BROWNIAN DYNAMICS STUDIES OF DNA INTERNAL MOTIONS

A Thesis
Presented to
The Academic Faculty

By

Benson Jer-Tsung Ma

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Computational Science and Engineering

Georgia Institute of Technology

December 2018

BROWNIAN DYNAMICS STUDIES OF DNA INTERNAL MOTIONS

Approved by:

Dr. Edmond Chow, Advisor
School of Computational Science and Engineering
Georgia Institute of Technology

Dr. Jeffrey Skolnick
School of Biological Sciences
Georgia Institute of Technology

Dr. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: November 27, 2018

Research is what I'm doing when I don't know what I'm doing.

Wernher von Braun

To my parents

ACKNOWLEDGEMENTS

The work presented in this thesis was carried out at the School of Computational Science and Engineering at Georgia Institute of Technology in Edmond Chow's Group between January 2017 and November 2018, and was supported by the National Science Foundation (NSF) under grant No. ACI-1147843. I thank the Partnership for an Advanced Computing Environment (PACE) as well as the National Energy Research Scientific Computing Center (NERSC) for providing the computational resources necessary for me to complete this project.

I would like to take the opportunity to thank my supervisor Dr. Edmond Chow, for his expertise, valuable time and unwavering support throughout my Masters study and for allowing me the opportunity to even undertake a Masters project within his group despite full knowledge that I am a Distance Learning student with a full-time job. It has been stressful and challenging at times but nevertheless extremely rewarding. My appreciation is extended to the members of my thesis committee, Drs. Jeffrey Skolnick and David Sherrill, who have taken the time out of their busy schedules to review my work. Similarly, I would like to thank the staff in the School of Computational Science and Engineering at Georgia Tech for their crucial administrative support.

I would also like to thank my longtime academic friends Drs. Narutoshi Kamiya and Junichi Higo of the Graduate School of Simulation Studies at the University of Hyogo, and Dr. Gert-Jan Bekker of the Institute of Protein Research at Osaka University, for their informal advice and fruitful discussions throughout the project.

Finally, I wish to thank my family, friends, and co-workers. Although they have at best only a vague idea of what I've been doing, they have supported me throughout these years, and I would not have completed the Masters program without their constant encouragement.

I am grateful for everything throughout these two years, and am now ready to face my next challenge.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

SUMMARY

Earlier studies by Chow and Skolnick [1] suggest that the internal motions of bacterial DNA may be governed by strong forces arising from being crowded into the small space of the nucleoid, and that these internal motions affect the diffusion of intranuclear protein through the dense matrix of the nucleoid. These findings open new questions regarding the biological consequences of DNA internal motions, and the ability of internal motions to influence protein diffusion in response to different environment factors. The results of diffusion studies of DNA based on coarse-grained simulations are presented. Here, our goals are to investigate the internal motions of DNA with respect to external factors, namely salt concentration of the solvent and intranuclear protein size, and to understand the mechanisms by which proteins diffuse through the dense matrix of bacterial DNA. First, a novel coarse-grained model of the DNA chain was developed and shown to maintain the fractal property of in vivo DNA. Next, diffusion studies using this model were performed through Brownian dynamics simulations. Our results suggest that DNA internal motions may be substantially affected by ion concentrations near physiological ion concentration ranges, with the diffusion activity increasing to a limit with increases in ion concentration. Furthermore, it was found that, for a fixed protein volume fraction, the motions of proteins in a DNA-protein system are substantially affected by the size of the proteins, with the diffusion activity increasing to a limit with decreasing protein radii, but the internal motions of DNA within the same system do not appear to change with changes to protein sizes.

CHAPTER 1
INTRODUCTION

This thesis describes the Brownian dynamics studies of protein and DNA diffusive motions in the nucleoid. The report extensively details the background and motivations for this work, as well as in-depth technical details of the experimental methods, and discusses the results in the greater research context of biological processes occurring in the nucleoid such as protein diffusion through the DNA matrix.

1.1 Motivation & Goals

One of the defining characteristics of the nucleoid is the high concentration of DNA that is packed within. The typical *E. coli* nucleoid space has a DNA volume fraction of 10–20% [2], leaving the average spacing between strands to be much less than the effective hydrodynamic diameters of many large DNA binding proteins (DBPs). This observation consequently brings up important questions regarding the role of DNA structure and dynamics in facilitating key biological functions that involve the DNA, such as target search by proteins, transcription, and replication, for it would seem difficult to impossible for some of these functions to be carried out at all in such a tightly packed environment. For example, how are DBPs able to diffuse through the dense matrix of DNA to bind to their target sites? The theory of facilitated diffusion was developed to reconcile the experimental observation that association rate of experimental DBPs can be two orders of magnitude faster than the rate allowed by the three-dimensional (3D) Smoluchowski diffusion limit [3], and hypothesizes three modes for target search that DBPs can perform [4, 5], but it is unclear how DBPs under these modes can successfully overcome the cage-like environment imposed by the packed DNA. Another question is, how can DNA replication and strand separation

successfully take place without a high likelihood of strand entanglement due to the packed nature of DNA? Questions such as these have yet to be addressed, despite the extensive work done in the field to understand the biological processes that can be impacted by DNA crowding. A brief overview of aspects of DNA structure that are relevant to understanding the problem space behind this work is presented in Chapter 2.

There has been some interest in a bulk property of packed DNA known as DNA diffusive motions, or "internal motions," which are suspected of having biological significance and may be key to answering some of the aforementioned open questions. Specifically, work by Ando and Skolnick [6], as well as Chow and Skolnick [1], demonstrate that DNA diffusive motions can accelerate the rates of 1D and 3D protein diffusion along the DNA strand and through the dense matrix of the nucleoid, respectively. These findings open new questions regarding the nature of DNA internal motions, and the ability of internal motions to influence other biological processes, such as strand separation after DNA replication, in response to different environment factors. However, to the best of our knowledge, there have been no further efforts since [1] to quantify DNA internal motions, and this is the focus of our work here. DNA diffusive motions may be a contributing mechanism through which these environment factors can indirectly regulate known biological processes involving the DNA, and in this work, we seek to understand whether or not salt concentration and protein size, two commonly studied parameters of protein diffusion, can have an effect on DNA internal motions, which have already been shown to consequently influence DBP diffusive motions.

1.2 Methodology

The Brownian dynamics simulation technique is a coarse-grained method in which explicit solvent molecules are replaced by an implicit solvent medium that exert a

stochastic force on the dispersed solute molecules, here referred to as "Brownian particles." The technique takes advantage of the fact that there is a large separation in time scales between the rapid motion of solvent molecules and the more sluggish motion of the large solutes, which enables for coarse-graining out the fast modes of the solvent. Brownian dynamics is used exclusively in this work so that the dynamics of bulk DNA can be studied over microsecond timescales, something that would have otherwise been computationally too expensive with finer-grained methods such as molecular dynamics. The theory behind the Brownian dynamics methodology is presented in Chapter 2.

1.3 Outline of Work

For this work, we first developed a novel coarse-grained model of the DNA chain that was shown to be able to maintain a fractal structure over long periods of time. We then performed coarse-grained Brownian dynamics simulations of a DNA-only and DNA-protein systems with different salt concentrations and different protein sizes subject to a constant protein volume fraction, respectively, and measured the diffusive motions of the DNA chain, which is modeled as a polymer chain of soft beads. Details of the DNA model construction and the implementation of the simulation are presented in-depth in Chapter 3, and the results of this work are summarized in Chapter 4. To summarize, results from the first study suggest that DNA internal motions may be substantially affected by ion concentrations near physiological ion concentration ranges, with the diffusion activity increasing to a limit with increases in ion concentration. On the other hand, results from the second study suggest that, for a fixed protein volume fraction, the diffusion activity of proteins in a DNA-protein system appear to increase to a limit with decreasing protein radii, but the internal motions of DNA within the same system do not appear to change with changes to protein sizes. The implications of these findings on our current understanding of

biological processes occurring in the nucleoid and suggestions for future directions of this research are discussed in Chapter 5.

CHAPTER 2
BACKGROUND & THEORY

2.1 DNA

DNA is a molecule with biological activities defined by its structure. This has been a well-established observation since the elucidation of the structure of DNA in 1953 by Watson and Crick [7]. In their landmark paper, Watson and Crick noted that the base pairing in the DNA's structure suggested a "possible copying mechanism for genetic material." Since then, many biological functions for DNA have been identified, but the detailed relationship between structure and function is still generally not well understood. A brief review of DNA structure is presented; see [8, 9] for a more thorough and general review of DNA.

2.1.1 DNA Conformations

The typical DNA molecule consists of two complementary polynucleotide chains that are multiply interwound, forming a double helix. DNA is polymorphic, with the prevailing conformation being B-DNA, which is a right-handed helix with a period of approximately 10.5 base pairs (bp) per turn under physiological conditions. The polymorphism of DNA was established as early as 1953 with the structure of B-DNA elucidated by Watson and Crick, followed soon after by the structure of A-DNA established by Franklin and Gosling from analyses of diffraction patterns obtained from DNA fibers at relatively lower hydration levels [10, 11]. Since then, other conformations of DNA have been discovered, such as the Z form [12], which is a left handed helix of approximately 12 bp per turn, the D form [13], and the C form [14], which is found in the presence of lithium. All of these DNA forms differ in characteristics

such as helical repeat, rise, and helix diameter (compiled and summarized in [15, 16]), and their structures are directly determined by the sequence, hydration levels, and cations in the surrounding solution. The B conformation accurately describes the overall structure of DNA *in vivo*; however, the biologically relevant forms of DNA include the A and Z conformations. A more detailed review of the different conformations of DNA and their functions can be found in [16].

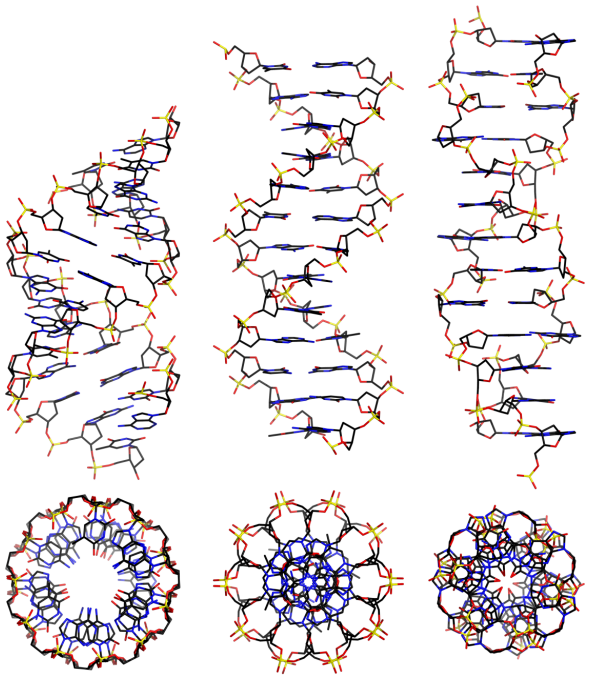


Figure 2.1: Side and top views of A (left), B (middle), and Z (right) conformations of DNA. Image from [17].

Biochemical, crystallographic, and computer simulation studies have indicated that the A-form of DNA plays many important biological functions. A-like conformations have been observed in the target site for the TATA-box binding protein (TBP), suggesting that A-DNA is an important intermediate step in forming a distorted structure required for binding by TBP [18, 19]. In addition, the compact nature of A-DNA relative to B-DNA has been shown to be necessary to reduce spatial distance between neighboring target sites to allow for proper binding, such as the *E. coli* cyclic AMP receptor protein (CAP) binding process [20]. Conformational

switches from the B-form to the A-form DNA at the DNA replication site by DNA polymerase [21] have been suggested to provide discrimination between correct and incorrect base pairing, thus increasing the reliability and fidelity of DNA replication [22]. Finally, it has been shown that nucleotide bases in A-DNA are an order of magnitude less susceptible to UV damage compared to those of B-DNA [23], which may explain its dominant presence in bacterial spores as protection against environmental factors [24].

Similar biological functions for Z-DNA have been observed, in particular its role in DNA transcription. Z-DNA is commonly associated with negative DNA supercoiling (discussed in Section 2.1.4) [25, 26], and while supercoiling is associated with both DNA transcription and replication, Z-DNA formation is primarily linked to the rate of transcription [27]. Analyses of genomic databases indicate that sequences with high affinity for forming Z-DNA are frequently found around transcription initiation sites [28], usually at approximately three helical turns after the promoter sequence [29]. This suggests that Z-DNA may influence transcription by modulating the supercoiling levels within a domain to provide torsional strain relief and thus lower the energy barrier for protein-DNA binding. In eukaryotes, nucleosome and Z-DNA formations are mutually exclusive [30], and it has been likewise hypothesized that, since nucleosomes interfere with DNA accessibility by proteins, [31], the Z conformation of DNA may play a role in regulating transcription, replication, and recombination [29, 32].

2.1.2 Polymer Model & Persistence Length

DNA is commonly modeled as an elastic rod with a persistence length l_p, bulk elasticity (Young's modulus) Y, torsional stiffness C, and twist-stretch coupling D [33], values for which have all been determined by single-molecule experiments [34–37]. We will focus on the persistence length here as it is most relevant to this study. In

polymer chemistry, the persistence length l_p is a basic mechanical property of a polymer that quantifies its stiffness, and is defined as the length over which correlations in the direction of the tangent are lost:

$$l_p = \frac{B_s}{k_B T} \tag{2.1}$$

where B_s is the flexural force constant (bending stiffness) of the polymer, k_B is Boltzmann's constant, and T is the temperature. B_s is defined as:

$$B_s = YI \tag{2.2}$$

where I is the moment of inertia of the polymer's cross section. In the case of a uniform rigid rod with radius a, I can be expressed as:

$$I = \frac{\pi a^4}{4} \tag{2.3}$$

The idea of the persistence length is that it defines the scale over which a polymer remains roughly unbent in solution like an elastic rod. At scales longer than l_p, thermal fluctuations result in spontaneous bending of the polymer, and the polymer can then only be described statistically as a worm-like chain (WLC). A WLC is a continuous version of the three-dimensional random walk model of a polymer, or Kratky-Porod model [38]. Related to the persistence length that is used in modeling polymers is the concept of the Kuhn length, which for WLCs is equivalent to twice the persistence length [39].

In an aqueous solution, the average persistence length of DNA is 46–50 nm, or 140–150 bp [40, 41], and the diameter of DNA is approximately 2.04 nm (effective hydrodynamic diameter is 2.2–2.6 nm) [42], although this can vary significantly depending on the conformation [43] and physicochemical properties of the environment

such as temperature [44]. The persistence length of a segment of DNA is dependent on its sequence, which can cause significant variation due to differences in base stacking energies and the residues which extend into the minor and major grooves of the DNA strand.

The relative flexibility of DNA has been observed to be important for gene regulation [45], as it has the potential to bring parts of genes into closer proximity with one another in three-dimensional space than would be possible if the DNA was perfectly linear. Furthermore, the entropic flexibility of DNA has been noted to be consistent with standard polymer physics models like the WLC model [46], allowing for computer studies of bulk DNA to be performed using coarse-grained models. However, for DNA segments shorter than the persistence length, the bending force is approximately constant, thus producing polymer behaviors that deviate from WLC model predictions.

2.1.3 *In Vivo* Dimensions

As this work is focused on studying the properties of prokaryotic DNA systems, we will focus on the *in vivo* dimensions of DNA for prokaryotes only. Prokaryotes generally possess one circular chromosome, which makes up most of the DNA of the cell. The DNA that comprises the prokaryotic chromosome is in a highly condensed conformation, known as chromatin, and occupies a defined region of the cell that is characterized by the absence of ribosomes. This chromatin-dense area forms a pseudo-compartment [47] that is analogous to the eukaryotic nucleus, and is generally referred to as the nucleoid. The nucleoid is generally observed to be less densely packed with proteins than the surrounding cytoplasm, presumably due to the internal crowding by the DNA.

The degree of DNA crowding in the nucleoid is estimated as follows. The typical *E. coli* K-12 cell has a nucleoid volume of 0.14 μm^3 [48], containing an average of 2.3

genomes, each $4,639,221$ base pairs long [49], or roughly 1.5 mm if expanded into perfectly linear form. The nucleoid volume is then assumed to be a cube with side lengths of about $\sqrt[3]{0.14} \approx 0.52$ μm, which is then uniformly stacked with the DNA that has been cut into segments of 0.52 μm. This approximation model gives an average spacing of about 6.4 nm between the centers of DNA strands [50, 51]. Taking the diameter of the DNA strand into account, this leaves a surface-to-surface spacing between DNA strands to be roughly 4.4 nm, which is less than the diameter of large transcription factors such as LacI.

The volume fraction of DNA in the nucleoid *in vivo* is not precisely known, with research literature quoting a range of 10–20% [2]. Furthermore, this volume fraction is subject to change during the life cycle of the cell [52].

The volume fraction of proteins in the nucleoid is also not precisely known, but can be roughly estimated as follows. The number of nucleoid proteins in an *E. coli* cell is estimated to be 50,000–200,000 [53, 54]; hereafter, the maximum of 200,000 proteins is assumed. Assuming that the proteins are modeled as perfect spheres with a radius of 2.19 nm, which is comparable to the hydrodynamic radius of proteins with similar weight as nucleoid-associated protein HU [6], then the total volume nucleoid protein volume is estimated to be $200,000 \cdot \frac{4}{3}\pi \cdot 2.19^3 = 879,9357.24$ nm^3, or 0.0087 μm^3. Given the nucleoid volume of 0.14 μm^3, this amounts to a protein volume fraction of approximately 6.28%.

2.1.4 Topological Organizations

Supercoiling

The characteristics and dimensions of DNA *in vivo* as described in Sections 2.1.2 and 2.1.3 necessitates extensive condensation and maintenance of a highly organized structure [55] so that entanglements are prevented during DNA replication. In fact, an approximately 103-fold compaction of the DNA is required in order for it to fit into

the nucleoid [56]. A fundamental structural aspect of *in vivo* prokaryotic DNA that makes this compaction possible is its supercoiling, namely negative supercoiling [57]. Observed in both plasmids and prokaryotic chromosomes, supercoiling is a property of topologically closed polymers (where the free rotation of the polymer ends is restrained), and has been hypothesized to support several critical biological functions, the most important of which is storing torsional tension energy for driving processes that require unwinding of DNA strands [58], such as transcription [59–61], replication [62] and recombination [63]. Other hypothesized functions of supercoiled DNA include bringing distant enhancer-promoter sequences to closer proximity to promote transcription [64], influencing the efficiency of target DNA site search by proteins [65], and inducing the formation of Z-DNA or other alternative conformations in regulatory regions to influence protein binding [16, 66, 67].

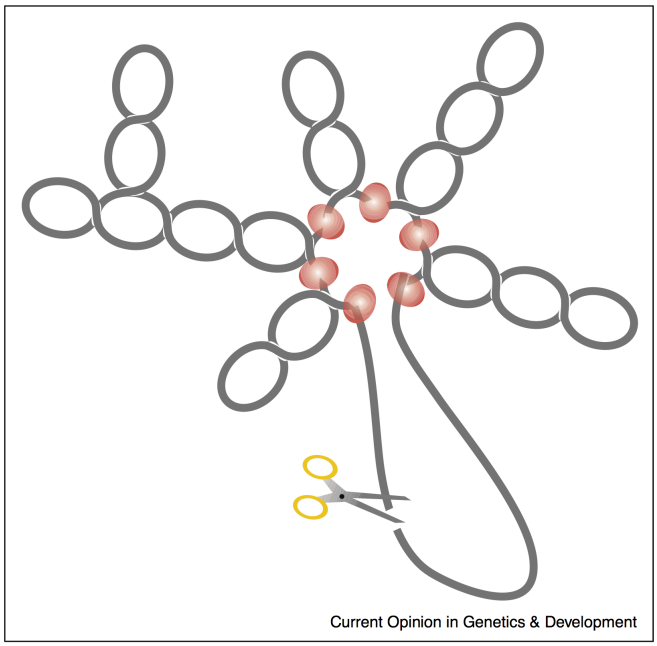


Figure 2.2: Topological organization of the bacterial chromosome. DNA nicking (indicated by scissors) relaxes only a single domain while the topological state of the remaining chromosome remains intact, due to the presence of supercoil diffusion barriers (red). Image from [68].

DNA supercoiling is a fragile process, and even a slight change in the overall superhelicity of chromosomal DNA is lethal [69–71]. However, studies have shown

that, unlike plasmids, prokaryotic chromosomes do not behave as a single topological unit but are composed of independently supercoiled domains, each with a topological state that is uncoupled from the rest of the chromosome by supercoil diffusion barriers (Figure 2.2) [72, 73]. These topological domains have a rosette-like appearance with topologically locked duplex loops (i.e. "plectonemic loops") emanating radially from a central core containing RNA and proteins [74, 75], and appears to be present across many prokaryote species [76]. As a consequence, a single loop could be relaxed by nicking without affecting the superhelicity of neighboring loops [77]. Furthermore, the establishment of supercoil domain boundaries appears to be essentially random throughout the entire nucleoid, with domain lengths exhibiting an exponential distribution with a mean length of approximately 10 kbp [60], and these boundaries can vary even within a single cell over time [78, 79]. More in-depth discussions of DNA supercoiling and supercoil-induced conformations of DNA can be found in [16, 68, 80–85].

Fractal Globules & Crumpling

Another structural aspect of *in vivo* DNA that makes DNA compaction possible is its crumpling into a fractal globule. A fractal globule, also known as a crumpled globule, describes a chain polymer state that has both compact local and global scaling [86]. This is in contrast to an equilibrium globule, which describes polymers in a densely knotted configuration and whose compaction exhibits poor scaling (Figure 2.3). Fractal globules emerge from a polymer condensing and collapsing upon itself iteratively over the whole chain as a result of topological constraints that prevent one region of the chain from crossing over to another [87]. First introduced in [88], this long-lived polymer state was proposed as a structural model of the human chromosome on the scale of up to 10 Mbp [89], based on experimental observations using chromosome conformational capture techniques [90]. Though the fractal globule model was first

proposed for eukaryotic DNA, it appears to be applicable to prokaryotic DNA as well. Computer and experimental studies have shown that long circular polymers tend to collapse from spread-out configurations into crumpled globules under high concentrations [87, 91], suggesting that bacterial plasmids might follow this behavior in the crowded environment of the cytoplasm. In addition, recent studies demonstrate that the spatial distance between any locus and the origin of replication in prokaryotic DNA correlates precisely linearly with the genomic distance between the two [92, 93]. Since this is scalar property exhibited by crumpled polymer conformations, the finding has lead to the proposal that prokaryotic DNA in nucleoid might also exhibit crumpling as well [94].

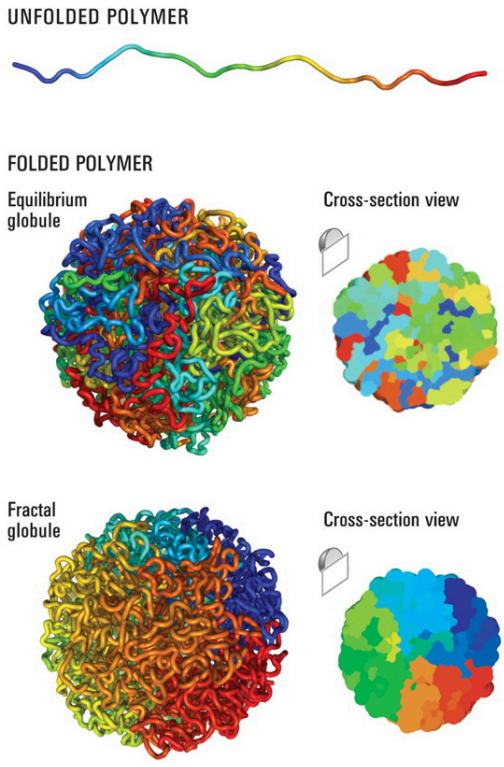


Figure 2.3: DNA loci clustering patterns in a fractal (left) vs. equilibrium (right) globule. Image from [89].

Because the dynamics of compact DNA opening up from an unentangled fractal conformation are very different from those of the knotted equilibrium conformation [94], the presence of fractal globules in the DNA chain has important implications for

its ability to promote biological activity. In fact, it has been observed that gene activation can cause rapid decondensations of large genomic regions (0.5–2.0 Mbp) [95], presumably to expose the DNA in those regions for DBPs to subsequently interact with in order to carry out transcription, replication, and other biological processes [96]. Conversely, for a DNA chain in an equilibrium globule conformation, no given region of the globule can fully open up as it would remain trapped by multiple local entanglements embedded in the structure. While the fractal globule, as opposed to equilibrium globule, is most consistent with experimental data, it is ultimately a long-lived intermediate that gradually converts into an equilibrium globule. Computer studies show that the loss of territorial organization in fractal globules is eventual, and can be either accelerated by the presence of active DNA re-modeling enzymes such as DNA topoisomerase II, or decelerated by the presence of crosslinking proteins that bind chromosomal loci [94], though it is not clear what the mechanisms that help maintain the DNA's fractal globule *in vivo* look like. Detailed reviews of the fractal globule model of DNA and its suitability as a model of chromatin organization can be found in [89, 94, 97, 98].

2.1.5 Salt Effects

DNA is highly negatively charged due to the presence of phosphate groups along the DNA backbone [7]. As such, salt (primarily Na^+ and Mg^{2+}) has been shown to strongly modulate DNA properties [99] and play a pivotal role in maintaining the stability of the DNA structure by regulating the charge neutrality and balance of the system. Here, stability is determined by the estimated melting temperature T_m of the molecule. According to the counter-ion theory proposed by Manning [100, 101], these cations act as shielding agents and effectively screen out the electrostatic repulsion exerted by the negatively charged backbone, allowing the complementary strands of DNA to stabilize into helix form, as well as the overall DNA strand to compact

onto itself. This proposal has been corroborated upon by multiple studies [102–106], which have shown that double-strand stability increases logarithmically with increasing cation concentration, regardless of strand length [107–110]. In addition, studies have shown that salt concentration can affect the transition of DNA from the B- to A-form [111, 112], as well as the stretching and unzipping behavior of the DNA strand [103, 113, 114], indicating that the mechanical stability of DNA also increases with the concentration of cations. It should be noted that all the aforementioned studies were carried under 0.1–1.0 M salt concentrations, and that there is a limited range of concentrations up to which cations contribute to the stability of the molecule, beyond which the stability of DNA will be shattered entirely [115, 116]. See [102, 117, 118] for further discussions on the effects of salt on DNA strand stability.

As DNA is a polyelectrolyte, the parameters that describe DNA as a polymer (see Section 2.1.2) are expected to be strongly dependent on salt conditions. In particular, theoretical studies predict that the electrostatic repulsion experienced by the backbone increases the persistence length l_p [119, 120]. This has been verified by single molecule experiments, which observed a reduction of l_p by approximately 20% when the salt concentration of the surrounding environment was increased from 20 mM to 300 mM (monovalent salt) [121, 122], though the exact nature and mechanism of this dependence is not well understood. Salt concentration also appears to strongly affect aspects of DNA supercoiling [123, 124]. At low salt concentrations, DNA supercoils appear to be more loosely interwound with the chain undergoing large scale opening and closing as well as rapid slithering, while at high salt concentrations, the bending and torsional components dominate over electrostatic terms as a result of charge screening, and the DNA supercoils appear highly compact and rigid [125]. More importantly, the same study revealed a critical point associated with the collapse of supercoils from the loose to tight conformations near the physiological salt concentration of 0.1 M, which suggests a potential regulatory role for salt on biological

functions associated with supercoiled DNA [125]. Other aspects that have been found to be dependent on salt concentration include the effective supercoil diameter [126–129], the critical linking number for buckling [36, 130], the number of plectonemes [131, 132], and torsional stiffness, though it is debated whether torsional stiffness is decreased [126, 133], increased [134], or not affected [135] by the presence of salt. It should be noted that the presence of salt also affects DNA supercoil dynamics [84], with dynamic light scattering measurements showing that the internal motions of superhelical DNA increase with salt concentration [136]. Finally, it appears that the dependence of supercoiling on salt is not just a function of concentration, but also the type and valence of the cations in solution [137, 138], with Na^+ showing a higher potential than K^+ in driving chromatin compaction for instance [139]. More reviews on the effect of salt concentration on supercoiling can be found in [84, 125, 135, 136, 140].

2.1.6 Crowding Effects

The typical *E. coli* cytosol contains roughly 300–400 mg/mL of macromolecules [141], resulting in a volume fraction of 20–40% [141–144]. Such an environment is called "crowded" as opposed to "concentrated," since no single macromolecular species occurs at high concentration, but together, the macromolecules reduce the volume of solvent available for other molecules in the solution. The crowding effect is dependent on shape of the macromolecule involved and is non-linearly dependent on the molecular mass [145]. A variety of effects have been observed that are a result of macromolecular crowding, such as different reaction rates experienced by the biomolecules in solution [142, 146], increased propensity of macromolecular association and assembly [147], and three- to tenfold reduction of diffusion coefficients of the constituent biomolecules compared to their values in water under infinite dilution [148–150].

Although the volume fraction of DNA in the nucleoid is lower than the 20–40%

volume fraction of macromolecules in the cytosol, macromolecular crowding plays equally important biological roles in the nucleoid. Crowding in the nucleoid is dominated by DNA binding proteins (DBPs) such as HU, which can cover up to 30% of genomic prokaryotic DNA [151], leaving around 80% of the genome to be accessible [53, 152]. Under crowding, the activity of water is decreased and hydration of DNA becomes unfavorable, and this has been shown to stabilize DNA and RNA duplexes. In particular, duplexes can be destabilized by small co-solutes and stabilized by large co-solutes [153] and stabilization is dependent on solute concentration [154] as well as DNA length [155]. On the secondary-structure level, crowding appears to favor the formation of non-canonical DNA structures, such as the triplex [156] and G-quadruplex [157] forms, over the canonical *B*-DNA form. It should be noted that theses studies were performed in biologically unfavorable salt concentration and temperature conditions, leading to the hypothesis that co-solute crowding can compensate for environment conditions unfavorable to non-canonical DNA structures by providing a mechanism for adaptation and buffering against it [158, 159]. Furthermore, crowding has been shown to increase the binding stability of proteins to DNA by effectively compacting the DNA-protein complex [160], thus promoting enzymatic activities such as transcription [161]. Similar to the case with non-canonical DNA structure formation, the sensitivity of DNA binding proteins, such as DNA polymerase, to salt can be reduced by the macromolecule buffering [162]. Suffice to say, crowding is also an important factor in controlling the conformation of supercoiled DNA [163]. More discussions on the effects of macromolecular crowding on DNA can be found in [142, 164].

2.1.7 Facilitated Diffusion

Critical cellular processes involving the DNA such as transcription and replication rely on recognition of the target sites located on the DNA and formation of complexes with

DNA by DBPs. But how are DBPs able to rapidly and efficiently find their target sites on the DNA given the dimensions of the DNA and nucleoid as described? This question was raised after the seminal work by Riggs et al., which observed *in vitro* that the lactose repressor was able to diffuse and locate its operator region around two orders of magnitude faster than the rate allowed by the three-dimensional (3D) Smoluchowski diffusion limit [3] (diffusion is discussed in Section 2.2). An additional diffusion mechanism was suggested to resolve this discrepancy, known as reaction rate enhancement by dimensional reduction, which is based on an earlier proposal by Adam and Delbrück [165]. The modified diffusion model, known as "facilitated diffusion," has since been expanded upon by Berg, Winter, and von Hippel [166], and describes three modes for target search that DBPs can perform: 1) 1D rotation-coupled sliding along the DNA strand through the helical major groove without dissociation, 2) 1D hopping along the DNA strand to nearby locations via a series of microscopic dissociation and association events, and 3) 3D diffusion in between the 1D diffusion events that effectively acts as large inter-segmental jumps along the DNA [4, 5]. Target search under this proposed model could then be faster than either 1D or 3D search by itself, since 1D diffusion would accelerate 3D search by exploring multiple sites in tandem along the DNA once the DBP is non-specifically bound, and 3D diffusion would optimize 1D search by cutting down revisits to the same sites.

Each of the three modes of diffusion has been confirmed experimentally *in vitro* [167–169] and *in vivo* [170, 171] using spectroscopic [172] and single molecule-based [173] techniques. Experiments at low salt concentrations [174] demonstrate that the 1D sliding of DBPs along the DNA is enabled by the electrostatic interaction between the positively charged protein subunits common in DBPs [175] and the negatively charged phosphate groups along the DNA backbone. The resulting affinity strongly depends on the salt concentration [125, 176, 177], and is typically several orders of magnitude lower than that of DBPs for their designated DNA target sites [178]. While

increasing the salt concentration will increase the number of hopping events due to its effective screening of DBP-backbone affinities, it does not appear to affect the 1D diffusion coefficient of the DBP itself [179]. Other environmental factors that have been observed to affect the 1D diffusion of DBPs include DNA conformation [180, 181], supercoiling [182, 183], DBP nonspecific binding energies [184–186], crowding [187, 188], DNA self-crowding [189], and hydrodynamic effects [190]. More reviews on facilitated diffusion can be found in [94, 166, 191–193].

2.1.8 Open Questions with the Facilitated Diffusion Model

Despite the intense work in this area to understand the mechanistic details and other determinants that might affect DNA target search, a holistic understanding of the process is yet to be achieved and many basic questions remain open, two of which we will cover here that are closely relevant to this work. The first question is the role of macromolecule crowders in affecting target search. Because the nucleoid environment is densely crowded, it is expected to hinder DBP diffusion and target search. However, experiments suggest that on the contrary, crowding may even facilitate the target search process and enhance their enzymatic activity [194, 195], thus hinting that the physicochemical side effects of crowding should be taken into account [196]. Colloid physics dictates that crowders increase the macroscopic viscosity of the solution while building a depletion layer between the crowders that effectively maintains a constant microscopic viscosity [176], but exactly how this dual-environment model affects target search is unclear, with conflicting reports confirming that it either promotes [194] or has no effect [197] on facilitated diffusion. Another point of contention is the mechanism by which crowding affects 1D search, with some studies proposing that interactions between the DBP and crowders promotes hopping, thus increasing the overall rate of diffusion [198], and others observing that crowding prevents escapes from 1D sliding to 3D diffusion, thus promoting 1D sliding [199, 200]. Finally, there

are other nonspecifically-bound DBPs diffusing independently along the DNA strand [5], and it is not well understood as to how DBPs overcome such obstacles during 1D sliding, though at least one study suggests that contact with bound crowders may even enhance target search [151]. More comprehensive discussions of the role of macromolecular crowding on target search can be found in [199, 201].

The second question is the efficacy of the facilitated model in explaining the efficiency of DNA target search. 1D sliding has been observed to be much slower compared to 3D diffusion [170], and so 1D hopping and 3D diffusion are generally considered to contribute more significantly to DNA target search [202]. However, it has been observed that DBPs spend a large majority of their time in 1D search compared to 3D search [167, 203]. For example, the lactose repressor LacI spends 90% of the time nonspecifically bound and diffusing in 1D along the DNA at a rate 65 times smaller than that for 3D diffusion [170]. Thus, it is unclear as to how a dominant slower mode of diffusion can boost overall target search kinetics to support the facilitated diffusion model [201], and it would appear that the inclusion of 1D sliding in the target search process is advantageous only in a dilute DNA solution environment, where the DNA is not multiply folded onto itself [1]. The optimal combination of 1D versus 3D diffusion is also unclear, with theoretical calculations showing that the optimal target search rate is achieved when DBPs spend an approximately equal amount of time in 3D diffusion and 1D diffusion [4], while computer studies suggest that the optimal ratio of time spent for 3D diffusion together with 1D hopping versus 1D diffusion is closer to an 80:20 ratio [179]. Reviews on the role of 1D sliding in the facilitated diffusion model can be found in [5, 192, 204, 205].

2.1.9 Previous Work and Motivations for this Study

The observation of a nucleoid that is highly concentrated with DNA as well as proteins consequently brings up important questions regarding the role of DNA structure

and dynamics in facilitating key biological functions that involve the DNA, such as target search by proteins, transcription, and replication, for it would seem difficult to impossible for some of these functions to be carried out at all in such a tightly packed environment. For example, how are DBPs able to diffuse through the dense matrix of DNA to bind to their target sites? As reviewed in Section 2.1.7, the theory of facilitated diffusion was developed to address this question, but it is unclear how DBPs diffusing under this model successfully overcome the cage-like environment imposed by the packed DNA. Another question that comes up is, how can DNA replication and strand separation successfully take place without a high likelihood of strand entanglement due to the packed nature of DNA? Questions such as these have yet to be addressed, despite the extensive work done in the field to understand the biological processes that can be impacted by DNA self-crowding.

There has been some interest in a bulk property of packed DNA known as DNA diffusive motions, or "internal motions," which are suspected of having biological significance and may be key to answering some of the aforementioned open questions. DNA internal motions were first proposed under the term "segmental diffusion" and studied analytically using the WLC model [206], where it was found that the association rates for a large protein to specific DNA sites can be dominated by DNA internal motions; more recent discussions can be found in [183]. Since then, some theoretical studies have attempted to incorporate this effect into the current understanding of DNA structure and function, such as the DNA target search problem for example. Specifically, Ando and Skolnick [190] carried out Brownian dynamics (BD) simulations of a flexible DNA chain with restrained endpoints to investigate the effects of hydrodynamic interactions on DBP 1D sliding, and demonstrated that DNA motions could at least double the 1D diffusion of DBPs along DNA. Subsequent work by Chow and Skolnick [1] demonstrated, using BD simulations of bulk DNA, that DNA diffusive motions can be substantial in a tightly packed environment and can increase

DBP diffusion when the DBPs are nonspecifically bound. The proposed mechanism for this phenomenon, which builds upon the facilitated diffusion model, is that as the DBP is performing a 1D search, the environment around the strand of DNA to which the DBP is nonspecifically bound changes due to the mobility of the strand. This will bring strands that were originally far from the DBP to within diffusion-reachable distance, in addition to effectively pushing DBPs around with a dynamic mesh. When the DBP unbinds, it can rebind to these new strands that would have been much less accessible to the freely diffusing DBP if the DNA strands were immobile. In this way, the dynamic structure of packed DNA, through its internal motions, accelerates protein target search by expanding the effective diffusion radius of a DBP.

The work on understanding DNA internal motions is still in a primordial state. To the best of our knowledge, there have been no recent efforts to quantify DNA diffusive motions with respect to environmental factors, and this is the focus of our study here. As reviewed in previous sections, many environmental factors, such as salt concentration, crowding, and temperature, have been shown in both *in vitro* and *in vivo* studies to influence biological processes, such as the association of DBPs for their target binding sites or the efficiency of DNA replication; however, the mechanisms for their influence and the role of DNA internal motions in all of this are often not considered. Moreover, most of these studies assume that the DNA strands are immobile, and very few studies have taken DNA diffusive motions into account. We believe that DNA diffusive motions may be a contributing mechanism through which these known factors can indirectly regulate many biological processes. Specifically for this work, we seek to understand whether or not salt concentration and protein size, two commonly studied parameters of protein diffusion, can have an effect on DNA internal motions, which have already been shown to consequently influence DBP diffusive motions.

In contrast to environmental factors, such as salt concentration, which are

carefully adjusted to reflect physiological conditions, most studies of DNA and biomolecules in general have been performed under dilute, as opposed to crowded, conditions [142]. For example, most facilitated diffusion studies use single, relatively short (on the order of a few thousand base pairs), straight segments of DNA for investigation (see references in Section 2.1.7). However, computer studies have suggested that the effects of crowding likely dominate in vivo macromolecular motion [6], hinting that the results of these studies may need to be revisited. The study by Chow and Skolnick [1] was one of the first to explicitly model the crowded conditions of the nucleoid; however, the generated model did not match the fractal property of DNA, which has been suggested to play a role in determining the target search efficacy of DBP 3D diffusion [65, 207]. Thus, a secondary goal of this study is to produce a coarse-grained DNA model that, in addition to reproducing the crowded environment of the nucleoid suitable for studying DNA internal motions, is able to maintain a fractal structure over long periods of time. Similar to [1], since this study is focused on understanding DNA strand dynamics in a packed DNA globule, the employed DNA and protein models are necessarily much coarser than other models previously used to study DNA and DBP diffusion where nucleoid packing was not considered.

2.2 Diffusion

Diffusion is the process by which particles redistribute themselves from regions of high chemical potential or particle concentration to regions of low chemical potential or particle concentration as a result of random motion. Diffusion as a generalized phenomenon is driven by one or more gradients of the diffusing species, though throughout this work, the gradient generally refers to the chemical gradient of the system. The distinguishing features of diffusion are that it depends on particles moving around in random walk, known as Brownian motion (see Section 2.3.1), and that un-

like bulk motion, or mass flow, it is dependent only on concentration gradients within a medium rather than pressure gradients of the medium itself. A brief overview of aspects of diffusion relevant to this work is presented; for a more detailed overview, see [208–210].

2.2.1 Stokes' Law

Stokes' law describes the frictional force, also called the drag force, that is exerted on a stationary spherical particle held in a viscous fluid moving with a steady velocity [211, 212], and is given by:

$$F_d = 6\pi\eta r v_d \tag{2.4}$$

where F_d is the frictional force (Stokes' drag) acting on the fluid-particle interface, η is the dynamic viscosity of the fluid, r is the radius of the spherical object (i.e. the Stokes radius), and v_d is the flow velocity of the fluid relative to the object. The dynamic, or shear, viscosity of a fluid expresses its resistance to forming shear flows, a phenomenon in which adjacent layers of the fluid move in parallel but at different speeds. By translation, Eq. 2.4 is also applied to a spherical particle moving with steady velocity v_d (also known as the terminal drift velocity) in an otherwise stagnant fluid.

The Reynolds number is the ratio of inertial forces to viscous forces within a fluid and is used to describe flow patterns in different fluid flow situations [213]. Flows at low Reynolds numbers tend to be laminar (sheet-like), while flows at high Reynolds numbers tend to be turbulent due to differences in the fluid's speed and direction. Stokes's law is derived by solving the Navier-Stokes equations in the limit where the Reynolds number is small (Re $\ll 1$), which is also known as the Stokes flow limit. The derivation details for Stokes' law and discussion of the Navier-Stokes equations are well beyond the scope of this work; see [214, 215] for reference.

In biophysics, Stokes' law is used to define the so-called effective Stokes radius for molecules. This is subtly different from the effective radius of a hydrated molecule in solution, as the Stokes radius of a molecule is equivalent to the radius of a hard sphere that diffuses at the same rate as the molecule. Since most molecules are not perfectly spherical, the Stokes radius is generally smaller than the effective, or rotational, radius of the molecule. Similarly, an extended molecule will have a larger Stokes' radius compared to a compact molecule with the same molecular weight. The Stokes radius is also referred to as the hydrodynamic radius of a molecule, but only in biophysical contexts; the term has a different meaning in polymer chemistry.

Rearranging Eq. 2.4 gives the mobility μ of the spherical particle, which is the ratio of the particle's terminal drift velocity to an applied force, or the drag force in this case:

$$\begin{aligned} \mu &= \frac{v_d}{F_d} \\ &= (6\pi\eta r)^{-1} \end{aligned} \tag{2.5}$$

In the limit of low Reynolds number, μ is the inverse of the drag coefficient γ of the particle in the moving fluid.

2.2.2 Einstein-Smoluchowski Diffusion

In statistical mechanics, the mean squared displacement (MSD) is a measure of the deviation of the position of a particle with respect to a reference position over the observation time frame t. Physically, it can be thought of as a measure of the amount of space "explored" by a particle moving around in the system. It is the most common measure of the spatial extent of random motion, also known as Brownian motion (discussed in depth in Sec 2.3.1), and major work on determining the MSD for a Brownian particle was done by Einstein [216], who derived it from Fick's laws of

diffusion [217] down to the following:

$$\langle (x_t - x_0)^2 \rangle = 2dDt \tag{2.6}$$

where x_0 is the reference, or starting, position of the particle, x_t is the position of the particle after time t relative to the reference position, d is the dimension of the space traveled by the particle and D is the particle's diffusion coefficient. Eq. 2.6 is obtained under the assumption that the mean particle displacement is zero in any direction (as is for Brownian motion), which holds true if particle inertia is neglected and $t \gg 0$. This type of diffusion is called Einstein-Smoluchowski, or normal, diffusion, and can be fitted into a generalized power law:

$$\langle (x_t - x_0)^2 \rangle \propto t^{\alpha} \tag{2.7}$$

where $\alpha = 1$.

When the relationship in Eq. 2.7 is non-linear, the diffusion process is called anomalous diffusion. If $\alpha > 1$, the phenomenon is called super-diffusion. An example of this is an active cellular transport process. If $\alpha < 1$, the phenomenon is called sub-diffusion, or dispersion. This usually occurs when a system is not yet equilibrated. In biophysical studies, the diffusion coefficient is measured by applying Eq. 2.6 to recorded observations of the MSD. However, due to the presence of anomalous diffusion at the beginning of an experiment, the diffusion coefficient is in practice estimated only from MSD measurements in the range where t is sufficiently large. See [218] for more discussions on anomalous diffusion.

In the process of deriving Eq. 2.6, Einstein also derived an expression for the diffusion coefficient, which was independently arrived at by Smoluchowski [219]:

$$D = \mu k_B T \tag{2.8}$$

where μ is the mobility of the particle. This is known as the Einstein-Smoluchowski Relation. Combining Eq. 2.5 into this equation consequently produces the following equation:

$$D = \frac{k_B T}{6\pi\eta r} \tag{2.9}$$

which is known as the Stokes-Einstein Relation. This equation relates the diffusion of a perfectly spherical Brownian particle directly to its radius and the viscosity of the surrounding fluid, and is commonly used to compute the input parameter D for computer-based studies of diffusion when D has not yet been experimentally determined (discussed in 2.3.6). Likewise, the Stokes-Einstein relation is used to compute the Stokes radius of a molecule given its experimentally-determined diffusion coefficient. For in-depth discussions of Einstein-Smoluchowski diffusion, see [220–222].

2.2.3 Calculating the Mean Squared Displacement

The MSD is measured over increasing time intervals and averaged over the number of particles to study the diffusive properties of the system, and is defined as a function of the observed system trajectory x:

$$MSD(t) = \langle (x_t - x_0)^2 \rangle = \frac{1}{N} \sum_{n=1}^{N} \Big(x_n(t) - x_n(0) \Big)^2 \tag{2.10}$$

In practice, both experiments and computer simulations may not be able to sample a long-enough time span to gather enough data points for computing the MSD for large values of t due to its high computational cost. However, this problem can be worked around by invoking the so-called ergodic hypothesis.

A random process is said to be ergodic if its statistical properties can be deduced from a single, sufficiently long, random sample of the process [223]. Ergodic systems exhibit the same behaviors averaged over time as averaged over the space of all the

system's microstates in its phase space, i.e., the ensemble average is equal to the time-average during the evolution of a system's trajectory. The ensemble average is defined as the mean of a quantity that is a function of the microstate of a system. In physics, ergodicity implies that a system satisfies the ergodic hypothesis of thermodynamics. This allows for steady-state properties of the system, which may be difficult to compute over the phase space, to be easily estimated instead by averaging the values collected from the system's trajectory over time.

Because the time-averaged properties of the system undergoing Brownian motion can be assumed to be the steady state properties according to the ergodic hypothesis [224–226], more data points for the squared displacement over a time frame t can be collected simply by shifting the observation window to obtain measurements under different reference positions for the particle. Assuming a discrete-time system trajectory x that has been measured up to time T, the formula for calculation of the MSD can be extended as follows:

$$MSD(t) = \langle (x_{\tau+t} - x_\tau)^2 \rangle = \frac{1}{N(T-t)} \sum_{\tau=0}^{T-t} \sum_{n=1}^{N} \Big(x_n(\tau + t) - x_n(\tau) \Big)^2 \tag{2.11}$$

More discussions on techniques for MSD calculations can be found in [227, 228].

2.3 The Brownian Dynamics Method

The Brownian dynamics (BD) simulation technique is a coarse-grained method in which explicit solvent molecules are replaced by an implicit solvent medium that exert a stochastic force on the dispersed solute molecules, here referred to as "Brownian particles." The technique takes advantage of the fact that there is a large separation in time scales between the rapid motion of solvent molecules and the more sluggish motion of the large solutes. The ability to coarse-grain out these fast modes of the solvent

allows one to simulate much larger time scales than those allowed by finer-grained methods such as molecular dynamics simulation. At the core of a BD simulation is a stochastic differential equation (SDE) that is numerically integrated forward in time to create a trajectory of the system, which consequently allows for the study of the temporal evolution and dynamics of large solutes in a complex medium. BD simulations are particularly well suited for studying the structure and flow dynamics of complex fluids and colloidal suspensions in non-equilibrium situations.

2.3.1 Brownian Motion

Brownian motion is the random motion of a small particle suspended in a fluid (a liquid or gas) that results from its collision with the fast-moving molecules in the fluid. While early investigations of this phenomenon were made on pollen grains, dust particles, and various other objects of colloidal size, the theory has also been applied and extended to situations where the "Brownian particle" is not a real particle at all, but instead some collective property of a macroscopic system. For example, in a Brownian-system description of the concentration of a component of a chemically reacting system near thermal equilibrium, the "Brownian motions" are defined to be the irregular fluctuations of this concentration over time.

Standard Brownian motion processes are continuous-time stochastic processes, and are also referred to as Wiener processes in probability theory [229]. A Wiener process $[W(t); t \geq 0]$ is defined by the following properties:

1. $W(0) = 0$.
2. $\langle W(t) \rangle = 0$ for $t \geq 0$.
3. For every $t > 0$, the future increments $W(t + \tau) - W(t)$, where $\tau \geq 0$, are independent of the past values $W(s)$, where $s < \tau$. In other words, $W(t)$ has stationary independent increments.
4. $W(t+\tau) - W(t)$ is normally distributed with mean 0 and variance proportional

to τ, i.e. $W(t+\tau) - W(t) \sim \mathcal{N}(0, \tau)$.

5. W is continuous everywhere, but differentiable nowhere.

A Wiener process can be constructed as the scaling limit of a random walk, i.e. the limit of a random walk on a lattice grid as the grid spacing becomes infinitesimally small. In fact, Wiener processes can be constructed from any discrete-time stochastic processes with stationary independent increments. This is known as Donsker's Theorem [230], and is a functional extension of the Central Limit Theorem [231], which asserts that the sum of independent random variables tends toward a normal distribution even if the original variables themselves are not normally distributed.

2.3.2 Equations of Motion

The description of classical dynamics can be generalized with the Lagrangian formulation approach. The equations of motion are expressed using generalized coordinates $\mathbf{q}(t)$ that span a high dimensional space. The Lagrangian of the system is defined as:

$$\mathcal{L}(\dot{\mathbf{q}}(t), \mathbf{q}(t), t) = K - U \tag{2.12}$$

where K and U are the kinetic and potential energies of the system, respectively, and $\dot{\mathbf{q}}(t)$ is the derivative of $\mathbf{q}$ with respect to time. While K and U can explicitly depend on $\mathbf{q}$, $\dot{\mathbf{q}}$, and t, a Cartesian coordinate (Newtonian) system is assumed here, where $\mathbf{q} = \mathbf{x}$, U is purely a function of $\mathbf{q}$, and K is purely a function of $\dot{\mathbf{q}}$ as given by:

$$K(\dot{x}) = \sum_i \frac{1}{2} m_i (\dot{x}_i)^2 \tag{2.13}$$

From the Lagrangian, an equation of motion can be derived using the Euler-Lagrange equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \frac{\partial \mathcal{L}}{\partial q_i} = 0, \qquad i = 1, ..., N \tag{2.14}$$

The Euler-Lagrange equation results from what is known as an action principle, and its derivation is a topic in the field of calculus of variations, which will not be covered here; see [232, 233] for proof. For a Newtonian system, the following equations hold by the chain rule and substitution of the definitions of K and U:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \dot{q}_i} &= \frac{\partial}{\partial \dot{x}_i} K(\dot{x}_i) - \frac{\partial}{\partial \dot{x}_i} U(x_i) \\ &= m_i \ddot{x}_i - 0\end{aligned} \tag{2.15}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial q_i} &= \frac{\partial}{\partial x_i} K(\dot{x}_i) - \frac{\partial}{\partial x_i} U(x_i) \\ &= 0 + f_i\end{aligned} \tag{2.16}$$

Combining Eqs. 2.15 and 2.16 into 2.14 reduces to the familiar Newton's equation of motion:

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = f_{total}(t) \tag{2.17}$$

Other descriptions of classical dynamics exist, most notably the Hamiltonian formulation. However, the Lagrangian formulation is the most useful for solving the equations of motion. Rigorous treatments of the subject can be found in [234, 235].

2.3.3 Thermal Equilibrium

In thermodynamics, a particle in one dimension is said to be at thermal equilibrium, i.e. does not experience acceleration, when its mean squared velocity is $\langle v^2 \rangle = k_B T/m$, where k_B is Boltzmann's constant, T is temperature, and m is the mass of the particle. This can be derived using the definition of kinetic energy from classical mechanics and the Equipartition Theorem from thermodynamics [236, 237], which states that the average kinetic energy associated with each degree of freedom

is $k_BT/2$:

$$\frac{1}{2}m\bar{v}^2 = \frac{1}{2}k_BT \tag{2.18}$$

2.3.4 Langevin Dynamics

Consider the one-dimensional motion of a colloidal-sized spherical particle of mass m in explicit solvent. Newton's equation of motion for the particle is given by Eq. 2.17, and can be numerically solved by molecular dynamics (MD). Interactions between the particles (atoms in this case) are given by pairwise potentials, with bonded interactions like angles and dihedrals being approximated by many-body potentials. These potential terms, collectively called the force field, are rough approximations of solutions given by the Schrödinger equation at best, and their quality is the main determinant of the accuracy of an MD simulation. However, to determine a force field, $f_{total}(t)$ needs to be defined, and it is usually not practical or even desirable to look for an exact expression for $f_{total}(t)$. Moreover, MD can be an unnecessarily expensive method for studying this type of system, because the target of interest is the solute molecule, which is generally orders of magnitude larger than the surrounding solvent molecules, and so the characteristic times of motion of the two are considerably different.

Consequently, a one-dimensional system consisting of a Brownian particle in implicit solvent is considered instead. In this model, $f_{total}(t)$ is mostly dominated by a frictional force $-\gamma v$ that is exerted by the solvent medium (i.e. collision frequency of the particle against the solvent) and is proportional to the velocity of the Brownian particle. γ in this case is the drag coefficient of the spherical particle in the moving fluid, which is the inverse of the mobility as defined by Stokes' law (Eq. 2.5) and can

be alternatively defined as:

$$\gamma = m\xi \tag{2.19}$$

where ξ is the friction constant and has units of inverse time (s^{-1}). If this was the full story, then the equation of motion for the Brownian particle becomes simply:

$$m\frac{\mathrm{d}v}{\mathrm{d}t} \cong -\gamma v \tag{2.20}$$

Since this is a linear first-order differential equation, it has the solution:

$$v(t) = v(0)e^{-\gamma t/m} \tag{2.21}$$

Eq. 2.21 implies that the velocity of the Brownian particle will decay to zero after a very long time. But that cannot be true, since Eq. 2.18 holds for a system in thermal equilibrium, and so the assumption that $f_{total}(t)$ is dominated by the frictional force needs to be modified. As suggested by the observed randomness of a Brownian trajectory, the appropriate modification is to add a "random," or fluctuating, force term into the equation of motion. This is the basis of Langevin dynamics.

A Langevin dynamics system arises from a classical system by removing degrees of freedom. The degrees of freedom which are removed exert conservative forces on the rest of the system. The frictional force, which is proportional to the velocity of the particle, is added to the conservative force and removes kinetic energy from the system. All other forces are assumed to add up to a random force (thermal noise), which adds kinetic energy back into the system. These frictional and random forces are caused by collisions of solvent molecules with the colloidal particle. The formal equation of motion for Langevin dynamics for a particle in a one-dimensional system

is:

$$m\ \mathrm{d}v = f(r)\ \mathrm{d}t - m\xi v\ \mathrm{d}t + \mathrm{d}W \tag{2.22}$$

where f is the force exerted by an external field and W is a Wiener process. The external field is usually empirically designed or chosen from a set of existing designs based on the system under study, and contains body forces such as sterics and electrostatics. The term $\mathrm{d}W$ is an additive noise term, i.e. it is independent of the state of the system. This is in contrast to multiplicative noise, which is state-dependent. Though it is possible to treat Langevin dynamics with memory in the noise term [238], we will not be doing so here.

Langevin equations are usually written in differential form, which can be obtained by dividing both sides of Eq. 2.22 by $\mathrm{d}t$. Note that this is formally incorrect, since $W(t)$ is not differentiable. The differential form of the Langevin equation in Cartesian coordinates, generalized for a system of N particles, is:

$$M\ddot{\mathbf{X}} = -\nabla U(\mathbf{X}) - M\Xi\dot{\mathbf{X}} + \mathbf{R} \tag{2.23}$$

where $\mathbf{X}$ is the positional trajectory of the system, $\mathbf{M}$ is the (diagonal) mass matrix, Ξ is a $3N \times 3N$ friction matrix that describes frictional interaction between each degree of freedom of each particle (a second-rank tensor [239]), and $\mathbf{R}$ is the random force consisting of $3N$ independent random variables acting on the system of interest due to the motion of solvent molecules. In practice, Ξ is often chosen to be block-diagonal, and furthermore, the 3×3 block diagonals of Ξ are each chosen to be isotropic for simplicity. Thus, the equation of motion is simplified to the following second-order stochastic differential equation (SDE) for a single Brownian particle i

in three dimensions:

$$m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \mathbf{f}_i(\mathbf{r}) - m_i \xi_i \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} + R_i \tag{2.24}$$

More extensive formal derivations of Langevin dynamics, including the treatment of Langevin dynamics with memory, can be found in Zwanzig [240–242] and Mori [243].

2.3.5 Fluctuation-Dissipation Theorem

The noise factor R is Gaussian distributed and δ-correlated in time:

$$\langle R(t) \rangle = 0, \qquad \langle R(t) R(t') \rangle = 2B\delta(t - t') \tag{2.25}$$

where B is a measure of the strength of the fluctuating force, and δ is the Dirac delta function indicating that there is no correlation between the noise values in any distinct time intervals $\mathrm{d}t$ and $\mathrm{d}t'$.

To resolve B, we first analytically solve the Langevin equation. Since Eq. 2.22 is a linear first-order differential equation, it has the solution:

$$v(t) = e^{-\gamma t} v(0) + \int_0^t \mathrm{d}t' e^{-\gamma(t-t')/m} R(t')/m \tag{2.26}$$

Substituting this into $\langle v(t)^2 \rangle$ gives:

$$\langle v(t)^2 \rangle = e^{-2\gamma t/m} v(0)^2 + \frac{B}{\gamma m}(1 - e^{-2\gamma t/m}) \tag{2.27}$$

The algebraic derivations for Eqs. 2.26 and 2.27 can be found in [244].

In the long time limit, the exponential terms drop out, and this quantity approaches $B/\gamma m$. But since the system approaches thermal equilibrium in the long time limit, the mean squared velocity must approach $k_B T/m$. Combining Eq. 2.18

with 2.27 consequently results in the following:

$$B = \gamma k_B T \tag{2.28}$$

This result is known as the Fluctuation-Dissipation Theorem. It relates the strength of the random noise B to the magnitude of the friction γ, and describes the balance between the two terms that were introduced into Langevin dynamics. The larger its value, the greater the influence of the surrounding fluctuating force exerted by the solvent. This balance is the requirement for a Langevin system to be in an equilibrium state.

2.3.6 Brownian Dynamics

Brownian dynamics (BD) is a simplified version of Langevin dynamics that corresponds to the limit where the friction is high, under which correlations in the velocity decay and no average acceleration takes place. Such a system is called an overdamped Langevin system. Since there is no average acceleration, the left-hand inertial term of Eq. 2.24 is ignored to give:

$$0 = \mathbf{f}_i(\mathbf{r}) - \gamma_i \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} + \sqrt{2B_i} G_i \tag{2.29}$$

where $\gamma_i = m_i \xi_i$ and G_i is the Gaussian random variable $\mathcal{N}(0, 1)$. This is known as the position Langevin equation.

Shifting the terms around and substituting in the Einstein-Smoluchowski relation

(Eq. 2.8) gives:

$$\begin{aligned}\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} &= \gamma_i^{-1}\mathbf{f}_i(\mathbf{r}) + \gamma_i^{-1}\sqrt{2B_i}G_i \\ &= \gamma_i^{-1}\mathbf{f}_i(\mathbf{r}) + \sqrt{2\gamma_i^{-1}k_BT}G_i \\ &= \gamma_i^{-1}\mathbf{f}_i(\mathbf{r}) + \sqrt{2D_i}G_i\end{aligned} \tag{2.30}$$

In order to run a BD simulation of a Langevin system, Eq. 2.30 needs to be transformed into an algebraic equation $\mathbf{r}_i(t+\Delta t)$ as a function of $\mathbf{r}_i(t)$, where Δt is the time step. This can be done by a time-ordered expansion using the Euler-Maruyama scheme [245], giving the following update equation for a Brownian particle in three dimensions:

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \gamma_i^{-1}\mathbf{f}_i(\mathbf{r}(t))\Delta t + \eta_i(t) \tag{2.31}$$

where $\eta_i(t)$ is the random displacement, defined as:

$$\eta_i(t) = \mathcal{N}(0, \sqrt{2D_i}) \tag{2.32}$$

This update formula is known as the Ermak-McCammon Method [246], and full details can be found in their landmark paper, which underpins the Brownian dynamics methodology. The steps for running a BD simulation are thus outlined in Algorithm 1. It should be noted that although one of the objectives of BD simulations is to estimate the diffusion constants of the particles of interest, an initial value for D must be supplied into the algorithm, which can simply be estimated by using the Stokes-Einstein relation (Eq. 2.9) for an ideal spherical particle.

Algorithm 1 Ermak-McCammon algorithm for BD simulations without hydrodynamic interactions

$t \leftarrow 0$;
$\Delta t \leftarrow$ Choose time step size;
$\boldsymbol{\gamma} \leftarrow$ Initialize the drag coefficients of all the particles;
$\mathbf{r}(0) \leftarrow$ Initialize positions of all the particles;
while `true` **do**
    $\mathbf{f}(\mathbf{r}(t)) \leftarrow$ Calculate the forces on each particle;
    $\eta(t) \leftarrow$ Generate a random displacement vector according to Eq. 2.32;
    $\mathbf{r}(t + \Delta t) \leftarrow$ Update the particle positions at the next time step according to Eq. 2.31;
    $t \leftarrow t + \Delta t$;
end while

2.3.7 Numerical Integration Schemes

While stability analysis has been well-studied for numerical solutions of SDEs, different definitions of stability have been introduced to quantify the term for different application contexts; this topic is covered in depth in [247–249]. The Euler-Maruyama (EM) scheme used in Eq. 2.31 is a simple first-order scheme that has been shown to be "numerically stable in the mean" by Saito and Mitsui [250, 251], but yields overestimations of calculated dynamic quantities, such as MSD values, that increases with increasing time step size Δt [252, 253]. This has been noted to be a fundamental issue with the original Brownian dynamics methodology, and several numerical methods of higher-order have been introduced since then to address this.

One such numerical method is the van Gunsteren-Berendesen (GB) Method [254], which proposes the following position update algorithm:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \frac{1}{2}\gamma_i^{-1}\Big(2\mathbf{f}_i(\mathbf{r}(t)) + \Delta t\dot{\mathbf{f}}_i(\mathbf{r}(t))\Big)\Delta t + \eta_i(t) \tag{2.33}$$

where the time derivative of the force is approximated by:

$$\dot{\mathbf{f}}_i(\mathbf{r}(t)) = \frac{\mathbf{f}_i(\mathbf{r}(t)) - \mathbf{f}_i(\mathbf{r}(t - \Delta t))}{\Delta t} \tag{2.34}$$

Since the Langevin system converges into a classical Newtonian system in the limit of zero friction, the GB scheme likewise converges to the Leap-Frog Integration scheme when applied to MD simulations.

Compared to the EM scheme, the GB scheme has been observed to give a larger deviation for the mean square displacement per time step, but yields a better estimate of static quantities, such as energies, than the EM scheme. This follows from the fact that although the GB scheme is of higher order than the EM scheme, it is not a true second order algorithm, as the deterministic part of the scheme is of order Δt^2 while the stochastic part involves only a term at the $\Delta t^{1/2}$ level.

In addition, higher-order integration schemes based on Runge-Kutta (RK) methods are also used for numerically solving SDEs [255–258]. Stochastic RK-based schemes have been observed to provide much more accurate estimates of both static and dynamic quantities than either the EM or GB schemes. However, as with their deterministic counterparts, they require more than one evaluation of the forces and/or higher-order derivative calculations per time step, which significantly reduces their computational efficiency and thus renders them advantageous mostly in simpler systems with a fewer variables. In-depth analyses and comparisons of these schemes can be found in [252, 253, 259, 260].

2.3.8 Advantages and Drawbacks

Due to its efficiency in handling large simulation volumes and particle numbers, as well as its ability to handle much longer time scales than those afforded by all-atom MD simulations, BD simulations are well suited for exploring many-protein scenarios that occur in biological cells, and are particularly useful for describing processes such as diffusion controlled reactions.

However, the Brownian dynamics methodology, or at least the simplified version that is described here, comes with a set of drawbacks. First, Langevin dynamics

mimics the viscous aspect of a solvent. Specifically, the Brownian regime of Langevin dynamics describes the diffusive limit of the Langevin equation where the motion is more random, and so uses very large values for the coefficient of friction γ. However, if one of the objectives of a simulation is to control temperature, one should be careful to use a small value for γ, and/or consider running Langevin dynamics simulations instead. See Table 2.1 for a comparison of fluid simulation methodologies.

Second, while Langevin dynamics removes degrees of freedom from the solvent, it does not fully model an implicit solvent; specifically, the model does not account for electrostatic screening and hydrophobic effects exerted by the solvent. More importantly, it does not capture hydrodynamic interactions for denser solvents. Though hydrodynamic interactions (HI) will not be explored in this work, the idea behind HI is that when a colloidal particle moves through solvent, it drags a part of the solvent with it, thus inducing a flow field that moves in parallel and affects neighboring particles. Hydrodynamic interactions add back the mechanical coupling between the colloidal particles that was lost in the implicit solvent approximation, albeit on a coarse-grained level. Depending on the system of study, the addition of HI may be required to correctly model the system [261].

Finally, the simplified version of formulation of BD assumes that the colloidal particles are spherical. However, in real systems, colloids such as proteins are in general not truly spherical, leading to colloidal dynamics that are different from those of perfect spheres. In addition, as consequence of the spherical particle assumption, rotational motion has been silently ignored. Similarly, the hydrodynamics of non-spherical colloids is different from those of perfect spheres. This can be addressed, however, by fine-graining the model of the colloids under study. For example, a protein can be modeled not as one but as a cluster of multiple Brownian particles bounded together by stretch and bend interactions, with each particle corresponding to a sub-region of the protein.

Table 2.1: Various granularity levels of physical descriptions for colloidal particles in a solution. $\tau_R = \xi^{-1}$ is the relaxation time of the particles. Table from [262].

Solvent **Model**	Microscopic **Variables**	Equation in the **Phase Space**	Equation of the **Trajectories**	**Simulation** Method
Discrete Solvent	$\{\mathbf{r}_{\text{solvent}}, \mathbf{p}_{\text{solvent}}, \mathbf{r}_{\text{solute}}, \mathbf{p}_{\text{solute}}\}$	Liouville	Newton	Molecular Dynamics
Continuous Solvent $m_{\text{solvent}} \ll m_{\text{solute}}$	$\{\mathbf{r}_{\text{solute}}, \mathbf{p}_{\text{solute}}\}$	Fokker-Planck	Langevin	Langevin BD
	$\{\mathbf{r}_{\text{solute}}\}$ $t > \tau_R$	Smoluchowski	Ermak-McCammon	Smoluchowski BD

2.3.9 Alternative Methods

BD represents a very coarse-grained model of a fluid system, and there are a number of other methods that range in accuracy and computational complexity from this model with the simplified implicit solvent all the way down to a fully-explicit solvent model employed by MD. As hinted earlier, one set of approaches involve adding in hydrodynamic interactions, such as BD with HI using the full Ermak-McCammon method [246] or Stokesian Dynamics [263, 264]. Another set of approaches is Fluid Particle methods, which involve numerically solving the Navier-Stokes equation [265, 266]. Other methods such as Dissipative Particle Dynamics [267–269], Multi-Particle Collision Dynamics (also known as Stochastic Rotation Dynamics) [270, 271], and the Lowe-Anderson thermostat [272] involve the use virtual particles to represent momentum "units" of the solvent. Yet another approach is the grid-based Lattice-Boltzmann simulation method, where a linearized Boltzmann equation is solved [273–277].

All of these methods introduce hydrodynamic interactions or some other form of softened interactions into the Langevin system, and include simplified but explicit models of the solvent. The solute particles are almost always treated by these methods as inertial systems. As a consequence, they are rather computationally expensive due to the addition of many more degrees of freedom for consideration and/or the long-

range nature of these interactions that require the use of elaborate numerical schemes (see [278] for a detailed comparison). Furthermore, while the theory of HI is rather well understood at low particle densities, much less is known at high particle densities [279], and so simple BD simulations without HI still has its place as a useful tool for studying colloidal dynamics in these conditions, in addition to circumventing the computational complexity of methods involving HI.

2.4 Electrostatics

The importance of electrostatic interactions in biological systems cannot be underestimated, and in an aqueous environment, the presence of electrolytes affects the functional, structural, and dynamic properties of every class of biomolecules, some of which that are relevant to this work have been described in Section 2.1.5. Thus, a brief overview of electrostatic theory for electrolytes in a dilute solution typical of the cellular environment is presented. Note that the centimeter-gram-second (cgs, or Gaussian) electrostatic units of measurement will be used throughout this section for consistency of notation.

2.4.1 Electrolyte Solution Model

Consider an electrolyte solution of volume V that is electrically neutral and consists of s different species of ions, with the i-th species having a concentration c_i and charge $q_i = z_i\boldsymbol{e}$, where z_i is i's charge valence and $\boldsymbol{e}$ is the charge of an electron (the elementary charge). It should be noted that only strong electrolytes, or those with high dissociation constants, such as Na^+Cl^-, are considered. Thus, ions will be considered as hard spheres that are distributed in a continuum of relative permittivity ε_r (formerly known as the "dielectric constant" [280]) and yielding an average charge density ρ. This system is governed by two equations: the Poisson equation, which defines the electric potential in terms of the charge density, and the Boltzmann equa-

tion, which defines the equilibrium charge distribution given an electric potential. When combined, these two equations form a self-consistent system.

The electroneutrality condition of the electrolyte solution requires that the charges of all the ions sum up to zero:

$$\begin{aligned} \frac{1}{V}\sum_{i=1}^{i=s} N_i z_i &= \sum_{i=1}^{i=s} c_i z_i \\ &= 0 \end{aligned} \tag{2.35}$$

where N_i is the total number of ions of species i.

We now fix a single ion j at the origin and calculate the equilibrium potential $\varphi_j(r)$ and charge distributions $c_i(r)$ of species i around it. First, we define the distance of closest approach a_d between ion j and a counter ion k:

$$a_d = r_j + r_k \tag{2.36}$$

where r_j and r_k are the radii of j and k, respectively.

The electroneutrality condition implies that the summation of all charges around j beyond r_j must be equal to the charge bore by j. Hence:

$$\int_{a_d}^{\infty} 4\pi r^2 \rho \mathrm{d}r = -z_j \boldsymbol{e} \tag{2.37}$$

The ions are assumed to be in thermodynamic equilibrium with each other and relatively free to move around in the solution, and so they obey Boltzmann statistics and form a Boltzmann distribution. The concentration c_i of species i is consequently altered from its bulk average value c_i^0 by a corresponding Boltzmann factor:

$$c_i(r) = c_i^0 \exp(-\beta U(r)) \tag{2.38}$$

where β is the inverse scale factor defined by Boltzmann's constant k_B and temperature T:

$$\beta = \frac{1}{k_B T} \tag{2.39}$$

The relationship in Eq. 2.38 is given by statistical mechanics, where the probability of a particle having energy U is proportional to the Boltzmann factor $e^{-\beta U}$. In the infinite temperature limit, the Boltzmann factor approximates to 1, and all ions are distributed uniformly without regard for their electrostatic interactions. Likewise, the distribution is smoothed out into a uniform distribution when $U = 0$.

In the absence of an external field, the energy U of an ion of species i with charge valence z_i in an electric potential field φ is equivalent to its electrostatic potential energy U_e, defined to be:

$$U_e(r) = z_i \boldsymbol{e} \varphi_j(r) \tag{2.40}$$

Substituting Eq. 2.40 back into Eq. 2.38 gives (abbreviating the dependence on r in the equation):

$$c_i = c_i^0 \exp(-z_i \boldsymbol{e} \varphi_j \beta) \tag{2.41}$$

This distribution function gives the probability of finding an ion of a single species i per volume element at a distance r from the central ion j. We can sum this result over all ion species i to express the overall charge density at distance r with respect to ion j:

$$\begin{aligned} \rho_j &= \sum_i c_i \boldsymbol{e} z_i \\ &= \sum_i c_i^0 \boldsymbol{e} z_i \exp(-z_i \boldsymbol{e} \varphi_j \beta) \end{aligned} \tag{2.42}$$

This is known as the Boltzmann Equation relating the charge distribution to the electric potential felt by ion j.

2.4.2 Poisson-Boltzmann Equation

We consider the Laplacian operator:

$$\Delta = \nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2} \tag{2.43}$$

which is also given in spherical coordinates by:

$$\Delta\varphi = \frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\varphi}{\mathrm{d}r}\right) \tag{2.44}$$

The Poisson Equation states that, at any point r, the Laplacian of the electric potential of a system is related to the charge density by the following formula:

$$\Delta\varphi = \frac{-4\pi}{\varepsilon_r}\rho \tag{2.45}$$

Thus, an ion will feel the electric potential created by other ions according to the above equation, and will adjust its spatial position in the presence of electrostatic boundary conditions, which can be either a constant surface potential or constant surface charge density. These boundary conditions are also referred to as Dirichlet and Neumann boundary conditions, respectively.

Combining Eq. 2.42 into Eq. 2.45 gives the following:

$$\Delta\varphi_j = \frac{-4\pi}{\varepsilon_r}\sum_i c_i^0 \boldsymbol{e} z_i \exp(-z_i \boldsymbol{e}\varphi_j\beta) \tag{2.46}$$

This is known as the full Poisson-Boltzmann equation.

2.4.3 Debye-Hückel Approximation

The full Poisson-Boltzmann equation is a partial differential equation of elliptic type with φ_j defined on both sides, and so is generally not analytically solvable, nor does it follow the principle of linear superposition for the relationship between the number of charges and the strength of the potential field. However, if the potential field is weak enough, the equation can be linearized and a solution can be approximated. This is the basis of Debye-Hückel theory [281]. The Debye-Hückel approximation assumes a sufficiently low concentration of ions such that $z_i \boldsymbol{e} \varphi_j \ll k_B T$, i.e. the energy derived from electrical forces is small compared to the thermal energy. Under this assumption, the exponential terms in Eq. 2.42 can be expanded using the Taylor series and the higher order terms (higher than $\mathcal{O}(\Delta\varphi_j^2)$) can be truncated, thus linearizing the equation. Doing so and subsequently removing the first expanded term by the electroneutrality condition (Eq. 2.35) gives:

$$\begin{aligned} \rho_j &= \sum_i c_i^0 \boldsymbol{e} z_i (1 - z_i \boldsymbol{e} \varphi_j \beta) \\ &= \boldsymbol{e} \sum_i c_i^0 z_i - \sum_i c_i^0 \boldsymbol{e}^2 z_i^2 \varphi_j \beta \\ &= - \sum_i c_i^0 \boldsymbol{e}^2 z_i^2 \varphi_j \beta \end{aligned} \tag{2.47}$$

Substituting this definition of ρ_j to the Poisson-Boltzmann equation, followed by rearrangement of the terms, simplifies the equation to:

$$\Delta\varphi_j = \kappa^2 \varphi_j \tag{2.48}$$

where κ is defined as:

$$\kappa^2 = 4\pi\varepsilon_r^{-1}\boldsymbol{e}^2\beta\sum_i c_i^0 z_i^2 \tag{2.49}$$

Eq. 2.48 is analytically solvable, and a general solution to this equation that satisfies the constraint that the potential converges to a finite value when $r \to \infty$ is:

$$\varphi_j = A\frac{e^{-\kappa r}}{r} \tag{2.50}$$

To obtain the value of the constant A, we first derive a new equation form for ρ_j, by substituting the left and right hand sides of 2.48 with 2.45 and 2.50, respectively, and subsequently rearranging the terms:

$$\rho_j = \frac{-\kappa^2\varepsilon_r}{4\pi}A\frac{e^{-\kappa r}}{r} \tag{2.51}$$

This form of ρ_j is then substituted into 2.37 to produce a new expression of the electroneutrality condition:

$$A\kappa^2\varepsilon_r\int_{a_d}^{\infty} re^{-\kappa r}\mathrm{d}r = z_j\boldsymbol{e} \tag{2.52}$$

From this equation, A can then be analytically solved using integration by parts, after which the following value is obtained:

$$A = \frac{z_j\boldsymbol{e}}{\varepsilon_r}\frac{e^{\kappa a_d}}{1+\kappa a_d} \tag{2.53}$$

In the case of a dilute solution with the Debye-Hückel approximation, $a_d \ll \kappa^{-1}$, and so the term $\frac{e^{\kappa a_d}}{1+\kappa a_d}$ approximates to 1, thus simplifying A and giving the final expression for the Debye-Hückel potential:

$$\varphi_j = \frac{z_j\boldsymbol{e}}{\varepsilon_r}\frac{e^{-\kappa r}}{r} \tag{2.54}$$

2.4.4 Debye-Hückel Screening Length

The value κ^{-1}, where κ is given from Eq. 2.49, is known as the Debye-Hückel screening length, or Debye length. Similarly, the Debye sphere for an ion j is defined as the volume surrounding j whose radius is j's Debye length κ_j^{-1}. Physically, the Debye length is the distance above which an ion's charge is effectively "shielded" by the surrounding ions in a solution. The excess charge within the Debye sphere of j is approximately equal to charge $-z_j\boldsymbol{e}$, thus canceling out the field of j's charge $+z_j\boldsymbol{e}$ at distances further than κ_j^{-1}. Symmetrically, as ions are located farther away from j, they are increasingly electrically screened off by ions closer to j. Consequently, the electric potential will decrease in magnitude by $1/\boldsymbol{e}$ with each κ_j^{-1} distance away from j. Thus, the Debye length is effectively a measure of an ion's net electrostatic effect in solution and how far out this effect persists, and is inversely dependent on the bulk concentration of electrolytes. For reference, the Debye length for an ion in a 1.0 M Na^+Cl^- solution is approximately 3 Å [282].

2.4.5 Advantages and Disadvantages of the Debye-Hückel Approximation

The Debye-Hückel treatment provides a simple description to the many-body interactions between ions by only assigning, for any given pair of ions separated by distance r, a pair-wise interaction that decays exponentially due to the screening by all neighboring ions surrounding the pair. Moreover, the treatment is claimed to be valid only at large distances between ions (i.e. greater than κ^{-1}) [283], which generally holds true when the salt concentration of the solution is low. Though the theory is at best an approximation of the distribution of the electric potential in solution, predictions made by Debye-Hückel theory have been shown to be rather accurate under physiological conditions, where the electrolyte strength is about 0.1 M [284], and so this model is used in our work. However, it should be noted that, while it produces good results for solutions of monovalent ions, the Debye-Hückel approximation and

Poisson-Boltzmann theory in general fails to reproduce some of the important features associated with multivalent ion solutions [282]. Thus, for solutions of higher ionic strengths, where Debye-Hückel theory can no longer accurately predict electrolyte behavior, alternative electrostatic models are considered instead, such as the Pitzer equations [285], the Davies equation [286], or specific ion interaction theory [287]. For more discussions on the theory of electrostatics applied to biological systems, see [281–283, 288].

CHAPTER 3
COMPUTATIONAL METHODS

3.1 *In Silico* Representation

The DNA chain was modeled as a chain of adjacent soft beads, each with radius $a = 1.59$ nm. This value has been used in earlier studies to reproduce the hydrodynamic properties of DNA [289], and corresponds to approximately 9.35 base pairs. The interaction forces of this DNA model follow those of Chow and Skolnick [1], which in turn are based on the model proposed by Jian, Schlick, and Vologodskii [290, 291], hereafter referred to as the JSV model. It should be noted that many alternative coarse-grained models of DNA exist, such as the elastic rod model by Swigon et al. [292], the mesoscale model by Knotts et al. [293], and the bead-rod model by Wang and Gao [294], which employs inextensible rods. However, the JSV model was chosen due to its suitability for simulating and roughly quantifying the diffusive motions of a large DNA strand ($>$ 100 kbp) by coarse-graining out the strand into a string of individual beads whose displacements over time can be easily measured. The forces of the JSV model are discussed in detail in Section 3.2.3 onwards.

Proteins were also modeled as soft beads, but without charge nor affinity to any specific DNA sites (i.e. DNA beads). For studies of the DNA-protein system, all protein beads in the given system are assigned the same radius, and the number of beads introduced to the system was set such that the protein volume fraction was fixed to a physiological protein volume fraction of 6% (see Section 2.1.3 for the estimation of this value).

3.1.1 Nondimensional Units

All simulations were performed using nondimensional units, with a nondimensional time step size of 10^{-4}, or 1.64 ps. Particle positions were recorded for analysis every 1000 time steps, which corresponds to 0.1 units of time (1.64 ns). In the context of this work, one unit of time is defined to be the characteristic time interval required for a particle in a monodisperse solution to escape its cage of immediate neighbors, i.e. transition from short-time to long-time diffusion.

MSD values are measured and reported in units of a^2, where a is the radius of a DNA bead. Likewise, diffusion constants are reported in units of D_0, where D_0 is defined to be the infinite-dilution diffusion constant for a particle of the same size as a DNA bead as given by the Stokes-Einstein relation:

$$D_0 = \frac{k_B T}{6\pi\eta a} \approx 154\ \mu\text{m}^2/\text{s} \tag{3.1}$$

where η is the viscosity of water under physiological conditions. Consequently, time is measured in units of a^2/D_0, which corresponds to 16.4 ns.

3.2 Simulation Technical Details

We developed a custom Brownian dynamics simulation package called `BDT` to run the simulations in this study. The following subsections describe the theory and technical details underlying the implementation of `BDT`.

3.2.1 Periodic Boundary Conditions

To approximate a large DNA system in the bulk range without needing to consider DNA-solvent interface effects, periodic boundary conditions (PBCs) were placed on the system. PBCs are a common technique in particle simulations, in which systems approximated by PBCs consist of an infinite number of unit cells, one of which is

the original simulation box being maintained by the simulation program, and the rest are copies called images [228, 295]. Properties of the original simulation box are recorded using the minimum-image convention, in which each individual particle in the simulation interacts with the closest image of the remaining particles in the system. The space made by a PBC is topologically a torus: as particles pass through one face of the simulation box, they (or rather, their image) re-enter through the opposite face.

While PBCs can be implemented by "folding back" the positions of the particles into the box when they leave, `BDT` does not perform positional fold-backs; instead, particles are allowed to exit the box volume, but interactions will be computed with the nearest images. The choice of PBC implementation has no effect on the course of the simulation beyond numerical precision errors; however, diffusion properties of the system such as mean displacements and diffusion lengths would become impossible to calculate if particles experienced unphysical positional displacements over time due to PBC fold-backs.

3.2.2 Neighbor Lists

In particle simulations, pair interaction energies between particles i and j are usually modeled in the following form [295]:

$$U_{ij} \propto r_{ij}^{-n} \tag{3.2}$$

where r_{ij} is the inter-particle distance. These interactions can be classified as either short-ranged ($n > 1$) or long-ranged ($n \leq 1$). While the exact computation of pair interactions between particles is an $\mathcal{O}(N^2)$ operation, it is possible to reduce the runtime complexity for computing short-ranged interactions by leveraging the observation that short-range interactions decay quickly, and skipping the calculation of pair interactions altogether for all particles j greater than a certain distance $r_{i,cutoff}$

from particle i.

To achieve this, neighbor lists (also known as Verlet lists) are used to filter out pairs of particles down to only those that are within a certain cutoff radius of each other. Neighbor lists can be naïvely built by periodically checking all possible pairs of particles, which is an $\mathcal{O}(N^2)$ procedure. However, this routine can be optimized into an $\mathcal{O}(N)$ algorithm using acceleration data structures. The accelerated neighbor list algorithm implemented in `BDT` is known as the cell-list-accelerated neighbor list method [228, 296, 297]. This method first spatially sorts particles into spatial bins, called cells, that are sized by the largest cutoff radius of all pair potentials in the system (Algorithm 2). To find neighbors, each particle then needs to search only its cell and the 26 surrounding cells, as opposed to the entire simulation domain (Figure 3.1). Since the number of particles per cell is roughly constant and particle binning is an $\mathcal{O}(N)$ operation, neighbor search using cell lists is effectively an $\mathcal{O}(N)$ operation.

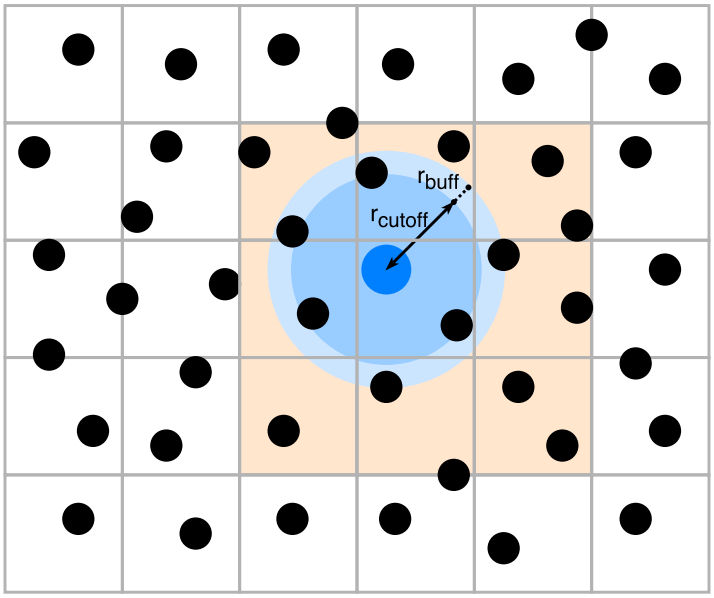


Figure 3.1: Neighbor search using cell lists in 2-dimensional space. Instead of searching the entire simulation domain, each particle needs to search only its resident cell and the surrounding cells (orange) for neighbors to add to its neighbor list.

A small buffer radius r_{buff} is typically added to the cutoff radius when searching for neighbor particles, so that the neighbor list can be computed less frequently. Doing

so will require the neighbor list to be rebuilt only when a particle diffuses over $r_{buff}/2$ distance, as opposed to after every time step in the simulation. However, increasing r_{buff} also increases the number of particles that are included in the neighbor list, which slows down the evaluation of pair interactions, and so r_{buff} must be chosen judiciously.

While the cell-list-accelerated neighbor list method is very efficient for systems with nearly monodisperse cutoffs, such as in the case of this study, the performance will degrade for systems with large cutoff radius asymmetries due to increased number of particles per cell and increased search volume. Studies have shown that the cell-list-accelerated neighbor list method works best when the asymmetry between the largest and smallest cutoff radius is less than a 2:1 ratio [296]. Neighbor list methods based on alternative acceleration data structures are found to be more suitable for these scenarios; these are discussed in detail in [296]. In addition, multiple neighbor lists can be built for calculating different types of pair interactions, each with its own appropriate r_{cutoff} and acceleration data structure. Building multiple neighbor lists is useful when there is a significant disparity in the pairwise cutoffs defined by different pair potentials, since each neighbor list will be built based on the exact r_{cutoff} of the corresponding pair potential, and pair interaction calculations for a particle i will not involve particles j that lie outside $r_{i,cutoff}$. This technique improves the performance of the pair interaction calculations, though at the expense of duplicate computations.

3.2.3 Steric Interactions

BD simulations [246] without hydrodynamic interactions (HI) model steric interactions between particles using the free-draining (FD) approximation, in which particles

Algorithm 2 Optimized neighbor list construction algorithm

```
{Assign particles into their appropriate cells}
for all particles i in the system do
    Calculate the index j of the cell that particle i currently resides in;
    Append i into cell j's particle list;
end for
{Apply Verlet procedure to build neighbor list}
for all particles i in the system do
    l ← cell number of particle i;
    for all cells m among cell l and neighbors of l do
        for all particles j in cell m do
            s_ij ← Compute the absolute distance between i and j;
            r_ij ← Apply PBC to s_ij to obtain the minimum image distance;
            if r_ij < r_i,cutoff + r_buff then
                Append j into the neighbor list of particle i;
            end if
        end for
    end for
end for
```

experience repulsive forces only when they overlap:

$$U_{steric} = \begin{cases} \frac{1}{2}k_s(r_{ij} - (a_i + a_j))^2, & \text{if } r_{ij} < a_i + a_j \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

where r_{ij} is the distance between the centers of the two particles, a_i and a_j are the radii of the two particles, and k_s is the steric force constant. This form can be normalized by dividing by the average of the particle radii:

$$U_{steric} = \begin{cases} \frac{1}{2}k_s(r' - 2)^2, & \text{if } r' < 2 \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

where $r' = 2r_{ij}/(a_i + a_j)$ is the normalized distance between two particles. In this work, a value of $k_s = 100k_BT/a^2$ was used, to match the DNA stretching force constant used (to be discussed in Section 3.2.4). Steric interactions between DNA beads that are adjacent and next-adjacent to each other were not computed, since

their interaction forces are already accounted for by the stretch and bend potentials, respectively (to be discussed in 3.2.4 and 3.2.5).

Because steric interactions are non-existent when the inter-bead distance is greater than $2a$, only a tiny subset of the $N-1$ other DNA beads needs to be accounted for when computing the forces acting on a single bead. Thus, `BDT` maintains a neighbor list with a cutoff of $2a' + \epsilon$, where a' is the radius of the largest particle in the system (DNA bead or protein) and ϵ is a small value relative to a'. Details of `BDT`'s neighbor list implementation can be found in Section 3.2.2. `BDT`'s use of neighbor lists not only reduced the steric forces computation routine from an $\mathcal{O}(N^2)$ to effectively $\mathcal{O}(N)$ operation, but also enabled the parallelization of the steric forces computation routine overall (Algorithm 3).

Algorithm 3 Parallel steric forces computation algorithm

```
1: {Run iterations of the outer for-loop in parallel}
2: for i = 1 to N do
3:     for all neighbors j of particle i do
4:         δ ← Compute the steric force between i and j;
5:         f_i ← f_i + δ;
6:     end for
7: end for
```

3.2.4 Stretch Interactions

The JSV model of DNA defines the stretch potential holding adjacent DNA beads together to form a chain to be:

$$U_{stretch} = \frac{h}{2}(r_{ij} - 2a)^2 \tag{3.5}$$

where r_{ij} is the center-to-center distance between the adjacent beads i and j, and a is the DNA bead radius. This form implies that the equilibrium stretch distance is the center-to-center distance between the beads at the point where they exactly touch. The stretch potential corresponds to the following force function that describes each

inter-bead stretch:

$$F_{stretch} = -\frac{\mathrm{d}U_{stretch}}{\mathrm{d}r_{ij}} = -h(r_{ij} - 2a) \tag{3.6}$$

where h is the force constant that has been experimentally determined [290] to be $h = 100k_BT/a^2$.

In `BDT`'s code, stretches are stored as `structs` holding the indices of the adjacent beads participating in the stretch and the stretch force constant. Computing the total contribution of the stretch forces to the overall forces is a trivial routine that can be optimized by parallelization. However, with the exception of those situated at the ends of the chain, all beads in the DNA chain participate in exactly two stretch interactions, and so care must be taken to avoid race conditions when updating the force values on each participating bead in parallel. If the stretch interactions are indexed in the order that they appear along the chain, then the topological structure of the chain lends itself to the straightforward algorithm implemented in `BDT`, Algorithm 4.

Algorithm 4 Parallel stretch forces computation algorithm

```
1: k ← 2;                          ▷ Set the number of rounds of parallelized operations
2: {Assume a sequential numbering of the stretch interactions}
3: for j = 0 to k − 1 do
4:     {Run iterations of the inner for-loop in parallel}
5:     for all stretch interactions i where i mod k = j do
6:         Compute and update forces for particles participating in stretch i;
7:     end for
8: end for
```

Under Algorithm 4, all even-indexed stretch interactions are computed together in parallel in the first round, followed by odd-indexed stretch interactions in the second round (Figure 3.2). This algorithm completely avoids data race conditions, as force values associated with each DNA bead are accessed and updated only once during each parallelization round (line 6 of Algorithm 4).

The general problem of finding the most optimal scheduling for the stretch forces

computation routine, can be stated as a problem of finding the minimum number of buckets k that the stretch interactions can be bucketed into, such that no two stretch interactions in each bucket set share a common participant bead. This is isomorphic to the Edge Coloring Problem in graph theory. While finding the optimal solution to this problem is known to be NP-hard, it has been shown that the edges of every simple undirected graph with at most Δ edges per vertex can be colored with at most $\Delta + 1$ colors. This is known as Vizing's Theorem [298]. Algorithm 4's construction is based on this theorem, and it can be clearly seen that in the case of the DNA chain model, it is *the* most optimal solution to the problem.

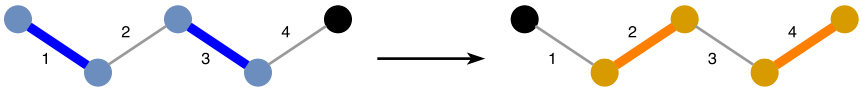


Figure 3.2: Computing stretch potentials in parallel using Algorithm 4. Stretches 1 and 3 are processed in parallel in the first round (blue) followed by 2 and 4 in parallel (orange).

3.2.5 Bend Interactions

In addition to the stretch potential, the JSV model also defines the bend potential between three adjacent DNA beads to model the elastic properties of a polymer. The bend potential is given as:

$$U_{bend} = \frac{g}{2}\theta_{ijk}^2 \tag{3.7}$$

where θ_{ijk} is the angle between the three beads. The bend potential uses a rigidity constant value of $g = 14.8k_BT$, which was selected using long simulations in infinitely dilute conditions to match the persistence length of DNA (50 nm) [290].

Similar to stretch interactions, bends in `BDT` are stored as `structs` holding the indices of the participating beads and the bend force constant. Likewise, it is possible to compute the total contribution of the bend forces to the overall forces in parallel, and Algorithm 4 can be leveraged for this, simply by setting $k = 3$ in line 1 of the

algorithm (Figure 3.3).

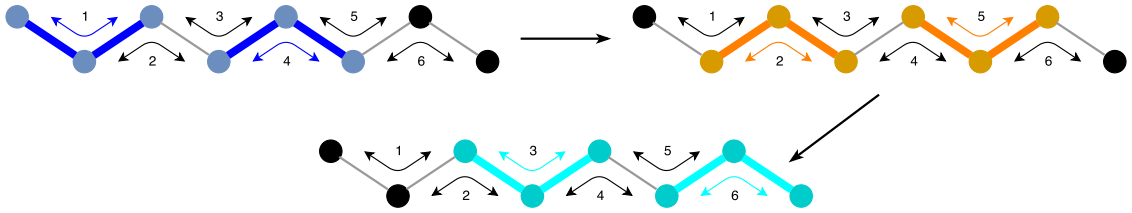


Figure 3.3: Computing bend potentials in parallel using Algorithm 4. Bends 1 and 4 are processed in parallel in the first round (blue), followed by 2 and 5 in parallel (orange), followed by 3 and 6 in parallel (cyan).

3.2.6 Debye-Hückel Interactions

Electrostatic interactions between DNA beads were modeled using the Debye-Hückel (DH) potential for an infinitely-thin charged cylinder [291, 299]. DNA beads are negatively charged, and each bead in the chain is assigned an effective point charge q based on the charge of a charged cylinder segment of equivalent size:

$$q = 2a\nu \tag{3.8}$$

where a is the radius of the DNA bead and ν is the effective linear charge density of the model DNA. ν corresponds to the best approximation for the Poisson-Boltzmann solution for DNA modeled as a charged cylinder, and extensive details on the method for deriving values for ν can be found in [300, 301]. Both literature and regression-estimated values for ν at different ion concentrations are shown in Table 4.1.

Applying Eq. 2.54, the DH potential experienced by a bead i of the DNA chain is then:

$$U_{DH} = \sum_{j>i+1} \frac{q^2}{\varepsilon_r} \frac{\exp(-\kappa r_{ij})}{r_{ij}} \tag{3.9}$$

where ε_r is the relative permittivity of the medium, and κ is the inverse of the Debye screening length. A symmetric monovalent electrolyte (e.g. Na^+Cl^-) solution is

assumed, in which the concentration of electrolyte pairs (i.e. the salt concentration) is equal the separate concentrations of the cations and anions. Thus, applying Eq. 2.49 gives:

$$\begin{aligned}\kappa^2 &= 4\pi\varepsilon_r^{-1}\boldsymbol{e}^2\beta(c_+^0 z_+^2 + c_-^0 z_-^2) \\ &= 4\pi\varepsilon_r^{-1}\boldsymbol{e}^2\beta(2c_s)\end{aligned} \tag{3.10}$$

or equivalently:

$$\lambda_D = \kappa^{-1} = \sqrt{\frac{\varepsilon_r k_B T}{8\pi\boldsymbol{e}^2 c_s}} \tag{3.11}$$

where c_s is the molar salt concentration of the solution. In this work, the systems modeled an aqueous medium under room temperature conditions ($\varepsilon_r = 80$, $T = 298$ K). The Debye screening lengths corresponding to the salt concentrations used in this work and following this model can be found in Table 3.1.

Table 3.1: Debye screening lengths for symmetric monovalent electrolytes in a monodisperse solution at different salt concentrations.

Salt Concentration (M)	**Debye screening Length λ_D (nm)**
0.0001	30.7
0.001	9.71
0.01	3.07
0.05	1.37
0.1	0.97
0.2	0.69
0.3	0.56
0.4	0.49
0.5	0.43
1.0	0.31
2.0	0.22

Like steric interactions, DH interactions are short-ranged, and so their calculations can be accelerated through the use of neighbor lists. However, since DH interactions do not simply truncate for $r_{ij} > 2a$ as steric interactions (as defined in Eq. 3.4) do, an

appropriate cutoff must be chosen for building a neighbor list suitable for calculating DH interactions (discussed in Section 3.4). In `BDT`, a single common neighbor list was used for calculating both steric and DH interactions as opposed to separate neighbor lists, as no significant computational time difference was observed between computing steric and DH interactions using a common list versus two separate neighbor lists. Similar to the case of steric interactions, DH interactions were not computed between DNA beads that were adjacent or next-adjacent to each other, since their interactions are already accounted for by the bend and stretch potentials.

3.3 Preparation of the Initial DNA(-Protein) System

3.3.1 Hilbert Curve Construction of the DNA Chain

To construct an initial DNA configuration for our studies, DNA beads were arranged along a Hilbert curve [302] with a spacing of $2a$ between the centers of each bead along the chain. The straight segments of the Hilbert curve were set to be at least 100 nm (twice the persistence length of DNA), which allows the DNA to start in a relaxed state. The chain was grown to 16,352 beads (approximately 152,891 base pairs) so that it would completely fill up the volume of a cube while having at least the length of 6,912 beads used for the model in [1] (Figure 3.4a). Details of the theory behind Hilbert curves and algorithm for generating Hilbert curves can be found in [303–305], and an example code implementation for 3D Hilbert curve generation can be found in [306]. Unlike the DNA model prepared in [1], the DNA chain prepared in this work was left to be an un-closed loop in order to preserve the fractal property of the chain generated by the Hilbert algorithm.

3.3.2 DNA Chain Compression

The size of the simulation box was chosen to approximate a physiological DNA volume fraction of 13% in order to be consistent with the study methods laid out in

[1]. A compression process was applied to reduce the box volume from 718.68^3 nm^3 to 128.4^3 nm^3 so that the desired volume fraction of 13% is reached. Volume reductions were proportional to the remaining volume difference between the current and target box volumes, and equilibration steps were added in between (Algorithm 5). The process naturally compacts and coils the DNA depending on the stretching and rigidity parameters of the DNA model, and its effect on the chain's internal structure and topology can be observed in Figure 3.4. Since compression is not an actual physical process, DH interactions were ignored during compression as an implementation optimization. A trajectory of the first 600 ns of the compression process for a DNA-protein system is shown in Movie S1.

Algorithm 5 DNA(-protein) system compression algorithm

```
1: n_steps ← 1000;
2: v_c ← Current box volume;
3: v_t ← Target simulation box volume that results in DNA volume fraction of 13%;
4: while v_c > v_t do
5:     {Equilibrate the system}
6:     for i = 1 to n_steps do
7:         {Advance single time step in BD simulation}
8:     end for
9:     {Shrink box volume by 1% of the difference between current and target volumes}
10:    Δv ← 0.01 · (v_c − v_t);
11:    v_c ← v_c − Δv;
12: end while
```

3.3.3 Addition of Proteins

For systems containing both DNA and protein, protein beads were inserted into the system at pseudorandom positions after the generation of the DNA chain but before the compression process (Figure 3.4a). The protein beads were placed such that they did not overlap with neighboring protein and DNA beads. For a given system, the radius was set to be uniform across all constituent proteins, and the number of proteins was pre-computed and assigned such that the target protein volume fraction

of 6% is reached along with the desired DNA protein fraction after the compression process.

3.4 Estimation of Debye-Hückel Interaction Cutoffs

A test was created to estimate appropriate cutoffs (r_{cutoff}) for building neighbor suitable for use with computing DH interactions. Using the magnitude of the inter-particle forces as a measurement quantity, we define $\zeta(r)$ as the proportion of contribution by DH interactions to the overall forces felt by the system:

$$\zeta(r) = \frac{\sum_N \left\| \vec{f}_{DH}(r) \right\|}{\sum_N \left\| \vec{f}_{steric} + \vec{f}_{stretch} + \vec{f}_{bend} + \vec{f}_{DH}(r) \right\|} \tag{3.12}$$

where N is the number of particles in the system, and $\vec{f}_{DH}(r)$ is the total DH interaction acting on a particle from its neighbors that are within distance r from it. The test finds an appropriate cutoff length by searching for the minimum value of r such that any marginal increase in r will result in a less than 0.01% increase in $\zeta(r)$:

$$\begin{aligned} &\min \quad r \\ &\text{s.t.} \quad \left\| \zeta(r + \Delta r) - \zeta(r) \right\| < 0.0001 \end{aligned} \tag{3.13}$$

where Δr is a small value relative to a (our work used $\Delta r = 0.1a$). To ensure safety against possible numerical errors, the test was set to terminate the search not at the exact minimum value r' that satisfied this condition, but at $r' + 3\Delta r$. This test was performed using the prepared initial DNA configuration to obtain suitable r_{cutoff} values for calculating DH interactions in systems at different salt concentrations (Table 4.2).

During the study, alternative schemes were tried as the test condition for searching the appropriate cutoff, shown in Eqs. 3.14 and 3.14 for completeness. These schemes have poor physical and mathematical basis; moreover, the cutoff values produced by

these schemes were often smaller than either the Debye lengths corresponding to the salt concentration level of the system and/or the cutoff values produced by the scheme in Eq. 3.13, suggesting that they may not be sufficient enough to accurately reflect DH interactions to within acceptable numerical error margin.

$$\frac{\overline{\|\vec{f}_{DH}(r+\Delta r)\|} - \overline{\|\vec{f}_{DH}(r)\|}}{\overline{\|\vec{f}_{steric}\|} + \overline{\|\vec{f}_{stretch}\|} + \overline{\|\vec{f}_{bend}\|} + \overline{\|\vec{f}_{DH}(r)\|}} < 0.0001 \tag{3.14}$$

$$\frac{\overline{\|\vec{f}_{DH}(r+\Delta r)\|} - \overline{\|\vec{f}_{DH}(r)\|}}{\overline{\|\vec{f}_{steric} + \vec{f}_{stretch} + \vec{f}_{bend}\|} + \overline{\|\vec{f}_{DH}(r)\|}} < 0.0001 \tag{3.15}$$

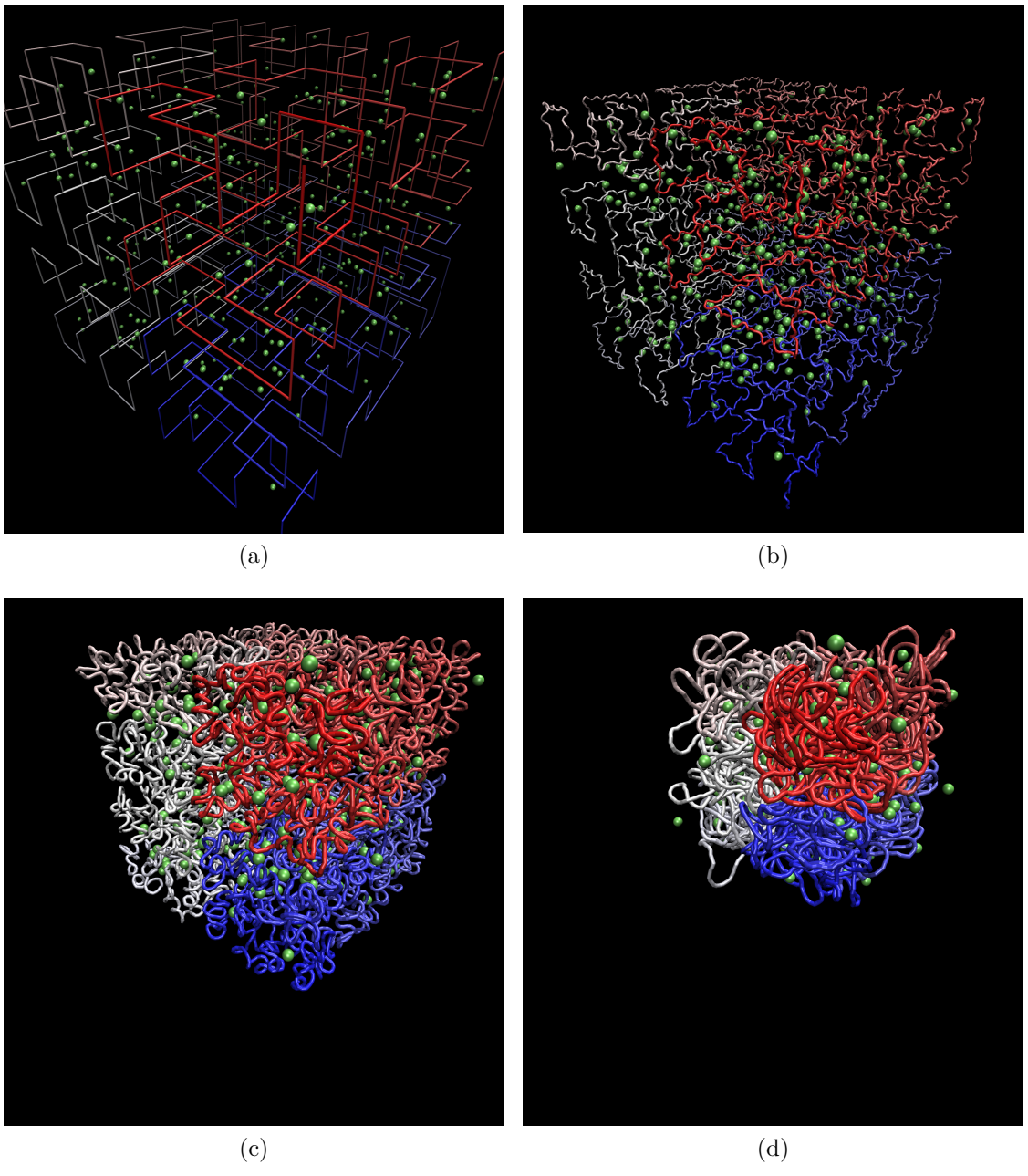


Figure 3.4: Preparation of the DNA(-protein) system, starting with generation of the DNA chain as a Hilbert curve (a), followed by compression (b), (c) into its final form (d). Figures were generated by VMD [307].

CHAPTER 4

RESULTS

4.1 Estimation of the Effective DNA Linear Charge Density

While the linear charge density of DNA ν can be implicitly solved by numerical integration methods as hinted in Section 3.2.6, a regression model was applied in this study instead to estimate ν. A log-log graph of ν against salt concentration was plotted using existing literature data, from which an exponential relationship between the two variables was observed (Figure 4.1). Accordingly, we fitted a linear regression model between $\log\log\nu$ and $\log c_s$ using the least squares approach, and built a model where the sum of the squared residuals was 0.0045. The small residual provided a level of confidence to the use of this regression model for estimating ν for the set of salt concentrations used in this study. In this work, existing literature values were used where possible, and regression-estimated values of ν were deferred to only where needed (Table 4.1).

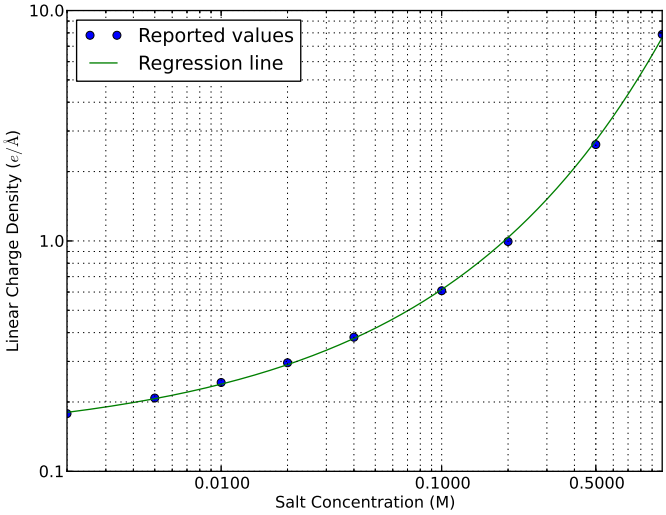


Figure 4.1: Log-log plot of reported DNA linear charge density values (ν) vs. salt concentration.

Table 4.1: Reported and extrapolated effective linear charge densities of DNA modeled under the Debye-Hückel potential at different salt concentrations. $\boldsymbol{e}$ is the charge of an electron.

Salt Concentration (M)	**Effective Charge Density ($\boldsymbol{e}$/Å)**
0.0001	0.147[a]
0.001	0.168[a]
0.002	0.178
0.005	0.208
0.01	0.243
0.02	0.296
0.04	0.382
0.05	0.415[a]
0.1	0.608
0.2	0.994
0.3	1.517[a]
0.4	2.078[a]
0.5	2.62[b]
1.0	7.88
2.0	30.47[a]

[a] Value is based on estimation using regression model against existing reported values.
[b] [300] reports 2.62 while [301] reports 2.59.

4.2 Effect of Salt Concentrations on DNA Internal Motions

To evaluate the effects of salt concentration on DNA internal motions, we performed BD simulations with the prepared bulk DNA-only system for a range of salt concentration values from 10^{-4} M to 2.0 M, as well as without Debye-Hückel potential (DH) interactions enabled (effectively corresponding to salt concentration $c_s = \infty$). Unless otherwise stated, all simulations were run for 2,000 units of time or 32.8 μs. Mean square displacement (MSD) analysis was performed on the output trajectories, and the MSD curves for the simulation trajectories at the different concentrations are overlaid and plotted in Figure 4.2. The anomalous diffusion regime was ignored, and the diffusion constants were measured only from a time interval of 80 (1.3 μs) onwards, where nearly linear behavior of the MSD curves was observed. The cutoff lengths used for building neighbor lists and estimated diffusion constants for DNA at

different salt concentrations are reported in Table 4.2. In addition, a semilog plot of estimated diffusion constants against salt concentration is prepared in Figure 4.3.

Our coarse-grained model of the DNA was able to reproduce the simulation results of earlier studies. In particular, the estimated diffusion constant for DNA under 0.01 M conditions ($D \approx 0.0094 D_0$) closely matches the number reported by Chow and Skolnick ($D \approx 0.0109 D_0$) [1], which suggests a level of robustness in our results, given that our methods modeled DNA length and configurations that are substantially different from those presented in [1].

To illustrate the effect of electrostatic interactions on the internal motions of the model chain, Figure 4.3 shows a trendline plot of the computed diffusion constants over the range of salt concentrations used in this study. The diffusion activity of DNA as a function of salt concentration follows an S-curve pattern, with a lower limit of $\approx 0.009 D_0$ ($\approx 1.45\ \mu\text{m}^2/\text{s}$) when the concentration is 10^{-4} M, and an upper limit of $\approx 0.021 D_0$ ($\approx 3.16\ \mu\text{m}^2/\text{s}$). This observation can be explained by Debye-Hückel theory: at extremely high salt concentrations, electrostatic interactions between the negatively charged phosphate groups along the DNA backbone are screened and thus internal motions are limited only by the self-crowding of DNA. At extremely low concentrations, there is little to no screening and the tapering of the diffusion constant is due to the electrostatic repulsion imposed by the DNA chain onto itself. Interestingly, the most rapid changes in DNA internal motions occur near the physiological salt concentration range of 0.1 M, the implications of which will be discussed in Section 5.1. We conclude that the internal motions of DNA in the nucleoid can be significantly affected by differences in the salt concentration of its environment, and that this change in mobility may in turn influence the ability of DBPs such as LacI to diffuse around DNA strands. A trajectory of the DNA-only system under physiological salt concentration 0.1 M is shown in Movie S2, which covers a simulation interval of 600 ns.

Table 4.2: Neighbor list cutoffs used and estimated diffusion constants of DNA (beads) for the DNA-only system at different salt concentrations.

Salt Concentration (M)	**Cutoff (a)**	**Estimated DNA (Bead) Diffusion Constant (D_0)**
0.0001	11.5	0.011157
0.001	9.4	0.009987
0.01	7.4	0.009445
0.05	5.5	0.010776
0.1	4.9	0.012781
0.2	4.3	0.015299
0.3	4.1	0.016402
0.4	3.9	0.016928
0.5	3.7	0.017645
1.0	3.2	0.018975
2.0	2.8	0.019798
∞ (no DH)	-	0.020522

4.3 Effect of Protein Size on DNA Internal Motions and Protein Diffusion in the Nucleoid

Next, we evaluate the effect of protein size on both DNA and protein diffusive motions in the nucleoid environment. We carried out a similar simulation protocol for a set of DNA-protein systems at a physiological salt concentration of 0.1 M, where each system contained proteins with an assigned common radius that was unique for that system. The number of beads introduced into each system was set such that the protein volume fraction remained fixed to 6%. The MSD curves for the simulation trajectories at the different protein radii are overlaid and plotted in Figure 4.5. Figure 4.6 shows a trendline plot of the computed diffusion constants for both DNA (beads) and proteins over the range of systems where the radii of the constituent proteins spanned from the hydrodynamic radius of DNA beads to that of a sphere with 4 times the hydrodynamic volume of a protein with similar molecular weight as LacI.

The diffusion values estimated from this study for a protein with a size compara-

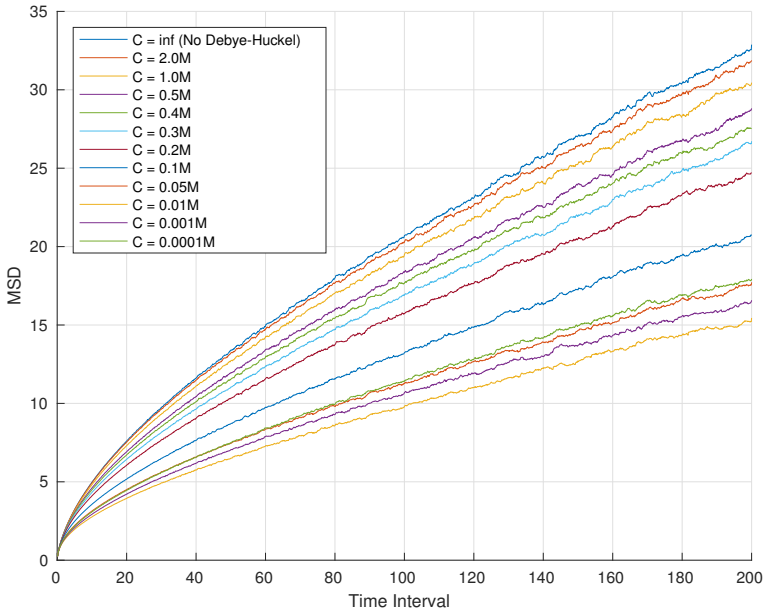


Figure 4.2: MSD curves of DNA (beads) for BD simulation trajectories of the DNA-only system at different salt concentrations.

ble to the effective hydrodynamic radius of LacI were slightly different from those in [1], which reported a diffusion value of $0.025D_0$ ($\approx 3.85\ \mu\text{m}^2/\text{s}$) that is close to the experimental values observed *in vivo* ($\approx 3\ \mu\text{m}^2/\text{s}$) [170]; our computer studies reported a value of $0.0504D_0$ ($\approx 7.76\ \mu\text{m}^2/\text{s}$) (Table 4). There may be two reasons for the difference in protein diffusion constants. The first is the difference in salt concentrations of the prepared systems. Since we used a physiological salt concentration of 0.1 M instead of the 0.01 M specified in [1], the increased effective electrostatic screening will allow the DNA chain to diffuse faster, and since DNA internal motions have been shown to increase protein motions, it may not be unexpected to consequently observe an increase in protein diffusion constants. The second is that our simulation studies did not incorporate hydrodynamic interactions (HI), which have been demonstrated in earlier studies to reduce DBP motions by nearly a factor of 2 [1]. However, since the inclusion of HI in the aforementioned study did not appear to significantly reduce DNA diffusive motions, we do not anticipate the addition of HI to significantly alter the diffusion trends found in the results, which we describe next.

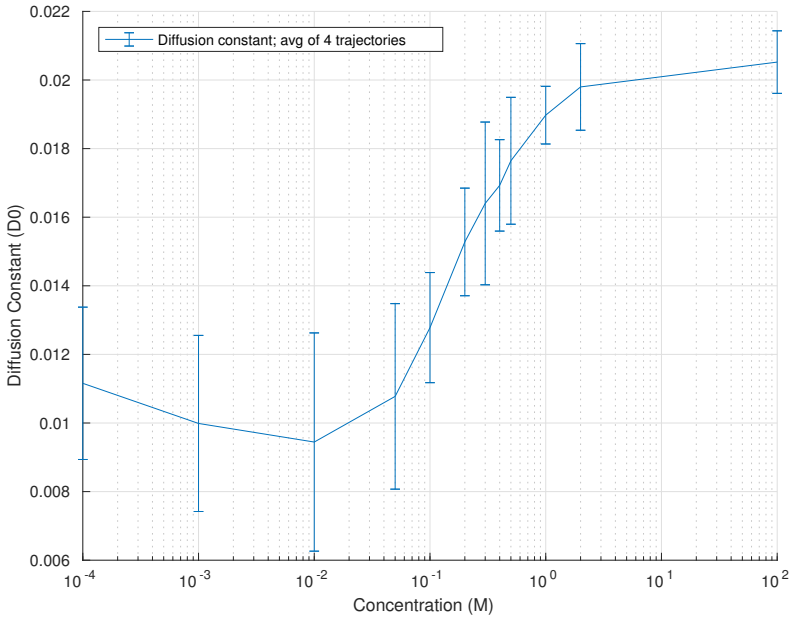
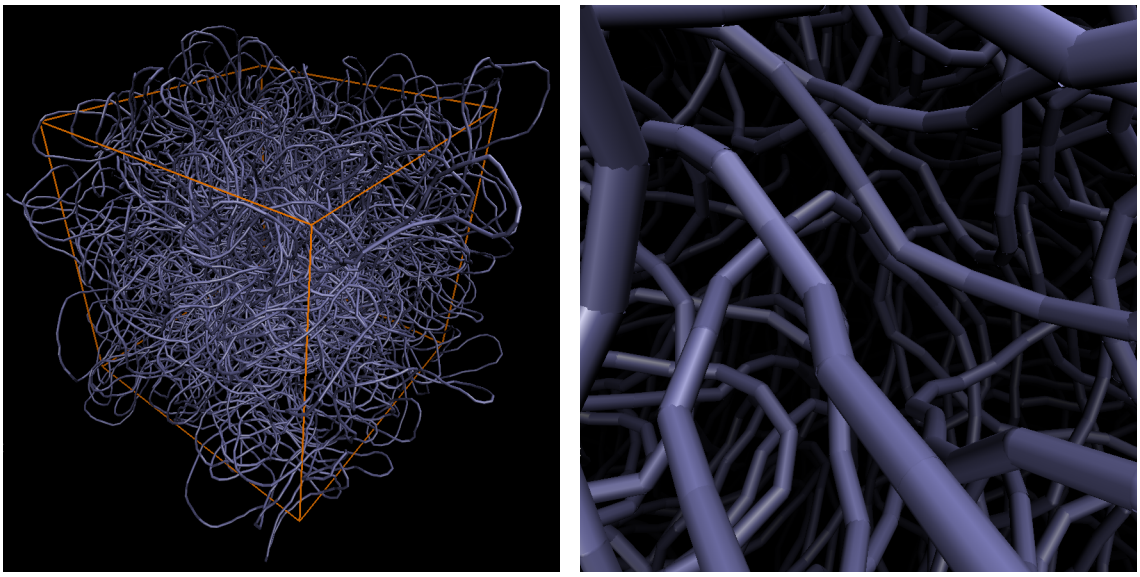


Figure 4.3: Estimated diffusion constants (mean ±1 standard deviation) of the DNA beads in the DNA-only system as a function of salt concentration, based on data from Table 4.2.

The diffusion activity of the proteins as a function of protein radii (for fixed volume fraction) follows an S-curve pattern (Figure 4.6). When the proteins are much larger than the average distance between DNA strands, they become effectively trapped by the mesh of strands, and their diffusion constants correspond with that of the relatively immobile DNA. When the proteins are smaller than the average inter-strand spacing, their diffusion increases to a limit, and it may be the case that their motions become increasingly governed by the macromolecular volume fraction. On the other hand, DNA internal motions did not appear to be affected by differences in the radii of the protein embedded in the DNA matrix. It may be that in the case of tightly-packed bulk DNA that is used in our studies, the effect of steric crowding from proteins on DNA internal motions is insignificant compared to that of DNA internal forces (as long as the volume fraction of proteins remains fixed). A trajectory of the DNA-protein system under 0.1 M conditions with proteins of 4.4 nm radii is shown in Movie S3, which covers a simulation interval of 600 ns.



(a) (b)

Figure 4.4: Full (a) and close-up (b) views of the DNA-only system under simulation. The thickness of the DNA strand is reduced for visualization purposes, and so the inter-strand spacing appears to be larger than that in the actual system under study.

4.4 Properties of the DNA Model

It is well established that the structure of DNA plays an important role in affecting the biological activities of the cell (see Sections 2.1.1, 2.1.4). Thus, it is imperative that the structure and conformation of the DNA model used in this study be quantitatively discussed here. Since this study is focused on the internal motions of DNA as bulk polymer, as opposed to small linear DNA strands or plasmids, our DNA model was prepared from a Hilbert curve conformation, as elongated fractal globules have been hypothesized [94] to be able to replicate the experimental observation that DNA genomic distance to a central loci in *E. coli* is linearly related to spatial distance [93]. This is in contrast to past studies [205], where models of *E. coli* DNA under crowded conditions use a conformation derived from a closed self-avoiding walk.

The estimated average center-to-center spacing between the DNA beads produced by this DNA model is approximately 6.381 nm, leaving $6.381 - (1.59 \cdot 2) = 3.201$ nm of surface-to-surface spacing. The center-to-center spacing nearly matches the

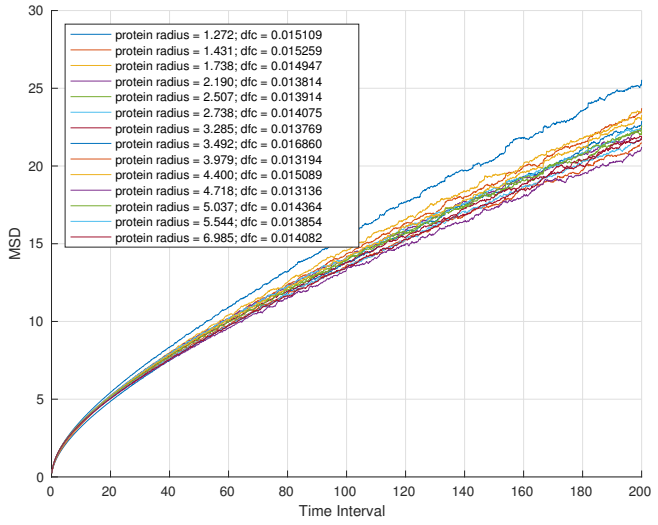
Table 4.3: Estimated diffusion constants for both DNA (beads) and proteins in the DNA-protein system with different-sized proteins, subject to a constant 6% protein volume fraction.

Protein Radius (nm)	**Number of Proteins**	**Estimated DNA (Bead) Diffusion Constant (D_0)**	**Estimated Protein Diffusion Constant (D_0)**
1.272	14397	0.0151	0.3557
1.431	10112	0.0153	0.3604
1.738	5644	0.0149	0.3545
2.190[a]	2821	0.0138	0.3187
2.507	1881	0.0139	0.2790
2.738	1444	0.0141	0.2496
3.285	836	0.0138	0.1638
3.492	696	0.0169	0.1339
3.979	471	0.0132	0.0768
4.400[b]	348	0.0151	0.0504
4.718	283	0.0131	0.0360
5.037	232	0.0144	0.0269
5.544	174	0.0139	0.0171
6.985	87	0.0141	0.0093

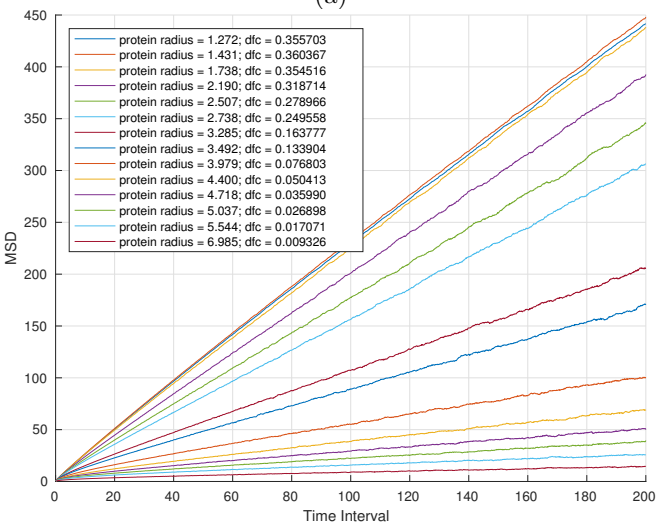
[a] 2.190 nm is comparable to the hydrodynamic radius of proteins with similar weight as the nucleoid-associated protein HU [6].
[b] 4.400 nm is comparable to the hydrodynamic radius of proteins with similar weight as *lac* repressor (LacI) [6].

experimentally-determined $\approx$ 6.4 nm value (Section 2.1.3), while the surface-to-surface spacing of this model is smaller than the corresponding experimental value by approximately 1.2 nm due to the large Stokes radius of the DNA beads in the model. This difference is not expected to affect the results of this study, since the modeled beads are soft beads with the steric interactions tuned to match the hydrodynamic properties of DNA [289, 290].

To quantify the structure of the DNA chain, we counted the number of interactions between loci (beads in this case) along the chain and effectively estimated the contact probability between two loci separated by a given sequential distance s. This is analogous to the measurements obtained by *in vitro* analysis of spatial organization of chromatin using chromosome conformation capture-based techniques [308, 309]. The probability values for s were plotted against the chain distance between the



(a)

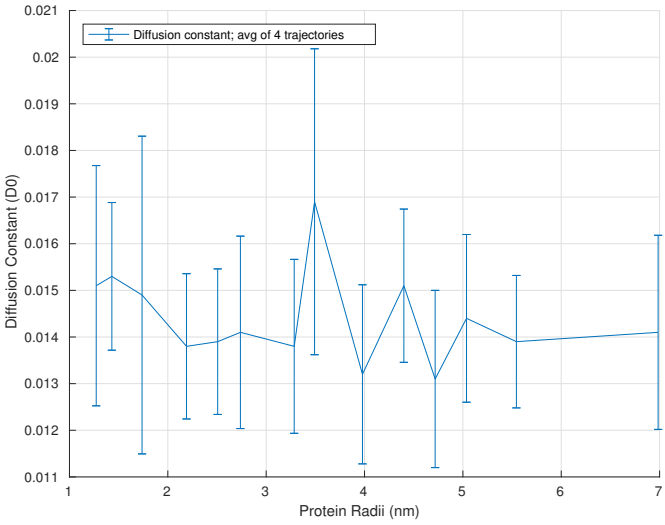


(b)

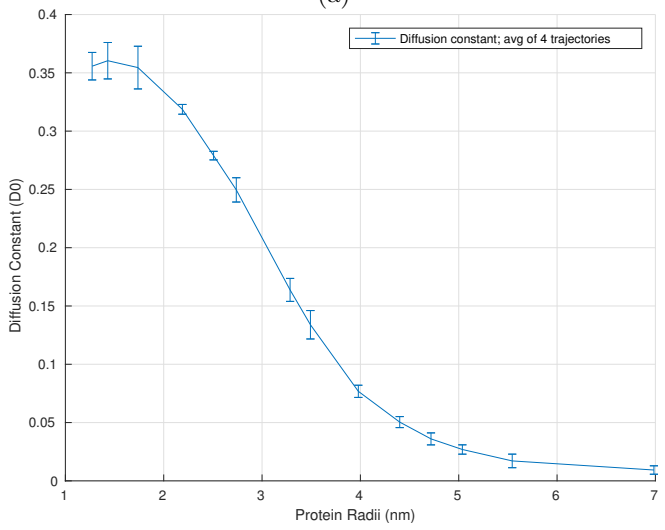
Figure 4.5: MSD curves of DNA (a) and proteins (b), respectively, for BD simulation trajectories of the DNA-protein system with different-sized proteins subject to a constant 6% protein volume fraction.

beads and fitted to the power law $s^{-\gamma}$. Unless otherwise stated, the threshold inter-bead distance, below which two DNA beads are considered to be in contact, was set to be $4a$, where a is the radius of the DNA bead.

A DNA conformation with γ close to 1 is consistent with that of a fractal (crumpled) globule, where genetic distance between two loci in the DNA chain correlates with their 3D spatial distance [88, 310]. For reference, the γ value for eukaryotic (human) DNA has been determined to be close to 1 [89]. On the other hand, a con-



(a)

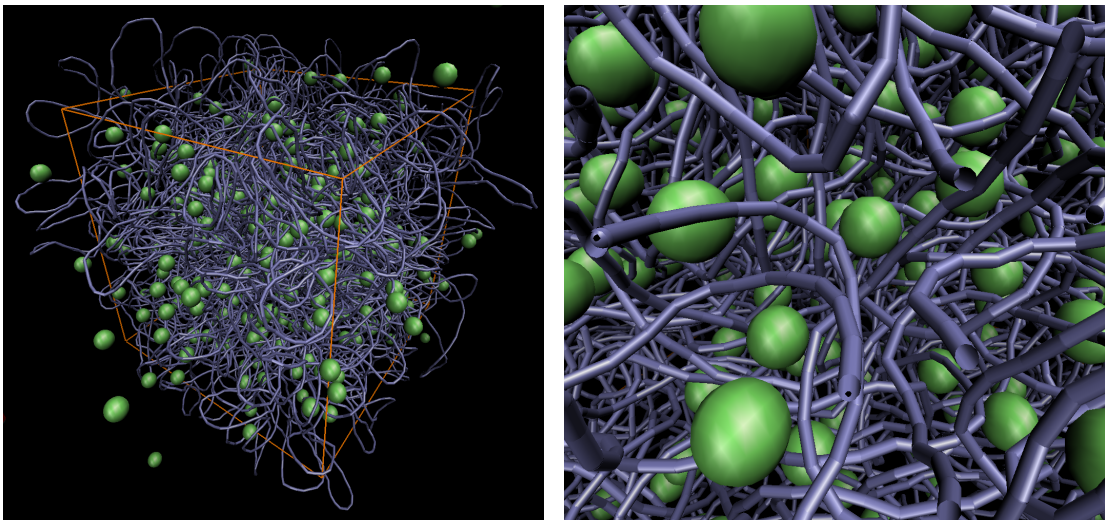


(b)

Figure 4.6: Estimated diffusion constants (mean ±1 standard deviation) of the DNA beads (a) and proteins (b), respectively, in the DNA-protein system as a function of protein size (subject to a constant 6% protein volume fraction), based on data from Table 4.3.mm

formation with γ value close to 1.5 corresponds to that of an equilibrium globule, in which there is no correlation between genetic and spatial distances of two DNA loci (Figure 2.3).

In both of the studies carried out in this work, DNA bead contact probabilities in the initial prepared DNA configuration (Figure 3.4d) were found to be more consistent with those of a fractal globule, with a scaling exponent of $\gamma \approx 1.04$ (Figure 4.8a) for the DNA-only system and $\gamma \approx 1.085$ (Figure 4.9a) for the DNA-protein system with

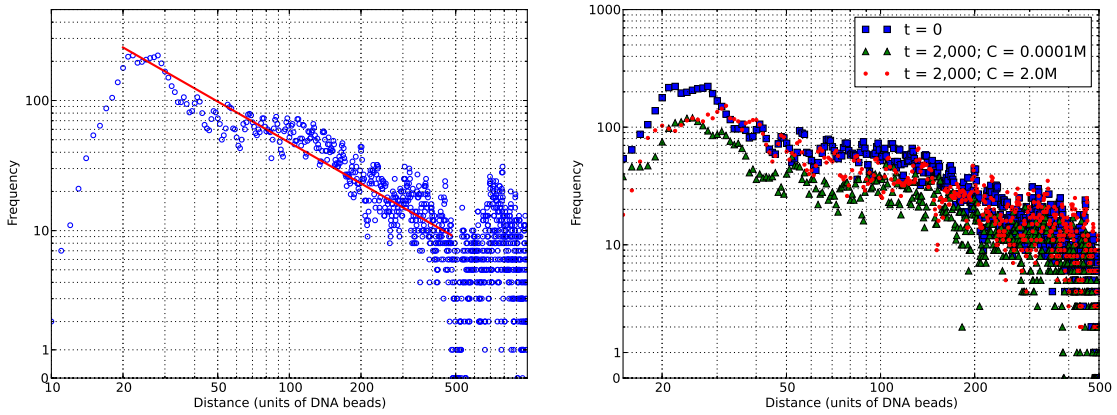


(a) (b)

Figure 4.7: Full (a) and close-up (b) views of a DNA-protein system with proteins of 4.4 nm radii under simulation. The thickness of the DNA strand is reduced for visualization purposes.

proteins having 4.4 nm radii. In contrast, the structure of the DNA model used in [1] was reported to have a γ value that is closer to 1.5. We note that the DNA preparation procedure used in [1] substantially differs from those used in this study. First, the DNA chain in [1] was prepared not as a Hilbert curve (a true fractal), but as a Peano curve [305, 311] that was afterwards modified to make the chain a closed loop in order to avoid chain entanglements [87]. In addition, the prepared chain in [1] was an order of magnitude shorter than the chain prepared in this study, thus lacking the length to build up to a 3D structure with sufficient levels of self-similarity. It is likely that the combination of these two factors account for the difference in γ values of the two generated DNA structures.

For both of the systems studied, the bead contact probabilities yielded similar γ values even when different threshold values were used for defining two beads to be in contact (Figure 4.10), suggesting that the prepared DNA model exhibits a sufficiently deep level of structural self-similarity. Moreover, the γ value for the DNA-only system measured after 4,000 time units (65.6 μs) of simulation under 0.1 M conditions is nearly identical to that of the initial DNA configuration, suggesting that the fractal



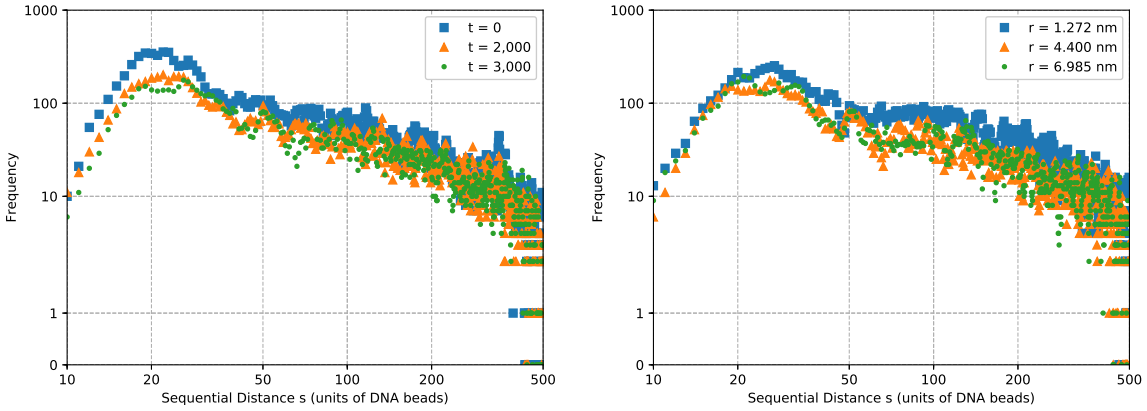
(a) (b)

Figure 4.8: (a): DNA bead contact frequency (proportional to probability) for the initial configuration of the DNA-only system. The solid line corresponds to a power-law scaling exponent of $\gamma = 1.04$. (b): DNA bead contact frequencies for the same system after 2,000 units of simulation time at 10^{-4} M ($\gamma \approx 0.99$) and 2.0 M ($\gamma \approx 1.02$) conditions.

structure may be long-lived for at least up to the time ranges explored by our computer studies (Figure 4.10b). Similar findings were observed in the DNA-protein system with proteins having 4.4 nm radii (Figure 4.9a). In light of the DNA preparation choices made in [1], the ability of the model prepared in this study to maintain a fractal globule for very long time intervals also suggests that, for the simulated time ranges at least, the formation of knots and other entanglements in an open-ended chain might not be a strong concern if the structure is sufficiently large and fractal.

In the studies involving the DNA-only system, it was observed that the DNA bead contact probabilities after simulation do not appear to be affected by differences in salt concentration levels, i.e. the peaks of the frequency plots and the γ values were found to be nearly the same (Figure 4.8b), at least for the time scales used in our studies. In combination with our findings regarding DNA internal motions over different salt concentration levels, this observation suggest that for bulk DNA, salt concentration predominantly affects the internal motions of DNA (kinetics), but not its overall structure (equilibrium).

Similarly, for the range of protein radii used in the studies involving the DNA-protein system, the DNA contact probabilities did not appear to be affected by size



(a) (b)

Figure 4.9: (a): DNA bead contact frequencies for a DNA-protein system containing proteins with 4.400 nm radii after 0, 2,000, and 3,000 units of simulation time at 0.1 M. The γ values are estimated to be 1.085, 0.9862, and 0.9720, respectively. (b): DNA bead contact frequencies after 2,000 units of simulation time for systems containing proteins with 1.272 nm ($\gamma \approx 0.99$) and 6.985 nm ($\gamma \approx 1.02$) radii.

differences of the proteins that were embedded within its structure (Figure 4.9b), suggesting that the internal motions of DNA allow it to accommodate proteins flexibly without losing its fractal structure. However, this study was carried out under a constant 6% protein volume fraction, and so it may be possible instead that there is a dependence of DNA contact probabilities on volume fraction.

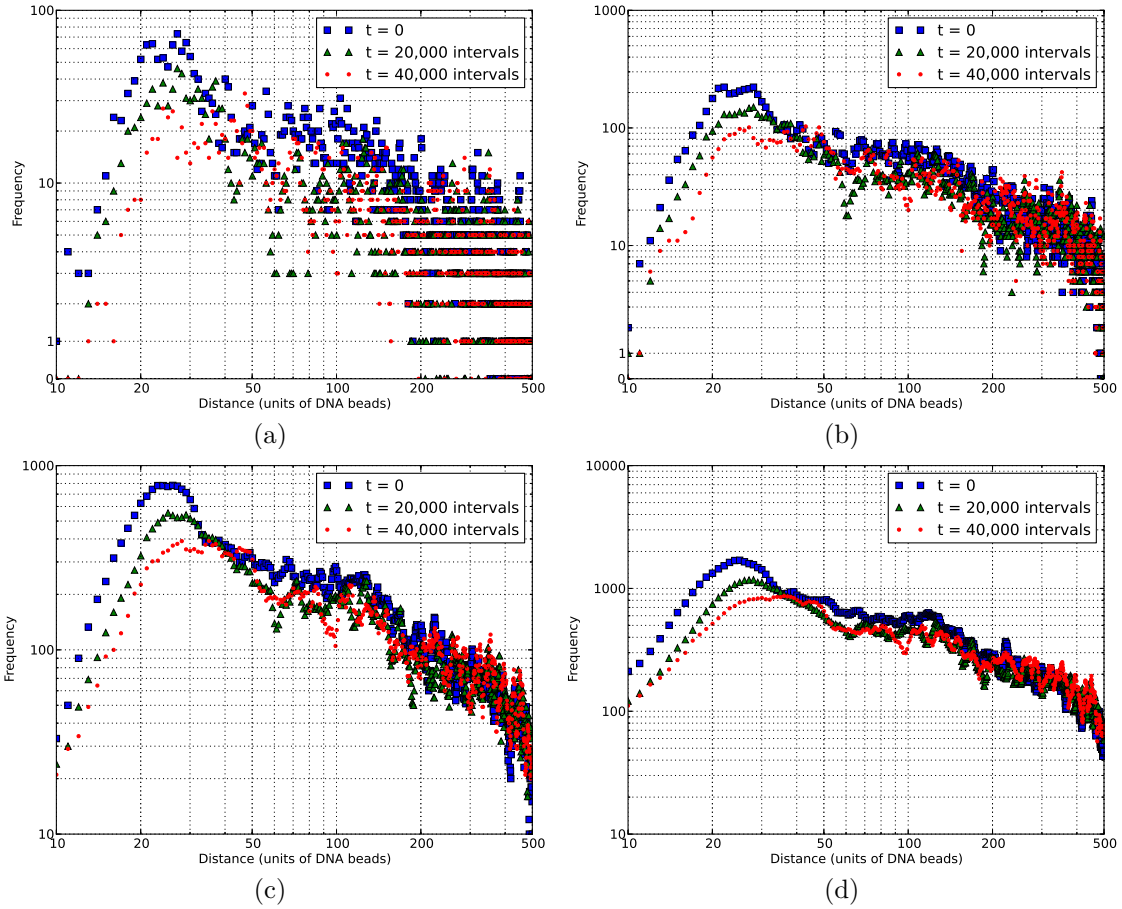


Figure 4.10: DNA bead contact frequencies for the DNA-only system after 2,000 and 4,000 units of simulation time under 0.1 M conditions, using $3.0a$ (a), $4.0a$ (b), $6.0a$ (c), and $8.0a$ (d) as the defined threshold for two beads to be considered in contact.

CHAPTER 5
DISCUSSION & FUTURE DIRECTIONS

The goal of this study was to evaluate the role of the physicochemical environment of the nucleoid on DNA internal motions. Specifically, two aspects were chosen as parameters of study, which are salt concentration and protein size. To begin to address this goal, we built a coarse-grained model of the DNA as a fractal chain of beads in systems with and without model proteins, and carried out diffusion studies through Brownian dynamics simulations with this model. A discussion of our findings is presented, along with suggestions for next steps in the direction of this research.

5.1 Role of Salt Concentration on DNA Internal Motions

In the first study, we examined the effect of salt concentration on DNA internal motions, and found that the diffusion constant nearly doubled as the salt concentration increased from 10^{-4} to 2.0 M. The most substantial variation in DNA diffusion behavior was observed in the 0.1 M to 0.5 M salt concentration range (Figure 4.3), which suggests that salt concentrations near 0.1 M can regulate DNA internal motions. This finding may be significant for several reasons. First, existing literature shows that intracellular concentration of ions in prokaryotes is at least 0.1 M [6, 312], suggesting that there may be biological significance to this value. Second, studies involving DNA supercoiling, another property that emerges from bulk DNA, demonstrate that supercoil formations are highly dependent on the salt concentration of its environment [135, 313, 314]. In particular, the critical point associated with rapid collapse of DNA from loose to tight supercoils appears to exist near 0.1 M [125]. While not explicitly modeled in this study, supercoils have been shown to form bubbles along the double-stranded DNA due to the torsional tension energy stored [58],

and in doing so, influence the transcription process and consequently gene regulation [59]. Though not well understood, it has been previously suggested that these supercoil formations themselves contribute to the DNA internal motions [313, 315, 316]. Finally, the activity of transcription and other structural proteins that bind DNA, such as nucleoid-associated protein HU, appears to be active at these salt concentration levels and promoted in highly negative-supercoiled DNA sites [58, 59, 317, 318]. If these observations are put together, then the role of salt concentration on DNA internal motions can be hypothesized as follows: at increased salt concentrations, the negative charges along the DNA backbone phosphates become effectively screened, which permits greater internal motions as well as closer contacts between the DNA strands, allowing for tight supercoiling. Tight supercoiling in turn opens up bubbles along the DNA strand to expose certain sites for interaction with proteins, while the internal dynamics of bulk DNA transport the proteins across this matrix to these sites, effectively complementing the promotion of biological activities in the nucleoid, e.g. transcription or replication. Verification or disproof of this hypothesis by future work in this area will contribute to our understanding of one of the key biological processes in the nucleoid, i.e. protein diffusion (see Section 2.1.8).

Since the results of the first study suggest that electrostatic interactions can be very influential to DNA dynamics in the crowded nucleoid, it may be useful for future studies to consider an electrostatic model that includes more explicit features than what is offered by the Debye-Hückel approximation, in order to better capture and understand the effects of salt on DNA diffusive motions. Namely, the model should take into account both the identity of the cations in solution as well as their valences, since both factors have been previously suggested to affect the formation of secondary-level structures of the DNA such as DNA supercoils (see Section 2.1.5).

5.2 Role of Protein Size on DNA and Protein Diffusive Motions

In the second study, we examined whether or not a protein's size had an effect on either DNA and/or protein diffusive motions in the nucleoid, and found that the rate of protein diffusion through the matrix decreases as a function of increasing protein size. One important observation made is that while large proteins are effectively trapped in the cages formed by the DNA matrix, proteins with diameters smaller than the average surface-to-surface DNA inter-strand spacing observe increases in their diffusion constants only to a certain limit, i.e. they do not behave as if the DNA strands were non-existent. To quantify this behavior, we carried out a simulation protocol similar to that specified for the DNA-protein studies, but for a protein-only system. In this short study, the protein volume fraction was kept at a constant 19%, this value being the sum of the DNA and protein volume fractions used in the DNA-protein study. For proteins with small radii, the results indicate an up to three-fold difference in diffusion constants between proteins diffusing inside a matrix of DNA strands and proteins diffusing in a similar setting where the DNA chain is replaced by other freely-diffusing proteins (Figure 5.1). Moreover, protein motions in this scenario appeared to decrease exponentially as the proteins' sizes are increased. The time scales of diffusion between the DNA and protein appear to be different enough that the DNA chain may appear to the proteins as a relatively-fixed mesh blocking their movements. From an evolutionary biology standpoint, the relatively-fixed cage-like nature of the DNA matrix may be a reason why DNA-involved biological processes such as transcription tend to involve complex assemblies of many small proteins at the site of interest on the DNA as opposed to fewer but larger proteins.

The results of the second study also suggest that DNA internal motions remain relatively unaffected by the size of the proteins diffusing within it, assuming a constant protein volume fraction. This observation is likely a result of both the properties of

the DNA employed, as well as the choice of low protein volume fraction. For tightly-packed bulk DNA, the diffusive motions of the DNA appear to be dominated primarily by the steric and bend forces imposed by the DNA model, and unless there is a dense-enough amount of proteins embedded in the matrix, the effects of steric crowding imposed by the protein onto DNA relative to those imposed internally by the DNA model are likely to be minimal at best. Thus, it may be more beneficial instead to investigate the effects of protein volume fraction on DNA internal motions, if any, as a follow-up to this work. This would be synonymous to investigating crowding effects in the nucleoid, though it should be noted that crowding in the nucleoid can be very different from crowding in the cytoplasm, where the freely diffusing macromolecules are not blocked by a relatively fixed and dense cage-like polymer, as suggested by the results of the aforementioned short study of a protein-only system.

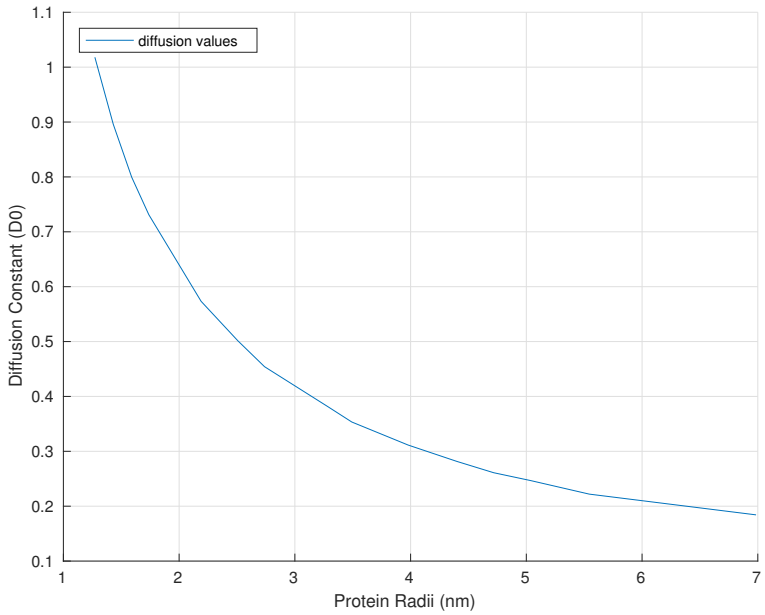


Figure 5.1: Estimated diffusion constants of proteins in a protein-only system as a function of protein size (subject to a constant 19% protein volume fraction).

5.3 DNA Model

In both studies, the prepared DNA-only and DNA-protein systems exhibited a tendency to maintain its fractal structure, at least for the microsecond-level time spans covered by the simulations. Moreover, the fractal property appeared to be maintained across a range of salt concentrations from 10^{-4} to 2.0 M, and also did not appear to be perturbed by the size of the proteins embedded within it, at least for proteins with radii ranging from 30% to 110% of the average inter-strand spacing distance (while the protein volume fraction was maintained at 6%). There are two possible factors behind the ability of this DNA model to maintain its fractal property. The first is that the model was constructed from a Hilbert curve prior to compression into its final compact form, and that the compression process was able to preserve the basic fractal property of the chain the chain's fractal property. The second is that the crowded conditions of the prepared system, along with the strong stretch and bend forces that are parameterized into the model, likely presents an unfavorable conformational and energy barrier that the DNA chain must overcome in order to evolve its conformation from that of a fractal globule to equilibrium globule. This corresponds to a previously-made proposal that the formation and maintenance of "chromosomal territories" in eukaryotic DNA requires topological constraints [319].

The original theory of the fractal globule model [88] suggested that the lifetime of a fractal globule is determined by the time $t \sim \mathcal{O}(N^3)$ required to thread the ends of the polymer through the globule, where N is the bead length of the polymer chain. However, prior simulations have shown that the actual times required for large DNA globules to converge to the equilibrium globule can actually exceed this asymptotic bound [94], with estimates suggesting that the process can take over 500 years [52]. This lifetime also depends on the stringency of the topological constraints, which can be enhanced by the presence of crosslinking proteins or violated by the presence

of DNA topology-altering DBPs such as DNA topoisomerase II (topo II) [94]. The proteins modeled in this work, however, do not interact with the DNA chain beyond sterically clashing against it. In either case, it is probably reasonable to claim that, for the time scales generally used to study DNA and protein diffusive motions in the nucleoid, the structure remains sufficiently fractal for the model to be useful.

There are some questions about the DNA model constructed for this work that remain open. The first question is whether or not chain entanglements will occur under this open-ended DNA chain model. Though no entanglements have been observed in the simulations carried out, there is no guarantee that this will not happen given enough simulation time. However, similar to the case of possible degeneration of the DNA chain into an equilibrium globule, it may be possible that the setup of the DNA model also makes entanglements and knot formations energetically-prohibitive.

The second question is whether or not the properties observed in the model constructed in this work can hold for shorter DNA chains, and if so, what is the lower-bound length before the chain cannot longer maintain its fractal form for a reasonable amount of time. The DNA chain in our work was constructed to fill up the volume of a cube while having at least the length of the model chain from [1] (6,912 beads long), but if the same properties can hold for a shorter model, then it would justify the use of a smaller system in future studies to reduce the time required for running simulations of the DNA chain. All of these questions should be addressed in future studies in order to refine and provide a better-understood model of DNA for which to perform future coarse-grained *in silico* DNA studies against.

Despite the efforts to reproduce a faithful model of *in vivo* DNA, the model itself is very rudimentary and fails to answer the most basic questions regarding DNA and proteins in the nucleoid. These questions include: how is the entropically-low fractal conformation of DNA achieved *in vivo* in the first place? Why is the formation of a fractal globule preferred to an equilibrium globule? What are the

mechanisms through which the DNA's fractal conformation is maintained over time? If the fractal conformation is largely maintained by crosslinking proteins binding to multiple strands, does their presence effectively slow down overall DNA diffusive motions? In addition, the model does not contain finer-grained features that can reproduce secondary structures found in *in vivo* DNA, such as DNA supercoils, which have been suggested to contribute to DNA internal motions [315, 316]. In fact, despite the amount of work on DNA secondary structures and topologies (see Section 2.1.4), it is still not clear yet how they fit into the current understanding of fractal DNA globules to provide a coherent description of DNA structure. These questions will need to be addressed by future work in order to grasp a better picture of the role of DNA structure and dynamics in biological processes involving the DNA.

5.4 Protein Model

While the results of the second study proved to be helpful for understanding the effects of protein size on its diffusion through the nucleoid, the soft-bead protein model used in the study needs several improvements to be able to provide better insights on the role of proteins in affecting DNA diffusive motions. In particular, the biggest observed drawback to the current model is its coarse-grained nature, in which proteins are represented by a single large soft-bead. Though single-bead approximations have been shown to be useful for analyzing *in vivo* macromolecular diffusion in the cytoplasm [6], the model ultimately does not reflect how proteins look and behave *in vivo*, and so fails to capture the microscopic details of proteins and protein-DNA interactions that would affect DNA internal forces. One consequence of this choice of model is that it is not possible to model proteins with complex shapes, nor proteins whose topologies constantly change as they diffuse through the DNA. Though ephemeral, changes in a protein's shape can be the difference between being able to diffuse through a gap between the strands and being trapped by the

cage imposed by the strands, since many DBPs such as LacI have sizes that are on the same order as the average inter-strand spacing of DNA in the nucleoid. A straightforward extension to the model to address this would be to fine-grain the representation of the protein from a single soft-bead to an assembly of smaller soft-beads held together by stretch and bend interactions. However, the optimal level of granularity that should be represented in a multi-bead assembly model and the force parameter values required to reproduce the *in vivo* behavior of proteins are neither well-understood nor agreed upon.

Protein electrostatics, whose fine-tuning are essential for protein binding and function [320], are another aspect of proteins that was not reflected in the protein model used in our studies. While the first study focused on solvent salt concentrations in affecting DNA internal motions, the presence of salt will likely have an effect on protein diffusive motions as well depending on the proteins' charge distributions. Electrostatic interactions are known to have large and pronounced effects on macromolecules in the cellular environment, especially under crowded conditions (see Section 2.1.5), and likewise, the presence of DNA-protein electrostatic interactions will likely affect DNA diffusive motions. Modeling electrostatics and partial charges in proteins can be achieved, for example, by leveraging the aforementioned finer-grained representation of the proteins to assign whole charges to the individual soft-beads that comprise a protein assembly.

Related to protein electrostatics that was also not considered in the protein model is the DNA site-specificity of DNA-binding proteins. Even if the bulk DNA in the nucleoid was assumed to be uniformly accessible for proteins to diffuse through (i.e. not compacted in certain regions), it is not likely for proteins to be uniformly dispersed across the matrix, since DBPs in general have binding preferences to certain sequences and sites along the DNA. This observation suggests that the diffusion constant of a protein near its preferred binding sequence can be lowered if its affinity to the site is

strong enough to prevent or delay its diffusive "escape" to another region of the DNA matrix. Similarly, the presence of a protein bound to a DNA segment will highly likely impact that segment's diffusive activity, due to the add-on weight imposed by the latched protein that effectively deters the mobility of the segment in solution. These phenomenon are not well understood yet, and so future *in silico* studies in this area should at least take DNA site-specificity into account when modeling the proteins.

5.5 Solvent Model

Our computer studies were performed on a coarse-grained system where the solvent is implicitly modeled. This solvent model needs to be improved to provide better accuracy in reproducing the *in vivo* fluid dynamics that both DNA and protein motions are affected by, since it is well known that the properties of a fluid on the nanometer scale, which is the scale of observation covered by this work, are different from those in the bulk [321, 322]. In particular, we did not take into account the HI experienced by the particles in our coarse-grained model. While earlier simulation studies have demonstrated that HI, along with crowding, likely dominates in-vivo macromolecular motion [6], it should be noted that in both the DNA-only and DNA-protein systems studied, the stretch and bend forces dominate the dynamics of highly compressed DNA, and so HI would appear to play a larger role in DNA dynamics only if DNA occupied a much higher volume fraction in the nucleoid (i.e. at least cytoplasm levels of 20–40%) and the DNA chain was initially set to be in a relaxed un-crumpled state. On the other hand, HI has been previously shown to reduce the motion of the proteins embedded in the compacted DNA chain by approximately a factor of 2 [1], confirming that the effects of HI on freely-diffusing macromolecules cannot be be neglected in the crowded environment of the nucleoid.

Alternative simulation methodologies are required if HI is to be accounted for in

future studies. Brownian dynamics with HI according to the McCammon algorithm [246] is usually the methodology of choice, but the HI in BD is highly simplified and generally incorporate only isolated body resistance. Stokesian dynamics (SD) is another option and is known for its rigorous inclusion and accurate handing of many-body HI, but can be computationally expensive – the classical SD algorithm requires $\mathcal{O}(N^3)$ operations [264], though subsequent developments have reduced the computational complexity of the method down to $\mathcal{O}(N^{1.25} \log N)$ [323, 324]. More recently, a hybrid method has been developed to capture the best of both approaches, in which average hydrodynamic effects computed using a mean field approach are included into the Brownian dynamics algorithm [325, 326]. This approach, hereafter referred to as the Mean Field Treatment (MFT) method, has been shown to accurately reproduce equilibrium properties of concentrated protein solutions as well as the decreased diffusive behavior of proteins caused by the crowding. The speed and accuracy of MFT would make the method applicable to very large-scale simulations of crowded solutions, such as the nucleoid in the case of this work.

5.6 Other Considerations

The model of the nucleoid used in our work is ideal and naïve at best even if the suggested improvements to the DNA, protein, and solvent models were included, because it does not fully reflect the dynamics of the nucleoid in the living cell.

First, we did not consider a model where the DNA conformations are changing as they normally do throughout the cell cycle [52]. Because replication and separation of the duplicate DNA strands requires the structure of DNA to change or break down significantly to accommodate this event, it is highly likely that an accurate simulation of the nucleoid as it undergoes such an event will provide us the crucial insights needed to understand the origins of the DNA's ability to build and maintain its fractal property in the first place. Second, we did not consider in our model intranuclear

proteins that cleave and re-join DNA segments, such as DNA topoisomerase II [327]. While topoisomerases are known to simplify the DNA topology at the strand and knot level [328], it is not clear how their activities can selectively bring about changes to the overall topology of large DNA chains [329], and future work with a model that takes into account DNA cleavages and re-joins will provide better details of the mechanisms through which DNA structure can form *in vivo*. Finally, while we have discussed possible improvements to the DNA model to account for secondary- and higher-level structures, we have not considered the role of external factors in the *in vivo* nucleoid environment that actively bend or block the motions of DNA, such as the presence of nucleoid-associated protein HU. While biophysical studies involving the nucleoid have included the treatment of crowder and blocker proteins [188, 197], they have mostly focused on understanding their effects on facilitated diffusion, and there have been virtually no concentrated efforts to specifically study the roles and effects of these proteins on DNA internal motions, if any. Such investigation may prove to be useful to understanding the flexibility of DNA dynamics in response to the physicochemical changes of its surrounding environment.

5.7 Software Development & Tooling

The implementation of a custom Brownian dynamics package for the purpose of this work came with its set of challenges, and one aspect of this project that could be drastically improved is the software tooling that is available for streamlining the progress of work.

First, there was an overwhelming need to write custom scripts in MATLAB [330], often from scratch, for ad hoc trajectory analyses. MATLAB was initially chosen since it came with a set of mathematical libraries useful for computing bulk statistical values. However, it appears that tools and libraries for trajectory analyses are already abundant but are generally published for more widely available programming

language environments such as Python [331]; see [332–334] for examples. While using these libraries requires that the input trajectories conform to an existing de facto file format, such as `.TRR` or `.XTC` [335], adding support into `BDT` for writing out to these formats is almost certain to be worth the investment, since existing molecular dynamics trajectory analysis tools like MDTraj [336] can subsequently be leveraged to perform both standard and ad hoc analyses, such as computing the MSD values over a range of time values or selecting certain trajectory frames to compute a specific system property, respectively.

Second, the agility of software development was much lower than desired as a result of the choice of implementing `BDT` in C++. While C++ is a high performance language that is able to compile down into very optimized programs, which is necessary for running simulations, the development productivity and maintenance costs associated with using C++ suffers from the lack of utility libraries, lack of in-language debugging facilities, build system non-portability across different computer platforms and compilers, and the absence of a package system for the language. Consequently, monumental effort is required for common development tasks such as adding new code, refactoring code, or importing an external library. Over the last few years, however, the Rust programming language [337] has gained attention as a promising programming language that advertises many of the features and benefits traditionally claimed only by C++, such as speed as demanded by `BDT`'s use case, while offering features not offered by C++ such as compiler-guaranteed memory safety and thread safety. More importantly, Rust comes with a strong tooling ecosystem around the language, such as its build and package systems, thus making the case to be a viable and ergonomic alternative to C++ for the development of high performance simulation software. Migration of `BDT` to Rust appears to be a promising direction for the evolution of the software package, and in fact, work has already begun on a re-write of `BDT` in Rust to investigate this direction.

REFERENCES

[1] E. Chow and J. Skolnick, "DNA internal motion likely accelerates protein target search in a packed nucleoid," *Biophysical Journal*, vol. 112, no. 11, pp. 2261–2270, 2017.

[2] S. Saberi and E. Emberly, "Chromosome driven spatial patterning of proteins in bacteria," *PLOS Computational Biology*, vol. 6, no. 11, pp. 1–10, Nov. 2010.

[3] A. D. Riggs, S. Bourgeois, and M. Cohn, "The *lac* represser-operator interaction: III. kinetic studies," *Journal of Molecular Biology*, vol. 53, no. 3, pp. 401–417, 1970.

[4] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, "How a protein searches for its site on DNA: The mechanism of facilitated diffusion," *Journal of Physics A: Mathematical and Theoretical*, vol. 42, no. 43, p. 434 013, 2009.

[5] A. Tafvizi, L. A. Mirny, and A. M. van Oijen, "Dancing on DNA: Kinetic aspects of search processes on DNA," *ChemPhysChem*, vol. 12, no. 8, pp. 1481–1489, 2011.

[6] T. Ando and J. Skolnick, "Crowding and hydrodynamic interactions likely dominate *in vivo* macromolecular motion," *Proceedings of the National Academy of Sciences*, vol. 107, no. 43, pp. 18 457–18 462, 2010.

[7] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[8] H. Lodish, *Molecular Cell Biology*. New York: W.H. Freeman, 2000, ISBN: 0-7167-3136-3.

[9] B. Alberts, *Molecular Biology of the Cell*. New York: Garland Science, 2008, ISBN: 978-0-8153-4105-5.

[10] R. E. Franklin and R. G. Gosling, "Molecular configuration in sodium thymonucleate," *Nature*, vol. 171, no. 4356, pp. 740–741, 1953.

[11] W. Fuller, W. Wilkins, H. Wilson, and L. Hamilton, "The molecular configuration of deoxyribonucleic acid," *Journal of Molecular Biology*, vol. 12, no. 1, 60–IN9, 1965.

[12] S. Arnott, R. Chandrasekaran, D. L. Birdsall, A. G. W. Leslie, and R. L. Ratliff, "Left-handed DNA helices," *Nature*, vol. 283, no. 5749, pp. 743–745, 1980.

[13] S. Arnott, R. Chandrasekaran, D. Hukins, P. Smith, and L. Watts, "Structural details of a double-helix observed for DNAs containing alternating purine and pyrimidine sequences," *Journal of Molecular Biology*, vol. 88, no. 2, pp. 523–533, 1974.

[14] D. Marvin, M. Spencer, M. Wilkins, and L. Hamilton, "The molecular configuration of deoxyribonucleic acid III. X-ray diffraction study of the *C* form of the lithium salt," *Journal of Molecular Biology*, vol. 3, no. 5, 547–IN14, 1961.

[15] W. Saenger, *Principles of Nucleic Acid Structure*. Springer New York, 1984.

[16] V. N. Potaman and R. R. Sinden, "DNA," in *DNA Conformation and Transcription*. Boston, MA: Springer US, 2005, pp. 3–17, ISBN: 978-0-387-29148-2.

[17] Wikipedia contributors, *A-DNA — Wikipedia, the free encyclopedia*, [Online; accessed 21-August-2018], 2018.

[18] D. Flatters, M. Young, D. L. Beveridge, and R. Lavery, "Conformational properties of the TATA-box binding sequence of DNA," *Journal of Biomolecular Structure and Dynamics*, vol. 14, no. 6, pp. 757–765, 1997.

[19] G. Guzikevich-Guerstein and Z. Shakked, "A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein," *Nature Structural Biology*, vol. 3, no. 1, pp. 32–37, 1996.

[20] V. I. Ivanov, L. E. Minchenkova, B. K. Chernov, P. McPhie, S. Ryu, S. Garges, A. M. Barber, V. B. Zhurkin, and S. Adhya, "CRP-DNA complexes: Inducing the *A*-like form in the binding sites with an extended central spacer," *Journal of Molecular Biology*, vol. 245, no. 3, pp. 228–240, 1995.

[21] A. Jacobo-Molina, J. Ding, R. G. Nanni, A. D. Clark, X. Lu, C. Tantillo, R. L. Williams, G. Kamer, A. L. Ferris, and P. Clark, "Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA," *Proceedings of the National Academy of Sciences*, vol. 90, no. 13, pp. 6320–6324, 1993.

[22] J. R. Kiefer, C. Mao, J. C. Braman, and L. S. Beese, "Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal," *Nature*, vol. 391, no. 6664, pp. 304–307, 1998.

[23] M. M. Becker and Z Wang, "B → A transitions within a 5 S ribosomal RNA gene are highly sequence-specific," *Journal of Biological Chemistry*, vol. 264, no. 7, pp. 4163–7, 1989.

[24] S. C. Mohr, N. V. Sokolov, C. M. He, and P. Setlow, "Binding of small acid-soluble spore proteins from *Bacillus subtilis* changes the conformation of DNA from B to A," *Proceedings of the National Academy of Sciences*, vol. 88, no. 1, pp. 77–81, 1991.

[25] S. C. Ha, K. Lowenhaupt, A. Rich, Y.-G. Kim, and K. K. Kim, "Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases," *Nature*, vol. 437, no. 7062, pp. 1183–1186, 2005.

[26] A. Rich and S. Zhang, "Z-DNA: The long road to biological function," *Nature Reviews Genetics*, vol. 4, no. 7, pp. 566–572, 2003.

[27] B. Wittig, T. Dorbic, and A. Rich, "Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2259–2263, 1991.

[28] G. P. Schroth, P. J. Chou, and P. S. Ho, "Mapping Z-DNA in the human genome. computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes," *Journal of Biological Chemistry*, vol. 267, no. 17, pp. 11 846–11 855, 1992.

[29] B. Wong, S. Chen, J.-A. Kwon, and A. Rich, "Characterization of Z-DNA as a nucleosome-boundary element in yeast *Saccharomyces cerevisiae*," *Proceedings of the National Academy of Sciences*, vol. 104, no. 7, pp. 2229–2234, 2007.

[30] J. Casasnovas and F. Azorin, "Supercoiled induced transition to the Z-DNA conformation affects the ability of a d(CG/GC)12 sequence to be organized into nucleosome-cores," *Nucleic Acids Research*, vol. 15, no. 21, pp. 8899–8918, 1987.

[31] R. T. Simpson, "Nucleosome positioning: Occurrence, mechanisms, and functional consequences," in *Progress in Nucleic Acid Research and Molecular Biology*, Elsevier, 1991, pp. 143–184.

[32] P. C. Champ, S. Maurice, J. M. Vargason, T. Camp, and P. S. Ho, "Distributions of Z-DNA and nuclear factor I in human chromosome 22: A model for coupled transcriptional regulation," *Nucleic Acids Research*, vol. 32, no. 22, pp. 6501–6510, 2004.

[33] F. Kriegel, N. Ermann, and J. Lipfert, "Probing the mechanical properties, conformational changes, and interactions of nucleic acids with magnetic tweezers," *Journal of Structural Biology*, vol. 197, no. 1, pp. 26–36, 2017.

[34] S. B. Smith, Y. Cui, and C. Bustamante, "Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules," *Science*, vol. 271, no. 5250, pp. 795–799, 1996.

[35] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, "Structural transitions and elasticity from torque measurements on DNA," *Nature*, vol. 424, no. 6946, pp. 338–341, 2003.

[36] F. Mosconi, J. F. Allemand, D. Bensimon, and V. Croquette, "Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers," *Physical Review Letters*, vol. 102, no. 7, 2009.

[37] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, "Magnetic torque tweezers: Measuring torsional stiffness in DNA and RecA-DNA filaments," *Nature Methods*, vol. 7, no. 12, pp. 977–980, 2010.

[38] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle," *Recueil des Travaux Chimiques des Pays-Bas*, vol. 68, no. 12, pp. 1106–1122, 2010.

[39] G. Strobl, *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior*. Berlin New York: Springer, 2007, ISBN: 3-540-25278-9.

[40] H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom, and P. A. Wiggins, "Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity," *Biopolymers*, vol. 85, no. 2, pp. 115–130, 2007.

[41] J. Bednar, P. Furrer, V. Katritch, A. Stasiak, J. Dubochet, and A. Stasiak, "Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA," *Journal of Molecular Biology*, vol. 254, no. 4, pp. 579–594, 1995.

[42] M. Mandelkern, J. G. Elias, D. Eden, and D. M. Crothers, "The dimensions of DNA in solution," *Journal of Molecular Biology*, vol. 152, no. 1, pp. 153–161, 1981.

[43] M. Vologodskaia and A. Vologodskii, "Contribution of the intrinsic curvature to measured DNA persistence length," *Journal of Molecular Biology*, vol. 317, no. 2, pp. 205–213, 2002.

[44] S. Geggier, A. Kotlyar, and A. Vologodskii, "Temperature dependence of DNA persistence length," *Nucleic Acids Research*, vol. 39, no. 4, pp. 1419–1426, 2010.

[45] R. Grosschedl, "Higher-order nucleoprotein complexes in transcription: Analogies with site-specific recombination," *Current Opinion in Cell Biology*, vol. 7, no. 3, pp. 362–370, 1995.

[46] C. Bouchiat, M. Wang, J.-F. Allemand, T. Strick, S. Block, and V. Croquette, "Estimating the persistence length of a worm-like chain molecule from force-extension measurements," *Biophysical Journal*, vol. 76, no. 1, pp. 409–413, 1999.

[47] T. Odijk, "Osmotic compaction of supercoiled DNA into a bacterial nucleoid," *Biophysical Chemistry*, vol. 73, no. 1-2, pp. 23–29, 1998.

[48] D. S. Goodsell, "Inside a living cell," *Trends in Biochemical Sciences*, vol. 16, pp. 203–206, 1991.

[49] F. R. Blattner, "The complete genome sequence of *Escherichia coli* K-12," *Science*, vol. 277, no. 5331, pp. 1453–1462, 1997.

[50] D. S. Goodsell, "*Escherichia coli*," *Biochemistry and Molecular Biology Education*, vol. 37, no. 6, pp. 325–332, 2009.

[51] A. Vendeville, D. Larivière, and E. Fourmentin, "An inventory of the bacterial macromolecular components and their spatial organization," *FEMS Microbiology Reviews*, vol. 35, no. 2, pp. 395–414, 2011.

[52] A. Rosa and R. Everaers, "Structure and dynamics of interphase chromosomes," *PLoS Computational Biology*, vol. 4, no. 8, S. Henikoff, Ed., e1000153, 2008.

[53] X. S. Xie, P. J. Choi, G.-W. Li, N. K. Lee, and G. Lia, "Single-molecule approach to molecular biology in living bacterial cells," *Annual Review of Biophysics*, vol. 37, no. 1, pp. 417–444, 2008.

[54] T. Ali Azam, A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama, "Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid," *J. Bacteriol.*, vol. 181, no. 20, pp. 6361–6370, 1999.

[55] D. J. Sherratt, "Bacterial chromosome dynamics," *Science*, vol. 301, no. 5634, pp. 780–785, 2003.

[56] V. F. Holmes and N. R. Cozzarelli, "Closing the ring: Links between SMC proteins and chromosome partitioning, condensation, and supercoiling," *Pro-*

ceedings of the National Academy of Sciences, vol. 97, no. 4, pp. 1322–1324, 2000.

[57] J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, "The twisted circular form of polyoma viral DNA," *Proceedings of the National Academy of Sciences*, vol. 53, no. 5, pp. 1104–1111, 1965.

[58] J.-H. Jeon, J. Adamcik, G. Dietler, and R. Metzler, "Supercoiling induces denaturation bubbles in circular DNA," *Physical Review Letters*, vol. 105, no. 20, 2010.

[59] H. M. Lim, D. E. A. Lewis, H. J. Lee, M. Liu, and S. Adhya, "Effect of varying the supercoiling of DNA on transcription and its regulation," *Biochemistry*, vol. 42, no. 36, pp. 10 718–10 725, 2003.

[60] L. Postow, C. D. Hardy, J. Arsuaga, and N. R. Cozzarelli, "Topological domain structure of the *Escherichia coli* chromosome," *Genes & Development*, vol. 18, no. 14, pp. 1766–1779, 2004.

[61] G. J. Pruss and K. Drlica, "DNA supercoiling and prokaryotic transcription," *Cell*, vol. 56, no. 4, pp. 521–523, 1989.

[62] B. E. Funnell, T. A. Baker, and A Kornberg, "Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome," *Journal of Biological Chemistry*, vol. 261, no. 12, pp. 5616–5624, 1986.

[63] H. A. Nash, "Bending and supercoiling of DNA at the attachment site of bacteriophage λ," *Trends in Biochemical Sciences*, vol. 15, no. 6, pp. 222–227, 1990.

[64] Y. Liu, V. Bondarenko, A. Ninfa, and V. M. Studitsky, "DNA supercoiling allows enhancer action over a large distance," *Proceedings of the National Academy of Sciences*, vol. 98, no. 26, pp. 14 883–14 888, 2001.

[65] T. Hu, A. Y. Grosberg, and B. Shklovskii, "How proteins search for their specific sites on DNA: The role of DNA conformation," *Biophysical Journal*, vol. 90, no. 8, pp. 2731–2744, 2006.

[66] M. Horwitz and L. Loeb, "An *E. coli* promoter that regulates transcription by DNA superhelix-induced cruciform extrusion," *Science*, vol. 241, no. 4866, pp. 703–705, 1988.

[67] W. A. Krajewski, "Enhancement of transcription by short alternating C.G tracts incorporated within a *Rous sarcoma* virus-based chimeric promoter: *in vivo* studies," *Mol. Gen. Genet.*, vol. 252, no. 3, pp. 249–254, 1996.

[68] M. Thanbichler, P. H. Viollier, and L. Shapiro, "The structure and function of the bacterial chromosome," *Current Opinion in Genetics & Development*, vol. 15, no. 2, pp. 153–162, 2005.

[69] E. L. Zechiedrich, A. B. Khodursky, and N. R., "Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in *Escherichia coli*," *Genes & Development*, vol. 11, no. 19, pp. 2580–2592, 1997.

[70] J. C. Wang, "DNA topoisomerases," *Annual Review of Biochemistry*, vol. 65, no. 1, pp. 635–692, 1996.

[71] A. Conter, "Plasmid DNA supercoiling and survival in long-term cultures of *Escherichia coli*: Role of NaCl," *Journal of Bacteriology*, vol. 185, no. 17, pp. 5324–5327, 2003.

[72] A. Worcel and E. Burgi, "On the structure of the folded chromosome of *Escherichia coli*," *Journal of Molecular Biology*, vol. 71, no. 2, pp. 127–147, 1972.

[73] R. R. Sinden and D. E. Pettijohn, "Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling," *Proceedings of the National Academy of Sciences*, vol. 78, no. 1, pp. 224–228, 1981.

[74] R. Kavenoff and B. C. Bowen, "Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*," *Chromosoma*, vol. 59, no. 2, pp. 89–101, 1976.

[75] R. Kavenoff and O. A. Ryder, "Electron microscopy of membrane-associated folded chromosomes of *Escherichia coli*," *Chromosoma*, vol. 55, no. 1, pp. 13–25, 1976.

[76] A. J. Bendich, "The form of chromosomal DNA molecules in bacterial cells," *Biochimie*, vol. 83, no. 2, pp. 177–186, 2001.

[77] H. Delius and A. Worcel, "Electron microscopic studies on the folded chromosome of *escherichia coli*," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 38, pp. 53–58, 1974.

[78] N. P. Higgins, X Yang, Q Fu, and J. R. Roth, "Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*," *Journal of Bacteriology*, vol. 178, no. 10, pp. 2825–2835, 1996.

[79] S. Deng, R. A. Stein, and N. P. Higgins, "Transcription-induced barriers to supercoil diffusion in the *Salmonella typhimurium* chromosome," *Proceedings of the National Academy of Sciences*, vol. 101, no. 10, pp. 3398–3403, 2004.

[80] S. M. Mirkin, "DNA topology: Fundamentals," in *eLS*. American Cancer Society, 2001, ISBN: 9780470015902.

[81] R. Sinden, *DNA Structure and Function*. Elsevier, 1994, ISBN: 978-0-08-057173-7.

[82] A. Vologodskii, *Topology and Physics of Circular DNA (1992)*. City: CRC Press, 2017, ISBN: 9781138562813.

[83] R. N. Irobalieva, J. M. Fogg, D. J. Catanese, T. Sutthibutpong, M. Chen, A. K. Barker, S. J. Ludtke, S. A. Harris, M. F. Schmid, W. Chiu, and L. Zechiedrich, "Structural diversity of supercoiled DNA," *Nature Communications*, vol. 6, no. 1, 2015.

[84] M. T. J. van Loenhout, M. V. de Grunt, and C. Dekker, "Dynamics of DNA supercoils," *Science*, vol. 338, no. 6103, pp. 94–97, 2012.

[85] D. Racko, F. Benedetti, J. Dorier, and A. Stasiak, "Are TADs supercoiled?" *Nucleic Acids Research*, 2018.

[86] T. Dewey, *Fractals in Molecular Biophysics*. Oxford New York: Oxford University Press, 1997, ISBN: 9780195084474.

[87] Grosberg, A. Yu., Nechaev, S.K., and Shakhnovich, E.I., "The role of topological constraints in the kinetics of collapse of macromolecules," *J. Phys. France*, vol. 49, no. 12, pp. 2095–2100, 1988.

[88] A Grosberg, Y Rabin, S Havlin, and A Neer, "Crumpled globule model of the three-dimensional structure of DNA," *Europhysics Letters (EPL)*, vol. 23, no. 5, pp. 373–378, 1993.

[89] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science*, vol. 326, no. 5950, pp. 289–293, 2009.

[90] J. Dekker, "Gene regulation in the third dimension," *Science*, vol. 319, no. 5871, pp. 1793–1794, 2008.

[91] M. Müller, J. P. Wittmer, and M. E. Cates, "Topological effects in ring polymers: A computer simulation study," *Physical Review E*, vol. 53, no. 5, pp. 5063–5074, 1996.

[92] E. Toro and L. Shapiro, "Bacterial chromosome organization and segregation," *Cold Spring Harbor Perspectives in Biology*, vol. 2, no. 2, a000349–a000349, 2010.

[93] P. A. Wiggins, K. C. Cheveralls, J. S. Martin, R. Lintner, and J. Kondev, "Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament," *Proceedings of the National Academy of Sciences*, vol. 107, no. 11, pp. 4991–4995, 2010.

[94] L. A. Mirny, "The fractal globule as a model of chromatin architecture in the cell," *Chromosome Research*, vol. 19, no. 1, pp. 37–51, 2011.

[95] M. R. Hübner and D. L. Spector, "Chromatin dynamics," *Annual Review of Biophysics*, vol. 39, no. 1, pp. 471–489, 2010.

[96] W. G. Müller, D. Walker, G. L. Hager, and J. G. McNally, "Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter," *The Journal of Cell Biology*, vol. 154, no. 1, pp. 33–48, 2001.

[97] P. A. Wiggins, K. C. Cheveralls, J. S. Martin, R. Lintner, and J. Kondev, "Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament," *Proceedings of the National Academy of Sciences*, vol. 107, no. 11, pp. 4991–4995, 2010.

[98] M. I. Bogachev, O. A. Markelov, A. R. Kayumov, and A. Bunde, "Superstatistical model of bacterial DNA architecture," *Scientific Reports*, vol. 7, no. 1, 2017.

[99] J. Lipfert, S. Doniach, R. Das, and D. Herschlag, "Understanding nucleic acid–ion interactions," *Annual Review of Biochemistry*, vol. 83, no. 1, pp. 813–841, 2014.

[100] G. S. Manning, "On the application of polyelectrolyte "limiting laws" to the helix-coil transition of DNA. I. excess univalent cations," *Biopolymers*, vol. 11, no. 5, pp. 937–949, 1972.

[101] ——, "Counterion condensation on a helical charge lattice," *Macromolecules*, vol. 34, no. 13, pp. 4650–4655, 2001.

[102] A. Singh and N. Singh, "Effect of salt concentration on the stability of heterogeneous DNA," *Physica A: Statistical Mechanics and its Applications*, vol. 419, pp. 328–334, 2015.

[103] ——, "Phase diagram of mechanically stretched DNA: The salt effect," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 9, pp. 2052–2059, 2013.

[104] A. Singh, B. Mittal, and N. Singh, "Force induced unzipping of dsDNA: The solvent effect," *Phys. Express*, vol. 3, no. 18, 2013.

[105] A. Krueger, E. Protozanova, and M. D. Frank-Kamenetskii, "Sequence-dependent basepair opening in DNA double helix," *Biophysical Journal*, vol. 90, no. 9, pp. 3091–3099, 2006.

[106] G. Weber, N. Haslam, J. W. Essex, and C. Neylon, "Thermal equivalence of DNA duplexes for probe design," *Journal of Physics: Condensed Matter*, vol. 21, no. 3, p. 034 106, 2008.

[107] R. Owczarzy, Y. You, B. G. Moreira, J. A. Manthey, L. Huang, M. A. Behlke, and J. A. Walder, "Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures," *Biochemistry*, vol. 43, no. 12, pp. 3537–3554, 2004.

[108] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder, "Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations," *Biochemistry*, vol. 47, no. 19, pp. 5336–5353, 2008.

[109] J. SantaLucia, H. T. Allawi, and P. A. Seneviratne, "Improved nearest-neighbor parameters for predicting DNA duplex stability†," *Biochemistry*, vol. 35, no. 11, pp. 3555–3562, 1996.

[110] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," *Proceedings of the National Academy of Sciences*, vol. 95, no. 4, pp. 1460–1465, 1998.

[111] S. Hormeño, F. Moreno-Herrero, B. Ibarra, J. L. Carrascosa, J. M. Valpuesta, and J. R. Arias-Gonzalez, "Condensation prevails over B-A transition in the structure of DNA at low humidity," *Biophysical Journal*, vol. 100, no. 8, pp. 2006–2015, 2011.

[112] S. Hormeño, B. Ibarra, J. M. Valpuesta, J. L. Carrascosa, and J. R. Arias-Gonzalez, "Mechanical stability of low-humidity single DNA molecules," *Biopolymers*, vol. 97, no. 4, pp. 199–208, 2011.

[113] R. Dong, X. Yan, and S. Liu, "The salt dependence of the stretching transition of double-stranded DNA molecules," *Journal of Physics A: Mathematical and General*, vol. 37, no. 18, pp. 4977–4984, 2004.

[114] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, "Single-molecule derivation of salt dependent base-pair free energies in DNA," *Proceedings of the National Academy of Sciences*, vol. 107, no. 35, pp. 15 431–15 436, 2010.

[115] S. Tomac, M. Sarkar, T. Ratilainen, P. Wittung, P. E. Nielsen, B. Nordén, and A. Gräslund, "Ionic effects on the stability and conformation of peptide nucleic acid complexes," *Journal of the American Chemical Society*, vol. 118, no. 24, pp. 5544–5552, 1996.

[116] I. Khimji, J. Shin, and J. Liu, "DNA duplex stabilization in crowded polyanion solutions," *Chemical Communications*, vol. 49, no. 13, p. 1306, 2013.

[117] A. Maity, A. Singh, and N. Singh, "Differential stability of DNA based on salt concentration," *European Biophysics Journal*, vol. 46, no. 1, pp. 33–40, 2016.

[118] Z.-J. Tan and S.-J. Chen, "Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length," *Biophysical Journal*, vol. 90, no. 4, pp. 1175–1190, 2006.

[119] T. Odijk, "Polyelectrolytes near the rod limit," *Journal of Polymer Science: Polymer Physics Edition*, vol. 15, no. 3, pp. 477–483, 1977.

[120] J. Skolnick and M. Fixman, "Electrostatic persistence length of a wormlike polyelectrolyte," *Macromolecules*, vol. 10, no. 5, pp. 944–948, 1977.

[121] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, "Ionic effects on the elasticity of single DNA molecules," *Proceedings of the National Academy of Sciences*, vol. 94, no. 12, pp. 6185–6190, 1997.

[122] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, "Salt dependence of the elasticity and overstretching transition of single DNA molecules," *Biophysical Journal*, vol. 82, no. 6, pp. 3160–3169, 2002.

[123] V. Rybenkov, "The effect of ionic conditions on DNA helical repeat, effective diameter and free energy of supercoiling," *Nucleic Acids Research*, vol. 25, no. 7, pp. 1412–1418, 1997.

[124] C. Maffeo, R. Schöpflin, H. Brutzer, R. Stehr, A. Aksimentiev, G. Wedemann, and R. Seidel, "DNA–DNA interactions in tight supercoils are described by a small effective charge density," *Physical Review Letters*, vol. 105, no. 15, 2010.

[125] T. Schlick, B. Li, and W. Olson, "The influence of salt on the structure and energetics of supercoiled DNA," *Biophysical Journal*, vol. 67, no. 6, pp. 2146–2166, 1994.

[126] S. Neukirch, "Extracting DNA twist rigidity from experimental supercoiling data," *Physical Review Letters*, vol. 93, no. 19, 2004.

[127] N. Clauvelin, B. Audoly, and S. Neukirch, "Elasticity and electrostatics of plectonemic DNA," *Biophysical Journal*, vol. 96, no. 9, pp. 3716–3723, 2009.

[128] R. Schöpflin, H. Brutzer, O. Müller, R. Seidel, and G. Wedemann, "Probing the elasticity of DNA on short length scales by modeling supercoiling under tension," *Biophysical Journal*, vol. 103, no. 2, pp. 323–330, 2012.

[129] T. Lepage, F. Képès, and I. Junier, "Thermodynamics of long supercoiled molecules: Insights from highly efficient Monte Carlo simulations," *Biophysical Journal*, vol. 109, no. 1, pp. 135–143, 2015.

[130] S. Neukirch and J. F. Marko, "Analytical description of extension, torque, and supercoiling radius of a stretched twisted DNA," *Physical Review Letters*, vol. 106, no. 13, 2011.

[131] M. Emanuel, G. Lanzani, and H. Schiessel, "Multiplectoneme phase of double-stranded DNA under torsion," *Physical Review E*, vol. 88, no. 2, 2013.

[132] J. F. Marko and S. Neukirch, "Global force-torque phase diagram for the DNA double helix: Structural transitions, triple points, and collapsed plectonemes," *Physical Review E*, vol. 88, no. 6, 2013.

[133] F. C. Oberstrass, L. E. Fernandes, and Z. Bryant, "Torque measurements reveal sequence-specific cooperative transitions in supercoiled DNA," *Proceedings of the National Academy of Sciences*, vol. 109, no. 16, pp. 6106–6111, 2012.

[134] J. Delrow, P. Heath, and J. Schurr, "On the origin of the temperature dependence of the supercoiling free energy," *Biophysical Journal*, vol. 73, no. 5, pp. 2688–2701, 1997.

[135] F. Kriegel, N. Ermann, R. Forbes, D. Dulin, N. H. Dekker, and J. Lipfert, "Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers," *Nucleic Acids Research*, vol. 45, no. 10, pp. 5920–5929, 2017.

[136] J. Langowski, "Salt effects on internal motions of superhelical and linear pUC8 DNA," *Biophysical Chemistry*, vol. 27, no. 3, pp. 263–271, 1987.

[137] P. Anderson and W. Bauer, "Supercoiling in closed circular DNA: Dependence upon ion type and concentration," *Biochemistry*, vol. 17, no. 4, pp. 594–601, 1978.

[138] J. A. Gebe, J. J. Delrow, P. J. Heath, B. S. Fujimoto, D. W. Stewart, and J. Schurr, "Effects of Na^+ and Mg^{2+} on the structures of supercoiled DNAs: Comparison of simulations with experiments," *Journal of Molecular Biology*, vol. 262, no. 2, pp. 105–128, 1996.

[139] A. A. Zinchenko and K. Yoshikawa, "Na^+ shows a markedly higher potential than K^+ in DNA compaction in a crowded environment," *Biophysical Journal*, vol. 88, no. 6, pp. 4118–4123, 2005.

[140] A. V. Vologodskii and N. R. Cozzarelli, "Conformational and thermodynamic properties of supercoiled DNA," *Annual Review of Biophysics and Biomolecular Structure*, vol. 23, no. 1, pp. 609–643, 1994.

[141] S. B. Zimmerman and S. O. Trach, "Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*," *Journal of Molecular Biology*, vol. 222, no. 3, pp. 599–620, 1991.

[142] R. Ellis, "Macromolecular crowding: Obvious but underappreciated," *Trends in Biochemical Sciences*, vol. 26, no. 10, pp. 597–604, 2001.

[143] K. Luby-Phelps, "Cytoarchitecture and physical properties of cytoplasm: Volume, viscosity, diffusion, intracellular surface area," in *International Review of Cytology*, Elsevier, 1999, pp. 189–221.

[144] S. R. McGuffee and A. H. Elcock, "Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm," *PLoS Computational Biology*, vol. 6, no. 3, J. M. Briggs, Ed., e1000694, 2010.

[145] A. P. Minton, "The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media," *Journal of Biological Chemistry*, vol. 276, no. 14, pp. 10 577–10 580, 2001.

[146] A. P. Minton, "How can biochemical reactions within cells differ from those in test tubes?" *Journal of Cell Science*, vol. 119, no. 14, pp. 2863–2869, 2006.

[147] H.-X. Zhou, G. Rivas, and A. P. Minton, "Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences," *Annual Review of Biophysics*, vol. 37, no. 1, pp. 375–397, 2008.

[148] S. B. Zimmerman and A. P. Minton, "Macromolecular crowding: Biochemical, biophysical, and physiological consequences," *Annual Review of Biophysics and Biomolecular Structure*, vol. 22, no. 1, pp. 27–65, 1993.

[149] R. Ellis, "Macromolecular crowding: An important but neglected aspect of the intracellular environment," *Current Opinion in Structural Biology*, vol. 11, no. 1, pp. 114–119, 2001.

[150] M. B. Elowitz, M. G. Surette, P.-E. Wolf, J. B. Stock, and S. Leibler, "Protein mobility in the cytoplasm of *Escherichia coli*," *Journal of Bacteriology*, vol. 181, no. 1, pp. 197–203, 1999.

[151] A. Marcovitz and Y. Levy, "Obstacles may facilitate and direct DNA search by proteins," *Biophysical Journal*, vol. 104, no. 9, pp. 2042–2050, 2013.

[152] R. C. Johnson, J. W. Schmidt, L. M. Johnson, and J. F. Gardner, "Major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome," in *The Bacterial Chromosome*, American Society of Microbiology, pp. 65–132.

[153] C. H. Spink and J. B. Chaires, "Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA†," *Biochemistry*, vol. 38, no. 1, pp. 496–508, 1999.

[154] L. A. Marky and D. W. Kupke, "Enthalpy-entropy compensations in nucleic acids: Contribution of electrostriction and structural hydration," in *Methods in Enzymology*, Elsevier, 2000, pp. 419–441.

[155] S. Nakano, H. Karimata, T. Ohmichi, J. Kawakami, and N. Sugimoto, "The effect of molecular crowding with nucleotide length and cosolute structure on DNA duplex stability," *Journal of the American Chemical Society*, vol. 126, no. 44, pp. 14 330–14 331, 2004.

[156] M. D. Frank-Kamenetskii and S. M. Mirkin, "Triplex DNA structures," *Annual Review of Biochemistry*, vol. 64, no. 1, pp. 65–95, 1995.

[157] J. T. Davis, "G-quartets 40 years later: From 5′-GMP to molecular biology and supramolecular chemistry," *Angewandte Chemie International Edition*, vol. 43, no. 6, pp. 668–698, 2004.

[158] R. Goobes, N. Kahana, O. Cohen, and A. Minsky, "Metabolic buffering exerted by macromolecular crowding on DNA-DNA interactions: Origin and physiological significance†," *Biochemistry*, vol. 42, no. 8, pp. 2431–2440, 2003.

[159] Z.-Y. Kan, Y. Yao, P. Wang, X.-H. Li, Y.-H. Hao, and Z. Tan, "Molecular crowding induces telomere G-quadruplex formation under salt-deficient conditions and enhances its competition with duplex formation," *Angewandte Chemie International Edition*, vol. 45, no. 10, pp. 1629–1632, 2006.

[160] M. Ganji, M. Docter, S. F. L. Grice, and E. A. Abbondanzieri, "DNA binding proteins explore multiple local configurations during docking via rapid rebinding," *Nucleic Acids Research*, vol. 44, no. 17, pp. 8376–8384, 2016.

[161] S. B. Zimmerman and B. Harrison, "Macromolecular crowding increases binding of DNA polymerase to DNA: An adaptive effect," *Proceedings of the National Academy of Sciences*, vol. 84, no. 7, pp. 1871–1875, 1987.

[162] B. Akabayov, S. R. Akabayov, S.-J. Lee, G. Wagner, and C. C. Richardson, "Impact of macromolecular crowding on DNA replication," *Nature Communications*, vol. 4, no. 1, 2013.

[163] X. Zhu, S. Y. Ng, A. N. Gupta, Y. P. Feng, B. Ho, A. Lapp, S. U. Egelhaaf, V. T. Forsyth, M. Haertlein, M. Moulin, R. Schweins, and J. R. C. van der Maarel, "Effect of crowding on the conformation of interwound DNA strands from neutron scattering measurements and Monte Carlo simulations," *Physical Review E*, vol. 81, no. 6, 2010.

[164] D. Miyoshi and N. Sugimoto, "Molecular crowding effects on structure and stability of DNA," *Biochimie*, vol. 90, no. 7, pp. 1040–1051, 2008.

[165] G. Adam and M. Delbruck, "Reduction of dimensionality in biological diffusion processes," in *Structural Chemistry and Molecular Biology*, A. Rich and N. Davidson, Eds., W. H. Freeman and Co. San Francisco, 1968, pp. 198–215.

[166] O. G. Berg, R. B. Winter, and P. H. V. Hippel, "Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory," *Biochemistry*, vol. 20, no. 24, pp. 6929–6948, 1981.

[167] Y. M. Wang, R. H. Austin, and E. C. Cox, "Single molecule measurements of repressor protein 1D diffusion on DNA," *Physical Review Letters*, vol. 97, no. 4, 2006.

[168] J. Gorman, A. Chowdhury, J. A. Surtees, J. Shimada, D. R. Reichman, E. Alani, and E. C. Greene, "Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6," *Molecular Cell*, vol. 28, no. 3, pp. 359–370, 2007.

[169] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, and X. S. Xie, "A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA," *Proceedings of the National Academy of Sciences*, vol. 103, no. 15, pp. 5752–5757, 2006.

[170] J. Elf, G.-W. Li, and X. S. Xie, "Probing transcription factor dynamics at the single-molecule level in a living cell," *Science*, vol. 316, no. 5828, pp. 1191–1194, 2007.

[171] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, "The *lac* repressor displays facilitated diffusion in living cells," *Science*, vol. 336, no. 6088, pp. 1595–1598, 2012.

[172] L. Zandarashvili, A. Esadze, D. Vuzman, C. A. Kemme, Y. Levy, and J. Iwahara, "Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble," *Proceedings of the National Academy of Sciences*, vol. 112, no. 37, E5142–E5 149, 2015.

[173] P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi, and X. S. Xie, "Nonspecifically bound proteins spin while diffusing along DNA," *Nature Structural & Molecular Biology*, vol. 16, no. 12, pp. 1224–1229, 2009.

[174] S. E. Halford, "An end to 40 years of mistakes in DNA–protein association kinetics?" *Biochemical Society Transactions*, vol. 37, no. 2, pp. 343–348, 2009.

[175] A. Marcovitz and Y. Levy, "Frustration in protein-DNA binding influences conformational switching and target search kinetics," *Proceedings of the National Academy of Sciences*, vol. 108, no. 44, pp. 17 957–17 962, 2011.

[176] A. Bhattacherjee and Y. Levy, "Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding," *Nucleic Acids Research*, vol. 42, no. 20, pp. 12 404–12 414, 2014.

[177] R. B. Winter and P. H. V. Hippel, "Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli lac* repressor-operator interaction: Equilibrium measurements," *Biochemistry*, vol. 20, no. 24, pp. 6948–6960, 1981.

[178] M. Record, J.-H. Ha, and M. A. Fisher, "[16] Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA," pp. 291–343, 1991.

[179] O. Givaty and Y. Levy, "Protein sliding along DNA: Dynamics and structural characterization," *Journal of Molecular Biology*, vol. 385, no. 4, pp. 1087–1097, 2009.

[180] L. Cai and H.-X. Zhou, "Theory and simulation on the kinetics of protein–ligand binding coupled to conformational change," *The Journal of Chemical Physics*, vol. 134, no. 10, p. 105 101, 2011.

[181] R. Murugan, "Theory of site-specific DNA-protein interactions in the presence of conformational fluctuations of DNA binding domains," *Biophysical Journal*, vol. 99, no. 2, pp. 353–359, 2010.

[182] B. van den Broek, M. A. Lomholt, S.-M. J. Kalisch, R. Metzler, and G. J. L. Wuite, "How DNA coiling enhances target localization by proteins," *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15 738–15 742, 2008.

[183] M. A. Lomholt, B. van den Broek, S.-M. J. Kalisch, G. J. L. Wuite, and R. Metzler, "Facilitated diffusion with DNA coiling," *Proceedings of the National Academy of Sciences*, vol. 106, no. 20, pp. 8204–8208, 2009.

[184] A.-M. Florescu and M. Joyeux, "Description of nonspecific DNA-protein interaction and facilitated diffusion with a dynamical model," *The Journal of Chemical Physics*, vol. 130, no. 1, p. 015 103, 2009.

[185] M. Bauer and R. Metzler, "Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states," *Biophysical Journal*, vol. 102, no. 10, pp. 2321–2330, 2012.

[186] M Sheinman, O Bénichou, Y Kafri, and R Voituriez, "Classes of fast and specific search mechanisms for proteins on DNA," *Reports on Progress in Physics*, vol. 75, no. 2, p. 026 601, 2012.

[187] C. Loverdo, O. Bénichou, R. Voituriez, A. Biebricher, I. Bonnet, and P. Desbiolles, "Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins," *Physical Review Letters*, vol. 102, no. 18, 2009.

[188] G.-W. Li, O. G. Berg, and J. Elf, "Effects of macromolecular crowding and DNA looping on gene regulation kinetics," *Nature Physics*, vol. 5, no. 4, pp. 294–297, 2009.

[189] A. Mondal and A. Bhattacherjee, "Searching target sites on DNA by proteins: Role of DNA dynamics under confinement," *Nucleic Acids Research*, vol. 43, no. 19, pp. 9176–9186, 2015.

[190] T. Ando and J. Skolnick, "Sliding of proteins non-specifically bound to DNA: Brownian dynamics studies with coarse-grained protein and DNA models," *PLoS Computational Biology*, vol. 10, no. 12, G. M. Clore, Ed., e1003990, 2014.

[191] P. H. von Hippel and O. G. Berg, "Facilitated target location in biological systems," *Journal of Biological Chemistry*, vol. 264, no. 2, pp. 675–678, 1989.

[192] S. E. Halford, "How do site-specific DNA-binding proteins find their targets?" *Nucleic Acids Research*, vol. 32, no. 10, pp. 3040–3052, 2004.

[193] O. Bénichou, C. Chevalier, B. Meyer, and R. Voituriez, "Facilitated diffusion of proteins on chromatin," *Physical Review Letters*, vol. 106, no. 3, 2011.

[194] S. L. Cravens, J. D. Schonhoft, M. M. Rowland, A. A. Rodriguez, B. G. Anderson, and J. T. Stivers, "Molecular crowding enhances facilitated diffusion of two human DNA glycosylases," *Nucleic Acids Research*, vol. 43, no. 8, pp. 4087–4097, 2015.

[195] P. Singh, S. Choudhury, S. Dutta, A. Adhikari, S. Bhattacharya, D. Pal, and S. K. Pal, "Ultrafast spectroscopy on DNA-cleavage by endonuclease in molecular crowding," *International Journal of Biological Macromolecules*, vol. 103, pp. 395–402, 2017.

[196] I. Kuznetsova, B. Zaslavsky, L. Breydo, K. Turoverov, and V. Uversky, "Beyond the excluded volume effects: Mechanistic complexity of the crowded milieu," *Molecules*, vol. 20, no. 1, pp. 1377–1409, 2015.

[197] C. A. Brackley, M. E. Cates, and D. Marenduzzo, "Intracellular facilitated diffusion: Searchers, crowders, and blockers," *Physical Review Letters*, vol. 111, no. 10, 2013.

[198] D. Krepel, D. Gomez, S. Klumpp, and Y. Levy, "Mechanism of facilitated diffusion during a DNA search in crowded environments," *The Journal of Physical Chemistry B*, vol. 120, no. 43, pp. 11 113–11 122, 2016.

[199] L. Liu and K. Luo, "Molecular crowding effect on dynamics of DNA-binding proteins search for their targets," *The Journal of Chemical Physics*, vol. 141, no. 22, p. 225 102, 2014.

[200] Y. Ma, Y. Chen, W. Yu, and K. Luo, "How nonspecifically DNA-binding proteins search for the target in crowded environments," *The Journal of Chemical Physics*, vol. 144, no. 12, p. 125 102, 2016.

[201] P. Dey and A. Bhattacherjee, "Role of macromolecular crowding on the intracellular diffusion of DNA binding proteins," *Scientific Reports*, vol. 8, no. 1, 2018.

[202] I. Bonnet, A. Biebricher, P.-L. Porté, C. Loverdo, O. Bénichou, R. Voituriez, C. Escudé, W. Wende, A. Pingoud, and P. Desbiolles, "Sliding and jumping

of single EcoRV restriction enzymes on non-cognate DNA," *Nucleic Acids Research*, vol. 36, no. 12, pp. 4118–4127, 2008.

[203] A. Graneli, C. C. Yeykal, R. B. Robertson, and E. C. Greene, "Long-distance lateral diffusion of human Rad51 on double-stranded DNA," *Proceedings of the National Academy of Sciences*, vol. 103, no. 5, pp. 1221–1226, 2006.

[204] A. B. Kolomeisky, "Physics of protein–DNA interactions: Mechanisms of facilitated target search," *Phys. Chem. Chem. Phys.*, vol. 13, no. 6, pp. 2088–2095, 2011.

[205] M. Bauer and R. Metzler, "*In vivo* facilitated diffusion model," *PLoS ONE*, vol. 8, no. 1, Y. K. Levy, Ed., e53956, 2013.

[206] O. G. Berg, "Diffusion-controlled protein-DNA association: Influence of segemental diffusion of the DNA," *Biopolymers*, vol. 23, no. 10, pp. 1869–1889, 1984.

[207] E. F. Koslover, M. A. D. de la Rosa, and A. J. Spakowitz, "Theoretical and computational modeling of target-site search kinetics *in vitro* and *in vivo*," *Biophysical Journal*, vol. 101, no. 4, pp. 856–865, 2011.

[208] J. Crank, *The Mathematics of Diffusion*. Oxford, Eng: Clarendon Press, 1975, ISBN: 0198534116.

[209] G. Allen, *Protein: A Comprehensive Treatise*. Greenwich, Conn: Jai Press, Inc, 1997, ISBN: 1559386711.

[210] I. L. Mostinsky, "DIFFUSION," in *A-to-Z Guide to Thermodynamics, Heat and Mass Transfer, and Fluids Engineering*, Begellhouse.

[211] K. Laidler, *Physical chemistry*. Menlo Park, Calif: Benjamin/Cummings Pub. Co, 1982, ISBN: 0805356827.

[212] R. M. Nedderman, "STOKES' LAW FOR SOLID SPHERES AND SPHERICAL BUBBLES," in *A-to-Z Guide to Thermodynamics, Heat and Mass Transfer, and Fluids Engineering*, Begellhouse.

[213] G. L. Shires, "REYNOLDS NUMBER," in *A-to-Z Guide to Thermodynamics, Heat and Mass Transfer, and Fluids Engineering*, Begellhouse.

[214] M. Itami and S.-I. Sasa, "Derivation of Stokes' law from Kirkwood's formula and the Green-Kubo formula via large deviation theory," *Journal of Statistical Physics*, vol. 161, no. 3, pp. 532–552, 2015.

[215] G. K. Batchelor, *An Introduction to Fluid Dynamics*. Cambridge University Press, 2000.

[216] A. Einstein, "Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen," *Annalen der Physik*, vol. 322, no. 8, pp. 549–560, 1905.

[217] A. Fick, "On liquid diffusion," *Journal of Membrane Science*, vol. 100, no. 1, pp. 33–38, 1995.

[218] I. M. Sokolov and J. Klafter, "From diffusion to anomalous diffusion: A century after einstein's brownian motion," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 15, no. 2, p. 026 103, 2005.

[219] M. von Smoluchowski, "Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen," *Annalen der Physik*, vol. 326, no. 14, pp. 756–780, 1906.

[220] G. G. Emch, "Diffusion, Einstein formula and mechanics," *Journal of Mathematical Physics*, vol. 14, no. 12, pp. 1775–1783, 1973.

[221] M. A. Islam, "Einstein-Smoluchowski diffusion equation: A discussion," *Physica Scripta*, vol. 70, no. 2-3, pp. 120–125, 2004.

[222] G. Peskir, "On the diffusion coefficient: The Einstein relation and beyond," *Stochastic Models*, vol. 19, no. 3, pp. 383–405, 2003.

[223] M. V. Reeks, "ERGODICITY," in *A-to-Z Guide to Thermodynamics, Heat and Mass Transfer, and Fluids Engineering*, Begellhouse.

[224] C. Derman, "Ergodic property of the Brownian motion process," *Proceedings of the National Academy of Sciences*, vol. 40, no. 12, pp. 1155–1158, 1954.

[225] G. Kallianpur and H. Robbins, "Ergodic property of the Brownian motion process," *Proceedings of the National Academy of Sciences*, vol. 39, no. 6, pp. 525–533, 1953.

[226] P. S. Doyle and P. T. Underhill, "Brownian dynamics simulations of polymers and soft matter," in *Handbook of Materials Modeling*, Springer Netherlands, 2005, pp. 2619–2630.

[227] X. Michalet, "Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium," *Physical Review E*, vol. 82, no. 4, 2010.

[228] D. Frenkel, *Understanding Molecular Simulation: From Algorithms to Applications*. San Diego: Academic Press, 2002, ISBN: 9780080519982.

[229] N. Wiener, "Differential-space," *Journal of Mathematics and Physics*, vol. 2, no. 1-4, pp. 131–174, 1923.

[230] M. D. Donsker, "An invariant principle for certain probability limit theorems," *Memoirs of the American Mathematical Society*, no. 6, pp. 1–10, 1951.

[231] P. Billingsley, *Probability and Measure*. New Delhi: Wiley India, 2012, ISBN: 8126517719.

[232] C. Lanczos, *The Variational Principles of Mechanics (Dover Books on Physics)*. Dover Publications, 1986, ISBN: 0486650677.

[233] A. R. Forsyth, *Calculus of Variations*. Dover Publications, 1960, ISBN: 0486322580.

[234] H. Goldstein, C. P. P. Jr., and J. L. Safko, *Classical Mechanics (3rd Edition)*. Pearson, 2001, ISBN: 0201657023.

[235] L. D. Landau and E. Lifshitz, *Mechanics: Volume 1 (Course of Theoretical Physics S)*. Butterworth-Heinemann, 1976, ISBN: 0750628960.

[236] J. J. Waterson, "On the physics of media that are composed of free and perfectly elastic molecules in a state of motion," *Proceedings of the Royal Society of London*, vol. 5, pp. 604–604, 1843.

[237] J. J. Waterston and L. Rayleigh, "On the physics of media that are composed of free and perfectly elastic molecules in a state of motion," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 183, pp. 1–79, 1892.

[238] L. Stella, C. D. Lorenz, and L. Kantorovich, "Generalized Langevin equation: An efficient approach to nonequilibrium molecular dynamics of open systems," *Physical Review B*, vol. 89, no. 13, 2014.

[239] F. X. Hart, "The coefficient of friction as a second rank tensor," *American Journal of Physics*, vol. 40, no. 3, pp. 475–476, 1972.

[240] R. Zwanzig, "Ensemble method in the theory of irreversibility," *The Journal of Chemical Physics*, vol. 33, no. 5, pp. 1338–1341, 1960.

[241] H. B. Callen, M. L. Barasch, and J. L. Jackson, "Statistical mechanics of irreversibility," *Physical Review*, vol. 88, no. 6, pp. 1382–1386, 1952.

[242] R. Zwanzig, "Memory effects in irreversible thermodynamics," *Physical Review*, vol. 124, no. 4, pp. 983–992, 1961.

[243] H. Mori, "Transport, collective motion, and brownian motion," *Progress of Theoretical Physics*, vol. 33, no. 3, pp. 423–455, 1965.

[244] R. Zwanzig, *Nonequilibrium Statistical Mechanics*. Oxford University Press, 2001, ISBN: 0195140184.

[245] G. Maruyama, "Continuous Markov processes and stochastic equations," *Rendiconti del Circolo Matematico di Palermo*, vol. 4, no. 1, p. 48, 1955.

[246] D. L. Ermak and J. A. McCammon, "Brownian dynamics with hydrodynamic interactions," *The Journal of Chemical Physics*, vol. 69, no. 4, pp. 1352–1360, Aug. 1978.

[247] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer Berlin Heidelberg, 1992.

[248] R. Khasminskii, *Stochastic Stability of Differential Equations*. Springer Berlin Heidelberg, 2012.

[249] K. Burrage, P. Burrage, and T. Mitsui, "Numerical solutions of stochastic differential equations - implementation and stability issues," *Journal of Computational and Applied Mathematics*, vol. 125, no. 1-2, pp. 171–182, 2000.

[250] Y. Saito and T. Mitsui, "Stability analysis of numerical schemes for stochastic differential equations," *SIAM Journal on Numerical Analysis*, vol. 33, no. 6, pp. 2254–2267, 1996.

[251] ——, "T-stability of numerical scheme for stochastic differential equations," in *Contributions in Numerical Mathematics*, World Scientific, 1993, pp. 333–344.

[252] A. C. Brańka, "On algorithms for Brownian dynamics computer simulations," *Computational Methods in Science and Technology*, vol. 4, no. 1, pp. 35–42, 1998.

[253] A. C. Brańka and D. M. Heyes, "Algorithms for Brownian dynamics computer simulations: Multivariable case," *Physical Review E*, vol. 60, no. 2, pp. 2381–2387, 1999.

[254] W. F. V. Gunsteren and H. J. C. Berendsen, "A leap-frog algorithm for stochastic dynamics," *Molecular Simulation*, vol. 1, no. 3, pp. 173–185, 1988.

[255] K. Burrage and P. Burrage, "High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations," *Applied Numerical Mathematics*, vol. 22, no. 1-3, pp. 81–101, 1996.

[256] E. Helfand, "Brownian dynamics study of transitions in a polymer chain of bistable oscillators," *The Journal of Chemical Physics*, vol. 69, no. 3, p. 1010, 1978.

[257] A. Iniesta and J. G. de la Torre, "A second-order algorithm for the simulation of the Brownian dynamics of macromolecular models," *The Journal of Chemical Physics*, vol. 92, no. 3, pp. 2015–2018, 1990.

[258] R. L. Honeycutt, "Stochastic Runge-Kutta algorithms. I. White noise," *Physical Review A*, vol. 45, no. 2, pp. 600–603, 2 1992.

[259] A. C. Brańka and D. M. Heyes, "Algorithms for Brownian dynamics simulation," *Physical Review E*, vol. 58, no. 2, pp. 2611–2615, 1998.

[260] A. Greiner, W. Strittmatter, and J. Honerkamp, "Numerical integration of stochastic differential equations," *Journal of Statistical Physics*, vol. 51, no. 1-2, pp. 95–108, 1988.

[261] T. Geyer, "Many-particle Brownian and Langevin dynamics simulations with the Brownmove package," *BMC Biophysics*, vol. 4, no. 1, p. 7, 2011.

[262] J.-F. Dufrêche, B. Rotenberg, V. Marry, and P. Turq, "Bridging molecular and continuous descriptions: The case of dynamics in clays," *Anais da Academia Brasileira de Ciências*, vol. 82, no. 1, pp. 61–68, 2010.

[263] L. Durlofsky, J. F. Brady, and G. Bossis, "Dynamic simulation of hydrodynamically interacting particles," *Journal of Fluid Mechanics*, vol. 180, no. -1, p. 21, 1987.

[264] J. F. Brady and G Bossis, "Stokesian dynamics," *Annual Review of Fluid Mechanics*, vol. 20, no. 1, pp. 111–157, 1988.

[265] H. Tanaka and T. Araki, "Simulation method of colloidal suspensions with hydrodynamic interactions: Fluid particle dynamics," *Physical Review Letters*, vol. 85, no. 6, pp. 1338–1341, 2000.

[266] Y. Nakayama and R. Yamamoto, "Simulation method to resolve hydrodynamic interactions in colloidal dispersions," *Physical Review E*, vol. 71, no. 3, 2005.

[267] R. D. Groot and P. B. Warren, "Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation," *The Journal of Chemical Physics*, vol. 107, no. 11, pp. 4423–4435, 1997.

[268] P Español and P Warren, "Statistical mechanics of dissipative particle dynamics," *Europhysics Letters (EPL)*, vol. 30, no. 4, pp. 191–196, 1995.

[269] P. J Hoogerbrugge and J. M.V. A Koelman, "Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics," *Europhysics Letters (EPL)*, vol. 19, no. 3, pp. 155–160, 1992.

[270] A. Malevanets and R. Kapral, "Mesoscopic model for solvent dynamics," *The Journal of Chemical Physics*, vol. 110, no. 17, pp. 8605–8613, 1999.

[271] G. Gompper, T. Ihle, D. M. Kroll, and R. G. Winkler, "Multi-particle collision dynamics: A particle-based mesoscale simulation approach to the hydrodynamics of complex fluids," in *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, C. Holm and K. Kremer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–87, ISBN: 978-3-540-87706-6.

[272] C. P Lowe, "An alternative approach to dissipative particle dynamics," *Europhysics Letters (EPL)*, vol. 47, no. 2, pp. 145–151, 1999.

[273] S. Chen and G. D. Doolen, "Lattice Boltzmann method for fluid flows," *Annual Review of Fluid Mechanics*, vol. 30, no. 1, pp. 329–364, 1998.

[274] A. J. C. Ladd and R. Verberg, "Lattice-Boltzmann simulations of particle-fluid suspensions," *Journal of Statistical Physics*, vol. 104, no. 5, pp. 1191–1251, 2001.

[275] M. E. Cates, K Stratford, R Adhikari, P Stansell, J.-C. Desplat, I Pagonabarraga, and A. J. Wagner, "Simulating colloid hydrodynamics with lattice Boltzmann methods," *Journal of Physics: Condensed Matter*, vol. 16, no. 38, S3903–S3915, 2004.

[276] R Adhikari, K Stratford, M. E Cates, and A. J Wagner, "Fluctuating lattice Boltzmann," *Europhysics Letters (EPL)*, vol. 71, no. 3, pp. 473–479, 2005.

[277] B. Dünweg and A. J. C. Ladd, "Lattice Boltzmann simulations of soft matter systems," in *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, C. Holm and K. Kremer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 89–166, ISBN: 978-3-540-87706-6.

[278] R. R. Schmidt, J. G. H. Cifre, and J. G. de la Torre, "Comparison of Brownian dynamics algorithms with hydrodynamic interaction," *The Journal of Chemical Physics*, vol. 135, no. 8, p. 084 116, 2011.

[279] A. Scala, T. Voigtmann, and C. De Michele, "Event-driven Brownian dynamics for hard spheres," *The Journal of Chemical Physics*, vol. 126, no. 13, p. 134 109, 2007.

[280] S. E. Braslavsky, "Glossary of terms used in photochemistry, 3rd edition (IUPAC recommendations 2006)," *Pure and Applied Chemistry*, vol. 79, no. 3, pp. 293–465, 2007.

[281] P. Debye and E. Hückel, "Debye-Hückel theory of electrolytes," *Phys. Z*, vol. 24, p. 185, 1923.

[282] D. Andelman, "Introduction to electrostatics in soft and biological matter," in *Scottish Graduate Series*, Taylor & Francis, 2006, pp. 97–122.

[283] P. Debye, *The Collected Papers of Peter J.W. Debye*. Woodbridge, Conn: Ox Bow Press, 1988, ISBN: 0918024587.

[284] H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions, 3rd edition*. New York: Reinhold Publishing, 1958, ISBN: 0278917291.

[285] K. S. Pitzer, "Thermodynamics of electrolytes. I. theoretical basis and general equations," *The Journal of Physical Chemistry*, vol. 77, no. 2, pp. 268–277, 1973.

[286] C. Davies, *Ion Association*. London: Butterworths, 1962, pp. 37–53.

[287] J. N. Brönsted, "Studies on Solubility. IV. The Principle of the Specific Interaction of Ions," *Journal of the American Chemical Society*, vol. 44, no. 5, pp. 877–898, 1922.

[288] M. T. Neves-Petersen and S. B. Petersen, "Protein electrostatics," in *Biotechnology Annual Review*, Elsevier, 2003, pp. 315–395.

[289] P. J. Hagerman and B. H. Zimm, "Monte Carlo approach to the analysis of the rotational diffusion of wormlike chains," *Biopolymers*, vol. 20, no. 7, pp. 1481–1502, Jul. 1981.

[290] H. Jian, A. V. Vologodskii, and T. Schlick, "A combined wormlike-chain and bead model for dynamic simulations of long linear DNA," *Journal of Computational Physics*, vol. 136, no. 1, pp. 168–179, 1997.

[291] H. Jian, T. Schlick, and A. Vologodskii, "Internal motion of supercoiled DNA: Brownian dynamics simulations of site juxtaposition," *Journal of Molecular Biology*, vol. 284, no. 2, pp. 287–296, 1998.

[292] D. Swigon, B. D. Coleman, and I. Tobias, "The elastic rod model for DNA and its application to the tertiary structure of DNA minicircles in mononucleosomes," *Biophysical Journal*, vol. 74, no. 5, pp. 2515–2530, 1998.

[293] T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, "A coarse grain model for DNA," *The Journal of Chemical Physics*, vol. 126, no. 8, p. 084 901, 2007.

[294] J. Wang and H. Gao, "A generalized bead-rod model for Brownian dynamics simulations of wormlike chains under strong confinement," *The Journal of Chemical Physics*, vol. 123, no. 8, p. 084 906, 2005.

[295] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd. New York, NY, USA: Oxford University Press, Inc., 2017, ISBN: 0198803206, 9780198803201.

[296] M. P. Howard, J. A. Anderson, A. Nikoubashman, S. C. Glotzer, and A. Z. Panagiotopoulos, "Efficient neighbor list calculation for molecular simulation of colloidal systems using graphics processing units," *Computer Physics Communications*, vol. 203, pp. 45–52, 2016.

[297] Z. Yao, J.-S. Wang, G.-R. Liu, and M. Cheng, "Improved neighbor list algorithm in molecular simulations using cell decomposition and data sorting method," *Computer Physics Communications*, vol. 161, no. 1-2, pp. 27–35, 2004.

[298] V. G. Vizing, "On an estimate of the chromatic class of a p-graph," *Diskret. Analiz.*, vol. 3, pp. 25–30, 1964.

[299] D. Stigter, "Interactions of highly charged colloidal cylinders with applications to double-stranded DNA," *Biopolymers*, vol. 16, no. 7, pp. 1435–1448, Jul. 1977.

[300] A. Vologodskii and N. Cozzarelli, "Modeling of long-range electrostatic interactions in DNA," *Biopolymers*, vol. 35, no. 3, pp. 289–296, Mar. 1995.

[301] D. Stigter and K. A. Dill, "Theory for second virial coefficients of short DNA," *The Journal of Physical Chemistry*, vol. 97, no. 49, pp. 12 995–12 997, Dec. 1993.

[302] H. Sagan, *Space-Filling Curves*. Springer New York, 1994.

[303] A. R. Butz, "Space filling curves and mathematical programming," *Information and Control*, vol. 12, no. 4, pp. 314–330, 1968.

[304] ——, "Alternative algorithm for Hilbert's space-filling curve," *IEEE Transactions on Computers*, vol. C-20, no. 4, pp. 424–426, 1971.

[305] M. Bader, "How to construct space-filling curves," in *Texts in Computational Science and Engineering*, Springer Berlin Heidelberg, 2012, pp. 15–30.

[306] J. Lawder, "Calculation of mappings between one and N-dimensional values using the Hilbert space-filling curve," Tech. Rep., Sep. 2000.

[307] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[308] J. Dekker, "Capturing chromosome conformation," *Science*, vol. 295, no. 5558, pp. 1306–1311, 2002.

[309] E. de Wit and W. de Laat, "A decade of 3C technologies: Insights into nuclear organization," *Genes Dev*, vol. 26, no. 1, pp. 11–24, 2012.

[310] L. Liu and C. Hyeon, "Contact statistics highlight distinct organizing principles of proteins and RNA," *Biophysical Journal*, vol. 110, no. 11, pp. 2320–2327, 2016.

[311] G. Peano, "Sur une courbe, qui remplit toute une aire plane," *Mathematische Annalen*, vol. 36, no. 1, pp. 157–160, 1890.

[312] M. L. J. Moncany and E. Kellenberger, "High magnesium content of *Escherichia coli* B.," *Experientia*, vol. 37, no. 8, pp. 846–847, 1981.

[313] M. Hammermann, C. Steinmaier, H. Merlitz, U. Kapp, W. Waldeck, G. Chirico, and J. Langowski, "Salt effects on the structure and internal dynamics of superhelical DNAs studied by light scattering and Brownian dynamics," *Biophysical Journal*, vol. 73, no. 5, pp. 2674–2687, 1997.

[314] V. V. Rybenkov, A. V. Vologodskii, and N. R. Cozzarelli, "The effect of ionic conditions on the conformations of supercoiled DNA. I. Sedimentation analysis," *Journal of Molecular Biology*, vol. 267, no. 2, pp. 299–311, 1997.

[315] T. Kalkbrenner, A. Arnold, and S. J. Tans, "Internal dynamics of supercoiled DNA molecules," *Biophysical Journal*, vol. 96, no. 12, pp. 4951–4955, 2009.

[316] R. Shusterman, T. Gavrinyov, and O. Krichevsky, "Internal dynamics of superhelical DNA," *Physical Review Letters*, vol. 100, no. 9, 2008.

[317] D. Kamashev, A. Balandina, A. K. Mazur, P. B. Arimondo, and J. Rouviere-Yaniv, "HU binds and folds single-stranded DNA," *Nucleic Acids Research*, vol. 36, no. 3, pp. 1026–1036, 2007.

[318] A. Balandina, D. Kamashev, and J. Rouviere-Yaniv, "The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids," *Journal of Biological Chemistry*, vol. 277, no. 31, pp. 27 622–27 628, 2002.

[319] J. Dorier and A. Stasiak, "Topological origins of chromosomal territories," *Nucleic Acids Research*, vol. 37, no. 19, pp. 6316–6322, 2009.

[320] N. Sinha and S. Smith-Gill, "Electrostatics in protein binding and function," *Current Protein & Peptide Science*, vol. 3, no. 6, pp. 601–614, 2002.

[321] B. Bhushan, J. N. Israelachvili, and U. Landman, "Nanotribology: Friction, wear and lubrication at the atomic scale," *Nature*, vol. 374, no. 6523, pp. 607–616, 1995.

[322] M. Benz, N. Chen, G. Jay, and J. Israelachvili, "Static forces, structure and flow properties of complex fluids in highly confined geometries," *Annals of Biomedical Engineering*, vol. 33, no. 1, pp. 39–51, 2005.

[323] A. Sierou and J. F. Brady, "Accelerated Stokesian dynamics simulations," *Journal of Fluid Mechanics*, vol. 448, 2001.

[324] A. J. Banchio and J. F. Brady, "Accelerated Stokesian dynamics: Brownian motion," *The Journal of Chemical Physics*, vol. 118, no. 22, pp. 10 323–10 332, 2003.

[325] P. Mereghetti and R. C. Wade, "Atomic detail Brownian dynamics simulations of concentrated protein solutions with a mean field treatment of hydrodynamic interactions," *The Journal of Physical Chemistry B*, vol. 116, no. 29, pp. 8523–8533, 2012.

[326] T. Frembgen-Kesner and A. H. Elcock, "Computer simulations of the bacterial cytoplasm," *Biophysical Reviews*, vol. 5, no. 2, pp. 109–119, 2013.

[327] J. C. Wang, "Moving one DNA double helix through another by a type II DNA topoisomerase: The story of a simple molecular machine," *Quarterly Reviews of Biophysics*, vol. 31, no. 2, pp. 107–144, 1998.

[328] J. Berger and A. Schoeffler, "Recent advances in understanding structure-function relationships in the type II topoisomerase mechanism," *Biochemical Society Transactions*, vol. 33, no. 6, p. 1465, 2005.

[329] A. Vologodskii, "Theoretical models of DNA topology simplification by type IIA DNA topoisomerases," *Nucleic Acids Research*, vol. 37, no. 10, pp. 3125–3133, 2009.

[330] *MATLAB version 9.1.0 (R2016b)*, The Mathworks, Inc., Natick, Massachusetts, 2016.

[331] G. Rossum, "Python reference manual," Amsterdam, The Netherlands, The Netherlands, Tech. Rep., 1995.

[332] P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. L. de Hoon, "Biopython: Freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, pp. 1422–1423, 2009.

[333] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations," *Journal of Computational Chemistry*, vol. 32, no. 10, pp. 2319–2327, 2011.

[334] S. Buchoux, "FATSLiM: A fast and robust software to analyze MD simulations of membranes," *Bioinformatics*, vol. 33, no. 1, pp. 133–134, 2016.

[335] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, pp. 19–25, 2015.

[336] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, "MDTraj: A modern open library for the analysis of molecular dynamics trajectories," *Biophysical Journal*, vol. 109, no. 8, pp. 1528–1532, 2015.

[337] N. D. Matsakis and F. S. Klock, "The Rust language," in *Proceedings of the 2014 ACM SIGAda annual conference on High integrity language technology - HILT'14*, ACM Press, 2014.