DNA to chromosome: developing a physical understanding of biological systems

Debasish Chaudhuri[1,*]

[1] *Institute of Physics, Sachivalaya Marg, Bhubaneswar 751005, India*

(Dated: May 5, 2022)

DNA constitutes the central information hub of the cell. It is a long polymer that carries instructions for life processes. The instructions must be regulated so that different types of cells with the same DNA display different properties. This is partly achieved by an intricate organisation of the DNA, with the help of associated proteins, into compact but malleable structures called chromosomes. A significant degree of compaction of DNA is also necessary due to the relatively small space in the cell available to them. We outline salient properties of this biologically crucial system and methods to build physical understanding.

I. INTRODUCTION

On 28 February 1953, Francis Crick, a physicist working at Cavendish laboratory, famously walked into The Eagle pub in Cambridge and declared "We have found the secret of life". The "secret" was DNA double-helix structure. Along with his collaborator, and later, fellow Nobel prize laureate James Watson, a biologist at the same lab, they figured it out. Watson and Crick mostly collected and analysed existing experimental data to build their model. The most crucial piece of evidence came from the lab of Rosalind Franklin in King's College, London. Franklin and her graduate student were performing X-ray diffraction experiments on DNA. One of their images, the famous 'image 51', showed a remarkably clear and striking X-shaped diffraction pattern [1] (see Fig. 1). Watson knew that Crick obtained similar structure factor earlier from his calculations considering double-helix molecules. Thus he immediately understood that the underlying structure of DNA is a double helix [2].

DNA was discovered much earlier, in the late 1860s by Swiss chemist Friedrich Miescher. It contains two kinds of bases, double-ringed Purines (Adenine, Guanine) and single-ringed Pyrimidines (Cytosine, Thymine). Erwin Chargaff found that in a DNA segment, the total amount of Adenine (A) is equal to Thymine (T) and that of Guanine (G) is equal to Cytosine (C). Watson and Crick used cardboard models to finally realize that the complementary bases A-T, G-C forming hydrogen bonds between themselves fit together perfectly with the DNA double helix structure of the sugar-phosphate backbones (Fig. 1). Another finding of Chargaff was that the amount of basepairs A-T and G-C differs in different DNA segments. After initial doubts and debates regarding whether the DNA or the proteins carry the genetic codes for hereditary information transfer, it has become clear that the base pair (bp) sequence in DNA contains the information for life [3].

Three consecutive bases on DNA act as a codon carrying information to produce a single amino acid, the basic unit of proteins. The protein production from DNA is known as the central dogma of biology. It involves two processes called transcription and translation. In transcription, RNA polymerase opens up specific DNA segments to generate RNAs with a sequence complementary to the DNA sequence. In turn, these RNAs produce amino-acid chains or proteins, the cell's machinery. The proteins are of various sizes and can contain up to 800 to 900 amino acids. The distribution of protein sizes typically has an asymmetric bell shape with a median between 300 to 400 amino acids. Thus approximately 1000 bases of DNA code a protein, on average.

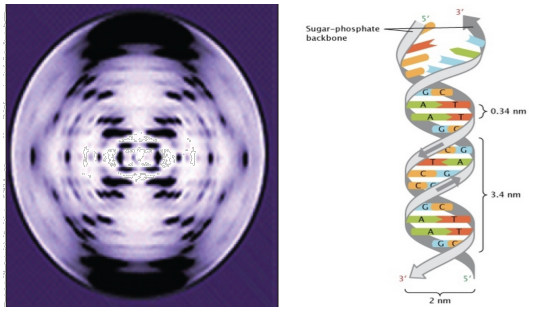

FIG. 1: (color online) DNA-double helix: (left) The figure-X diffraction pattern of DNA adapted from Ref. [1]. (right) The DNA double helix model adapted from Ref. [4].

In Fig. (1), we show the figure-X diffraction pattern mentioned above and the model of DNA double helix. Note that DNA is a long chain-like molecule called a polymer. In it, the two sugar-phosphate backbones form two helices (hence the term double-helix) run anti-parallel to each other. The helical pitch is 3.4 nm; it is the DNA length over which each helix performs one complete turn. The width of the DNA is about 2 nm. Lengthwise, a one-nanometre segment of DNA contains about three base pairs. At such short lengths, DNA behaves like a rigid rod. It has a persistence length of 50 nm, a length scale beyond which the DNA starts to become flexible allowing bending. In this sense, DNA is considered to be a semi-flexible polymer, which can have enormous length. DNA chain length in typical bacterial cells is millimetres, containing $\sim 10^6$ basepairs or 1 Mega basepair (Mbp). In

*Electronic address: debc@iopb.res.in

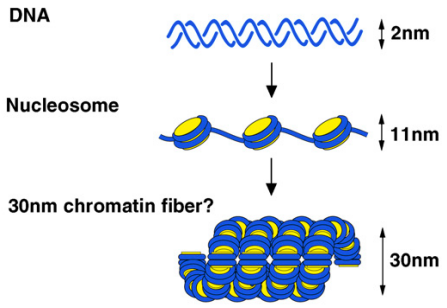

FIG. 2: (color online) The steps of DNA folding into 30 nm chromatin structure. The image is adapted from Ref [6].

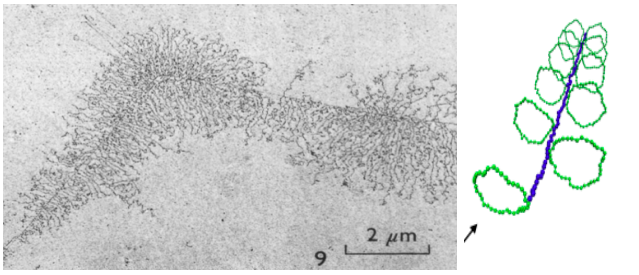

FIG. 3: (color online) Electron micrograph of lampbrush chromosome from oocyte nucleolei, adapted from Ref. [7]. The image on the right shows a simple self-avoiding polymer model that we constructed based on such observations [8].

mammals, the DNA length is even longer; it is ∼metres containing 1 Giga basepair (Gbps). As per the approximate estimate that about 1000 bps code for a protein, the bacterial DNA can produce ∼ 1000 proteins while the mammalian DNA can produce ∼ 10^6 proteins. Not all the proteins that a DNA has the code of are always produced. There are intricate physical and chemical processes that regulate transcription. One can imagine if all the cells in our body were creating all the proteins they potentially can, we would not have different organs. Even single-cell bacteria need to produce particular proteins depending on the cell cycle and the immediate environment.

Although the DNA can be so long, the space it needs to get accommodated is tiny. The millimetre long DNA in bacteria must be compacted a thousandfold to get housed in a cell of one-micrometre size. On the other hand, the 23 pairs of meter-long DNAs are packed in a micrometre-sized cell nucleus in humans. The amount of compaction it requires is equivalent to fitting an 80 km long chain inside a football. Despite such enormous compaction, they still allow gene expression requiring access of DNA to RNA-polymerase [3, 5].

II. CHROMOSOME

Such enormous compaction and delicate organization are maintained by several DNA associated proteins that also mediate gene regulation. The resultant compact structures involving DNA are called chromosomes. The most basic structural unit of the chromosome is the nucleosome (Fig. 2). They are formed by DNA segments wrapped around disk-shaped histone octamers and has a size ∼ 11 nm. The nucleosomes connected by DNA forms the so-called beads-on-a-string structure. A hierarchy of a structural organization is found from experiments on eukaryotic cells. The beads-on-a-string system folds into the 30 nm (width) chromatin fibre (Fig. 2). The chromatin fibre is further folded into loops [3]. Chromatin folding into loops was first observed in growing egg-cell nuclei by Walther Flemming in 1882. He called them lampbrush chromosomes (Fig. 3). It took a century to study their morphology clearly [7]. The simple protein-mdeiated organization is conserved in evolution. In bacteria, histone-like nucleoid structuring (HNS) proteins are found.

Two types of proteins can maintain the loops, (a) Nucleoid associated proteins (NAP) that act as passive cross linkers (CL) and (b) Structural maintenance of chromosome (SMC) proteins that often act like active loop extruding factors (LEF) [9]. The CLs diffuse around the environment to bind two segments of DNA if they come to spatial proximity. Their binding and unbinding can determine chromosomal compaction. In human chromosomes, ∼ 92% of DNA is euchromatin which is loosely compacted and accessible for transcription. The rest is densely packed heterochromatin. However, in prokaryotes like bacteria, all of the chromosomes is euchromatin. This suggests that heterochromatin evolved later, possibly as a mechanism to handle increasing genome size.

A. Passive cross-linking

The CLs bind to different chromatin segments to form loops with topologically simple and complex structures (loops inside loops) and zippers. As was found out from the numerical and analytical calculations by Amit Kumar, one of the graduate students in our group, at very high concentrations of CL (value?) chromatin gets into a highly compact and stable configuration [10]. In this state, transcription will be suppressed as it is energetically costly to access DNA segments. A living cell belonging to a particular tissue type can use such compaction for portions of DNA required to suppress protein production as in heterochromatin completely. On the other hand, at an intermediate concentration of CL, the chromatin fibre gets into a loosely packed critical state in which it is highly susceptible to slight variations in local CL concentration such that it can easily get into either compact or open shape switching between suppression and expression of DNA transcription. This behavior is similar to the so-called euchromatin. The scaling of chro-

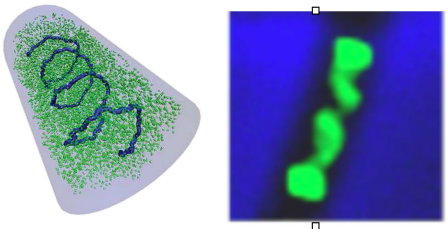

FIG. 4: (color online) (Left) The emergent helicoid shape of the backbone (blue) due to cylindrical confinement and side-loops (green), obtained from theoretical modeling and numerical simulations [8]. The model polymer shown in Fig. 3 spontaneously gets into the helicoid shape under cylindrical confinement provided by bacterial cells of E.coli or B.subtilis. (Right) Experimentally observed helicoid shape in B.subtilis chromosome adapted from Ref. [11].

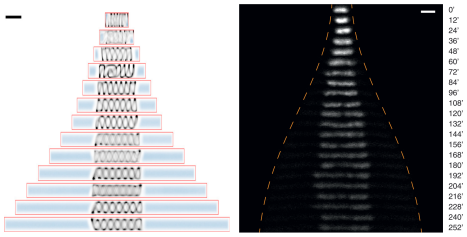

FIG. 5: (color online) The changing size and shape of nucleoid at constant crowder density in growing cell. The images show our simulation results (left) that agree well with growing E.coli cells in experiments (right) [17]. The scale bar size is 2 μm.

matin extension as a function of actual DNA length obtained for human chromosomes agrees well with Amit's prediction for the model chromosome at criticality, a concept central to continuous phase transitions studied in Physics. This suggests that the living cells may regulate CL concentration to stay near criticality in order to access easy switching between compact and open states.

B. Active loop extrusion

The other mechanism forming chromatin loops involves active loop extrusion by the so-called loop extruding factors (LEF), e.g., cohesin and CTCF complex. In interphase, the energy-consuming (ATPase activity) loop extrusion by cohesin grows symmetrically in both directions before it stops when its motion is stalled by CTCF attached to chromatin [12, 13]. This mechanism is found in eukaryotes and is preserved across evolution. In metaphase, a different kind of complex, condensins, lead to an active asymmetric loop extrusion [13].

Cohesin stays on DNA for $\sim$ 10 minutes in a processive manner, in the presence of Wapl that enhances cohesin detachment [14]. With the removal of Wapl, the processivity increases up to $\sim$ 6 hours and longer loops are extruded. This is observed in high-C maps showing longer-ranged contacts. A high probability of contact appears at CTCF locations pausing loop extrusion. The loop extrusion hypothesis [15] assumes cohesins stop at CTCF depending on their orientation.

The cellular or nuclear confinement, molecular crowding, passive CL and active LEF – all are responsible for the complex hierarchical organization of chromosomes. Earlier, Amit Kumar of our group studied the role of CLs in chromatin looping. Chitrak Karan, a current graduate student in the group, is studying the influence of active LEFs on chromosomal organization.

C. Emergent morphology of bacterial chromosome

This section mainly focuses on the emergent morphology of bacterial chromosomes. Appearing early in evolution, bacterial cells are much simpler than eukaryotes. Unlike eukaryotes, they do not possess membrane-bound organelles or motor-protein-based transport. In the absence of a nucleus, a single bacterial chromosome floats in the cytosol encapsulated by the cell membrane and wall. However, due to the tiny cell size of $\sim 1\,\mu$m, direct optical microscopy of bacterial chromosomes remained challenging as the size approaches the diffraction limit of optical wavelengths. With the advent of super-resolution microscopy and fluorescence labeling, live-cell imaging of bacteria has become feasible during the last two decades. Several interesting observations were made. Even in the absence of a bounding membrane, bacterial chromosomes have a compact shape occupying only a sub-volume of bacteria. Such a membrane-less organelle is called a nucleoid. It raises a question as to what maintains the compact structure of nucleoids. Moreover, the chromosome in different types of bacteria having cylindrical cell shapes showed a clear helicoid morphology [11, 16]. This poses the question of if there is a generic physical mechanism behind the emergence of helicoid shapes.

1. *Chromosome shape*

It turns out that a simple physical mechanism can explain the emergent helicoid shapes [8]. As we outlined before, chromatin fibre can form loops and zippers with the help of proteins. Side-loops and plectonemes (a loop of helices twisted together) can provide an effective bending rigidity to the chromatin backbone in the following manner. They can not occupy the same space due to local repulsion. In presence of such repulsions bent conformations that can potentially generate a higher probability of loop overlap are suppressed. Even contour-wise distant segments do not like to come too close spatially due to the local repulsion between loops. As a result, the chromosome assumes specific packing within the cellular

confinement. The emergent bending rigidity and packing, along with the shape of the cellular confinement, decide the emergent shape of the chromosome. In cylindrical confinement, the resultant structure is a helicoid. The emergent shape can be described entirely in terms of an effective entropy maximization of the chromatin fiber [8]. In other kinds of cell shapes, the emergent morphology is different, e.g., in disk-shaped cells, the same physical mechanism predicts that the circular DNA of bacteria will show a ring-like morphology.

2. Chromosome size

This leaves us with the other question as to what stabilizes the nucleoid size in the absence of a nuclear membrane. The chromatin and confinement alone suggest that the chromosome will expand to take the total cell volume in physiological cell volumes. To develop a better understanding, along with our experimentalist collaborators, we studied the change in size and shape of the chromosome in E.coli bacteria [17, 18]. Using ten times longer cells than the naturally occurring wild-type cells, we found that although the nucleoid is longer, it occupies only a fraction of the cell volume. What was missing was the consideration of cytosolic crowding in the cell. Once that is considered, it is straightforward to see that a force balance develops between the crowders and the chromosome that compresses the chromosome to the compact nucleoid. In a growing cell, continuous protein production maintains the crowder density. With increasing cell size, the nucleoid size first increases linearly to saturate at longer cells (Fig. 5). There are some subtleties involving various possible crowder sizes [19]. For example, a simple depletion between crowders and the chromosome will not be able to describe such compression. Crowder molecules with a size smaller than the typical mesh size can penetrate the chromosome volume and expand the chromosome so that it will touch the cell boundaries. However, the usual macromolecular crowders in a cell have a bigger radius of gyration than the typical chromosomal mesh size. As a result, it can generate the compression necessary to stabilize nucleoids.

3. Chromosome positioning

The central location of the chromosome, as can be observed from Fig. 5, can be maintained by the molecular crowders in the cytosol. The transcription of the DNA followed by translation generates new proteins. They are produced predominantly around the chromosome itself. In the absence of any directional bias, symmetric protein production lead to a perfect force balance between the cytosolic crowders to the left and right of the nucleoid. This symmetry and force balance keep the chromosome located at the center of the cell.

III. DISCUSSION

We described the complex structure of chromosomes starting from DNA and associated proteins. The complexity arises due to the involvement of several length scales and time scales. DNA has a width of 2 nm and persistence length of 50 nm, whereas the total length of DNA is ~ 1 mm in bacteria and 1 m in human cells. Such a long chain is housed inside a micron-sized bacterial cell or cell nucleus of size $\sim 10\,\mu$m.

In developing much of the physical understanding we employed numerical simulations. In principle, atomistic molecular dynamics of DNA can be performed, but even with the best computational resources, they are restricted by length scales of a few Angstroms to nanometers and time scales of a few femtosecond to microseconds. In contrast, the biologically relevant processes involve length scales of mm to m and times scales of several minutes to hours. As a result, one needs to coarse-grain the system to perform any reasonable computation. We used coarse-graining at the chromatin level. The calculations necessarily involve stochasticity due to thermal noise and the ongoing active biological processes at smaller lengths and time scales. The theoretical calculations involve techniques derived from statistical mechanics, both equilibrium and non-equilibrium, and soft matter physics including that of polymers and colloids. Mean-field calculations are routinely performed to gain insights, e.g., to obtain analytic estimates for the change in chromosome extension [18].

[1] R. E. Franklin and R. G. Gosling, Nature **171**, 740 (1953).
[2] J. D. Watson and F. H. C. Crick, Nature **171**, 737 (1953).
[3] B. Alberts *et al.*, *Essential Cell Biology*, 4th ed. (Garland Science, New York, USA, 2015).
[4] L. Pray, Nature Education **1**, 100 (2008).
[5] A. Wolffe, *Chromatin: Structure and function*, 3rd ed. (Academic press, London, 2012).
[6] K. Maeshima, S. Hihara, and M. Eltsov, Current opinion in cell biology **22 3**, 291 (2010).
[7] U. Scheer, W. W. Franke, M. F. Trendelenburg, and H. Spring, J. Cell Sci. **22**, 503 (1976).
[8] D. Chaudhuri and B. M. Mulder, Physical Review Letters **108**, 268305 (2012).
[9] I. F. Davidson *et al.*, Science (80-.). **366**, 1338 (2019).
[10] A. Kumar and D. Chaudhuri, J. Phys. Condens. Matter **31**, 354001 (2019).
[11] I. A. Berlatzky, A. Rouvinski, and S. Ben-Yehuda, Proc. Natl. Acad. Sci. U. S. A. **105**, 14136 (2008).
[12] I. F. Davidson and J. M. Peters, Nat. Rev. Mol. Cell Biol. **22**, 445 (2021).
[13] S. Golfier, T. Quail, H. Kimura, and J. Brugus, eLife **9**, e53885 (2020).
[14] A. Tedeschi *et al.*, Nature **501**, 564 (2013).

[15] G. Fudenberg *et al.*, Cell Rep. **15**, 2038 (2016).
[16] J. K. Fisher *et al.*, Cell **153**, 882 (2013).
[17] F. Wu *et al.*, Curr. Biol. **29**, 2131 (2019).
[18] P. Swain, B. M. Mulder, and D. Chaudhuri, Soft Matter **15**, 2677 (2019).
[19] A. Kumar, P. Swain, B. M. Mulder, and D. Chaudhuri, EPL (Europhysics Lett. **128**, 68003 (2020).