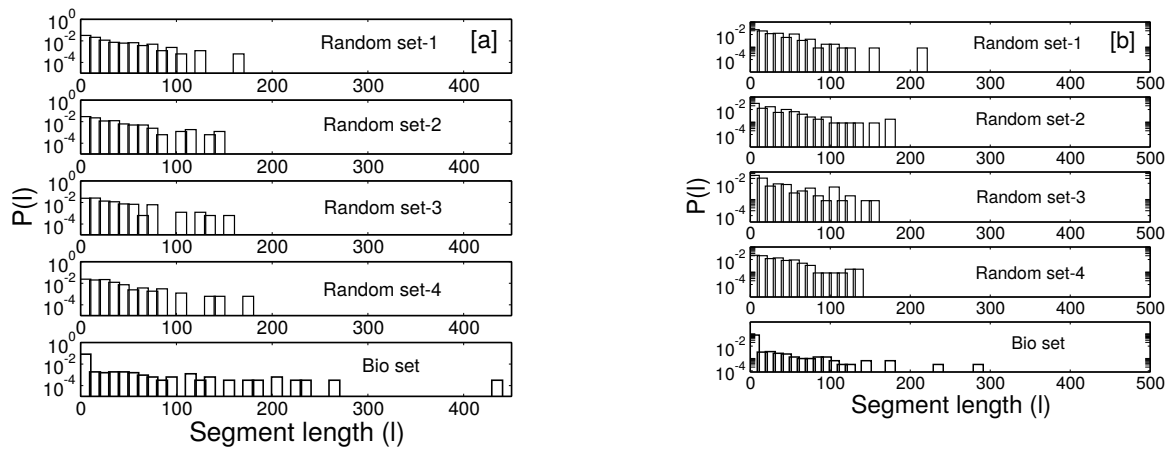
Fig. 2.18 The figures (a) and (b) represent the distribution of segment length between two adjacent CLs for BC-2 set and 4 different random CL sets corresponding to RC-2 for the DNA polymer of bacteria *E. coli* and *C. crescentus*, respectively. The x-axis shows the length of the segments between two CLs, and the y-axis shows the frequency density of the segment's length.
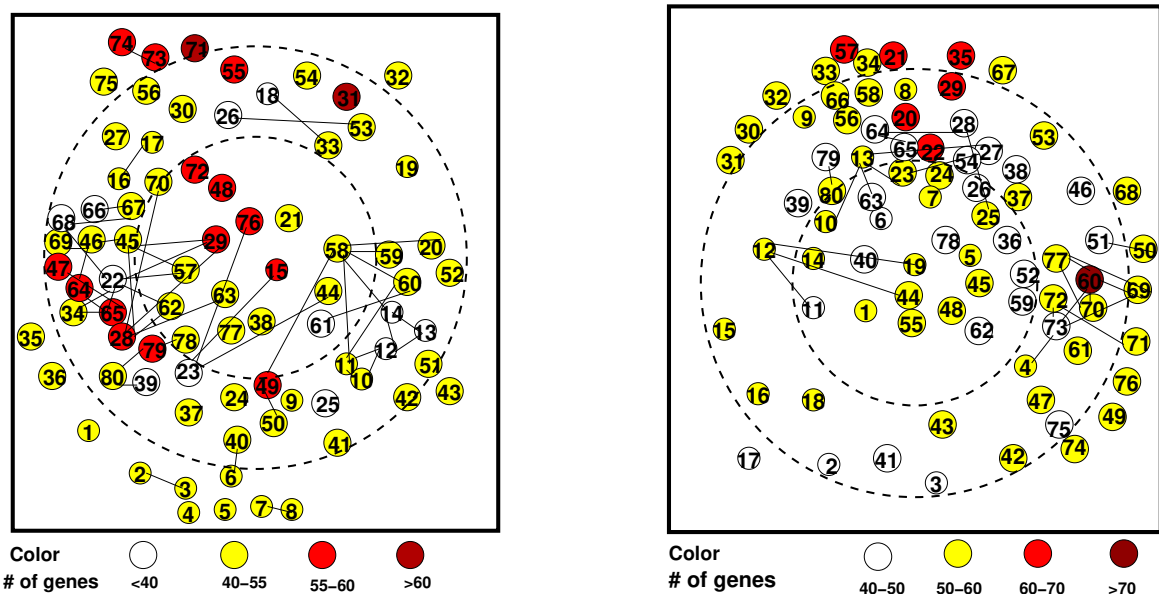
Fig. 2.19 The 2-d maps of DNA organization of *C. crescentus* and *E. coli* bacteria above the 100 nm length scale using BC-2. This is obtained using the statistical data presented in this chapter. The links between different circles indicate that colormaps show these segments to be spatially close to each other though they look separated on the map. The number inside the circle is the segment index; thus, we know the relative positions of different segments with respect to each other. The circles are colored red, blue or white depending on the number of genes in the corresponding segment
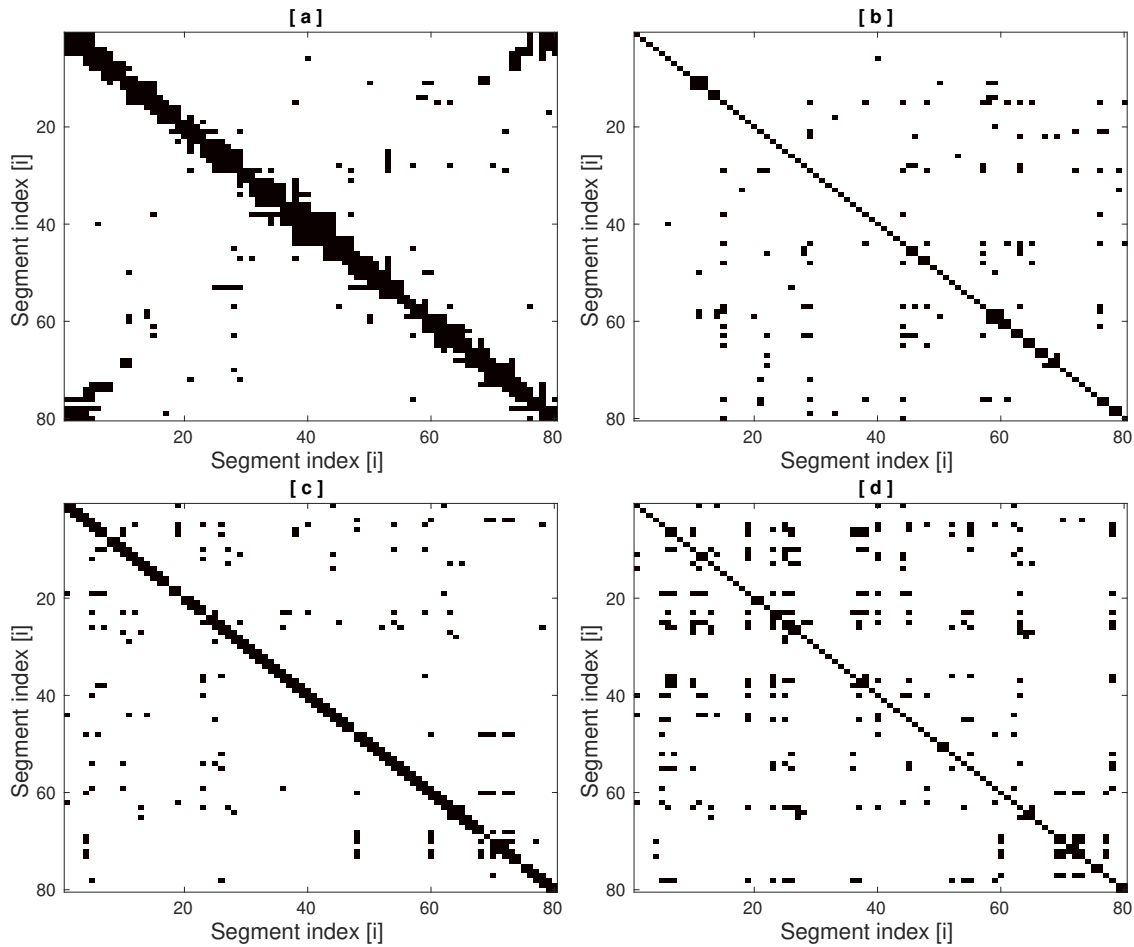
Fig. 2.20 Plots (a), (b) show binary version of the coarse-grained experimental C-map and color map obtained from simulations respectively, for *C. crescentus*. Plots (c),(d) are corresponding data for *E. coli*. To be able to compare easily by the eye, we have color coded all the frequencies *f* [probabilities *p*] with value $f > 0.0001$ [$p > 0.5$] as black for experimental [simulation] colormaps.

2.4 Conclusions

The primary conclusions of the study in this chapter is that if particular sets of monomers in a DNA-ring polymer are held together by suitable proteins (cross-links at specific points in our model polymer), it leads to an organization of the polymer coil. A minimal number of CLs are required to be able to claim that there is a distinct organization of the DNA-polymer since we do not obtain a well-defined structure with 47 bio-CLs (equivalently 27 *effective* CLs). We have also tried to look for an organization with the number of CLs in between BC-1 and BC-2, but we cannot identify well defined structure. The number of *effective* CLs is 82 for

a ring polymer of 4642 monomers, or approximately 2% of the polymer chain. Moreover, the monomers which are cross-linked in the bacterial DNA are not randomly chosen from the length of the contour and lead to an organization of the ring polymer into a particular organization which is very different and distinct compared to what is obtained using an equal number of random cross-links. Of course, the DNA polymer undergoes local conformational fluctuations due to the thermal energy, but overall, the structure is maintained in a statistical sense. We can deduce the presence of distinctive mesoscale organization of DNA from the calculation of three quantities: (a) radial distribution of segments, (b) positional correlations between segments, and (c) angular correlations between segments. Thus we have much more detailed information about the organization of different segments than that can be obtained from pair correlation function. We have used 159/153 CLs for our simulation of DNA-polymer, but these should be considered as only 82/60 *effective* CLs. Release of topological constraints through chain crossing could play a crucial in the above organization. We predict the 2-d arrangement of different segments relative to each other with the statistical quantities. Note that we have used minimal information from the (experimental) contact map to predict the global organization, and we have cross-checked that our prediction matches the data from the experimental contact map. We find the clusters of CLs towards the center of the coil, and these CLs are pulling different segments of the chain towards the center, and many loops on the periphery, which we interpret as the rosette-like structure. We have given a possible argument of how and why the structure with relatively well-localized DNA-polymer segments is achieved in a polymer, but a full understanding and systematic methodology of the choice of CL-positions from the view of polymer physics can be developed only in future, when we will have access to the larger number of contact maps of many DNAs.

Chapter 3

Role of release of topological constraints and molecular crowders in the bacterial chromosome organization

3.1   Introduction

In chapter 2, we have seen that the few cross-links at specific positions along the chain contour gives the unique and well-defined organization to the DNA polymer for bacteria *E. coli* and *C. crescentus*. Furthermore, we allowed the chains to cross itself in the study of chapter 2 to model the role of enzyme topoisomerase, which helps to release the topological constraints by suitably cutting and rejoining the chains. We also did not incorporate the effect of crowding environment by the molecular crowders (other molecules in the cytoplasm of the cell, e.g., proteins, enzymes, RNAs, etc.) in the organization of the chromosome in chapter 2.

Thus in this chapter, we study the effect of the release of topological constraints and the effect of molecular crowders in the organization of the bead-spring polymer with few specific cross-links. Next, we will be presenting our model and results.

3.2   Model

In this section, we describe our polymer model to study the role of the release of topological constraints by chain crossing and the effect of the molecular crowders in the organization of the bacterial DNA polymer. As in the previous chapter, We model the bacterial chromosome as a cross-linked bead spring polymer, where adjacent bead along the polymer contour interact with harmonic potential with spring constant $\kappa = 200 k_B T / a^2$ and bond-length $a = 1$.

The cross-links are chosen from the contact map of the chromosome of bacteria *C. crescentus* by setting a suitable frequency cut-off as in the previous chapter. We work with the BC-2 set of cross-links (study in chapter 2) corresponding to the bacteria *C. crescentus* for the study in this chapter. The cross-links are again modeled by the harmonic springs with $\kappa = 200 k_B T/a^2$ and the bond length $a = 1$, similar to the study in chapter 1. However, the excluded volume interaction is modeled differently in the two studies and is explained in the following paragraphs.

3.2.1 Chromosome Model: to study the role of the release of topological constraints

In this case, the excluded volume between non-adjacent beads is modeled by the WCA potential V, same as in the previous chapter, where the strength of the potential ε is chosen to be 1. The diameter of the bead σ is varied from $\sigma = 0.2, 0.3, ..., 0.5$ to study the different extent of release of topological constraints. In other words, chain crossing is not allowed for the higher value of bead diameter ($\sigma > 0.4a$), while the chains can cross itself for $\sigma \leq 0.4$.

3.2.2 Chromosome Model: to study the role of presence of molecular crowders

For this study, the excluded volume interaction between the non-adjacent monomers along the chain countour is not modeled by the same potential V, as in the previous chapter, but is modeled by the Lennard-Jones potential V' suitably cut at $r_c = 3.0\sigma$ and then shifted.

$$
V' = \begin{cases} 4\varepsilon \left[\left(\dfrac{\sigma}{r} \right)^{12} - \left(\dfrac{\sigma}{r} \right)^6 \right] + V_{LJ}(r = 3\sigma), & \text{if } r < 3\sigma \\ 0, & \text{otherwise} \end{cases}
$$

Here, the potential V_{LJ} is of the form: $V_{LJ} = 4\varepsilon \left[\left(\dfrac{\sigma}{r} \right)^{12} - \left(\dfrac{\sigma}{r} \right)^6 \right]$. We take the value of the bead diameter $\sigma = 0.2a$ to be fixed, same as in the previous chapter, which we chose to enable the chain crossing. The parameter ε which is the depth of the potential, decides the strength of the interactions between the monomers. The potential V' is chosen to model the the effect of molecular crowders by taking a small attraction (compared to the thermal energy $k_B T$) between the monomers at short length-scales.

In this chapter, we will study the organization of the polymer with BC-2 sets of cross-links for different values of parameter $\varepsilon = 0.1, 0.2..., 0.5k_B T$. We do not want to investigate the

collapse of the DNA-polymer with changing solvent conditions. Thus, we restrict ourselves to the values of $\varepsilon/k_B T < 1$, such that the local density of monomers is less than that observed for a collapsed globule state and the monomers can still explore phase space without getting kinetically trapped.

To equilibrate the system, we use the same protocol as described in the previous chapter. We also start our Monte-Carlo runs from the 9 initial independent conditions, and after the equilibration runs, start collecting the average structural quantities to identify the organization of the model chromosomes. We next compare these structural quantities for different values of parameter ε.

3.3 Results

Before we present how (a) changing the excluded volume parameter σ (b) changing the value of ε affects the organization of the chromosome, we first present the analysis of how the increase in monomer diameter σ changes the size of the polymer. This and other statistical quantities that we first discuss will help us in understanding and analyze our subsequent results.

3.3.1 Effect of varying the value of excluded volume parameter σ.

For Gaussian linear polymer chains (without considering excluded volume (EV) interactions between monomers), the radius of gyration $R_g \sim N^{0.5}$, whereas if the polymer is self-avoiding then $R_g \sim N^{0.6}$. For the self-avoiding ring polymers $R_g \sim N^{0.65}$ [121]. Self-avoidance can be modeled by a bead-spring model of a polymer chain with EV interactions modeled by a suitably truncated Lennard Jones potential between beads to retain only the repulsive part of the potential. It has been shown for linear polymers that any value of $\sigma/a \neq 0$ gives the value of scaling exponent to be 0.6 though the absolute value of R_g decreases [143] with the decrease in the value of parameter σ. An estimate of the size of the polymer is given by the radius of gyration R_g of the chains. The radius of gyration is calculated as $\sqrt{(I_1 + I_2 + I_3)/2M}$, where I_1, I_2, I_3 are the eigenvalues of the moment of inertia matrix and $M = mN$; we have taken $m = 1$. For our studies, we have a ring polymer with cross-links at very specific points along the chain; we cannot expect the scaling which is relevant for linear or a ring polymer without CLs. For our studies, we focus on the increase in the size of the DNA ring-polymer of fixed length and specific CLs with the increase in the EV parameter σ.

To this end, we plot R_g versus σ in Fig. 3.1(a) for the polymer with BC-2 and RC-2 CLs in a log-log plot. The RC-2 CL set has the same number of CLs as the number of effective

CLs in BC-2, but the monomers which are cross-linked are chosen randomly along the chain contour (also refer appendix-1). The error bars in Fig. 2.1(a) show the standard deviation in the value of R_g for the polymer starting from 9 independent initial conditions. From Fig. 2.1(a) we see that the value of R_g increases with σ for polymer with the BC-2 and RC-2 from $9a$ to $15a$ and $7a$ to $13a$, respectively. The slope of the graph gives the exponent for the scaling relation. From the graph, we see that the value of R_g scales as $R_g \sim \sigma^{0.5}$ for the polymers for both BC-2 and RC-2 set of CLs.

To analyze this, we note that in our previous studies, we observed that with the BC-2 set of CLs, the CL monomers had clustered towards the center of the coil, and there were lengthier loops of monomers which were in the peripheral region of the globule [136, 137]. On the other hand, for the RC-2 CL set, the CL monomers were distributed along the contour and in space randomly. As a consequence, relatively longer loops were not present on the periphery of the polymer globule. Due to this, there was higher compaction of the globule, and the value of R_g in case of polymer with RC-2 CLs was less compared to the polymer with BC-2 CLs. Now the question is, with the increase in the value of σ will the increase in R_g is due to the overall swelling of the chain with BC-2 CLs or is the amount of swelling is affected by the location of CL monomers in the chain contour? Since the exponent in Fig. 2.1 remains the same for both RC-2 and BC-2 set of CLs; it appears that there is an overall swelling of the chain. We conclude this because the increase in the size of the polymer does not depend on the locations of the CL monomers in the chain contour.

The value of R_g gives only an estimate of the overall extent of the polymer but does not give the information about the internal reorganization of monomers due to the increase in σ. In the case of polymer with BC-2 CL set, the clusters of cross-links pull a large number of other monomers near the center of the coil and increase the monomer density in the inner part of the globule. To check how the radial distribution of monomers changes with the parameter σ, we calculate the number density of monomers from the center of mass (CM) of the polymer globule. The next question is whether relatively taut stretches of polymer between two individual CLs hold these clusters of CLs near the center of the coil? In this scenario, if σ is increased, the average distance between CL clusters and their spatial locations could remain relatively unaffected, and the effect of swelling could be significant only at the peripheral loops. In the other scenario, the core region will swell along with the peripheral regions of the polymer coil.

To investigate this, in Fig. 3.2(a),(b) we plot the number density of the monomers $n_M(r)$ and the cumulative number of CLs $c\text{-}n_M^*(r)$, normalized by total number of CLs, respectively, as a function of the radial distance r from the center of mass (CM) of the globule. The quantities $n_M(r)$ and $c\text{-}n_M^*(r)$ are plotted for different values of the parameter σ. The error
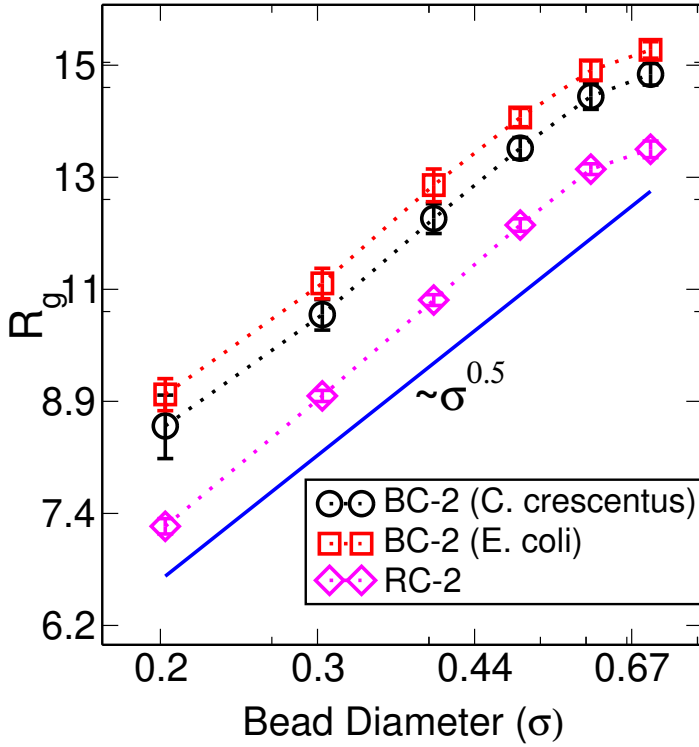
Fig. 3.1 The plot shows the value of the radius of gyration (R_g) in a log-log plot for different values of bead diameter σ. Different lines correspond to the R_g of polymer with BC-2 sets of CLs chosen from the experimental contact map of *C. crescentus*, *E. coli* and with the RC-2 set of CLs. The average is taken over 9 independent initial conditions and the standard deviation is shown by the error bars. The blue line corresponds to the scaling $R_g \sim \sigma^{0.5}$ and is the guide to the eye.

bars show the standard deviation from the average value in 9 independent initial conditions. From Fig. 3.2(a), it can be confirmed that as we increase the value of parameter σ, the value of $n_M(r)$ decreases near the center of the coil (for low values of r). This indicates that there is a swelling of the polymer core with the increase in the value of σ. But $n_M(r)$ has non-zero values even at larger r when the the value of σ is increased. The normalized cumulative number of CLs, $c\text{-}n_M^*(r)$ is shown in the Fig. 3.2(b). From the figure, we see that for low values of σ (say $\sigma = 0.2a$) the value of $c\text{-}n_M^*(r)$ approaches the value 1 at $r = 11a$ from the CM of the chain. Comparatively, $c\text{-}n_M^*(r)$ reaches 1 at $r = 15a$ for $\sigma = 0.5a$. These results again indicate that the increase in the size of the polymer is because of the overall swelling of the chain.

We also check for the average distance between the CL monomers for different values of parameter σ. This is shown in the Figs. 3.2(c),(d) as the colormaps. The x and y-axis represent the CL index, and the color represents the average distance between the monomers
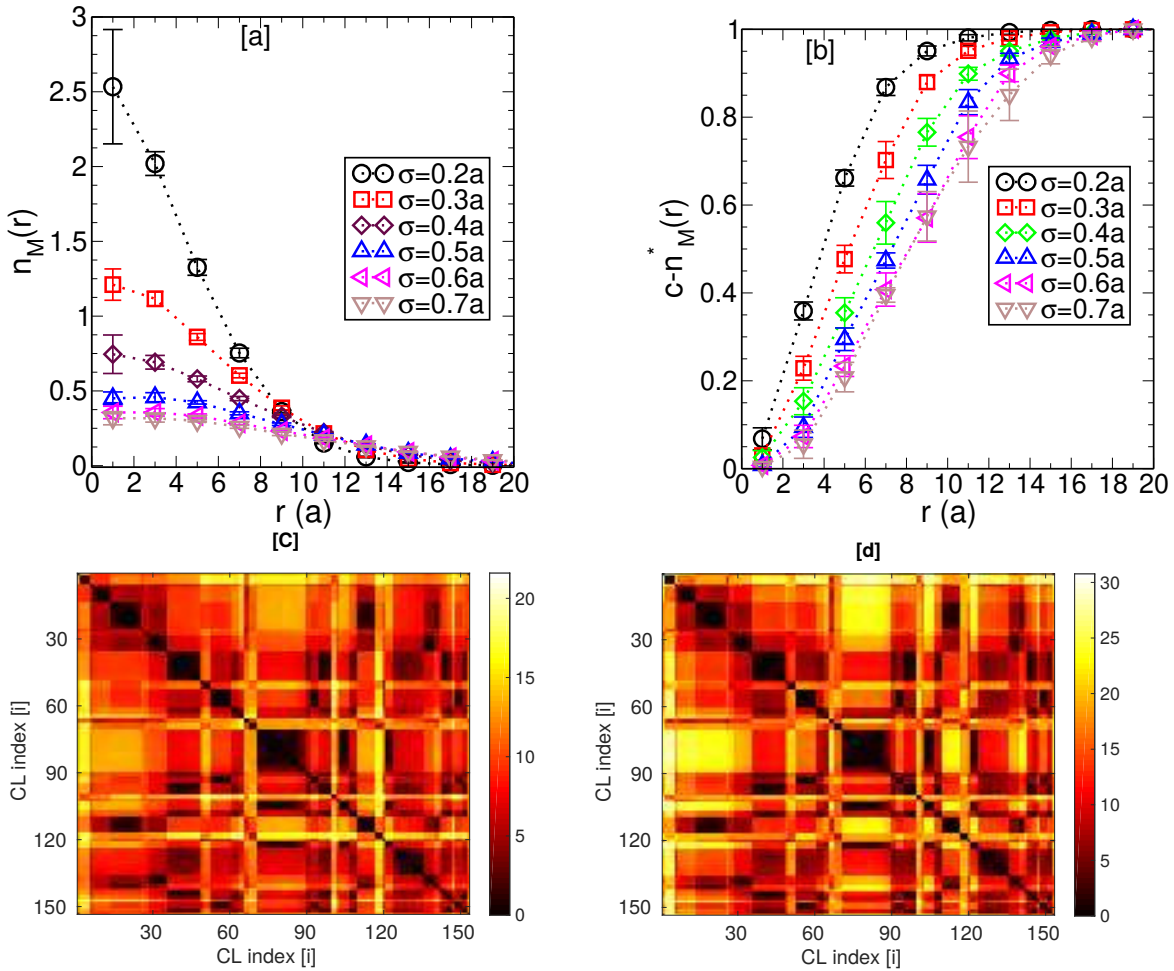
Fig. 3.2 (a) The plot shows the number density of monomers $n_M(r)$ as a function of the distance (r) from the center of mass of the globule for different values of parameter σ. (b) The graph shows the cumulative number of CLs normalized by the total number of CLs as a function of distance from the CM of the polymer for different values of bead diameter σ. The colormaps (c) and (d) represent the average distance in units of a between the CL monomers i and j for $\sigma = 0.3a$ and $0.7a$, respectively. The color denotes the average distance between CLs i and j. The graphs correspond to the model chromosome with the BC-2 set of CLs of bacteria *C. crescentus*.

constituting the CL, note that the range of color bars is different. The dark color represents the two CLs that are in proximity, and the bright color represents the two CLs that are far from each other. The average value of the distance is given in the color bar. The two representative colormaps in the figure correspond to $\sigma = 0.3a$ and $\sigma = 0.7a$, respectively. From the figure, we find that the average distance between the CL monomers increases with σ, but the pattern of dark and bright pixels in the colormap remains the same. This suggests that there is an

overall swelling of the CL cluster with the increase in the value of EV parameter σ hence the increase in the value of R_g.
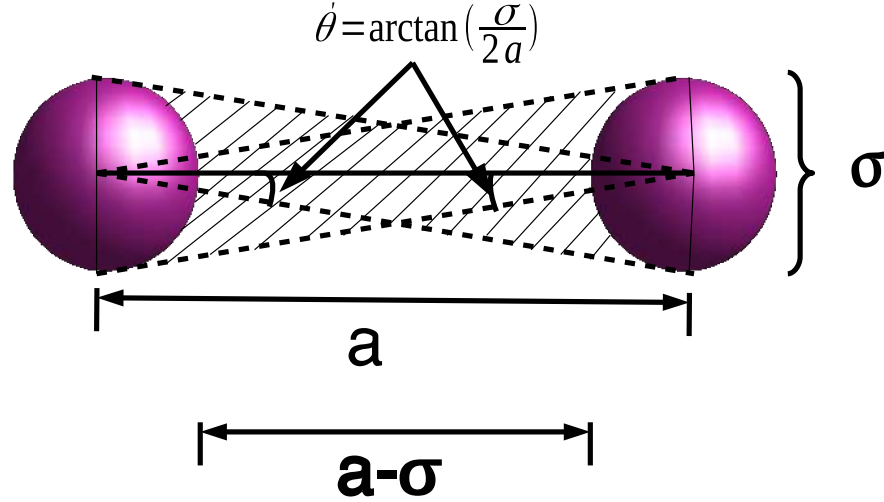

Fig. 3.3 The schematic diagram shows the shaded region between the two neighboring monomers (along the chain contour) where the third (non-neighboring) monomer can come. We calculate the average number of monomers whose centers lie in the shaded region to calculate $\langle n_{CC} \rangle$. For details, refer to the text.

We expect that the crossing of chains should not be feasible once the value of the parameter $\sigma > 0.4a$. Even if we introduce Monte-Carlo attempts with trial large-displacement of monomers to encourage chain crossing, chain crossings will not occur as large trial displacements MC attempts leads to a high energy cost arising from large extensions of harmonic bonds. For smaller values of σ, the chain crossings can occur frequently, and the probability P_{CC} of chain-crosses should decrease as we increase the value of σ. It is difficult to calculate the exact frequency of chain crossing in our simulations. So, to estimate the frequency of chain crossing in each Monte-Carlo step for different values of EV parameter σ, we calculate the quantity $\langle n_{CC} \rangle$, which gives the average number of monomers (from the other parts of the chain) between two neighboring monomers along the chain. It is a reasonable estimate because monomers that are in between two adjacent monomers along the chain can either cross the chain in the next MC attempt or move away.

We calculate the quantity, $\langle n_{CC} \rangle$ as follows. The distance between the nearest surfaces of two neighboring monomers along the chain i and $i+1$ is $a - \sigma$, refer to the schematic diagram of Fig. 3.3. Suppose another monomer j comes in between the two monomers i and $i+1$. The angle θ' between the vector joining the center of the monomer i and j (or alternatively between $i+1$ and j) and the vector joining the center of the monomers i and $i+1$ should be $\theta' \leq tan^{-1}(\frac{\sigma}{2a})$. It is shown as shaded region in the schematic diagram of

Fig. 3.3. We calculate the number of monomers that are satisfying the above condition and normalize it by the total number of monomers N to obtain $\langle n_{CC} \rangle$. The quantity $\langle n_{CC} \rangle$ is plotted in Fig. 3.4 versus the EV parameter σ. It is averaged over 9 MC runs starting from 9 independent initial conditions. From the figure, we see that on increasing the value of the parameter σ the value of $\langle n_{CC} \rangle$ decreases and for $\sigma \geq 0.5$ the value of $\langle n_{CC} \rangle$ becomes zero, as expected. From this data, we can estimate that the frequency of chain crossing drops rapidly with an increase in the value of σ.
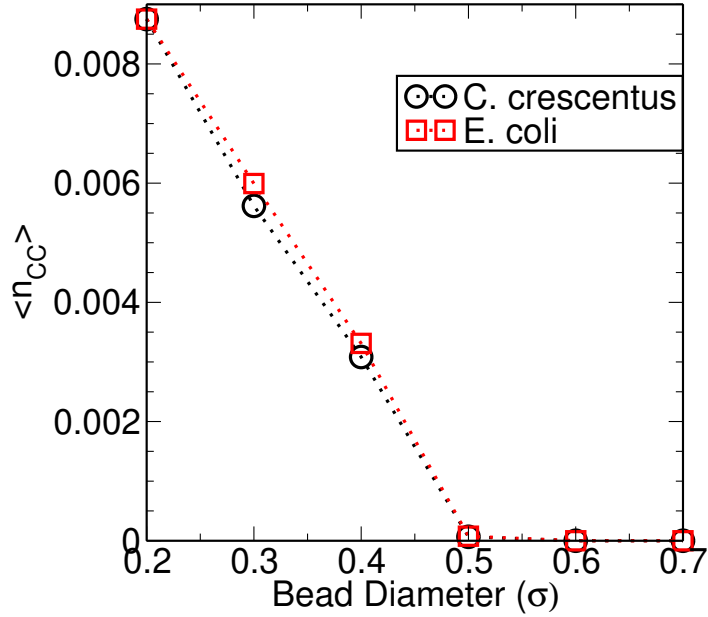

Fig. 3.4 The plot shows the bead diameter σ on the x-axis, and the quantity $\langle n_{CC} \rangle$ which is an estimate of chain crossing is plotted on the y-axis. Refer text for the precise definition of the quantity $\langle n_{CC} \rangle$.

Now with this background, we would like to focus on how increasing σ affects the overall *organization* of the DNA-polymer. Instead of calculating radial distribution functions $g(r)$ of different segments of the chain, we calculate the positions of different chain segments with respect to each other to quantify and analyze the organization of a chain. For this, we calculate the positional correlation between different segments, as we did in our previous studies [136, 137]. We determine the position of the center of mass (CM) of each segment (50 monomers each) and calculate the probability of the CM of two segments to be at a distance less than a cutoff distance, R_c. In our previous papers, we chose the value of $R_c = 5a$ to be nearly half of the value of $R_g = 9a$ for $\sigma = 0.2a$. Since R_g scales as $R_g \sim \sigma^{0.5}$ with different values of the EV parameter σ, we choose the values of R_c for this study, such that the ratio $R_c/R_g \approx 0.55$ is maintained consistent with our previous studies [136, 137]. (Note that we incorrectly calculated the value of R_g to be $\approx 7a$ for $\sigma = 0.2a$ in [137], as we used

the expression $R_g = \sqrt{(I_1 + I_2 + I_3/3M)}$. The correct formula is $R_g = \sqrt{(I_1 + I_2 + I_3/2M)}$. Using this correct expression, $R_g \approx 9a$.)

The positional correlations between the CMs of different segments are shown in the colormaps of Fig. 3.5 for $\sigma = 0.2a, 0.4a, 0.7a$, respectively. In the colormaps, the x-axis and the y-axis represent the segment index corresponding to the CM of the 80 segments. The color represents the probability of the two segments to be within a cut-off distance of R_c. The bright color shows a higher probability of two segments to be within a cut-off distance of R_c. From Figs.3.5, we see that as we increase the value of the parameter σ from $0.2a$ to $0.7a$, the rectangular patch like-pattern of the colormaps start disappearing and for $\sigma = 0.7a$ the rectangular patch-like pattern completely disappears. The rectangular patch like-pattern (where all the neighboring pixels are of the same color) in the colormaps show that the neighboring segments along the chain contour of a particular segment-i (50 monomers) come close in the 3D space to another segment j with the similar probabilities as that of i, where i and j are far apart along the contour. Also, the number of bright pixels in the colormaps decreases in-spite of choosing the higher values of R_c with increasing R_g in each case. This has also been quantified later in the result section. Hence, there are indications to conclude that with the increase in the value of σ, there is a loss of organization or structure of the DNA-polymer. We also get similar qualitative results for the colormaps of the polymer with the BC-2 set of cross-links chosen from the experimental contact map of bacteria *E. coli*.

Till now, we have been identifying the organization from the visual inspection of the color-maps, but we need more suitable quantities to quantify and compare the level of the organization. We can claim that a particular ring polymer gets organized in the presence of CLs only if the colormaps from different MC runs (starting from different initial conditions) relax to the same structure. As seen previously, the BC-2 CLs lead a unique organization/structure of the polymer, and all the nine independent runs give the same positional correlations colormap for $\sigma = 0.2a$. With the increase of the value σ, this will remain true unless the polymer gets kinetically stuck in different configurations. To quantify the similarity of colormaps, we calculate the Pearson correlation between the colormaps of positional correlations from the runs starting from 9 independent initial conditions. A high value of Pearson correlation on comparing positional correlation colormaps from different runs is indicative of the same resultant organization of the polymer chain due to CLs. This, in turn, will lead to a statistical measure that will quantify how well-defined a structure is, in spite of the presence of thermal fluctuations.

For a particular value of σ, the 9 independent runs give $^9C_2 = 36$ comparisons between pairs of colormaps, and thereby 36 values of Pearson correlation. We can calculate the average value of pc, and also get the SD (standard deviation) from the mean. Low values
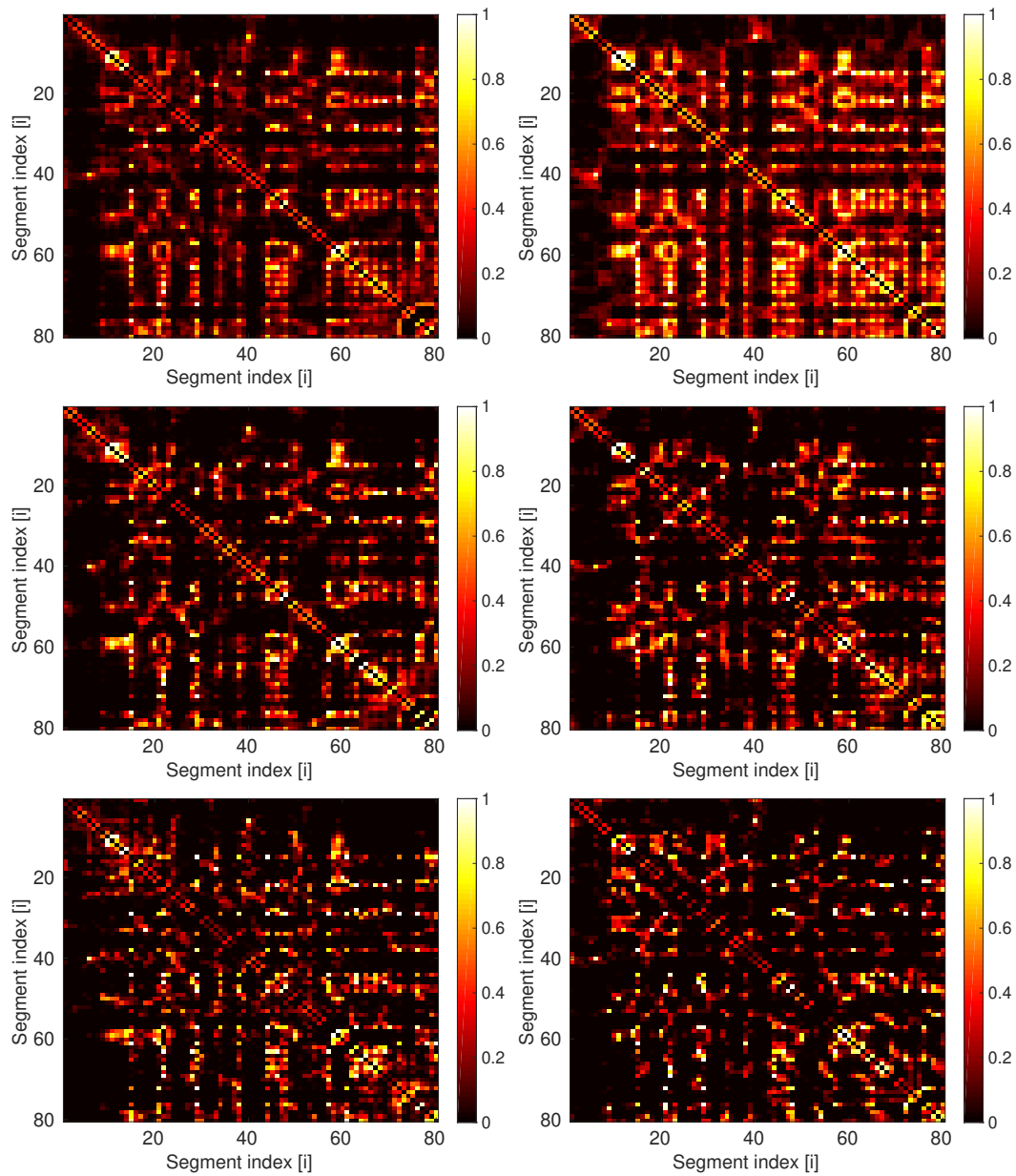
Fig. 3.5 The plots show the positional correlation colormaps of different polymer segments
for three value of bead diameter σ for the model chromosome of bacteria *C. crescentus*. The
colormaps in the uppermost row are for $\sigma = 0.2a$, middle ones are for $\sigma = 0.4a$ and the
colormaps in the lowermost row correspond to $\sigma = 0.7a$. We present only one colormap for
each value of σ, though we have 9 other colormaps from independent runs for each value of
σ. The colormaps in each row represents the positional correlations of different segments
from two independent runs starting from independent initial conditions. While calculating
the positional correlations we keep the value of $R_c/R_g = 0.55$. Refer, the text for details.

Fig. 3.6 The y-axis shows the average value of the Pearson correlation ($\langle pc \rangle$) of positional correlation colormaps with parameter σ on the x-axis. The average is taken over $^9C_2 = 36$ values of Pearson correlation. Error bars represent the SD from the mean value. The correlation is calculated over only those pixels (i, j) for which $p_{ij} > 0.05$ in at least one of the runs. This is done to prevent large contribution to $\langle pc \rangle$ from the regions which are dark in all color maps.
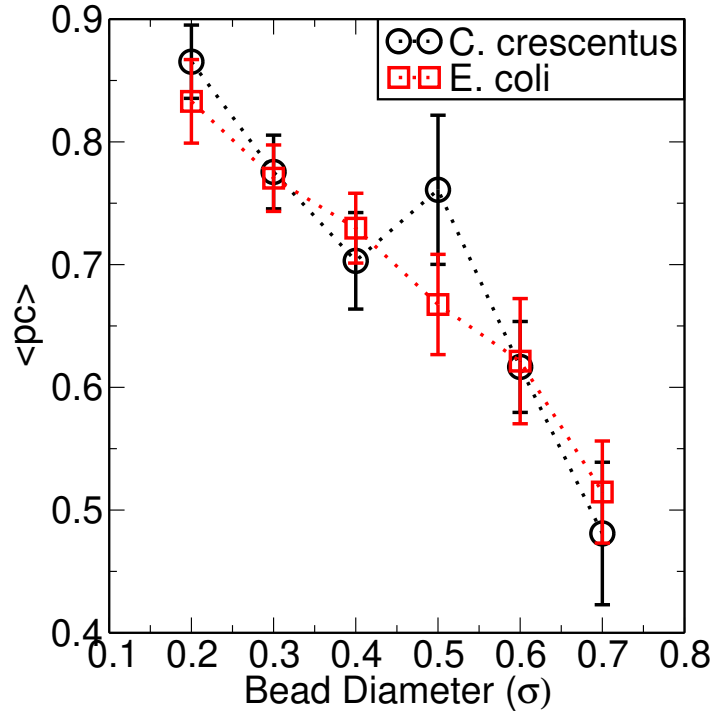
of pc and large SD values represents positional correlation colormaps are less correlated and the structure of the polymer is different across different runs. We calculate the Pearson correlation between two colormaps by the following formula.

$$pc(\alpha\beta) = \frac{\langle (p_{ij}^{\alpha} - \langle p_{ij}^{\alpha} \rangle)(p_{ij}^{\beta} - \langle p_{ij}^{\beta} \rangle) \rangle}{\sqrt{(\langle (p_{ij}^{\alpha} - \langle p_{ij}^{\alpha} \rangle)^2 \rangle)((p_{ij}^{\beta} - \langle p_{ij}^{\beta} \rangle)^2 \rangle)}}$$

Here p_{ij}^{α} and p_{ij}^{β} correspond to the probability of two segments i and j to be within the cutoff R_c in the colormaps obtained from the runs (run index α, β) starting from two different initial conditions. The average $\langle ... \rangle$ has been taken over all values of i and j which have probability $p_{ij} > 0.05$ in at least one of the 9 independent runs. We chose the pixels with probability $p > 0.05$ to avoid the bias from the large parts of dark areas of the colormap, which can result in a high value of $\langle pc \rangle$.

The average value of the Pearson correlation is plotted in Fig. 3.6 versus the parameter σ for the positional colormaps of bacteria *C. crescentus* and *E. coli*. From the figure 3.6 we see

that as we increase the parameter σ the value of $\langle pc \rangle$ decreases and the SD also increases. This suggests that the polymer organizes in different structures in the runs starting from the different initial conditions on increasing the value of parameter σ. Note that even for high values of the parameter σ the value of $< pc >$ is nearly $0.5 - 0.6$ and not lower, this is because the monomers which constitute the CLs will come closer to each other for every value of σ as they are connected by the springs. But the rest of the polymer is not able to organize in a particular structure because it cannot overcome the topological constraints.

To quantify the differences between the colormaps of Fig. 3.5 we calculate f_i the number of segments which are at a distance $< R_c$ from segment i with probability $p_{ij} > 0.05$. We can then calculate $f = \sum_i f_i / N_{seg}^2$, $N_{seg} = 80$. The value of the cutoff probability p is the same as the value we chose in our previous work [137, 136] and correspond to the deep red color in the colormap. We then calculate the average value of $\langle f \rangle$ and the standard deviation from $\langle f \rangle$ for 9 independent initial conditions. In Fig. 3.7 $\langle f \rangle$ is plotted on the y-axis with the parameter σ on the x-axis. From the figure, we observe that with the increasing value of σ, $\langle f \rangle$ is decreasing. This means less number of segments are found within distance R_c with each other. This is a non-trivial result because it indicates that the polymer is not only swelling (as was indicated by the previously calculated statistical quantities), but the internal structure of the polymer (as measured by positional correlation colormaps) is also decreasing which in turn leads to fewer bright pixels in the colormaps at higher values of σ. We get the qualitatively similar results for the positional colormaps corresponding to the polymer with the BC-2 set of CLs of bacteria *E. coli*. This can be seen in Fig. 3.6, where the $\langle pc \rangle$ is plotted corresponding to the colormaps of the model chromosome of *E. coli*.

Thus we can conclude from here that the release of the topological constraints are necessary for a polymer with CLs to organize itself into a particular structure.

3.3.2 Effect of varying the attraction strength ε

We now systematically study the role of the parameter ε in the organization of the DNA polymer. As mentioned earlier, for this study we have kept the value of σ to be fixed at $0.2a$, and set the cutoff of LJ potential to $r_c = 3\sigma = 0.6a$ and vary ε from $0.1k_BT$ to $0.5k_BT$. When we show data for attraction strength $\varepsilon = 0$, it implies that the r_c is $2^{1/6}\sigma$ with $\varepsilon = 1$ such that the interaction between the monomers is purely excluded volume. We first calculate the radius of gyration R_g as we vary parameter ε to estimate how the polymer with CLs shrinks in size. With the increase in the value of ε, the value of R_g should decrease because
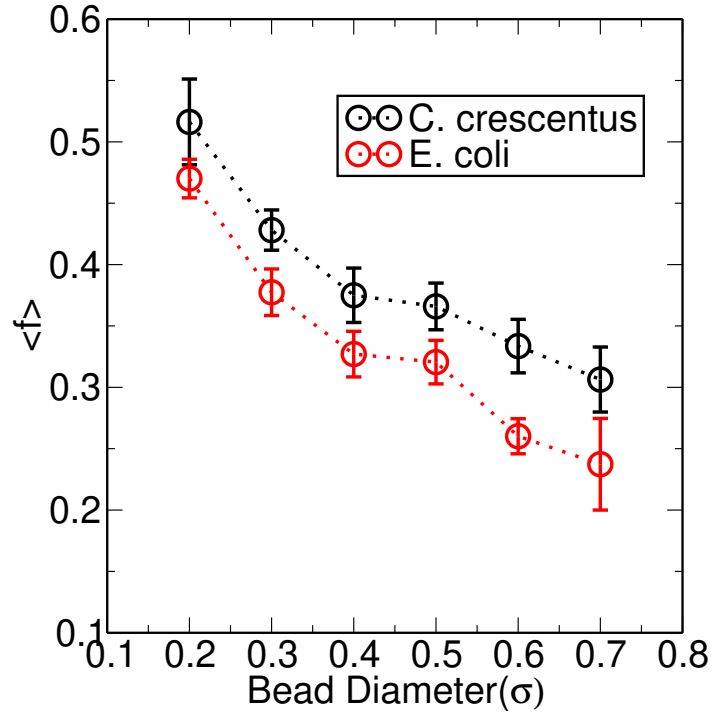
Fig. 3.7 The y-axis in the figures shows the number of segments $\langle f \rangle$ (normalized by the total number of segments) which are at a distance $< R_c$ with other segments with probability $p > 0.05$. The x-axis in the plot shows the parameter σ. The average is taken over 9 independent initial conditions, and the standard deviation is shown by the error bars.

the attraction between the monomers will lead to the collapse of the polymer at higher values of ε which will lead the polymer to form a polymer-globule.

The decrease in the value of R_g with increasing ε is shown in Fig. 3.8. The error bars represent the SD from the average value of R_g for the polymer with BC-2 CLs starting from 9 independent initial conditions. For the higher value of the parameter ε (e.g., $\varepsilon = 0.5k_BT$), we see that the error bars are larger. To understand this, we observe the snapshots of the polymer with CLs starting from the different initial conditions. From the snapshots, we observe that the bigger loops (which are in the peripheral region for the low value of ε) are unable to come on the periphery for higher values of ε in some initial conditions. This is because of the strong attraction between the monomers. It leads to the small value of R_g for some initial conditions and higher for others hence the larger error bar for $\varepsilon = 0.5k_BT$. Previous theoretical studies have reported that the R_g of the flexible polymer decreases monotonically with the increase of the volume fraction of crowders. [144, 145].

The attraction between the monomers leads to the further compaction of polymer globule in addition to the compaction by CLs. Hence, we expect the number density $n_M(r)$ of monomers to increase in the innermost part as we increase the value of parameter ε. But if
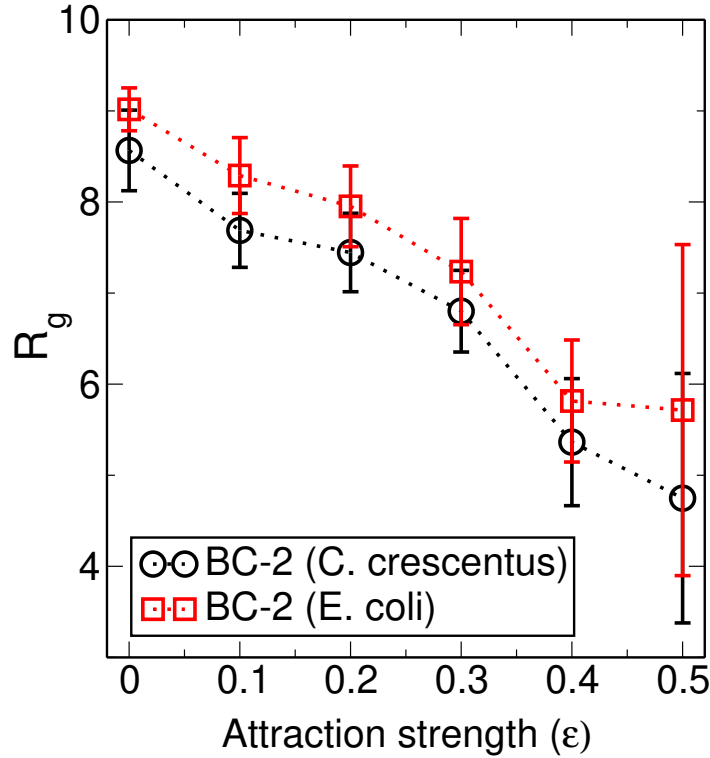
Fig. 3.8 The plot shows the average value of radius of gyration with the parameter ε for polymer with BC-2 CL-set of bacteria *C. crescentus* and *E. coli*, respectively. The SD is shown as error bars. The $\varepsilon = 0$ refers to the case where we cut off LJ at $r_c = 2^{1/6}\sigma$, such that only repulsive forces act between non-neighboring spheres.

the packing is too high for relatively large values ε, then the monomers near the center of the globule will be unable to explore the configuration space effectively. This could lead to the polymer getting kinetically stuck. This will also result in very different coarse-grained positional correlation contact maps in different runs. On the other hand, a weak attraction could help get monomers together and increase the positional correlation between different segments of the chromosome. We want to check if there is an optimum value of ε, which helps the polymer to get compact, but at the same time, the polymer should be able to explore the configuration space and reach its organized state.

To start, we plot the number density of monomers with distance from the CM of the globule in Fig. 3.9. In Fig. 3.9 different lines correspond to the different values of parameter ε varying from $0, 0.1, 0.2.., 0.5 k_B T$. From the figure, we see that as we increase the value of parameter ε, the number density of monomers increases by order of magnitude in the inner core region for $\varepsilon = 0.4, 0.5 k_B T$ compared to when $\varepsilon = 0, 0.1 k_B T$. But it decays very rapidly with the distance r from the CM of the globule. This suggests that the inner core is quite dense with monomers for large values of the parameter ε and the monomers in the innermost

part of the globule should be less mobile because of the less phase space availability. With this understanding of polymer structure with parameter ε, we next investigate the effect of the parameter ε on the internal organization of the polymer.
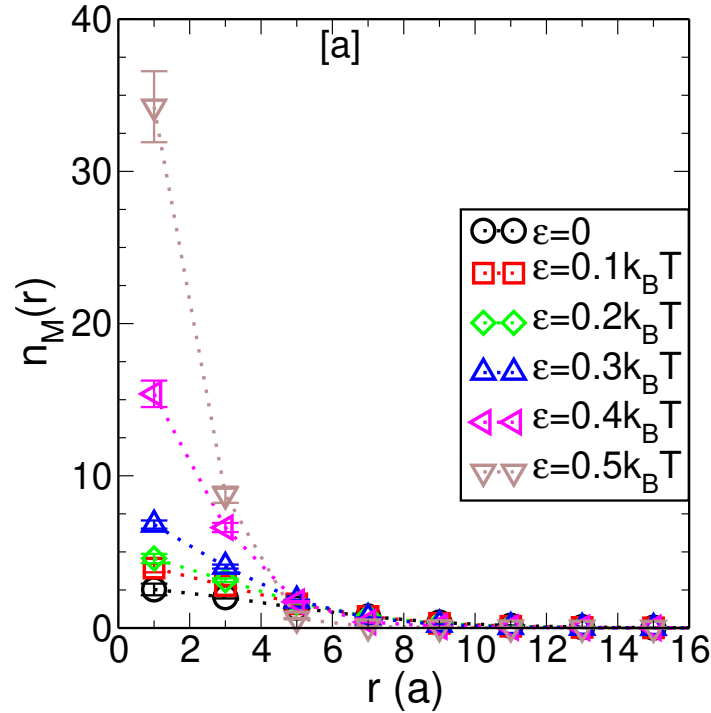

Fig. 3.9 The figure shows the number density of monomers as a function of distance from the center of mass of the polymer globule. The error bar represent the s.d. from the average value across 9 independent initial conditions.

For this, we calculate the positional correlation of different segments of the polymer, as was done in the previous case. We calculate the positional correlation again by choosing a cut-off R_c such that the value of $R_c/R_g = 0.55$ remains constant and the same as before. Then we calculate the probability of the CM of the two segments to be within a distance of R_c. This is an 80×80 matrix shown as colormaps in Fig. 3.10 for three different values of ε, where the x-axis and y-axis represent the segment index, and color represents the probability. Bright color corresponds to the higher probability of the CM of two segments being within distance R_c, and the light color represents the lesser value of probability. From Fig. 3.10 we notice that as we increase the value of the parameter ε, the patch-like pattern in the colormaps becomes more prominent and clear. But after a certain value of the parameter ε, the whole colormap becomes nearly uniformly brighter. This is because the polymer collapses for higher values of ε, and each segment is near many other segments with higher probability. The rectangular and square patch-like pattern in the colormap represent that the neighboring segments are coming close to other segments with equal probability, e.g., if segments i and j

come within cut-off distance with probability p_{ij} then neighboring segments of i and j, i.e.,
$i+1, i+2$ or $j+1, j+2$ are also coming within cut-off distance R_c with frequency $\approx p_c$ thus
giving the pixels in a rectangular patch of the same color. We do not show the colormaps for
the model chromosome of bacteria *E. coli*, but the outcome and conclusions are qualitatively
similar to that from Fig. 3.10.

If the polymer has a unique organization, then the MC runs starting from independent
initial conditions should give statistically similar positional correlation colormaps. To quan-
tify this, we calculate the average of Pearson correlation $\langle pc \rangle$ of the positional correlations
among different runs as was done for the study of polymer organization with various values
of the parameter σ. If the average Pearson correlation has a higher value, then we can say that
all independent conditions are leading to the same organization of the polymer. This is shown
in Fig. 3.11 where y-axis and x-axis represent the value $\langle pc \rangle$ and parameter ε, respectively.
From the figure, we see that the value of $\langle pc \rangle$ increases slightly as we increase the value of
the parameter ε till 0.3, and after that, it decreases. This can be because for the larger value
of parameter ε polymer from different initial conditions may be kinetically stuck in different
states and are not able to reach the same structure. This can also be confirmed from the plots
of number density, where for the high value of ε, the number density in the inner core is
very high. Moreover, we have also tested that after the equilibration runs of 10^6 MCS, the
average distance between the pair of monomers, which constitute a CL, remains a in all the
independent runs. This suggests that the CL monomers are near each other in all independent
runs, but other segments of the polymer are not able to organize themselves into a particular
structure across different runs starting from independent initial conditions. The plausible
reason for this can be the relatively strong attraction between the monomers, because of the
strong attraction between the monomers some segments (especially the segments in the inner
core) of the polymer are not able to explore the configuration space, thus giving the different
positional correlation colormaps for independent runs. The conclusions from the model
chromosome of bacteria *E. coli* are similar and can be confirmed from Fig. 3.11, where we
observe that the value of $\langle pc \rangle$ decreases after $\varepsilon = 0.3 k_B T$.

We also quantify the differences in the colormaps of positional correlation using the same
quantity $\langle f \rangle$, which gives the number of pixels with probability $p > 0.05$ the same as the
previous case. This quantity $\langle f \rangle$ is plotted in Fig. 3.12 with parameter ε. The average is
taken over the colormaps from 9 independent initial conditions. From the graph, we see that
the increase in the value of $\langle f \rangle$ for lower value of parameter ε is not significant but from
$\varepsilon = 0.3$ to $\varepsilon = 0.4$ it increases from 0.7 to 0.9 and for $\varepsilon = 0.5$ the value of $\langle f \rangle$ becomes
nearly 1. The value of $\langle f \rangle = 1$ signifies that all the segments are coming closer to other
segments. This can also be confirmed from the bottom colormap of positional correlation
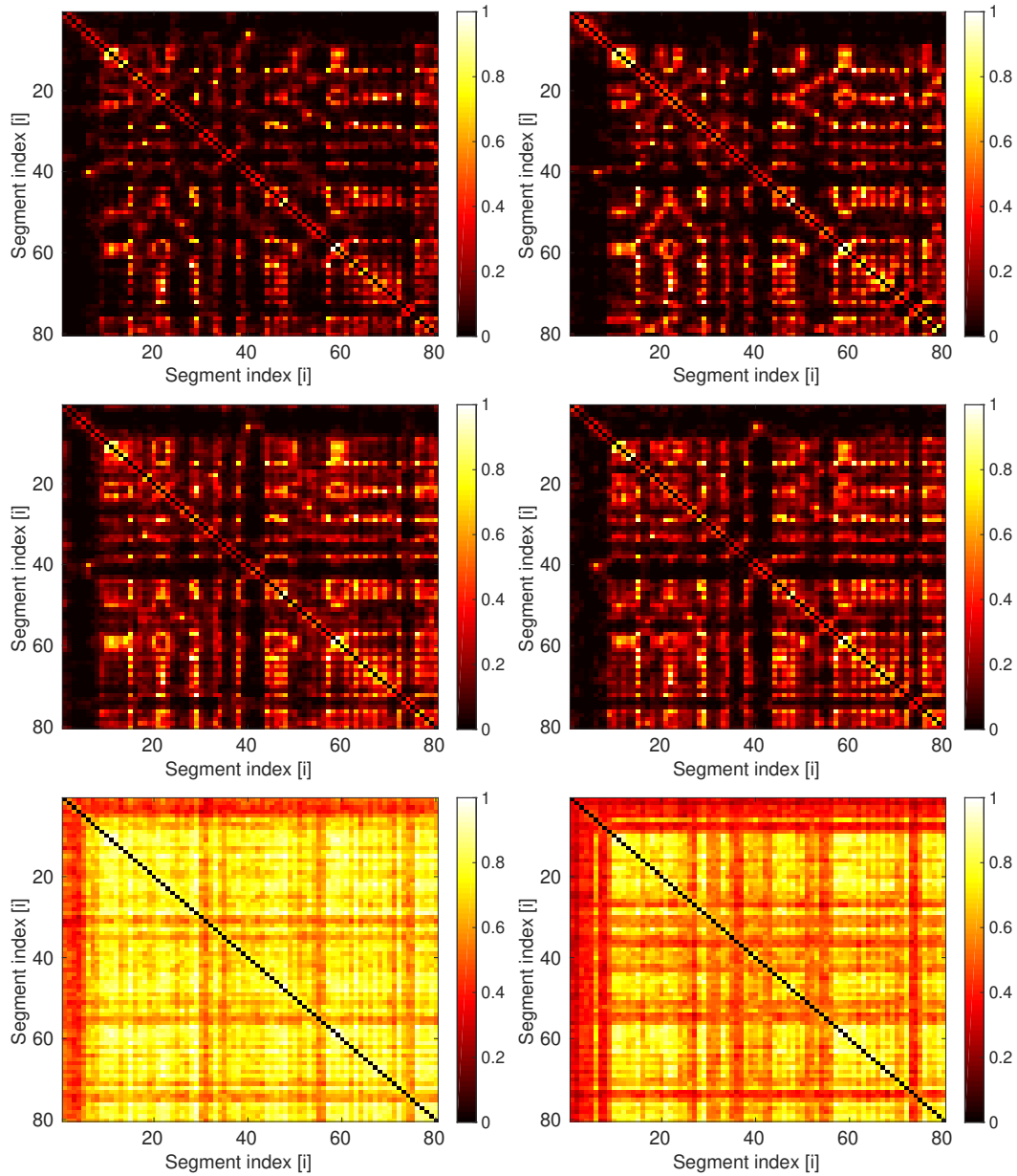
Fig. 3.10 The colormaps show the positional correlation between the center of mass (CM) of different segments of the polymer for different values of parameter ε. There are 80 segments, each with 50 monomers. The top, middle and bottom figures correspond to the $\varepsilon = 0.1, 0.3, 0.4 k_B T$, respectively. The colormaps in each row are from two independent runs starting from two different initial conditions.

in Fig. 3.10, which consists of all the bright pixels. We obtain the same conclusions for the colormaps of the polymer with BC-2 set of CLs of *E. coli*, see Fig. 3.12.

Fig. 3.11 The plot shows the average value of Pearson correlation for different values of parameter ε for the model chromosome of bacteria *C. crescentus* and *E. coli*. The SD from the average value is denoted by the error bars.
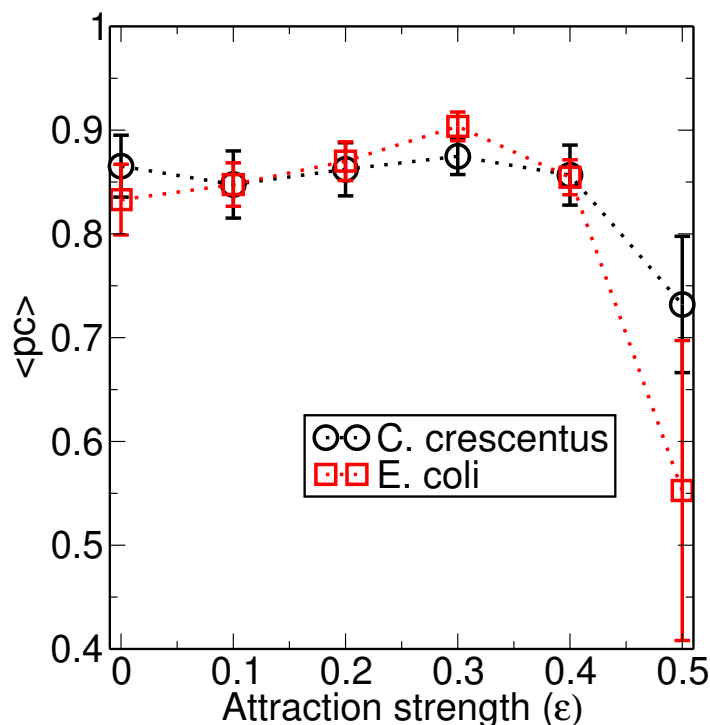
So from these results, we can say that there is an optimum value of attraction strength between the monomers, which helps the polymer to organize. Since for smaller attraction strengths ε, the polymer will be able to explore configuration space and can obtain a particular organization, but for the very high value of attraction strengths, the polymer can get kinetically stuck into different states and would not be able to explore the configuration space.

3.4 Discussions

We study the effect of the release of topological constraints and molecular crowders in the organization of the polymer with very few specific permanent cross-links ($\approx 3\%$ of the monomers) taken from the experimental contact map of bacteria *E. coli* and *C. crescentus*. We showed that the release of topological constraints is crucial for the polymer to organize into a particular structure since the model chromosomes from 9 independent runs are able to organize into a particular structure when we allow the chains to cross itself and release the topological constraints. We get a relatively higher value of Pearson correlations of at least $\langle pc \rangle \approx 0.7$ (and up to 0.85 across runs starting from independent initial conditions, for
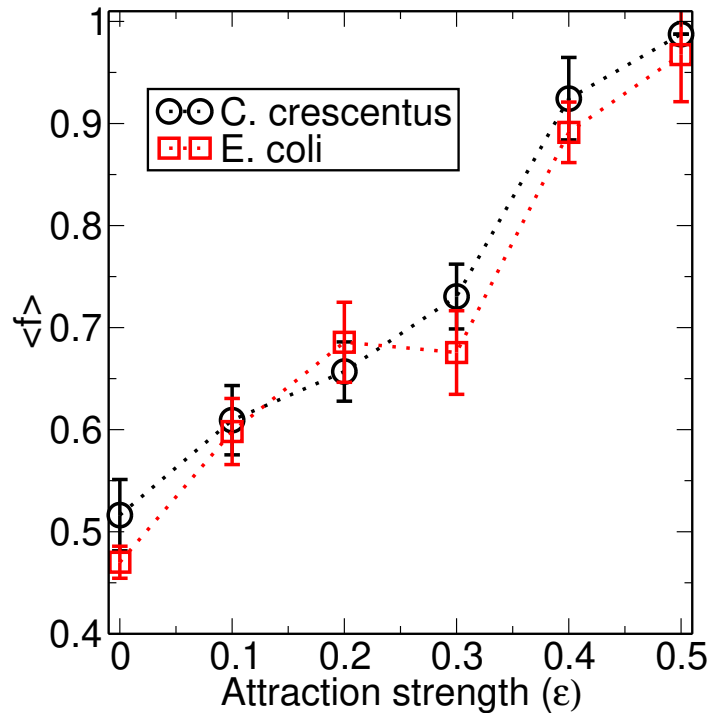
Fig. 3.12 The plot shows the quantity $< f >$ with the parameter ε.

the cases in which chain crossing is allowed compared to the cases when the chain crossing is not allowed. This leads us to believe that the activity of enzyme topoisomerase can play a vital role in the organization of the chromosome at large-length scales by allowing the chains to cross itself. Furthermore, an increase in the diameter of the monomers leads to an overall swelling of the DNA-polymer coil and the radius of gyration (R_g) scales as $R_g \sim \sigma^{0.5}$ for the ring polymers with BC-2 CL-sets corresponding to the bacteria *C. crescentus* and *E. coli*. The same behavior is also seen for the random set of CLs (RC-2), which had the same number of CLs and coil length as that of the model polymer of *C. crecentus*. We also observe that the positions of the CLs along the chain do not play any role in the swelling behavior of the polymer.

We also show that optimum compaction due to molecular crowders in the bacterial cytoplasm of the cell is possible so that the DNA polymer is able to explore the different configurations as well as should be able to organize better into a particular structure. We find the values of effective weak attraction between the monomers to be $\varepsilon = 0.1, 0.2, 0.3 k_B T$ for which the polymer from different initial conditions is able to obtain a unique organization.

We wanted to study the role of individual factors in the organization of the bacterial chromosome, thus we do not introduce the effect of cell wall confinement in the present chapter but, we study the effect of confinement with the effect of these factors by taking the

optimum values of the parameter from this study and present it in the next chapter. There we also compare our predicted organization with experimental contact maps to establish that the factors we identify are truly relevant for bacterial DNA organization. We emphasize that though we start from the contact map of the bacterial chromosomes, we actually use a very small fraction of the information present in contact-maps to identify the position of a minimal number of monomers, which are cross-linked permanently. We then try to predict the polymer's overall 3-D organization.

Chapter 4

Role of confinement, special cross-links and molecular crowders in the organization of bacterial DNA polymer

4.1 Introduction

In the previous chapter, we study the role of the release of topological constraints and the molecular crowders in the organization of the bacterial DNA polymer in the presence of few specific cross-links taken from the Hi-C contact map. Till now, we did not incorporate the effect of the confinement of the cell wall in our studies of the chromosome organization. However, it is shown in various previous studies that confinement of different topology (sphere, cylinder, nanochannels, etc.) plays a crucial role in the dynamics and organization of the polymer.

In this chapter, we will present our model and study of the role of confinement of the cell wall in the organization of the bacterial DNA polymer with few cross-links. In the end, we also introduce the effect of molecular crowders in the model chromosome and study the combined effect of specific cross-links, confinement, and molecular crowders in the organization of the DNA polymer.

4.2 Model

The interaction between the adjacent monomers along the chain and between the cross-linked monomers is modeled by harmonic potential with spring constant $\kappa = 200 k_B T / a^2$. The excluded volume interaction between the neighboring monomers (excluding the nearest

neighbors along the chain) is modeled by the WCA potential V described in chapter 2 of the thesis. The value of the parameter ε has been taken as $1k_BT$. We set the value of the parameter σ to be $0.2a$ from the systematic study in chapter 3. We next incorporate the effect of cylindrical confinement of the bacterial cell. The shape of the bacterial cell is a capsule (spherocylinder). We approximate this shape by a cylinder with planar ends for the sake of simplicity with an aspect ratio (diameter: length) $\approx 1 : 7.5$. We intentionally chose a longer aspect ratio than the expected $\approx 1 : 5$ to check whether the final organization ends up in a configuration with a ratio closer to $\approx 1 : 5$. To fix the degree and shape of confinement in our model, we take the help of the previous study. It has been suggested previously [8] that the chromosome of bacteria *C. crescentus* could have a bottle brush structure with plectonemes (super-coiled segments of the DNA) emanating out in different directions from a central backbone. It is strongly believed that plectonemes play a significant role in DNA organization at large length scales [8, 131]. Though we do not explicitly consider plectonemes in our model, the CLs that we use from the experimental contact maps may incorporate the effect of plectonemes. This is because the DNA segments within plectonemes, which are in spatial proximity in the 3D space, will be reflected in the experimental contact maps. Also, it is known from the previous experimental studies that the average length of the plectoneme segment is ≈ 10-15 kilo-BP [8], which is up to 15 monomers in our model. Since a plectoneme is coiled on itself, the maximum spatial length of a plectoneme can be $\approx 7a$ in our model units. We take $7a$ to be the radius of the cylinder, and the diameter $D = 14a$ allows two plectonemes to be radially opposite to each other in the cylinder. The length of our confining cylinder was then fixed at $108a$ to maintain the given aspect ratio.

In the studies of chapter 2 and chapter 3 of the thesis, we started our Monte-Carlo simulations from specially designed 9 independent initial conditions. But in the presence of confinement, we cannot start the simulation with the same initial conditions as described in chapter 2, as those initial configurations will violate the constraint of the confinement. Further, if we start our simulation with the random initial configuration of the ring polymer and with the BC-2 set of CLs with 153/159 CLs, the polymer forms a blob at the center of the cylinder, and the polymer will be unable to relax. Hence, to systematically investigate the role of confinement, we first start with a ring polymer (without CLs) in the cylinder and then observe how entropy makes the polymer spread out. We start from 12 different initial conditions, which we describe later in this chapter. Further, it is known for bacteria *C. crescentus* that the DNA segment where the replication starts (origin of replication termed as *ori*) is attached to one end of the cylinder; while for the bacteria *E. coli* there is no such tethering of the *ori* to the cylinder's boundary. We incorporate this constraint for the model chromosome of *C. crescentus* at this stage by attaching the monomer, which we have given

the index 1, to one pole of the cylinder. The positions of the first monomer remain unchanged throughout the simulation. We first observe the organization of the ring polymer (without CLs), which is confined within a cylinder with the first monomer tethered at one pole of the cylinder.

Next, we introduce the CLs at specific positions to this spread-out configuration of the polymer, allowing the system to relax and reach equilibrium and carry out further analysis at the second stage of our computations. The effect of CLs are added in the model by introducing a quadratic spring potential between a specific pair of monomers with very low strength, and then slowly increase the strength as we allow the system to relax. At the end of this step, even if two cross-linked monomers are far away in space in the initial starting configuration, they come close to each other at the end of equilibration and, thereafter, maintain a distance of a. The cross-linked monomers drag the adjacent segments along with them. However, confinement already restricts the possible polymer configurations in the 3D space. Hence instead of 153/159 CLs (BC-2 CL-set in [137]) for bacteria *C. crescentus* and *E. coli*, we just add 60/70 and 49/47 number of CLs at specific locations along the chain; we refer these as $BC^{'}$ and BC-1 set of CLs. The list of CLs corresponding to the $BC^{'}$ CL set for the chromosome of *C. crescentus* and *E. coli* are given in the table C.3 in Appendix B of the thesis. Note that the minimal number of CLs which were required to obtain organization in the DNA ring polymers without confinement was 159 and 153 CLs for *E. coli* and *C. crescentus*, respectively [137, 136]. These correspond to 82 and 60 *effective* CLs, respectively, and we refer to that set as BC-2 set of CLs.

The details of initial conditions, equilibration protocol, and the results corresponding to this model will be described later in this chapter.

We next study the polymer organization with $BC^{'}$ set of CLs in the presence of molecular crowders and confinement. For this, we introduce a weak attraction between the monomers in addition to the cylindrical confinement and the $BC^{'}$ sets of cross-links. To introduce the weak attraction between the monomers, we model the excluded volume interaction by the LJ potential $V^{'}$, which is cut at 3σ and suitably shifted (see chapter 3). We chose the values of parameters σ and ε of the potential $V^{'}$ to be $\sigma = 0.2a$ and $\varepsilon = 0.3k_BT$ from the study of chapter 3 of this thesis. Again, We will start our simulation from the 12 independent initial conditions of the polymer and then compare the organization of the polymer across 12 runs.

4.3   Results

We start by reporting the effect of cylindrical confinement in the organization of DNA ring polymer of the bacteria *C. crescentus*. Before presenting our results, we give a brief

description of the generation of initial conditions. First, we keep all the monomers in two different arrangements inside a cylinder of length $108a$ and diameter $14a$, as shown in Fig. 4.1. The two neighboring monomers along the chain contour remain at a distance a from each other in the two arrangements. The harmonic-spring interaction between the CL monomers is kept switched off. We keep the position of the first monomer of the chain fixed at one end of the cylinder. The last monomer is not placed adjacent to the first monomer in both the cases, though for a ring polymer, the first and last monomers should be at a distance of a.

Next, we evolve the system using Monte-Carlo simulations for 10^6 MCS keeping only repulsive part of the Lennard Jones potential (WCA) between all the monomers. To equilibrate the system, initially, we take the spring constant between the first and the last monomer $\kappa' = 0.2k_BT/a^2$ but keep the value of spring constant of the other springs (connecting the nearest neighbors along the contour) fixed at $200k_BT/a^2$. We increase κ' by $0.2k_BT/a^2$ after every 1000 MCS. After 10^6 MCS, κ' will have the value of $200k_BT/a^2$ similar to other spring constants. After this, we generate 12 independent initial conditions over the next 5×10^6 iterations. To generate 12 different initial conditions, we do the following. Starting from the configuration at the end of 10^6 MCS, we again equilibrate the system for a further 5×10^6 MCS with 6 different random number seeds for each of the two configurations. After 5×10^6 MCS, we get 12 independent configurations of the ring polymer inside the cylinder, which we use as the initial conditions for the next set of runs from which we calculate statistically averaged quantities to determine the organization of the polymer. We use these 12 initial configurations as the starting points for our studies of a confined ring polymer for 3 cases: (a) ring polymer without CLs (b) ring polymer with CLs (c) ring polymer with CLs and weak attraction. Also, the first monomer (Ori) is tethered at the boundary of the cylinder throughout the simulations and in all the independent runs for the model chromosome of bacteria *C. crescentus*. We start our MC simulations from these independent initial conditions and compare ensemble-averaged statistical quantities from independent runs to draw conclusions.

To ensure that the configurations we obtain after 5×10^6 MCS, are independent of each other we calculate the longest relaxation time of the polymer [128] by calculating the autocorrelation function of the z-component of the vectors connecting monomer pairs numbered 1 and 2008, 1004 and 3012, 2510 and 502, 1506 and 2514 in 4017 monomers chain in the presence of cylindrical confinement and with $\sigma = 0.2a$. The axis of the cylinder lies along the z-axis. We chose the vectors between the monomers which are at a maximum distance from each other along the chain contour. The autocorrelation function is calculated using the formula:

$$C_{ij}(w-w_o) = \frac{\langle (z_{ij}(w) - \langle z_{ij} \rangle)(z_{ij}(w_o) - \langle z_{ij} \rangle) \rangle}{\langle (z_{ij}(w) - \langle z_{ij} \rangle) \rangle^2},$$
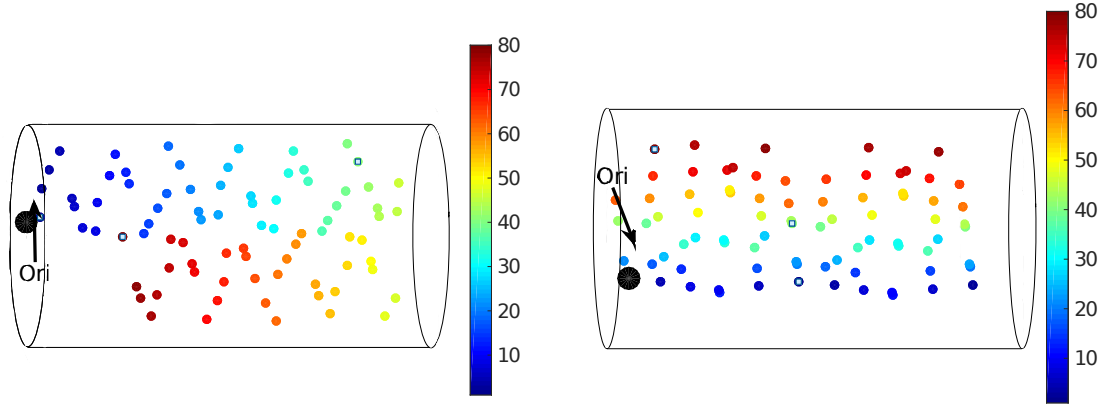
Fig. 4.1 The figure shows the 2 different initial arrangements of the polymer inside the cylinder. The monomers are represented by the color according to their index in the contour from blue to red. For better visualization we only plot the positions of every 50th monomer. Thus the index 80 in the colorbar corresponds to the monomer numbered 4000. The first monomer of the chain is colored black and marked as ori.

where z_{ij} is the longitudinal component of the vector joining i^{th} and j^{th} monomer. The average $\langle .. \rangle$ is taken over 5×10^6 MCS. Then we calculate the average value of the autocorrelation function of all the 4 vectors. The function $C_{ij}(w - w_o)$ is plotted in the Fig. 4.2 versus $(w - w_o)$. Each $\Delta w = 1$ corresponds to 100 MCS, we collect data after every 100 MCS. We see from the graph that the correlation function of the longitudinal component of vector decays exponentially, and around $w - w_o = 2000$ (i.e. 2×10^5 MCS) the value of $C_{ij}(w - w_o)$ is $1/e$ of its initial value. Hence, we can expect the configurations we get after 5×10^6 MCS are independent of each other.

To obtain statistical averages, we further evolve the confined polymer without CLs for 29×10^6 MCS using Monte-Carlo simulations, but start collecting the data to calculate the statistical quantities only after the first 5×10^6 MCS. Thus the data presented in Fig.4.3 and other data is collected over 24×10^6 MCS for each initial condition, where the data is sampled after every 2×10^5 MCS (since the polymer relaxation time is 2×10^5 MCS).

Starting from the above mentioned 12 independent initial configurations, we also introduce CLs to study the polymer configurations with 49 CLs (BC-1 CL-set) and 60 CLs (BC' CL set). In the 12 different initial conditions generated, the positions of the monomers which constitute the CLs can be greater than the bond-length a. Thus, we use the same strategy as mentioned previously to relax the system. Firstly, we take a small value of spring constant $\kappa_c = 0.2 k_B T / a^2$ between the CL monomers and increase the value by $0.2 k_B T / a^2$ in every 1000 MCS till it reaches the $\kappa_c = 200 k_B T / a^2$. This will lead to the cross-linked monomers to come near each other at the end of the equilibration, without affecting the stability of the
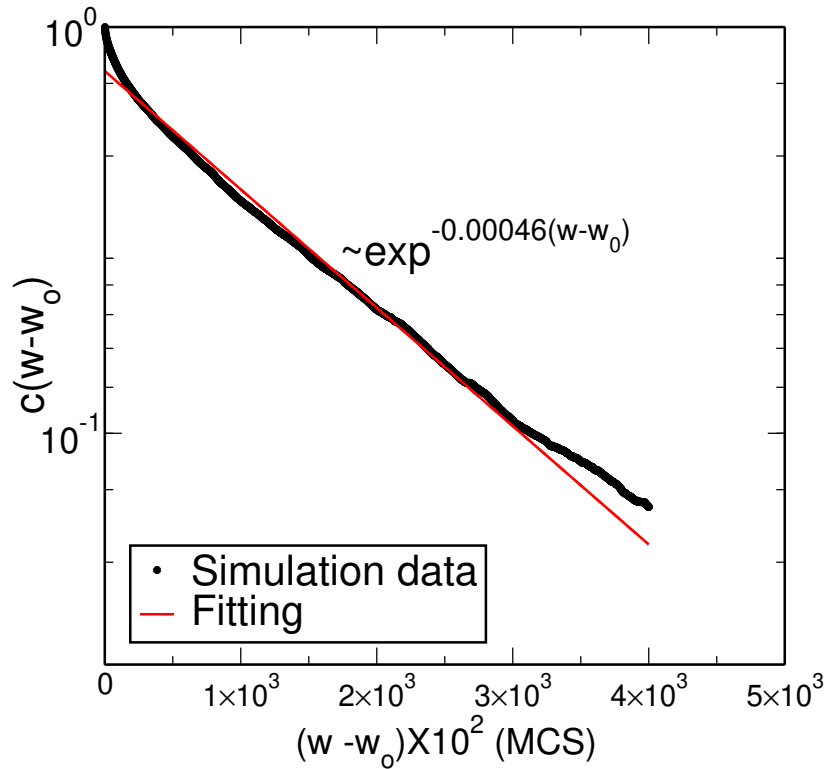
Fig. 4.2 The plot shows the average correlation function of the longitudinal component of the vector joining monomers numbered 1 & 2008, 1004 & 3012, 3012 & 502 and 1506 & 2514 for a polymer with $w - w_0$ in a semi-log plot. Each w represents 100 MCS. The red line corresponds to the exponential fitting to the correlation function from the simulation.

computation. As before, we evolve the system for 5×10^6 MCS for equilibration and then start collecting data for the next 2.4×10^7 MCS to calculate the statistical quantities across 12 independent runs.

Firstly, we estimate the spread of the polymer inside the cylindrical confinement with and without CLs. For this, we calculate the moment of inertia tensor of the polymer chain from its center of mass. Then we diagonalize the tensor to get the principal moments I_1, I_2, I_3, where $I_1 > I_2 > I_3$. The ratio of the principal moments I_1 and I_3 gives the idea about the asymmetry of the polymer chain. Inside a cylinder, the chain is extended along the length of the cylinder and will be squeezed in the other directions.

If the ring polymer is tethered at one end without any CLs, the polymer should extend along the length of the cylinder to maximize the entropy and the value of the ratio I_1/I_3 is expected to be very large. But in the presence of the intra-chain cross-links, the polymer should be less extended along the cylinder-axis because more constraints (CLs) lead to the compaction of the polymer. Thus, to get the estimate of the asymmetry and compaction of ring polymer along the direction of the cylinder's length, we calculate the ratio I_1/I_3 for the

three cases a) Ring polymer with no cross-links. b) polymer with 47 CLs and c) polymer with 60 CLs. The average values are given in table 4.1, where the average has been taken over the value obtained from 12 independent initial conditions of the polymer. From the table, we see that with the increase in the number of CLs, the value of I_1/I_3 decreases leading to the compaction of the polymer along the longitudinal direction of the cylinder, as expected. We also calculate the asphericity, $A_s = 0.5 \left[\dfrac{3(I_1^2 + I_2^2 + I_3^2)}{(I_1 + I_2 + I_3)^2} - 1 \right]$. The average values of A_s for three cases are given in table 4.1. For a perfect sphere(rod) the value of $A_s = 0(1)$. From the table, we see that as we increase the no. of CLs, the value of A_s decreases. This is because of the CLs compacts the polymer extension in the longitudinal direction. But the values of A_s and I_1/I_3 does not change significantly, as we increase the no. of CLs from BC-1 to BC'.

No. of CLs	No. of *effective* CLs	I_1/I_3	A_s
1 (ring)	1	50.38	0.23
49	26	14.50	0.20
60	33	13.89	0.19

Table 4.1 The table shows the average value of the moment of inertia ratio I_1/I_3 and the asphericity for a polymer with a different number of CLs. The average is taken over the configurations of the polymer starting from 12 initial conditions.

Further, to understand the distribution of the monomers inside the cylinder, we calculate the number density of all monomers in radial and longitudinal directions of the cylinder for the polymer with a different number of CLs. The average values of densities are shown in Fig. 4.3(a) and (b), where the average is taken over the value obtained from the 12 independent initial conditions, and the error bars show the s.d. from the average value. Different lines in the plot represent the number density of the polymer with a different number of CLs. From the plot, we see that the number density of monomers in the radial direction does not change with the increase in the number of CLs, which suggests that the increase in the number of CLs is not helping in the compaction of the polymer in the radial direction. But we see that the number density of monomers significantly changes in the longitudinal direction as we change the number of CLs. In the absence of any CLs, the polymer extends with an approximately uniform density of monomers along the length of the cylinder. But as we increase the number of CLs to 49, the density at the center increases and gives a high value at the center of the cylinder. Further increasing the number of CLs from 49 to 60, there is a slight increase in the height of the peak. This suggests that the increase in the number of CLs helps in the compaction of the polymer along the longitudinal direction while the
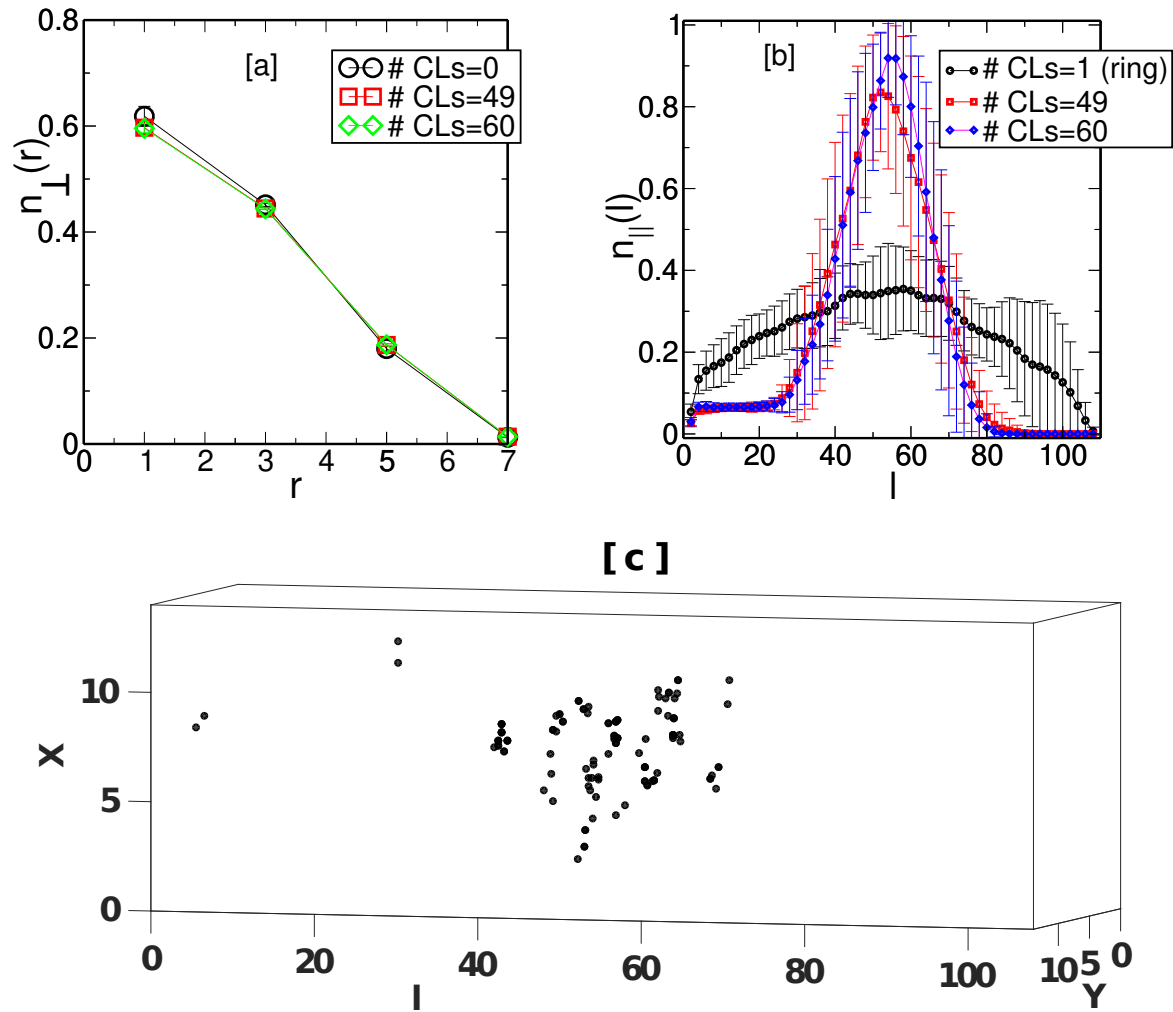
Fig. 4.3 The figures (a) and (b) show the number density of the monomers in the radial and longitudinal directions with the distance. Different lines correspond to the polymers with a different number of CLs. Error bars show the s.d. from the average value across 12 independent initial conditions. To get the radial number density, we calculate the number of monomers present between the concentric cylindrical shell ($dr = 2a$) from the axis of the cylinder and divide it by the volume of the shell. The longitudinal number density represents the number of monomers within sections of the cylinder with radii $7a$ and length $2a$ divided by the total volume of the cylinder section. (c) Snapshot from the simulation shows the position of the monomers, which constitute the CLs. The bounding box is given to read the positions of the CLs easily and does not represent the confinement geometry of the polymer, which is a cylinder.

radial directions remain unaffected. There are almost no monomers near the wall, and all monomers are within a 10a spread along the radial axis of the cylinder.
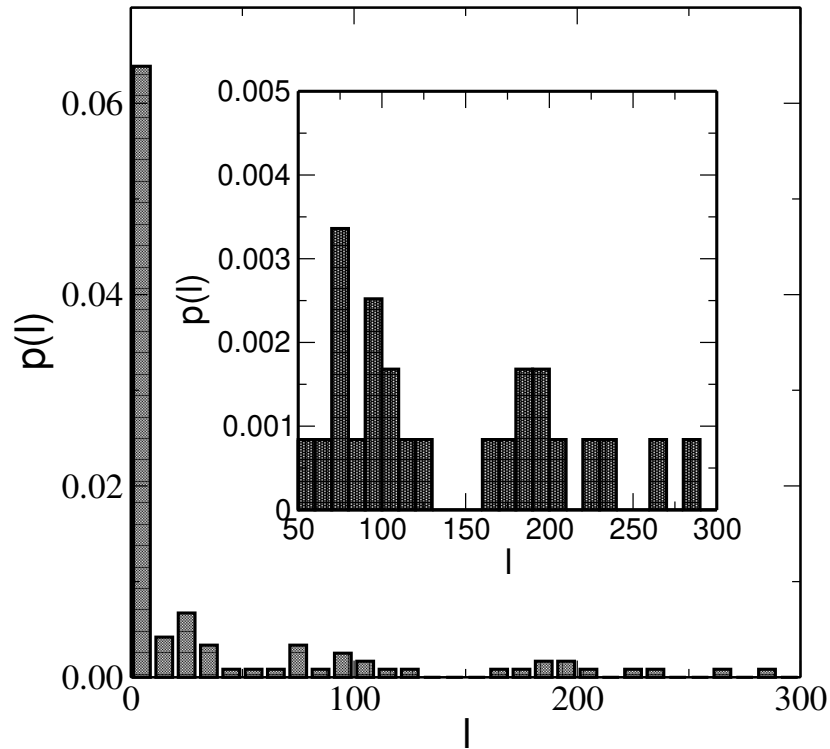
Fig. 4.4 The figure shows the distribution $p(l)$ of segments such that the segment has l monomers between two neighboring cross-links along the contour. The data is for BC$'$ CL-set with 60 CLs. The probability density function $p(l) = N(l)/(N*dl)$. Here, N(l) represents the number of segments of length l (the bin size $dl = 10$ monomers has been chosen) and N represents the total number of segments between the CLs; $N = 60 * 2 - 1 = 119$. From the plot, we see a peak at $l = 1 - 10$. This is because we implement all the 60 CLs instead of 33 *effective* CLs in our calculations. Whenever neighbouring monomers i and $i+1$ or i and $i+2$ are cross-linked to other neighbours j and $j+1$ or j and $j+2$ as per our criteria of choosing position of CLs, these segments gives a peak at $p(l = 1 - 10)$. The inset shows the zoomed in version of the $p(l)$ of the segments with lengths (l) from 50 to 300.

To have a better understanding of why the radial distribution of monomers does not shrink with the introduction of CLs, we note in Fig.4.4 that there are a significant number of segments with more than 100 monomers along the chain contour without an intervening CL. There are ≈ 15 segments with $l > 100$ monomers, such that a part of these segments can be thought to consist of one (or more) polymer blob(s) of diameter 10a which can span the diameter of the cylinder. If we use blob diameter $\zeta \approx al^{3/5}$ scaling for a polymer chain with excluded volume interactions, the number of real-chain blobs which are of size 10a are even more. A real-chain blob with 60 monomers has a diameter $\zeta \approx 11a$; thus, a polymer segment with 200 monomers can have three real-chain blobs each spanning the diameter of the cylinder. This formation of a large number of blobs, even in the presence of CLs, is a

plausible reason why the radial distribution of monomers remains relatively unchanged even on the introduction of CLs.

We can also see that the longitudinal number density from $l = 0$ to $l = 26$ is very low, and the error bars are relatively smaller for the polymer with CLs. To investigate why this is the case, we plot only the positions of the CL monomers from the simulation, which is shown in Fig. 4.3(c). From the snapshot, we see that most of the CLs are clustered around $l = 50$, thus giving a very high value of density at $l = 50$. The number of monomers that constitute the CLs is relatively less in the region $l = 0$-26, thus giving the value of number density and error bars to be less in that region.

With this information, now we want to understand how does the internal organization of the ring polymer change in the presence of cylindrical confinement with zero CLs, as well as with BC-1 and BC' set of CLs. For this, we calculate the positional correlation of different segments of the polymer with a different number of CLs. We can estimate the positional correlation as the probability of the CMs of the two segments (50 monomer in each segment) to be within cutoff R_c. A ring polymer with 4017 monomers will have 80 segments. We chose the value of cutoff R_c to be $5a$, which we estimate according to the value of polymer's R_g in the absence of confinement with BC-2 set of CLs, refer [138]. We choose R_c such that the value of $R_c/R_g = 0.55$. The positional correlations are shown in Fig. 4.5, 4.6 and 4.7 as four representative colormaps from four independent runs for different cases. In the colormap, the x- and y-axis represent the segment index, and the color signifies the calculated probability. Bright color denotes the higher probability, and the dark color corresponds to the lower probability.

To explain the data of positional correlation of the colormaps for polymer without CLs, we define two parts of the ring polymer as arm-1 and arm-2. The arm1 consists of monomers numbered 1 to 2008, and arm2 consists of monomer numbered 2009 to 4017. The colormap of Fig. 4.5 shows the positional correlation of the ring polymer under cylindrical confinement without CLs. In the colormap of Fig. 4.5 we see the two diagonals. The main diagonal (top-left of the figure to bottom-right) signifies that the segments which are neighbors along the contour of the chain are also spatially close to each other in the 3D space. For, e.g., segment number 40 is close to segment number $38, 39, 41, 42$, etc. These correspond to the intra-arm-interactions of the segments of arm-1. The presence of other diagonal means that the monomers of arm-1 and arm-2 are coming closer to each other. This suggests that the two arms are arranged nearly parallel to each other along the long axis of the cylinder. These diagonals are also present in the experimental contact map of the bacteria *C. crescentus* and suggest that the two arms of the chromosome are arranged parallel to each other in-vivo as was also reported in previous study [8]. But by comparing the simulation colormap with the
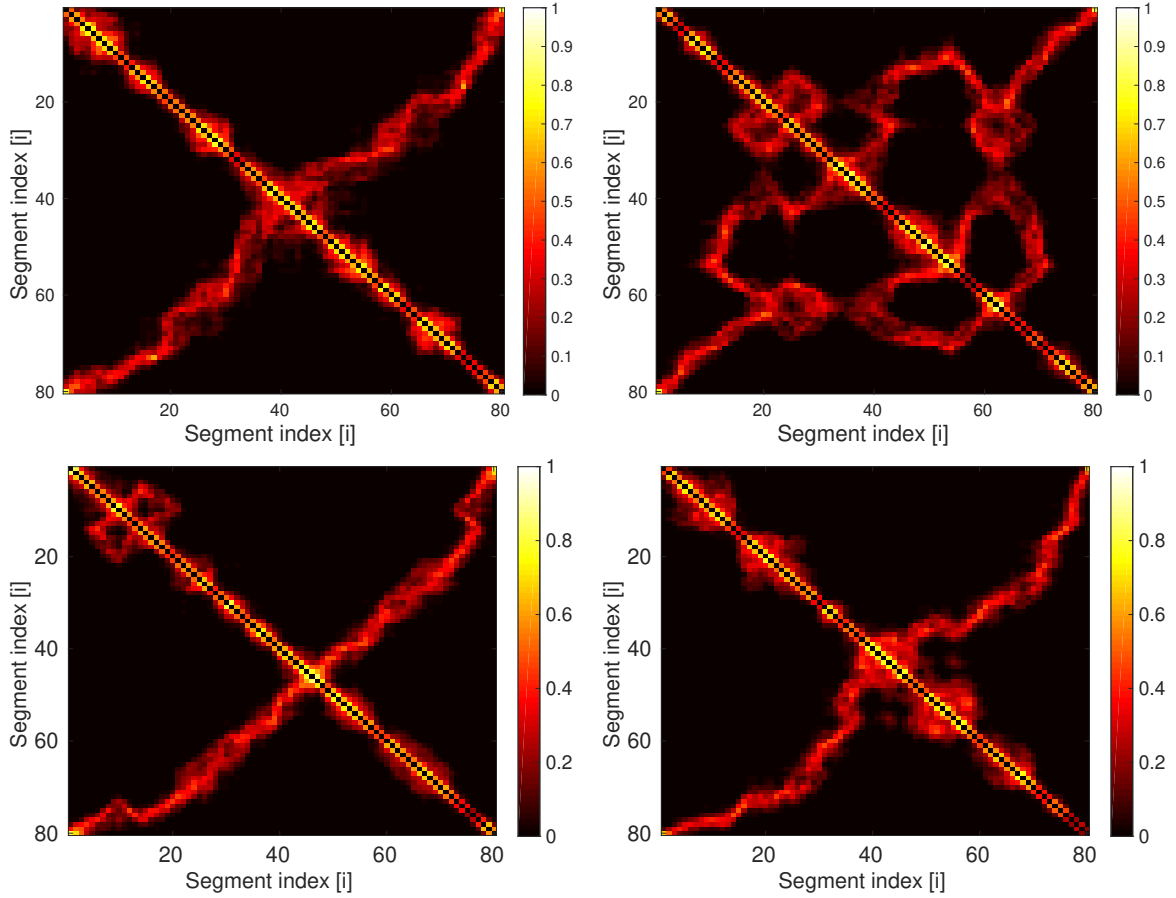
Fig. 4.5 The colormaps show the probability of the CMs of different segments of the polymer to be within a distance $R_c < 5a$ in the absence of any cross-links. The four colormaps are for the polymer starting with four different initial conditions, and the probability is an average, calculated over 120 independent snapshots over 2.4×10^7 MCS for each run.

experimental coarse-grained colormap of Fig.4.9 (e), we observe that the off-diagonal points which are present in the experimental contact map, are missing in the simulation colormap. Also, the colormaps from 12 independent conditions do not match with each other, as can be seen from the two representative colormaps of Fig.4.5. This suggests that the confinement of the cell helps in arranging the two arms of the polymer to arrange parallel to each other but does not organize the polymer completely into a particular structure as different initial conditions show the different positional correlation of the segments.

We next investigate how is cross-linking the specific pairs of monomers help in the organization of the polymer? Also, the positional correlations from our model with cross-links should be consistent with the experimental coarse-grained contact map, which has two diagonal as well as several off-diagonal bright pixels. We reiterate that we cross-link 49 (BC-1 CL set) and 60 (BC' CL-set) pair of specific monomers, which are 26 and 33 effective

CLs, respectively. Note that the number of cross-links is significantly less than the number of
CLs in BC-2 CL-set (153 CLs), which we used in our previous studies [137, 136] to obtain
the organization of DNA-polymer without confinement. The two representative colormaps
using data from two independent MC runs, starting from different initial conditions, are
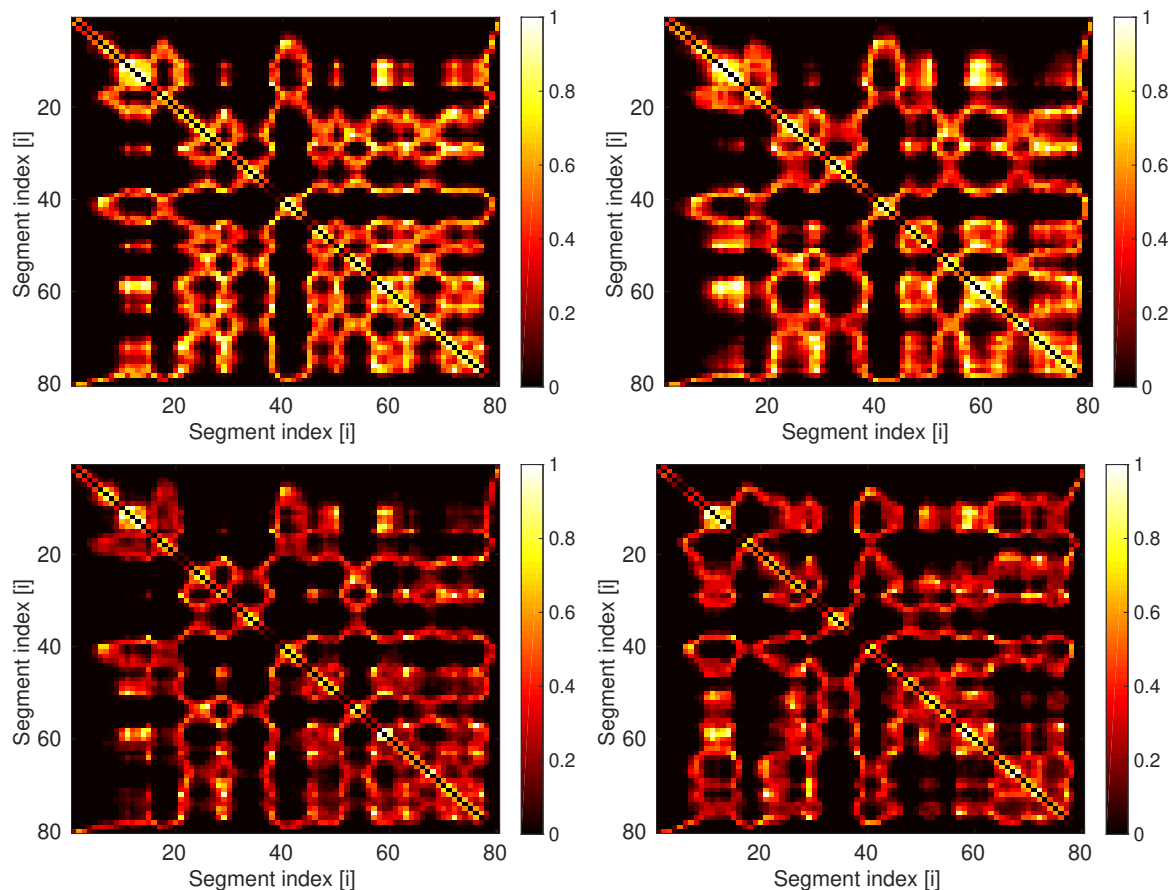shown in Fig. 4.6 and Fig. 4.7 for BC-1 and BC' CL sets, respectively.

Fig. 4.6 The colormaps show the probability $p(i, j)$ of the CMs of two different segments
i, j of the polymer with BC-1 set of CLs (49 CLs with 27 *effective* CLs) to be at a distance
$R_c < 5a$. The four colormaps are for the runs starting from different initial conditions.

From the figures, we see that in the colormap of the polymer with BC-1 CL set, we
get the off-diagonals pixels as well as the bright diagonal pixels, which are also present
in the experimental contact map [137]. We also quantify the number of bright pixels in
the diagonals, data for which is given later in this section. Also, the colormaps from the 2
independent initial conditions look statistically similar. We also check the colormaps from
12 initial conditions, and all the colormaps look statistically similar to the two in Fig. 4.6.
But on increasing the number of CLs from 49 to 60, we see that colormap does not change
significantly, though we, of course, get additional bright patches in the colormap. Hence,
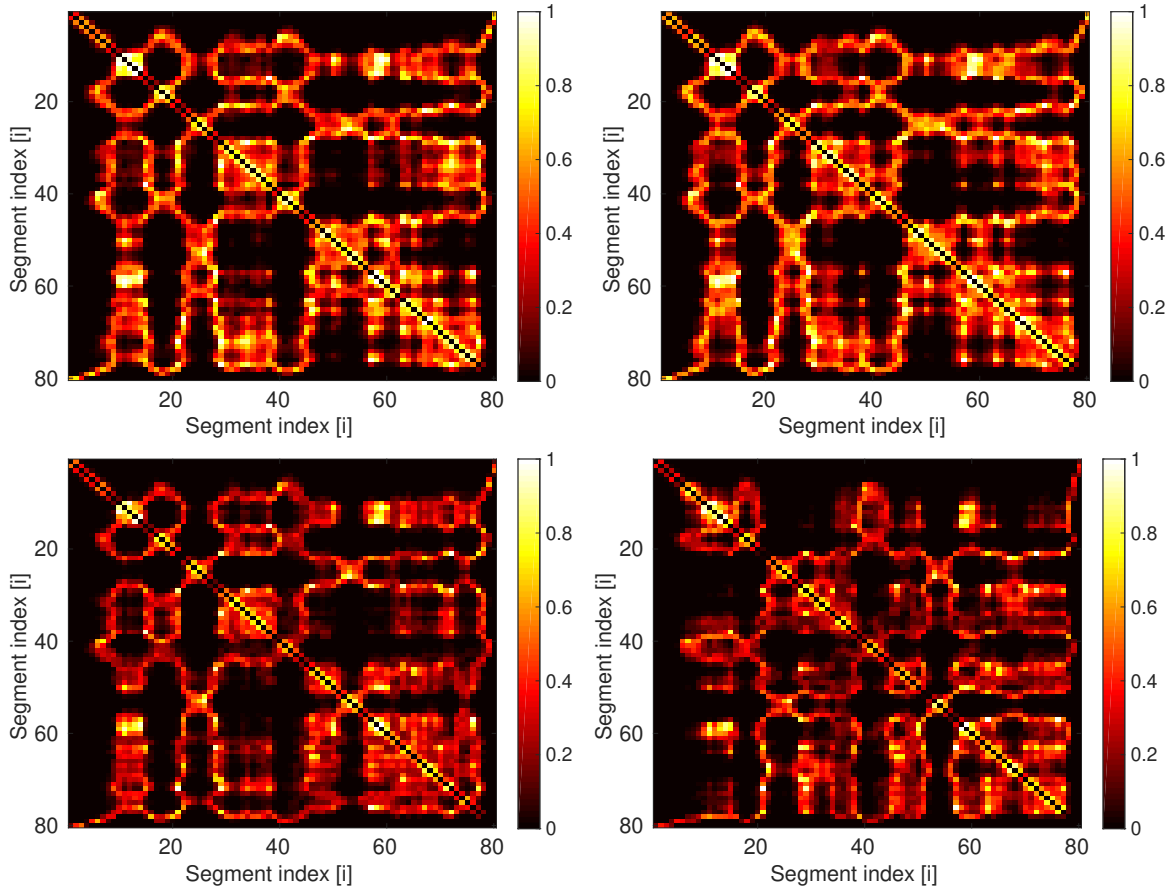
Fig. 4.7 The colormaps show the probability $p(i, j)$ of the CMs of two different segments i, j of the polymer with BC' set of CLs (60 CLs) to be within at $R_c < 5a$. The four colormaps are calculated from four independent MC runs, starting with four different initial configurations of the polymer.

we can confidently claim that 60 CLs (33 effective CLs) can organize the polymer into a particular structure. Also, note that in the absence of confinement the positional colormaps for the polymer with BC-1 set of CLs were predominantly black in the study of chapter 2.

Now we present our result for the polymer with BC' CL set, cylindrical confinement with a small attraction between the monomers using the LJ potential mentioned in the model section with parameter $\varepsilon = 0.3k_BT$. Starting from the 12 initial conditions of the ring polymers without CLs, we introduce the weak attraction at the same time as we introduce the CLs. Before starting to collect data to calculate the average of statistical quantities, we allow the polymer to relax using the same protocol as described before.

As mentioned before, the small attraction between the monomers incorporates the effect of compaction due to molecular crowding. The attraction between the monomers should further help in the compaction of the polymer chain. To check whether this is the case, first we

calculate the number density of the monomers in radial and longitudinal directions and then we calculate the values of the ratio of principal moment of inertia, I_1/I_3 and asphericity, A_s and compare it with the polymer with BC' CLs under confinement but no attraction between the monomers. The no. densities of monomers in the radial and longitudinal directions are shown in Fig. 4.8. On comparing the no. densities in Figs.4.3 and 4.8 (a), we see that the attraction between the monomers helps in the compaction of the polymer in the radial directions as well. Note that, without attraction, on increasing the number of CLs, the radial density did not change. But the small attraction between the monomers helps the polymer to get compact in the radial as well as in the longitudinal directions.
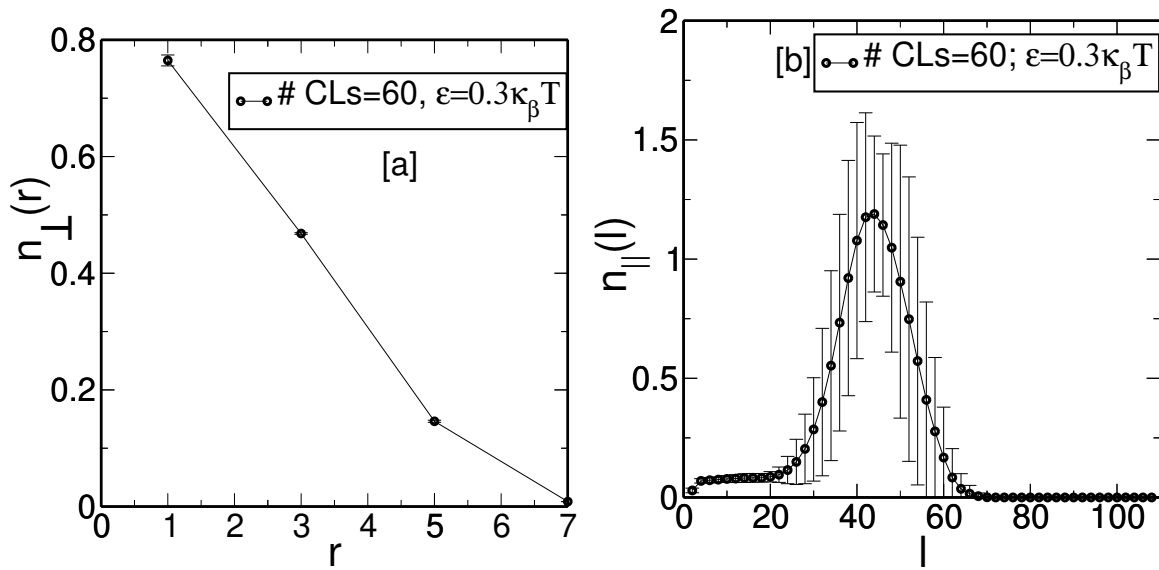

Fig. 4.8 The plots (a) and (b) show the number density of the monomers in the radial and longitudinal directions with the distance in the presence of attraction between the monomers. Error bars show the s.d. from the average value across 12 independent initial conditions.

Further, in the presence of attraction, the average values of the ratio I_1/I_3 and A_s are 11.38 and 0.18, respectively. On comparing with the values in table 4.1 for the polymer with BC' set of CLs, we see that the attraction between the monomers indeed compact the polymer and decreases its asphericity.

To investigate the internal organization of the polymer in the presence of small attraction between monomers with the value of parameter $\varepsilon = 0.3k_BT$, we calculate the positional correlation of different segments as before. To estimate the value of R_c, we take the help of our previous study of the polymer with different attraction strength parameter ε, with BC-2 sets of CLs and without confinement [138]. The value was taken to be $R_c = 4a$ to keep the value of R_c/R_g to be 0.55. The positional correlations are represented as a colormap in Fig. 4.9(a) and (b). The two colormaps correspond to the polymer starting from the 2 independent

initial conditions. From the colormaps, we see that the two colormaps look statistically similar. Furthermore, the overall positional correlations in the colormaps of Fig. 4.9 does not look significantly different from the colormaps of Fig. 4.7. But note that the cutoff R_c we chose to calculate the positional correlation has been decreased to $4a$ in this case, which was chosen to be $R_c = 5a$ to plot the colormap of Fig. 4.7. So, the small attraction between the monomers leads to the overall compaction of the polymer but does not change the internal organization of the polymer significantly in the presence of cylindrical confinement. Also, for the better visualization of the off-diagonal and diagonal pixels in the simulation colormaps of Fig. 4.9(a),(b), we plot the diagonal pixels with $p > 0.05$ and off-diagonal pixels with $p > 0.5$ separately in the plots of Fig. 4.9(c) and (d) in the binary form, respectively.

On comparing the plots Figs.4.9 (c) and (d) with the experimental coarse-grained colormap of Fig. 4.9 (e), we see that the diagonal and off-diagonal pixels, which are present as the bright pixels in the coarse-grained experimental contact map, are also present in the simulation colormap. The red markers in the plot Fig.4.9 (d) of Fig. 4.9 represents the segment-segment correlations, which could occur because those segments consist of the monomers constituting a cross-link. We observe that most of the segment-segment correlations in plot Fig.4.9(d) are due to the cross-links, but there are some segments that do not consist of the cross-linked monomers and correlated with each other with high probability ($p > 0.5$). These segment-segment correlations are also present in the experimental coarse-grained map of plot Fig.4.9(e). This is a validation of our model with the experimental data. The procedure to generate the plot 4.9 (e) is described in the supplementary section.

Next, we check for the statistical equivalence of organization obtained from 12 independent MC simulations as observed in the colormaps, which will establish that the few specific cross-links, the effect of molecular crowders and confinement lead to the unique structure of the polymers. Thus, we calculate the average value of the Pearson correlation of the positional correlation colormaps for the polymer (a) with no CLs (b) with BC-1 set of CLs (c) with BC' set of CLs and (d) with BC' set in the presence of small attraction between the monomers . Corresponding to runs starting from the 12 initial conditions, there will be $^{12}C_2 = 66$ values of Pearson correlation (pc) for each pair of graphs. We calculate the average value of Pearson correlation $\langle pc \rangle$ for each of the 4 different studies. The values of $\langle pc \rangle$ have been listed in the table 4.2. For details of the calculation procedure of $\langle pc \rangle$, refer [138].

From the table 4.2, we see that the value of Pearson correlation increases as we increase the number of CLs. We get the relatively higher value of $\langle pc \rangle$ if we introduce the small attraction between the monomers.

No. of CLs	Monomer Attraction	$\langle pc \rangle$
1(ring)	absent	0.51
49	absent	0.54
60	absent	0.60
60	present	0.70

Table 4.2 The table shows the average value of Pearson correlation $\langle pc \rangle$ for polymer with different value of CLs and in the presence and absence of attraction between the monomers. Note, to calculate the value of pc for a pair of colormaps, we consider only those pixels corresponding to segments i and j, for the probability $p(i, j) > 0.05$ in at least one of the colormaps from different runs. This is to avoid the bias in calculating pc, from the large number of pixels which are completely dark.

In our previous study of the DNA ring-polymer with BC-2 (159) CL-set and in the absence of the confinement we obtained fewer bright pixels along the Diagonal-2 in our simulations, unlike what is seen in the coarse-grained experimental contact map of *C. crescentus* (refer Fig. 4.9(e)). However, there was a good match of the off-diagonal points in the experimental and simulation contact maps [137]. In the current study, i.e., in the presence of confinement, we observe the presence of a Diagonal-2 in the computationally calculated colormaps, refer Fig.4.9(a),(b) and (c). The presence of the diagonal in Fig 4.9(c) can be compared with the diagonal in the experimental data shown in Fig 4.9(e). Fig 4.9(c) has plotted only the diagonals in binary colors to help the eye identify the diagonal, which is more difficult to identify in Fig.4.9(a),(b). In addition, we get a good match of computed off-diagonal bright pixels with the bright pixels in the coarse-grained experimental contact map; compare 4.9(d) with 4.9 (e) or (f).

To quantify and compare the presence of pixels along the Diagonal-2 in our current computations with the computations in chapter 2 and 3 (in the absence of confinement), we calculate the number of bright pixels (n_d) with $p > 0.05$, normalized suitably along the Diagonal-2 which are present in the colormap of Fig.4.9. We compare it with the positional correlation colormap of polymer with the BC-2 set of CLs in the absence of confinement in [137]. We choose the pixels along the diagonal-2 such that it consists of 5 nearest neighboring pixels (in the vertical and horizontal direction) of a particular segment index in the simulation colormaps. We find that the average value of n_d is significantly high, i.e., $n_d = 0.68$ for the colormaps of Fig. 4.9(a) and (b) compared to the relatively smaller value of n_d equal to 0.44 in the colormap of Fig. 2(a) in the [137].

In Fig. 4.10, we plot the positional correlation colormaps corresponding to the model chromosome of bacteria *E. coli* in the presence of small attraction between the monomers with

strength $\varepsilon = 0.3k_BT$. We use the same procedure that was used to calculate the colormaps of *C. crescentus*. The four colormaps in Fig.4.10 are from the independent MC runs starting from the four different initial conditions. We again see that the colormaps from independent runs look statistically similar to each other, and the average value of $\langle pc \rangle = 0.75$ is relatively high across independent runs starting from 12 initial conditions. To compare the simulation colormaps of Fig. 4.10 with the experimental coarse-grained colormap, we plot the binary version of the simulation colormap of Fig.4.10 in Fig.4.11, Where the black color represents the pixels with probability $p > 0.5$. We see that all the pixels which are present in the experimental contact map are also present in the colormaps obtained from our simulations. This again validates our model of the polymer with few CLs at specific positions. But note, now we have used fewer CLs (i.e. only 77 CLs $\equiv$ 38 *effective* CLs) compared to the *E. coli* BC-2 CL-set (159 CLs, effectively 82 CLs), which we used before [137, 136].

To predict the overall organization of the polymer in the presence of CLs at specific positions, confinement, and the small attraction between the monomers, we calculate the radial and longitudinal probability distribution of positions of the center of mass of different segments for two bacterial chromosomes. Both the longitudinal and radial distribution of positions of CMs of different segments is averaged over the runs starting from 12 independent initial condition. The longitudinal distribution of the positions of CMs of the segments is shown as the colormap in Fig. 4.12(a), (c) for *C. crescentus* and *E. coli*, respectively. The color shows the probability of the z-component of a particular segment to be found in the region numbered $1, 2.., 10$ for *C. crescentus*, or Left-R1, Left-R2.. for *E. coli*. For the model chromosome of bacteria *C. crescentus*, we measure the position from the location of the monomer numbered 1, which is tethered at one end of the cylinder. For the bacteria *E. coli*, the Ori is not tethered at the boundary of the cylinder; thus, we measure the longitudinal distribution of the segments from the center of mass of the longitudinal (z) component.

In Fig. 4.12 (a) the x-axis represents the segment index, and the y-axis represents the different regions numbered $1, 2, ..10$ along the length of the cylinder starting from the ori position. Each region corresponds to a length of $l = 10a$, i.e., Region-1 corresponds to $l = 0$-$10a$, Region-2 corresponds to $l = 10a$-$20a$ and so on. In Fig. 4.12(c) x-axis shows the segment index $i = 1$-80 and the y-axis represents the different regions measured from the center of mass (CM) of the polymer.
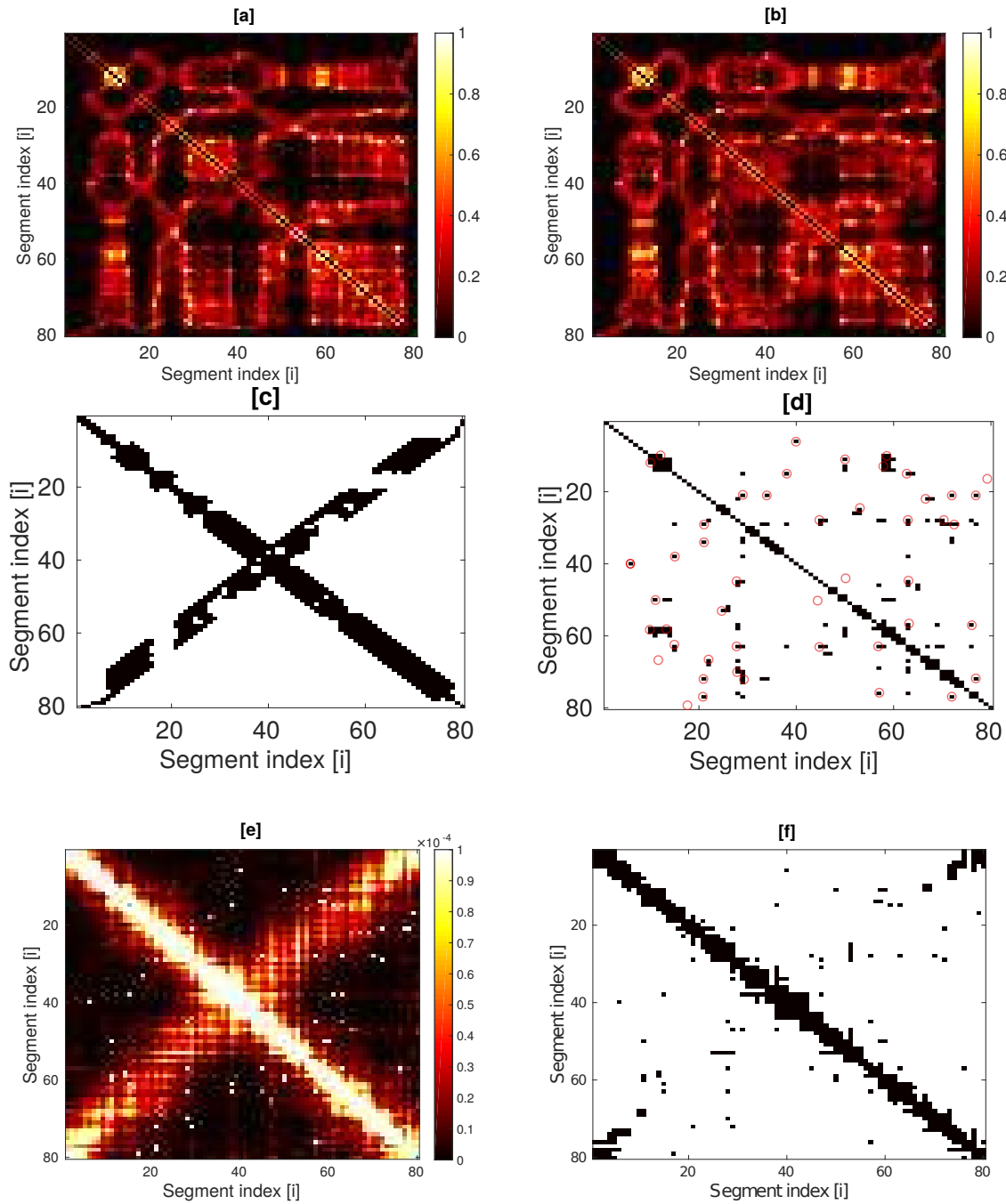
Fig. 4.9 The plots (a), (b) show the positional correlation colormaps for the center of mass of different polymer segments in the presence of weak attraction ($\varepsilon = 0.3k_BT$) between the monomers and with BC' set of CLs for *C. crescentus*. The value of the cutoff R_c to plot the colormap is set as $R_c = 4a$ to keep $R_c/R_g \approx 0.55$. The two colormaps are from the runs starting from two different initial conditions. To aid visualization of the diagonal and off-diagonal pixels of the plots (a) and (b), we plot only the diagonal pixels with $p > 0.05$ of the positional correlation maps and off-diagonal pixels with $p > 0.5$ in a black and white format in plots (c) and (d), respectively. The red markers in the plot (d) show the correlations between the segments which have monomers which are cross-linked. The plot (e) show the experimental coarse-grained colormap of 80 segments. For the better visualization of the off-diagonal pixels of the experimental colormap of plot (e), we only plot the pixels with probability $p > 0.0001$ in the plot (f).
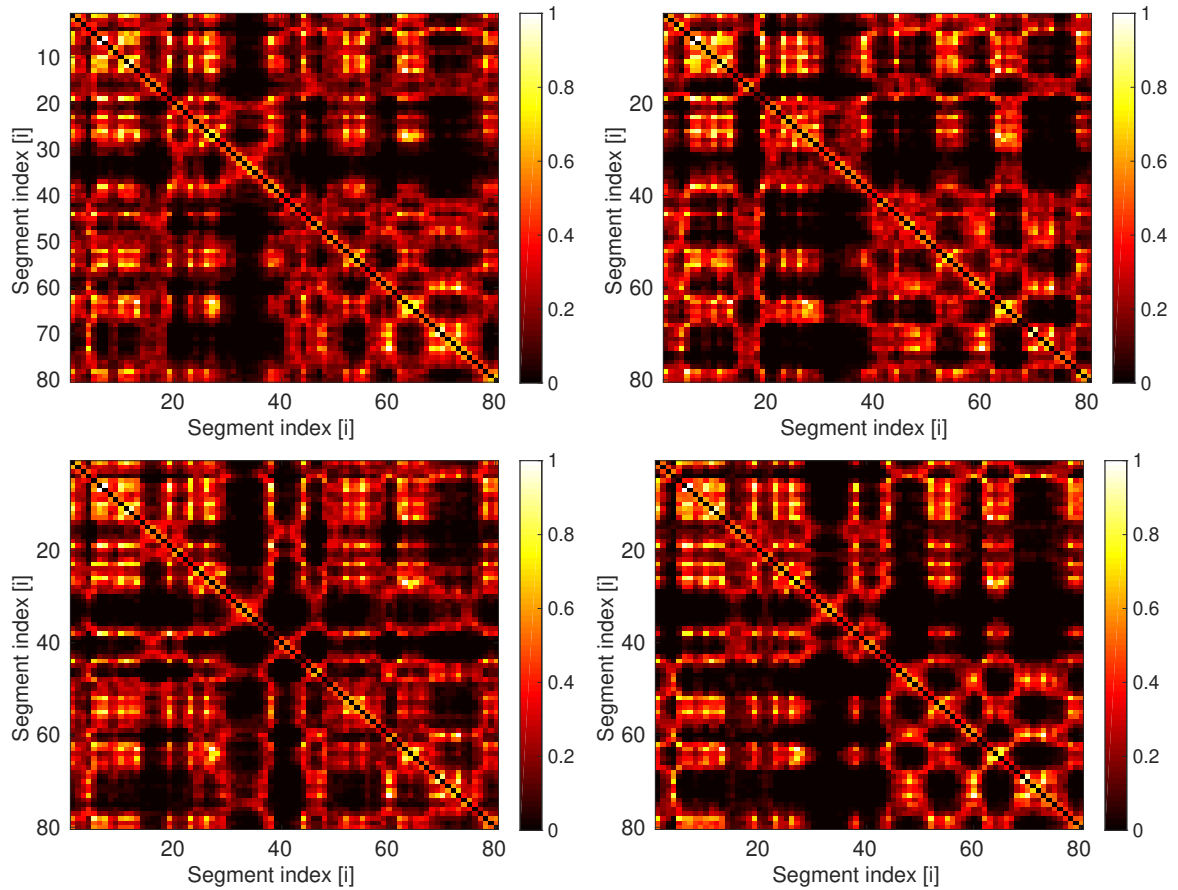
Fig. 4.10 The figure show the positional correlation colormaps for the center of mass of different polymer segments in the presence of weak attraction ($\varepsilon = 0.3k_BT$) between the monomers and with 77 CLs (38 *effective* CLs) taken from the contact map of bacteria *E. coli*. The value of the cutoff R_c to obtain the colormap is set as $R_c = 4a$. The four colormaps are from the runs starting from four different initial conditions.
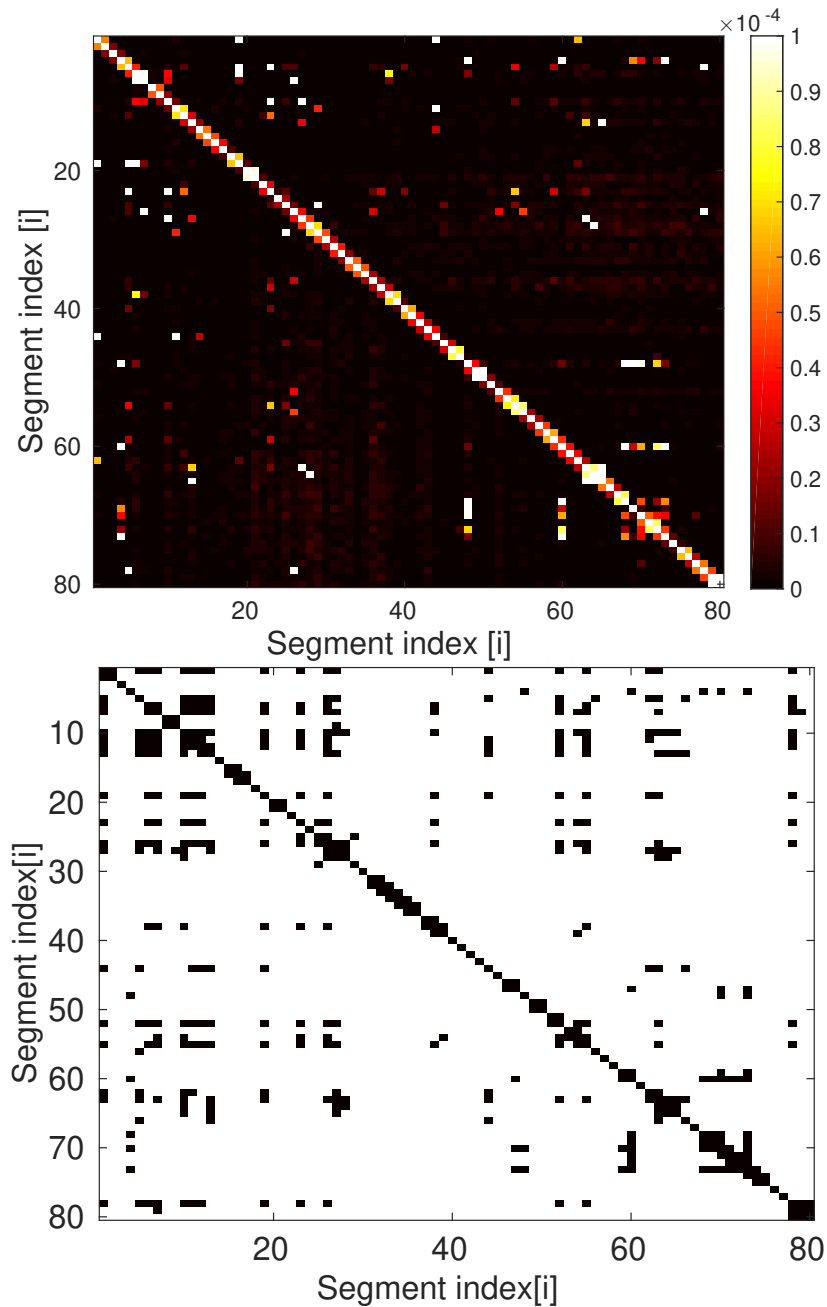
Fig. 4.11 The first figure show the coarse-grained experimental contact maps of bacteria
E. coli. The x-axis and y-axis represent the segment index numbered $1, 2...80$ and the color
represents the probability of the two segments to be found close to each other in Hi-C
experiments. The graph has been reproduced with permission from Europhysics letters, **121**,
18004 (2018). Copyright 2018 EPLA. The second figure shows the binary version of the
Fig.4.10 where the black points represent the pixels with probability $p > 0.5$. From the
above two figures we see that the pixels which are present in the experimental coarse-grained
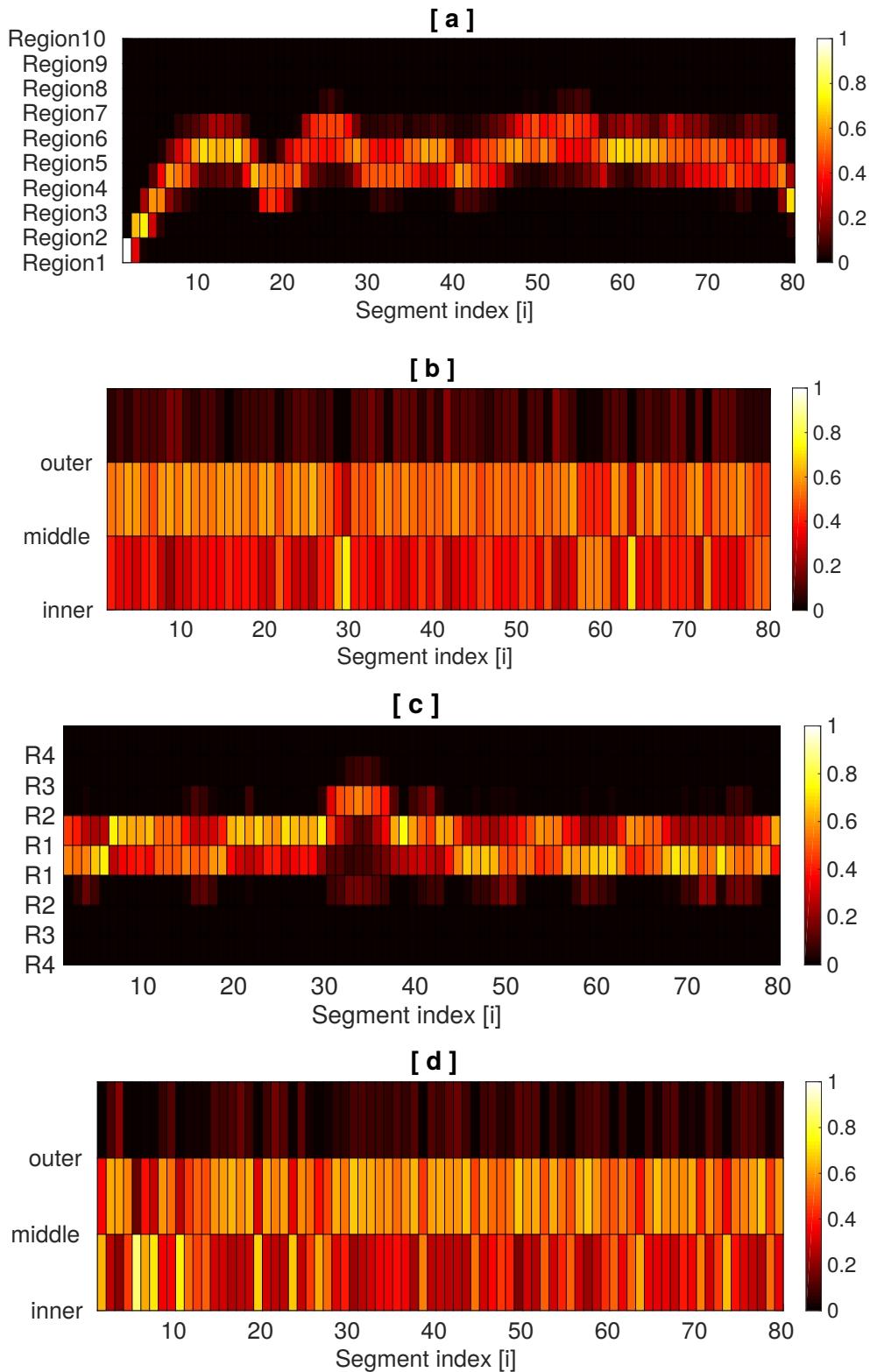colormap are also present in the simulation colormap.

Fig. 4.12 The colormaps (a),(c) show the longitudinal and (b), (d) the radial distributions of the positions of the center of mass of different segments for model chromosome of bacteria *C. crescentus* and *E. coli*, respectively. The color bar denotes the probability to find a segment in a particular region. Refer text for description of how the different regions are specified. In essence, this figure gives our prediction of the 3D organization of the different segments of the DNA-polymer within the cylinder for both the bacteria *C. crescentus* and *E. coli* using the thesis that few CLs at specific positions along the chain contour, effective attraction between the monomers and confinement effects are enough to organize the bacterial chromosome in the space within the cell.

The region left-R1 corresponds to the region at a value of z-component between 0 and $-10a$, and the region right-R1 represents the region at a value of z-component between 0 and $+10a$ from the CM.

For the radial distribution, we define three regions inner ($r \leq 2a$), middle ($2a < r \leq 4a$) and outer ($r > 4a$) from the axis of the cylinder as monomer numbered 1 is tethered at the axis and one end of the cylinder for the bacteria *C. crescentus*. For the bacteria *E. coli* the ori is not tethered at the boundary of the cylinder, we define the three regions from the CM of radial components (*x* and *y*). We then calculate the probability of the radial component of the CM of the segments to be found in the inner, middle, and outer region. Then we average the probability over independent runs starting from the 12 different initial conditions. It is shown as the colormaps in Fig. 4.12(b) and (d). The x-axis represents the segment index, and the y-axis corresponds to the three regions inner, middle, and outer. The color shows the probability of a segment to be found in the three regions.

The positional colormaps, the longitudinal, and radial distribution of the polymer segments give the overall 3D organization of the model chromosome. In our previous work, we observed that for the model chromosome of bacteria *E. coli*, the polymer segments with highly expressed genes were found in the outer region of the globule with high probability [137]. Here also, we check that the chromosome segments of bacteria *E. coli*, which consist of highly expressed genes, are found in the middle and outer regions, and they have a very less or zero probability to be found in the inner region. Hence, our results are also consistent with our previous prediction of the organization of the chromosome without confinement.

We also present the snapshots from the simulations in Fig. 4.13 for the model chromosomes of two bacteria. The monomers are colored from blue to red according to their positions along the chain contour. For the model chromosome of *E. coli* where the first monomer is not tethered to the boundary of the cylinder, the polymer occupies the center of the cylinder whereas for the chromosome of *C. crescentus* we see that most of the monomers occupy the center of the cylinder while the first monomer is tethered at the boundary of the cylinder. We also mark the positions of the CL monomers with black color. From the snapshots, we can confirm that the CLs are clustered in space for both the model chromosomes.

4.4 Discussions

We study the organization of bacterial DNA-polymers in cylindrical confinement and in the presence of optimally chosen weak attraction between the monomers with the different number of cross-links at specific positions along the chain contour. These positions of specific
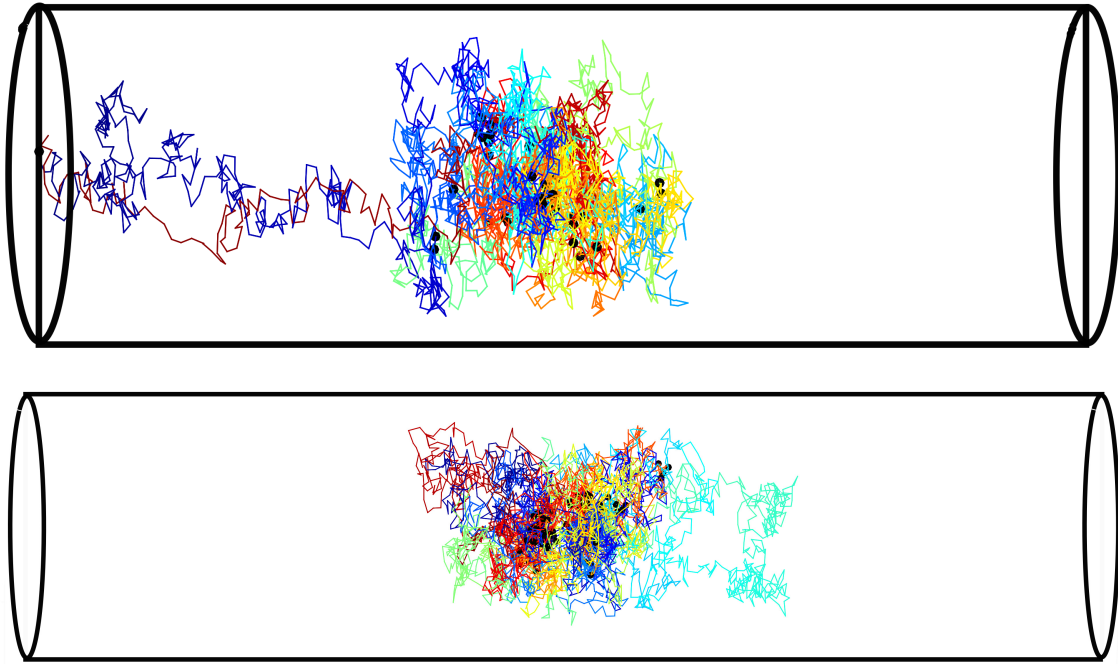
Fig. 4.13 The snapshots from the simulation for the model chromosome of bacteria *C. crescentus* and *E. coli*. The positions of CLs are represented by the black points.

cross-links are taken from the contact maps of bacterial DNA. In the studies of chapter 2 and chapter 3, we had shown that in the absence of confinement and molecular crowders, we were able to get the structure of the DNA-polymer with 159&153 cross-links (82&60 effective cross-links) [137, 136] for *E. coli* & *C. crescentus*, respectively. In this study, we show that we get the organization of the DNA polymer with only 38&33 effective cross-links, respectively, for both the bacterial chromosomes when we incorporate the effects of molecular crowders, ability to release topological constraints and confinement. So, in effect, we show that the very few specific cross-links, confinement, and the effect of molecular crowders play the pivotal roles for the polymer to acquire a particular micron-scale organization. The Pearson correlation on comparing positional-correlation contact-maps from independent MC runs (starting from the independent initial conditions) are $= 0.7$ for *C. crescentus* and 0.75 for *E. coli*. The ability to release topological constraints also plays a crucial role in the organization of the DNA polymer.

We obtain a better agreement of our predicted organization with the coarse-grained experimental contact map of *C. crescentus*, viz., the pixels along the second diagonal of the positional correlation colormaps which were absent in our previous studies [137] are now

present for the polymer in the presence of BC' set of cross-links, confinement and crowders. The organization of the chromosome of *E. coli* in the presence of these added effects are consistent with our previous prediction of the organization of the chromosome, which in turn was in agreement with the coarse-grained experimental contact map for *E. coli*. But, now we can predict the approximate positions of the center of mass of different polymer segments each consisting of 50&58 monomers (50&58 kilobase pairs) for *C. crescentus* and *E. coli*, respectively, in the 3D space within the cylinder. We hope that our predictions can be validated in future experiments.

In this study, we took a larger aspect ratio (1 : 7.5) of the cylinder compared to the aspect ratio ($\approx$ 1 : 5) as used by other researchers for their studies [8, 131]. Even though we use a longer cylinder (of length $108a$ and rather than the expected $70a$ for a cylinder diameter of $14a$), we observe that the polymer is collapsed in the cylinder, and the monomer density is zero beyond $l = 75a$. We also check the organization of polymer with BC' set of CLs inside a cylinder with different diameters, i.e., for cylinder's diameter $D = 10, 12$, and we do not observe a significant difference in the positional correlation colormaps of different polymer segments. For *C. crescentus*, the Ori is fixed at one end of the cylinder, whereas the Ori for *E. coli* is at the center of the coil. The special position of the *C. crescentus* Ori contributes to the presence of the diagonal seen in the contact map for *C. crescentus*. The consistency of the *E. coli* organization with our previous results means that the gene-dense regions and the active regions are found on the peripheral regions of the chromosome, as was pointed out earlier in [137]. This is expected to have important biological consequences.

In our studies, we have considered the cross-linked monomers to be at fixed positions along the contour of the polymer. Other studies [81, 88, 96] consider the binders (cross-links) between different segments of the chain to have the ability to diffuse around, as is known to be the case experimentally. But we have considered only 33 (&38) *effective* cross-links in a 4 (&4.6) million base pair DNA chain, modeled with 4 (and 4.6) $\times 10^3$ beads in a bead-spring polymer model. Other than the ones we have considered, there can be other binding proteins which can diffuse around and play a role in the local and more detailed organization. Secondly, even if the position of the cross-links in our study moves by a few (4 or 5) monomers on either side of its present position, the obtained organization cannot be significantly different, within the statistical fluctuations. As each monomer represents 1000 base pairs, a variation of the position of the cross-link over 8 monomers represents a diffusion of the DNA-binding protein along the contour over $\approx$ 8 kilobase pairs.

In our prediction of the 3D organization, we list the segments which are likely to be found in the peripheral regions with higher probability or the segments which are likely to be found in the central core region. Moreover, for *C. crescentus*, we predict the relative distance from

the Ori, which is fixed at one end of the cell. However, for *E. coli*, we measure the distance from the center of the coil (center of mass), and from the data, we can experimentally validate which segments are likely to be opposite each other in space along the axis. We further observe in our simulations that the CLs are clustered at the center of the cylinder and the coil; this is consistent with the prevalent understanding that DNA organization at large-length scales in bacteria is seen by the formation of long loops "emanating" out from clusters of DNA-binding proteins [42].

We observe that the relatively very few cross-links can organize a ring polymer in the presence of confinement. We are aware that at smaller length-scales (smaller than 1000BP), Nucleoid associated proteins (NAP) and super-coiling play a significant role in locally compacting the DNA-chain. We believe our model is simpler than other present models that optimize multiple parameters to predict the detailed contact map. The only parameters we have optimized in our preceding manuscript and used here are the weak attraction $\varepsilon = 0.3k_BT$ and the bead diameter $\sigma = 0.2a$, which accounts for the role of crowders and the ability of the chains to cross itself.

Chapter 5

Conclusions and Future Directions

In this thesis, we study the role of cross-links in the organization of the polymer using the bead-spring model of polymer physics. We show that $< 3\%$ of the cross-links can organize the polymer into a particular structure, while we know that the polymer without cross-links can take all the configurations in the 3D space and does not have a unique organization. We take the specific positions of the cross-links along the polymer contour from the Hi-C contact map of the DNA. In a separate study, we also chose the positions of the cross-links randomly. Using Monte-Carlo simulation, we generate the different configurations of the polymer with cross-links at specific or random positions along the contour. On comparing the organization between the biological cross-linked polymer and randomly cross-linked polymer, we find that the nature of the organization in two cases is different. Release of topological constraints through chain crossing could play a crucial in the above organization. Thus, we conclude that the biological cross-links are at specific positions along the contour and give rise to the unique organization to the DNA polymer. We work with the contact map of two bacteria, namely, *E. coli* and *C. crescentus*, and using the bead-spring polymer model and the data from Hi-C contact map we predict the 3D organization of the DNA polymer. We also match our predicted organization with the available experimental data. We find the clusters of CLs towards the center of the coil. These CLs are pulling different segments of the chain towards the center, and many loops on the periphery. We interpret it as the rosette-like structure.

We next study the effect of the release of topological constraints and molecular crowders in the organization of the polymer with very few specific permanent cross-links taken from the experimental contact map of bacteria *E. coli* and *C. crescentus*. We showed that the release of topological constraints is crucial for the polymer to organize into a particular structure since, in the presence of chain crossing, the model chromosomes from 9 independent runs organize into a particular structure. We get a relatively higher value of Pearson correlations of at least $\langle pc \rangle \approx 0.7$ (and up to 0.85 across runs starting from independent initial conditions, for the

cases in which chain crossing is allowed compared to the cases when the chain crossing is not allowed. This leads us to believe that the activity of enzyme topoisomerase can play a vital role in the organization of the chromosome at large-length scales by allowing the chains to cross itself. We also show that optimum compaction due to molecular crowders in the bacterial cytoplasm of the cell is possibly required, so that the DNA polymer is able to explore the different configurations and organize into a particular structure. We find the values of effective weak attraction between the monomers to be $\varepsilon = 0.1, 0.2, 0.3 k_B T$ for which the polymer from different initial conditions is able to obtain a unique organization.

Finally, we study the organization of bacterial DNA-polymers in cylindrical confinement and with the optimally chosen weak attraction between the monomers with the different number of cross-links at specific positions along the chain contour in chapter 4. We show that in the presence of molecular crowders, the number of cross-links required for the polymer to get organized is $< 1\%$. So, in effect, we show that the very few specific cross-links, confinement, and the effect of molecular crowders play the pivotal roles for the polymer to acquire a particular micron-scale organization. The Pearson correlation on comparing positional-correlation contact-maps from independent MC runs (starting from the independent initial conditions) are $= 0.7$ for *C. crescentus* and 0.75 for *E. coli*. The ability to release topological constraints also plays a crucial role in the organization of the DNA polymer. We obtain a better agreement of our predicted organization with the coarse-grained experimental contact map of *C. crescentus*. As the pixels along the second diagonal of the positional correlation colormaps which were absent in our studies of chapter 2 are now present for the polymer in the presence of BC' set of cross-links, confinement, and crowders. The organization of the chromosome of *E. coli* in the presence of these added effects are consistent with our previous prediction of the organization of the chromosome, which in turn was in agreement with the coarse-grained experimental contact map for *E. coli*. We hope that our predictions can be validated in future experiments.

In our prediction of the 3D organization, we list the segments which we find in the peripheral regions with higher probability or the segments found in the central core region. Moreover, for *C. crescentus*, we predict the relative distance from the Ori, which is fixed at one end of the cell. However, for *E. coli*, we measure the distance from the center of the coil (center of mass), and from the data, we can experimentally validate Which segments are likely to be opposite to each other in space along the axis. We further observe in our simulations that the CLs cluster at the center of the cylinder and the coil; it is consistent with the prevalent understanding that DNA organization at large-length scales in bacteria is seen as the formation of long loops "emanating" out from clusters of DNA-binding proteins [42].

For this study, we have studied the chromosome organization in the stage of the cell cycle, where the DNA is not replicating. But it is essential to consider the organization of the chromosome over the period of cell-cycle. We have reported some experimental results in Appendix A, to obtain the organization of chromosome of *E. coli* during multifork replication. But, understanding the mechanisms of the large-scale organization during multifork replication and for the cells growing with different rates can be done only in the future with close dialogue between modeling and experiments.

References

[1] J. D. WATSON and F. H. C. CRICK. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, April 1953. doi: 10.1038/171737a0. URL https://doi.org/10.1038/171737a0.

[2] Julie Theriot Rob Phillips, Jane Kondev. *Physical Biology of the Cell*. Garland Science., 2008.

[3] Anjana Badrinarayanan, Tung B.K. Le, and Michael T. Laub. Bacterial chromosome organization and segregation. *Annual Review of Cell and Developmental Biology*, 31 (1):171–199, November 2015. doi: 10.1146/annurev-cellbio-100814-125211. URL https://doi.org/10.1146/annurev-cellbio-100814-125211.

[4] Marc Joyeux. Compaction of bacterial genomic dna: clarifying the concepts. *Journal of Physics: Condensed Matter*, 27(38):383001, 2015. URL http://stacks.iop.org/0953-8984/27/i=38/a=383001.

[5] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson. *Molecular Biology of the Cell*. Garland, 4th edition, 2002.

[6] Gary Felsenfeld and Mark Groudine. Controlling the double helix. *Nature*, 421(6921): 448–453, January 2003. doi: 10.1038/nature01411. URL https://doi.org/10.1038/nature01411.

[7] Xindan Wang, Paula Montero Llopis, and David Z. Rudner. Organization and segregation of bacterial chromosomes. *Nature Reviews Genetics*, 14(3):191–203, February 2013. doi: 10.1038/nrg3375. URL https://doi.org/10.1038/nrg3375.

[8] Tung B. K. Le, Maxim V. Imakaev, Leonid A. Mirny, and Michael T. Laub. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, 342(6159):731–734, 2013. doi: 10.1126/science.1242059.

[9] Talukder Ali Azam, Sota Hiraga, and Akira Ishihama. Two types of localization of the DNA-binding proteins within the escherichia coli nucleoid. *Genes to Cells*, 5(8):613–626, August 2000. doi: 10.1046/j.1365-2443.2000.00350.x. URL https://doi.org/10.1046/j.1365-2443.2000.00350.x.

[10] R. T. Dame. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Research*, 28(18):3504–3510, September 2000. doi: 10.1093/nar/28.18.3504. URL https://doi.org/10.1093/nar/28.18.3504.

[11] Remus T. Dame. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology*, 56(4):858–870, March 2005. doi: 10.1111/j.1365-2958.2005.04598.x. URL https://doi.org/10.1111/j.1365-2958.2005.04598.x.

[12] Andreas Hofmann and Dieter W. Heermann. The role of loops on the order of eucaryotes and prokaryotes. *FEBS Letters*, 589:2958–2965, 2015. doi: 10.1016/j.febslet.2015.04.021.

[13] Vincenzo G Benza, Bruno Bassetti, Kevin D Dorfman, Vittore F Scolari, Krystyna Bromek, Pietro Cicuta, and Marco Cosentino Lagomarsino. Physical descriptions of the bacterial nucleoid at large scales, and their biological implications. *Reports on Progress in Physics*, 75(076602), 2012. doi: https://doi.org/10.1088/0034-4885/75/7/076602.

[14] Martijn S. Luijsterburg, Malcolm F. White, Roel van Driel, and Remus Th. Dame. The major architects of chromatin: Architectural proteins in bacteria, archaea and eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology,*, 43:393–418, 2008. doi: 10.1080/10409230802528488.

[15] Shane C. Dillon and Charles J. Dorman. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, 8:185–195, 2010. doi: 10.1038/nrmicro2261.

[16] Miriam Fritsche, Songling Li, Dieter W. Heermann, and Paul A. Wiggins. A model for escherichia coli chromosome packaging supports transcription factor-induced DNA domain formation. *Nucleic Acids Research*, 40(3):972–980, oct 2011. doi: 10.1093/nar/gkr779. URL https://doi.org/10.1093/nar/gkr779.

[17] Fabai Wu, Pinaki Swain, Louis Kuijpers, Xuan Zheng, Kevin Felter, Margot Guurink, Jacopo Solari, Suckjoon Jun, Thomas S. Shimizu, Debasish Chaudhuri, Bela Mulder, and Cees Dekker. Cell boundary confinement sets the size and position of the e. coli chromosome. *Current Biology*, 29(13):2131–2144.e4, July 2019. doi: 10.1016/j.cub.2019.05.015. URL https://doi.org/10.1016/j.cub.2019.05.015.

[18] Wei-Shao Tung, Russell J. Composto, Robert A. Riggleman, and Karen I. Winey. Local polymer dynamics and diffusion in cylindrical nanoconfinement. *Macromolecules*, 48 (7):2324–2332, apr 2015. doi: 10.1021/acs.macromol.5b00085. URL https://doi.org/10.1021/acs.macromol.5b00085.

[19] James J. Champoux. DNA topoisomerases: Structure, function, and mechanism. *Annual Review of Biochemistry*, 70(1):369–413, June 2001. doi: 10.1146/annurev.biochem.70.1.369. URL https://doi.org/10.1146/annurev.biochem.70.1.369.

[20] Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, I. Amit T. Ragoczy, A. Telling, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, B. Bernstein R. Sandstrom, M. A. Bender, J. Stamatoyannopoulos M. Groudine, A. Gnirke, L. A. Mirny, and J. Dekker E. S. Lander. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009. doi: 10.1126/science.1181369.

[21] Cedric Cagliero, Ralph S. Grand, M. Beatrix Jones, Ding J. Jin, and Justin M. O'Sullivan. Genome conformation capture reveals that the escherichia coli chromosome is organized by replication and transcription. *Nucleic Acids Res.*, 41(12): 6058–6071, 2013. doi: 10.1093/nar/gkt325.

[22] B. Youngren, H. J. Nielsen, S. Jun, and S. Austin. The multifork escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes & Development*, 28(1):71–84, January 2014. doi: 10.1101/gad.231050.113. URL https://doi.org/10.1101/gad.231050.113.

[23] H. J. NIELSEN and F. G. HANSEN. An automated and highly efficient method for counting and measuring fluorescent foci in rod-shaped bacteria. *Journal of Microscopy*, 239(3):194–199, August 2010. doi: 10.1111/j.1365-2818.2010.03374.x. URL https: //doi.org/10.1111/j.1365-2818.2010.03374.x.

[24] Henrik J. Nielsen, Yongfang Li, Brenda Youngren, Flemming G. Hansen, and Stuart Austin. Progressive segregation of the escherichia coli chromosome. *Molecular Microbiology*, 61(2):383–393, July 2006. doi: 10.1111/j.1365-2958.2006.05245.x. URL https://doi.org/10.1111/j.1365-2958.2006.05245.x.

[25] Xindan Wang, Xun Liu, Christophe Possoz, and David J. Sherratt. The two escherichia coli chromosome arms locate to separate cell halves. *Genes & Development*, 20(13): 1727–1731, July 2006. doi: 10.1101/gad.388406. URL https://doi.org/10.1101/gad. 388406.

[26] X. Wang, C. Lesterlin, R. Reyes-Lamothe, G. Ball, and D. J. Sherratt. Replication and segregation of an escherichia coli chromosome with two replication origins. *Proceedings of the National Academy of Sciences*, 108(26):E243–E250, June 2011. doi: 10.1073/pnas.1100874108. URL https://doi.org/10.1073/pnas.1100874108.

[27] Shun Adachi, Toru Fukushima, and Sota Hiraga. Dynamic events of sister chromosomes in the cell cycle of escherichia coli. *Genes to Cells*, 13(2):181–197, January 2008. doi: 10.1111/j.1365-2443.2007.01157.x. URL https://doi.org/10.1111/j. 1365-2443.2007.01157.x.

[28] David Bates and Nancy Kleckner. Chromosome and replisome dynamics in e. coli: Loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, 121(6):899–911, June 2005. doi: 10.1016/j.cell.2005.04.013. URL https://doi.org/10.1016/j.cell.2005.04.013.

[29] Jay K. Fisher, Aude Bourniquel, Guillaume Witz, Beth Weiner, Mara Prentiss, and Nancy Kleckner. Four-dimensional imaging of e. coli nucleoid organization and dynamics in living cells. *Cell*, 153(4):882–895, May 2013. doi: 10.1016/j.cell.2013. 04.006. URL https://doi.org/10.1016/j.cell.2013.04.006.

[30] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, M. Hu Y. Li, Y. Shen, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485:376–380, 2012. doi: 10.1038/nature11082.

[31] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148:458–472, 2012. doi: 10.1016/j.cell.2012.01.010.

[32] Jennifer E. Phillips-Cremins, Michael E.G. Sauria, Amartya Sanyal, Tatiana I. Gerasimova, Bryan R. Lajoie, Joshua S.K. Bell, Chin-Tong Ong, Tracy A. Hookway, Changying Guo, Yuhua Sun, Michael J. Bland, William Wagstaff, Stephen Dalton, Todd C. McDevitt, Ranjan Sen, Job Dekker, James Taylor, and Victor G. Corces. Architectural protein subclasses shape 3d organization of genomes during lineage commitment. *Cell*, 153(6):1281–1295, 2013.

[33] Reza Kalhor, Harianto Tjong, Nimanthi Jayathilaka, Frank Alber, and Lin Chen. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnology*, 30:90–98, 2012. doi: 10.1038/nbt. 2057.

[34] Siyuan Wang, Jun-Han Su, Brian J. Beliveau, Bogdan Bintu, Jeffrey R. Moffitt, Chao-ting Wu, and Xiaowei Zhuang. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, 353(6299):598–602, 2016. ISSN 0036-8075. doi: 10.1126/science.aaf8084. URL http://science.sciencemag.org/content/353/6299/598.

[35] Daan J.W. Brocken, Mariliis Tark-Dame, and Remus T. Dame. The organization of bacterial genomes: Towards understanding the interplay between structure and function. *Current Opinion in Systems Biology*, 8:137 – 143, 2018. ISSN 2452-3100. doi: https://doi.org/10.1016/j.coisb.2018.02.007. URL http://www.sciencedirect.com/science/article/pii/S2452310017302019. • Regulatory and metabolic networks • Special Section: Single cell and noise.

[36] Fabai Wu, Pinaki Swain, Louis Kuijpers, Xuan Zheng, Kevin Felter, Margot Guurink, Debasish Chaudhuri, Bela Mulder, and Cees Dekker. Cell boundary confinement sets the size and position of the e. coli chromosome. jun 2018. doi: 10.1101/348052. URL https://doi.org/10.1101/348052.

[37] Peiwu Qin, Mahmut Parlak, Cem Kuscu, Jigar Bandaria, Mustafa Mir, Karol Szlachta, Ritambhara Singh, Xavier Darzacq, Ahmet Yildiz, and Mazhar Adli. Live cell imaging of low- and non-repetitive chromosome loci using crispr-cas9. *Nature Communications*, 8(14725), 2017. doi: 10.1038/ncomms14725.

[38] Mahipal Ganji, Indra A. Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H. Haering, and Cees Dekker. Real-time imaging of DNA loop extrusion by condensin. *Science*, 360(6384):102–105, feb 2018. doi: 10.1126/science.aar7831. URL https://doi.org/10.1126/science.aar7831.

[39] Guy Nir, Irene Farabella, Cynthia Pérez Estrada, Carl G. Ebeling, Brian J. Beliveau, Hiroshi M. Sasaki, S. Dean Lee, Son C. Nguyen, Ruth B. McCole, Shyamtanu Chattoraj, Jelena Erceg, Jumana AlHaj Abed, Nuno M. C. Martins, Huy Q. Nguyen, Mohammed A. Hannan, Sheikh Russell, Neva C. Durand, Suhas S. P. Rao, Jocelyn Y. Kishi, Paula Soler-Vila, Michele Di Pierro, José N. Onuchic, Steven P. Callahan,

John M. Schreiner, Jeff A. Stuckey, Peng Yin, Erez Lieberman Aiden, Marc A. Marti-Renom, and C. ting Wu. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLOS Genetics*, 14(12):e1007872, December 2018. doi: 10.1371/journal.pgen.1007872. URL https://doi.org/10.1371/journal.pgen.1007872.

[40] Bogdan Bintu, Leslie J. Mateo, Jun-Han Su, Nicholas A. Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N. Boettiger, and Xiaowei Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, October 2018. doi: 10.1126/science.aau1783. URL https://doi.org/10.1126/science.aau1783.

[41] Yingxin Ma, Mingxiu Wang, Wei Li, Zhiping Zhang, Xiaowei Zhang, Tianwei Tan, Xian-En Zhang, and Zongqiang Cui. Live cell imaging of single genomic loci with quantum dot-labeled TALEs. *Nature Communications*, 8(1), May 2017. doi: 10.1038/ncomms15318. URL https://doi.org/10.1038/ncomms15318.

[42] Joel Stavans and Amos Oppenheim. Dna–protein interactions and bacterial chromosome architecture. *Physical Biology*, 3(4):R1, 2006. URL http://stacks.iop.org/1478-3975/3/i=4/a=R01.

[43] J Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3 (1):371–394, October 1949. doi: 10.1146/annurev.mi.03.100149.002103. URL https://doi.org/10.1146/annurev.mi.03.100149.002103.

[44] R. E. Buchanan. Life phases in a bacterial culture. *Journal of Infectious Diseases*, 23(2):109–125, August 1918. doi: 10.1086/infdis/23.2.109. URL https://doi.org/10.1086/infdis/23.2.109.

[45] Suckjoon Jun, Fangwei Si, Rami Pugatch, and Matthew Scott. Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. *Reports on Progress in Physics*, 81(5):056601, February 2018. doi: 10.1088/1361-6633/aaa628. URL https://doi.org/10.1088/1361-6633/aaa628.

[46] P R Painter and A G Marr. Mathematics of microbial populations. *Annual Review of Microbiology*, 22(1):519–548, October 1968. doi: 10.1146/annurev.mi.22.100168.002511. URL https://doi.org/10.1146/annurev.mi.22.100168.002511.

[47] Hans Bremer and Patrick P. Dennis. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1), September 2008. doi: 10.1128/ecosal.5.2.3. URL https://doi.org/10.1128/ecosal.5.2.3.

[48] E. O. POWELL. Growth rate and generation time of bacteria, with special reference to continuous culture. *Journal of General Microbiology*, 15(3):492–511, December 1956. doi: 10.1099/00221287-15-3-492. URL https://doi.org/10.1099/00221287-15-3-492.

[49] Sattar Taheri-Araghi, Serena Bradde, John T. Sauls, Norbert S. Hill, Petra Anne Levin, Johan Paulsson, Massimo Vergassola, and Suckjoon Jun. Cell-size control and homeostasis in bacteria. *Current Biology*, 25(3):385–391, February 2015. doi: 10.1016/j.cub.2014.12.009. URL https://doi.org/10.1016/j.cub.2014.12.009.

[50] Fangwei Si, Dongyang Li, Sarah E. Cox, John T. Sauls, Omid Azizi, Cindy Sou, Amy B. Schwartz, Michael J. Erickstad, Yonggun Jun, Xintian Li, and Suckjoon Jun. Invariance of initiation mass and predictability of cell size in escherichia coli. *Current Biology*, 27(9):1278–1287, May 2017. doi: 10.1016/j.cub.2017.03.022. URL https://doi.org/10.1016/j.cub.2017.03.022.

[51] Charles E. Helmstetter. Rate of DNA synthesis during the division cycle of escherichia coli b/r. *Journal of Molecular Biology*, 24(3):417–427, March 1967. doi: 10.1016/0022-2836(67)90228-8. URL https://doi.org/10.1016/0022-2836(67)90228-8.

[52] Stephen Cooper and Charles E. Helmstetter. Chromosome replication and the division cycle of escherichia coli. *Journal of Molecular Biology*, 31(3):519–540, February 1968. doi: 10.1016/0022-2836(68)90425-7. URL https://doi.org/10.1016/0022-2836(68)90425-7.

[53] H. Bremer and G. Churchward. An examination of the cooper-helmstetter theory of DNA replication in bacteria and its underlying assumptions. *Journal of Theoretical Biology*, 69(4):645–654, December 1977. doi: 10.1016/0022-5193(77)90373-3. URL https://doi.org/10.1016/0022-5193(77)90373-3.

[54] E. Toro and L. Shapiro. Bacterial chromosome organization and segregation. *Cold Spring Harbor Perspectives in Biology*, 2(2):a000349–a000349, January 2010. doi: 10.1101/cshperspect.a000349. URL https://doi.org/10.1101/cshperspect.a000349.

[55] Michael T. Laub, Lucy Shapiro, and Harley H. McAdams. Systems biology of-Caulobacter. *Annual Review of Genetics*, 41(1):429–441, December 2007. doi: 10.1146/annurev.genet.41.110306.130346. URL https://doi.org/10.1146/annurev.genet.41.110306.130346.

[56] M. Schaechter, O. MaalOe, and N. O. Kjeldgaard. Dependency on medium and temperature of cell size and chemical composition during balanced growth of salmonella typhimurium. *Journal of General Microbiology*, 19(3):592–606, December 1958. doi: 10.1099/00221287-19-3-592. URL https://doi.org/10.1099/00221287-19-3-592.

[57] N. O. Kjeldgaard, O. MaalOe, and M. Schaechter. The transition between different physiological states during balanced growth of salmonella typhimurium. *Journal of General Microbiology*, 19(3):607–616, December 1958. doi: 10.1099/00221287-19-3-607. URL https://doi.org/10.1099/00221287-19-3-607.

[58] Stephen Cooper. On the fiftieth anniversary of the schaechter, maaløe, kjeldgaard experiments: implications for cell-cycle and cell-growth control. *BioEssays*, 30(10):1019–1024, October 2008. doi: 10.1002/bies.20814. URL https://doi.org/10.1002/bies.20814.

[59] P. A. Wiggins, K. C. Cheveralls, J. S. Martin, R. Lintner, and J. Kondev. Strong intranucleoid interactions organize the escherichia coli chromosome into a nucleoid filament. *Proceedings of the National Academy of Sciences*, 107(11):4991–4995, March 2010. doi: 10.1073/pnas.0912062107. URL https://doi.org/10.1073/pnas.0912062107.

[60] J. Pelletier, K. Halvorsen, B.-Y. Ha, R. Paparcone, S. J. Sandler, C. L. Woldringh, W. P. Wong, and S. Jun. Physical manipulation of the escherichia coli chromosome reveals its soft nature. *Proceedings of the National Academy of Sciences*, 109(40): E2649–E2656, September 2012. doi: 10.1073/pnas.1208689109. URL https://doi.org/10.1073/pnas.1208689109.

[61] Julio Mateos-Langerak, Manfred Bohn, Wim de Leeuw, Osdilly Giromus, Erik M. M. Manders, Pernette J. Verschure, Mireille H. G. Indemans, Hinco J. Gierman, Dieter W. Heermann, Roel van Driel, and Sandra Goetze. Spatially confined folding of chromatin in the interphase nucleus. *PNAS U.S.A.*, 106(10):3812–3817, 2009. doi: 10.1073/pnas.0809501106.

[62] Job Dekker, Marc A. Marti-Renom, and Leonid A. Mirny. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet.*, 14(6):390–403, 2013. doi: 10.1038/nrg3454.

[63] Rolf Ohlsson, Marek Bartkuhn, and Rainer Renkawitz. Ctcf shapes chromatin by multiple mechanisms: the impact of 20 years of ctcf research on understanding the workings of chromatin. *Chromosoma*, 119(4):351–360, 2010. doi: 10.1007/s00412-010-0262-0.

[64] Mathieu Rousseau, James Fraser, Maria A Ferraiuolo, Josée Dostie, and Mathieu Blanchette. Three-dimensional modeling of chromatin structure from interaction frequency data using markov chain monte carlo sampling. *BMC Bioinformatics*, 12 (414), 2011. doi: 10.1186/1471-2105-12-414.

[65] Davide Baù, Amartya Sanyal, Bryan R Lajoie, Emidio Capriotti, Meg Byron, Jeanne B Lawrence, Job Dekker, and Marc A Marti-Renom. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nature Structural & Molecular Biology*, 18(1):107–114, 2011.

[66] Tom Misteli. Beyond the sequence: Cellular organization of genome function. *Cell*, 128(4):787–800, 2007.

[67] Wendy A. Bickmore and Bas van Steensel. Genome architecture: Domain organization of interphase chromosomes. *Cell*, 152(6):1270–1284, 2013.

[68] Job Dekker and Leonid Mirny. The 3d genome as moderator of chromosomal communication. *Cell*, 164(6):1110–1121, 2016.

[69] H. Tjong, K. Gong, Chen, L., and F Alber. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.*, 22: 1295–1305, 2012.

[70] Leonid A. Mirny Maxim V. Imakaev, Geoffrey Fudenberg. Modelling chromosome: Beyond pretty pictures. *FEBS Letters*, 589:3031–3036, 2015. doi: 10.1016/j.febslet.2015.09.004.

[71] Debasish Chaudhuri and Bela M. Mulder. Spontaneous helicity of a polymer with side loops confined to a cylinder. *Phys. Rev. Lett.*, 108(268305), 2012. doi: 10.1103/PhysRevLett.108.268305.

[72] Debasish Chaudhuri and Bela M. Mulder. Size and shape of excluded volume polymers confined between parallel plates. *Phys. Rev. E*, 83(031803), 2011. doi: 10.1103/PhysRevE.83.031803.

[73] Balaji V. S. Iyer and Gaurav Arya. Lattice animal model of chromosome organization. *Phys. Rev. E*, 86(011911), 2012. doi: 10.1103/PhysRevE.86.011911.

[74] Kurt Kremer Jonathan D. Halverson, Jan Smrek and Alexander Y. Grosberg. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Reports on Progress in Physics*, 77(2), 2014.

[75] N. Ramakrishnan, Kripa Gowrishankar, Lakshmi Kuttippurathu, P. B. Sunil Kumar, and Madan Rao. Active remodeling of chromatin and implications for in-vivo folding, 2015.

[76] Angelo Rosa and Ralf Everaers. Structure and dynamics of interphase chromosomes. *PLOS Computational Biology*, 4:e1000153(8), 2008. doi: 10.1371/journal.pcbi.1000153.

[77] Maxim V. Imakaev, Konstantin M. Tchourine, Sergei K. Nechaev, and Leonid A. Mirny. Effects of topological constraints on globular polymers. *Soft Matter*, 11:665–671, 2015. doi: 10.1039/C4SM02099E.

[78] Leonid A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.*, 19:37–51, 2011.

[79] Alexander Vologodskii. Theoretical models of dna topology simplification by type iia dna topoisomerases. *Nucleic Acids Res.*, 37(10):3125–3133, 2009. doi: 10.1093/nar/gkp250.

[80] M. Muller, J. P. Wittmer, and M. E. Cates. Topological effects in ring polymers: A computer simulation study. *Phys. Rev. E*, 53(5):5063–5074, 1996.

[81] A. Pombo and M. Nicodemi. Models of chromosome structure. *Current Opinion in Cell Biology,*, 28(90), 2014.

[82] Andrea M. Chiariello, Carlo Annunziatella, Simona Bianco, Andrea Esposito, and Mario Nicodemi. Polymer physics of chromosome large-scale 3d organisation. *Nature Scientific Reports*, 6(29775), 2016.

[83] R. K. Sachs, G. van den Engh, B. Trask, H. Yokota, , and J. E. Hearst. A random-walk/giant-loop model for interphase chromosomes. *PNAS U.S.A.*, 92(7):2710–2714, 1995.

[84] Davide Marenduzzo, Cristian Micheletti, and Peter R. Cook. Entropy-driven genome organization. *Biophysical Journal*, 90(10):3712–3721, 2006.

[85] Manfred Bohn and Dieter W. Heermann. Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS ONE*, 2010. doi: 10.1371/journal.pone.0012218.

[86] Daniel Jost, Pascal Carrivain, Giacomo Cavalli, and Cédric Vaillant. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucl Acids Res*, 42(15):9553–9561, 2014. doi: 10.1093/nar/gku698.

[87] Mariano Barbieri, Mita Chotalia, James Fraser, Liron-Mark Lavitas, Josée Dostie, Ana Pombo, and Mario Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *PNAS U.S.A.*, 109(40):16173–16178, 2012. doi: 10.1073/pnas.1204799109.

[88] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. Formation of chromosomal domains by loop extrusion. *Cell Reports*, 15(9):2038–2049, 2016. doi: 10.1016/j.celrep.2016.04.085.

[89] Kurt Kremer and Gary S. Grest. Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, April 1990. doi: 10.1063/1.458541. URL https://doi.org/10.1063/1.458541.

[90] Luca Giorgetti, Rafael Galupa, Elphège P. Nora, Tristan Piolot, France Lam, Job Dekker, Guido Tiana, and Edith Heard. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157(4), 2014. doi: 10.1016/j.cell.2014.03.025.

[91] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, and Nezar Abdennur. Formation of chromosomal domains by loop extrusion. *Cell Reports*, 15 (9):2038–2049, 2016.

[92] Anton Goloborodko, John F. Marko, and Leonid A. Mirny. Chromosome compaction by active loop extrusion. *Biophysical Journal*, 110(10):2162–2168, 2016.

[93] S. Jun and B. Mulder. Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proceedings of the National Academy of Sciences*, 103(33):12388–12393, August 2006. doi: 10.1073/pnas.0605305103. URL https://doi.org/10.1073/pnas.0605305103.

[94] Alexander Y. Grosberg. Extruding loops to make loopy globules? *Biophysical Journal*, 110(10):2133–2135, 2016.

[95] Andrea M. Chiariello, Carlo Annunziatel, Simona Bianco, Andrea Esposito, and Mario Nicodemi. Polymer physics of chromosome large-scale 3d organisation. *Scientific Reports*, 6(29775), 2016. doi: 10.1038/srep29775.

[96] Nick Gilbert, Marenduzzo, and Davide. Genome organization: experiments and modeling. *Chromosome Research*, 25(1):1–4, 2017. ISSN 1573-6849. doi: 10.1007/s10577-017-9551-2. URL http://dx.doi.org/10.1007/s10577-017-9551-2.

[97] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. Formation of chromosomal domains by loop extrusion. *Cell Reports*, 15(9):2038 – 2049, 2016. ISSN 2211-1247. doi: https://doi.org/10.1016/j.celrep.2016.04.085. URL http://www.sciencedirect.com/science/article/pii/S2211124716305307.

[98] Kazuhiro Maeshima, Saera Hihara, and Mikhail Eltsov. Chromatin structure: does the 30-nm fibre exist in vivo? *Current Opinion in Cell Biology*, 22(3):291 – 297, 2010. ISSN 0955-0674. doi: https://doi.org/10.1016/j.ceb.2010.03.001. URL http://www.sciencedirect.com/science/article/pii/S0955067410000256.

[99] Elnaz Alipour and John F. Marko. Self-organization of domain structures by dna-loop-extruding enzymes. *Nucleic Acids Research*, 40(22), 2012. doi: 10.1093/nar/gks925.

[100] Natalia Naumova, Maxim Imakaev, Geoffrey Fudenberg, Ye Zhan, Bryan R. Lajoie, Leonid A. Mirny, and Job Dekker. Organization of the mitotic chromosome. *Science*, 342(6161):948–953, 2013. ISSN 0036-8075. doi: 10.1126/science.1236083. URL http://science.sciencemag.org/content/342/6161/948.

[101] Jill M. Brackley, Chris A.and Brown, Dominic Waithe, Christian Babbs, James Davies, Jim R. Hughes, Veronica J. Buckle, and Davide Marenduzzo. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biology*, 17(1):59, 2016. doi: 10.1186/s13059-016-0909-0. URL https://doi.org/10.1186/s13059-016-0909-0.

[102] Nirmalendu Ganai, Surajit Sengupta, and Gautam I. Menon. Chromosome positioning from activity-based segregation. *Nucleic Acids Research*, 42(7):4145–4159, 2014. doi: 10.1093/nar/gkt1417. URL +http://dx.doi.org/10.1093/nar/gkt1417.

[103] Fabrizio Benedetti, Julien Dorier, and Andrzej Stasiak. Effects of supercoiling on enhancer–promoter contacts. *Nucleic Acids Research*, 42(16):10425–10432, 2014. doi: 10.1093/nar/gku759. URL +http://dx.doi.org/10.1093/nar/gku759.

[104] Angelo Rosa, Nils B. Becker, and Ralf Everaers. Looping probabilities in model interphase chromosomes. *Biophysical Journal*, 98(11):2410 – 2419, 2010. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2010.01.054. URL http://www.sciencedirect.com/science/article/pii/S0006349510002195.

[105] Suckjoon Jun and Andrew Wright. Entropy as the driver of chromosome segregation. *Nature Reviews Microbiology*, 8(8):600–607, August 2010. doi: 10.1038/nrmicro2391. URL https://doi.org/10.1038/nrmicro2391.

[106] Youngkyun Jung, Chanil Jeon, Juin Kim, Hawoong Jeong, Suckjoon Jun, and Bae-Yeun Ha. Ring polymers as model bacterial chromosomes: confinement, chain topology, single chain statistics, and how they interact. *Soft Matter*, 8(7):2095–2102, 2012. doi: 10.1039/c1sm05706e. URL https://doi.org/10.1039/c1sm05706e.

[107] Tyler N. Shendruk, Martin Bertrand, Hendrick W. de Haan, James L. Harden, and Gary W. Slater. Simulating the entropic collapse of coarse-grained chromosomes. *Biophysical Journal*, 108(4):810–820, February 2015. doi: 10.1016/j.bpj.2014.11.3487. URL https://doi.org/10.1016/j.bpj.2014.11.3487.

[108] Juin Kim, Chanil Jeon, Hawoong Jeong, Youngkyun Jung, and Bae-Yeun Ha. A polymer in a crowded and confined space: effects of crowder size and poly-dispersity. *Soft Matter*, 11(10):1877–1888, 2015. doi: 10.1039/c4sm02198c. URL https://doi.org/10.1039/c4sm02198c.

[109] Guillaume Le Treut, François Képès, and Henri Orland. A polymer model for the quantitative reconstruction of chromosome architecture from HiC and GAM data. *Biophysical Journal*, 115(12):2286–2294, December 2018. doi: 10.1016/j.bpj.2018. 10.032. URL https://doi.org/10.1016/j.bpj.2018.10.032.

[110] Amit Kumar and Debasish Chaudhuri. Cross-linker mediated compaction and local morphologies in a model chromosome. *Journal of Physics: Condensed Matter*, 31 (35):354001, June 2019. doi: 10.1088/1361-648x/ab2350. URL https://doi.org/10. 1088/1361-648x/ab2350.

[111] Marc Joyeux. Preferential localization of the bacterial nucleoid. *Microorganisms*, 7 (7):204, July 2019. doi: 10.3390/microorganisms7070204. URL https://doi.org/10. 3390/microorganisms7070204.

[112] Ajoy Maji, Jahir Ahmed, Shubhankar Roy, Buddhapriya Chakrabarti, and Mithun K. Mitra. A lamin associated chromatin model for chromosome organization. October 2019. doi: 10.1101/788497. URL https://doi.org/10.1101/788497.

[113] Michael Rubinstein and Ralph H. Colby. *Polymer physics*. Oxford University Press, Oxford, 2003.

[114] Alexander Yu. Grosberg and Alexei R. Khokhlov. *Statistical physics of macro-molecules*. AIP, New York, 1994.

[115] Pierre-Gilles de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, Ithaca, 1979.

[116] Nigel Goldenfeld. *Lectures On Phase Transitions And The Renormalization Group*. Perseus Books Publishing, Readig Massachussets, 1992.

[117] Masao Doi. *Introduction to Polymer Physics*. Oxford University Press, 1996.

[118] Alexander Y Grosberg and Alexei R Khokhlov. *Giant Molecules Here, There and Everywhere*. World Scientific Publishing., 2011.

[119] J.P. Cotton. Polymer excluded volume exponent v : An experimental verification of the n vector model for n = 0. *Journal de Physique Lettres*, 41(9):231–234, 1980. doi: 10.1051/jphyslet:01980004109023100. URL https://doi.org/10.1051/jphyslet: 01980004109023100.

[120] J.C. Le Guillou and J. Zinn-Justin. Accurate critical exponents from field theory. *Journal de Physique*, 50(12):1365–1370, 1989. doi: 10.1051/jphys: 0198900500120136500. URL https://doi.org/10.1051/jphys:0198900500120136500.

[121] Marvin Bishop and J. P. J. Michels. Scaling in two-dimensional linear and ring polymers. *The Journal of Chemical Physics*, 85(2):1074–1076, jul 1986. doi: 10. 1063/1.451300. URL https://doi.org/10.1063/1.451300.

[122] Javier Arsuaga, Reyka G. Jayasinghe, Robert G. Scharein, Mark R. Segal, Robert H. Stolz, and Mariel Vazquez. Current theoretical models fail to predict the topological complexity of the human genome. *Frontiers in Molecular Biosciences*, 2, August 2015. doi: 10.3389/fmolb.2015.00048. URL https://doi.org/10.3389/fmolb.2015.00048.

[123] Mark A. Umbarger, Esteban Toro, Matthew A. Wright, Gregory J. Porreca, Davide Baù, Sun-Hae Hong, Michael J. Fero, Lihua J. Zhu, Marc A. Marti-Renom, Harley H. McAdams, Lucy Shapiro, Job Dekker, and George M. Church. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Molecular Cell*, 44(2):252–264, October 2011. doi: 10.1016/j.molcel.2011.09.010. URL https: //doi.org/10.1016/j.molcel.2011.09.010.

[124] Jaeoh Shin, Andrey G Cherstvy, and Ralf Metzler. Mixing and segregation of ring polymers: spatial confinement and molecular crowding effects. *New Journal of Physics*, 16(5):053047, 2014. URL http://stacks.iop.org/1367-2630/16/i=5/a=053047.

[125] M. Kojima, K. Kubo, and K. Yoshikawa. Elongation/compaction of giant DNA caused by depletion interaction with a flexible polymer. *The Journal of Chemical Physics*, 124(2):024902, jan 2006. doi: 10.1063/1.2145752. URL https://doi.org/10.1063/1. 2145752.

[126] Zimmerman Steven B. and Murphy Lizabeth D. Macromolecular crowding and the mandatory condensation of dna in bacteria. *FEBS Letters*, 390(3):245–248. doi: 10.1016/0014-5793(96)00725-9. URL https://febs.onlinelibrary.wiley.com/doi/abs/10. 1016/0014-5793%2896%2900725-9.

[127] Sean R. McGuffee and Adrian H. Elcock. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Computational Biology*, 6(3):e1000694, March 2010. doi: 10.1371/journal.pcbi.1000694. URL https://doi.org/10.1371/journal.pcbi.1000694.

[128] Axel Arnold, Behnaz Bozorgui, Daan Frenkel, Bae-Yeun Ha, and Suckjoon Jun. Unexpected relaxation dynamics of a self-avoiding polymer in cylindrical confinement. *The Journal of Chemical Physics*, 127(16):164903, oct 2007. doi: 10.1063/1.2799513. URL https://doi.org/10.1063/1.2799513.

[129] Youngkyun Jung, Suckjoon Jun, and Bae-Yeun Ha. Self-avoiding polymer trapped inside a cylindrical pore: Flory free energy and unexpected dynamics. *Physical Review E*, 79(6), jun 2009. doi: 10.1103/physreve.79.061912. URL https://doi.org/10.1103/ physreve.79.061912.

[130] Hongsuk Kang, Young-Gui Yoon, D. Thirumalai, and Changbong Hyeon. Confinement-induced glassy dynamics in a model for chromosome organization. *Phys. Rev. Lett.*, 115:198102, Nov 2015. doi: 10.1103/PhysRevLett.115.198102. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.198102.

[131] William C. Hacker, Shuxiang Li, and Adrian H. Elcock. Features of genomic organization in a nucleotide-resolution molecular model of the escherichia coli chromosome. *Nucleic Acid Research*, 45(13):7541–7554, 2017. doi: 10.1093/nar/gkx541.

[132] Bae-Yeun Ha and Youngkyun Jung. Polymers under confinement: single polymers, how they interact, and as model chromosomes. *Soft Matter*, 11(12):2333–2352, 2015. doi: 10.1039/c4sm02734e. URL https://doi.org/10.1039/c4sm02734e.

[133] Yanwei Wang, Douglas R. Tree, and Kevin D. Dorfman. Simulation of DNA extension in nanochannels. *Macromolecules*, 44(16):6594–6604, aug 2011. doi: 10.1021/ ma201277e. URL https://doi.org/10.1021/ma201277e.

[134] Liang Dai, Johan van der Maarel, and Patrick S. Doyle. Extended de gennes regime of DNA confined in a nanochannel. *Macromolecules*, 47(7):2445–2450, mar 2014. doi: 10.1021/ma500326w. URL https://doi.org/10.1021/ma500326w.

[135] Andrey Milchev. Single-polymer dynamics under constraints: scaling theory and computer experiment. *Journal of Physics: Condensed Matter*, 23(10):103101, feb 2011. doi: 10.1088/0953-8984/23/10/103101. URL https://doi.org/10.1088/0953-8984/23/10/103101.

[136] Tejal Agarwal, G P Manjunath, Farhat Habib, Pavana Lakshmi Vaddavalli, and Apratim Chatterji. Role of special cross-links in structure formation of bacterial dna polymer. *Journal of Physics: Condensed Matter*, 30(3):034003, 2018. URL http://stacks.iop.org/0953-8984/30/i=3/a=034003.

[137] Tejal Agarwal, G. P. Manjunath, Farhat Habib, and Apratim Chatterji. Origin of spatial organization of dna-polymer in bacterial chromosomes. *EPL (Europhysics Letters)*, 121(1):18004, 2018. URL http://stacks.iop.org/0295-5075/121/i=1/a=18004.

[138] Tejal Agarwal, G. P. Manjunath, Farhat Habib, and Apratim Chatterji. Bacterial chromosome organization. i. crucial role of release of topological constraints and molecular crowders. *The Journal of Chemical Physics*, 150(14):144908, April 2019. doi: 10.1063/1.5058214. URL https://doi.org/10.1063/1.5058214.

[139] Tejal Agarwal, G. P. Manjunath, Farhat Habib, and Apratim Chatterji. Bacterial chromosome organization. II. few special cross-links, cell confinement, and molecular crowders play the pivotal roles. *The Journal of Chemical Physics*, 150(14):144909, April 2019. doi: 10.1063/1.5058217. URL https://doi.org/10.1063/1.5058217.

[140] Samuel Karlin, Jan Mrázek, Allan Campbell, and Dale Kaiser. Characterizations of highly expressed genes of four fast-growing bacteria. *Journal of Bacteriology*, 183 (17), 2001. doi: 10.1128/JB.183.17.5025-5040.2001.

[141] https://www.ncbi.nlm.nih.gov.

[142] Jack A. Bryant, Laura E. Sellars, Stephen J. W. Busby, and David J. Lee. Chromosome position effects on gene expression in escherichia coli k-12. *Nucleic Acids Research*, 42(18), 2014. doi: 10.1093/nar/gku828.

[143] A. Baumgartner and K. Binder. Monte carlo studies on the freely jointed polymer chain with excluded volume interaction. *The Journal of Chemical Physics*, 71(2541), 1979. doi: 10.1063/1.438608.

[144] Hongsuk Kang, Philip A. Pincus, Changbong Hyeon, and D. Thirumalai. Effects of macromolecular crowding on the collapse of biopolymers. *Physical Review Letters*, 114(6), feb 2015. doi: 10.1103/physrevlett.114.068303. URL https://doi.org/10.1103/physrevlett.114.068303.

[145] Hongsuk Kang, Ngo Minh Toan, Changbong Hyeon, and D. Thirumalai. Unexpected swelling of stiff DNA in a polydisperse crowded environment. *Journal of the American Chemical Society*, 137(34):10970–10978, aug 2015. doi: 10.1021/jacs.5b04531. URL https://doi.org/10.1021/jacs.5b04531.

[146] H. Li and R. Durbin. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics,*, 26(5):589–595, 2010.

Appendix A

Spatiotemporal organization of *E. coli* chromosome: a experimental study

A.1   Introduction

Here, I present the experimental study of chromosome organization of bacteria *E. coli* during multifork replication. I work with the different bacterial strains of *E. coli* in which fluorescent proteins tag different segments of the chromosome. Then, we analyze the positions of different sections marked by fluorescent proteins along the long and short axis of the cylindrical cell using fluorescent microscopy in a large number of individual cells ($\sim 10^6$). The long and short axis distributions of the positions of different loci in the population of cells reveal the well-defined spatiotemporal organization of the chromosome.

I start with the experimental procedure and then present the results. Firstly, I show the results of the spatiotemporal organization of *E. coli* chromosome for two overlapping cell-cycles, which I have reproduced by repeating the experiments of the paper [22]. Next, I show the experimental results for the chromosome organization of bacteria *E. coli* in faster growth conditions with three overlapping cell-cycles.

A.2   Experimental procedure

A.2.1   Strains

I work with 3 sets of strains of bacteria *E. coli* (derivative of strain MG1655) in which the two positions on the chromosome are tagged by two independent fluorescent markers used in the paper[22]. The two positions of the fluorescent markers are at $22'$ (reference marker on the right arm of the chromosome) and at one of the three positions, including Ori and
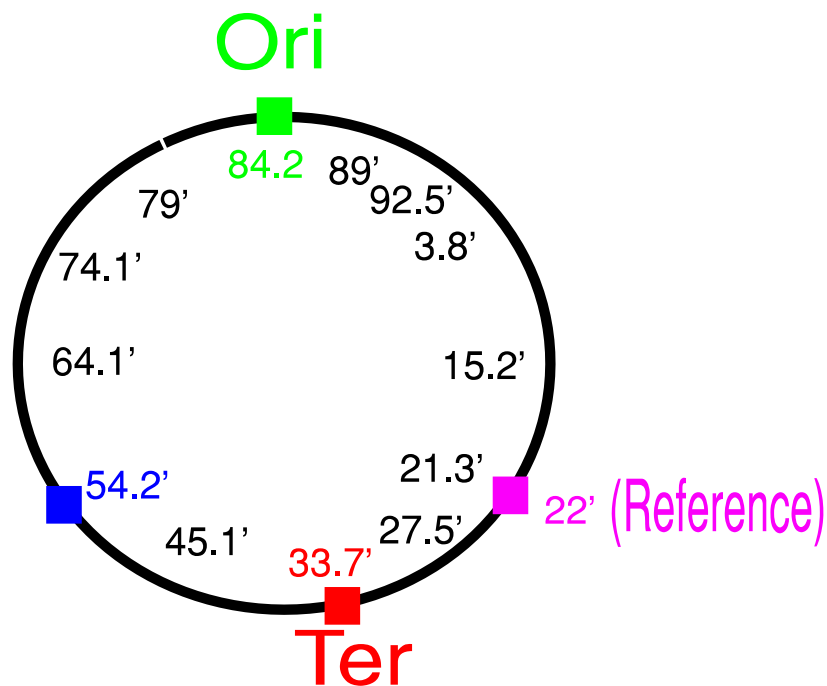
Fig. A.1 The figure shows the positions of the fluorescent markers around the genetic map of *E. coli* chromosome, used in this study. We work with 3 different strains of the *E. coli*. In each of the strain, the chromosome is marked at two positions: reference position $22'$ (pink) and one of the 3 positions marked as red, blue and green in the schematic, repectively.

Ter, distributed around the genetic map of the chromosome, respectively (see Fig.A.1). The strains have parS sequence (pMT1-parS) at $22'$ (reference marker), and another type of parS sequence (P1-parS) at one of the 13 positions on the genetic map. The parS sequences are the palindromic sequences in the chromosome of bacteria which with other two proteins parA and parB DNA binding proteins, control the mechanism of chromosome segregation in bacteria. Each strain contains a pair of parS sequence and a matching pair of ParB proteins fused to Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) (CFP-P1DOParB and yGFP-pMT1D23ParB) [22].

A.2.2 Growth media

We performed the experiments for the *E. coli* strains in two different growth media with cell doubling time 55 minutes (slow growth) and 22.5 minutes (fast growth) at 30° and 37° Celcius, respectively [22, 49]. We grow the cells in AB minimal media supplemented with 0.2% glucose, $1\mu g/mL$ thiamin, 0.05% Casamino acids, $100\mu g/mL$ ampicillin and $1\mu g/mL$ uracil at 30° Celcius with constant shaking for the aeration with cell doubling time

$\sim$ 55 minutes [22]. For 22.5 minutes doubling time, the cells were grown in mops based synthetic rich media with 0.2% glucose as a carbon source and supplements given in [49] at 37° Celcius.

A.2.3   Sample preparation

The sample for the microscopy is prepared in 3 steps. Firstly, one particular cell strain (taken from $-80°$ Celcius glycerol stock) is grown in LB agar plates overnight at 37° degree Celcius. For the seed culture, the cells (in a single colony) from agar plate is inoculated in the LB media and cultured overnight at 30° or 37° for slow or fast growth conditions, respectively. Then the cells are back diluted 1000 folds into the $1-2mL$ experimental growth media (AB minimal or mops synthetic rich) and are grown at 30° or 37° Celius with constant shaking. The cells are taken out from the culture at optical density $OD \sim 0.1$. I pipette $2\mu L$ of the cells from the cell-culture on the agarose pad (0.2% w/v agarose). Next, the flipped agarose pad is put on the wilco disc and covered with the glass coverslip for the microscopy.

A.2.4   Microscopy and image acquisition

We use Nikon Ti-E inverted microscope with perfect focus (PFS 2), 60× oil immersion objective, light-emitting diode (LED) fluorescent light (Lumencor), and Andor NEO sCMOS camera (Andor Technology) for the transmission and fluorescent microscopy [49] (Jun lab, UCSD, USA). The typical field of view of the imaged cells is shown in Fig.A.2. We set the exposure time to be 200ms with the full transmission. For each set, we analyze $5000-25000$ cells to get statistically significant data.

A.2.5   Image analysis

The image analysis is done by the custom made software in python using the opencv2 library. The procedure includes subtracting the background noise from the black and white phase-contrast image. Then search for the dark objects on the bright background using the *connected_components* function of opencv2 library. Next, fit the rectangle on the searched object to obtain the length of the cell. We use the blob detection function, namely Laplacian of Gaussian (LOG) of the Sci-kit image library of python, to search for the fluorescent foci (bright foci on dark background) inside the cell. After that, the longitudinal and radial positions of the foci can be measured along the cylindrical cell's poles and axis.
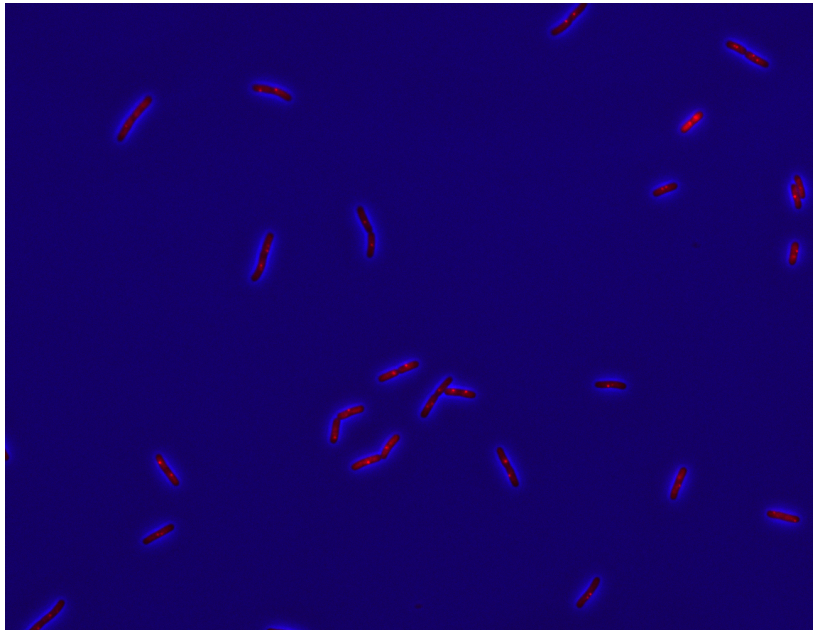
Fig. A.2 The image show the typical field of view of the imaged *E. coli* cells. The red color shows the position of the reference marker (marked with yGFP) at $22'$ in the cells grown in mops rich media at $37°$ degree celcius (doubling time ~ 23 minutes).

A.3　Results

I first present the results for the chromosome organization in the case of two overlapping cell-cycles and then show the data for the spatiotemporal organization in three overlapping cell-cycles cases.

A.3.1　Spatiotemporal organization of the chromosome during two overlapping cell cycles

From the analysis of the phase-contrast images, we can obtain the lengths of the individual cells in the population. Thus, I first present the length distribution of the cells in the three strains, in which the chromosome is marked at Ori, Ter, and $54'$ position in the genetic map of the *E. coli* in Fig.A.3. The x-axis in the figures of Fig.A.3 represents the cell length, and the y-axis shows the probability $p(l)$ of the cell having length l, i.e., the number of cells having length l in the population divided by the total number of cells in the population. From Fig. A.3 we see that the three different strains have similar length distribution, which ensures the reproducibility of the data.
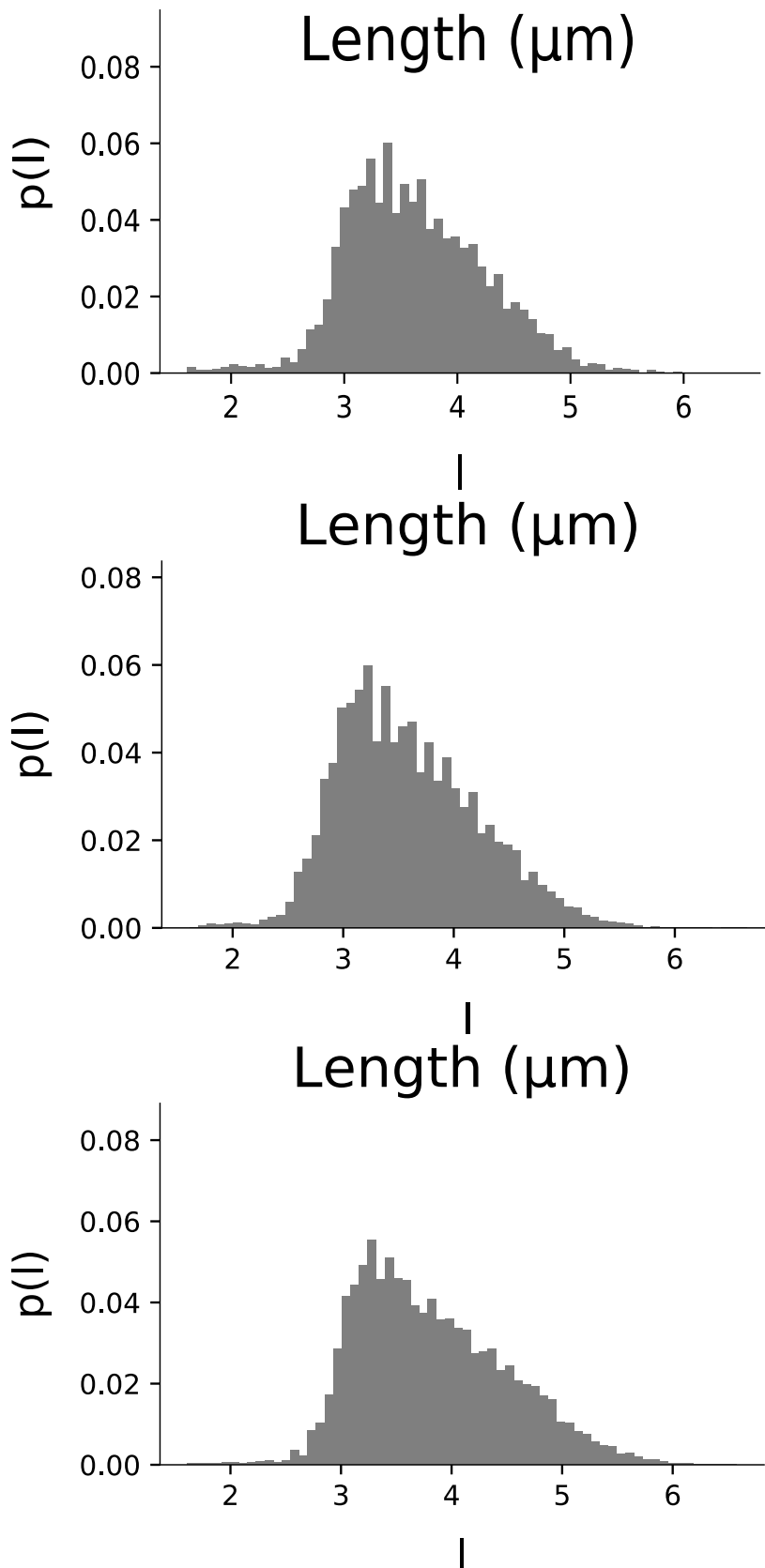
Fig. A.3 The figures show the length distribution of the cells in three bacterial strains. In three strains, the chromosome is marked at the Ori (Top figure), Ter (Middle figure), and $54'$ (Bottom figure) positions, respectively by the CFP and at $22'$ reference position by GFP.

We also see that most of the cells in the distribution has a length equal to 3μm (newborn cells), and very fewer cells have a length of about 6μm (the old cells which are about to divide). It is because, at any time in the cell population, the number of newborn cells is more than the number of old cells in the exponentially growing population. The short tail on the left of the length distribution represents the stochasticity in the birth size of the cell. We next calculate the distributions of different fluorescent foci, namely, Ori, Ter, and 54' along the length and radial axis of the cylindrical cell from the analysis of the fluorescent images. To plot the positional distribution of the cells along the cell's length, we sort the cells according to the normalized cell's age and the number of foci present in the cell. The normalized age is approximated from the cell's length, assuming the linear relationship between the two, i.e., the longest cell in the population is the oldest cell, which is about to divide and the shortest cell in the population is newly born.

$$t_a = \frac{l - l_b}{l_o - l_b}$$

Here, l_b is the length of the newly born cell, $l_o = 2l_b$ is the length of the oldest cell, which is about to divide, and t_a represents the normalized cell age. We chose the value of $l_b \sim 3\mu$m from the peak in the length distribution.

For the positional distribution of different loci along the length of the cell we plot the distribution of relative position of the foci from one of the cell pole similar to [22], i.e., the distribution of $\delta l/l$, where δl is the distance of the foci from the pole of the cell and l is the length of the cell. The length distribution of different foci is shown in Fig.A.4, Where green, blue, and the red color represents the distributions of Ori, 54', and Ter foci, respectively. We use the same color scheme to represent the statistical data for different foci throughout the text. From the figure, we observe that for the newborn cells (age between 0-0.2), most of the cells have one Ter focus, which is found in the middle of the cell. While most of the cells are born with 2 Oris at quarter-quarter positions along the length of the cell, also, the position of the 54' marker in most of the newly born cells can be seen as the two resolved peaks near the cell center in the histograms of Fig.A.4. As time progresses, DNA replicates (thus each focus gets doubled), and we observe that the one peak distribution of Ter focus becomes two-peak distribution with each peak lies near the cell center. Similarly, the two peak distribution of Ori becomes four peaks, where each peak is located at $(1/8)^{th}, (3/8)^{th}, (5/8)^{th}$ and $(7/8)^{th}$ positions along the length of the cell. It indicates that the chromosome shows a well-defined organization along the length of at different stages of the cell cycle consistent with the results in the paper [22].
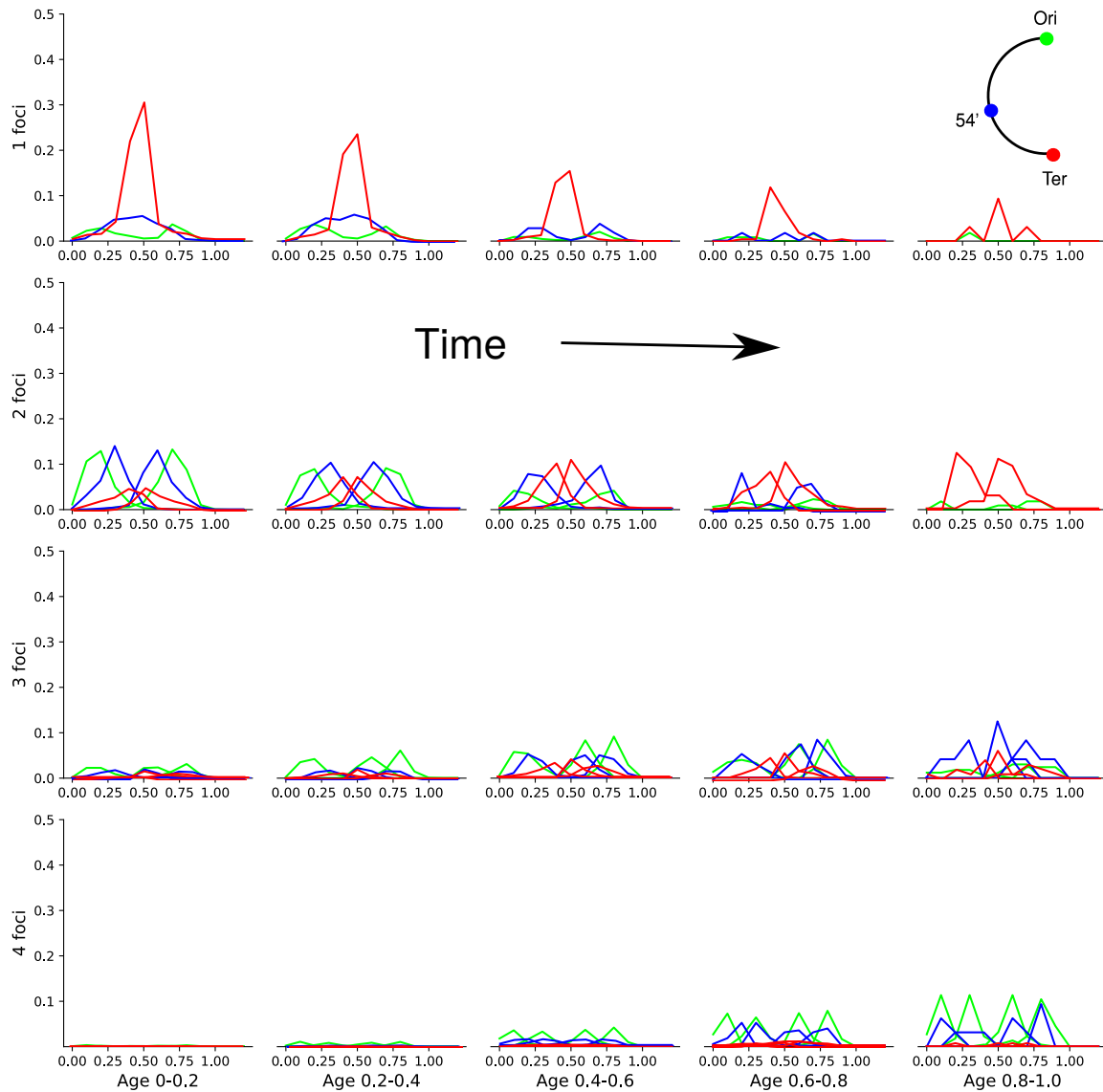
Fig. A.4 The figure shows the positional distributions of the different viz. Ori, Ter, and 54' along the cell's long axis. The histograms are sorted according to the age of the cell and the number of foci in the cell. The green, red, and blue color represents the distributions of Ori, Ter, and 54' marker, respectively. The time progresses towards the right in the figure. The x-axis in each histogram represents the relative length of the foci from the cell's pole, and the y-axis represents the probability of the foci to be found at some relative length from the cell's pole. In the length distribution of different foci we see that the green (ORI) and blue (54') lines are missing in the first row while only the red lines (TER) are present. The absence of blue, green lines and the presence of the red line in the first row of the figure suggest that the most of the cells are born with two ORI, two foci marked at 54' position but only one TER foci. We also see that with time one/two peak distribution becomes two/four peak distribution indicating the replication of the chromosomal foci with time.
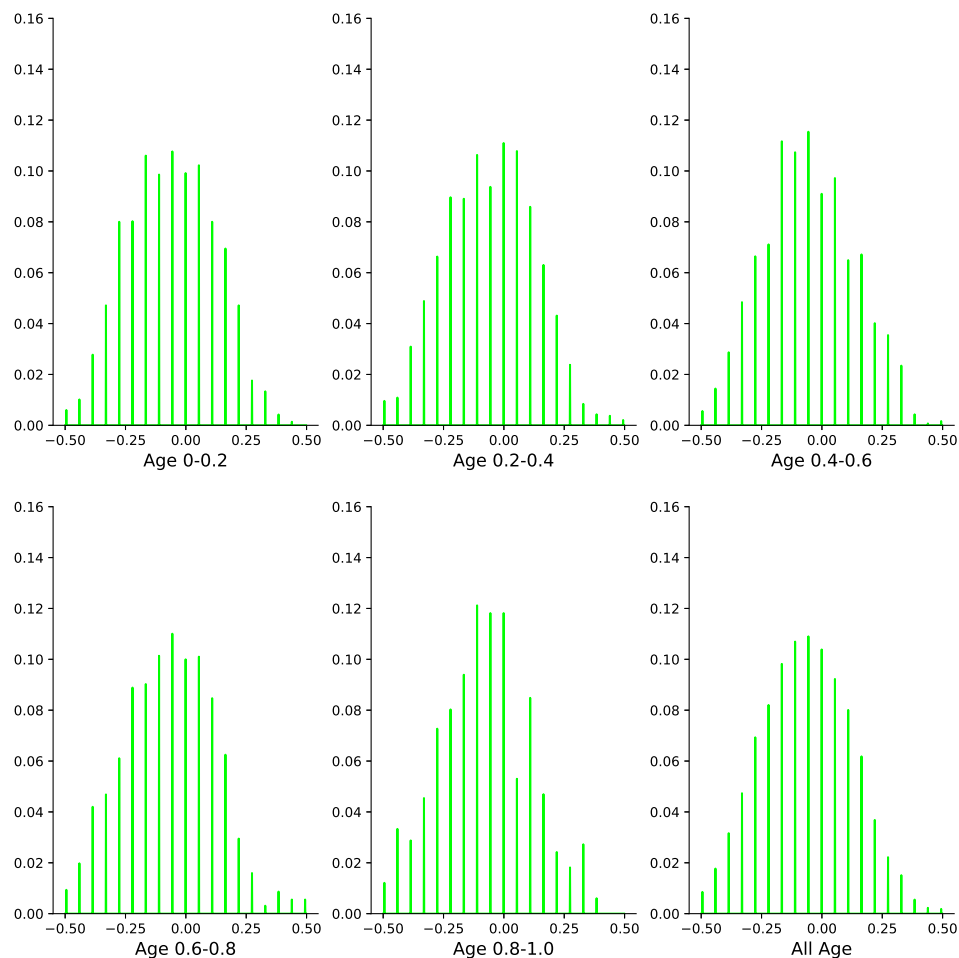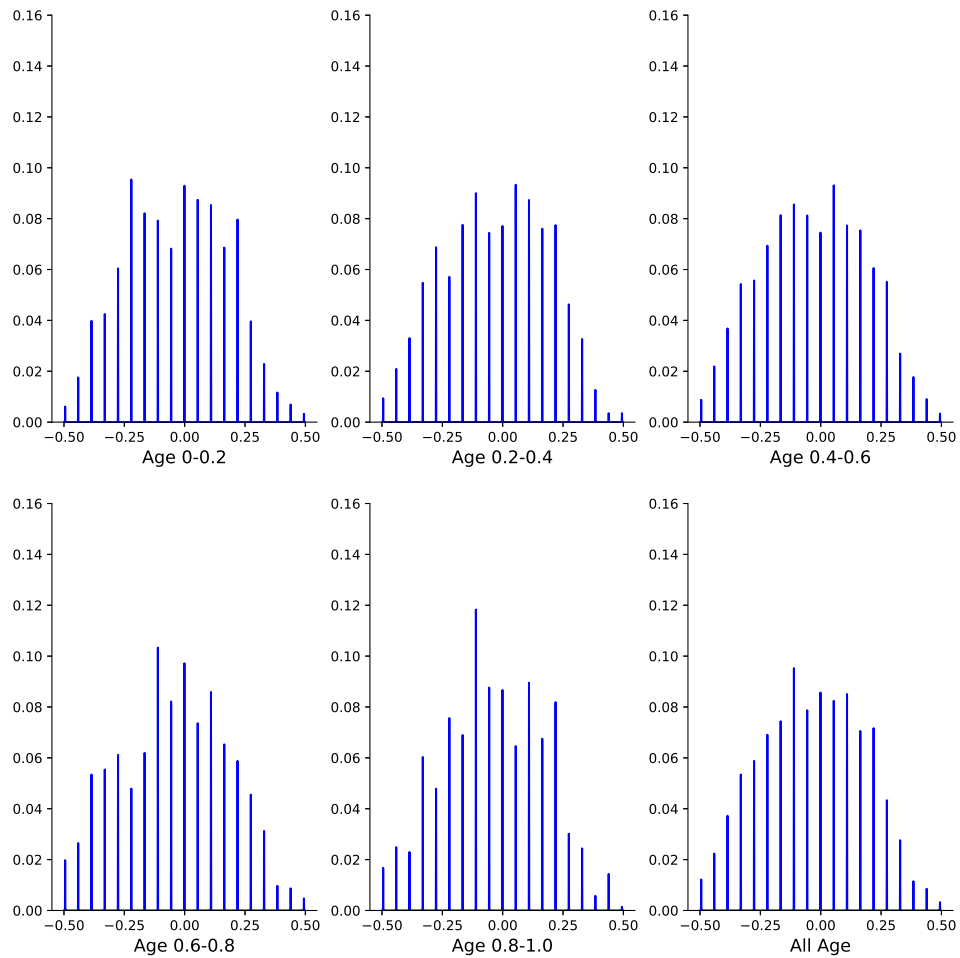
Fig. A.5 The figure shows the positional distribution of Ori along the radial axis of the cell. The x-axis represents the radial axis of the cell, and the y-axis represents the probability of the foci to be located at some position along the radial axis. To calculate the radial distribution of the foci we consider the axis of the cylinder as the origin. The foci which are present in the one half of the cell takes negative values of the distance along the radial axis and the foci which are present in another half of the cell take the positive value of distance along the radial axis. The different histograms are for different time (ages) over the cell cycle. The last histogram represents the average radial distribution of Ori for all the ages.
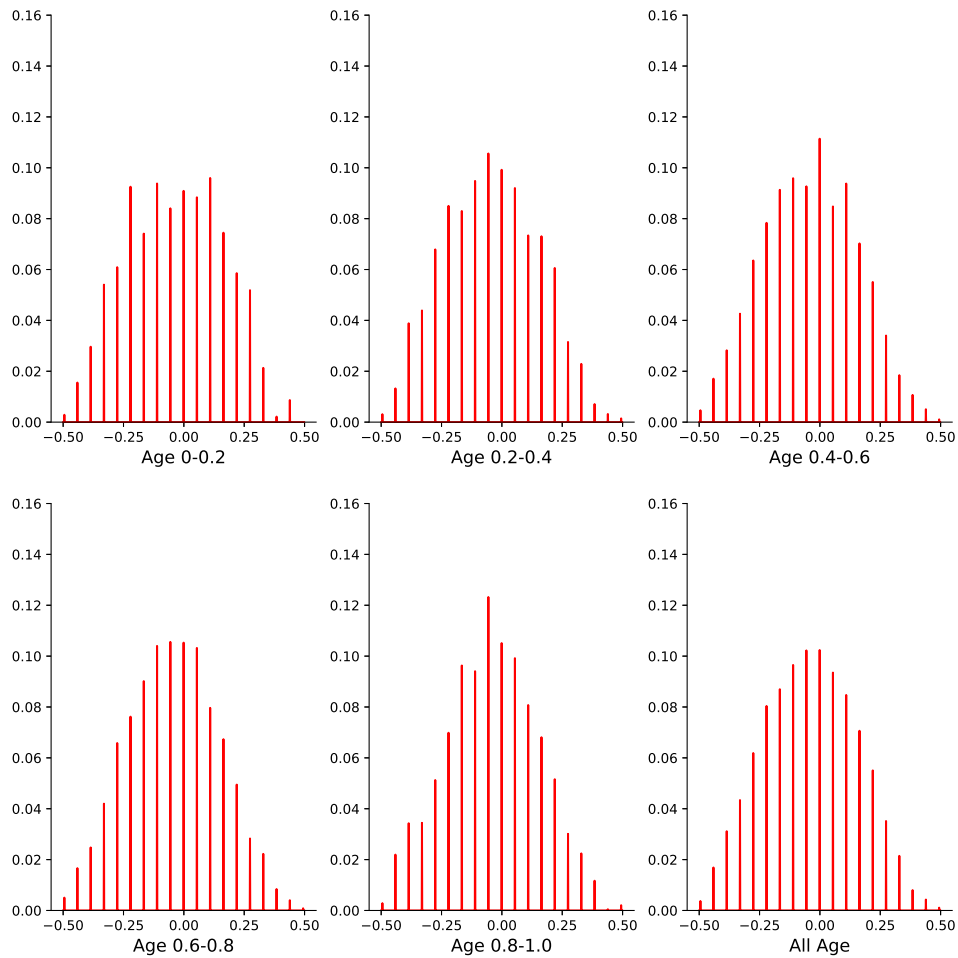
Fig. A.6 The figure shows the positional distribution of 54' along the radial axis of the cell. The x-axis represents the radial axis of the cell, and the y-axis represents the probability of the foci to be located at some position along the radial axis. The different histograms are for different time (ages) over the cell cycle. The last histogram represents the average radial distribution of 54' for all the ages.

Fig. A.7 The figure shows the positional distribution of Ter along the radial axis of the cell. The x-axis represents the radial axis of the cell, and the y-axis represents the probability of the foci to be located at some position along the radial axis. The different histograms are for different time (ages) over the cell cycle. The last histogram represents the average radial distribution of Ter for all the ages.

We next plot the positional distribution of different foci along the radial axis of the cylinder. It is shown in Figs. A.5, A.6, A.7 for foci Ori, 54' and Ter, respectively at different ages of cells. The x-axis in the radial positional distribution represents the position of the foci along the radial axis of the cell. The y-axis represents the probability of finding the foci at some position along the radial axis of the cell.

From the plots A.5, A.6, A.7, we see that the foci 54' shows the two peak distribution and the foci 'Ori' shows one peak distribution for all ages of the cell. We also observe that the 'Ter' foci show a flat peak distribution at age 0-0.2, and this distribution changes to sharp peak distribution with over the period of the cell cycle. The two peak distribution of 54' foci suggests that the chromosome segment corresponding to the 54' foci lies either in the right or left cell halves along the radial axis of the cell while the single peak distribution of Ori indicates that the Ori is found near the axis of the cylindrical cell. For Ter foci, we observe single peak distribution with a flat top for the newly born cells with age 0-0.2, which suggests that the Ter foci are also found in the left and right halves along the axis of the cylinder. Later in the cell cycle, Ter foci are located near the cell center, resulting in the sharp single peak distribution at the cell's center. The results of the radial distribution of foci match with the results of the paper [22].

A.3.2   Spatiotemporal organization of the chromosome during three overlapping cell cycles

To obtain the spatiotemporal organization of the *E. coli* chromosome during three overlapping cell cycle, I will calculate the same statistical quantities, viz., the length distribution of the cells, and positional distribution along the length and radial axis of the cell. The length distributions of the cells from three strains corresponding to the marker positions at Ori, 54', and Ter are shown in Fig. A.8. From the figure, we see that the length distributions from three strains look statistically similar, with mean 5.36μm, 5.21μm, & 5.21μm and standard deviation 1.03μm, 1.07μm, & 1.05μm, respectively. It again gives us confidence that the data is reproducible as the three distributions match with each other statistically.

I next plot the positional distribution of different foci along the length of the cell in Fig. A.9, A.10, and A.11 for Ori, 54' and Ter strains, respectively. As before, I calculate the relative position of the foci along the cell's length and plot the distribution of foci's position along the cell length. I calculate the cell's age from the cell's length, assuming the linear relationship between the length and age, as in A.3.1. Here, we take the value of l_b to be $l_b \sim 4.7\mu$m from the peak of the distributions. The histograms in one particular column correspond to the cells of the same age, and different columns represent the cells at different stages of the cell cycle. The x-axis in each histogram represents the relative position of the foci along the cell's longitudinal axis, and the y-axis shows the probability of the foci to be found at a particular relative position along the cell's length. From Fig. A.9, we see in most of the newly born cells (age 0-0.2) have four Ori foci with relative positions at $1/8^{th}$, $3/8^{th}$, $5/8^{th}$, and $7/8^{th}$ along the cell's length, respectively.

The one and two peak distributions of Ori foci in the newly born cell is the artifact of the bad signal to noise ratio in the fluorescent images. We also see from Fig. A.9 that as the cell cycle progresses the 4 Ori foci duplicate and become 8 at the end of the cell cycle, and it acquires the positions $1/32, 3/32, ...$ along the length of the cell as can be seen from the 8 equally spaced peaks in the positional distribution of foci along the length of the cell. We can also check from Fig. A.10 and A.11 that most of the newly born cells have only 2, 54' foci and 1 or 2 Ter foci at quarter positions and at middle of the cell, respectively. At the later stage of the cell cycle first, 54' foci duplicate, and we start getting the 4 peak distribution of 54' foci, and the foci are located at octa positions along the cell length. Next, the Ter focus duplicates, and we get the 2 or 4 peak distribution of Ter foci, which acquire the quarter-quarter position along the cell length at the end of the cell cycle. From the Figs. A.10 and A.11, we see that the rows 6,7 and 8 are mostly empty. This suggests that the at any point of time most of the cells have atmost 5 foci (marked at 54') present.
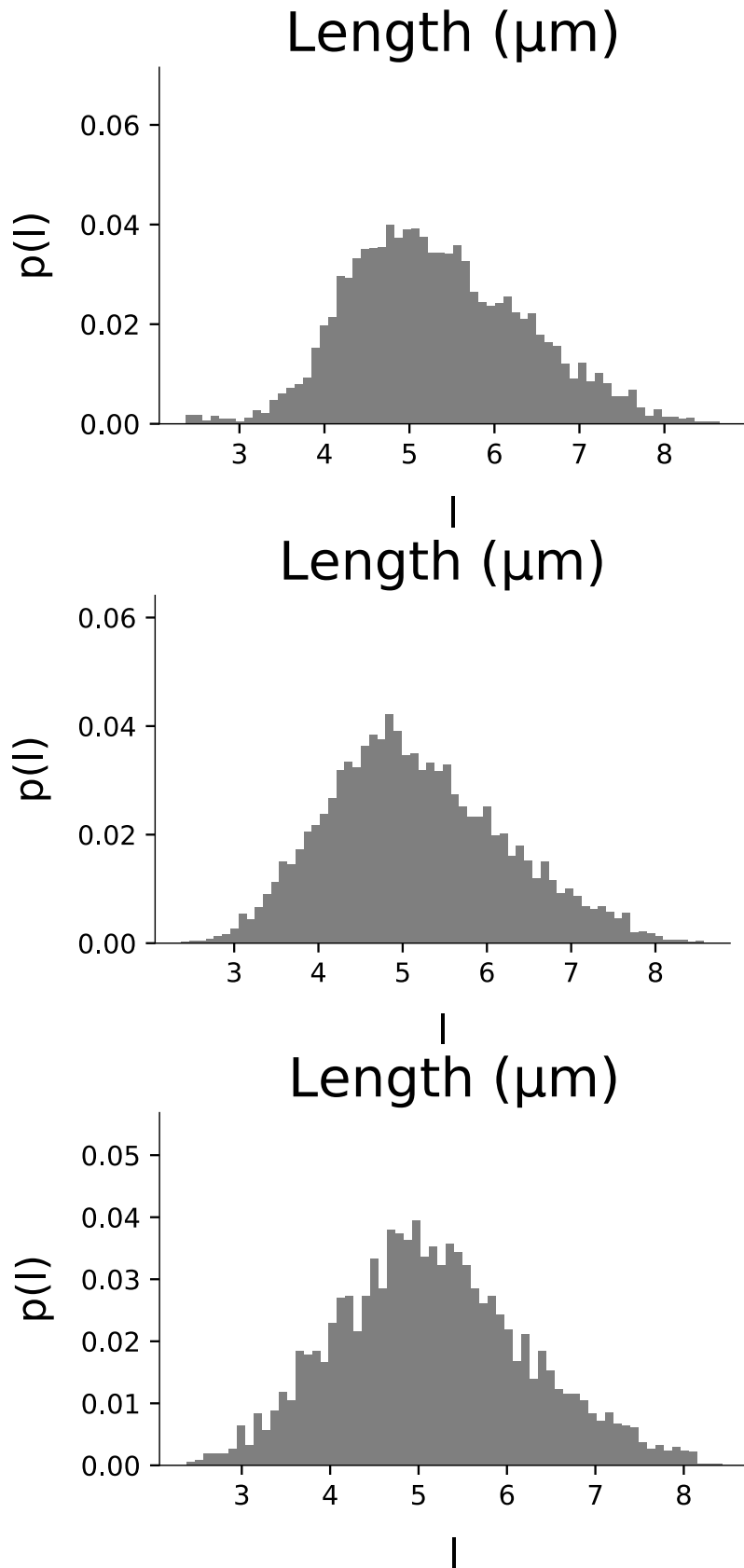
Fig. A.8 The figures show the length distribution of the cells in three bacterial strains. In three strains, the chromosome is marked at the Ori, Ter, and 54$'$ positions, respectively by the CFP and at 22$'$ reference position by GFP.
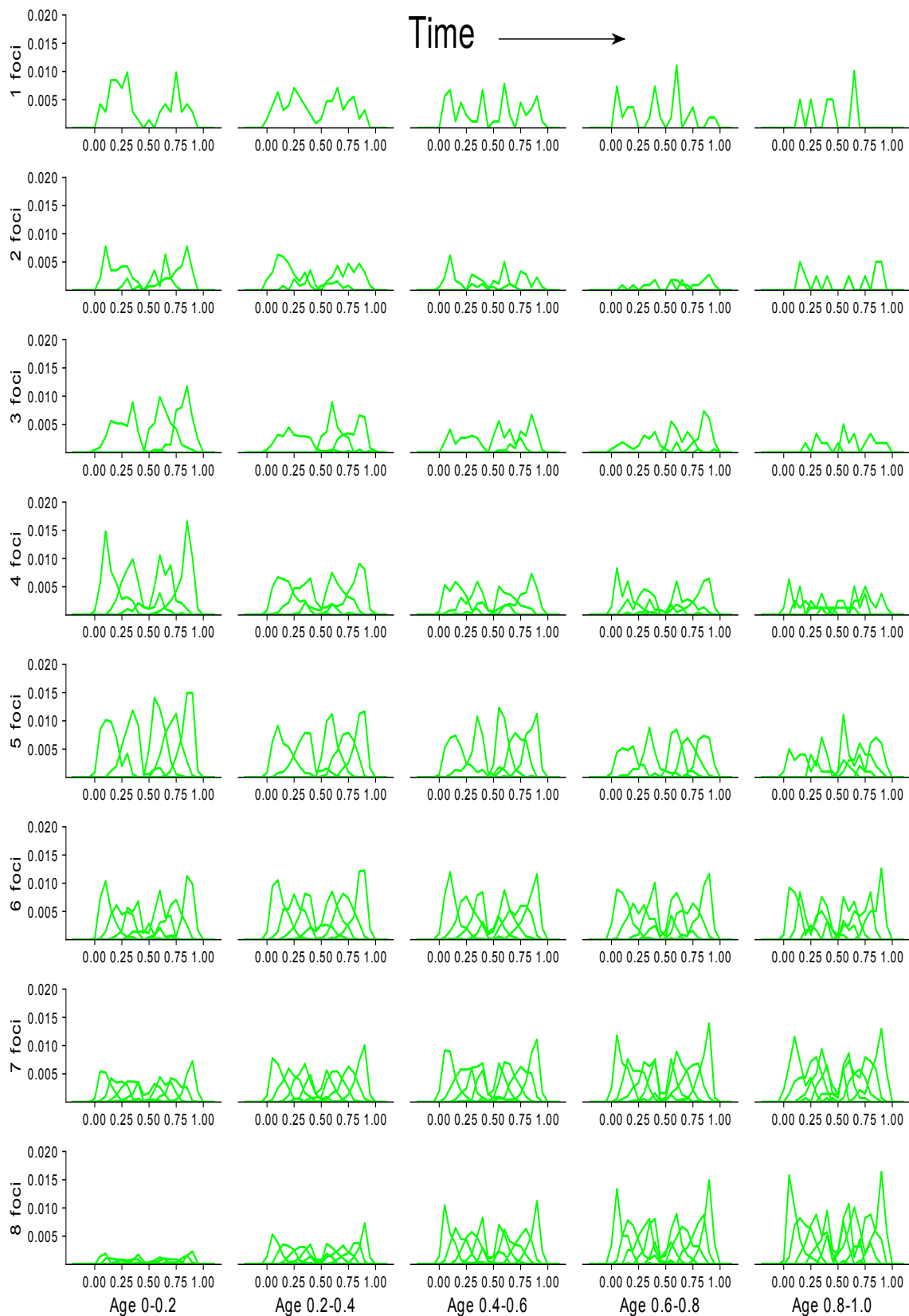
Fig. A.9 The figure shows the positional distributions of thre Ori along the cell's long axis. The histograms are sorted according to the age of the cell and the number of foci in the cell. The time progresses towards the right in the figure. The x-axis in each histogram represents the relative length of the foci from the cell's pole, and the y-axis represents the probability of the foci to be found at some relative length from the cell's pole.
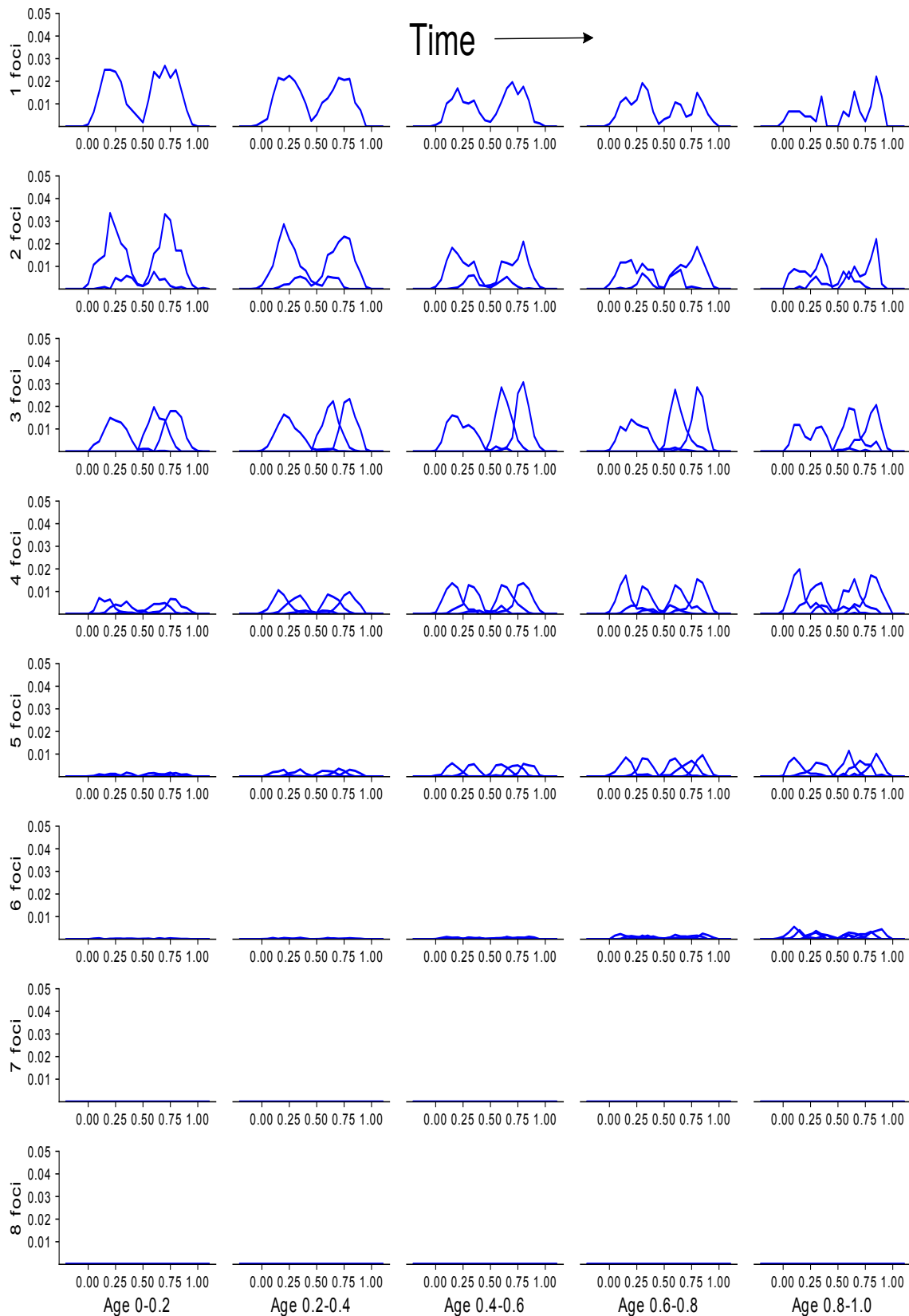
Fig. A.10 The figure shows the positional distributions of the 54' foci along the cell's long axis. The histograms are sorted according to the age of the cell and the number of foci in the cell. The time progresses towards the right in the figure. The x-axis in each histogram represents the relative length of the foci from the cell's pole, and the y-axis represents the probability of the foci to be found at some relative length from the cell's pole.
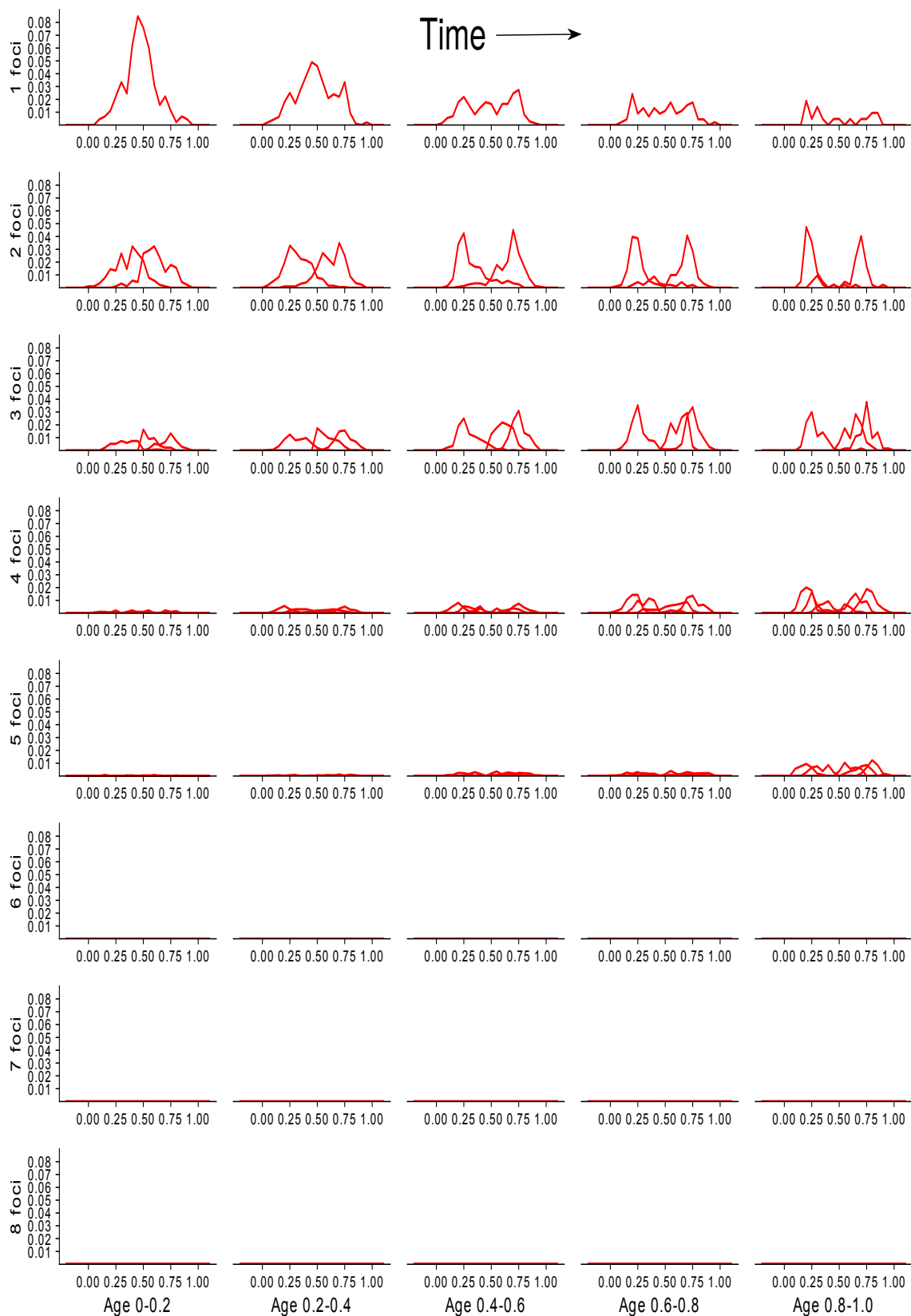
Fig. A.11 The figure shows the positional distributions of chrome Ter along the cell's long axis. The histograms are sorted according to the age of the cell and the number of foci in the cell. The time progresses towards the right in the figure. The x-axis in each histogram represents the relative length of the foci from the cell's pole, and the y-axis represents the probability of the foci to be found at some relative length from the cell's pole.
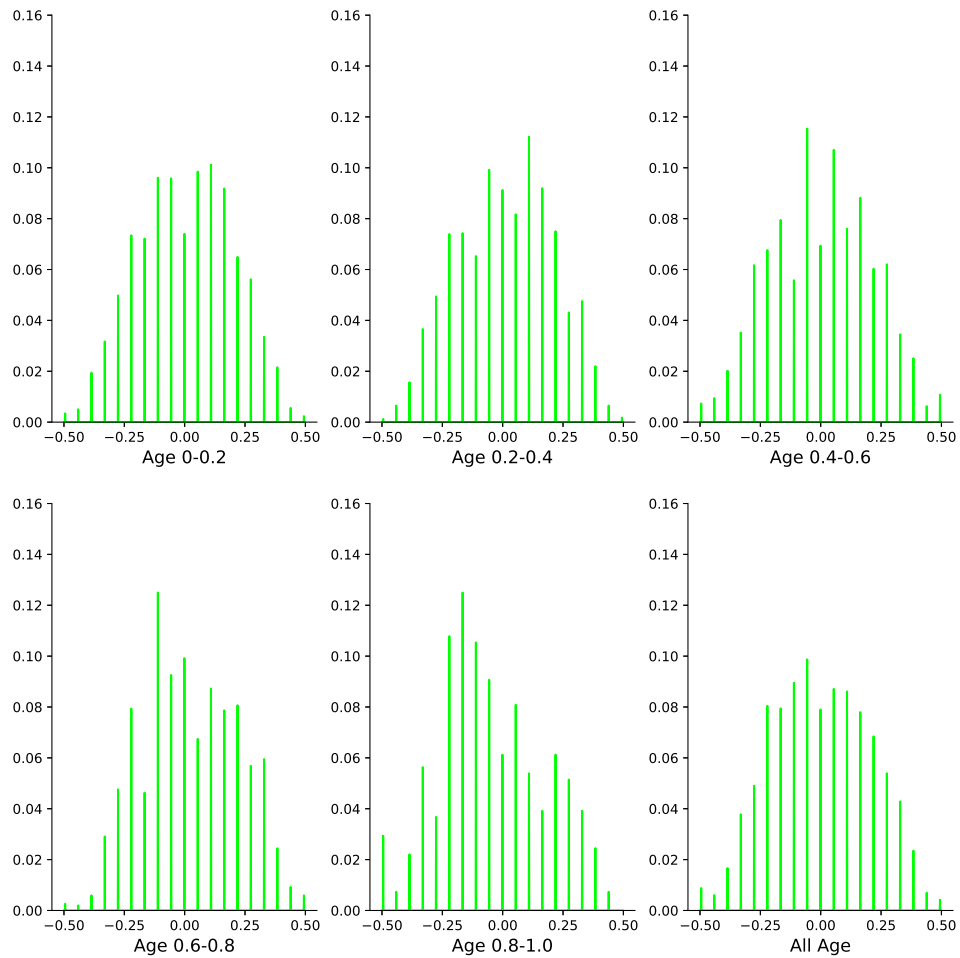
Fig. A.12 The figure shows the positional distribution of Ori along the radial axis of the cell. The x-axis represents the radial axis of the cell, and the y-axis represents the probability of the foci to be located at some position along the radial axis. The different histograms are for different time (ages) over the cell cycle. The last histogram represents the average radial distribution of Ori for all the ages.
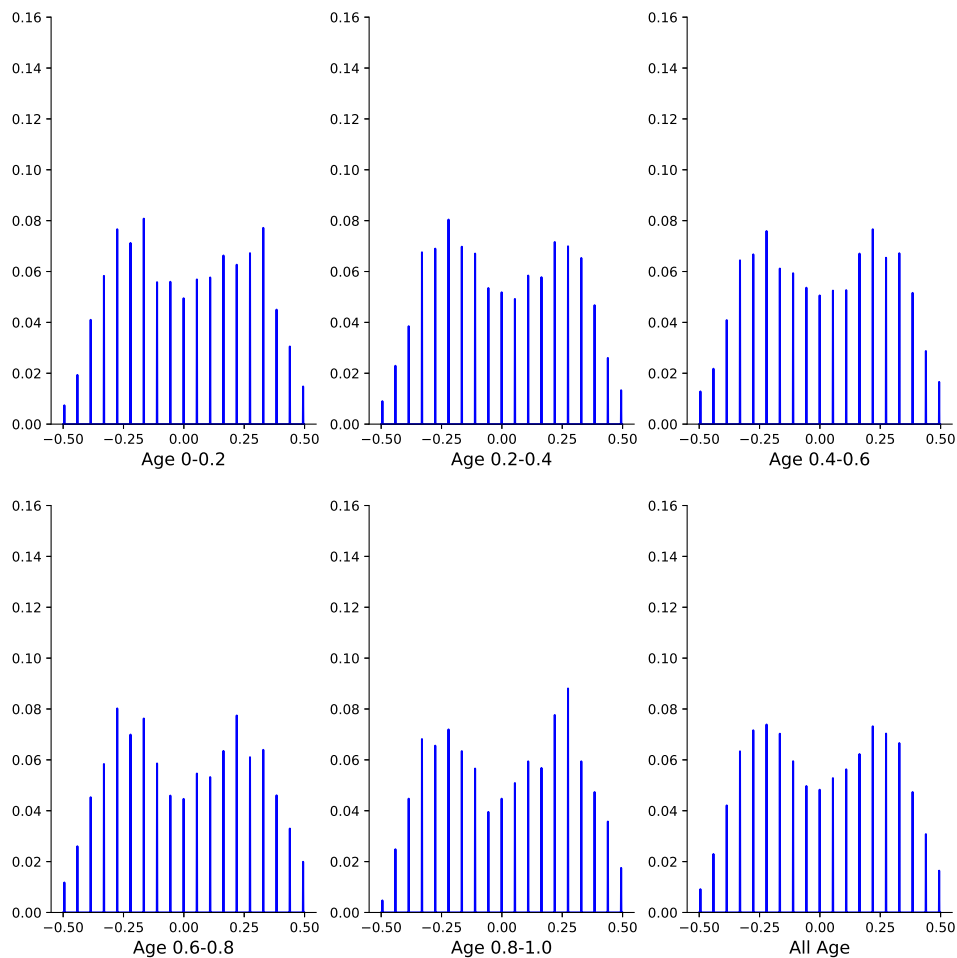
Fig. A.13 The figure shows the positional distribution of 54' along the radial axis of the cell. The x-axis represents the radial axis of the cell, and the y-axis represents the probability of the foci to be located at some position along the radial axis. The different histograms are for different time (ages) over the cell cycle. The last histogram represents the average radial distribution of 54' for all the ages.
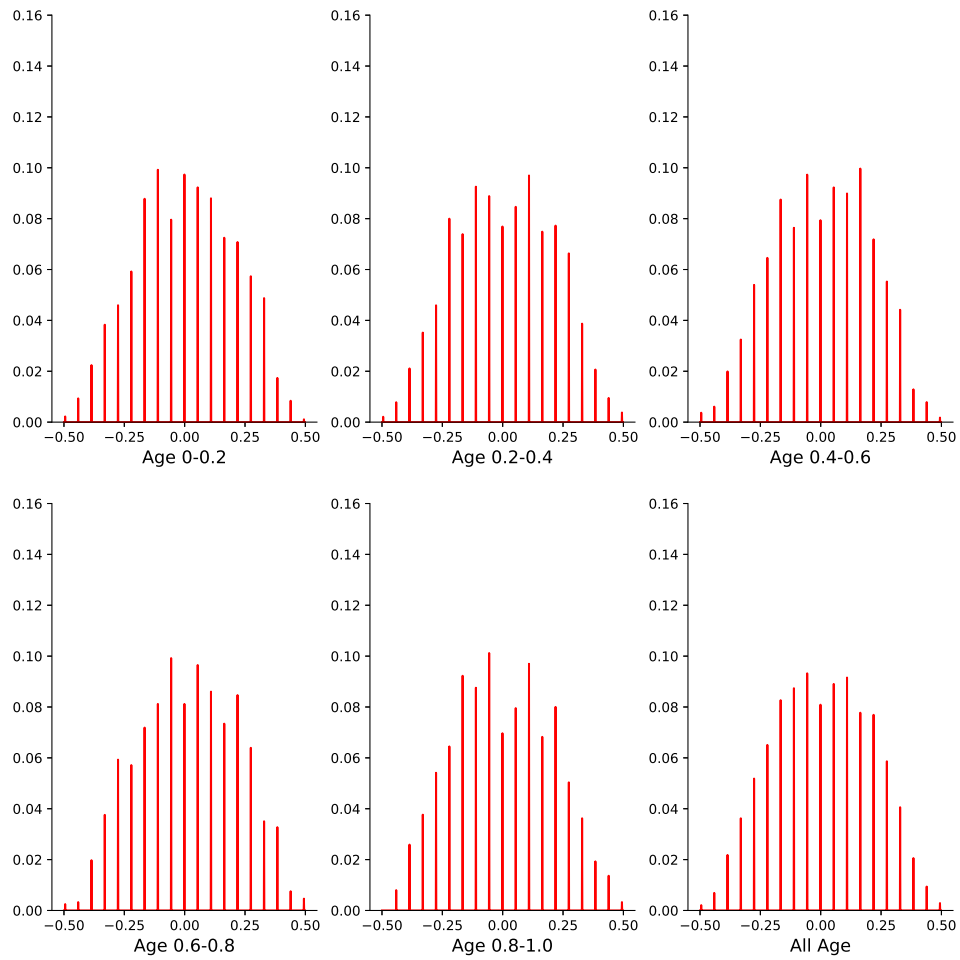
Fig. A.14 The figure shows the positional distribution of Ter along the radial axis of the cell. The x-axis represents the radial axis of the cell, and the y-axis represents the probability of the foci to be located at some position along the radial axis. The different histograms are for different time (ages) over the cell cycle. The last histogram represents the average radial distribution of Ter for all the ages.
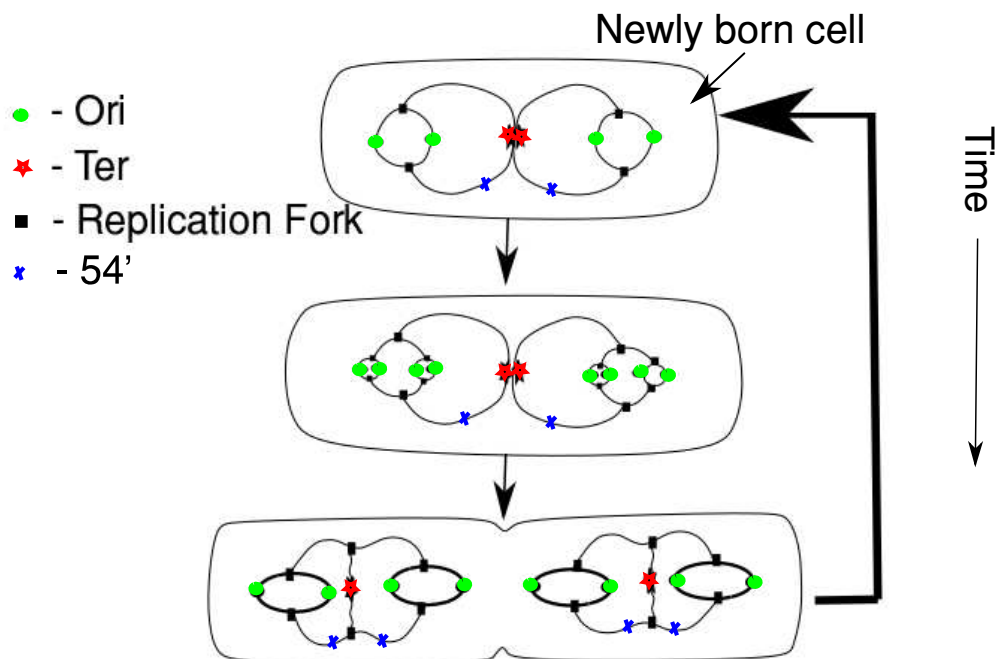
Fig. A.15 The schematic show the spatiotemporal organization of the chromosome over the period of the cell cycle. The schematic figure is deduced from the radial and longitudinal distribution of different foci.

We next plot the radial distribution of different foci, viz., Ori, 54', and Ter in the cylindrical cell as before. It is shown in Figs. A.12, A.13, and A.14, respectively. From figure A.12, we see that Ori foci show single peak distribution along the radial axis of the cell, and the peak is centered at the axis of the cell at all stages of the cell cycle. The radial distribution of the foci 54' in A.13 shows two-peak distribution where the peak lies in the left half or right half along the axis of the cell. It suggests that the segments corresponding to 54' foci lie in the right or left half of the cell from the axis of the cell. Also, the radial distribution of Ter foci in figure A.14 indicates that the Ter is found near the axis of the cell throughout the cell cycle.

Next, we combine the data from the radial and length distribution of different foci and provide the spatiotemporal organization of the chromosome during three overlapping cell-cycles. It is shown in Fig. A.15, where each focus is placed according to its most probable positions in the radial and length distribution histograms.

A.4   Discussion and future plans

In this appendix chapter, we show the experimental results to obtain the spatiotemporal organization of the *E. coli* chromosome over the period of the cell cycle during multifork replication. In particular, we use fluorescent microscopy technique to get the radial and longitudinal organization of the different fluorescently marked segments of the chromosome. We first reproduced the results of the paper [22] about the chromosome organization for two overlapping cell-cycles and then obtained the chromosome organization for three overlapping cell cycles. In particular, we find that all the foci are well-organized in space at different stages of the cell cycle. For, e.g., all the Ori foci are equally spaced spatially along the length of the cell and found at $1/4, 3/4$ or $1/8, 3/8, 5/8, 7/8$ positions along the cell's length for two or three overlapping cell-cycles conditions, respectively. Our results of the chromosome organization for two overlapping cell cycle is consistent with the results in paper [22]. Our results suggest that the chromosome organization in three overlapping cases can be thought of as an extrapolation to the chromosome organization for two overlapping cell-cycles. We provide the spatiotemporal organization of the chromosome for three overlapping cell cycle in Fig. A.15. In this chapter, we do not study the mechanism leading to the well-defined spatiotemporal organization of the chromosome during multifork-replication, which can be done in the future. The polymer models and the principles of statistical mechanics can help shed some light on the problem in the future. Also, as the best of my knowledge, there is no existing generic polymer model that can explain all the experimental data about the spatiotemporal dynamics of the chromosome in the cells growing at different rates. It can only be done in the future.

Appendix B

Generation of contact frequency map

In the field of bioinformatics, a sequence database is a collection of computerized nucleic acid sequences. Paired-end sequencing allows researchers to sequence both ends of a DNA-fragment to generate high-quality, alignable sequence data. Paired-end sequencing facilitates detection of genomic reorganization and repetitive sequence elements. In a paired end-sequencing run, the distance between the alignments of the two fragments is the length of the DNA fragment being sequenced. Aligners (software tools) use this information to better align reads when faced with a read that aligns to multiple regions such as those that may lie in a repeat region. To avoid this behavior, the reads are aligned in single end mode while keeping track of the pairs. We have employed the BWA [146] aligner to align reads as it has the best sensitivity among short read aligners. The aligned reads are then binned at the desired resolution (or the minimum distance between restriction sites). A 2D matrix with the required number of bins is initialized. Large fractions of the reads in a 3C library were from fragments that were not cross-linked and fall into the same bin or bins adjacent to each other. These read pairs were filtered out. The counter in the bin with the coordinates indicated by the alignment of each read in the pair is incremented for all the remaining reads. The filled matrix gives the total number of contacts between different parts of the genome and the resulting matrix is called the contact map. To be able to compare between different runs, the contact map is normalized, so that effect of varying number of sequenced reads is accounted for. Each sum of the number of contacts in each row and column in the matrix was normalized to 1. This provides a normalized contact map, which can now be used to elucidate the 3D structure of the genome and compare changes across different conditions. The above analysis has been done using the raw data obtained from [21, 8].

We next present the graphs to show the probabilty of different segments/CLs to be found in the inner, middle and outer region of the globule.
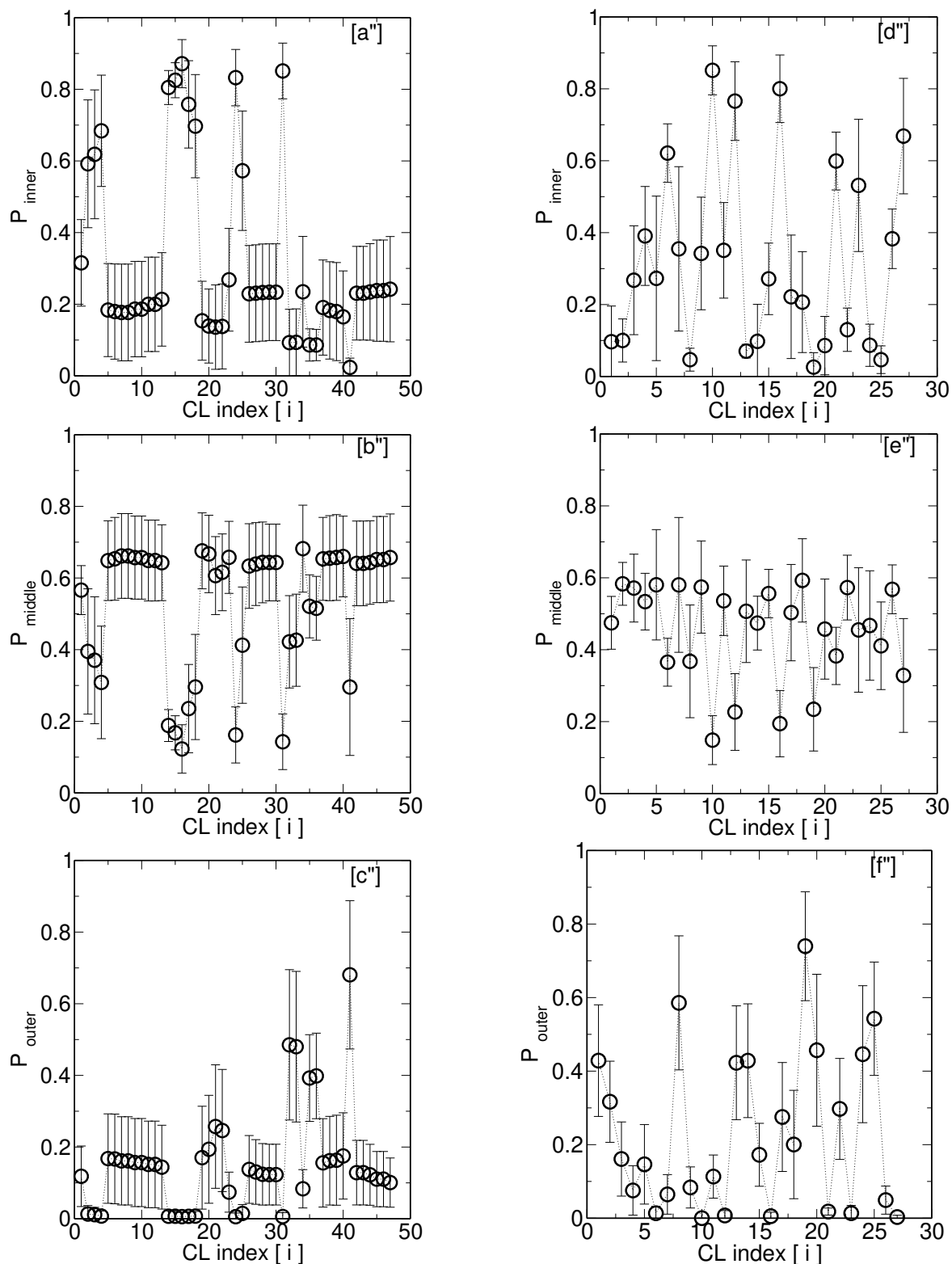
Fig. B.1 Subplots (a"), (b") and (c") (on left column) show the probabilities of individual CLs to be found in the inner, middle and outer region for CL-set BC-1. The x-axis is an index for CLs. The average is taken over 9 independent runs starting from different initial conditions, and standard deviation from the average is shown as an error bar. Small error bars indicate that the probability of finding CL i is the same across different runs. Data (d"), (e"), (f") on the right column is for the set of random CLs RC-1. The number of CLs correspond to the BC and RC CL-sets of bacteria *E. coli*.
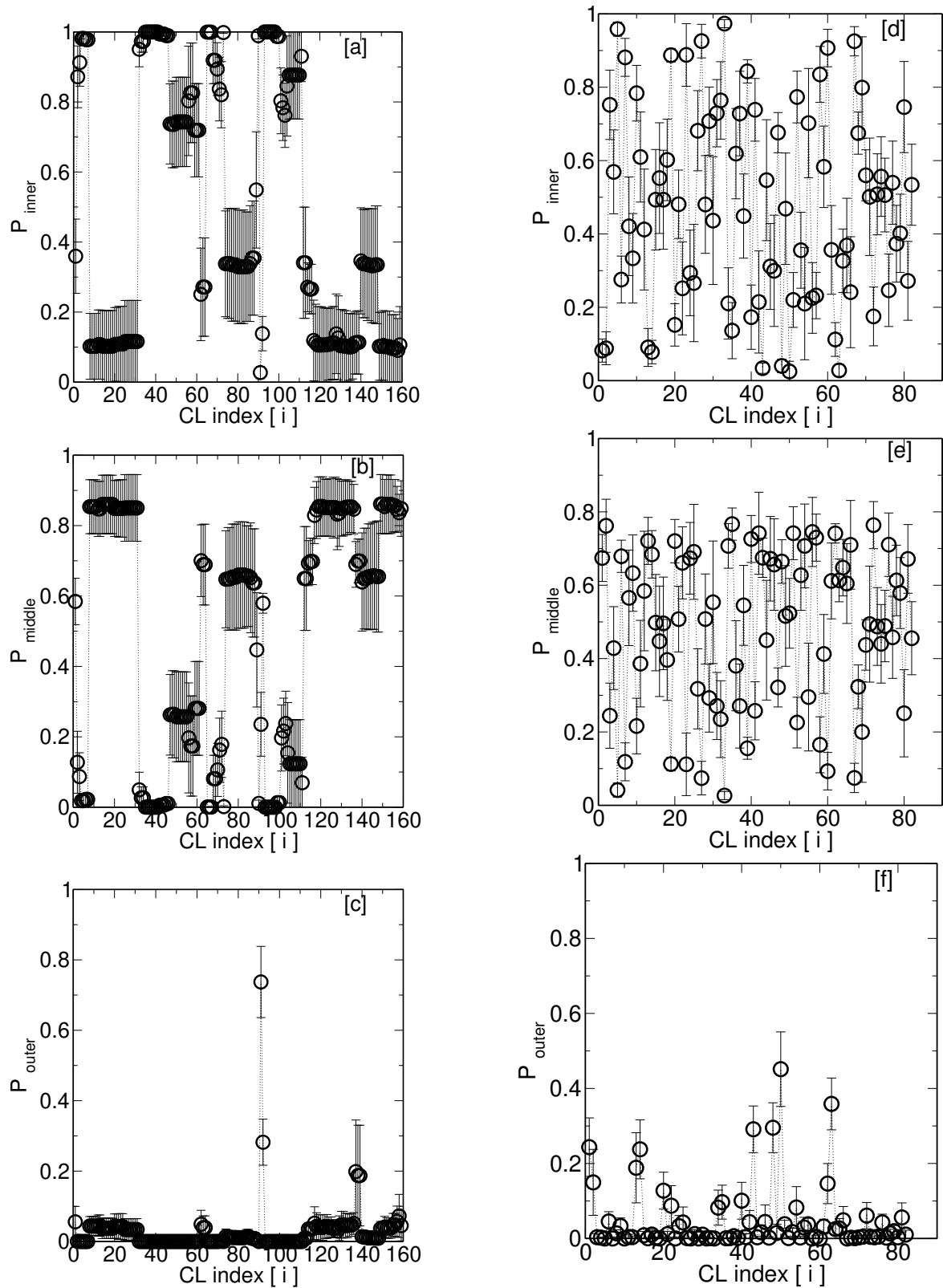
Fig. B.2 Subplots (a), (b) and (c) (on left column) show the probabilities of individual CLs to be found in the inner, middle and outer region for CL-set BC-2. The x-axis is an index for CLs. The average is taken over 9 independent runs starting from different initial conditions, and standard deviation from the average is shown as an error bar. Small error bars indicate that the probability of finding CL i is the same across different runs. Data (d), (e), (f) on the right column is for the set of random CLs RC-2. The set of 159 CLs for *E. Coli* are referred to as BC-2.
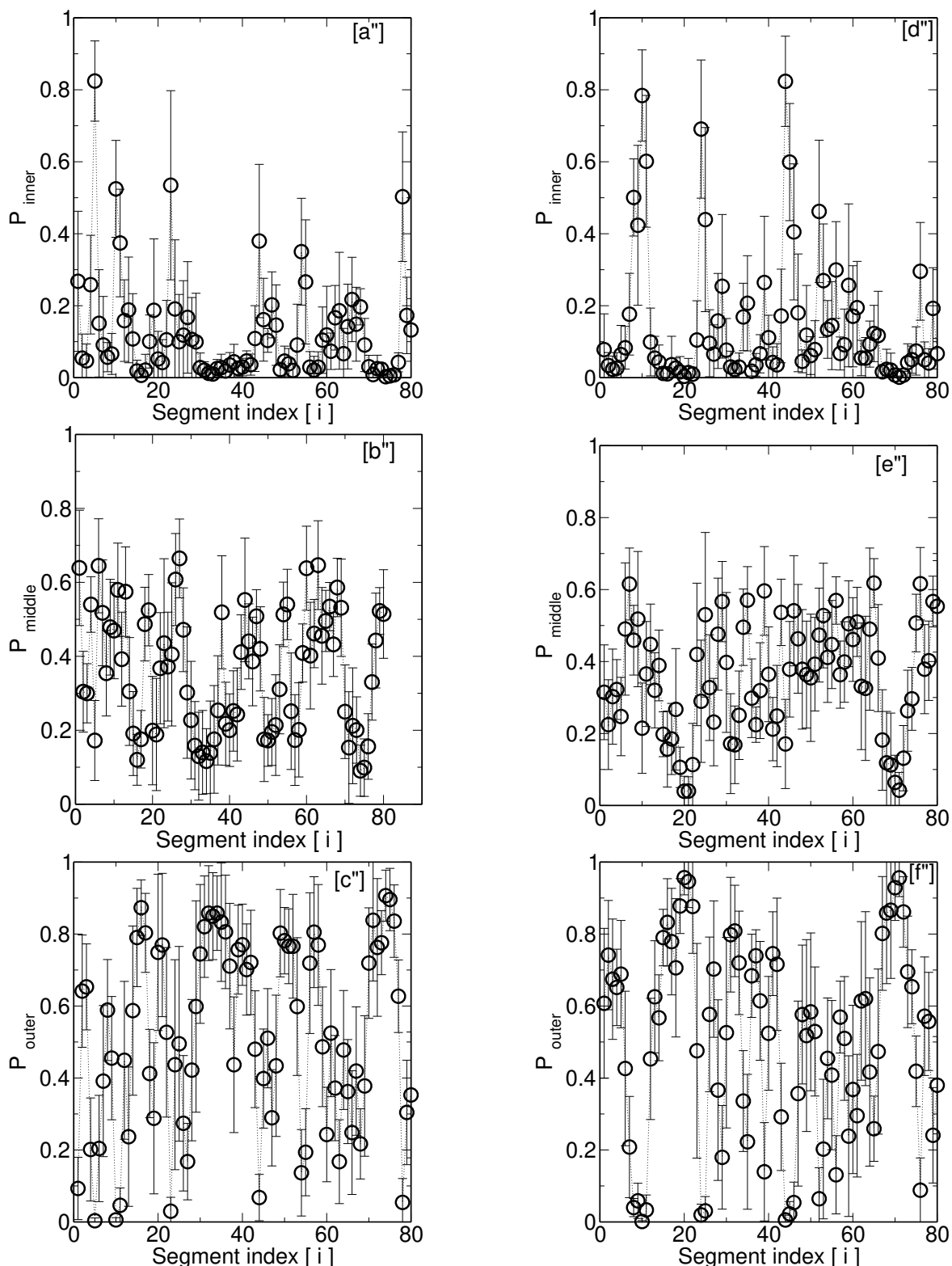
Fig. B.3 Subplots (a"), (b") and (c") (in the left column) shows the probabilities of the center of mass (CM) of 80 DNA-polymer segments to be found in the inner, middle and outer region of *E. Coli* DNA coil with BC-1 set of CLs. The x-axis is segment index. In each case average, values of P_{inner}, P_{middle} and P_{outer} are taken over nine runs starting from independent initial conditions, deviation from the average is shown as error bars. Small error bars indicate that the probability of finding CM of segment *i* in a particular region is nearly the same across different runs. Data on subplots (d"), (e"), (f") are for the random choice of cross-link position (set RC-1) with 82 CLs in a chain with 4642 monomers. Each segment has 58 monomers.

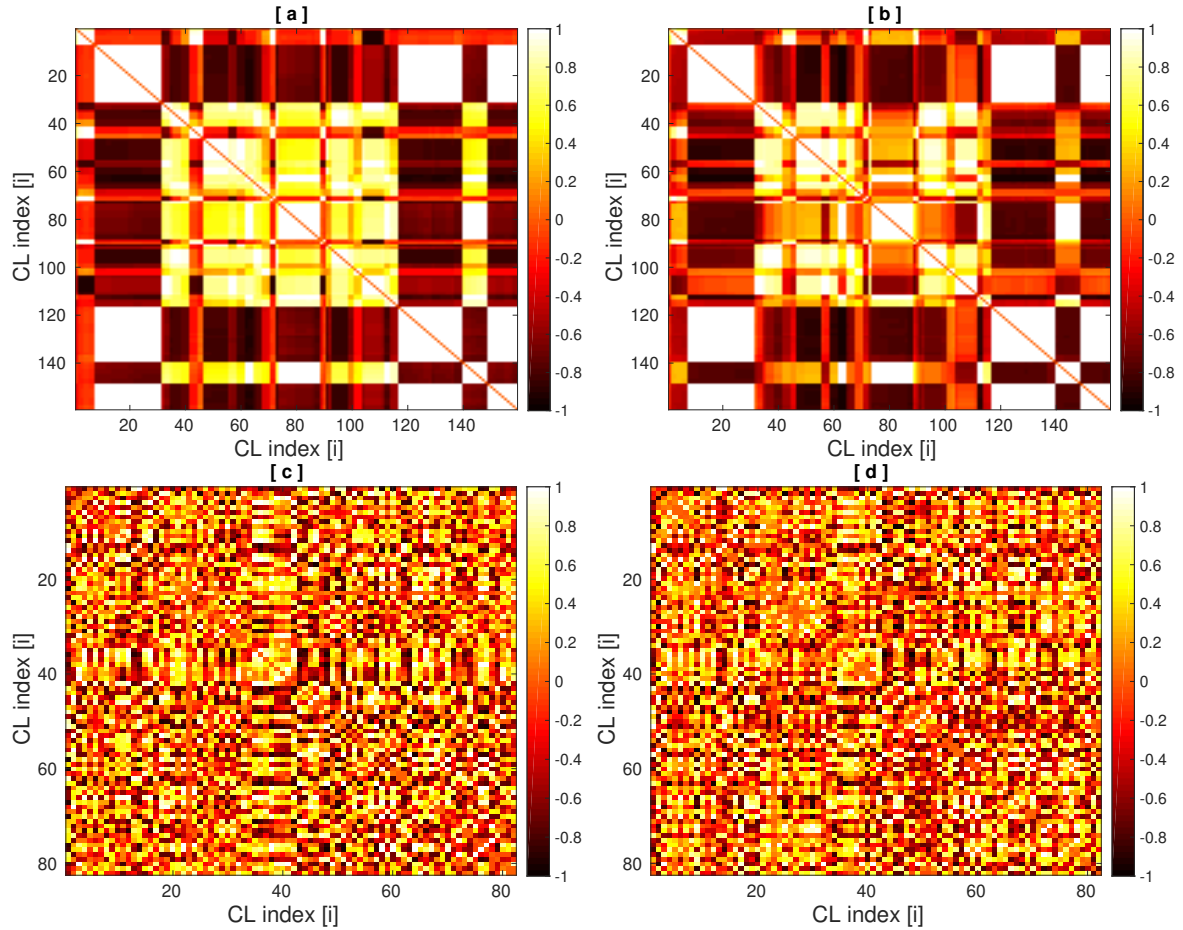# B.1 Angular correlation between different CLs

Fig. B.4 Colormaps to investigate the angular location of different CLs with respect to each other. Subplots (a),(b) are for BC-2 and (c),(d) for RC-2 with different initial conditions, respectively. The colormaps correspond to the DNA polymer of bacteria *E. coli*.

Fig. B.5 Colormaps to investigate the angular location of different DNA-polymer segments respect to each other. Subplots (a),(b),(a"), (b") are for BC-1 and (c),(d),(c"),(d") for RC-1 with different initial conditions, respectively. Each segment represent 58 monomers corresponding to the bacterial polymer of *E. coli* in 4642 monomer chain.
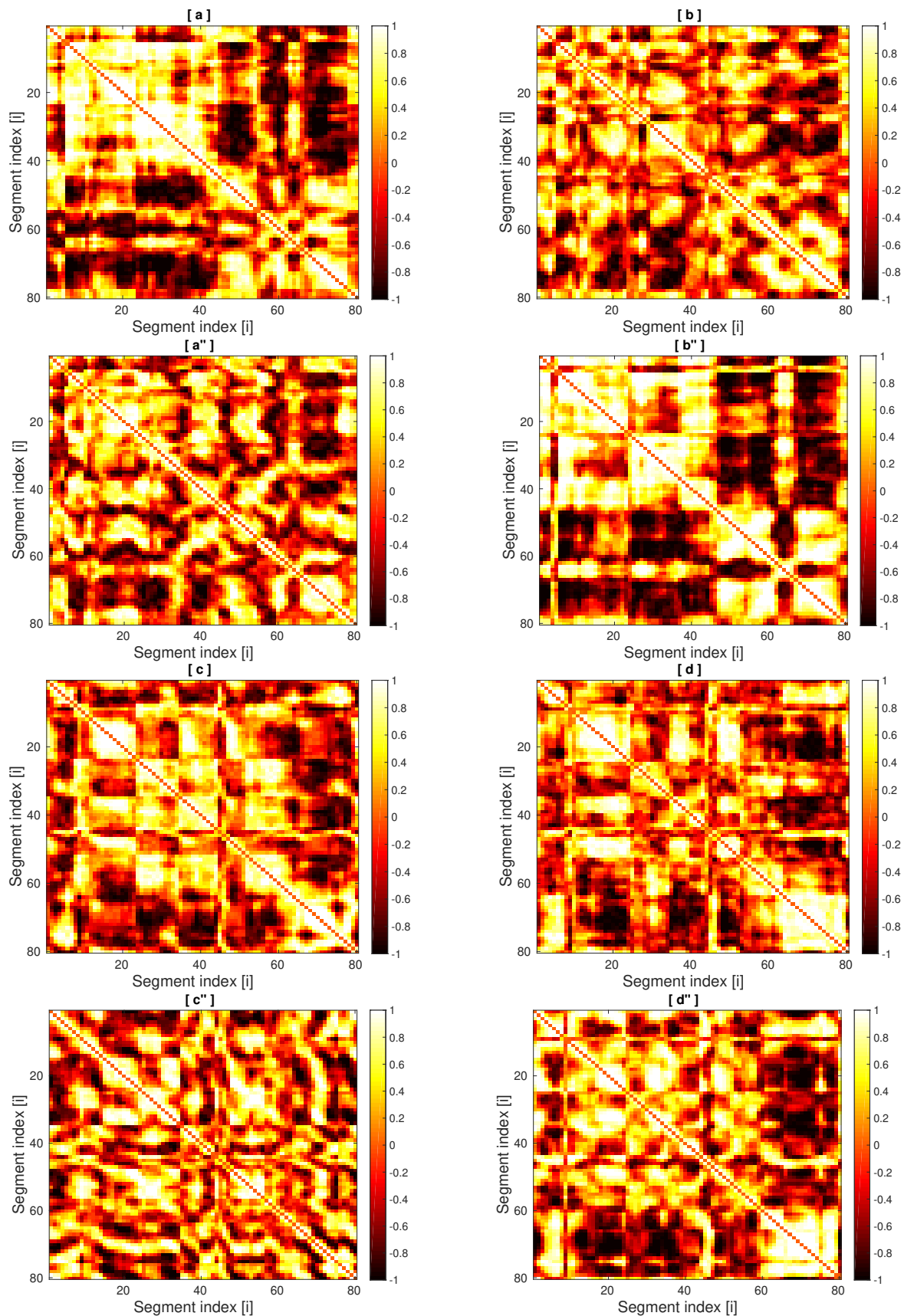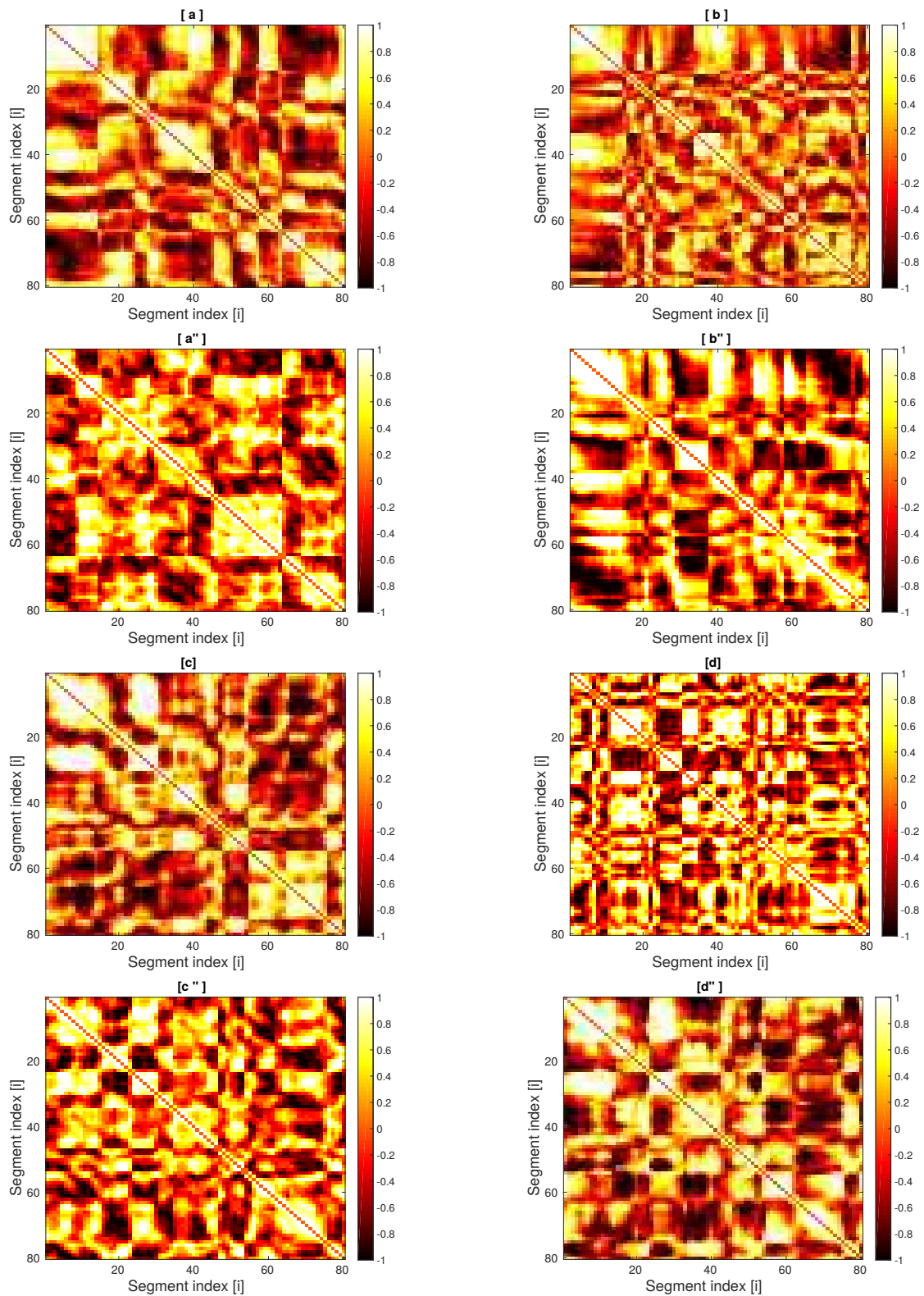
Fig. B.6 Colormaps to investigate the angular location of different DNA-polymer segments with respect to each other. Subplots (a),(b),(a"),(b") are for BC-1 and (c),(d),(c"),(d") for RC-1 with different initial conditions, respectively. Each segment represent 50 monomers corresponding to the bacterial polymer of *C. crescentus* in 4017 monomer chain.

Appendix C

List of cross-linked monomers in our simulations

In the following tables, we list the monomers which are cross-linked to model the constraints for the DNA of bacteria *E. coli* and *C. crescentus*, respectively. Note that for random cross links (CL) set-1 and set-2 (RC-1, RC-2) we have fewer number of CLs, as there are fewer *effective* CLs in the list of CLs.

In particular while counting the number of independent CLs, one should pay special attention to the points listed below. As a consequence, 47/49 CLs of BC-1 should be counted as only 27/26 independent CLs. Hence, we use just 27/26 CLs in RC-1, when we compare organization of RC-1 and BC-1. Correspondingly, we have just 82/60 CLs in RC-2, instead of 159/153 CLs in BC-2.

- The rows corresponding to independent cross-links of set BC-1 are marked by *, one can observe that the next row of CLs are adjacent to the monomers marked just previously by *. These cannot be counted as independent CLs.

- The rows marked by $^+$ are not the independent CLs at all, monomers $-$ and $-$ are trivially close to each other by virtue of their position along the contour.

C.1 Specific/ Randomly chosen cross-links for the chromosome of *E. coli*

This table has been generated by analysis of raw data obtained from C. Cagliero et. al., Nucleic Acids Res, **41**, 6058-6071 (2013).

-	BC-1		RC-1		BC-2		RC-2	
S. No.	Monomer index-1	Monomer index-2	Monomer index-1	Monomer index-2	Monomer Index-1	Monomer Index-2	Monomer Index-1	Monomer Index-2
1	1*	4642	1	4642	1	4642	1	4642
2	16*	2515	3739	4531	16	2515	3739	4531
3	17	2516	3011	1610	17	2516	3011	1610
4	20*	1051	2582	4367	20	1051	2582	4367
5	224*	2731	3370	1680	20	3584	3370	1680
6	225	2731	556	2622	21	1050	556	2622
7	226	2730	1676	1426	21	3584	1676	1426
8	226*	3428	998	2741	224	2731	998	2741
9	227	2728	474	2233	224	3429	474	2233
10	227	2729	2522	533	224	4208	2522	533
11	228	2727	1967	2490	224	4209	1967	2490
12	228	2728	2536	616	225	2730	2536	616
13	229*	2727	769	4614	225	2731	769	4614
14	271*	4509	2	2023	226	2729	2	2023
15	272	4508	3494	2484	226	2730	3494	2484
16	275*	1300	2534	1365	226	3427	2534	1365
17	280*	1051	3053	2256	226	3428	3053	2256
18	291*	1051	3779	2647	226	4038	3779	2647
19	316*	393	4199	4452	226	4169	4199	4452
20	317*	2172	2839	1309	227	2728	2839	1309
21	382*	1469	1385	449	227	2729	1385	449
22	383	1469	4398	371	228	2727	4398	371
23	527*	1529	522	1434	228	2728	522	1434
24	575*	1301	3676	320	228	3946	3676	320
25	609*	2515	178	4317	229	2727	178	4317
26	730*	3763	3220	515	229	3424	3220	515
27	731	3764	527	2992	229	3947	527	2992
28	732	3765	-	-	229	3948	2391	1402
29	733+	735	-	-	229	4172	284	4086
30	733*	3766	-	-	229	4213	4311	2243
31	1301*	3132	-	-	229	4214	283	1687
32	1433*	1635	-	-	271	4509	1599	2420

33	1434	1634	-	-	272	1471	4445	3365
34	1533*	3626	-	-	272	4508	1523	739
35	1571*	3667	-	-	274	1301	2721	113
36	1572	3668	-	-	275	1300	1371	4360
37	2728*	3945	-	-	275	1301	4137	2593
38	2729	3945	-	-	275	3130	1026	2807
39	2730	3943	-	-	276	2291	3007	2767
40	3429*	3942	-	-	276	3130	1576	1282
41	3471*	4177	-	-	276	3367	3041	3010
42	3620*	3763	-	-	280	292	2558	2709
43	3620	3764	-	-	280	1050	2156	3872
44	3621	3764	-	-	280	1051	945	4229
45	3622	3765	-	-	291	1051	4465	2873
46	3622	3766	-	-	291	2760	1943	4488
47	3623*	3766	-	-	315	393	4286	881
48	-	-	-	-	316	391	3282	3882
49	-	-	-	-	316	392	3555	2445
50	-	-	-	-	316	393	1196	40
51	-	-	-	-	317	392	1997	3918
52	-	-	-	-	317	568	4178	1595
53	-	-	-	-	317	569	678	3768
54	-	-	-	-	317	1094	3519	164
55	-	-	-	-	317	1095	2979	4115
56	-	-	-	-	317	2172	2871	3747
57	-	-	-	-	382	1469	3930	4263
58	-	-	-	-	383	1468	2787	2654
59	-	-	-	-	383	1469	1101	831
60	-	-	-	-	393	567	2785	1485
61	-	-	-	-	393	1096	3477	1069
62	-	-	-	-	393	2171	2345	795
63	-	-	-	-	526	1529	4037	3848
64	-	-	-	-	527	1529	395	1040
65	-	-	-	-	527	1530	328	930
66	-	-	-	-	575	1301	1926	2551
67	-	-	-	-	576	3130	4440	1484
68	-	-	-	-	576	3367	3799	4456

69	-	-	-	-	581	1636	4129	837
70	-	-	-	-	581	1637	1500	1352
71	-	-	-	-	582	1636	3197	947
72	-	-	-	-	608	2515	263	3435
73	-	-	-	-	609	2515	2272	277
74	-	-	-	-	688	1301	4276	702
75	-	-	-	-	730	3763	3405	978
76	-	-	-	-	731	3763	388	3658
77	-	-	-	-	731	3764	2796	1022
78	-	-	-	-	732	3621	3411	1122
79	-	-	-	-	732	3764	861	2185
80	-	-	-	-	732	3765	3564	1606
81	-	-	-	-	733	735	1860	1447
82	-	-	-	-	733	3623	904	3577
83	-	-	-	-	733	3765	-	-
84	-	-	-	-	733	3766	-	-
85	-	-	-	-	734	3765	-	-
86	-	-	-	-	734	3766	-	-
87	-	-	-	-	738	1533	-	-
88	-	-	-	-	738	3626	-	-
89	-	-	-	-	782	2522	-	-
90	-	-	-	-	1051	3585	-	-
91	-	-	-	-	1208	1210	-	-
92	-	-	-	-	1269	1271	-	-
93	-	-	-	-	1301	1398	-	-
94	-	-	-	-	1301	2102	-	-
95	-	-	-	-	1301	2289	-	-
96	-	-	-	-	1301	3132	-	-
97	-	-	-	-	1301	3366	-	-
98	-	-	-	-	1301	3652	-	-
99	-	-	-	-	1397	2573	-	-
100	-	-	-	-	1397	3118	-	-
101	-	-	-	-	1433	1635	-	-
102	-	-	-	-	1434	1634	-	-
103	-	-	-	-	1435	1633	-	-
104	-	-	-	-	1469	2071	-	-

105	-	-	-	-	1470	2071	-	-
106	-	-	-	-	1470	2998	-	-
107	-	-	-	-	1470	3186	-	-
108	-	-	-	-	1470	4498	-	-
109	-	-	-	-	1470	4508	-	-
110	-	-	-	-	1470	4509	-	-
111	-	-	-	-	1471	4508	-	-
112	-	-	-	-	1533	3625	-	-
113	-	-	-	-	1533	3626	-	-
114	-	-	-	-	1571	3667	-	-
115	-	-	-	-	1572	3667	-	-
116	-	-	-	-	1572	3668	-	-
117	-	-	-	-	2726	4172	-	-
118	-	-	-	-	2727	4172	-	-
119	-	-	-	-	2728	3945	-	-
120	-	-	-	-	2728	4039	-	-
121	-	-	-	-	2728	4171	-	-
122	-	-	-	-	2729	3945	-	-
123	-	-	-	-	2729	4038	-	-
124	-	-	-	-	2730	3943	-	-
125	-	-	-	-	2730	4038	-	-
126	-	-	-	-	2731	3942	-	-
127	-	-	-	-	2731	4036	-	-
128	-	-	-	-	2732	3942	-	-
129	-	-	-	-	3424	4172	-	-
130	-	-	-	-	3426	3945	-	-
131	-	-	-	-	3426	4156	-	-
132	-	-	-	-	3427	3944	-	-
133	-	-	-	-	3427	4038	-	-
134	-	-	-	-	3428	3943	-	-
135	-	-	-	-	3428	3944	-	-
136	-	-	-	-	3429	3942	-	-
137	-	-	-	-	3471	4177	-	-
138	-	-	-	-	3472	4176	-	-
139	-	-	-	-	3472	4177	-	-
140	-	-	-	-	3619	3763	-	-

141	-	-	-	-	3620	3763	-	-
142	-	-	-	-	3620	3764	-	-
143	-	-	-	-	3621	3764	-	-
144	-	-	-	-	3621	3765	-	-
145	-	-	-	-	3622	3765	-	-
146	-	-	-	-	3622	3766	-	-
147	-	-	-	-	3623	3766	-	-
148	-	-	-	-	3623	3768	-	-
149	-	-	-	-	3942	4167	-	-
150	-	-	-	-	3942	4209	-	-
151	-	-	-	-	3943	4038	-	-
152	-	-	-	-	3944	4038	-	-
153	-	-	-	-	3944	4169	-	-
154	-	-	-	-	3944	4210	-	-
155	-	-	-	-	4036	4167	-	-
156	-	-	-	-	4036	4209	-	-
157	-	-	-	-	4037	4443	-	-
158	-	-	-	-	4041	4214	-	-
159	-	-	-	-	4172	4214	-	-

Table C.1   Table shows the list of pair of monomers which constitute the CLs for *E. coli*, these CLs are used as an input to the simulation by constraining these monomers to be at a distance *a* from each other. The first monomer with label 1 and the last monomer labelled 4642 are linked together because the chromosome is a ring polymer.

C.2   Specific/ Randomly chosen cross-links for the chromosome of *C. crescentus*

This table has been generated by analysis of raw data obtained from Tung B. et. al., Science, **342**, 731-734 (2013).

-	BC-1		RC-1		BC-2		RC-2	
S. No.	Monomer Index-1	Monomer Index-2	Monomer Index-1	Monomer index-2	Monomer Index-1	Monomer Index-2	Monomer Index-1	Monomer Index-2
1	1*	4017	1	4017	1	4017	1	4017
2	289*	1985	23	2743	289	1985	23	2743
3	289	1986	2348	3956	289	1986	2348	3956

4	290	1986	2602	3884	290	1986	2602	3884
5	290	1987	3724	2295	290	1987	3724	2295
6	468*	564	2675	2406	438	3797	2675	2406
7	469	565	1972	2520	468	563	1972	2520
8	470	566	3779	1729	468	563	3779	1729
9	470	567	3022	3962	469	565	3022	3962
10	471	567	1093	2574	469	566	1093	2574
11	541*	2494	2739	3649	470	566	2739	3649
12	541*	2907	958	3944	470	567	958	3944
13	541*	2957	2611	1071	471	567	2611	1071
14	541	2958	512	1466	540	955	512	1466
15	641*	683	229	1385	540	2494	229	1385
16	693*	2875	3206	679	540	2906	3206	679
17	693	2876	2471	3744	540	2957	2471	3744
18	693*	2945	663	2183	541	2493	663	2183
19	710*	1890	1906	3786	541	2494	1906	3786
20	710*	3145	2029	1691	541	2906	2029	1691
21	733*	1890	1218	199	541	2907	1218	199
22	847*	3912	1478	2675	541	2957	1478	2675
23	1032*	1437	906	1605	541	2958	906	1605
24	1032*	3851	1124	3776	542	2907	1124	3776
25	1033**	3579**	1656	881	542	2958	1656	881
26	1261*	2613	2183	332	542	2907	2183	332
27	1369*	2250	-	-	640	683	3083	1374
28	1370	2249**	-	-	641	683	2086	742
29	1370	2250	-	-	693	2875	2145	314
30	1393*	3119	-	-	693	2876	748	841
31	1398*	3455	-	-	693	2945	3555	3745
32	1399	3454	-	-	693	2946	1687	2814
33	1399	3454	-	-	694	2876	1723	520
34	1437*	3580	-	-	694	2944	2480	2903
35	2249**	2803**	-	-	694	2975	1722	3828
36	2493*	2907	-	-	710	733	525	2133
37	2494	2907	-	-	710	1890	3900	1691
38	2494	2907	-	-	710	1891	831	1958
39	2581$^+$	2584	-	-	710	3037	1538	2521

40	2582$^+$	2584	-	-	710	3038	826	2225
41	2803**	3119	-	-	710	3145	2078	2734
42	2837*	3767	-	-	710	3146	2208	1592
43	2838	3768	-	-	733	1890	3914	557
44	2839	3769	-	-	733	3037	1838	3442
45	2842*	3772	-	-	733	3038	3230	939
46	2875*	2945	-	-	733	3145	1519	785
47	2875	2946	-	-	734	1890	2927	2645
48	3348*	3377	-	-	734	3038	1024	2151
49	3579**	3850	-	-	847	3912	1855	3704
50	-	-	-	-	874	876	375	3836
51	-	-	-	-	875	1611	2626	54
52	-	-	-	-	876	1611	1493	2839
53	-	-	-	-	1032	1437	2158	3728
54	-	-	-	-	1032	3579	1389	2943
55	-	-	-	-	1032	3580	100	2045
56	-	-	-	-	1032	3850	2130	2899
57	-	-	-	-	1032	3851	1214	480
58	-	-	-	-	1033	1438	1162	1808
59	-	-	-	-	1033	3579	981	1120
60	-	-	-	-	1033	3850	2492	1058
61	-	-	-	-	1057	1103	-	-
62	-	-	-	-	1057	3240	-	-
63	-	-	-	-	1057	3331	-	-
64	-	-	-	-	1069	3418	-	-
65	-	-	-	-	1069	3419	-	-
66	-	-	-	-	1162	1164	-	-
67	-	-	-	-	1163	1165	-	-
68	-	-	-	-	1261	2613	-	-
69	-	-	-	-	1261	2614	-	-
70	-	-	-	-	1262	2613	-	-
71	-	-	-	-	1369	1394	-	-
72	-	-	-	-	1369	2250	-	-
73	-	-	-	-	1369	2804	-	-
74	-	-	-	-	1369	3119	-	-
75	-	-	-	-	1370	1393	-	-

76	-	-	-	-	1370	1394	-	-
77	-	-	-	-	1370	2249	-	-
78	-	-	-	-	1370	2250	-	-
79	-	-	-	-	1370	2803	-	-
80	-	-	-	-	1370	3119	-	-
81	-	-	-	-	1393	2249	-	-
82	-	-	-	-	1393	2803	-	-
83	-	-	-	-	1393	3119	-	-
84	-	-	-	-	1394	2250	-	-
85	-	-	-	-	1394	2804	-	-
86	-	-	-	-	1394	3119	-	-
87	-	-	-	-	1398	3455	-	-
88	-	-	-	-	1399	3454	-	-
89	-	-	-	-	1399	3455	-	-
90	-	-	-	-	1421	1657	-	-
91	-	-	-	-	1421	2385	-	-
92	-	-	-	-	1437	3579	-	-
93	-	-	-	-	1437	3580	-	-
94	-	-	-	-	1437	3851	-	-
95	-	-	-	-	1890	3038	-	-
96	-	-	-	-	1890	3145	-	-
97	-	-	-	-	1891	2177	-	-
98	-	-	-	-	1891	2178	-	-
99	-	-	-	-	1891	3958	-	-
100	-	-	-	-	2041	2046	-	-
101	-	-	-	-	2059	2061	-	-
102	-	-	-	-	2141	2143	-	-
103	-	-	-	-	2249	2803	-	-
104	-	-	-	-	2250	2803	-	-
105	-	-	-	-	2250	2804	-	-
106	-	-	-	-	2250	3119	-	-
107	-	-	-	-	2303	3240	-	-
108	-	-	-	-	2303	3958	-	-
109	-	-	-	-	2385	3958	-	-
110	-	-	-	-	2493	2907	-	-
111	-	-	-	-	2493	2958	-	-

112	-	-	-	-	2494	2906	-	-
113	-	-	-	-	2494	2907	-	-
114	-	-	-	-	2494	2957	-	-
115	-	-	-	-	2495	2906	-	-
116	-	-	-	-	2495	2957	-	-
117	-	-	-	-	2581	2583	-	-
118	-	-	-	-	2581	2584	-	-
119	-	-	-	-	2582	2584	-	-
120	-	-	-	-	2655	3912	-	-
121	-	-	-	-	2803	3118	-	-
122	-	-	-	-	2803	3119	-	-
123	-	-	-	-	2804	3119	-	-
124	-	-	-	-	2837	3767	-	-
125	-	-	-	-	2837	3768	-	-
126	-	-	-	-	2838	3767	-	-
127	-	-	-	-	2838	3768	-	-
128	-	-	-	-	2838	3769	-	-
129	-	-	-	-	2839	3769	-	-
130	-	-	-	-	2839	3770	-	-
131	-	-	-	-	2840	3770	-	-
132	-	-	-	-	2841	3770	-	-
133	-	-	-	-	2842	3773	-	-
134	-	-	-	-	2842	3772	-	-
135	-	-	-	-	2842	3773	-	-
136	-	-	-	-	2874	2946	-	-
137	-	-	-	-	2875	2945	-	-
138	-	-	-	-	2875	2946	-	-
139	-	-	-	-	2876	2945	-	-
140	-	-	-	-	2906	2957	-	-
141	-	-	-	-	2907	2958	-	-
142	-	-	-	-	3008	3011	-	-
143	-	-	-	-	3037	3145	-	-
144	-	-	-	-	3038	3145	-	-
145	-	-	-	-	3241	3591	-	-
146	-	-	-	-	3241	3957	-	-
147	-	-	-	-	3348	3377	-	-

148	-	-	-	-	3579	3850	-	-
149	-	-	-	-	3579	3851	-	-
150	-	-	-	-	3580	3851	-	-
151	-	-	-	-	3591	3957	-	-
152	-	-	-	-	3764	3958	-	-
153	-	-	-	-	4007	4011	-	-

Table C.2 Table showing the list of pair of monomers which constitute the CLs for *C. Crescentus*, these CLs are used as an input to the simulation. The first monomer with label 1 and the last monomer labelled 4017 are linked together because the chromosome is a ring polymer.

C.3 List of cross-linked monomers in BC' CL sets

-	BC' (*C. crescentus*)		BC' (*E. coli*)	
S. No.	Monomer index-1	Monomer index-2	Monomer index-1	Monomer index-2
1	1	4017	1	4642
2	289	1985	16	2515
3	289	1986	17	2516
4	290	1987	20	1051
5	290	1987	21	1050
6	468	564	21	3584
7	469	565	224	2731
8	470	566	224	4209
9	470	567	225	2730
10	471	567	225	2731
11	541	2494	226	2729
12	541	2907	226	2730
13	541	2957	226	3428
14	541	2958	227	2728
15	641	683	227	2729
16	693	2875	228	2727
17	693	2876	228	2728

18	693	2945	229	2727
19	694	2876	229	4213
20	710	1890	229	4214
21	710	3145	271	4509
22	733	1890	272	4508
23	847	3912	274	1301
24	1032	1437	275	1300
25	1032	3580	280	1050
26	1032	3850	280	1051
27	1032	3851	291	1051
28	1033	3579	316	392
29	1033	3850	316	393
30	1057	3331	317	392
31	1261	2613	317	1095
32	1369	2250	317	2172
33	1370	1393	382	1469
34	1370	2249	383	1469
35	1370	2250	393	567
36	1393	3119	526	1529
37	1398	3455	527	1529
38	1399	3454	527	1530
39	1399	3455	575	1301
40	1421	1657	609	2515
41	1437	3579	688	1301
42	1437	3580	730	3763
43	2249	2803	731	3764
44	2250	2804	732	3765
45	2303	3240	733	735
46	2493	2907	733	3765
47	2494	2906	733	3766
48	2494	2906	735	3766
49	2581	2583	1208	1210

50	2581	2584	1269	1271
51	2582	2584	1301	3132
52	2803	3119	1433	1635
53	2837	3767	1434	1634
54	2838	3768	1435	1633
55	2839	3769	1470	2998
56	2842	3772	1470	3186
57	2875	2945	1470	4509
58	2875	2946	1533	3626
59	3348	3377	1571	3667
60	3579	3850	1572	3667
61	-	-	1572	3668
62	-	-	2728	3945
63	-	-	2729	3945
64	-	-	2729	4038
65	-	-	2730	3943
66	-	-	3427	4038
67	-	-	3429	3942
68	-	-	3471	4177
69	-	-	3472	4176
70	-	-	3620	3763
71	-	-	3620	3764
72	-	-	3621	3764
73	-	-	3622	3765
74	-	-	3622	3766
75	-	-	3623	3766
76	-	-	3623	3768
77	-	-	3944	4210

Table C.3 The table shows the list of pair of monomers which constitute the CLs for *E. coli* and *C. crescentus* corresponding to the BC' CL sets mentioned in the main paper.

Appendix D

List of highly expressed genes for the bacteria *E. coli*

Serial no.	gene name	Start point (in BP)	End point (in BP)	Segment index	Location of segment
1	rpsA	961995	963668	17	outer
2	rpsB	189874	190599	4	middle
3	rpsC	3449182	3449883	60	middle,outer
4	rpsD	3441055	3441675	60	middle,outer
5	rpsE	3444726	3445229	60	middle,outer
6	rpsI	3377815	3378207	59	inner
7	rpsM	3442115	3442471	60	middle,outer
8	rplA	4178879	4179583	73	middle,outer
9	rplC	3452297	3452926	60	middle,outer
10	rplB	3450543	3451364	60	middle,outer
11	rplD	3451681	3452286	60	middle,outer
12	rplE	3446899	3447438	60	middle,outer
13	rplK	4178447	4178875	73	middle,outer
14	rplM	3378223	3378651	59	inner
15	rplN	3447778	3448149	60	middle,outer
16	rplP	3448759	3449169	60	middle,outer
17	rplQ	3439616	3439999	60	middle,outer
18	rplT	1799393	1799749	32	outer
19	dnaK	12163	14079	1	middle,inner
20	ptsG	1157869	1159302	20	middle,outer

21	pnp	3309033	3311168	58	middle
22	fusA	3471400	3473514	41	middle,outer
23	tufA	3470145	3471329	60	middle,outer
24	tsf	190857	191708	4	middle
25	rpoB	4181245	4185273	73	middle,outer
26	rpoC	4185350	4189573	72	middle,outer
27	deaD	3305971	3307860	57	middle,outer
28	fbaA	3070165	3071244	52	inner,middle
29	pgk	3071459	3072622	53	middle,outer
30	eno	2906643	2907941	51	middle,outer
31	gapA	1862771	1863766	33	outer
32	pckA	3532818	3534440	61	middle,outer
33	aceE	123017	125680	3	middle,outer
34	aceF	125695	127587	3	middle,outer
35	lpdA	127912	129336	3	middle,outer
36	adhE	1295446	1298121	23	inner
37	atpA	3918316	3919857	68	middle,outer
38	atpD	3915993	3917375	68	middle,outer
39	aspA	4366891	4368327	76	outer
40	pta	2414747	2416891	42	outer
41	purA	4404687	4405985	76	outer
42	ilvC	3957970	3959445	69	middle,outer
43	ompA	1019013	1020053	18	middle,outer
44	ompC	2311646	2312749	40	inner
45	ompF	985894	986982	17	outer
46	pal	779067	779588	14	inner,middle
47	typA	4058407	4060230	70	middle,outer

Table D.1 Table shows the 47 highly expressed genes in *E. coli* and there positions in the DNA polymer. Start point is the location of the base pair at which gene is starting and end point is the location of the last base pair of that gene. Fifth column shows the segment index corresponding to that gene in our simulation polymer. Last column shows the location of the segment in our simulation from chapter 2.