To Davşan

PROTEIN-DNA DISSOCIATION KINETICS AND CHROMOSOME ORGANIZATION IN A MODEL BACTERIAL CONFINEMENT

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

MATERIALS SCIENCE AND NANOTECHNOLOGY

By

ZAFER KOŞAR

September 2021

PROTEIN-DNA DISSOCIATION KINETICS AND CHROMOSOME ORGANIZATION IN A MODEL BACTERIAL CONFINEMENT

By Zafer Koşar

September 2021

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Aykut Erbaş (Advisor)

Asst. Prof. Seymur Jahangirov

Asst. Prof. Esra Yüca Yılmaz

Approved for the Graduate School of Engineering and Science:

Prof. Ezhan Karaşan

Director of the Graduate School

ABSTRACT

PROTEIN-DNA DISSOCIATION KINETICS AND CHROMOSOME ORGANIZATION IN A MODEL BACTERIAL CONFINEMENT

Zafer Koşar

M.S. in Materials Science and Nanotechnology

Advisor: Dr. Aykut Erbaş

September 2021

Transcriptional initiation and repression require the temporal interactions of transcription factors with DNA. Recent experiments showed that the interaction lifetime is crucial for transcriptional regulation. Relevantly, *in vitro* single-molecule studies showed that nucleoid-associated proteins (NAPs) dissociate rapidly from DNA through facilitated dissociation (FD) with the increasing phase-solution protein concentration. Nevertheless, it is not clear whether such a concentration-dependent mechanism is functional in bacterial confinement, in which NAP levels and the 3D chromosomal architecture are coupled. Here, we employ extensive coarse-grained molecular simulations, where we model the dissociation of specific and nonspecific dimeric NAPs from a high-molecular-weight circular DNA polymer in a rod-shaped structure constituting the cellular boundaries. Our simulations indicate that the peak cellular protein concentrations result in highly compact chromosomal conformations. Such compactions lead to shorter DNA-residence times but only for NAPs demonstrating sequence-specificity, such as the factor for inversion stimulation (Fis). On the other hand, the dissociation rates of nonspecific NAPs decrease with the

increasing protein concentrations, exhibiting an inverse FD behavior. Another set of simulations utilizing restrained chromosome models reveal DNA-segmental fluctuations as the cause of this reversed response, suggesting that moderate chromosomal compaction promotes protein dissociation. Together, our findings suggest that cellular quantities of structural DNA-binding proteins could be highly influential on their residence times and the chromosome architecture.

Keywords: Facilitated Dissociation, Transcriptional Regulation, Chromosome Organization, Molecular Dynamics Simulations

ÖZET

MODEL BAKTERİ HÜCRE SINIRLARINDA PROTEİN-DNA AYRILMA KİNETİKLERİ VE KROMOZOM ORGANİZASYONU

Zafer Koşar

Malzeme Bilimi ve Nanoteknoloji, Yüksek Lisans

Tez Danışmanı: Dr. Aykut Erbaş

Eylül 2021

Transkripsiyonel aktivasyon ve baskılama protein-DNA komplekslerinin geçici etkileşimine bağlıdır. Son çalışmalar bu etkileşimlerin yaşam sürelerinin de transkripsiyon süreci için hayati önem taşıdığını göstermektedir. Benzer olarak *in vitro* tek-molekül çalışmaları proteinlerin yığın derişimi kolaylaştırılmış ayrılma (FD) yoluyla arttıkça nükleoid-bağlı proteinlerin (NAP) DNA'dan hızla ayrıldığını göstermiştir. Bununla beraber söz konusu konsantrasyona bağlı mekanizmanın NAP seviyeleri ve 3D kromozom yapısının sıklıkla birbirine bağlı olduğu bakteri hücresinde fonksiyonel olup olmadığı bilinmemektedir. Bu çalışmada, spesifik ve spesifik olmayan dimerik NAP'ların yüksek moleküler ağırlıklı halkasal DNA molekülünden kopması, kapsamlı iri taneli moleküler simülasyonlar kullanılarak bakteriyel kromozomun hücre sınırlarını taklit eden çubuk biçimli bir yapı kullandık. Simülasyonlarımız fizyolojik olarak ilgili pik protein seviyelerinin çok sıkışık kromozom yapılarına sebep olduğunu göstermektedir. Bu tür çökmeler proteinlerin DNA'da kalma süresinin daha kısa olmasına sebep olsa da bu olay yalnızca DNA'ya spesifik olarak bağlanan inversiyon stimülasyon faktörü (Fis) gibi NAP'larda gerçekleşmektedir. Buna karşın nonspesifik

NAP’ların kopma hızının protein miktarı arttıkça düşmesi ters bir FD yapısına işaret etmektedir. Sabitlenmiş kromozon modeliyle çalıştırılan simülasyonlar söz konusu ters cevabın DNA’nın segmental dalgalanmalarına bağlı olduğunu ve ılıman kromozom çökmesinin protein ayrılmasını hızlandırdığını göstermektedir. Genel anlamda sonuçlarımız DNA’ya bağlanan proteinlerin hücre içindeki sayılarının bu proteinlerin DNA’ya bağlı kalma süresi ve kromozom mimarisyle yakından ilişkili olduğunu göstermektedir.

Anahtar Kelimeler: Kolaylaştırılmış Ayrılma, Kromozom Organizasyonu, Transkripsiyonel Regülasyon, Moleküler Dinamik Simülasyonları

Acknowledgements

First and foremost, I would like to thank my Supervisor Dr. Aykut Erbaş for his support throughout this study. He believed in me to complete this project even though I had no prior experience in anything computational.

For their invaluable comments and feedbacks, I am thankful to my MS jury members Dr. Seymur Jahangirov and Dr. Esra Yüca Yılmaz.

I want to thank my trainee Ali Göktuğ Attar for his extensive help. He followed me when I switched from wet-lab to a computational discipline.

I would also like to thank my previous trainees Alper Gençeroğlu, Fatma Chafra, Leyla Yalçınkaya, Mükrime Altun, Niran Çağıl, Tutku Muratoğlu, and Zeynep Yücekaya.

I want to thank my friends Julian Ostaku, Çişil Köksaldı, and Gökçe Özkul for their scientific and beverage supports.

I would also like to thank UNAM for their financial support throughout my master's studies.

Finally, I want to thank Zeynep Gültekin for her everlasting support through everything.

Contents

List of Figures

1 INTRODUCTION

The cells maintain their genetic information within their genomes. While the genome stores genetic information as DNA sequences in chromosomes, the gene expression depends on the transcription process. Transcriptional regulation governs the flow of genetic information through coordinating transcription in response to environmental or intracellular stimuli to ensure cellular survival. Among the many constituents of transcriptional regulation are transcription factors and chromosome architecture[1–7].

1.1 Transcription Factors as the Regulators of Transcription

As the central constituent of the central dogma, transcription is tightly regulated. The main constituents of transcriptional regulation are transcription factors (TFs). As DNA binding proteins, TFs demonstrate sequence-specific recognition and binding behaviors. They regulate the transcription process via promoting or repressing transcription by enhancing RNA polymerase recruitment or preventing its binding to the promoter via the temporal formation of TF-DNA complexes[1,8] (Figure 1).

Since there are several binding sites for TFs along the genome[3,8], their cytoplasmic concentration is also a significant regulatory factor. At low or vanishing concentrations, TFs have significantly reduced chances of binding to their specific sites because their localization on cis-regulatory elements on DNA depends on the diffusion caused by the thermal fluctuations in the cells. On the other hand, it is more likely for a binding sequence to be occupied by a TF at higher concentrations. Therefore, in the case of activation, transcriptional activity is expected to have a higher yield of RNA, thus, protein (vice versa in repression) in the abundance of TFs

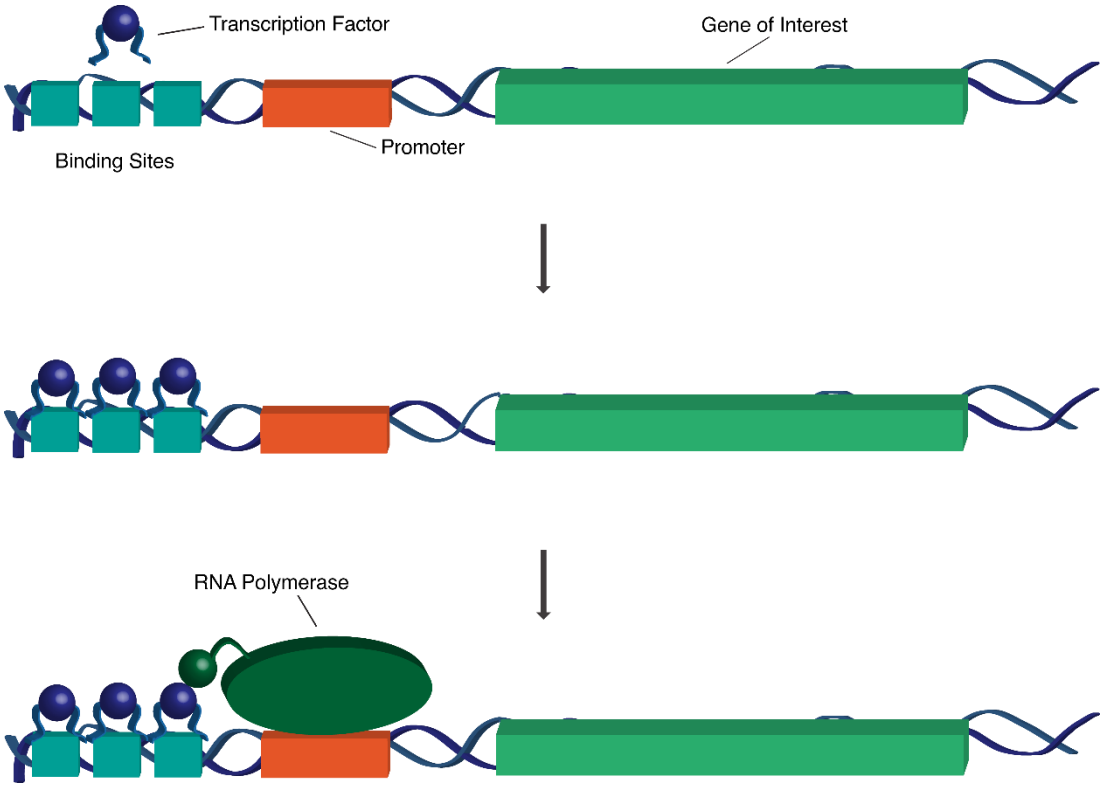

Figure 1 Schematic representation of RNA polymerase recruitment by transcription factors.

Indeed, this partly linear increase would reach its peak when all the binding sites regions are occupied. However, recent studies demonstrated that the duration a TF remains bound to its binding site could impact its transcriptional activity. Both for repression and activation, a TF should maintain its residence on DNA to allow proper initiation or repression of transcription. This duration of TF residence on DNA is called residence time (RT)[9,10].

Meanwhile, a different set of works demonstrated that a concentration-dependent mechanism, facilitated dissociation (FD), significantly affects the DNA residence times of various bacterial DNA-binding proteins. FD describes a molecular mechanism in which the competition between the DNA-bound and solution-phase (i.e., free) DNA binding proteins (e.g., transcription factors) for the same DNA site leads to enhanced dissociation rates of proteins[11–19]. These findings contradict the previous understandings on receptor-ligand kinetics, where the dissociation rates are considered constant for ligands[20].

1.2 Chromosome Architecture as a Regulator of Transcription

A chromosome primarily consists of double-stranded DNA (dsDNA), which forms a double helix arrangement. Due to its negatively charged phosphate backbone, DNA is self-repulsive[21,22]. Furthermore, it has a high persistence length of 150 base pairs (bps) or $\sim 50nm$ in its double-stranded form[23–25]. This combination forces DNA to have a highly relaxed conformation. Indeed, a completely relaxed or linearized chromosome would not fit within the cellular confinement boundaries. Nevertheless, histones in eukaryotes[26–28] and nucleoid-associated proteins (NAPs)[29–34] in prokaryotes are able to maintain the condensed formation of chromosomes.

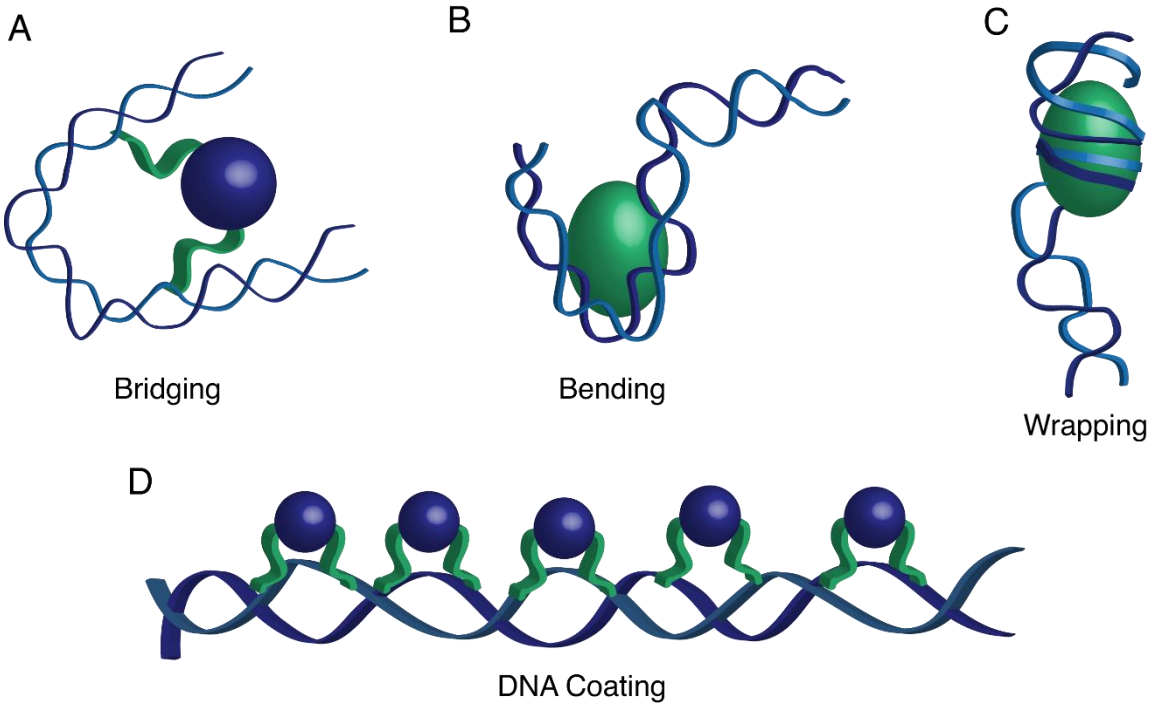

Figure 2 Examples of Nucleoid Associated Proteins Structuring DNA sequences.

Additionally, the nucleus constitutes a physical barrier of confinement for the genome serving a similar purpose for eukaryotes [35].

Although a chromosome collapses at every site, not every region has the same level of compaction. In highly compacted regions of the chromosome, TF binding sites are less exposed, reducing the likelihood of TF association. Chromosome condensation can also prevent RNA polymerase recruitment, thus impacting protein expression. In eukaryotes, heterochromatin regions are the more compact regions of the chromosomes, and they are transcriptionally less active compared to more relaxed euchromatin regions[36,37]. Moreover, the heterochromatin regions differ among the cell types emphasizing the role of chromosome architecture in transcription and thus in cellular behavior.

1.3 Nucleoid Associated Proteins

Unlike eukaryotes, bacteria have no organelle that compartmentalizes their genomes. Instead, the bacterial genome is confined within the cytoplasm. In prokaryotes, genome and associated proteins are maintained in the nucleoid[11]. Nucleoid Associated Proteins (NAPs) are a group of DNA binding proteins involved in chromosome structuring (Figure 2). Some NAPs can also act as transcription factors via binding specific regions along the chromosome and inhibiting or activating transcription at that region[5,6]. The NAPs that are directly involved in transcription are called dual-purpose NAPs[4].

The NAPs can bind to DNA specifically and nonspecifically. For instance, HU is a NAP that interacts with DNA only nonspecifically[38], and it has roles in chromosome

compaction, replication, and recombination in *E. coli* (*Escherichia coli)*. On the other hand, another *E. coli* NAP that is H-NS (heat-stable nucleoid-structuring protein), can bind to specific sequences of DNA and repress transcription in addition to its activity on nucleoid structuring[4,34].

The NAPs, regardless of whether or not they have additional functions by binding DNA, affect the chromosomal architecture by changing the trajectory of DNA via one or more processes of bridging, wrapping, and bending. Consequently, they alter the 3D conformation of the nucleoid, forming denser or more relaxed segments along the chromosome[33]. Therefore, in addition to their activity as TFs, they can also impact transcriptional regulation via chromosomal compaction or relaxation.

1.4 Fis as a Model NAP

One of the most extensively studied and well-characterized NAP is the Factor for inversion stimulation (Fis). Fis is a dual-purpose NAP of *E. coli*, and it binds to its target DNA in its homodimer form. Through bending, wrapping of DNA, and bridging of different regions of the chromosome, Fis plays a vital role in nucleoid structuring[13,14,30,39].

In the early exponential phase, when the Fis reaches its peak level (Figure 3D), the looped domains in the chromosome are most abundant, implicating the relationship between Fis concentration and nucleoid structure[17,40]. Fis can bind DNA nonspecifically throughout the genome, which could enable it to shape the chromosome effectively. It also binds to DNA in a sequence-specific manner with much greater affinity than that of nonspecific binding.

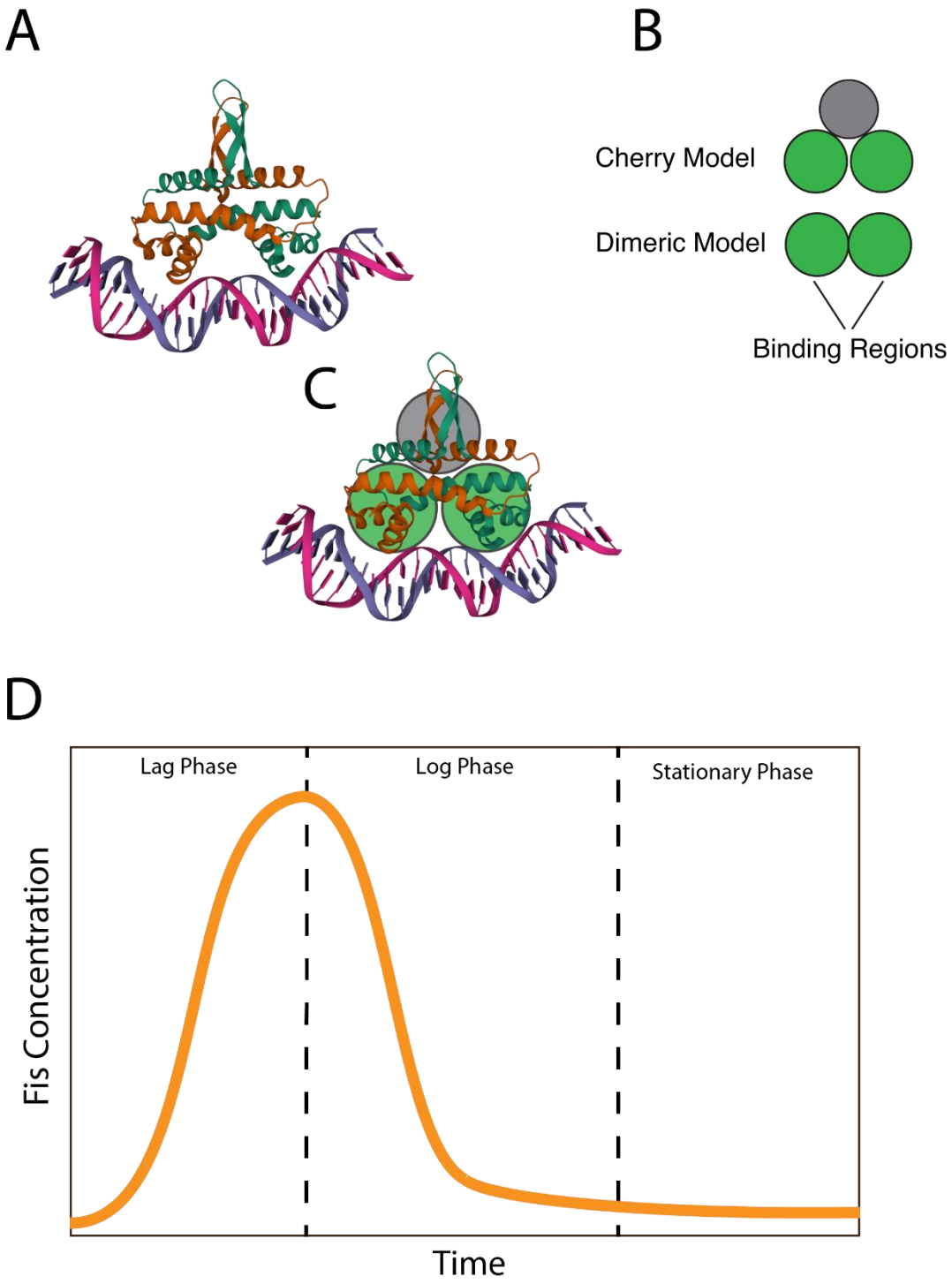

Figure 3 Coarse-grained model and the fluctuating levels of Fis. (A) Crystal Structure of Fis-DNA interaction. (B) Our cherry model and the previous dimeric model. (C) Depiction of cherry model and Fis resemblance. (D) The fluctuating levels of Fis.

On those binding sites, Fis could contribute to transcriptional regulation as either transcriptional activator or repressor. As a transcription factor, Fis regulates more than 200 genes, including itself (i.e., auto-regulation)[17]. Therefore, understanding the kinetics of binding and dissociation of Fis as a model transcription factor would further enhance our understanding of transcriptional regulation.

1.5 Dissociation Rates of DNA Binding Proteins

Receptor-ligand binding kinetics are defined via k_{on} (binding rate), k_{off} (dissociation rate) (Figure 4), and the ratio between them $k_d = k_{off}/k_{on}$, which is the equilibrium dissociation constant[20]. Although these rates are beneficial for understanding receptor-ligand binding kinetics, they neglect the aforementioned impact of concentration on dissociation rates.

Taking FD into consideration, DNA-protein binding kinetics severely change, rendering the notion of constant off-rates pointless. Since the concentrations of the previously mentioned NAPs (i.e., Fis, HU, H-NS) fluctuate from several micromolar to >50μM along a cell's life cycle[41,42], their off-rates should also fluctuate in accord with the cellular phases.

In very low concentrations (i.e., $< 1\mu M$), there is almost no competition for the binding sites. In such cases, TFs depend on the spontaneous dissociation (SD) to unbind from DNA. Since the dissociation via SD is slow, residence times of proteins should be higher in low concentrations [19].

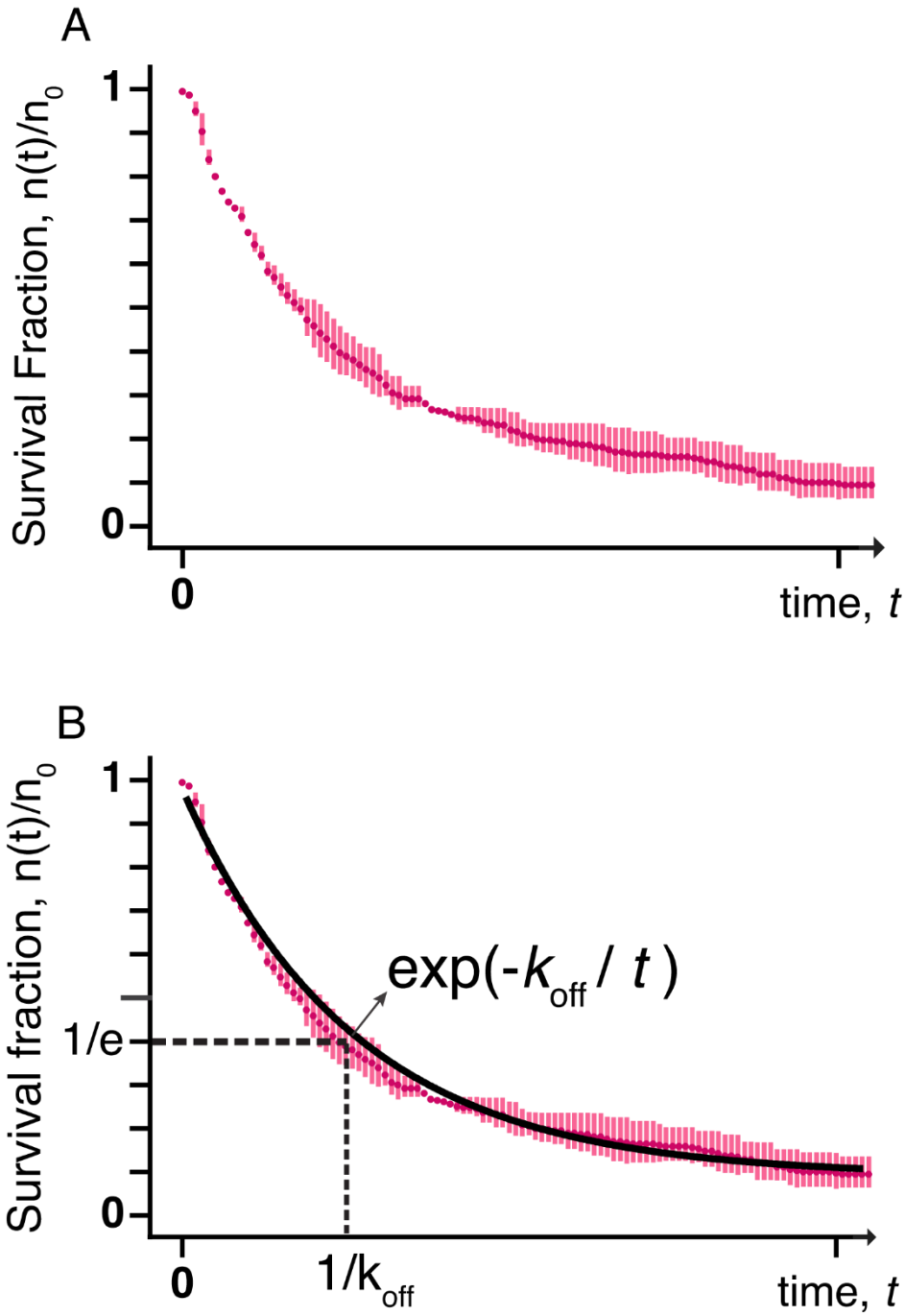

Figure 4 A sample survival fraction graph to obtain off-rates. (A) A decaying dataset of $n(t)/n_0$. (B) Fitting a single exponential decay curve to the dataset.

In high concentrations (i.e., tens of micromolar), transcription factors compete for the same binding site forming a ternary structure. This competition may result in partial (or microscopical) and eventual dissociation of the initially bound transcription factor[19,43]. Since the higher concentration of transcription factors suggests more competition, the dissociation rates increase with increasing concentration through the process of facilitated dissociation (FD). Consequently, the residence times of the proteins should drop drastically. Consecutively, the reduced residence times could lead to a decreased level of transcriptional activation or inhibition depending on the nature of the TF-DNA interaction (i.e., inhibitory or activatory)[9].

In the case of NAPs, an increase in their concentration also affects chromosomal compaction. As the NAP concentration reaches its peak levels, nucleoid compaction becomes more apparent[4,11]. As mentioned previously, chromosomal condensation could significantly affect transcriptional regulation through changing residence times or RNA polymerase accession. Therefore, the change in concentrations of the NAPs may not only alter residence times by competition but also by modifying the nucleoid architecture.

1.6 The System for the Molecular Dynamics Simulations

Here, we introduce a design of *E. coli* bacterium for Molecular Dynamics (MD) simulations with a novel coarse-grained cherry model of Fis resembling its crystal structure[39] (Figure 3) and a system where the chromosome is adequately[44] condensed and confined within a rod-like membrane structure constituting boundary (Figure 5).

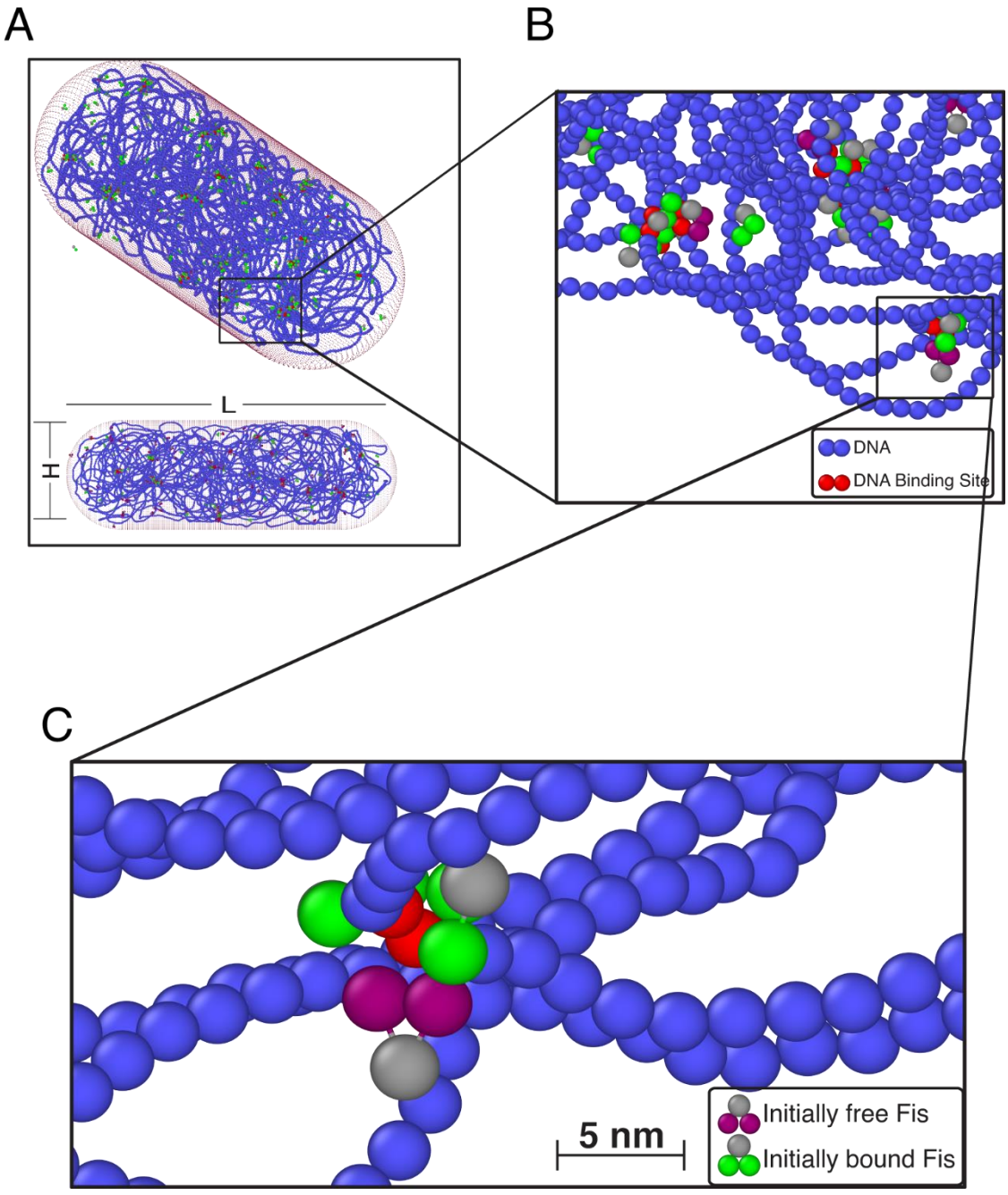

Figure 5 Coarse-grained *E. coli* model. (A) An overall picture of the system with membrane dimension where L = 3H. (B) Close-up image of the chromosome depicting specific and nonspecific sites. (C) A zoomed-in image of proteins (Fis) interacting with DNA binding sites.

In order to enable the testing of FD in bacterial confinement, the circular chromosome includes specific binding sites, transcription factors that are initially bound to those sites, and free transcription factors that are initially not in contact with binding sites. While the number of bound transcription factors remained the same in all simulations, the number of free transcription factors changed with the desired concentration and volume. We expected a change in residence times of initially bound transcription factors via changing free transcription factor concentration through facilitated dissociation. Additionally, the system allowed us to investigate the effect of transcription factor concentration, affinity, and the role of nonspecific protein-DNA interactions in chromosome compaction. Furthermore, we were also able to investigate the impact of chromosome condensation on the residence times.

1.7 Previous Studies and the Significance of Our Study

Previous studies show that NAPs, including Fis, H-NS, and HU, undergo a concentration-dependent dissociation mechanism, FD [11–19]. It was also demonstrated that NAPs could affect the 3D conformation of the nucleoid via forming dense DNA-multiprotein complexes or clusters[2,45]. Therefore, the fluctuating concentrations of NAPs alter the very environment from which they dissociate. That, of course, complicates TF dissociation and FD effectiveness at the cellular level.

Notably, Fis proteins dissociate faster (i.e., higher off-rates) with increasing concentration from extended (or relaxed) DNA segments or sparsely protein-occupied DNA binding sites[16,46,47]. However, when the Fis proteins are involved in the formation of DNA looping or coating, they form highly stable protein-DNA

complexes[48]. Different experiments switched the competitor molecules from Fis to DNA segments. That, in fact, also increased the dissociation rates of Fis molecules exhibiting a DNA-segmental type of concentration-dependent dissociation (segmental FD)[49]. This behavior suggests a relationship between nucleoid architecture and residence times.

Accordingly, metalloregulator ZuR and CueR exhibited chromosome conformation-dependent dissociation rates. However, unlike NAPs, they are not responsible for chromosome compaction since their cellular concentration levels do not exceed a few micromolar concentrations[50,51].

In addition to Fis, HU, a nonspecifically interacting NAP, exhibits a similar behavior where its FD rates depend on the conformation of the DNA polymer[52]. H-NS, another nonspecifically interacting NAP, showed a similar pattern where its dissociation relies on the structure of the DNA-protein complex containing the H-NS molecules[53]. Combined, these findings suggest that the chromosomal conformation at the cells would affect the TF dissociation rates. However, the influence of NAP concentration in the interplay between nucleoid compaction and FD is beyond the reach of those studies. Thus, it requires another set of systemic research.

Here, we designed an *in-silico* system to reveal a latent transcriptional regulation mechanism consisting of NAP concentration, chromosome organization, and FD effectiveness in a native-like environment. To do so, we utilize a large-scale coarse-grained model of *E. coli* bacterium for molecular dynamics (MD) simulations, most prominently to reveal the network of connections of chromosome structure, dissociation rates, and NAP concentrations (Figure 6). Although similar MD models

have been utilized investigating chromosome organization, dissociation kinetics were not a part of those studies[53–55].

In our simulations, we employ a dimeric *cherry* model resembling specific and nonspecific NAPs that are Fis and HU, respectively. From zero to up to several thousands of free NAPs (i.e., $0 - 200\mu M$) competes for 120 binding sites, which are initially occupied by the same kind of NAPs. Here, NAPs can dynamically interact with the self-avoiding DNA polymer (chromosome) and alter its conformation. The system enables us to track all the proteins spatiotemporally (i.e., positions with respect to time). Therefore, it allows us to monitor NAP dissociations and chromosome organization simultaneously.

Examining several NAP concentrations and affinity levels demonstrated that FD could take place in cellular confinement, but its effectiveness depends on nucleoid compaction, which relies on NAP concentration and affinities. At high nonspecific affinities, even with $60\mu M$ NAP concentrations, chromosomes formed well-ordered DNA-protein structures, including donut-shaped, filamentous, and network forming clusters.

We inspect several NAP concentrations with changing affinities, trying to shed light on FD behavior on cellular confinement. We also demonstrate the effect of chromosomal architecture on FD on those conditions. Altogether, NAP concentration, through FD and nucleoid structuring, could be a regulatory factor for transcription.

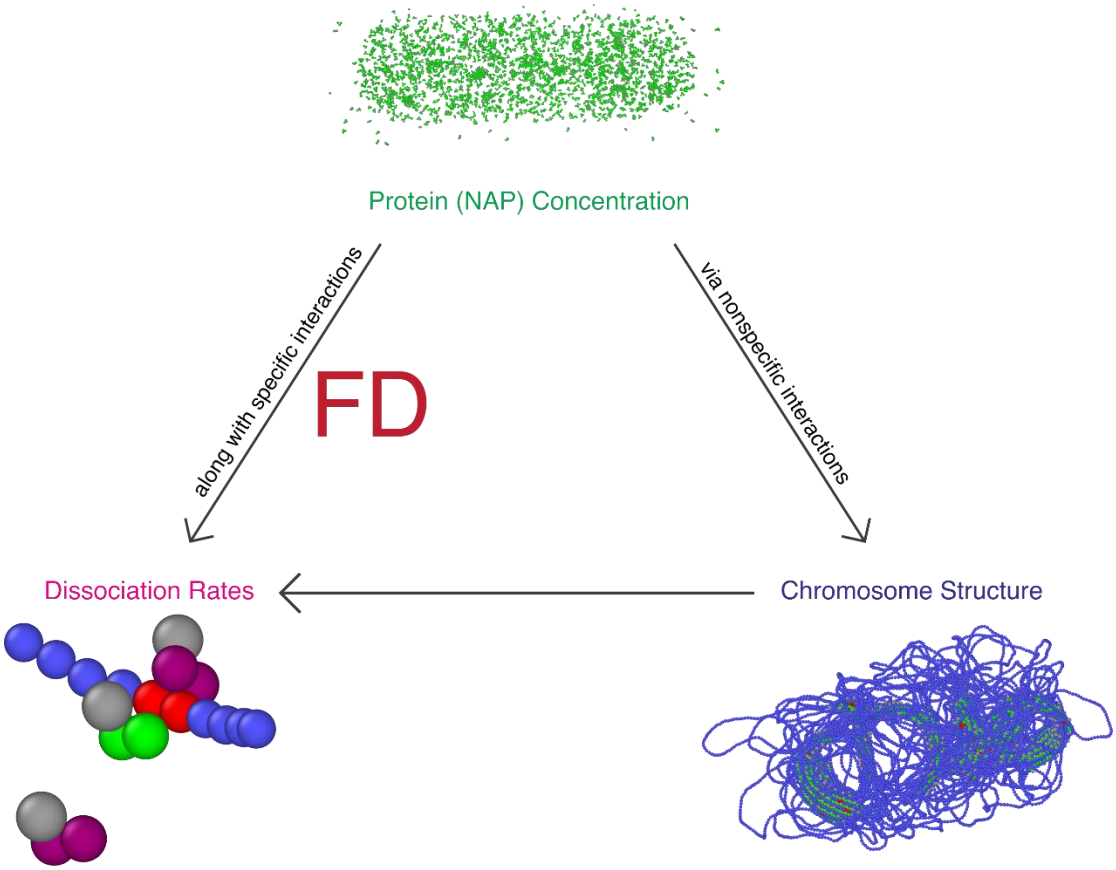

Figure 6 Graphical abstract of our study demonstrating the network of relations among the protein concentration, chromosome organization, and the dissociation rates.

2 METHODOLOGY

2.1 Chromosome Design

2.1.1 Bacterial Chromosome Model

A DNA polymer is formed via the covalent attachment number of nucleotides. Even though there are some exceptions, DNA mainly exists in its double-stranded (i.e., dsDNA) form in which one strand is complementary to the other[21]. However, when exposed, the end of the dsDNA is treated as DNA damage by the intracellular DNA repair mechanisms[56]. Therefore, both in eukaryotic and prokaryotic cells, dsDNA ends must be closed (or sealed).

Telomere regions of human chromosomes form a complex structure to enclose ends of the chromatins in eukaryotes[57]. Contrarily, prokaryotes such as *E. coli* utilize a much simpler mechanism. As the ends of the dsDNA are compatible, merging those ends is possible and renders the *E. coli* chromosome a circular dsDNA structure[33]. Consequently, all models of the *E. coli* chromosome should be of circular nature.

Another concern with the bacterial chromosome model is that every nucleotide is composed of many atoms. Although their chemical compositions are very well-known[58], using their full atomistic model would be computationally costly. Therefore, as molecular dynamics (MD) simulations allow, we can use much more simplistic DNA models.

In the coarse-grained models, complex structures could be reduced to simple representative models. However, this simplification process should be done so that

the model should resemble the structural features of the molecule. Since a bacterial chromosome (e.g., *E. coli* chromosome) is a dsDNA[33], and the parts of dsDNA are structurally similar, several base pairs (nucleotides with their complementary nucleotides) could be represented as one bead in a simulation system.

Our system initially aimed to characterize the facilitated dissociation (FD) of the Factor for inversion stimulation (Fis). The FD requires multivalency (i.e., binding with multiple sites) for binding site competition and partial (or microscopical) dissociation[18,43]. Therefore, both the Fis binding site (i.e., part of the DNA that Fis binds more firmly) and binding domains of Fis protein should be at least two beads. Accordingly, we used a two-bead coarse-grained model to represent DNA binding sites. For Fis protein, these binding sites are around 20 base pairs (bp)[59,60]. Thus, our model uses one bead per $\sim 10bp$ approach. The approach, of course, was also applied to the rest of the chromosome.

Even though one bead per $10bp$ strategy provides a great deal of simplicity, our model organism *E. coli* has 4.6 million base pairs ($4.6Mbp$) in its chromosome[61]. We would have had to work with around 460 thousand beads even with this approach, which again would be computationally expensive. Thus, we also needed to reduce the number of base pairs we represent in our model.

Initially, we decided to use an $N = 2400$ bead model, which is equivalent to $24{,}000\ bps$. With 20 bead spaces between binding sites, we had 120 binding sites, which could provide statistical significance in dissociation rates of proteins.

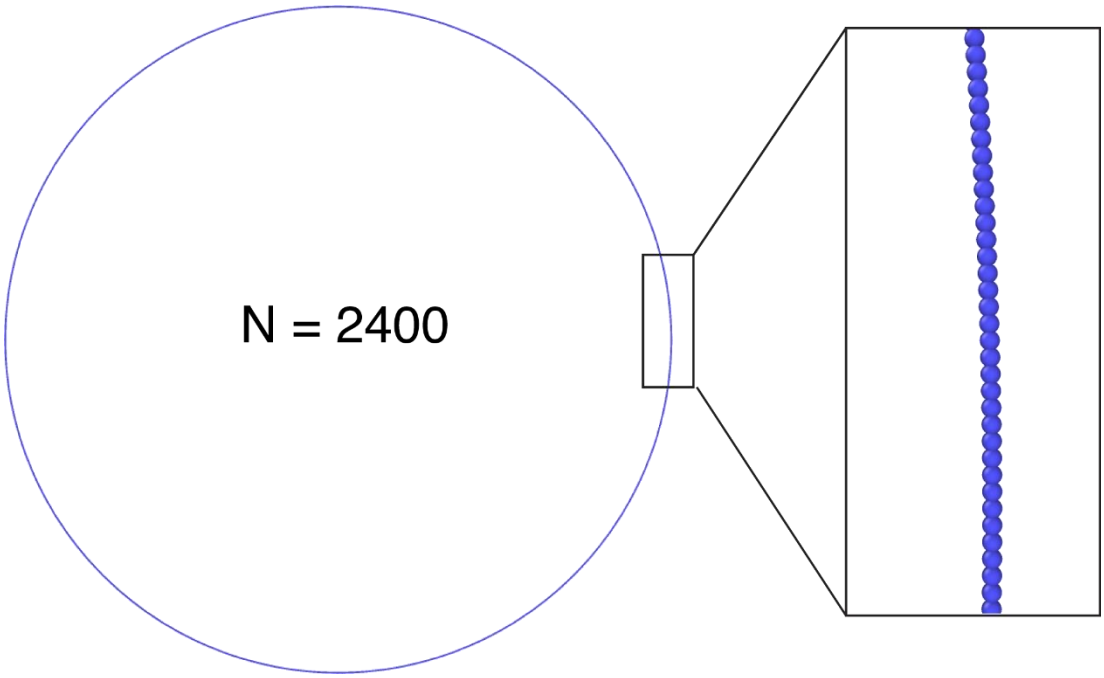

Figure 7 The snapshot of our initial circular chromosome and depiction of its constituents.

2.1.2 Designing and Collapsing the Chromosome

Chromosomes -whether in eukaryotes or prokaryotes- are highly condensed structures. For instance, the *E. coli* genome is around $1.5cm$ long if linearized. However, *E. coli* is around $2\mu m$ in length and $0.5\mu m$ in width[62] (Figure 8). Therefore, the *E. coli* genome must be folded at least a thousand times to fit within the cellular boundaries. Accordingly, our chromosome model should follow that pattern. Unfortunately, directly creating a fully collapsed structure would be mathematically complex and inevitably deterministic.

Rather than creating the chromosome structure in its final form, we utilized a different, much simpler, and more straightforward approach. We first made a data file (i.e., a file where the coordinates of the atoms are stored) of a perfectly circular polymer (Figure 7) using the following equation.

$$x = \cos(\theta) \times r$$

$$y = sin\ (\theta) \times r$$

while $-\pi < \theta \leq \pi$ with $2\pi r$ times increment to ensure the distance between the neighboring beads is a σ which corresponds to $3.4nm$ in our coarse-grained model.

This initial polymer was $N = 2400$ beads in length and had a radius of $1200\sigma/\pi$ (i.e., $1.3\mu m$), considering the distance between each bead is one σ (or $3.4nm$). Similar to the bacterial chromosome, the bead indexed first attaches to the last bead completing the circle. Due to its perfectly circular nature, the initial DNA polymer was

too loose, thus too big to fit within the necessary boundaries. Therefore, we needed to collapse it.

Simply lowering the distance between beads from one σ to much lower values didn't provide solutions, as the LAMMPS molecular dynamics simulation package gives out an error when the beads are initially too close to each other. That is because when the beads are too close, they repulse each other. This repulsion is so extensive that they move away to a distance which the FENE (finitely extensible nonlinear elastic) bond doesn't allow due to its finitely extensible nature, as the name suggests.

Another strategy is to utilize the LAMMPS MD package to simulate circular DNA polymer and collapse it onto itself. Accordingly, using the data file for circular DNA and input file (i.e., a file where the instructions, parameters, etc., for LAMMPS MD simulations, are stored) with high self-attraction affinities of $17kT$ among the DNA monomers, the circular polymer was forced collapsed onto itself. The simulation ran in 2×10^7 simulation steps with a timestep of $\Delta t = 0.005\tau$, which in total corresponds to $25ms$ considering a τ is approximately $250ns$. Those parameters were sufficient for the polymer condensation. Consequently, we were able to obtain a fully collapsed circular DNA polymer (Figure 9).

2.1.3 Chromosome Relaxation

Previously obtained fully condensed circular polymer (or chromosome) satisfied the condition to fit within the membrane boundaries. However, this condensation is extensive and thus, contradicting the partially relaxed nucleoid of *E. coli*[44]. Therefore, we needed to relax the collapsed polymer.

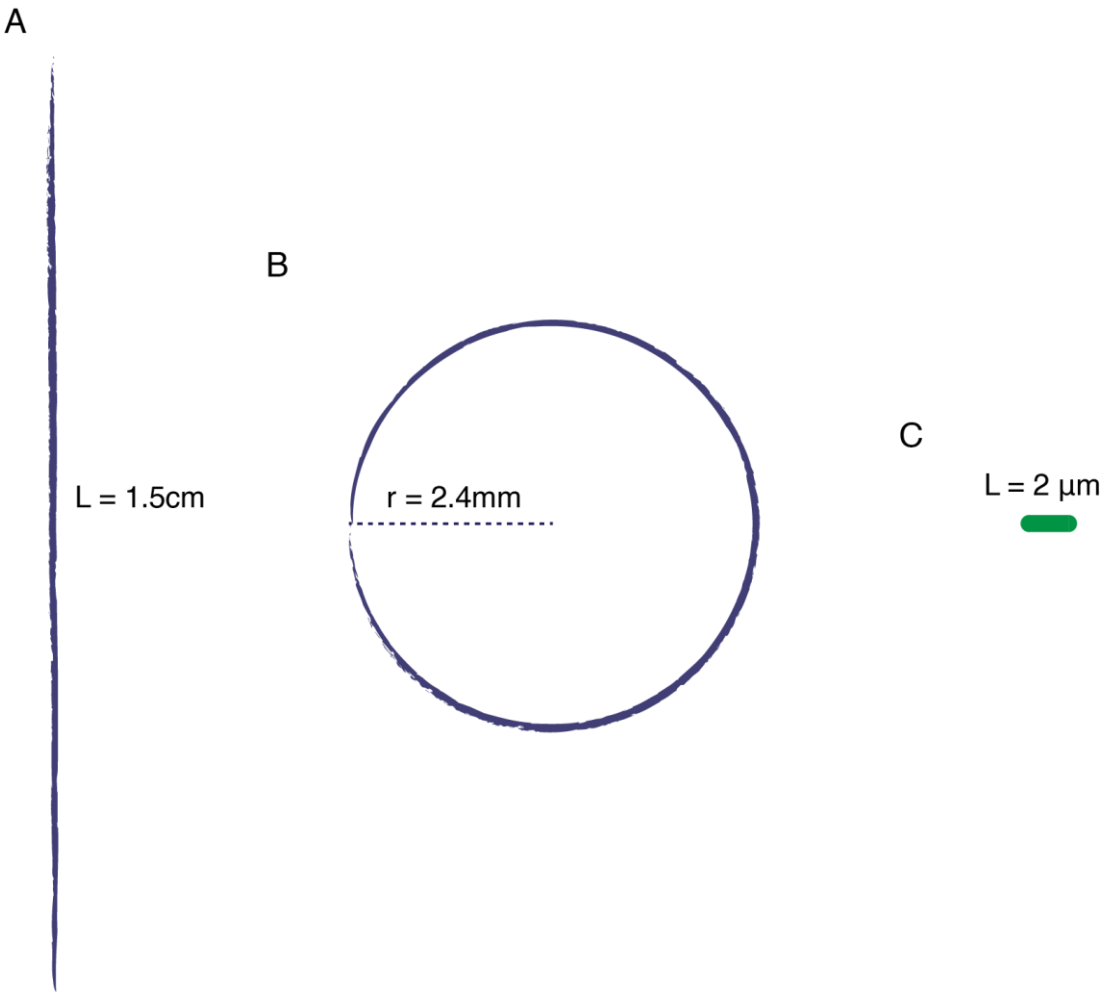

Figure 8 Size comparison of the *E. coli* chromosome (A) if it is linearized, (B) in its fully circular form, and (C) the size of an average *E. coli* bacterium. Note that the drawings are just representative and not to scale.

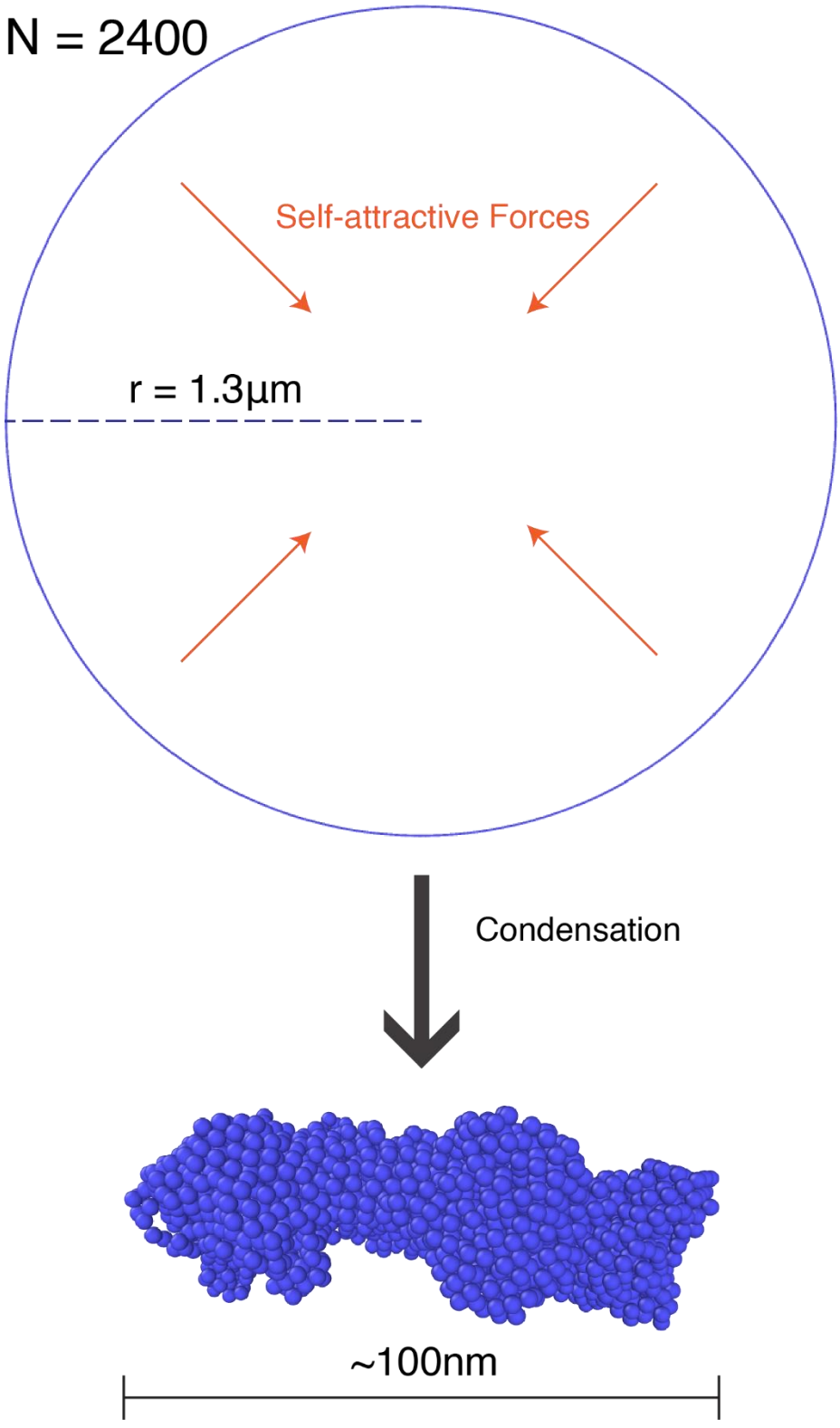

Figure 9 The process for obtaining a highly condensed chromosome via the self-attractions to fit chromosome within the cellular confinement boundaries.

To do so, we ran simulations of the chromosome with self-repulsive interactions among the chromosome monomers. Self-repulsive interactions were appointed via setting interaction distance threshold to $r = \sqrt[6]{2}\sigma$ or 1.122σ a distance below which the repulsive interactions start taking over according to Lennard-Jones potential represented by the following equation[63,64].

$$4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

With less than 6×10^4 simulation steps and a timestep of $\Delta t = 0.005\tau$, we could relax the initial $N = 2400$ bead model enough to mimic the natural condensation level of *E. coli* nucleoid, which occupies around 40-60% of the cytoplasmic volume [44].

2.1.4 Polymer Extension

Although $N = 2400$ bead model provided a sufficient resemblance to the *E. coli* chromosome, one binding site per 20 beads model restricted other sites from moving freely due to the excessive bridging of the binding sites via transcription factor bindings. Therefore, we decided on employing a much longer polymer.

We first tried to utilize our initial strategy mentioned in which we begin with a perfectly circular polymer then collapse it using the LAMMPS MD package. However, this approach did not work well with the extended polymer $N = 12000$. Although there were local collapsed regions, the polymer did not collapse onto itself entirely. Even after 2×10^8 simulations steps (i.e., ten times more than the initial model) with a timestep of $\Delta t = 0.005\tau$, we were unable to condense the polymer further than obtaining some local clusters.

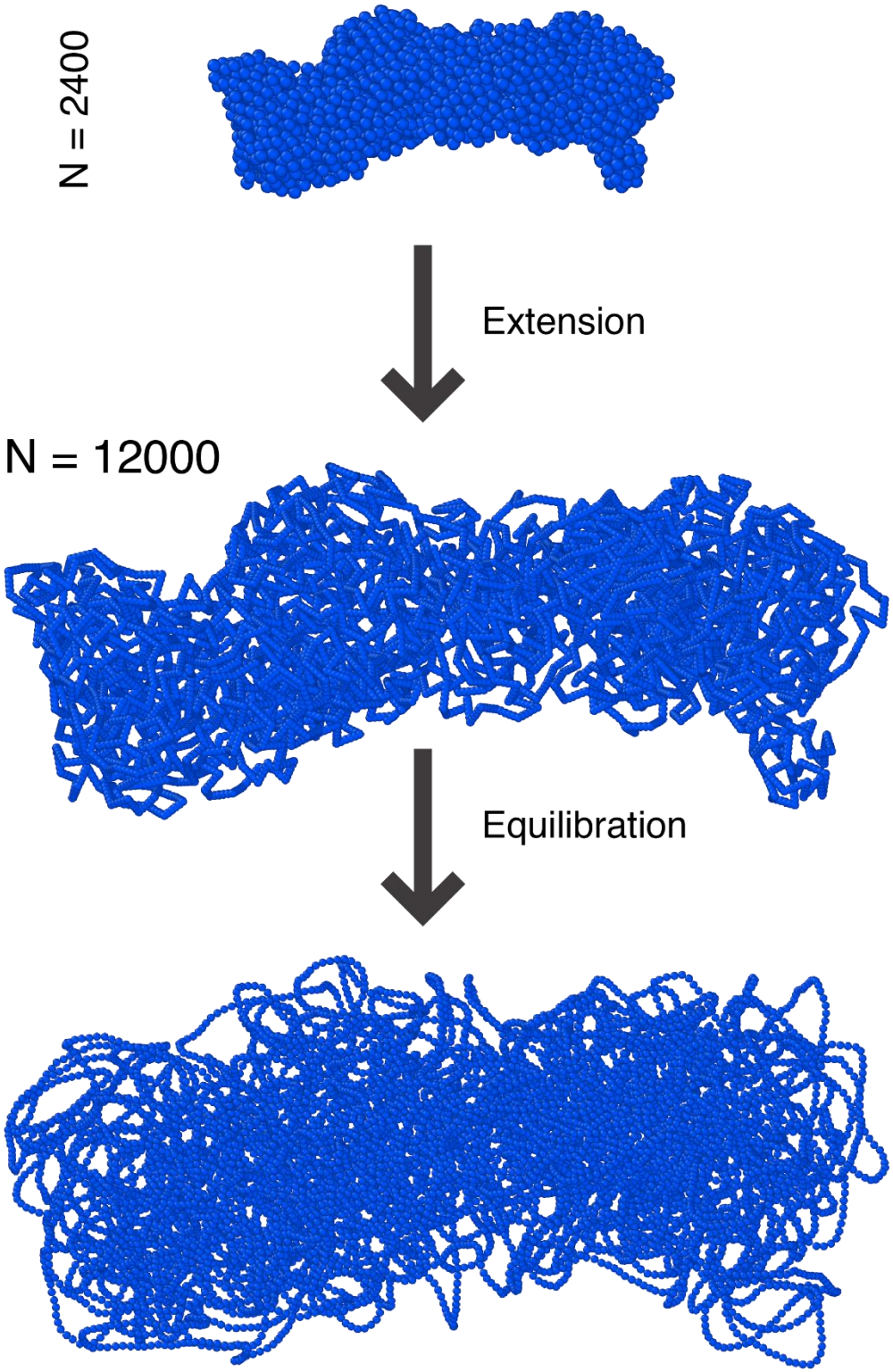

Figure 10 The schematic demonstration of obtaining the mildly-relaxed extended DNA polymer.

Failure to obtain a collapsed extended polymer meant that we needed to employ a different strategy. Rather than increasing the polymer size before condensing it, we realized we could extend the previously collapsed $N = 2400$ to $N = 12000$ model. To do so, we first needed to acquire the positions of the collapsed polymer. Since the coordinates of the beads were already present in a LAMMPS data file, obtaining the locations was quite a straightforward process.

Next, we needed to place the $N = 12000$ beads to their relevant positions. Therefore, our following challenge was to figure out the coordinates for those beads. Since the structure was already too dense, there was no way to position $N = 12000$ beads to $N = 2400$ bead space. Thus, we needed to enhance the occupied space. Hence, we chose to multiply x, y, z coordinates with a number proportionate to the extension factor for all the beads in the collapsed $N = 2400$ bead polymer. To obtain enough space for $N = 12000$ bead polymer, we multiplied coordinates with $\sqrt[2/3]{5}$ where 5 is the expansion ratio $(12000/2400)$ (Figure 10).

$$(x_m, y_m, z_m) = x_m, y_m, z_m \times \sqrt[2/3]{5}$$

Then, we shifted the IDs of the beads via multiplying them with the extension factor (bead number 1 became number 5 and so on). This process enabled the integration of extra four beads to the space between the initially neighbor beads. The filler beads were evenly spaced and positioned between the beads with IDs n and $n + 5$ using the following calculations.

Changes in coordinates are,

$$\Delta x = x_{n+5} - x_n, \ \Delta y = y_{n+5} - y_n, \ \Delta z = z_{n+5} - z_n$$

Therefore, the spacing required between each bead is,

$$\Delta\vec{r} = (\frac{\Delta x}{4}, \frac{\Delta y}{4}, \frac{\Delta z}{4})$$

Thus, the positions of the filler beads are,

$$P_{n+1} = P_n + \Delta\vec{r}$$

.

.

$$P_{n+4} = P_n + 4\Delta\vec{r}$$

Although this approach creates some undesired straightness, the issue is solved via simulating the resulting data file using the LAMMPS MD simulation package. The intermediate step for polymer dispersion was also the equilibration step which is necessary for all our simulations. Therefore, we did not have to employ any extra procedure solely for disrupting the local straightness. Additionally, since we already expanded the positions of the beads, we did not need to relax the polymer further.

In this expanded polymer, we placed evenly spaced 120 dimeric binding sites on DNA similar to the initial $N = 2400$ construct. Contrary to the initial 20 bead spacing in the $N = 2400$ system, the expanded $N = 12000$ system has 100 bead spacing between the dimeric DNA binding sites. Therefore, there is sufficient spacer DNA to mimic the flexibility of the *E. coli* chromosome while preserving a statistically adequate number of binding sites to analyze dissociation rates.

2.2 Bacterial Cell Wall and Membrane

2.2.1 Membrane Architecture

Bacterial morphology (i.e., the cell shape) varies between species. The most common three shapes are coccus (spherical), spiral, and bacillus or rod-shaped. Our model organism, *E. coli*, has a rod-shaped morphology due to its cell wall architecture[65].

To obtain such morphology in a coarse-grained model, first, we need to simplify the rod shape of the bacterial membrane to an architecture represented via the basic geometric shapes. In its most simplistic form that preserves its unique morphology, the *E. coli* membrane could be formed via two semi-spheres covering the ends of a cylinder in a coarse-grained model (Figure 11). As expected, radius (r) is the same for both the cylinder and the semi-spheres.

2.2.2 Membrane Calculations

The bacterial inner membrane (or cytoplasmic membrane) is a part of the cell that designates the cytoplasmic borders. Bacterial chromosomal and extrachromosomal DNA (i.e., plasmids) and nucleoid associated proteins are located within the cytoplasm[4]. Therefore, for simplicity, we integrated neither the outer membrane nor the periplasm separately into our system. Instead, we approached the bacterial cell wall as a whole and focused on ensuring the bp/V (base pair to volume) ratio similar to that of *E. coli*.

Although genomic base pair number slightly varies among the strains of *E. coli*, one of the most widely used strains, K-12, has approximately 4.6 million base pairs

($4.6\ Mbp$) in its chromosome[66]. Accordingly, we calculated the percentage volume occupied by the chromosome. To do so, we considered double-stranded DNA as a cylinder with a radius of $r = 1nm$ and a length of $0.34nm \times 4.6 \times 10^6$ (i.e., the distance between base pairs[67] multiplied by the number of the base pairs in the *E. Coli* genome). Next, based on the literature, we took *E. coli* cytoplasmic volume as $0.67\mu m^3$ [68].

Therefore,

$$V_{cylinder} = \pi r^2 h$$

$$\mathrm{V_{chromosome}} = \pi(10^{-9}m)^2 \times (0.34 \times 10^{-9}m) \times (4.6 \times 10^6)$$

$$= 4.91 \times 10^{-21}m^3$$

Thus,

$$\text{Chromosomal density} = \frac{4.91 \times 10^{-21}m^3}{0.67 \times 10^{-18}m^3} \times 100\%$$

$$= 0.73\%$$

Note that DNA the $Volume/Volume$ (V/V)percentage in *E. coli* is 1% in the literature[69]; however, we are only taking chromosomal DNA into account. Therefore, this slight difference is the result of the exclusion of extrachromosomal DNA.

Next, we needed to decide the cylinder length to radius ratio to suit the *E. coli* rod shape structure in the closest way possible. After several trials and visual inspection of the models, we decided to use the $4r = L$ ratio as it resembled the *E. coli* shape

the best. Therefore, in parallel with our model in which a cylinder is joined by two semi-spheres,

$$\mathrm{V_{system}} = V_{cylinder} + V_{sphere}$$

$$= \pi r^2 4r + \frac{4}{3}\pi r^3$$

$$= \frac{16}{3}\pi r^3$$

Then, as we aimed to preserve V/V ratio of DNA to the cytoplasm of *E. coli*, the system volume depended on the base pair number. Since the only variable that determines the system volume is the radius, we could calculate the radius with respect to the number of base pairs (N_{bp}) using the following equation.

$$V_{system} = \frac{V_{chromosome}}{0.0073}$$

$$\frac{16}{3}\pi r^3 = \frac{0.34nm \times N_{bp} \times \pi(1.0nm)^2}{0.0073}$$

Thus,

$$r = 2.06nm \times \sqrt[3]{N_{bp}}$$

Inputting the $4.6\,Mbp$, the length of *E. coli* K-12 genome to this radius equation would give a radius of $r = 343\,nm$. Considering $W = 2r$ and $L = 6r$ are the measures of our system, a system with $4.6\,Mbp$ would have a width of $W = \sim 0.7\,\mu m$ and a length of $L = \sim 2\,\mu m$ which satisfies *E. coli* measures[70], thus verifying our system boundary calculations.

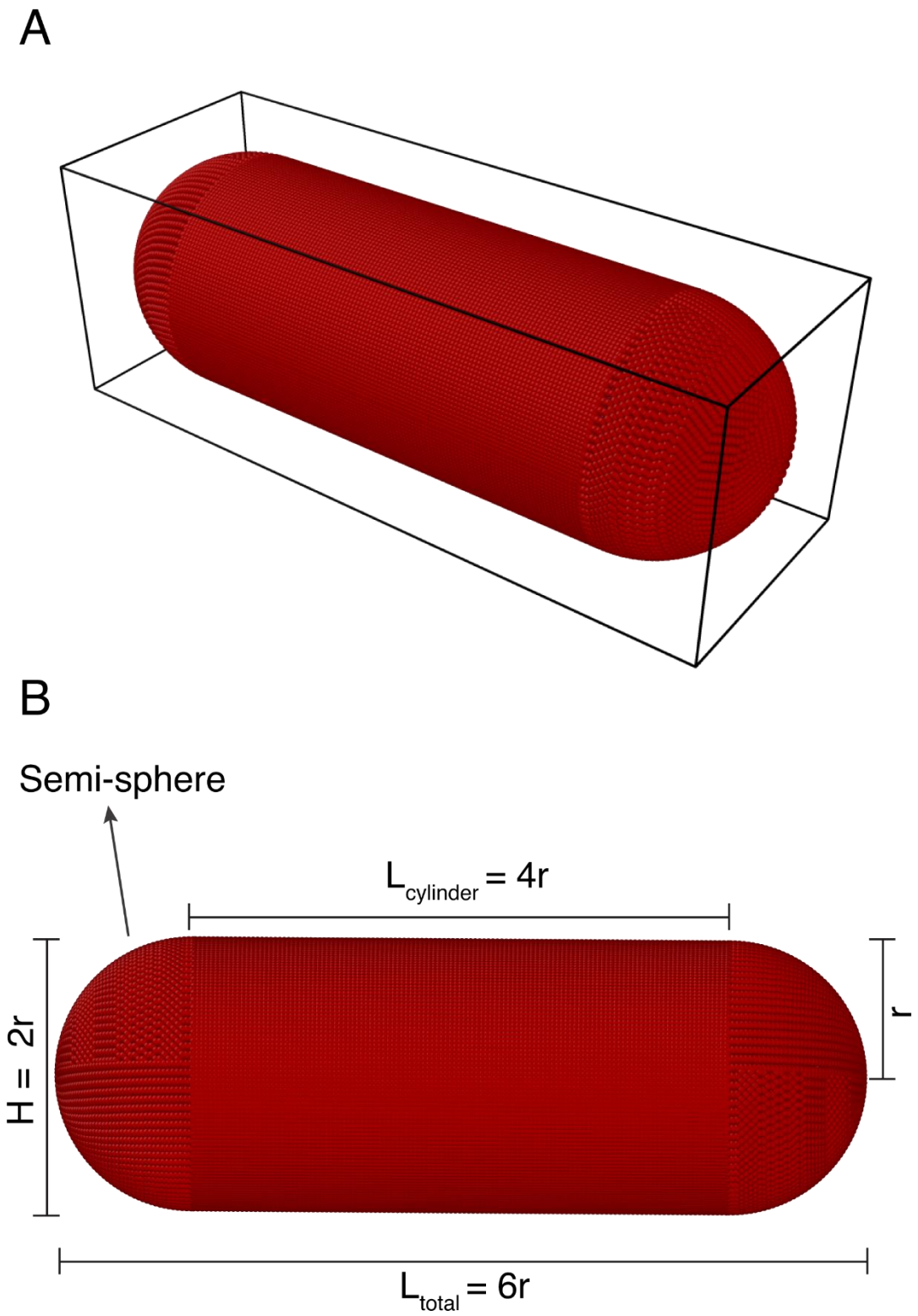

Figure 11 The coarse-grained model of *E. coli* confinement boundaries. (A) A perspective view of the model. (B) The dimensions of the model.

Another, more straightforward approach for such boundary calculations would be to utilize base pair to volume ratio (N_{bp}/V). This method, which we employed, eliminates the need to calculate the DNA volume for both sites that ultimately cancel each other out. Such equation could be used like the following,

$$\frac{N_{bp}}{\frac{16}{3}\pi r^3} = \frac{4.6 \times 10^6}{0.67\mu m^3} = 6.9 \times 10^{12} bp/\, m3$$

This equation, of course, outputs the exact radiuses with the former equation for the same N_{bp}. Therefore, using any radius calculation equations provides the necessary radius for the cell wall boundaries. However, the latter is less complicated. Thus, we utilized the latter in our programs.

Since our primary experimental model was a 12000 bead model of the bacterial chromosome (i.e., $120\ Kbp$), our system radius is $r = 100nm$ for most of our simulations.

2.2.3 Cylinder

After figuring out the measurement to work with, what remains was to create the beads forming the membrane. As discussed previously, the middle part of the membrane was formed via a cylinder. A cylinder is essentially multiple circles (or rings) layered on top of each other. Utilizing this approach, we constructed the circles using,

$$y = \cos(\theta) \times r$$

$$z = \sin(\theta) \times r$$

while $-\pi < \theta \leq \pi$ with $2\pi r$ times increment to ensure the distance between the neighboring beads is a σ which corresponds to $3.4nm$ in our coarse-grained model. Next, we created those circles, so they layer on each other in the x direction consistent with the required length of the cylinder, which is $4r$—using $-2r < z \leq 2r$ with $4r$ times increment to again ensure one σ distance between the circles, we completed the cylinder.

2.2.4 Semi-Spheres or Caps

A cylinder alone does not complete our pursuit to construct a rod-shaped membrane model. The finishing touch requires the joining of two semi-spheres to both ends of the cylinder. Similar to a cylinder, a semi-sphere is formed via circles layering on top of each other. However, unlike a cylinder, creating a semi-sphere is quite a challenge. Firstly, each ring should have a smaller radius than the previous one. Additionally, decrementation of radius cannot be linear since it would construct a cone. Instead, the radius should change according to the function, $r_c = r \times \cos(\theta)$, where $0 < \theta \leq \pi/2$.

The circles are placed through the $-2r > z \geq -3r$ at one side and $2r < z \leq 3r$ at the other side. Decrement or increment in z direction was done r times to provide around a σ distance between the rings. The resulting membrane model sufficiently resembled the cellular confinement and membrane architecture of *E. coli*. Functionally, it could maintain its content (i.e., the chromosome and the proteins) with only less than 1% leakage of proteins in $8 \times 10^5 \tau$, which is the equivalent of

$20ms$. Although $20ms$ may seem like a short duration, it was more than enough to obtain protein dissociation rates.

2.3 Design and Positioning of the Transcription Factors

Previously, studies showed that the concentration-dependent dissociation (i.e., facilitated dissociation) heavily relied on the dimeric transcription factors' multivalent binding to the DNA[18,43,46]. The suggested mechanism is that partial dissociation could happen due to the dissociation of only one monomer while the other one stays bound. This partial dissociation allows an intermediate state that lowers the energy barrier, thus facilitating the dissociation process. Therefore, to simulate facilitated or concentration-dependent dissociation, transcription factors (TFs) and Nucleoid Associated Proteins (NAPs) must be multimeric[43].

2.3.1 Initial Dimer Model of Transcription Factors

Our model NAP, the Factor for inversion stimulation (Fis), is a dimeric protein. Thus, our coarse-grained model for Fis must be dimeric. Initially, we decided on using a two-bead system in which both beads are binding sites and covalently bound to each other. Although this model was sufficient to model multivalency and FD, it had some limitations.

Firstly, the covalent bond between the binding domains restricts their flexibility; however, dimeric TFs are relatively flexible[71], which helps with their interaction with the DNA. Furthermore, TFs are not composed solely of binding domains since they also contain non-binding domain(s)[72]. Therefore, the two-bead model would not

allow the modeling of steric interactions. Although the preliminary results obtained using this model combined with non-collapsed DNA polymer indicated that this model is helpful for modeling FD, we abandoned it due to its restrictions.

2.3.2 The Cherry Model

Even though our primary objective was to model facilitated dissociation in confinement, we realized we could also characterize steric interactions along the way. However, the two-bead model is not suitable for such characterization. To do so, we needed to utilize a better coarse-grained model, which should also address the problems of the previous model.

Through visual inspection of the crystal structure of Fis[39] and utilizing the coarse-graining function of the VMD (Visual Molecular Dynamics), we have come up with the cherry model. This cherry model of Fis is a three-bead design where the two binding beads are joined by a third that acts as a hinge. The angle formed by those three beads is $\theta = 50°$. Therefore, the model resembles a "V" shape or a cherry (Figure 12).

In addition to its consistency with the crystal structure of Fis, the cherry design includes a non-binding region providing flexibility for binding domains, thus resolving both of our previous concerns at once. Therefore, the cherry model is a much better candidate to characterize DNA-Protein binding and dissociation kinetics. Consequently, we used the cherry design of Fis in all of our simulations.

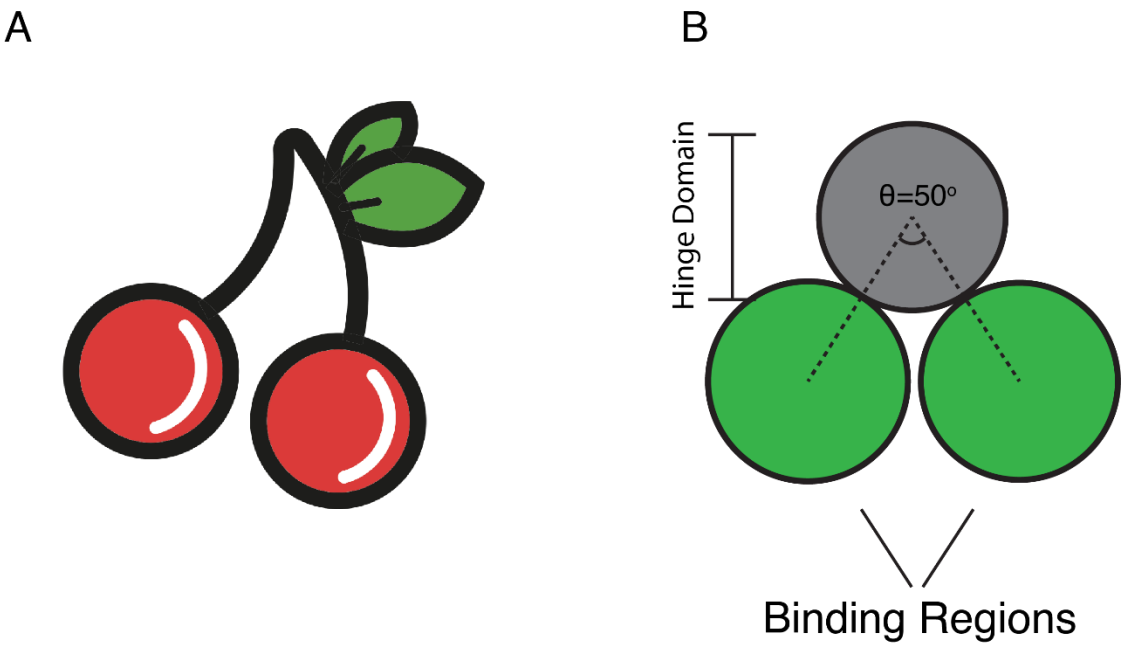

Figure 12 The coarse-grained model of Fis protein. (A) Inspiration source of the name. (B) The parts of the Cherry model of Fis with $\theta = 50°$ angle.

2.3.3 Positioning the Bound Transcription Factors

The primary purpose of our research is to characterize the change in dissociation rates of DNA-binding proteins under various conditions. However, for a protein to dissociate from the DNA, it should be attached to it, to begin with; thus, we first need to place some TFs to the proximity of dimeric binding sites. This approach requires information on the coordinates of 120 binding sites out of the $N = 12000$ polymer.

Previously, when placing the dimeric binding sites to the DNA polymer, we employed a strategy to position them every once in a 100^{th} bead starting from 0 (i.e., $0, 100, 200, \dots, 12000$). Therefore, we only needed the positions of the monomers, which we obtained via polymer expansion. After finding the locations of the dimeric binding sites, we placed one TF per binding site (120 in total) at a 0.6σ or $\sim 2nm$ distance from their respective binding sites (Figure 13). Henceforward, we will address that 120 TFs bound TFs.

2.3.4 Positioning the Free Transcription Factors

After bound TFs, we also needed to place some free TFs in the system. Those free TFs act as competitors for DNA binding sites necessary to observe facilitated or concentration-dependent dissociation of bound TFs. We utilized the concentration calculations mentioned in the following section to determine the required number of free TFs to satisfy the desired concentration.

Next, we needed to decide the space where the free TFs would be located. Due to the complex architecture of our simulation space dictated by the membrane (see membrane subsection), drafting the whole volume as the possible coordinates for

the free TFs would be rather complicated. Instead, we employed a much simpler approach where we utilized the largest cuboid possible within the membrane boundaries.

The most voluminous cuboid fitting inside the membrane is also the largest to fit within the cylinder compartment of the system. Considering the cylinder radius $r_c = r$ and the length $L_c = 4r$, the height H of the cuboid therefore is,

$$H_{cuboid} = 2l$$

$$r^2 = l^2 + l^2$$

$$l = r/\sqrt{2}$$

and since the length does not change $L_{cuboid} = L_c$. Thus, the volume of the cuboid,

$$V_{cuboid} = H \times L^2$$

$$= 4r \times \left(\frac{2r}{\sqrt{2}}\right)^2$$

$$= 8r^3$$

While this volume does not cover all the space present in the system, it provides a good enough distribution of the free TFs for the following equilibration step.

After calculating the possible interval of the coordinates, we placed the free TFs at random positions within those x, y, z intervals. Additionally, we did not try to evade overlap with the DNA polymer in this process since its actual volume is $< 1\%$ of the system.

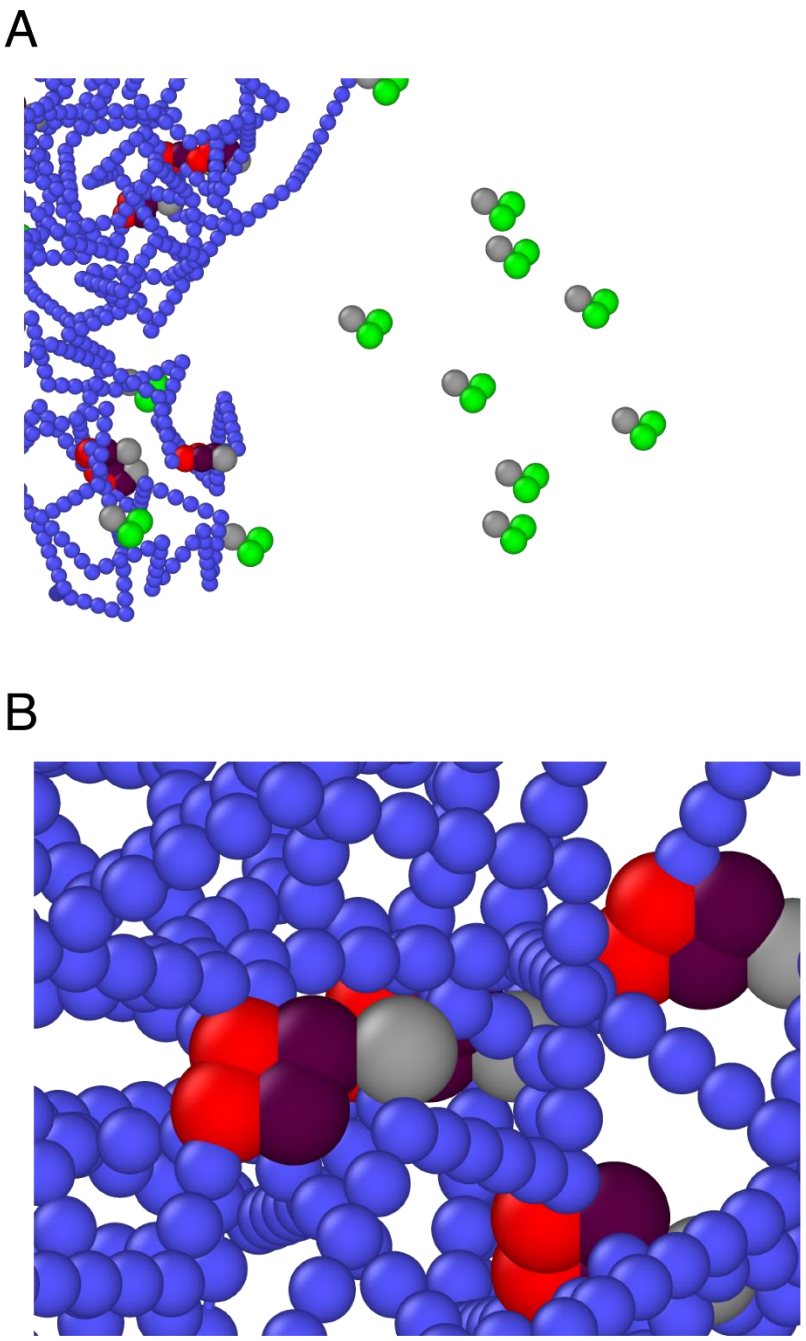

Figure 13 The localization of the proteins. (A) Free proteins (green) were located at random positions. (B) Proteins to act as initially bound proteins (purple) were positioned near DNA binding sites (red).

2.3.5 Concentration Calculations for the Free Transcription Factors

One of the variables in our simulations is the concentration of the free TFs. Consequently, we needed the capability to change the free TF concentration on command. Thus, we implanted a system to allow the user to input the desired concentration in micromolars. Then, the concentration is converted to a number via the following calculations.

$$M\ (Molarity) = \frac{mol}{L}$$

where,

$$mol = \frac{n}{6.022 \times 10^{23}}$$

$$L = 10^{-3} m^3$$

$$V_{system} = \frac{16}{3}\pi r^3$$

Therefore, the equation to obtain the number of free TFs,

$$n_{freeTF} = M \times (\frac{16}{3}\pi r^3) \times 10^3 \times (6.022 \times 10^{23})$$

Here, the only two variables are the radius r and the molarity M. Considering our $N = 12000$ bead system has the radius of $r = 102nm$, $10\mu M$ corresponds to 105 proteins.

The peak level of the Fis protein in *E. coli* reaches 60,000 copies per cell, an equivalent of the $\sim 150\mu M$ cytoplasmic concentration. Thus, we ran our simulations with

different concentrations of free TFs ranging from 0 to $200\mu M$ to analyze the effect of concentration on dissociation rates.

2.4 Analyzing the Simulation File for Dissociation Rates

2.4.1 Conditions and Parameters for Dissociation

Following the runs of our simulations, we needed to analyze the resulting simulation files to find dissociation rates for each simulation. Therefore, we first needed to quantify the TFs remained attached to their relative dimeric binding sites. Accordingly, TFs should not only be attached, but they must not have been dissociated at all to be considered bound. Hence, once a TF dissociates, it will never be counted as bound even if it rebinds to its original binding site.

After deciding how to separate bound TFs from unbound TFs, we needed to choose a threshold distance above which the TFs would be considered dissociated. Thus, we first needed to select the points from which the length between a TF and its binding site should be calculated. Accordingly, we decided to use the center of masses of both the dimeric DNA binding site and TF binding domains as the reference points. This process excludes the hinge domain of the TF since it has no direct role in binding.

Next, we decided on a threshold distance of $r = 2.5\sigma$, or $8.5nm$ to consider a protein fully dissociated. This length could seem more than necessary; however, considering the half-dissociated states of TFs, $r = 2.5\sigma$ is near-optimal as the threshold distance.

2.4.2 Analysis of Transcription Factor Dissociation

After concluding the parameters and the conditions for TF dissociation, we needed to check how many proteins dissociate at each simulation step. Accordingly, we had to utilize the dump file from the LAMMPS MD simulation. This file type contains all the positions of all the beads at each simulation step.

One approach to detect unbound TFs would be the visual inspection of dump files using visualization tools such as OVITO and VMD. These programs allow the user to view the simulations as movies. However, considering that this process requires a human to follow 120 TFs for 1250 simulation steps, this method would not be feasible or accurate even for one simulation. Since we ran hundreds of simulations for various scenarios, this approach would be impossible to employ.

Of course, the more straightforward approach here is utilizing computational tools. A dump file contains all the information necessary for distance calculation. Therefore, our program first reads the dump file. Then, it separates each timestep into 2D arrays where one dimension holds the bead IDs and the second contains their respective positions.

Next, our program finds the positions of the dimeric binding sites and their respective TFs. Then, it calculates the distance between them using the following calculations.

$$X_{dimer} = \frac{X_{monomer1} + X_{monomer2}}{2}$$

$$X_{tf} = \frac{X_{binding\ domain1} + X_{binding\ domain2}}{2}$$

$$d(tf, dimer) = \sqrt{(x_{tf} - x_{dimer})^2 + (y_{tf} - y_{dimer})^2 + (z_{tf} - z_{dimer})^2}$$

If a TF is 2.5σ away from its respective binding site, our code marks it as detached. This process is repeated for every TF for all the timesteps. Then, the bound TF number is calculated for each time step, and this count is stored in a .csv file with their respective timesteps. In the following sections, we will discuss how we utilize those files to find out residence times or dissociation rates of the TFs for each case.

2.4.3 Survival Fraction of the Bound Transcription Factors

Survival fraction is the ratio of the survivors to initial components for a chosen time point. It can be simply represented as $n(t)/n_0$. In our work, the components are the initially bound transcription factors. The previous subsection described how we obtained the $n(t)$, the number of TFs that remained attached to their binding sites for each timestep. Since all of our kinetics simulations have $n_0 = 120$ initially bound TFs, the survival fraction in any given timestep is $n(t)/120$.

Next, we make the graphs of survival fractions with respect to time (t). For each case where the concentration or affinities changes, we ran three replicates to ensure the reliability of our results. Thus, in the survival fraction graphs, dots represent the average $n(t)/n_0$ of three simulations, and the error bars are shown in lighter colors than their data points.

Our primary goal in using the survival fraction graphs is to extract the residence times of TFs for each case. Survival fraction graphs of TF dissociation have the decay

characteristics of the exponential decay where the graph decays with a constant exponential rate. A single exponential decay is represented as the following equation:

$$n(t) = n_0 e^{-\lambda t}$$

Here, the lambda (λ) is the exponential decay constant, and e is Euler's number. Since we deal with dissociation rates, we will employ k instead of lambda (λ) due to the convention. This preference slightly changes the above equation to the following.

$$n(t) = n_0 e^{-kt}$$

2.4.4 Curve Fitting the Survival Fraction Graphs

Even though the raw survival fractions graphs exponentially decay, the transition between time points is not smooth. Therefore, we needed to fit a single exponential graph for each survival fraction data. Initially, we used the Python *SciPy* library *optimize* module for the fitting. However, fits were somewhat not convenient. After realizing that it could negatively affect our results, we abandoned it. Fortunately, the graph illustration program we utilize, *GraphPad Prism*, has a curve fitting tool that works very well. We used it to obtain curves and the relative data such as residence times and dissociation rates for each case.

2.4.5 Half-Life and Mean Average Lifetime

A half-life describes the time it takes for a decaying system to reach $n_0/2$ (Fig 14). Half-life is vital for describing behaviors of many decaying systems, including carbon dating[73]. However, the half-life does not provide information on the average duration of the components remaining bound.

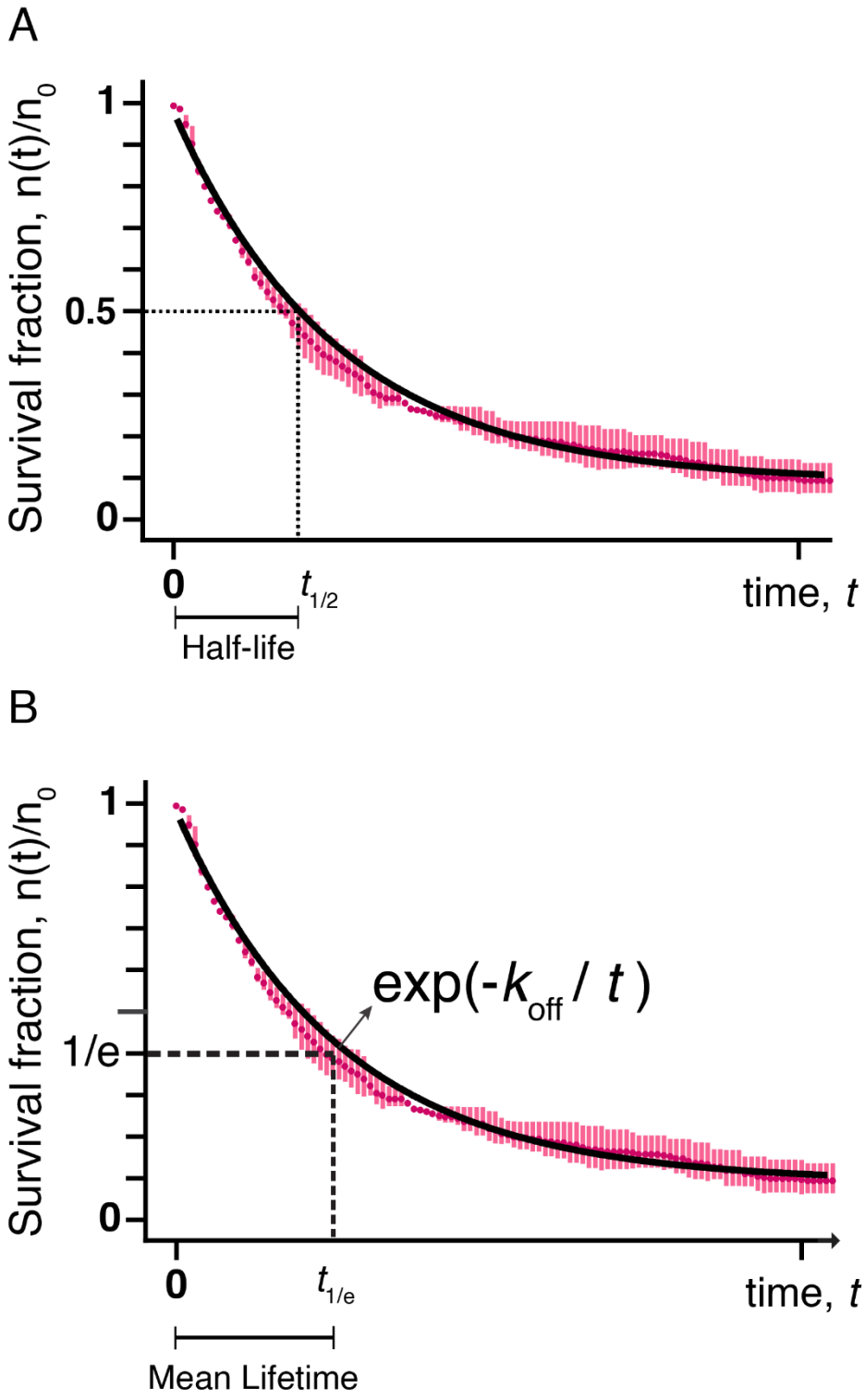

Figure 14 The distinction between (A) half-life and (B) mean lifetime concepts.

The average period of an element remaining bound is described by the mean lifetime. Analogous to the half-life notion, mean lifetime is the time point where the decaying system reaches n_0/e (Figure 14). Therefore, the residence time of a TF on a dimeric binding site is its mean lifetime. Here, the time (t) is calculated as the following.

$$n(t) = n_0 e^{-kt}$$

$$n(t) = n_0/e$$

Hence,

$$n_0/e = n_0 e^{-kt}$$

$$e^{-kt+1} = 1$$

Thus,

$1 - kt = 0$

$$t = 1/k$$

Therefore, the residence time for a transcription factor is the reverse of its exponential decay constant or dissociation constant. Obviously, fitting an exponential decay curve to a graph requires estimating the dissociation constant (k). Consequently, calculating the residence times of proteins is a straightforward process.

2.4.6 Calculating the Dissociation Rates

Dissociation rate defines how fast detachment of the bound molecule takes place. It is described by the number of dissociating molecules per unit time. For our case, it is the number of TF dissociating per second. Since the residence time and the dissociation rates are complete opposites, the dissociation rate could be represented as $1/t$. As stated previously, $t = 1/k$ and thus, dissociation rate is equal to decay constant (k) as expected.

2.4.7 Facilitated Dissociation in Mutated Transcription Factors

Genetic mutations are the driving forces of evolution. All living cells and viruses depend on mutations to adapt to conditions around them. However, genetic mutations are generally random; thus, they can bring diverse modifications to the genes[74].

One such alteration could be the change in TF binding affinity to the specific DNA sites. Those types of mutation can either decrease or increase the affinity of a TF towards its target site. We wanted to test the difference in residence times in case of such a mutation while the wild-type (WT) TFs are still in the cell.

Therefore, we changed the binding affinity of the free TFs by one kT, lower or higher, whilst keeping the affinities of the bound TFs remain the same. In this part, nonspecific binding strength for all the TFs is $u_{ns} = 2\ kT$. Also, we used specific binding affinity $u_{sp} = 9\ kT$ for all bound TFs. For the test groups, free TFs have either the specific attraction potential $u_{sp} = 8kT$ or $u_{sp} = 10kT$. Simulations with free TF

affinity $u_{sp} = 9\ kT$ are used as negative controls. Cases are run at four different concentrations, $20,60,120$, and $200\mu m$ with three replicates to ensure data quality.

2.4.8 Bound Fraction of the Transcription Factors

Transcription factors and nucleoid-associated factors spend a significant portion of their lifetimes bound to the DNA. This portion varies between proteins depending on their affinity to their binding sites and the rest of the DNA[4,5,45]. However, we also wanted to check the effect of multivalency on this phenomenon. Therefore, using the distance calculation approach mentioned in the dissociation section, we calculate whether a TF is free or bound to the DNA.

For TF dissociation from their specific binding sites, we used 2.5σ or $8.5nm$ as the threshold distance for complete dissociation. However, on bound fraction calculations, we need only the certainly bound TFs. Thus, the threshold distance should be lower. After experimenting with different lengths, we decided on using 1.1σ or $3.74nm$ as the threshold distance. Therefore, transcription factors with their binding sites closer than 1.1σ to any bead of the chromosome are considered bound. Both for monomeric and dimeric cases, we employed 0 and $200\mu m$ of TF concentration with the affinities $u_{sp} = 9kT$ and $u_{ns} = 2kT$. Then, we ran the simulation as described in the *Simulation Model* section and inspected only the final stage of the simulations to analyze the bound fractions of TFs.

2.4.9 Residence Times of Transcription Factors on Chromosome

In our initial work set, we inspected the kinetics of the TFs dissociating from their binding sites. There, we analyzed how fast a TF detaches from a region that it is bound initially. Even though this approach is essential for dissociation kinetics, it neglects the following binding and unbinding kinetics, thus providing a limited perspective.

A protein can rebind after its initial dissociation and dissociate following its rebinding. It is fundamental to comprehend how long a TF or a NAP stay bound and unbound to the DNA to grasp the binding/dissociation kinetics of TFs fully. In turn, the comprehension of residence times on DNA could provide a broader understanding of transcriptional regulation by TFs or NAPs.

To figure out residence times of TFs on the chromosome, we first need to equilibrate the system to minimize the effects of the initial fluctuations. We already apply an equilibration step to all our simulations to relax the system. Therefore, rather than utilizing another equilibration step, we simply ignore the first $4ms$ of the simulation. Since the total duration for each simulation is 20ms, the remaining $16ms$ provides sufficient time for analysis of residence times.

Our program first reads the exported simulation file (i.e., dump file) similar to the dissociation section. Then, it calculates the center of mass of binding domains for each TF. Next, it computes the distance between each TF binding site with each DNA monomer. When a TF is near a DNA monomer within a distance of $r = 1.2$, it is marked as bound and given a value of 1 or otherwise 0 as a reference to Boolean (i.e., true or false) data type.

This process is repeated for all the proteins and for every timestep. This information is stored in two 2D arrays. One of the arrays holds the information for specific binding sites of DNA, while the other holds the data for the rest of the DNA polymer. According to whether they are bound to dimeric binding sites of DNA or the rest of the chromosome, they are given 1 or 0 in their respective arrays for that particular timestep. Subsequently, we can distinguish the residence times of the TFs on nonspecific DNA and the dimeric binding sites.

Then, another program analyzes how long TFs remain bound or free uninterrupted utilizing the abovementioned arrays. Next, all the durations for every protein are pooled together. Finally, this program computes the abundance of the residence times in each time interval (e.g., there are n instances where the TF remained bound to a binding site for 15-20 timestep).

2.4.10 Facilitated Dissociation Efficiency

As previously described, facilitated dissociation is defined as the increase in dissociation rates in a concentration-dependent manner. Spontaneous dissociation, on the other hand, is the dissociation without any competitors[46]. In other words, it is the dissociation of the molecules from their binding sites without any external factor. In our work, SD occurs when there is no free TF and FD reaches its maximum effect at our highest experimental concentration $200\mu m$.

Therefore, we defined FD efficiency as the ratio of FD to SD. In any case, SD exerts its effect; thus, FD is the difference between dissociation rates of the $200\mu m$ and $0\mu m$ cases. Therefore, FD efficiency is calculated as the following.

$$SD = k_{0\mu m}$$

$$FD = k_{200\mu m} - k_{0\mu m}$$

$$\eta_{FD} = \frac{k_{200\mu m} - k_{0\mu m}}{k_{0\mu m}}$$

2.5 Nucleoid Architecture and Chromosome Compaction

Chromosomes can condense and relax throughout their lifespans[48,51]. This change in chromosome formation depends on the histone proteins in eukaryotes[26–28] and NAPs in prokaryotes[4,11,45]. Chromosome condensation is a regulatory factor for transcription[51,62]. Thus, understanding chromosomal architecture is crucial for untangling the mystery of transcriptional regulation.

2.5.1 Chromosome Condensation and Radius of Gyration

In this study, we deal with various NAP concentrations and affinities that are likely to result in different levels of chromosomal condensations. Therefore, we need to analyze the change in chromosome density of diverse cases in our simulations.

The *E. coli* nucleoid is commonly sustained at 40-60% volume of the cytoplasm[44]. However, the chromosome in our main system ($N = 12000$) occupies around 30% of the cytoplasmic volume after the equilibration process. Although this ratio is not consistent with *E. coli* nucleoid, it turned out to be advantageous to detect relaxations and condensations in the simulations.

In its simplest form, the radius of gyration (R_g) is the root mean square of distances between the monomers of a polymer and the polymer's rotation axis when the monomers have the same mass. Therefore, R_g is defined as the following

$$R_g^2 = (r_1^2 + r_2^2 + \cdots + r_n^2)/N$$

where N stands for the number of monomers.

Here, we use R_g to define chromosome condensation level. After the initial equilibration process, the nucleoid with $N = 12000$ beads has an R_g of 32σ or $\sim 100nm$. That leaves space to relax or condense, considering the upper and lower limits for the chromosome are 52σ and 22σ respectively.

For the R_g calculation, we used the LAMMPS MD package's *compute* command for the radius of gyration. Therefore, for each step, the LAMMPS exports the R_g of the DNA polymer. For simplicity, we only calculated the radius of gyration for nonspecific sites of the DNA as the binding sites are only 2% of the whole polymer; thus, they would have no significant effect on the calculations.

2.5.2 Fixed Chromosome

At high protein concentrations and nonspecific protein-DNA affinity levels, the chromosome collapses onto itself. The resulting configuration either does not allow dissociation at all or leads to a DNA-segmental FD[49], where the DNA segments compete for TFs, thus removing them from their original binding sites. Although it is the reverse of TF-based FD, DNA-segmental FD is very effective when the nucleoid is compact.

Therefore, we restrain or fix the DNA polymer at several relaxation levels to decouple nucleoid condensation from TF dissociation. To achieve those levels, we ran the equilibration simulations with repulsive interactions with increasing durations. Then, the resulting chromosome structure was fixed, thus restraining its movements in our main simulations.

2.5.3 Snapshots of the Simulations

Following the realization that the chromosome architecture differs under various simulation conditions, we wanted to have a closer look at this phenomenon. The previous section described how we quantified the chromosomal condensation via the radius of gyration. However, Rg does not provide information on local structures within the nucleoid.

A broader comprehension of the chromosome architecture requires visual inspection of the nucleoid. Therefore, we utilized the simulation visualization tools to obtain snapshots of the systems (Figure 15). These tools are namely VMD [75](Visual Molecular Dynamics) and OVITO[76] (*Open Visualization Tool*).

For a clear vision of the nucleoid, the membrane is made semi or fully transparent. Instead of the full membrane, we used a 2D membrane as a frame to help the viewer to comprehend chromosomal density. The frame is simply constructed via deleting all the atoms except for the ones with $-1.5 < y < 1.5$ range.

After layering the frame on top of nucleoids, we obtained the snapshots with an aspect ratio of $16{:}\,9$ and resolution of $1920 \times 1080p$. Then, the resulting images

were exported as PNG (Portable Network Graphics) files with transparent backgrounds via OVITO.

On the other hand, we utilized VMD for obtaining snapshots of the local nucleoid structures such as protein clusters.

2.5.4 Multiprotein Complexes and Clusters

We focused on chromosomal architecture and its dependency on several factors along with the transcription factor (TF) dissociation kinetics from regulatory DNA sequences on a prokaryotic supercoiled chromosome. As discussed in previous sections, these factors included TF (or Nucleoid Associated Protein (NAP) in our case) concentration, the affinity between TFs and specific binding sites, and nonspecific DNA sequences or namely specific and nonspecific interactions.

After realizing the importance of chromosomal condensation on dissociation rates (i.e., the reverse of residence times), we decided to visually investigate our simulations to construct a comprehensive and reasonable explanation for this phenomenon. We realized that the chromosome, a self-repelling molecule, along with the proteins, formed highly collapsed structures, particularly above a nonspecific interaction strength of $5kT$ and high protein concentration of $\sim 60\mu m$. As discussed in detail in its own section, global chromosome condensation affects the dissociation by forming heavily compact structures where TFs are trapped within, and thus, cannot dissociate from their binding sites. However, global condensation is not the only parameter that impacts dissociation via changing DNA architecture. Local multi-protein complexes are generally formed via bridging of multiple specific binding

sites and the binding of several TFs to each binding site. The resulting clusters vary in shape, the number of proteins, and quantity with each simulation of different concentrations or affinity.

The physiological significance of these protein clusters is unclear. However, in our simulations, such multi-protein complexes formed local condensed regions along the chromosome. As we know from our previous results, chromosomal condensation significantly impacts dissociation rates; therefore, it may also affect transcriptional regulation. Similar to global condensation, we hypothesize that the local protein-DNA complexes entrap TFs with a similar mechanism to the global chromosome condensation, affecting dissociation rates and, consequently, transcription rates.

Due to their hypothesized significance, we decided to quantify the multi-protein complexes. In the quantification process, we wanted to count both the number of clusters and TFs in each cluster. There are several clustering algorithms available for clustering problems. For instance, a widespread clustering algorithm is k-NN (k-Nearest Neighbor), widely used in machine learning[77]. Although this algorithm applies to a wide range of practices, it requires the cluster number beforehand from the user. In other words, the user needs to know the number of clusters before running the algorithm. In our case, the quantities of the protein clusters vary drastically with different conditions, even among the replicas of the same conditions. Therefore, it would not be practical to use such an algorithm. Another option was to use a density-based algorithm[78]. However, those types of algorithms are not compatible with varying cluster formations.

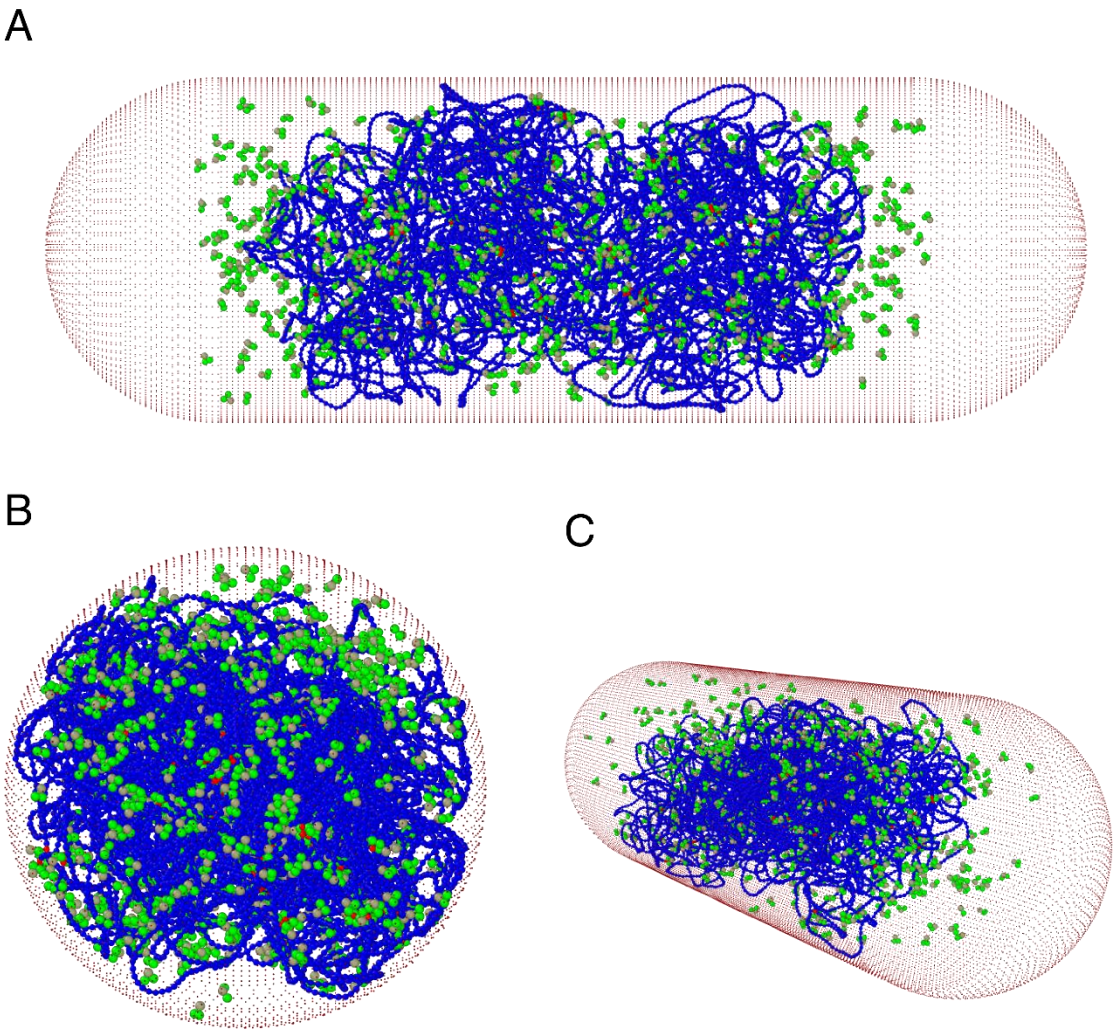

Figure 15 Representative snapshots of the system obtained via OVITO. (A, B) Side and (C) perspective views of the system.

Here, the most suitable approach was to use Stillinger's definition of the cluster. In his 1962 paper, Stillinger defines a cluster as units in the range of each other for a specified radius[79]. Therefore, clusters can grow in any direction, thus forming various shapes and sizes.

Investigating how many clusters there are and which protein belongs to which cluster requires a multistep strategy. First, we listed all the neighbors for each protein. We immediately eliminated the TFs with no neighboring TF to reduce complexity, and thus, the computation runtime. Then, each neighbor list is compared with the rest. The lists with common proteins are pooled together, establishing the clusters.

The combined lists represent the clusters. Therefore, the number of elements in a combined list is the equivalent of cluster size. Also, the number of the lists gives the cluster count in a system. We decided on the protein count threshold as 5. Consequently, any protein complex with less than five proteins is not considered a cluster.

Conventionally, the threshold distance for neighboring is 1.5 times the bead diameter. However, we found it more suitable to use 2.1 times the size of the beads (or 2.1σ) following the visual inspection and quantification of the clusters using a data file.

Most prominently, we used the $U_{ns} = 12kT, U_{ns} = 5$ at $200\mu m$ in which the chromosome and almost all the proteins form one giant cluster. Thus, we expected to have one cluster as our results. Initially starting with 1.6σ, we gradually increased the threshold radius by 0.1σ. Eventually, at 2.1σ (i.e., $\sim 6.7nm$), we encountered

only one cluster as desired. Expanding the threshold radius could have joined proteins to the clusters, which are essentially not part of the multi-protein complex. Also, decreasing the radius 2.1σ even by 0.05σ excluded the beads that belong to the complex.

2.6 Simulation Model

We employed Kremer-Grest (KG) coarse-grained polymer model[80] for our models of DNA and proteins in implicit solvent conditions. In our primary system, DNA polymer consists of 12000 KG beads corresponding to 1.2×10^5 bps. The DNA polymer, TFs, and the membrane are formed by identical beads with the size of $r = 1\sigma$ $(3.4nm)$ and mass of $m = 1$ in LJ (Lennard Jones) units.

Truncated and shifted LJ potential is employed to model steric interactions among the beads.

$$V^{LJ}(r) = \begin{cases} 4u[(\sigma/r)^{12} - (\sigma/r)^6 + v_s] & r \le r_c \\ 0 & r > r_c \end{cases}$$

Unless otherwise noted, we set interaction strength to $u = 1kT$, cut-off distance to $r_C = 2^{1/6}\sigma$, and shift factor to $v_s = 1/4$ for repulsive interaction conditions where k is the Boltzman constant, and T is the absolute temperature.

On the other hand, we set interactions between DNA and TF to the interaction strengths of $u \gg 1$ for attractive interactions. Here, the cut-off distance is $r_c = 2.5\sigma$, and the shift factor is $v_s = 0$. We utilized those parameters only to model specific and nonspecific interactions between TF and DNA. Therefore, all other interactions are repulsive.

We ran all the simulations using the LAMMPS MD package. We utilized reduced temperature $T_r = 1.0$ and constant volume V. Also, we kept the temperature constant via a thermostat coefficient of $\gamma = 0.5\tau^{-1}$ with Langevin thermostat style. Additionally, we use a viscosity prefactor of 0.5 to replicate cytoplasmic conditions.

In our primary model, we set simulation box boundaries and volume $L_x \times L_y \times L_z = 184 \times 64 \times 64\ \sigma^3$ leaving a 4σ distance from the end of the cellular confinement. Periodic boundary conditions were applied in all directions.

2.6.1 Bonds

The following equation of nonlinear FENE potential[80] governs the bonds between the monomers of the proteins as well as the DNA.

$$V^{Bond}(r) = -0.5kr_0^2 \ln\left[1 - (r/r_0)^2\right]$$

where the bond stiffness is $k = 30kT/\sigma^2$ and the bond distance is $r = |\boldsymbol{r}|$ or $r = \sigma$ and the maximum allowed bond distance is $r_0 = 1.5\sigma$ among the DNA monomers. In the meantime, bond distance is $r = 1.3\sigma$, and the maximum distance is $r_0 = 2\sigma$ between TF hinge and binding domains.

We ran three replicates for each simulation case (i.e., changing concentrations and affinities). The average of the replicates is presented as the data points, and the error bars on graphs represent standard deviation.

We did not utilize replicates for $0\mu m$ free TF cases since the only difference between the replicates are initial positions of the free TFs, and without any free TF, all replicates are essentially the same

2.6.2 Angles and Bending

Angles and the bending potential are regulated by the following equation of harmonic bending potential to model the semi-flexible nature of biopolymers.

$$V^{bend}(\theta) = k_\theta(\theta - \theta_0)^2$$

For transcription factors, we set the angle to $\theta_0 = 50°$ and the angle stiffness to $k = 12kT/rad^2$ to provide and preserve the cherry shape. On the other hand, we set the angle to $\theta_0 = \pi$ and the angle stiffness to $k = 15kT/rad^2$ to achieve the persistent length of 15 monomers. Here, 15 beads translate to $150bps$ or $50nm$ which is consistent with the persistent length of dsDNA[25].

2.6.3 Conversion of Simulation Time to Metric Units

As previously described, we ran our simulations with the time step of $\Delta t = 0.005\tau$. Here, τ is the unit time scale for the simulations in LJ units. However, the unit τ does not agree with our understanding of the real world. Therefore, we need to convert simulation time to the metric unit of time, seconds.

Using the Einstein diffusion equation[81],

$$\tau = \frac{d^2}{6D}$$

where our sample molecule has the size of $d = 6.8\,nm$ and $D = 0.5kT/2\pi\eta d$. The viscosity of water $\eta = 10^{-3}N \times s/m^2$ and the $kT = 4 \times 10^{-21}N \times m$. Thus,

$$\tau = \frac{\pi\eta d^3}{kT}$$

Inputting variables results in $\tau \cong 250ns$. Therefore, our 2×10^7 simulation steps correspond to $T_{max} = 25ms$. However, we mainly analyzed the first $20ms$ as it is sufficient for dissociation analysis.

3 RESULTS AND DISCUSSION

3.1 FD Can Occur in the Bacterial Confinement

Our primary goal was to reveal whether a facilitated dissociation could take place in bacterial confinement. Following the construction of our initial $N = 2400$ system (i.e., small system), we ran hundreds of simulations, including three replicates for each case (i.e., different concentration or binding strength). In this small model, we utilized three different DNA-protein attraction strengths, which are 6, 6.5, and $7kT$ with 1.6σ nonspecific DNA-protein affinity for each case. For each affinity level, we employed 10 different concentrations within the interval of $0 - 200\mu M$. At $u_{sp} = 7kT$, we observed a clear pattern of concentration-mediated dissociation rates through FD (Figure 16C). However, at lower DNA-protein interaction levels, FD behavior is ambiguous, at least for the low concentrations (i.e., $< 20\mu M$)(Figure16A-B). This contradiction between the FD behavior not only demonstrates that FD is possible in bacterial confinement but also suggests FD requires somewhat high binding strength.

Nevertheless, the small system restrained the flexibility of non-binding regions of the DNA polymer due to its relatively small spacers (i.e., 20 bead or 200bps). This lack of flexibility is inconsistent with the *E. coli* chromosome, thus putting the reliability of our results to question. Therefore, to provide more reliable findings, we scaled up

our DNA polymer from $N = 2400$ to $N = 12000$ beads. Indeed, the rest of the system, including the system volume, was also scaled up to maintain the consistency of ratios. Then, we also changed the interaction levels to $u_{sp} = 20kT$ and $u_{ns} = 13kT$ to be consistent with the Fis-DNA binding energies. However, these high binding energies limited dissociation rates below such a level that residence times far exceeded simulation durations rendering off-rate analysis impossible. Therefore, we scaled down the affinities to $u_{sp} = 20kT; u_{ns} = 13kT$, $u_{sp} = 16kT; u_{ns} = 9kT$, and $u_{sp} = 9kT; u_{ns} = 2kT$ maintaining the $7kT$ difference between specific and nonspecific interaction strengths.

Except for the $u_{sp} = 9kT, u_{ns} = 2kT$ system, time limitation was still an issue even though they demonstrated an FD pattern. Thus, we continued with the $u_{sp} = 9kT, u_{ns} = 2kT$ system employing 6 different NAP concentrations. Investigation of the dissociation rates demonstrated an apparent concentration-mediated dissociation behavior. In Figure 17, the upper-most curve represents the spontaneous dissociation (SD) in which there is no competitor (i.e., no free DNA binding protein); thus, the dissociation solely depends on SD. In the lower datasets and curves, there are competitor proteins. With the increasing concentrations, the curves become steeper, suggesting that the dissociation rates increase in a concentration-dependent manner. At peak concentrations (i.e., $\geq 40\ \mu M$), the dissociations rates are a hundred folds higher than that of SD alone (no competing protein case). That suggests FD could, in fact, decrease residence times from hours to minutes in the cellular confinement consistent with the previous findings.

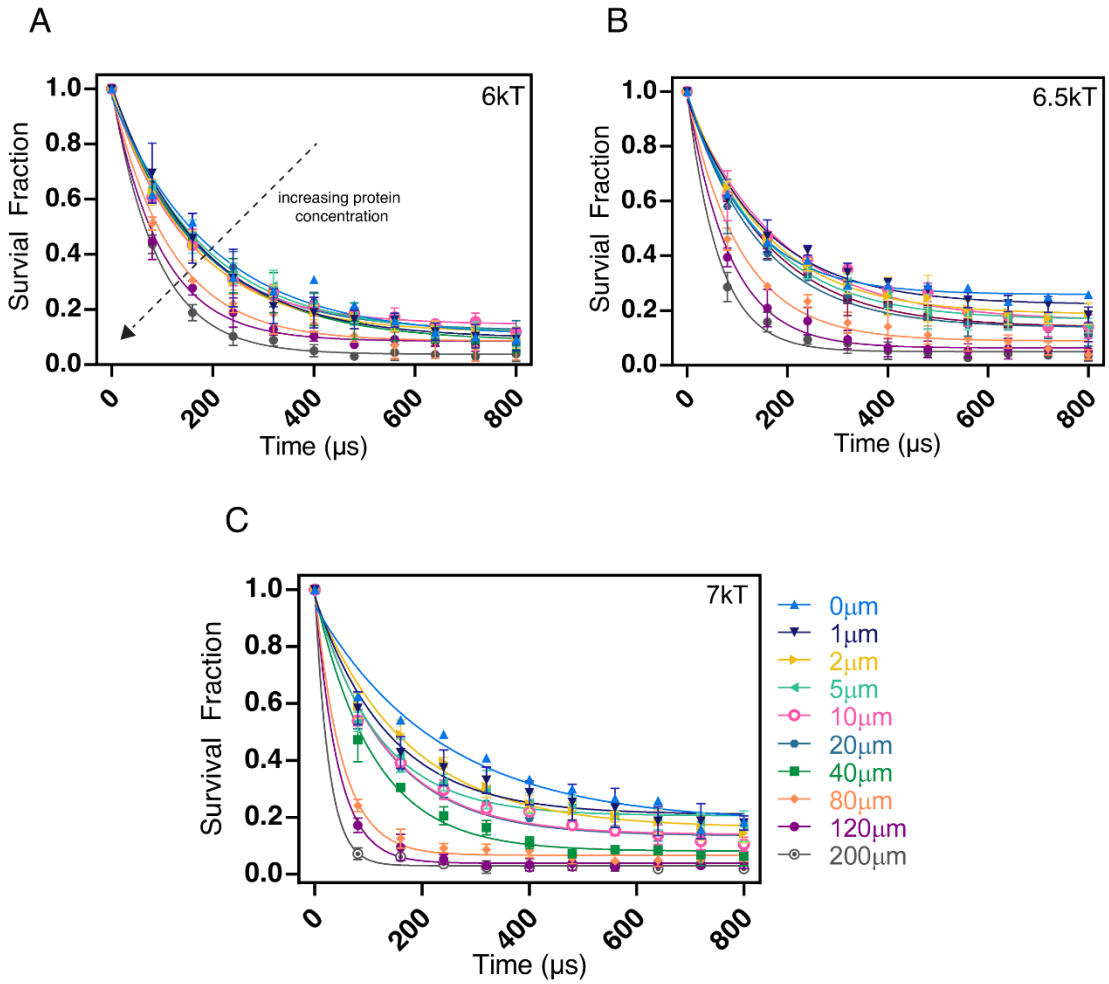

Figure 16 Survival fractions in the small (N = 2400) system with the given concentrations at (A) $6kT$ (B) $6.5kT$ (C) $7kT$. For all cases nonspecific interaction potential $u_{ns} = 1.6kT$. Error bars indicate standard deviation. Curves were fitted to datasets with matching colors.

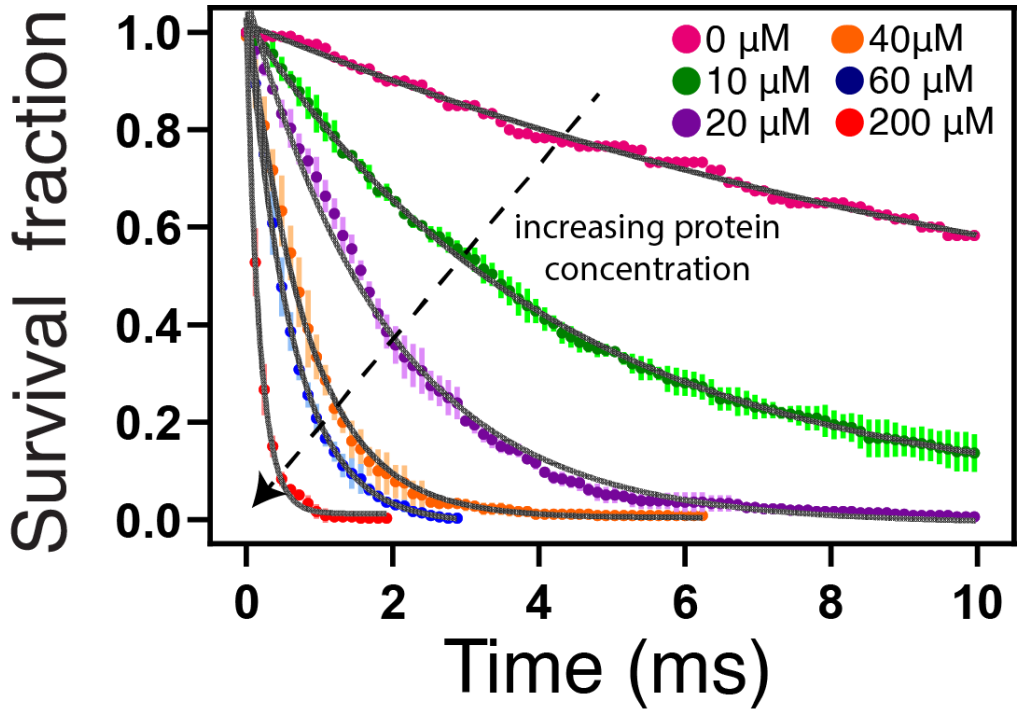

Figure 17 Survival fractions in the large (N = 12000) system at the given concentrations with the specific interaction potential of $u_{sp} = 9kT$ and NSI potential of $u_{ns} = 2kT$. Error bars, indicated with lighter colors, depict standard deviation. Gray curves show the exponential decay curves fitted to datasets.

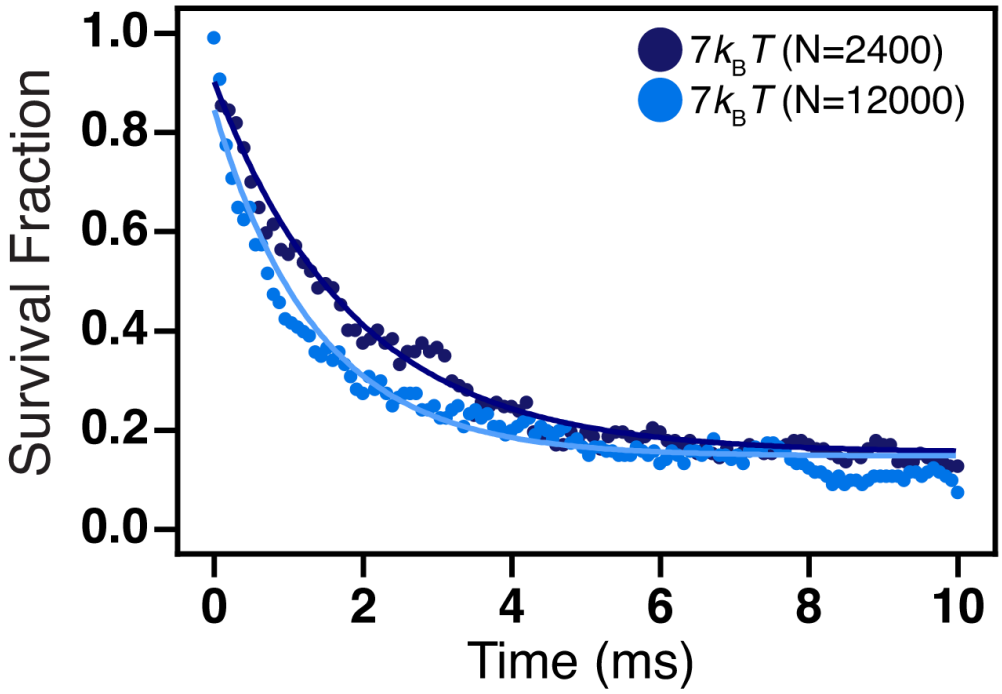

Figure 18 Survival fraction graph comparing the large system (N = 12000) and small system (N = 2400) at $u_{sp} = 7kT$, $u_{ns} = 1.6\ kT$. Curves were fitted to datasets with matching colors.

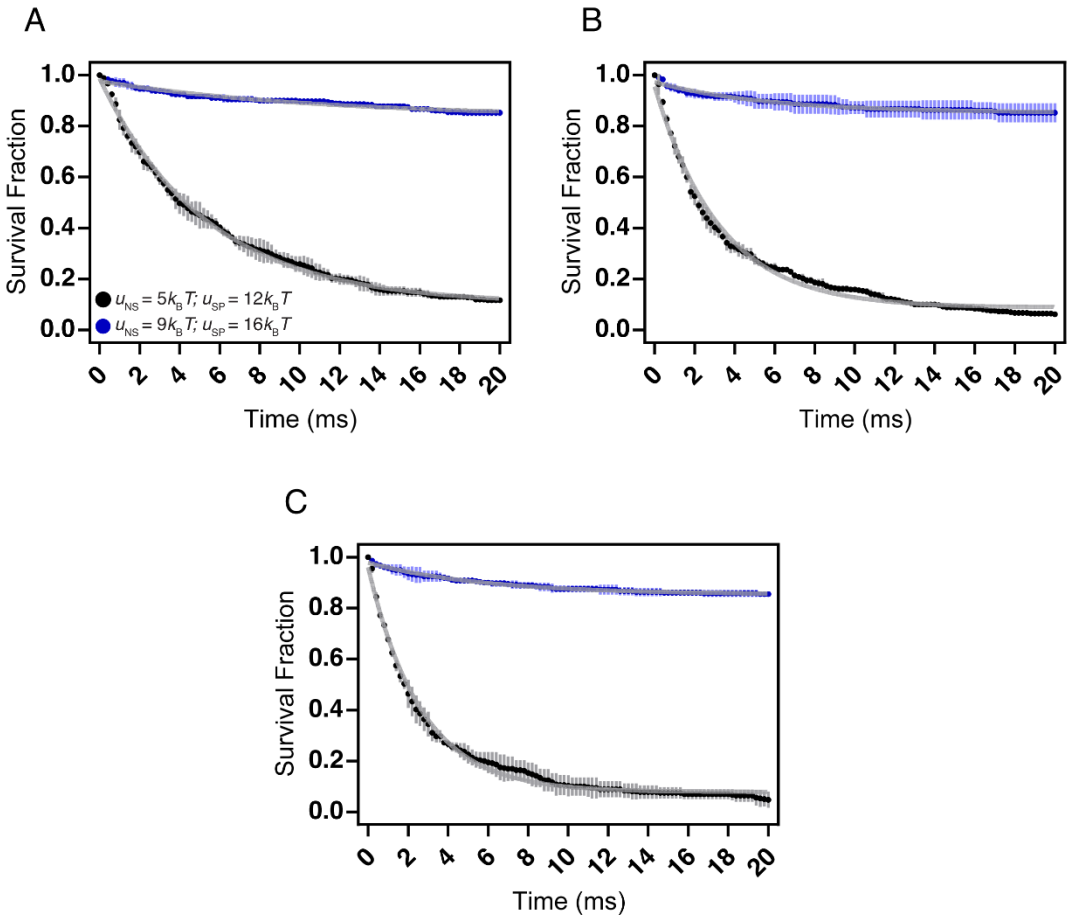

Figure 19 Comparison of various specific and nonspecific binding energies showing the survival fractions of two different attraction sets, $u_{sp} = 16kT$, $u_{ns} = 9kT$ (blue) and $u_{sp} = 12kT$,$u_{ns} = 5kT$ (black) each of which is higher than the case used in the main text. Initially, unbound protein concentrations are 60, 120, and $200\mu M$, respectively.

Various off-rates with changing concentrations also implicate that fluctuating cellular levels of NAPs or TFs may influence residence times differently depending on the growth phase.

3.2 Off-rates are Similar for the Small and the Large Systems

To verify the consistency of the small systems with the large system, we compared the survival fractions at $u_{sp} = 7kT, u_{ns} = 1.6kT$, where u_{sp} and u_{ns} represent specific and nonspecific bindings potentials, respectively. Here, the comparison was made in the absence of competitor proteins since, at $u_{sp} = 7kT; u_{ns} = 1.6kT$ SD was sufficient for the dissociation. The survival datasets and the curves are very similar for $N = 2400$ and $N = 12000$ systems. Thus, the scaling of the system does not affect the off-rates, at least for the relatively lower affinities. However, this data does not provide any information on the relation between scaling and dissociation rates for high affinities and concentrations. That is because high binding potentials and concentrations may cause chromosomal compactions, which drastically affect off-rates.

3.3 Facilitated Dissociation Relies on Multivalency

As suggested by the previous findings, facilitated dissociation relies on multivalency (i.e., interaction via multiple sites)[43,46]. Their finding implied that multivalent binding allows partial or microscopical dissociations. With the increasing concentrations and competition, partially dissociated proteins are easier to remove. Therefore, a concentration-dependent mechanism is supposed to amplify its impact in the presence of multivalent interactions. However, we wanted to check whether such

behavior could occur in our system. Thus, we employed a single bead protein model to represent monovalent binding. Since our primary protein model (i.e., the cherry model) already provided multivalency, we did not need to construct an additional model (Figure 20A).

Next, we compared their off-rates at $200\mu M$ with the same binding affinities. The outcome of the simulations demonstrated a clear disparity between the dissociation rates. Dimeric proteins (i.e., multivalent binders) dissociated with a rate ~5 fold higher than monomeric proteins (i.e., monovalent binders). That difference is very significant according to the t-test we applied to the datasets. Additionally, the error bars showing the standard deviation of three replicates are tiny, validating the consistency of the behavior (Figure 20B). Although it was possible to run similar simulations at different concentrations, the apparent distinction and the small standard deviations rendered additional simulations unnecessary.

3.4 DNA Occupancy by DNA Binding Proteins Decreases with Multivalency

DNA binding proteins, including nucleoid-associated proteins (NAPs) and transcription factors (TFs), spend significant portions of their lifetimes attached to the chromosomes. Consequently, a significant fraction of DNA binding proteins should be bound to the DNA at any given time. Note that, here, there is no distinction between being attached to a specific or a nonspecific site on DNA. Experimental data suggest that this fraction varies among the NAPs and with respect to the cellular stage.

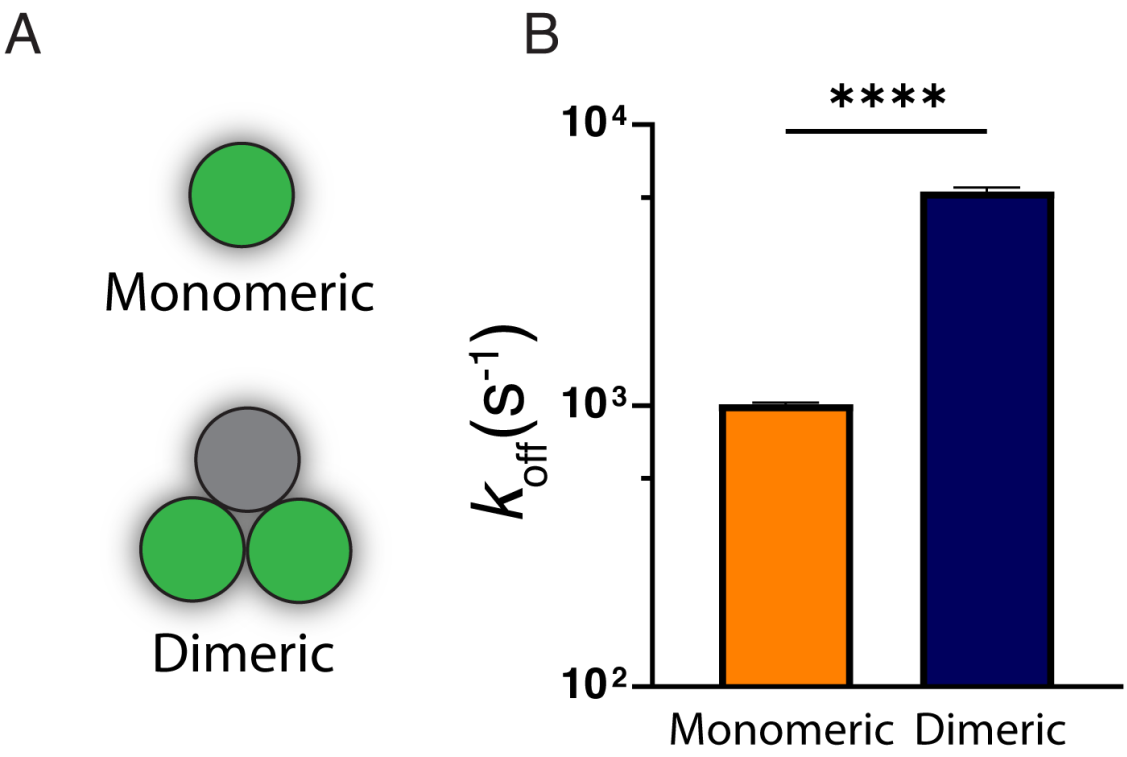

Figure 20 Comparison of FD rates between monomeric and dimeric proteins. (A) Coarse-grained models of monomeric and dimeric DNA-binding proteins. (B) Dissociation rates of monomeric and dimeric proteins at $200\mu M$ and $u_{sp} = 9kT, u_{ns} = 2kT$. We applied an unpaired two-tailed Student's t-test for statistical comparison.

Intuitively, we recognized that the fraction of DNA-bound proteins should depend on several factors, including affinity, concentration, and the nature of binding between DNA and protein. Here, as an extension to multivalency studies, we tested whether multivalently and monovalently binding proteins exhibit contrasting behaviors. Instead of running another set of simulations, we used the previous datasets through which we tested the effect of multivalency on off-rates.

Analysis of the bound fractions of NAPs revealed that, at the same protein concentration ($200\mu M$) and affinity ($u_{sp} = 9kT, u_{ns} = 2kT$), the fraction of bound proteins is significantly higher for monovalently interacting proteins. Dimeric proteins were found $\sim 58\%$ on the chromosome. At the same time, the bound fraction of monovalent proteins reached near 72% at the quasi-equilibrium state of the simulations (Figure 21). Another way to interpret this data is that 42% of the dimeric proteins are free (i.e., not attached to DNA), whereas only 28% of the monomeric proteins are separated from the chromosome. Although these two distinct perspectives convey the same message, the latter emphasizes the drastic change between the samples.

This variation of DNA occupancy is also consistent with the effect of multivalency on FD. As we described in the previous section, dissociation rates are significantly higher for multimeric proteins compared to monomeric proteins. Thus, we expect proteins with higher dissociation rates to have less DNA occupancy. Consistently, dimeric proteins were found attached to DNA in fewer quantities. Additionally, relatively minor deviations (depicted by error bars) from the averages confirm the outcomes' reliability. Thus, no further analysis was required.

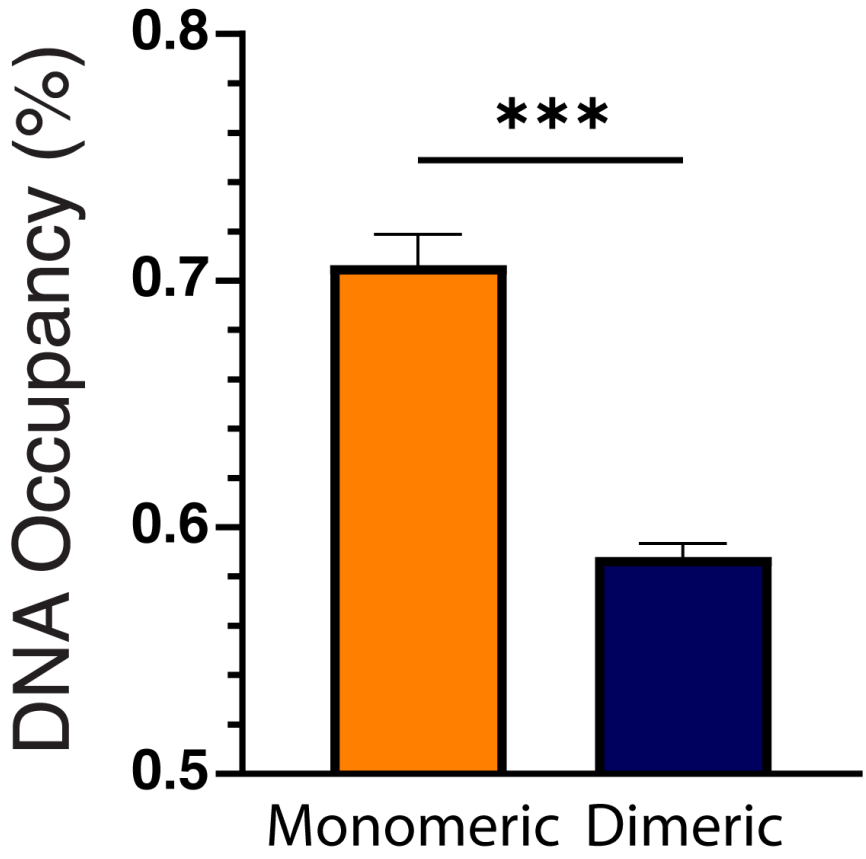

Figure 21 Fractions of the bound proteins to total proteins in the case of monomeric and dimeric proteins in the presence of $200\mu M$ free proteins. We applied an unpaired two-tailed Student's t-test for statistical comparison.

3.5 Nonspecific Interactions Alter the Transcription Factor Dissociation Rates

DNA binding proteins can interact with DNA specifically and nonspecifically. In specific interactions, protein has higher affinities towards specific DNA sequences, including promoters, operators, and enhancers. However, DNA binding proteins can also bind to the rest of the chromosome in a nonspecific manner. One of our model NAPs, Fis, binds to DNA specifically and acts as a transcription factor also interacts with the rest of the chromosome, serving as a structural protein similar to the other NAPs. NAPs have a diverse range of nonspecific interactions (NSIs), impacting TF residence times and chromosome architecture. Relevant to the primary purpose of our study, we first wanted to investigate the off-rates with respect to changing nonspecific affinity. To do so, we kept the specific interaction potential at $u_{SP} = 9kT$ and had two sets of NSIs that are $u_{ns} = 2kT$ (i.e., the main simulation parameter) and $u_{ns} = 5kT$. We ran the simulations at 9 different concentrations with three replicates for each case.

In the absence of competitor proteins, higher NSI strength (i.e., $u_{ns} = 5kT$) lead to faster dissociations. However, at $10\mu M$ both NAPs exhibited very similar off-rates. Remarkably, with the increasing free protein concentrations, NAPs with lower NSI strength (i.e., $u_{ns} = 2kT$) demonstrated much higher off-rates. Moreover, the difference between the dissociation rates increased linearly with the increasing competitor concentration. Consistently, at peak concentrations ($\geq 120\mu M$), NAPs with $u_{ns} = 2kT$ have more than 10-fold higher off-rates (Figure 22).

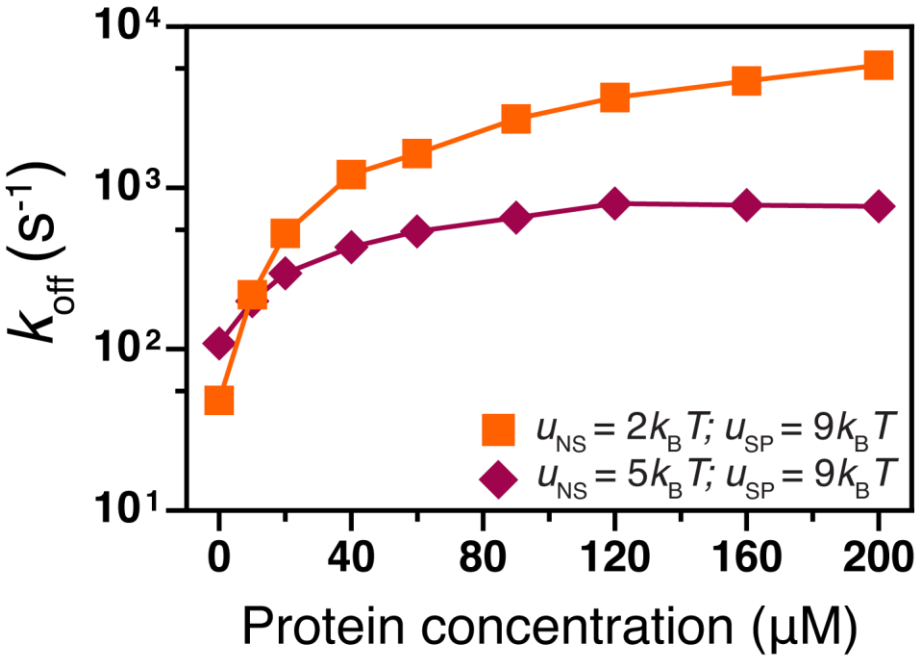

Figure 22 Dissociation rates of proteins at different NSI energies with nine different concentrations. Error bars are not shown when under the symbol. Data points are joined to guide the eye.

These results suggest that NSIs may have a critical role in residence times of DNA binding proteins. Therefore, they might also be a regulatory factor in the transcription process. However, dissociation rates alone do not provide enough information as to how NSIs stabilize DNA-protein interactions. Also, linearly increasing off-rate disparities indicate the presence of a third factor in the play. Thus, the effect of nonspecific protein-DNA interactions requires further systemic studying.

3.6 Increasing Nonspecific Interactions Leads to a Concentration-Mediated Biphasic Dissociation Behavior

After realizing the impact of nonspecific interactions at dissociation, we wanted to diversify the range of NSIs whilst keeping specific interaction potential at $u_{sp} = 9kT$. Therefore, we gradually (by $1kT$) increased the nonspecific attraction potential from $u_{ns} = 0$ to $u_{ns} = u_{sp} = 9\ kT$. Along with the 10 different nonspecific binding strengths, we employed two concentrations that are $20\mu M$, and $60\mu M$. Indeed, these concentrations were used to mimic the fluctuating quantities of several NAPs[4].

At nonspecific interaction potential range $0kT \leq u_{ns} \leq 3kT$, both $20\mu M$ and $60\mu M$ concentrations show only minor fluctuations in their dissociation rates (Figure 23). However, at this interval, $60\mu M$ case has 3-fold higher off-rates, which is significant and consistent with our previous findings. Then, as the NSI potential increases to the range $4kT \leq u_{ns} \leq 6kT$, off-rates for $60\mu M$ case drops gradually but drastically. By the time we reach $u_{ns} = 6kT$, dissociation rates drop 3-fold in comparison to off-rates at $0 \leq u_{ns} \leq 3$.

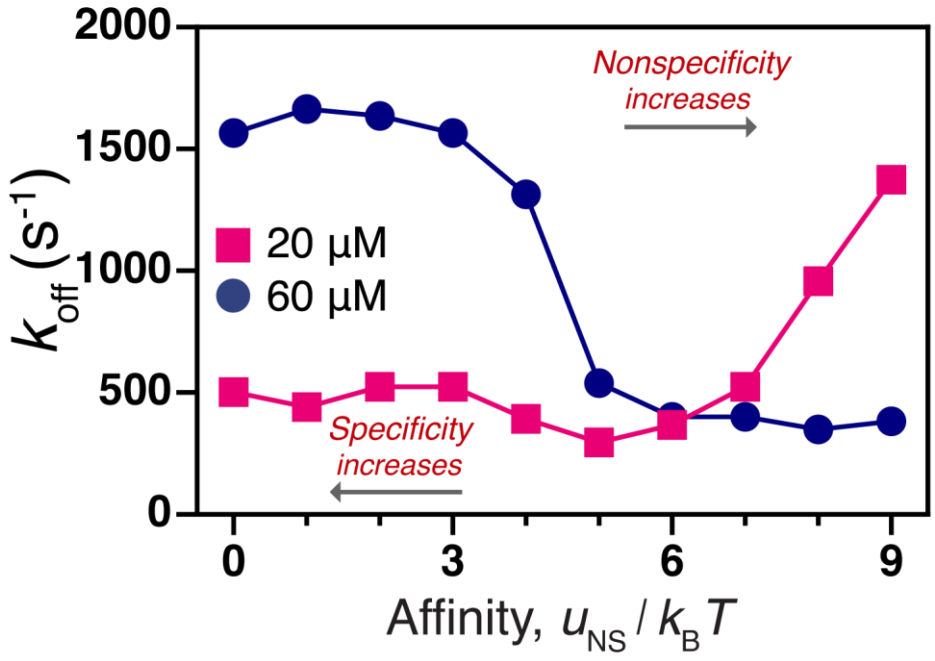

Figure 23 Biphasic dissociation behavior with increasing NSI affinity at two different concentrations with a fixed $u_{sp} = 9kT$. Data points are joined to guide the eye. Error bars are not shown when under the symbol.

Remarkably, $20\mu M$ case shows only small insignificant fluctuations in its off-rates until $u_{ns} > 6$. As we will describe in another section, high NSIs and protein concentration cause the chromosome to collapse onto itself, entrapping the proteins, which, in turn, increase the residence times thus, dropping the off-rates. Therefore, the drastic decrease of $60\mu M$ case at $4kT \leq u_{ns} \leq 6kT$ might be a result of nucleoid compaction. However, at the same NSI interval, $20\mu M$ concentration is insufficient for chromosome compaction, explaining the relative flatness.

The most drastic change takes place as the nonspecific interaction potential approaches the specific interaction energies. As expected, highly collapsed chromosome for $60\mu M$ case still results in high residence times (or low off-rates) at $u_{ns} > 6$. Outstandingly, at the same NSI potentials, $20\mu M$ case shows a reversed pattern in a biphasic manner. As we gradually increase the NSI affinity towards the specific affinity (i.e., $u_{ns} \rightarrow u_{sp}$), $20\mu M$ case exhibited a remarkable increase in its off-rates. At $u_{ns} = u_{sp} = 9\ kT$, its dissociation rate is almost 3-fold higher than its initial rates, arriving at the level of dissociation of $60\mu M$ at low NSI strengths.

One possible explanation is that the chromosome is not totally collapsed at $20\mu M$; thus, the NAPs are free to move along the DNA or among the chromosome segments. Additionally, there are a hundred times more nonspecific DNA sites for a protein to interact in the $N = 12000$ system. With the increasing NSI affinity, proteins can bind to those regions just as strongly. Consequently, proteins can dissociate from their original binding sites more rapidly to migrate to other sites along the DNA polymer.

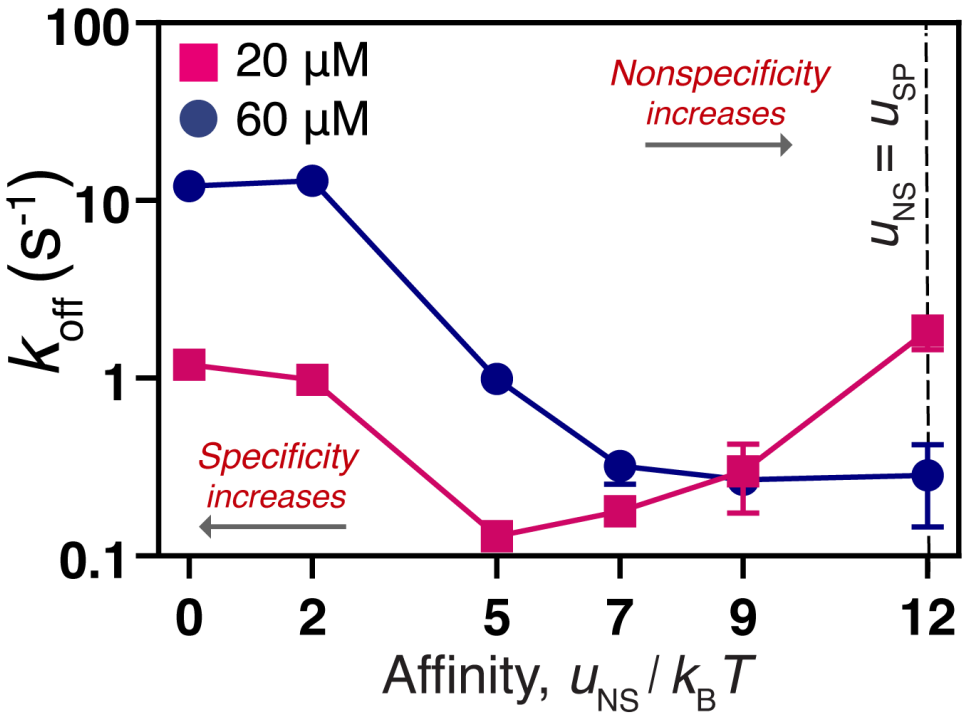

Figure 24 Biphasic dissociation behavior with increasing NSI affinities at two different concentrations also with a higher $u_{sp} = 12kT$ are shown. Here, Langevin and NVT conditions applied together. Data points are joined to guide the eye. Error bars are not shown when under the symbol.

Another but a related interpretation is that DNA segments of the polymer are free to move and fluctuate within the confinement. Therefore, those segments could encounter bound proteins at any given time. Consequently, segments can remove bound proteins from their original binding sites as they fluctuate. Here, the segments act as competitors for proteins in a similar but inversed mechanism to facilitated dissociation. Therefore, this process is referred to as segmental FD.

Additionally, we employed another thermostat (i.e., NVT plus Langevin) and specific affinity potential of $u_{sp} = 12kT$. This time, we only used 6 different nonspecific interaction potentials, again approaching to specific interaction strength ($u_{ns} \rightarrow u_{sp}$). In addition to 10-fold lower off-rates than $60\mu M$ case, $20\mu M$ sample showed around a 10-fold decrease in its off-rates at $u_{ns} = 5kT$. More prominently, the biphasic pattern persisted in this setup, suggesting that this behavior could be observed for various DNA binding proteins (Figure 24).

Note that NVT (i.e., constant Number (N), Volume (V), and Temperature (T)) and Langevin algorithms work similarly. Therefore, in the rest of the simulations, we applied Langevin along with NVE, where "E" stands for energy.

3.7 Nonspecifically Interacting Proteins Demonstrate an Inversed Facilitated Dissociation Pattern

Previous sections described how nonspecific affinity strength might have distinct effects on dissociation rates in a concentration-dependent manner. Following those findings, we wanted to focus on how nonspecifically interacting proteins ($u_{ns} = u_{sp}$) behave in terms of facilitated dissociation. Thus, we utilized three distinct NSI

potentials that are 0, 5, and $9kT$ while maintaining u_{sp} at $9kT$. The proteins with no affinity towards nonspecific DNA sites are the specific proteins. On the other, nonspecific proteins ($u_{ns} = u_{sp} = 9kT$) interact with every part of DNA polymer with the same affinity.

In this setup, we employed 9 different concentrations for each case, similar to the previous FD setups. Consistent with the previous FD data, $u_{sp} > u_{ns}$ samples exhibited an apparent concentration dependency in their off-rates. Even though $u_{ns} = 5kT$ and $u_{ns} = 0kT$ had very similar off-rates in the absence of competitors, with the increasing protein concentrations, specific proteins (i.e., $u_{sp} = 9kT; u_{ns} = 0kT$) demonstrated around 10-fold higher off-rates compared to $u_{ns} = 5kT$. That is again consistent with our previous findings on the relation between NSIs and dissociation rates. Additionally, $u_{sp} = 9kT; u_{ns} = 5kT$ sample's dissociation rates reach a plateau above $80\mu M$ consistent with chromosomal compaction at that affinity and concentration (Figure 25).

Despite further validation of our previous results, the most critical finding here is about nonspecific proteins ($u_{ns} = u_{sp} = 9kT$). In specifically interacting proteins ($u_{sp} > u_{ns}$), we observe an FD pattern where off-rates rise with increasing concentration. For nonspecific proteins, we observe very high dissociation rates (i.e., comparable to that of specific proteins at $200\mu M$) in the absence of competitors. In contrast with the specifically interacting proteins, the gradual increase in protein concentrations exerts an inverse effect on dissociation rates of the nonspecific proteins. As the protein concentration is increased, dissociation rates drop gradually.

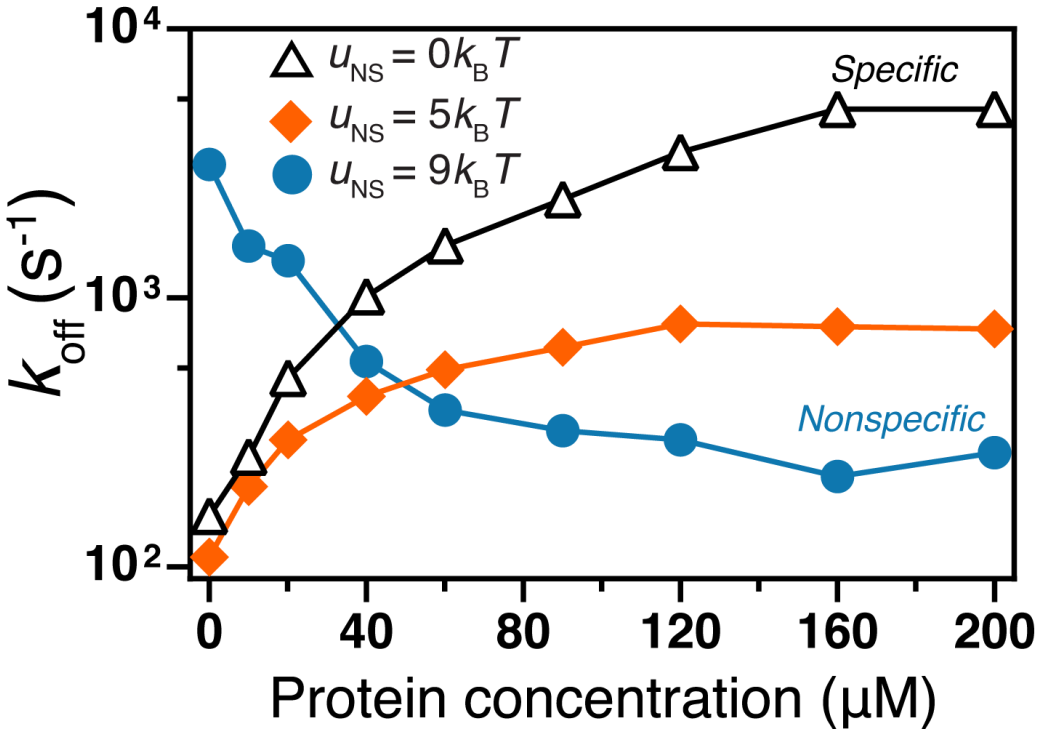

Figure 25 Dissociation rates of proteins and the reverse FD pattern of the nonspecific proteins. Off-rates are depicted with respect to concentration for three distinct NSI potentials. Here, the specific attraction strength is $u_{sp} = 9kT$ for all cases.

At high protein concentrations ($> 80\mu M$), off-rates become drastically lower by more than 10-fold compared to relatively low concentrations. It could also be interpreted that residence times of nonspecific proteins significantly increase with their rising concentrations. Therefore, concentration has a stabilizing role in nonspecific bindings.

Also-notably, nonspecific protein off-rates reach a plateau after $80\mu M$ similar to the specifically interacting proteins, suggesting that this chromosome compaction-dependent behavior of dissociation rates is also the case for nonspecific proteins.

3.8 Nonspecific Protein-DNA interactions are the Driving Forces of Chromosomal Compaction in a Concentration-Dependent Manner

Thus far, we have only focused on the off-rates with respect to DNA-protein affinities and protein concentrations. Notably, at high NSI potentials and protein concentrations, we observed that the effect of FD diminishes, and dissociation rates reach a plateau (i.e., no significant variation along the x-axis). Therefore, we wanted to investigate the reason behind such behavior.

First, we started with a visual investigation of the simulations. Since the LAMMPS molecular dynamics simulations *dump* the positions of the beads at every simulation step, we could visualize the simulation data using OVITO (*Open Visualization Tool*). Pretty obviously, some chromosomes formed highly compact nucleoid structures, whereas some were mostly relaxed.

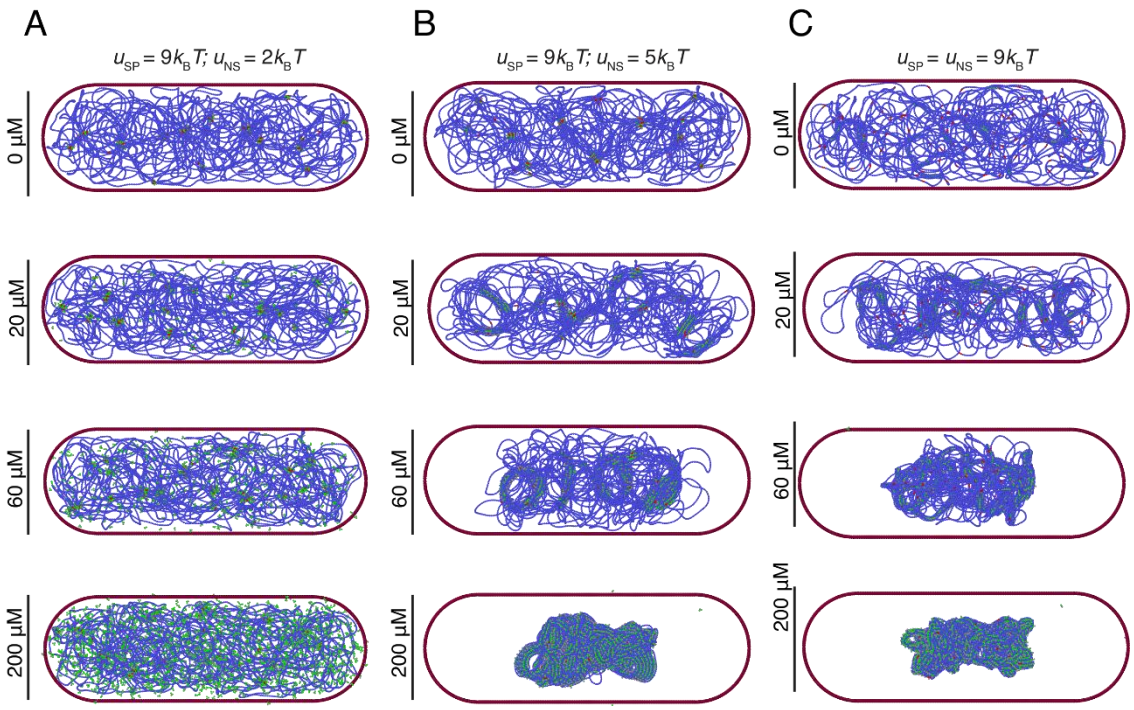

Figure 26 Snapshots of the varying chromosome architectures at four distinct protein concentrations and at NSI potentials of (A) $u_{ns} = 2kT$, (B) $u_{ns} = 5kT$, and (C) $u_{ns} = 9kT$.

Our investigation of three distinct NSIs energies that are $u_{ns} = 2kT$, $u_{ns} = 5kT$, and $u_{ns} = 9$ along with specific protein-DNA interaction strength of $u_{sp} = 9kT$, revealed various chromosome conformations.

First of all, in the absence of competitor proteins, the chromosomes occupy $\sim 100\%$ of the confined volume in all three samples. Note that, even without free proteins, initially bound proteins have a cellular concentration of $\sim 12\mu M$. This occupation pattern is persistent among all the concentrations of low NSI cases (i.e., $u_{ns} = 2kT$). As we set the free protein concentration to $20\mu M$, chromosomes started to compact slightly in the other two samples ($u_{ns} = 5kT$ and $u_{ns} = 9kT$). However, there is no clear distinction between their compaction levels at this concentration. Next, we increased the protein concentration to $60\mu M$. Notably, both chromosomes collapsed drastically; nevertheless, the compaction rate was significantly higher for $u_{ns} = 9kT$. At the peak protein levels of $200\mu M$, both samples resulted in highly condensed nucleoids (Figure 26).

Distinct responses to the same protein concentrations revealed the cause of nucleoid compaction as nonspecific interactions. Even with $200\mu M$ free proteins ($212\mu M$ total concentration including the bound proteins), low NSI set (i.e., $u_{ns} = 2kT$) was unable to condense the chromosome not even slightly. Even though two other samples have the same specific binding potential (i.e., $u_{sp} = 9kT$), their relatively higher nonspecific interaction strengths (i.e., $u_{ns} \geq 5$) allowed them to initiate chromosome condensation. Consistently, the sample with the highest nonspecific interaction potential compacted the chromosome more effectively.

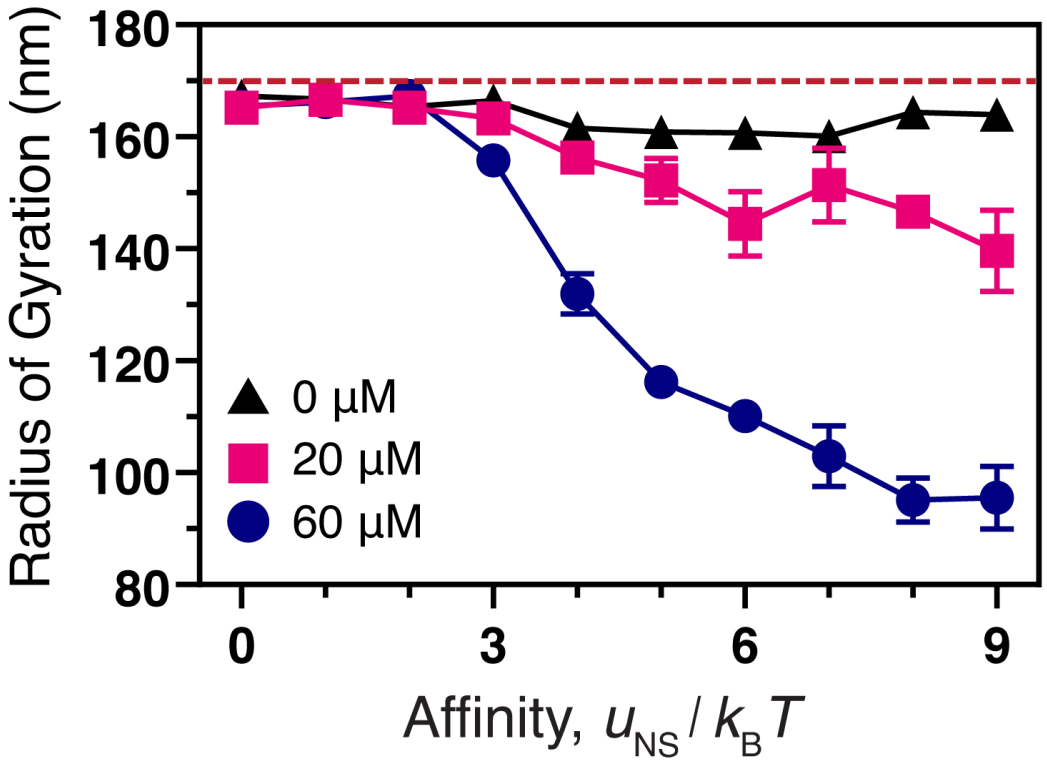

Figure 27 Radius of gyrations with respect to NSI affinity for three distinct free protein concentrations. The dashed red line depicts the maximum radius of gyration possible. Here, the specific attraction strength is $u_{sp} = 9kT$ for all cases.

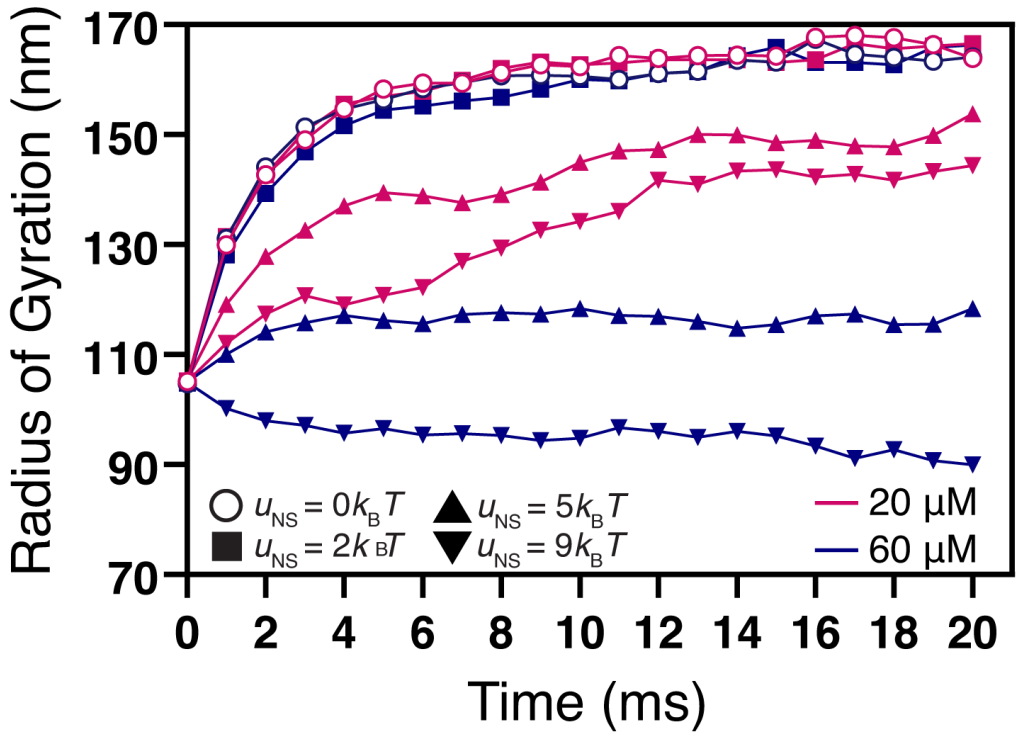

Figure 28 Radius of gyrations with respect to time for different NSI affinities and free protein concentrations. Here, the NSI potentials and concentration are depicted by symbols and concentrations, respectively. Initial Rg is the same for all the cases. Here, the specific attraction strength is $u_{sp} = 9kT$ for all cases.

Although the visual inspections alone revealed nonspecific interactions as the driving force of nucleoid structuring, they provide limited reliability. Accordingly, we quantified the radius of gyrations (R_g) with respect to gradually increasing NSI potentials with three different free protein concentrations.

In the absence of competitor proteins, there are only initially bound proteins at a relatively low concentration of $12\mu M$. At this quantity, the DNA polymer is highly relaxed, approaching the maximum Rg the confinement allows that is $170nm$ (Figure 27). Even with $u_{ns} \rightarrow u_{sp}$, the radius of gyration does not drop below $160nm$, suggesting that low DNA-binding proteins concentrations are not sufficient for nucleoid compaction.

At $20\mu M$ free protein concentration ($32\mu M$ of protein in total), relatively low NSI potentials ($u_{ns} \leq 3$) did not provide any chromosomal condensations. However, as the NSI potential is increased, the chromosome started to condense. Although the decrease in Rg is not very smooth, the $20\mu M$ case reaches its R_{g_min} where $u_{ns} = u_{sp} = 9kT$.

Thus far, the relation is still obscure. However, there is a definitive gradual decrease in Rg along the x-axis for the $60\mu M$ case while $u_{ns} > 3$. These findings suggest that chromosome compaction is a concentration-mediated process and it requires relatively high nonspecific interaction energies.

3.9 Specific and Nonspecific NAPs Have Opposite Responses to Nucleoid Condensation

Our NSI studies revealed that NSIs are crucial for nucleoid structuring in a concentration-dependent manner. Additionally, we observe that increasing protein concentrations do not further enhance off-rates above $80\mu M$ for relatively high NSI strengths. We suspected that these two distinct findings might be related. To test whether these behaviors are intertwined, we analyzed the dissociation rates with respect to the radius of gyration.

Since the low NSI potential (i.e., $u_{ns} = 2kT$) case is unable to collapse the DNA polymer to any extent, we could not deduce any relation between Rg and off-rates. However, a relatively high nonspecific interaction energy (i.e., $u_{ns} = 5kT$) shows that with the decreasing Rg, dissociation rates increase (Figure 29). This result may seem contradictory to what we discussed in the previous sections. However, note that decreasing Rg implies chromosome condensation, which depends on the protein concentrations and at high concentrations, FD is more effective. Hence, $u_{ns} = 5kT$ case alone does reveal a mechanism. Nevertheless, nonspecific NAPs ($u_{ns} = u_{sp} = 9kT$) demonstrate a completely opposite response to changing chromosome relaxation levels. The decreasing radius of gyrations lead to lower dissociation rates for the nonspecific NAPs. This distinction between the specific and the nonspecific proteins suggests that the residence times of DNA-binding proteins depend on the interaction type.

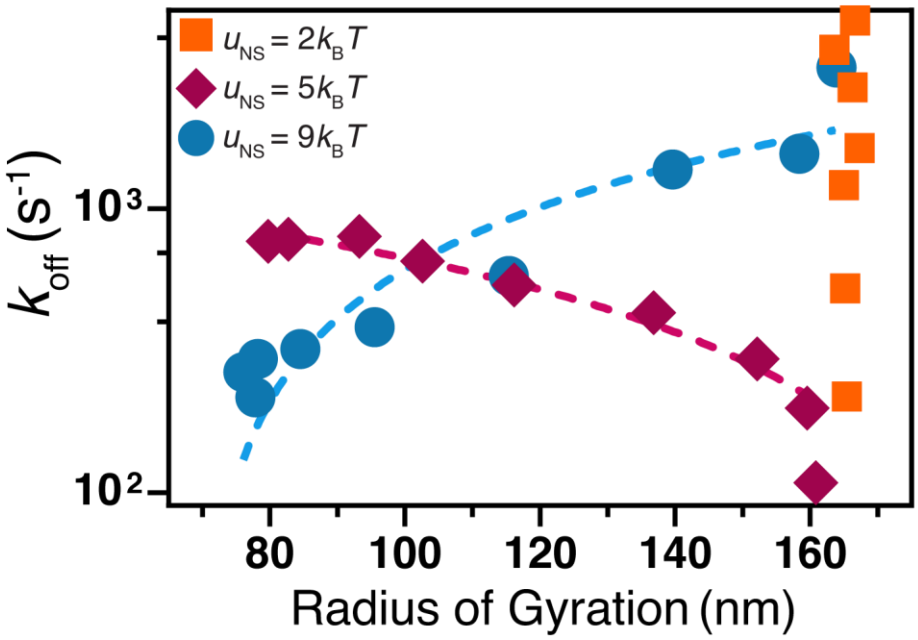

Figure 29 Off-Rates with respect to the radius of gyrations for three NSI levels. Dashed lines are the nonlinear fits to show the pattern. Here, the specific attraction strength is $u_{sp} = 9kT$ for all cases.

3.10 Restraining Chromosomal Fluctuations Decouples Segmental FD from Protein-Mediated FD

High protein concentrations and nonspecific protein-DNA potentials enable the chromosome to collapse onto itself, forming a more compact conformation. The resulting configuration restricts protein dissociation and allows a DNA-segmental FD[49], where the DNA segments compete for DNA-binding proteins, removing them from their original binding sites.

Despite the switched roles in comparison to protein-mediated FD, DNA-segmental FD is very effective when the nucleoid is compact. Therefore, we restrained the movement of the chromosome to decouple these two effects. In our primary experimental setup, DNA segments are able to roam within the membrane freely. Restricting the fluctuations of the chromosome segments removes any possibility of DNA-segmental FD. Thus, the only type of facilitated dissociation here is the protein-mediated FD.

In this setup, we utilized only one NAP concentration that is $60\mu M$, along with three distinct NSI potentials (i.e., 2, 5, and $9kT$). Furthermore, we employed four chromosomes with diverse compaction levels, varying from highly condensed ($R_g \cong 85nm$) to highly relaxed ($R_g \cong 165nm$) with moderately compact or relaxed DNA polymers ($R_g \cong 100$ and $R_g \cong 125nm$) in between. All three cases have their highest dissociation rates at $R_g \cong 100$, where $u_{ns} = 5kT$ case have more than 10-fold increase in its dissociation rates compared to other R_g levels (Figure 30).

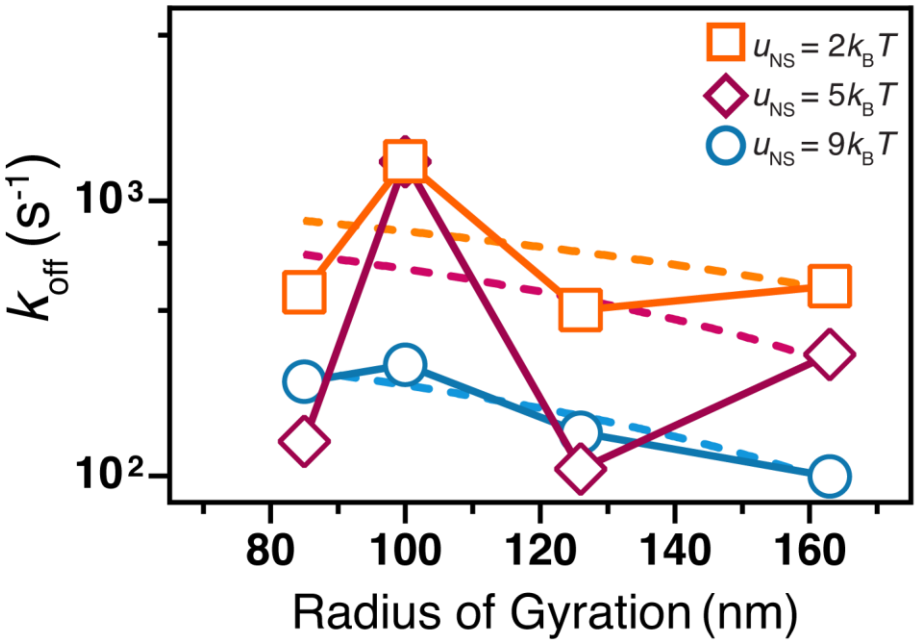

Figure 30 Dissociation rates with respect to four distinct fixed Rg for three NSI levels when the chromosome is restrained. Dashed lines are the nonlinear fits to show the pattern. Here, the specific attraction strength is set to $u_{sp} = 9kT$ for all cases.

Except for the nonspecific proteins ($u_{ns} = u_{sp} = 9kT$), samples show a drastic decrease (5 to 10-fold) in their off-rates followed by significant escalation at $R_g \cong 125nm$ and $R_g \cong 165nm$, respectively. On the other hand, nonspecific proteins exhibit only slight drops in their off-rates at the same compaction levels.

Here, the first interpretation is that the nucleoid architecture might impact off-rates extensively. All three cases reaching their maximum off-rates by far in the presence of moderately collapsed DNA polymer indicates that the protein-mediated FD could be most efficient with such chromosome architecture. A possible explanation for this phenomenon is that a highly dense chromosome entraps proteins, thus increasing residence times and decreasing off-rates. In the meantime, the DNA segments are sparsely located in a relaxed polymer. Therefore, proteins cannot migrate to the neighboring segments conveniently. However, with a moderately collapsed nucleoid, these dissociation rate reducing effects vanish. Consequently, the DNA residence times of proteins decline drastically. Moreover, these findings also suggest that the nonspecific (e.g., HU) and the specific (e.g., Fis) DNA-binding proteins could respond to chromosomal conformations differently. Overall, chromosome compaction is a possible regulator of protein residence times.

3.11 DNA Binding Proteins Could be Removed by Lower or Higher Affinity Proteins

In all living cells, the genome is susceptible to mutations. As the driving forces of evolution, mutations enable cells to adapt to their environment. Potential mutations within the binding domains of DNA-binding proteins could affect their binding affinities towards their binding sites.

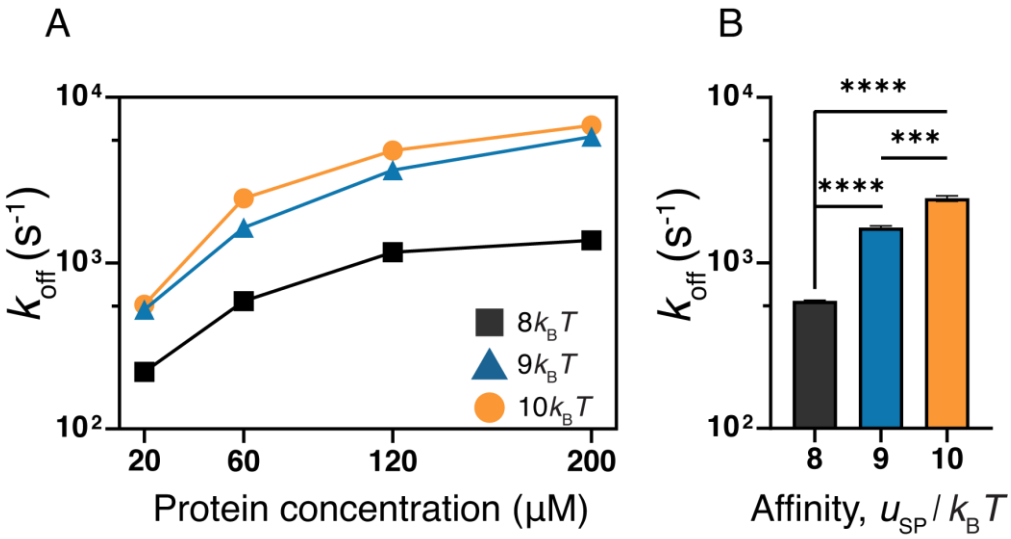

Figure 31 Dissociation rates with respect to concentration in the presence of higher ($u_{sp} = 10kT$) and lower affinity ($u_{sp} = 8kT$) competitors. Here, specific attraction strength for bound proteins is set to $u_{sp} = 9kT$ whereas, NSI affinity is $u_{ns} = 2kT$ for all cases.

To investigate how the proteins with higher or lower binding energies alter the FD behavior, we utilized another setup. In this case, we employed three competitor binding affinities that are 8, 9, and $10kT$ while maintaining a $9kT$ binding potential for the initially bound proteins along with $u_{ns} = 2kT$ for all proteins.

With all four concentrations that we utilized, both the lower affinity and higher affinity cases were able to remove bound proteins from their binding sites. Additionally, there are significant distinctions of protein off-rates among the cases. As expected, competitor proteins with higher affinities ($u_{sp} = 10kT$) demonstrated the highest removal rates and vice versa. That implicates mutated proteins of the same kind, as well as various DNA-binding proteins with an affinity towards the same binding sites, are also able to participate in FD in an affinity-based manner.

3.12 DNA-Protein Interactions Lead to the Formation of Multi-Protein Complexes

In our system, even though both the proteins and the DNA polymer are self-repulsive, the chromosome and the proteins can interact with each other. Similar to the NAPs, our model proteins were able to condense the chromosome in a concentration and affinity-dependent manner. In our visual investigations of chromosome organizations for a previous part, we encounter several distinct protein-DNA formations. With our primary parameters (i.e., $u_{sp} = 9kT; u_{ns} = 2kT$), we only observed multi-protein DNA bridges where multiple binding sites from distinct DNA segments are joined together by multiple proteins (Figure 32A). However, nonbinding regions of the chromosome maintained their distance to one another due to low nonspecific potential.

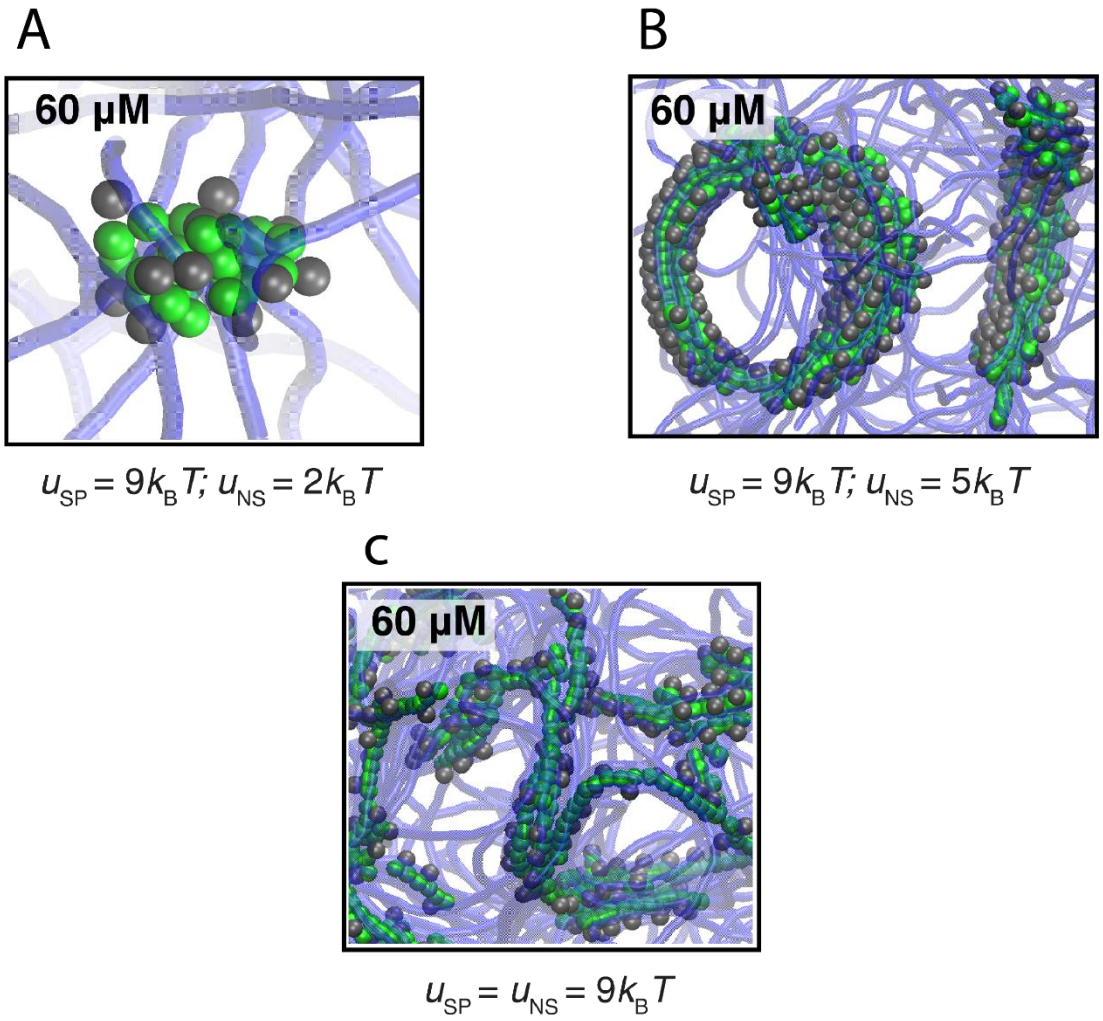

Figure 32 Snapshots of the DNA-protein clusters at $60\mu M$ free protein conctration with NSI potentials of (A) $u_{ns} = 2kT$, (B) $u_{ns} = 5kT$, and (C) $u_{ns} = 9kT$. Snapshot are obtained using VMD, and DNA is made semi-transparent for better visualization.

In case of a relatively high NSI potential of $u_{ns} = 5kT$, we see much larger clusters with interesting conformations as the chromosome is relatively collapsed; thus, the segments are closer. Most noticeably, we observed a donut-shaped cluster formed by hundreds of proteins along with multiple turns of DNA segments (Figure 32B). On the other hand, nonspecific proteins ($u_{ns} = u_{sp} = 9kT$) formed filamentous clusters due to lack of specificity in their DNA bindings, scattering proteins all around the chromosome (Figure 32C). Although the physiological significance of such cluster formations is unclear, they might participate in chromosome organization and lead to stable complexes, thus increasing protein residence times.

3.13 Clusters Size and Numbers Vary Depending on the Affinity and Concentration

Following the investigation of cluster architectures, we wanted to quantify both the number of the clusters and the size of each cluster in terms of protein count. Our setup included three distinct free protein concentrations that are 0, 20, and $60\mu M$ along with five different NSI potential at the range of $0kT \leq u_{ns} \leq 9kT$ with fixed specific interaction energy of $u_{sp} = 9\ kT$. For each case, we used three replicates. Then, for each case, we have taken the average of the number of clusters and pooled all the cluster sizes from the replicates.

Consistent at all five NSI potentials, the average cluster size grows with the increasing protein concentrations and affinities. However, the average number of the clusters formed fluctuates and exhibits no apparent pattern due to the formation of large clusters decreasing the number of clusters at high affinities and concentrations.

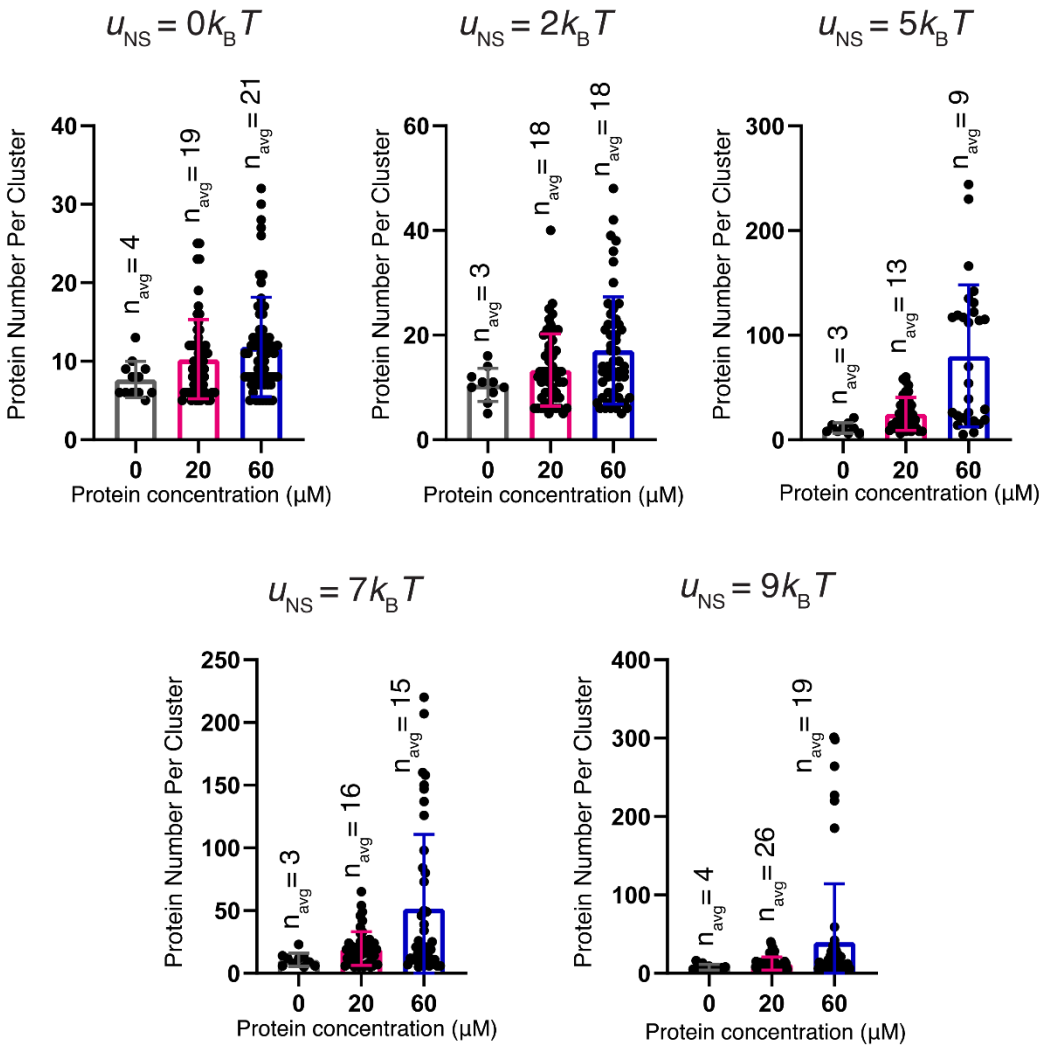

Figure 33 Cluster sizes and numbers with changing protein concentrations and NSI levels. Each black dot depicts a cluster, and the averages (n_{avg}) are taken from three replicates for each case.

3.14 Distribution of Residence Times Drastically Change with Nonspecific Interaction Affinity

Our previous setups focused on FD, where we only checked the first dissociation of the initially bound proteins. However, we wanted to analyze how long each protein stays bound to the DNA polymer. Therefore, we utilized previous simulation files of $u_{ns} = 2kT$ and $u_{ns} = 5kT$ at $20\mu M$ and $60\mu M$ where $u_{sp} = 9kT$ for each case. For each protein -whether initially bound or free- we checked whether the protein is attached to a binding site or not. Then, we analyzed the durations they reside on DNA binding sites and how many times that particular duration is encountered.

At the cases with our main parameters ($u_{ns} = 2kT$, $u_{sp} = 9kT$), we observe the highest occurrence rates for very short durations (i.e., $< 2ms$). There are fewer and fewer encounters of protein residing on DNA for longer the duration goes in an apparent decay pattern both for $20\mu M$ and $60\mu M$ (Figure 34A). Also, the higher turnover rates of $60\mu M$, where shorter durations are more frequent, are consistent with FD. In the case of high NSI potential of all proteins spends all of their lifetimes on DNA; thus, their duration frequencies are the number of the total proteins. That indicates that the high NSI potentials entrap proteins and increasing their RTs.

Additionally, we compared the percentages of the time they spend on DNA to the simulation duration (Figure 34B). As also suggested by Figure 34A, high NSI affinity cases spent 100% of their lifetimes on DNA. Contrarily, at $u_{ns} = 2kT$, $20\mu M$ spent more of their lifetimes (78%) compared to $60\mu M$ case' 72%, which is consistent with FD. Moreover, these findings suggest that our main parameters were, in fact, very optimal for studying concentration-dependent dissociation rates.

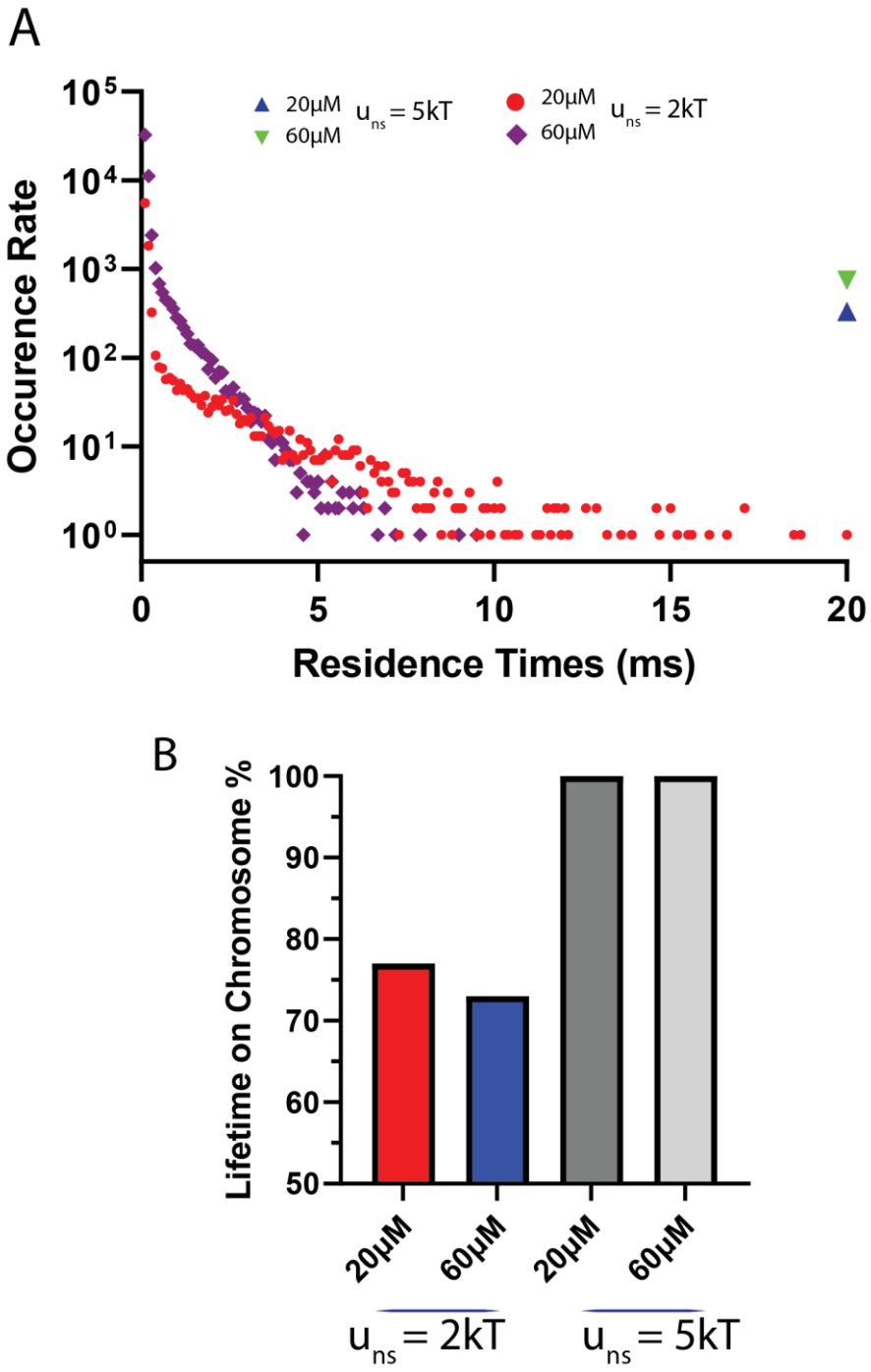

Figure 34 Residence time distribution of proteins with different NSI affinities at two different concentrations. (A)Number of occurrences in the given residence times or how many times proteins resided on DNA for the given duration. (B) The percentages of proteins residing on DNA to the total simulations time. Specific interaction affinity is $u_{sp} = 9kT$ for each case.

4 Conclusion

Transcription and its regulation are crucial for any living cell. The initiation or the inhibition of transcription depends on the temporal formation of the TF-DNA complex. Recent findings suggest that transcriptional activation and repression depend on the residence times of the transcription factors. Experimental and simulation models showed that the residence of time of a protein is intertwined with its concentration[1,3,9,10]. In light of these findings, we employed a coarse-grained model of *E. coli* to test whether the concentration-dependency of the dissociation rates persist in bacterial confinement with bacterial chromosome-structuring proteins, NAPs.

Initially, we demonstrated that FD could also occur in bacterial confinement and with a circular chromosome using a cherry model of Fis protein. Along the way, we discovered more than we aimed for at the beginning. Most prominently, we show that chromosome compaction depends on the nonspecific protein-DNA interactions rather than specific interactions. Moreover, extensive nucleoid condensations entrap protein within, drastically increasing the residence times of the proteins. That also suggests that highly compacted DNA-protein complexes might block the access of RNA polymerase to promoter regions analogous to eukaryotic heterochromatin regions. Therefore, NSIs could also be another regulatory factor of transcription.

Additionally, moderate chromosome compaction enhances protein dissociation via DNA-segmental dissociation. The most prominent example of that is the reversed FD pattern of nonspecific proteins. Contrary to specific proteins, nonspecific protein dissociation rates decreased with the increasing concentrations. Remarkably,

nonspecific proteins also exhibited a biphasic behavior in the presence of two distinct physiologically relevant concentrations, indicating dissociation behavior is complex and dependent on multiple factors. Also notably, restrained DNA polymer provided insight into DNA-segmental fluctuation and FD. When the DNA segments are restrained, protein dissociation solely depended on the protein-mediated FD. In that case, FD was most efective in the presence of moderately collapsed nucleoid rather than highly collapsed or relaxed formations, suggesting that chromosome compaction is another regulatory factor for protein off-rates.

Furthermore, changing NSI potentials resulted in the formation of diversely-shaped multi-protein complexes. Most notably, we observed donut-shaped DNA-protein clusters with a relatively high NSI affinity. These clusters could participate in local chromosomal compactions, further influencing protein off-rates and repressing RNA polymerase access. Consequently, the multi-protein cluster might be another transcriptional regulator.

Overall, our findings reveal a network of connections among the DNA-binding protein concentration, chromosome architecture, and dissociation rates. Moreover, we suggest that this network might be a part of transcriptional regulation and participate in cellular behavior (Figure 35).

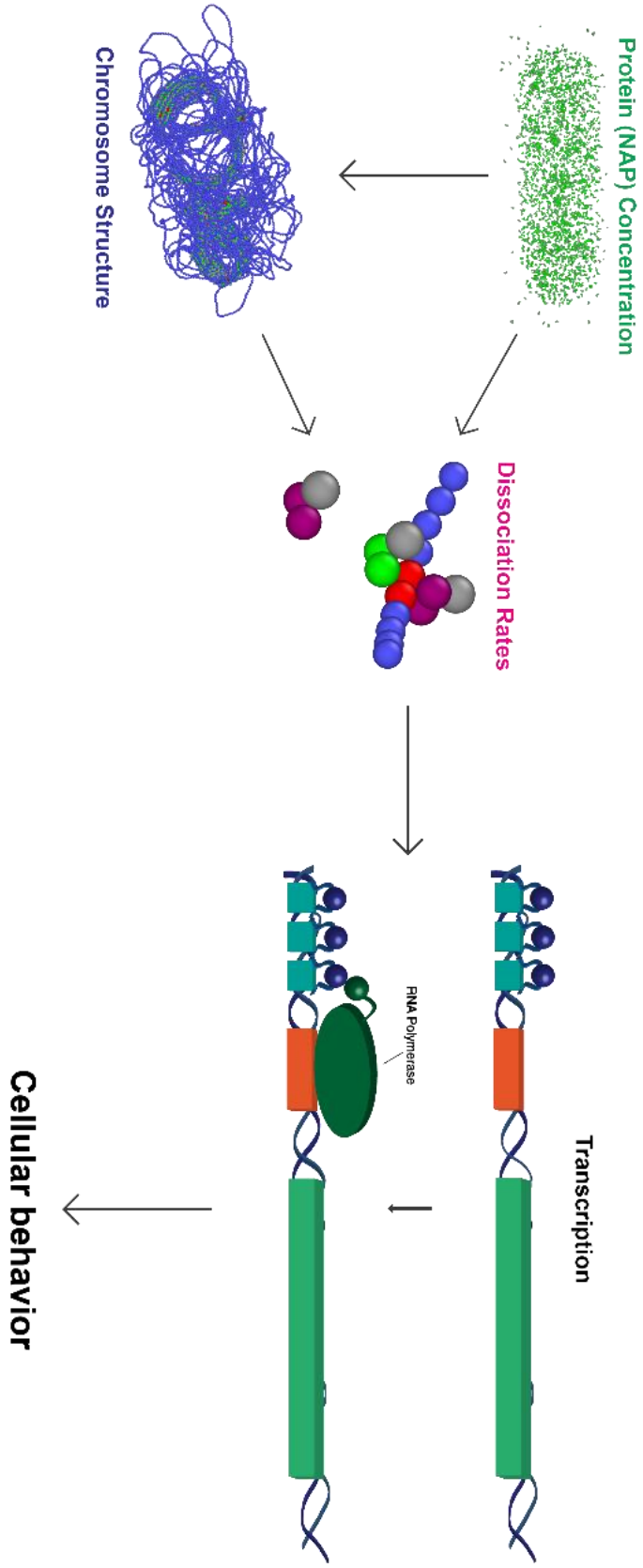

Figure 35 Possible implications of our findings.

5 References

1. DF Browning, S. B. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65 (2004).

2. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology 2016 14:10* **14**, 638–650 (2016).

3. A, M.-A. & J, C.-V. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current opinion in microbiology* **6**, 482–489 (2003).

4. Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology 2010 8:3* **8**, 185–195 (2010).

5. Dorman, C. J. Co-operative roles for DNA supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription. *Biochemical Society Transactions* **41**, 542–547 (2013).

6. Dorman, C. J. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nature Reviews Microbiology 2013 11:5* **11**, 349–355 (2013).

7. Slattery, M. *et al.* Absence of a simple code: How transcription factors read the genome. *Trends in Biochemical Sciences* **39**, 381–399 (2014).

8. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics 2012 13:9* **13**, 613–626 (2012).

9. Clauß, K. *et al.* DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Research* **45**, 11121 (2017).

10. Azpeitia, E. & Wagner, A. Short Residence Times of DNA-Bound Transcription Factors Can Reduce Gene Expression Noise and Increase the Transmission of Information in a Gene Regulation System. *Frontiers in Molecular Biosciences* **0**, 67 (2020).

11. Luijsterburg, M. S., Noom, M. C., Wuite, G. J. L. & Dame, R. T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *Journal of Structural Biology* **156**, 262–272 (2006).

12. Joshi, C. P. *et al.* Direct substitution and assisted dissociation pathways for turning off transcription by a MerR-family metalloregulator. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15121–15126 (2012).

13. Auner, H. *et al.* Mechanism of transcriptional activation by FIS: Role of core promoter structure and DNA topology. *Journal of Molecular Biology* **331**, 331–344 (2003).

14. Travers, A., Schneider, R. & Muskhelishvili, G. DNA supercoiling and transcription in Escherichia coli: The FIS connection. *Biochimie* **83**, 213–217 (2001).

15. Ross, W., Thompson, J. F., Newlands, J. T. & Gourse, R. L. E. coli Fis protein activates ribosomal RNA transcription in vitro and in vivo. *EMBO Journal* **9**, 3733–3742 (1990).

16. Graham, J. S., Johnson, R. C. & Marko, J. F. Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. *Nucleic Acids Research* **39**, 2249–2259 (2011).

17. Bradley, M. D., Beach, M. B., de Koning, A. P. J., Pratt, T. S. & Osuna, R. Effects of Fis on Escherichia coli gene expression during different growth stages. *Microbiology* **153**, 2922–2940 (2007).

18. Sing, C. E., de La Cruz, M. O. & Marko, J. F. Multiple-binding-site mechanism explains concentration-dependent unbinding rates of DNA-binding proteins. *Nucleic Acids Research* **42**, 3783–3791 (2014).

19. Kamar, R. I. *et al.* Facilitated dissociation of transcription factors from single DNA binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E3251–E3257 (2017).

20. Attie, A. D. & Raines, R. T. Analysis of Receptor–Ligand Interactions. *Journal of chemical education* **72**, 119 (1995).

21. Travers, A. & Muskhelishvili, G. DNA structure and function. *The FEBS Journal* **282**, 2279–2295 (2015).

22. Keller, N., delToro, D., Grimes, S., Jardine, P. J. & Smith, D. E. Repulsive DNA-DNA Interactions Accelerate Viral DNA Packaging in Phage Phi29. *Physical Review Letters* **112**, 248101 (2014).

23. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences* **94**, 6185–6190 (1997).

24. GS, M. Is a small number of charge neutralizations sufficient to bend nucleosome core DNA onto its superhelical ramp? *Journal of the American Chemical Society* **125**, 15087–15092 (2003).

25. Manning, G. S. The Persistence Length of DNA Is Reached from the Persistence Length of Its Null Isomer through an Internal Electrostatic Stretching Force. *Biophysical Journal* **91**, 3607 (2006).

26. Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A. & Landsman, D. Histone structure and nucleosome stability. *Expert review of proteomics* **2**, 719 (2005).

27. Martire, S. & Banaszynski, L. A. The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology 2020 21:9* **21**, 522–541 (2020).

28. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology 2015 16:3* **16**, 178–189 (2015).

29. ST Arold, P. L. G. P. J. L. H-NS forms a superhelical protein scaffold for DNA condensation. *Proc. Natl Acad. Sci. USA* **107**, 15728–15732 (2010).

30. Schneider, R. An architectural role of the Escherichia coli chromatin protein FIS in organizing DNA. *Nucleic Acids Res.* **29**, 5107–5114 (2001).

31. RT Dame, C. W. N. G. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.* **28**, 3504–3510 (2000).

32. Kahramanoglou, C. Direct and indirect effects of H-NS and Fis on global gene expression control in Escherichia coli. *Nucleic Acids Res.* **39**, 2073–2091 (2011).

33. Dame, R. T., Rashid, F.-Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nature Reviews Genetics 2019 21:4* **21**, 227–242 (2019).

34. Grainger, D. Structure and function of bacterial H-NS protein. *Biochem. Soc. Trans.* **44**, 1561–1569 (2016).

35. AL, M., CR, H. & J, L. Squish and squeeze-the nucleus as a physical barrier during migration in confined environments. *Current opinion in cell biology* **40**, 32–40 (2016).

36. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology 2017 19:4* **19**, 229–244 (2017).

37. Decombe, S. *et al.* Epigenetic rewriting at centromeric DNA repeats leads to increased chromatin accessibility and chromosomal instability. *Epigenetics & Chromatin* **14**, (2021).

38. Castaing, B., Zelwer, C., Laval, J. & Boiteux, S. HU Protein of Escherichia coli Binds Specifically to DNA That Contains Single-strand Breaks or Gaps (∗). *Journal of Biological Chemistry* **270**, 10291–10296 (1995).

39. SP Hancock, S. S. D. C. R. J. DNA sequence determinants controlling affinity, stability and shape of DNA complexes bound by the nucleoid protein Fis. *PLOS ONE* **11**, (2016).

40. Ball, C. A., Osuna, R., Ferguson, K. C. & Johnson, R. C. Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli. *Journal of Bacteriology* **174**, 8043–8056 (1992).

41. Ball, C. A., Osuna, R., Ferguson, K. C. & Johnson, R. C. Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli. *Journal of Bacteriology* **174**, 8043–8056 (1992).

42. Azam, T. A., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *Journal of Bacteriology* **181**, 6361–6370 (1999).

43. Tsai, M.-Y., Zhang, B., Zheng, W. & Wolynes, P. G. Molecular Mechanism of Facilitated Dissociation of Fis Protein from DNA. *Journal of the American Chemical Society* **138**, 13497–13500 (2016).

44. Wu, F. *et al.* Cell Boundary Confinement Sets the Size and Position of the E. coli Chromosome. *Current Biology* **29**, 2131-2144.e4 (2019).

45. Dame, R. T. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology* **56**, 858–870 (2005).

46. Kamar, R. I. *et al.* Facilitated dissociation of transcription factors from single DNA binding sites. *Proceedings of the National Academy of Sciences* **114**, E3251–E3257 (2017).

47. Graham, J. S., Johnson, R. C. & Marko, J. F. Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. *Nucleic Acids Research* **39**, 2249–2259 (2011).

48. Skoko, D. *et al.* Mechanism of Chromosome Compaction and Looping by the Escherichia coli Nucleoid Protein Fis. *Journal of Molecular Biology* **364**, 777–798 (2006).

49. RD, G. *et al.* DNA-Segment-Facilitated Dissociation of Fis and NHP6A from DNA Detected via Single-Molecule Mechanical Response. *Journal of molecular biology* **427**, 3123–3136 (2015).

50. Jung, W., Sengupta, K., Wendel, B. M., Helmann, J. D. & Chen, P. Biphasic unbinding of a metalloregulator from DNA for transcription (de)repression in Live Bacteria. *Nucleic Acids Research* **48**, 2199–2208 (2020).

51. Chen, T.-Y. *et al.* Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells. *Nature Communications 2015 6:1* **6**, 1–10 (2015).

52. Dunja Skoko, ‡,§, ben Wong, ‖,⊥, Reid C. Johnson, ‖,# and & John F. Marko*, ‡. Micromechanical Analysis of the Binding of DNA-Bending Proteins HMGB1, NHP6A, and HU Reveals Their Ability To Form Highly Stable DNA−Protein Complexes†. *Biochemistry* **43**, 13867–13874 (2004).

53. Walthers, D. *et al.* Salmonella enterica Response Regulator SsrB Relieves H-NS Silencing by Displacing H-NS Bound in Polymerization Mode and Directly Activates Transcription *. *Journal of Biological Chemistry* **286**, 1895–1902 (2011).

54. Jun, S. & Mulder, B. Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proceedings of the National Academy of Sciences* **103**, 12388–12393 (2006).

55. Brackley, C. A., Taylor, S., Papantonis, A., Cook, P. R. & Marenduzzo, D. Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proceedings of the National Academy of Sciences* **110**, E3605–E3611 (2013).

56. Bhattacharya, S. & Rattan, S. I. S. Primary Stress Response Pathways for Preconditioning and Physiological Hormesis. *The Science of Hormesis in Health and Longevity* 35–51 (2019) doi:10.1016/B978-0-12-814253-0.00003-6.

57. Wai, L. K. Telomeres, Telomerase, and Tumorigenesis -- A Review. *Medscape General Medicine* **6**, (2004).

58. Lodish, H. *et al.* Structure of Nucleic Acids. (2000).

59. Hengen, P. N., Bartram, S. L., Stewart, L. E. & Schneider, T. D. Information Analysis of Fis Binding Sites. *Nucleic Acids Research* **25**, 4994–5002 (1997).

60. Shao, Y., Feldman-Cohen, L. S. & Osuna, R. Functional Characterization of the Escherichia coli Fis–DNA Binding Sequence. *Journal of Molecular Biology* **376**, 771–785 (2008).

61. Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. Topological domain structure of the Escherichia coli chromosome. *Genes & Development* **18**, 1766 (2004).

62. Krogh, T. J., Møller-Jensen, J. & Kaleta, C. Impact of Chromosomal Architecture on the Function and Evolution of Bacterial Genomes. *Frontiers in Microbiology* **0**, 2019 (2018).

63. On the determination of molecular fields. —II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **106**, 463–477 (1924).

64. Adams, J. B. Bonding Energy Models. *Encyclopedia of Materials: Science and Technology* 763–767 (2001) doi:10.1016/B0-08-043152-6/00146-7.

65. Yang, D. C., Blair, K. M. & Salama, N. R. Staying in Shape: the Impact of Cell Shape on Bacterial Survival in Diverse Environments. *Microbiology and Molecular Biology Reviews* **80**, 187–203 (2016).

66. Blattner, F. R. *et al.* The Complete Genome Sequence of Escherichia coli K-12. *Science* **277**, 1453–1462 (1997).

67. Alberts, B. *et al.* The Structure and Function of DNA. (2002).

68. *Escherichia coli and Salmonella : cellular and molecular biology /*. *Escherichia coli and Salmonella: cellular and molecular biology* vol. 2 (ASM Press, 1996).

69. Sundararaj, S. *et al.* The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli. *Nucleic Acids Research* **32**, D293 (2004).

70. Microorganisms, N. R. C. (US) S. G. for the W. on S. L. of V. S. Correlates of Smallest Sizes for Microorganisms. (1999).

71. Miller, M. The Importance of Being Flexible: The Case of Basic Region Leucine Zipper Transcriptional Regulators. *Current protein & peptide science* **10**, 244 (2009).

72. Gonzalez, D. H. Introduction to Transcription Factor Structure and Function. *Plant Transcription Factors: Evolutionary, Structural and Functional Aspects* 3–11 (2016) doi:10.1016/B978-0-12-800854-6.00001-4.

73. Arfken, G. B., Griffing, D. F., Kelly, D. C. & Priest, J. NUCLEAR STRUCTURE AND NUCLEAR TECHNOLOGY. *International Edition University Physics* 842–869 (1984) doi:10.1016/B978-0-12-059858-8.50048-0.

74. Lynch, M. *et al.* Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics 2016 17:11* **17**, 704–714 (2016).

75. W, H., A, D. & K, S. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33–38 (1996).

76. Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering* **18**, 015012 (2009).

77. Taunk, K., De, S., Verma, S. & Swetapadma, A. A brief review of nearest neighbor algorithm for learning and classification. *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019* 1255–1260 (2019) doi:10.1109/ICCS45141.2019.9065747.

78. Sander, J. Density-Based Clustering. *Encyclopedia of Machine Learning* 270–273 (2011) doi:10.1007/978-0-387-30164-8_211.

79. Jr., F. H. S. Rigorous Basis of the Frenkel-Band Theory of Association Equilibrium. *The Journal of Chemical Physics* **38**, 1486 (2004).

80. Kremer, K. & Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics* **92**, 5057 (1998).

81. Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* **322**, 549–560 (1905).

6 APPENDIX

The codes used in this study are available at https://github.com/ZaferKosar/FD.git .