EXPERIMENTAL AND THEORETICAL STUDIES OF DNA-MACROION INTERACTIONS

by

Shi Yu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in the University of Michigan
2015

Doctoral Committee:

Professor Ronald G. Larson, Chair
Professor L. Jay Guo
Professor Michael J. Solomon
Professor Robert M. Ziff

DEDICATION

To Mama and Baba.

ACKNOWLEDGMENT

I would like to express my deepest appreciation to my thesis advisor, Professor Ronald G. Larson. I would like to thank him for giving the opportunity to work on this PhD project in his group and giving me support and technical advice, from fundamental research direction to code debugging strategy, throughout this time. I admire his broad spectrum of knowledge in both theoretical physics and experimental techniques for biophysics. Without his guidance and persistent help this thesis would not have been possible.

I would like to thank my committee members, Professor Michael J. Solomon, Professor L. Jay Guo, and Professor Robert M. Ziff, who have given me suggestions and constructive criticism on this project. The light scattering experiments would not have been done without the help from Professor Michael J. Solomon and his student Dr. Mahesh Ganesan. I would like to thank Professor L. Jay Guo and his student Ashwin Panday for helping me in imaging experiments. And I would like to thank Professor Robert M. Ziff for his great instructions in statistical physics classes where I learned many simulation techniques.

I would like to thank Professor Nils Walter and his student Dr. Erika Cline, Professor Seok Ki Choi, Dr. Ming-Hsin Li for their tremendous help on biophysics experiments. I am thankful to Professor Ioan Andricioaei and Emel Ficici for insightful discussions. And I am particularly thankful to Dr. Shihu Wang in Professor Ronald G. Larson group for countless

stimulating discussions and great help in tough debugging process.

I would also like to thank both past and present Larson group colleagues. Indranil Saha Dalal, Xueming Tang, Zhicheng Long, Xue Chen, Lei Jiang, Priyanka Desai, Jun Liu were sources of valuable technical assistance during the course of my research.

I am also indebted many thanks to friends in the department, who have made my experience as doctorate candidate a truly memorable one. Among these individuals, I would like to thank Jaesung Lee, Hao Chen, Yu Chen, Jason Huang, Yingda Lu.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

The thesis uses experiments and simulations to examine the interactions of DNA molecules with proteins and protein-like nanoparticles, with applications to protein search and targeting of DNA sequences, and to DNA complexation in chromatin and for gene delivery. Two topics are covered in depth. In the first topic, kinetic Monte Carlo simulations, one dimensional reaction-diffusion equations, and analytical methods are used to determine rate at which DNA-binding proteins (e.g. transcription factors) can find the target sequences in long DNA molecules through a combination of sequence-dependent 1D diffusion and sequence-independent 3D diffusion. We quantify how thousands of "decoy sites" which have similar base pair sequences as target sites slow down the protein targeting process dramatically. We find the conditions under which the protein targeting process can be sped-up, including the effect of a "two-state" protein model, allowing for both rapid diffusion and accurate searching.

In the second topic, we investigate how the surface charge density of a poly(amido amine) (or PAMAM) dendrimer affects its ability to condense on DNA, using light scattering, circular dichroism, and single molecule imaging of dendrimer-DNA complexes combed onto surfaces and tethered to those surfaces under flow. This study is important not only for understanding how to condense dsDNA to facilitate its penetration into cell membranes for

non-viral gene therapy, but also because PAMAM dendrimers provide an ideal biomimic of DNA-binding proteins (e.g. histones). To describe DNA compaction by dendrimers, we develop a mesoscale model combining a coarse-grained DNA model of de Pablo and coworkers which resolves the DNA double helix structure with a coarse-grained dendrimer model of Muthukumar and coworkers. The predictions of our new model on effects of dendrimer generation, dendrimer surface charge density, and salt concentration on dendrimer-DNA complexes formation are consistent with both experiments and potential of mean force results from all-atom molecular dynamics simulations, but give much more detail regarding the structure of the complex. The model predicts that DNA wraps a fully charged G5 (generation 5) or G6 dendrimer at low salt concentration (10 mM) similarly to a histone octamer, and for the G5 dendrimer, DNA super helices with both handednesses occur. At salt concentrations above 50 mM, or when a high fraction of dendrimer surface charges are neutralized by acetylation, DNA adheres, but does not compactly wrap, the dendrimer, in agreement with experimental findings. We are also able to simulate pairs of dendrimers binding to the same DNA strand. Thus, this mesoscale simulation could be a good starting point of understanding chromatin formation. Moreover, this model can be extended to other cationic macroion-DNA systems which are also of great interest, such as, polylysine, micelles, and colloidal particles.

CHAPTER I

Introduction

1.1 DNA-protein interaction

Deoxyribonucleic acid (DNA) - protein interaction plays a vital role in many biological reactions, such as DNA packaging, replication, genetic recombination, DNA repair, and transcription [1]. Although the hereditary information is mainly stored in DNA, to survive in an ever-changing environment living cell needs to know how and when to transfer the hereditary information from DNA to RNA and finally to protein. In other words, the realization of central dogma [2] relies on DNA-protein interaction.

Generally, the DNA-protein interaction can be classified into two categories: non-specific interaction and specific interaction. In eukaryotic cells, the double stranded DNAs are packaged into chromatin by histones [3] which can bind to DNA segments with various base pair sequences. Such base pair sequence-independent interaction is one example of non-specific interaction. For specific interactions, e.g. transcription initiation, DNA-binding protein (transcription factor) can find a targeting DNA segment with a specific sequence and initiate transcription. This targeting segment, also known as promoter, usually contains a TATA box (5'-TATAAA-3') for transcription factors to bind [4].

Since manipulation of DNA-protein interaction can regulate the expression of genes, various methods have been developed to study the DNA-protein interaction. For example, electrophoretic mobility shift assays is used to determine the potential of a DNA sequence to bind a DNA-binding protein [5]. This technique is based on the observation of electrophoretic mobility of protein bound to DNA fragments in no denaturing polyacrylamide or agarose gels, as protein-DNA binding reduces its electrophoretic mobility. Some other techniques like DNase I footprinting [6], scanning probe microscope (SPM) [7], surface plasmon resonance technology (SPR) [8] are also widely used for determining the binding affinity and kinetics of protein-DNA interaction.

To study and image the DNA-protein interaction in real time, a series of novel techniques called *single-molecule experiments* have been developed. Some thorough reviews on single-molecule experiments were published recently [9, 10]. Many single-molecule imaging techniques use fluorescence microscopy to break the diffraction-limited resolution barrier of conventional optical microscope. The novel super-resolution microscopies, such as stochastic optical reconstruction microscopy (STORM) [11], photoactivated localization microscopy (PALM or FPALM) [12], even have nanometer resolution. Thus, those super-resolution microscopy can be extremely useful to resolve the protein and other biological macromolecules dynamics in living cells. In 2014, Eric Betzig, Stefan W. Hell, and William E. Moerner received Nobel Prize in Chemistry for the development of super-resolved fluorescence microscopy.

Optical tweezers [13] and magnetic tweezers [14] are usually used to hold DNA or protein for fluorescence microscope imaging *in vitro*. Simple techniques like DNA combing [15] and DNA curtains [16] were also used to immobilize DNA molecules on surface for further experiments. As shown in Fig.1.1.1, fluorescent λ-DNA molecules which were spin-

coated onto a polystyrene coated surface were imaged by epi-fluorescence microscope, as the water-air surface tension was able to stretch DNA on polystyrene surface during spin-coating DNA solution.

For protein labeling, protein can either be conjugated to quantum dots [17] or fluores-

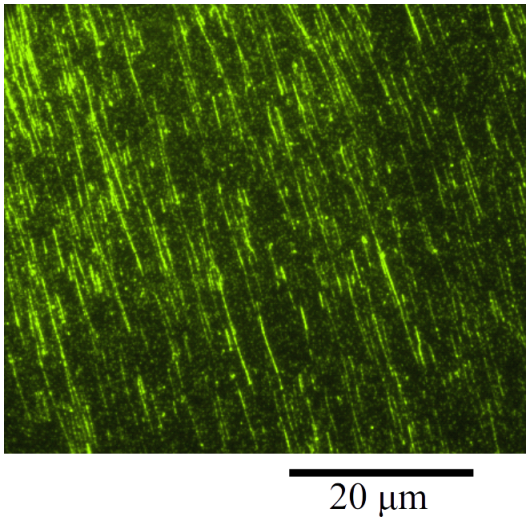

Figure 1.1.1: epi-fluorescence microscopy image of spin-coated fluorescent λ-DNA on polystyrene coated glass surface

cent antibodies [18]. Images of CdTe quantum dots taken by conventional fluorescence microscope and total internal reflection fluorescence microscope (TIRFM) are shown in Fig.1.2.1. Although fluorescence microscope imaging has many advantages in studying protein-DNA interaction in real time, to resolve the finer structures of DNA-protein complexes, especially chromatin fibers, cryogenic electron microscopy (cryo-EM) must be used [19].

In addition, extensive theoretical studies on DNA-protein interaction, especially the protein targeting process, have been carried out recently. How a protein finds its target on DNA will be discussed in details in Chapter 2.

1.2 DNA-macroion interaction

Deoxyribonucleic acid (DNA) is a highly charged semiflexible polyelectrolyte. Therefore, it not only can bind positively charged protein (e.g. histone) but also can bind negatively charged polyeletrolyte [20], dendritic polymer or dendrimer [21], positively charged nanoparticles (e.g., positively charged CdTe quantum dots, see Fig.1.2.1), and microparticles (see Appendix C).

Since polycation is a promising DNA delivery vector for gene therapy, many experiments

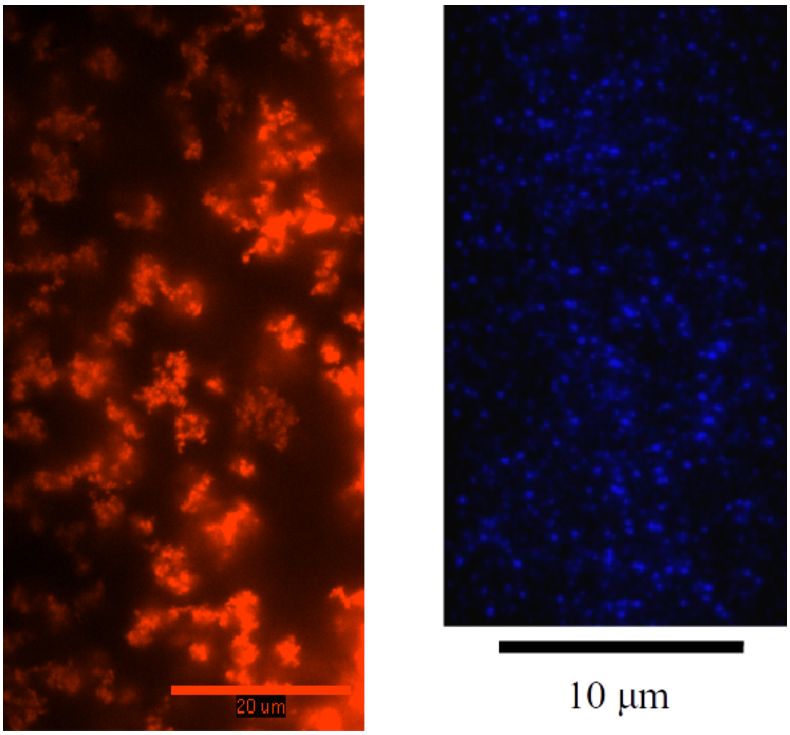

Figure 1.2.1: epi-fluorescence microscopy image of positively charged CdTe nanoparticles (*left*); total internal reflection fluorescence microscopy image of negatively charged CdTe nanoparticle (*right*)

were performed to study DNA-macroion (specifically, polycation) interaction. PAMAM dendrimers which can condense dsDNA and penetrate cell membranes have a great potential

for gene delivery. How the size and surface charge density of PAMAM dendrimer affect its capability of condensing DNA double strands will be investigated by both experiments and simulation which will be discussed in details in chapter 3 and chapter 4.

And as will be discussed in chapter 2, many DNA-binding protein can slide along DNA molecule. So, using polycations like quantum dots or PAMAM dendrimers as a mimic of DNA-binding protein diffusing on DNA is an interesting topic. Relevant experimental results can be found in Appendix B.

Moreover, polyeletrolytes adsorption onto curved surface is an very interesting problem in physics. Theoretically, it can be solved by Poisson-Boltzmann equation (PB) (Eq.1.2.1) [22–24].

$$\nabla^2\psi = -\frac{\rho_e}{\varepsilon\varepsilon_0} \tag{1.2.1}$$

where ψ is electrostatic potential, ρ_e is the local electric charge density in C/m^3, ε is the dielectric constant of the solvent, and ε_0 is the permittivity of free space.

However, the continuity assumption for PB equation might fail for this DNA-polycation interaction, as DNA and polycation are not much larger in size than water molecules or salt ions. Therefore, modeling technique with higher resolution like molecular dynamics simulation might be necessary for tackling this problem.

1.3 Simulations of DNA-protein/macroion interactions

Due to the accumulation of resolved protein and DNA structures by NMR and XRD, many force field which can be used for all-atom molecular dynamics (AAMD) simulations of nucleic acid-protein interaction are available [25]. However, AAMD simulation of DNA-

protein/macroion interaction is extremely computational expensive. Since DNA, a stiff long polymer chain, is much more difficult to be equilibrated by MD simulation compared to globular protein. So, developing coarse-grained model for DNA-protein interaction becomes very crucial. One successful coarse-grained model which we adopted to simulate DNA-PAMAM dendrimer interaction is 3SPN.1 model, a newer version 3SPN.2 force field is published recently [26, 27]. Using this model, Freeman et al. [28] demonstrated that histone octamers have different binding affinity to different DNA sequences, as the DNA shape can vary slightly according to the local base pair sequences. This 3SPN.2 force field can be used to simulate both A-form DNA, classical B-form DNA, and DNA with explicit salt ions (as shown in Fig.1.3.1). Detailed information about this model can be found in chapter 4.

Another recent research [29] provided a new method to determine the partial charge on

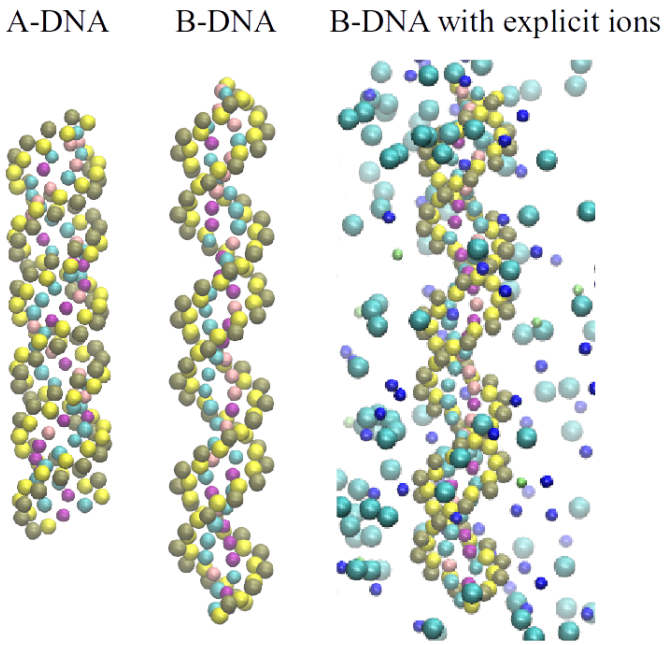

Figure 1.3.1: 3SPN.2 model of different DNA molecules

the charged amino acid residues of a protein molecule, and this method was used to study the interaction between a short nucleic acid (3SPN.2 force field) and various coarse-graind proteins.

Zhang et al. [30] have developed a discrete surface charge optimization (DiSCO) algo-

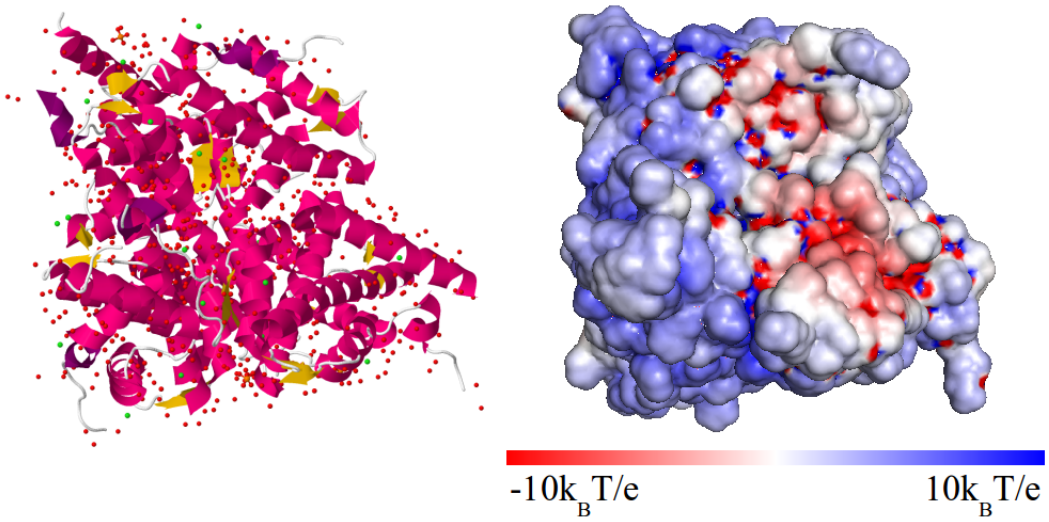

Figure 1.3.2: Native histone octamer (*left*); electrostatic potential of native histone octamer (*right*)

rithm which can uses hundreds of point charges to represent a macromolecular surface of almost the same electrostatic potential distribution on the surface. For example, as shown in Fig.1.3.2 & 1.3.3, the electrostatic potential of native histone [31] and T7 RNA polymerase [32] can be computed using PDB2PQR [33] and APBS [34]. To represent the surface electrostatic potential of these biomacromolecules, about 10^6 grid points are used. Using the DiSCO method, about 400 point charges are sufficient to model these charged macromolecules, where the variance of electrostatic potential field is less than 5%.

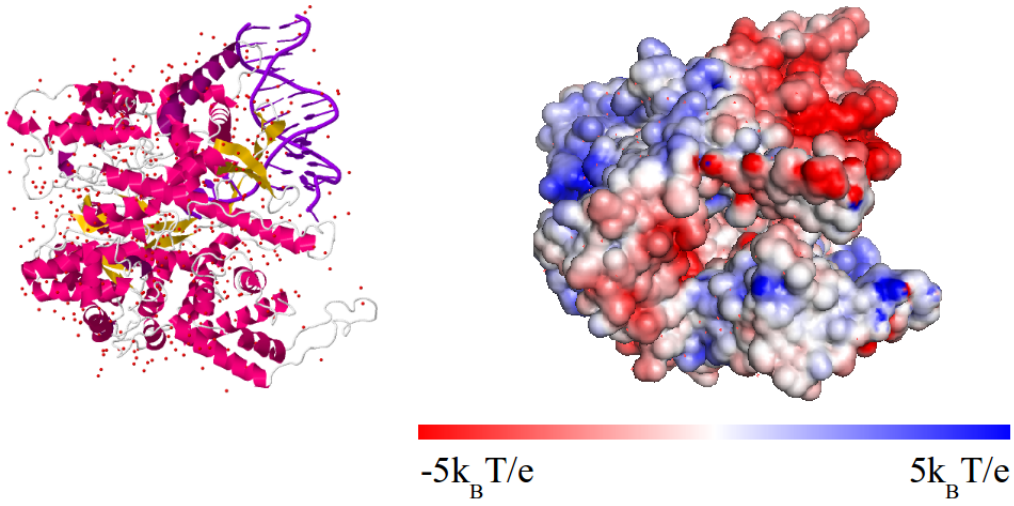

Figure 1.3.3: T7 RNA polymerase-promoter complex (*left*); electrostatic potential of T7 RNA polymerase-promoter complex (*right*)

1.4 Summary

In sum, in chapter 2, various diffusion-reaction models will be discussed to provide a new explanation of how the DNA-binding protein overcomes the "search speed-stability" paradox. In chapter 3, experimental results of interactions between PAMAM dendrimer acetylated to different extents and DNA molecules will be presented. In chapter 4, a new coarse-graind model will be proposed to simulate the PAMAM dendrimer-DNA interactions which can give us more structural information of DNA-dendrimer complexes then experiments.

CHAPTER II

Proteins searching for their target on DNA by one-dimensional diffusion: overcoming the "speed-stability" paradox

The sequence dependence of DNA-protein interactions that allows proteins to find the correct reaction site also slows down the 1D diffusion of the protein along the DNA molecule, leading to the so-called "speed-stability paradox", wherein fast diffusion along the DNA molecule is seemingly incompatible with stable targeting of the reaction site. Here, we develop diffusion-reaction models that use discrete and continuous Gaussian random 1D diffusion landscapes with or without a high-energy cut-off, and two-state models with a transition to and from a "searching" mode in which the protein diffuses rapidly without recognizing the target. We show the conditions under which such considerations lead to a predicted speed-up of the targeting process, and under which the presence of a "searching" mode in a two-state is nearly equivalent to the existence of a high-energy cut-off in a one-state model. We also determine the conditions under which the search is either diffusion-limited or reaction-limited, and develop quantitative expressions for the rate of successful targeting as a function of the site-specific reaction rate, the roughness of the

DNA-protein interaction potential, and the presence of a "searching" mode. In general, we find that a rough landscape is compatible with a fast search if the highest energy barriers can be avoided by "hopping" or by the protein transitioning to a lower-energy "searching" mode. We validate these predictions with the results of Brownian dynamics, kinetic Metropolis, and Kinetic Monte Carlo simulations of the diffusion and targeting process, and apply these concepts to the case of T7 RNA polymerase searching for its target site on T7 DNA.

2.1 Introduction

The long-standing problem of proteins (repressor, polymerase, transcription factor, etc.) seeking their target sites among millions of nonspecific sites along DNA is attracting increasing interest. To help explain how proteins might find their targets faster than allowed by ordinary three dimensional diffusion, in 1981 O. G. Berg et al. [36, 37] proposed a "facilitated diffusion" model for this process. Their model, as further elaborated by others, includes four mechanisms: (a) one-dimensional diffusion or "sliding" of a protein along the DNA contour, (b) "hopping" of a protein from one site on DNA to a nearby site a few base pairs away, (c) "jumping" of the protein to a distant site, mediated by 3D diffusion through solvent, and (d) "intersegment transfer" and "intersegmental jumps" [38] of the protein to a site on a DNA sequence close by in 3D space, but possibly many base pairs distant along the DNA contour (since the DNA typically has a coiled configuration). Since then, a large number of experiments [17, 18, 39–45] have been carried out to test this model, both *in vitro* and *in vivo*. Many models have been developed to describe the protein targeting, through a combination of 1D and 3D diffusion [46–48].However, since the 1D diffusion

coefficient of a protein moving along DNA is two or three orders smaller than the 3D diffusion coefficient, long-range sliding of a protein along DNA may slow down rather than speed up the targeting process relative to the predictions of the Smoluchowski equation for three-dimensional diffusion-limited targeting [49,50], a result that has been confirmed by Brownian dynamics (BD) simulations [51, 52]. Recently, Halford [53] reexamined the early experimental results [54] and claimed that the targeting speed of DNA-binding proteins (DBPs) does not exceed the true 3D diffusion limit, if one includes in the 3D diffusion the effect of electrostatic attraction between the cationic proteins and the anionic DNA in the Smoluchowski equation. In other words, the so-called "facilitated diffusion" mechanism does not really facilitate the targeting, but rather impedes it. Halford's conclusion remains controversial, however, and in any event single- molecule imaging experiments leave little doubt that proteins do actually slide along DNA [18, 41, 55]. Hence, the 1D diffusion of the protein along DNA remains a very interesting and important problem to understand in greater detail, regardless of whether it actually speeds up the target search or not. Most DNA-binding proteins (DBPs) interact with 5 to 20 base pairs at a time when forming nonspecific complexes with DNA. Both statistical mechanical analysis [56] and experiments [57] demonstrate that the interaction energies between a DBP and each base pair are almost independent of each other and additive. Because of the irregularity of the DNA base pair sequence, it is assumed that the sequence-dependent binding energy can be approximated by an uncorrelated (or random) Gaussian distribution [58]. The problem of 1D diffusion of particles in a random potential is well studied problem [59, 60] and it has also been used to model the motion of dislocations in disordered crystals [61], the unzipping of DNA [62], protein folding [63], and other problems. Zwangzig [64] determined theoretically the simple exponential relationship between the 1D diffusion coefficient D

in a random Gaussian one-dimensional potential and the standard deviation, σ, or roughness, of the potential (Eq. 2.1.1), which was subsequently rediscovered by Slutsky and Mirny [58, 65]:

$$D = D_0 exp\left[-\left(\frac{\sigma}{k_B T}\right)^2\right] \tag{2.1.1}$$

Based on (Eq. 2.1.1), Slutsky and Mirny noted that a fast search process requires that the roughness of the one-dimensional potential be small ($\sigma = 1 \sim 2\ k_B T$), which seemingly conflicts with the requirement that the DNA-protein specific complex be thermodynamically stable once it forms ($\sigma \sim 5\ k_B T$) [58]. To resolve this so-called "search-speed/stability paradox", Slutsky and Mirny proposed a "two-state model", which allows the protein to change its conformation to perform fast diffusion in a "searching" state and bind stably to the target site in a "recognition" state. A recent single-molecule experiment [17] has confirmed that p53, and fragments of p53 that contain either the "recognition domain" or the "sliding domain", diffuse differently along DNA in a manner consistent with this "two-state" mechanism.

However, since the DBP finds its target by a combination of 1D and 3D diffusion and the transition from 1D to 3D diffusion should be sequence-dependent, especially for a rough potential, a Gaussian random potential might be an unrealistic model of the actual 1D potential landscape that the protein experiences when sliding along the DNA major groove. One important source of non-randomness is the likely relatively high-energy barrier to motion from one site to the next along the DNA, since the protein-DNA interaction is dominated by highly localized hydrogen bonds between DNA and protein that must be broken for the protein to slide a distance of one base pair along the DNA. Since these barriers must be overcome whether the roughness of the landscape is small or large, Eq. 2.1.1 may overestimate the effects of potential roughness on the 1D diffusion coefficient. Another

consideration is that the sequence-independent electrostatic interaction might decay more slowly in the direction perpendicular to the DNA than do sequence-dependent forces, and hence this relatively weak electrostatic force might by itself be sufficient to reduce the dimensionality of protein Brownian motion from 3D to 1D [66]. A protein can diffuse a very short distance away from the DNA surface to get over a high-energy barrier resulting from sequence-dependent forces, while easily staying close enough to the DNA that electrostatic forces can maintain the 1D diffusion. Thus, the real 1D potential landscape along DNA that the protein experiences during sliding might be better approximated by imposing a maximum-energy threshold onto an otherwise Gaussian-distributed positionally random potential. So, in this thesis, we discuss how introducing a threshold into the 1D random potential can reduce the dependence of 1D diffusivity of an over-damped particle on the strength of a Gaussian disordered potential. We consider both a discrete 1D potential landscape and also propose a continuous random 1D potential based on Barbiâs hydrogen-bonding model [67,68] for the sequence-dependent interaction potential, which we will show reconciles fast sliding with stability of targeting. We also compare the diffusivities along random potentials with wells of various shapes to determine the sensitivity of 1D diffusivity to the shape, as well as the depths, of random potential wells. We obtain diffusivities for these landscapes in three different ways and show that they agree with each other. Using these diffusivities, we determine using both theory and simulations the rate at which the protein finds its target in reaction- and diffusion-limited regimes.

Unlike synthetic nanoparticles, a protein can change its conformation frequently. We therefore also propose a simple two-state model, similar to that of Mirny and coworkers, to explore whether the sliding motion of DBPs can be accelerated substantially by switching their conformations. In our simple model, the transitions between the two states on differ-

ent sites are controlled by a constant activation barrier; in this respect our model differs from most previous two-state models [65, 69]. In addition, we choose transition rates between states that satisfy detailed balance, guaranteeing that our model does not violate thermodynamic requirements [70]. We then carry out kinetic Monte Carlo simulations to see if this simple two-state model improves the targeting process.

Before discussing our diffusion models, we address some general conditions that govern the rate of target search and the conditions under which this search is controlled by the overall one-dimensional diffusivity along the DNA, or by the rate of reaction at the target site, or by both. The "search-speed/stability paradox" mentioned above arises only if both the one-dimensional diffusion coefficient and the rate of reaction at the target site both influence the overall targeting rate. If the rate of reaction at the target site is slow compared to the rate of diffusion, then most proteins will not react the first time they reach the target site, but will occupy sites along the DNA with probability given by simple Boltzmann statistics, governed by the site-specific free energy. The rate of reaction will then be controlled by this occupancy probability and the reaction rate constant, and not by the diffusion speed. On the other hand, if the reaction rate at the target site is very fast, then any protein reaching the target will react immediately, irrespective of whether the protein remains on the site for very long, and the "stability" of the binding site will be unimportant, while the speed of diffusion dominates the targeting rate. To demonstrate this quantitatively, in the next section we derive the targeting rate for a simple 1D reaction-diffusion model, and then show how the effective diffusion coefficient and effective reaction rate are influenced by properties of the one-dimensional potential.

2.2 Theory and models

2.2.1 Diffusion-limited vs. reaction-limited targeting

We assume that away from the binding site, the targeting proteins can associate with DNA with overall rate constant k_{on}, and dissociate from DNA with rate constant k_{off}. The bulk concentration of protein can be considered to be a constant c_{bulk} . Then, the concentration profile of the protein on DNA can be approximated simply by a steady-state analysis balancing flux of proteins to and from the DNA and 1D diffusion along the DNA molecule. At steady state, we have:

$$D\frac{\partial^2 c}{\partial x^2} + k_{on}c_{bulk} - k_{off}c = 0 \quad (x \leq 0) \tag{2.2.1}$$

where *D* is the effective 1D diffusion coefficient of protein along DNA, in units of inverse time, c is the protein concentration (or probability of occupancy) along the DNA, and x is a dimensionless distance along the DNA molecule in units of base pairs. To simplify the analysis, we divide the one-dimensional domain into a "far field", where we can take the diffusivity D, and k_{on}, and k_{off} to be constants that are pre-averaged over the free-energy landscape, and the "near field", which is the target site, where we will take account of the strength of binding of the DNA, which will affect its probability of reacting there. Within this approximation (which we will later test using simulations), the concentration of protein along the DNA can be obtained by applying the boundary conditions:

$$\begin{aligned} c = c_\infty = (k_{on}/k_{off})\, c_{bulk} \quad (x \to \infty) \\ -2D\frac{\partial c}{\partial x} - \Omega^{eff}c = 0 \quad (x = 0) \end{aligned} \tag{2.2.2}$$

where Ω^{eff} is the effective reaction rate constant at the target site $(x = 0)$. We only consider proteins to the left (negative side) of the target site $(x \in (-\infty, 0])$ and account for diffusion from the positive side to the target site at $x = 0$ by the factor of 2 multiplying the diffusivity in Eq. 2.2.2. The effective reaction rate $\Omega^{eff} = \Omega exp\left[\frac{\mu - E_T}{k_B T} - \frac{1}{2}\left(\frac{\sigma}{k_B T}\right)^2\right]$ (see Appendix A) accounts for the residence time of the protein at the target site as well as rates of transition to and from one "state" to the other in the case of a two-state model, as will be discussed below. We also assume that $k_{off} \ll D$, so that 1D diffusion is an important part of the target-searching. (If $k_{off} > D$, the proteins will not remain on the DNA long enough to slide one dimensionally and so will reach the target site primarily by three-dimensional diffusion.) Since we take $k_{off} \ll D$, we can neglect the terms involving k_{off} and k_{on} in the second boundary condition above, so that we disallow the protein to hop on or off the DNA at the target site. (The small contribution of this could be easily included, if desired.) Solving Eq. 2.2.1 and Eq. 2.2.2, the protein concentration along the DNA is:

$$c = \frac{k_{on}}{k_{off}} c_{bulk} - \frac{k_{on}}{k_{off}} c_{bulk} \left(\frac{\Omega^{eff}}{2D\sqrt{\frac{k_{off}}{D}} + \Omega^{eff}}\right) exp\left(x\sqrt{\frac{k_{off}}{D}}\right) \tag{2.2.3}$$

where we only consider $x < 0$. Note in the above, that the typical diffusion distance traveled by the protein between its adsorption from bulk solution and its subsequent desorption, in units of base pairs, is given by $\sqrt{\frac{D}{k_{off}}}$. The reaction rate is given by:

$$r = \Omega_{eff} c_{x=0} = k_{on} c_{bulk} \left(\frac{2\sqrt{D/k_{off}}}{1 + \frac{2\sqrt{D k_{off}}}{\Omega^{eff}}}\right) \tag{2.2.4}$$

When $\Omega^{eff} \gg \sqrt{Dk_{off}}$, the targeting process is diffusion limited, the concentration of protein at the target site is very small, and the reaction rate is given by:

$$r = 2k_{on}c_{bulk}\sqrt{\frac{D}{k_{off}}} = 2\frac{(k_{on}/k_{off})\,c_{bulk}D}{\sqrt{D/k_{off}}} = \frac{2c_{\infty}D}{\sqrt{D/k_{off}}} \tag{2.2.5}$$

This rate, r, can also be written as the product of the diffusivity D times the 1D concentration gradient of protein, which is the concentration of protein far from the target, namely $c_{\infty} \equiv (k_{on}k_{off})\,c_{bulk}$, divided by the diffusion distance $\sqrt{Dk_{off}}$. The factor of two again arises becomes the protein can diffuse to the target from either the right or the left. We performed a series of Metropolis Monte-Carlo simulations over a periodic landscape of 2,000 base pairs with different roughness using the Metropolis scheme to verify Eq. 2.2.5. The simulations were performed by randomly introducing a protein along the periodic landscape, and simulating its diffusion, until it either reached the target site and reacted, or dissociated from the DNA. The sites were treated as discrete with energies E_i distributed with Gaussian distribution, $E_i = N(0, \sigma)$, and hopping from one site to the next was accomplished by the usual Metropolis method. Thus, randomly selected diffusion moves between neighboring sites were accepted with probability given by the Boltzmann factor for the energy difference between the sites, if the move is unfavorable, or unity otherwise. The probability of dissociation $k_{off,i}$ was also site-specific with probability given by a Boltzmann of the energy difference between the site and the energy in free solution, i.e., $k_{off,i} = exp\left[-\left(E_{bulk} - E_i\right)/k_BT\right]$. We adjusted the energy in free solution E_{bulk} according to the roughness of the potential along DNA by setting $E_{bulk} = -ln\left(D_0/250^2\right) - (\sigma/k_BT)^2$ to fix the value of the diffusion distance at $\sqrt{D/\overline{k_{off}}} \approx 250$, since $D \approx exp\left[-\left(\frac{\sigma}{k_BT}\right)^2\right]$, and $k_{off} \approx \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} k_{off,i} = exp\left[-E_{bulk} + (\sigma/k_BT)^2\right]$, as discussed further below. The

reaction rate is just the inverse of the average time between reaction events times the protein "concentration", which is 1/2,000 in this case, since proteins are tracked one at time over a landscape with 2,000 sites. The results, plotted in Fig. 2.2.1, show that Eq. 2.2.5 remains correct even for a rough potential landscape with standard deviation up to $\sigma = 2k_BT$. Note that in the upper insert, the occupancy of sites along the DNA is very "noisy" even for a relatively small roughness of $\sigma = 0.5k_BT$, not due to insufficient averaging over multiple proteins, but due to this roughness, which strongly favors some sites over neighboring ones. As the potential landscape becomes rougher, the noise of both the site occupancy and the local diffusion coefficient increases greatly, but Eq. 2.2.5 holds up to at least $\sigma = 2k_BT$, showing that our use of a far-field average diffusivity to compute the rate of diffusion to the reaction site remains valid at least to this level of roughness. (The occupancy of sites for $\sigma = 2k_BT$ is even "noisier" than for $\sigma = 0.5k_BT$ as shown in the bottom inset of Fig. 2.2.1.) For σ higher than this, the diffusion becomes too slow to obtain adequate averaging; however, the use of a far-field average diffusivity will eventually break down for large enough σ. We can estimate the point at which this will occur by considering the landscape-dependent variance of D_n, the diffusion coefficient averaged over the diffusion distance $n = \sqrt{D/k_{off}}$. As shown in the Appendix A, the variance of the D_n/D_0 can be estimated by $2\sigma^4/n$. So, when the diffusion distance is about 100 ($n \approx 100$, typical value for DBP's), the variance of Dn over this diffusion distance will be significant for σ larger than $3k_BT$. Thus, for $\sigma > 3k_BT$, instead of using the far-field 1D diffusion coefficient for D in Eq. 2.2.5, we would need to compute D_n from the local base pair sequence neighboring the target to determine the reaction rate in the *diffusion-limited* regime.

On the other hand, when $\Omega^{eff} \ll \sqrt{Dk_{off}}$, the targeting process becomes *reaction limited*

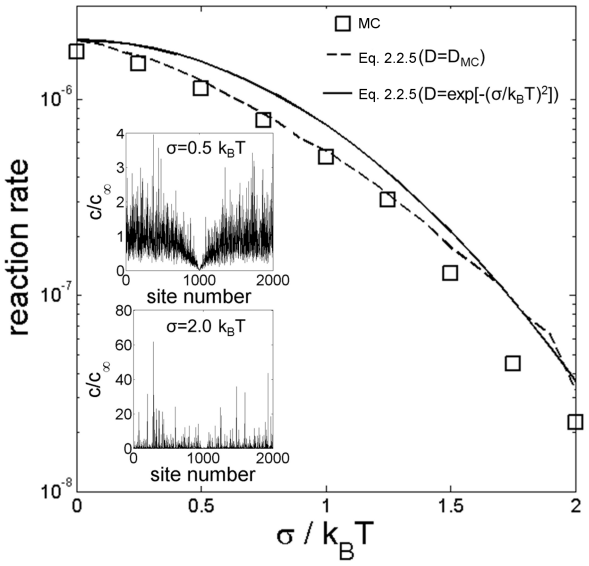

Figure 2.2.1: Steady-state reaction rate of a protein at the target site (site number = 1,000), assuming instantaneous reaction at the reaction site, for a discrete Gaussian-distributed landscape, obtained by Metropolis Monte-Carlo simulations (squares). Also shown is the reaction rate from Eq. 2.2.5, using MC simulations to determine the diffusion coefficient (dashed line); and using $D = exp\,[-(\sigma/k_BT)^2]$ to approximate the diffusion coefficient (continuous line). The insets show the rough concentration profiles over the 2,000 base pairs for $\sigma = 0.5k_BT$ and $\sigma = 2.0k_BT$

and the reaction rate is given by:

$$r = (k_{on}/k_{off})c_{bulk}\Omega^{eff} \tag{2.2.6}$$

Thus, in the reaction-limited regime, the protein targeting speed is only determined by the effective reaction rate at the target, and by the overall concentration of protein on the DNA, which is given by $(k_{on}/k_{off})c_{bulk}$, and not by the diffusivity. Note that the effective reaction rate must include the effect of the residence time of the protein at the target site, and thus by the stability of binding to the target. In the reaction-limited regime, diffusion

is fast enough compared to reaction that the protein concentration, or probability distribution, along the DNA reaches thermodynamic equilibrium. At the target site, the protein finds a local free energy minimum, and so its concentration there will be higher than elsewhere according to the ratio of Boltzmann factors for the target site relative to the average over all sites.

Recent experiments [44] have shown that lac repressor frequently ($> 90\%$) diffuses over its natural operator several times before binding, which indicates the targeting of lac repressor, is not diffusion-limited. Therefore, determination of conditions for diffusion-limited vs. reaction-limited targeting, as described above, is an important element of the analysis of the target search on DNA. One aspect of targeting *in vivo* not considered here is the role of proteins, such as histones, which can block 1D diffusion along DNA in the cell. Our current theory is applicable only to *in vitro* single-molecule experiments that lack such blockages to diffusion, but in the future, the effect of blocking proteins will need to be considered.

2.2.2 Diffusion models

In this section, we limit ourselves to the diffusion-limited case, where the protein targeting speed is determined by the 1D diffusion coefficient, and hence the targeting speed is given in Eq. 2.2.5. We will determine the "far-field" protein diffusivity for models with three different 1D potentials, namely 1) a Gaussian-distributed square-well potential with a threshold, 2) a Gaussian-distributed square-well potential with a threshold and regular barriers, and 3) a continuous disordered potential. These potentials are illustrated in Fig. 2.2.2a below. We will also consider 4) a two-state model (Fig. 2.2.2b), where the protein

can switch between a fast "searching state" with a weak or flat potential and a "recognition state" where it feels a rough Gaussian-distributed random potential.

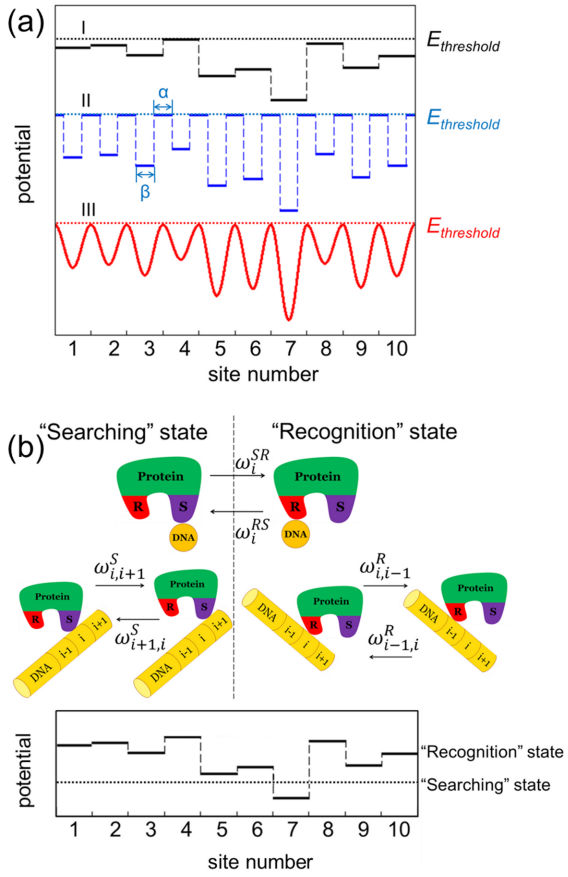

Figure 2.2.2: (a) 1D disordered potential landscapes (from *top* to *bottom*): I. Gaussian random potential with threshold; II. Threshold Gaussian random potential with activation barriers between each site; III. Sinusoidal disordered potential with threshold. (b) two-state model with transition rates between different sites or states (given by the ω's) determined by Eq. 2.2.18. "R" is the recognition state and "S" the search state. We will show later that in some cases the two-state model gives a diffusivity nearly identical to a one-state model with a low threshold set by the searching state energy as shown in the bottom diagram

We can calculate the diffusion coefficient along a random energy landscape using two different approaches. One is to compute the mean first passage time (MFPT) for the protein

to diffuse from a starting point $y = 0$ (where $y = 0$ is also a reflecting point to prevent particles from diffusing towards negative y) to a finishing point $y = L$ along a randomly generated landscape, and then average this MFPT over multiple randomly generated landscapes. This gives an average time tave to diffuse a given distance L, and the diffusion coefficient can then be estimated from the standard formula $D \approx L^2/2t_{ave}$. This MFPT can be obtained from the more general formula for a particle starting at position y [71]:

$$t(y) \propto \frac{1}{D_0} \int_y^L dy_1 e^{U(y_1)/k_B T} \int_0^{y_1} e^{-U(x)/k_B T} dx \tag{2.2.7}$$

and taking $y = 0$ in our calculation.

The second approach to obtain an analytical solution for the long-time D is to use the formula for a periodic potential derived by Festa and d' Agliano [72], namely:

$$D = \frac{D_0}{\langle e^{U/k_B T} \rangle \langle e^{-U/k_B T} \rangle} \tag{2.2.8}$$

where the brackets indicate an average over the period of the potential:

$$\langle e^{U/k_B T} \rangle = \frac{1}{L} \int_0^L e^{U(x)/k_B T} dx \tag{2.2.9}$$

This formula can be used for a non-periodic random potential by using quasi-random potentials that are random over a finite distance, beyond which they are periodic. The result for a random potential is then obtained by taking the limit of a long repeat distance L, i.e., taking $L \to \infty$.

Gaussian-distributed square-well potential with threshold

As discussed above, the hopping/jumping of a protein can be highly sequence-dependent.

Once the protein encounters a high free-energy barrier, it can retreat back, diffuse over the barrier, or be expelled into bulk solution. Because of this third possibility, the real energy landscape that the protein actually experiences cannot strictly be Gaussian disordered, since the high-energy tail of the distribution will have energy higher than the protein in free solution, and the protein will likely escape to free solution if a sequence with high enough energy in this tail is encountered. To simplify our model, we therefore assume that the protein will never feel a net energy that is higher than energy $E_{threshold}$. Thus, after generating an energy landscape from a Gaussian distribution, we set any energy higher than $E_{threshold}$ to $E_{threshold}$. This assumes that when the threshold energy is encountered, the protein briefly loses attachment to the DNA and then quickly binds to a neighboring site, rather than disappearing into the bulk. Hence, we are only including very short "hopping" motions of the protein, but not longer-range "jumping" or "intersegment transfer". We will also show below that in some cases, the diffusivity for the "two-state" model introduced below can be approximated by a one-state model with a low threshold energy, which provides additional motivation for studying this simple model. Thus, for these cases, our potential landscape can be approximated by random square-well potential with threshold, given below:

$$E_i = min\left[N(\mu, \sigma), \quad E_{threshold} + \mu\right], \quad i = 1, 2, 3, ... \tag{2.2.10}$$

We now compute the diffusivity over this potential using both the MFPT and the Festa and d' Agliano expression Eq. 2.2.8 and Eq. 2.2.9 above, with $L \to \infty$, and taking the mean site free energy $\mu = 0 k_B T$. For a Gaussian random potential with threshold Eq. 2.2.10,

this yields:

$$D = \frac{D_0}{\{e^{\widetilde{\sigma}^2/2}\left[1-\Phi\left(\frac{-\widetilde{\nu}+\widetilde{\sigma}^2}{\widetilde{\sigma}}\right)\right]+e^{\widetilde{\nu}}\left[1-\Phi\left(\frac{\widetilde{\nu}}{\widetilde{\sigma}}\right)\right]\}\{e^{\widetilde{\sigma}^2/2}\Phi\left(\frac{\widetilde{\nu}+\widetilde{\sigma}^2}{\widetilde{\sigma}}\right)+e^{-\widetilde{\nu}}\left[1-\Phi\left(\frac{\widetilde{\nu}}{\widetilde{\sigma}}\right)\right]\}} \quad (2.2.11)$$

where we have made the equation more compact by defining $\nu = E_{threshold}$, $\widetilde{\nu} = \nu/k_BT$, $\widetilde{\sigma} = \sigma/k_BT$. In the following discussion, all parameters with a tilde will represent these parameters normalized by k_BT. In the above, Φ is the cumulative distribution function of the normal distribution:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt = \frac{1}{2}\left[1+erf\left(\frac{x}{\sqrt{2}}\right)\right] \quad (x \in R) \quad (2.2.12)$$

Eq. 2.2.11, which is based on the equation of Festa and d' Agliano, agrees very well with the numerical results we obtain from the MFPT, which are plotted in Fig. 2.2.3a. When $\widetilde{\sigma}$ is large ($\sigma \gg |\nu|$), Eq. 2.2.11 can be reduced to:

$$D = 2D_0 exp\left(-\widetilde{\nu} - \frac{1}{2}\widetilde{\sigma}^2\right) \quad (2.2.13)$$

As shown in Fig. 2.2.3b, Eq. 2.2.11 and Eq. 2.2.13 agree with each other for $\sigma \gg |\nu|$, whether ν is positive or negative. We will later show that a very similar formula also results from analysis of the diffusivity in the two-state model, which we will discuss later. For a Gaussian disordered potential with a threshold (Eq. 2.2.10), the diffusion coefficient depends strongly on both $E_{threshold}$ (or ν) and σ. We note here that σ is the standard deviation of the Gaussian energy distribution before applying the cut-off, and the standard deviation of the actual potential is therefore less than σ. When $E_{threshold}/\sigma$ is large, only a very small fraction of the Gaussian-distributed energies are cut off by

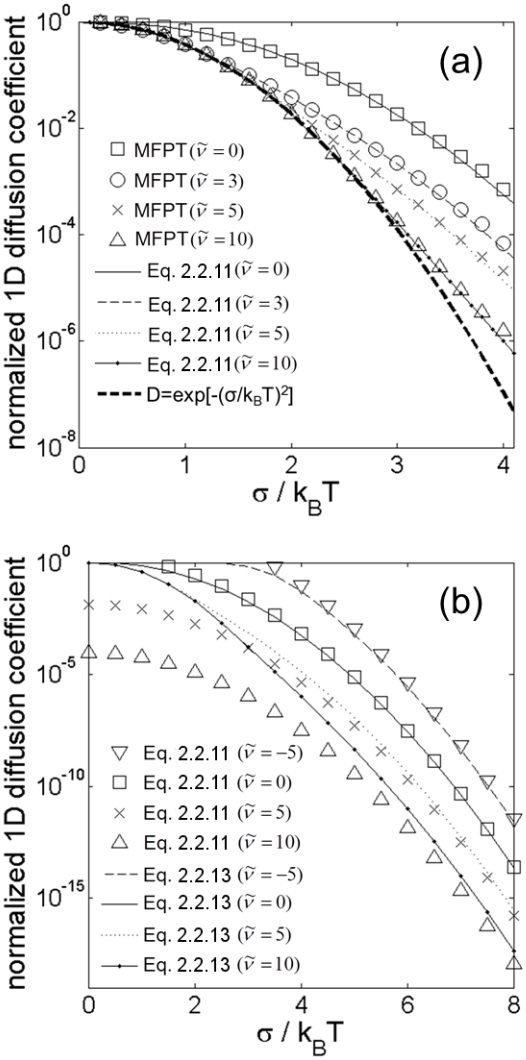

Figure 2.2.3: (a) Normalized diffusivity D/D_0 vs. Gaussian standard deviation σ for Model I, namely a Gaussian-distributed potential with a threshold (Eq. 2.2.10). Symbols are obtained numerically from the MFPT, while lines are calculated from Eq. 2.2.11. Note that σ describes the standard deviation of the Gaussian potential, while the true standard deviation is lower than this because of the cut-off applied to the tail of the Gaussian. (b) Normalized diffusivity D/D_0 vs. Gaussian standard deviation σ for different energy threshold, *lines* are obtained by Eq. 2.2.11, *symbols* are obtained by Eq. 2.2.13

$E_{threshold}$. Therefore, for small σ, Eq. 2.2.11 almost converges to Eq. 2.1.1 ($E_{threshold} > 0$ k_BT). The difference between Eq. 2.2.11 and Eq. 2.1.1 increases with $sigma$. According

to a model of Slutsky and Mirny [58], σ should be at least $5k_BT$ to maintain the stability of the DNA-protein specific complex. In this case, the value of D given by Eq. 2.2.11 can be as much as 10^4 times higher than that given by Eq. 2.1.1, if $E_{threshold}$ is small enough ($3k_BT$); see Fig. 2.2.3a. By calculating the partition function numerically, we find that the fraction of time that the protein spends on the target site when system is in equilibrium for small $E_{threshold}$ ($E_{threshold} = 3k_BT$) and large roughness ($\sigma = 5k_BT$) is about 0.25, which is the same value as obtained for the Gaussian-distributed potential without threshold [58], because the threshold only affects high-energy states that contribute little to the partition function. Therefore, if the highest-energy states along the DNA can be eluded by hopping/jumping of the protein short distances along an otherwise Gaussian-distributed random potential (or by transitioning to a searching mode as discussed below), then the diffusion coefficient can be increased greatly without compromising the fidelity of targeting. Thus, the "speed-stability paradox" might be overcome by brief excursions of the protein from the DNA surface when the high-energy states are encountered, thus reducing the effective roughness of the landscape. We also note that these few high-energy states might be selected against by mutation of the few base pairs responsible for them, and this would be favored by natural selection if they indeed have a large negative effect on protein mobility. In any case, we might expect that since the high-energy states have a large effect on mobility but a small effect on binding stability, we should consider the possibility that they are somehow avoided, and we have modeled this avoidance by a simple Gaussian landscape with high energy-states truncated by a threshold.

Gaussian-distributed square-well potential with threshold and regular barriers

During translocation from one site to the next site, the protein may have to overcome activation barriers (E_a) because of charge mismatch or hydrogen bond deformation and

breakage. Thus, we consider a random potential augmented by a periodic potential to add activation barriers into our model. All activation barriers E_a and site potential wells are modeled by square barriers with width α and square wells of the width β, respectively ($\alpha + \beta = 1$).

$$E_{2j} = min[N(\mu, \sigma), E_{threshold} + \mu], \quad E_{2j+1} = E_a \quad (j = 0, 1, 2, 3, ...) \tag{2.2.14}$$

where $N(\mu, \sigma)$ is again a normal (or Gaussian) distribution with mean μ and standard deviation σ. For simplicity, let $\mu = 0$, $E_{threshold} = E_a = \nu$ (shown in Fig. 2.2.2a II). By applying Eq. 2.2.8 to this random potential, we have:

$$D = \frac{D_0}{\left\{\alpha + \beta\left\{e^{-\widetilde{\nu}+\widetilde{\sigma}^2/2}\left[1 - \Phi\left(\frac{-\widetilde{\nu}+\widetilde{\sigma}^2}{\widetilde{\sigma}}\right)\right] + 1 - \Phi\left(\frac{\widetilde{\nu}}{\widetilde{\sigma}}\right)\right\}\right\}\left\{\alpha + \beta\left[e^{\widetilde{\nu}+\widetilde{\sigma}^2/2}\Phi\left(\frac{\widetilde{\nu}+\widetilde{\sigma}^2}{\widetilde{\sigma}}\right) + 1 - \Phi\left(\frac{\widetilde{\nu}}{\widetilde{\sigma}}\right)\right]\right\}} \tag{2.2.15}$$

The relations between 1D diffusion coefficients and the roughness of different random potentials are plotted in Fig. 2.2.4. Numerical results obtained using the MFPT agree well with analytical results based on the equation of Festa and d' Agliano (Eq. 2.2.15). When high activation barriers are introduced into this random potential in Model II, the sliding of the protein is dominated by these activation barriers. Thus, the random potential has even less effect on the 1D diffusion. Such periodic barriers are probably important for DNA-protein systems, because many DNA-protein nonspecific complexes are stable enough for the structure of those complexes to be resolved [73]. While such barriers make the diffusivity less sensitive to landscape roughness, they will slow down the overall one dimensional diffusion, relative to that for bulk 3D diffusion. (Note that the diffusivity plotted in Fig. 2.2.4 is normalized by the diffusion coefficient for no roughness, but with barriers.)

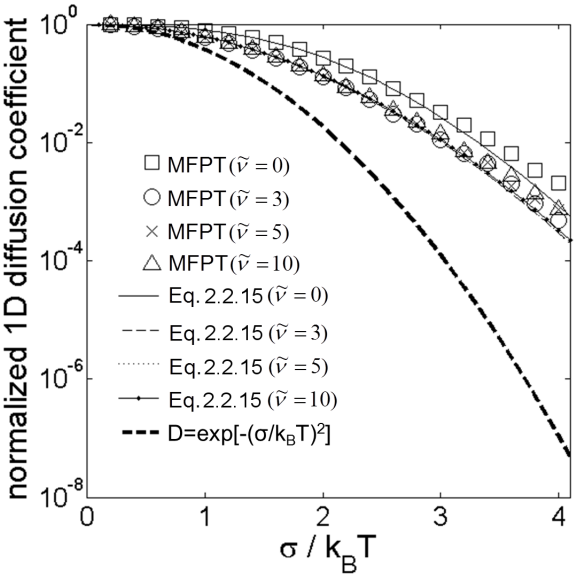

Figure 2.2.4: Normalized diffusivity D/D_0 vs. roughness σ for Model II, which is a Gaussian-distributed potential with a threshold and an energy barrier (Eq. 2.2.14), where the threshold and barrier are the same height. *Symbols* are obtained from the MFPT while lines are obtained from Eq. 2.2.15, with $\nu = E_{threshold}$, $\alpha = \beta$

Sinusoidal disordered potential

In reality, of course, the DNA-protein interaction energy landscape is not discrete. So, here we also consider a continuous potential landscape with continuous first derivative, which is given by:

$$E(x + i \cdot l) = E_i \left[sin^{2n} \left(\pi \frac{x}{l} \right) \right], \quad (0 \leq x < l, \quad i = 1, 2, 3, ...) \tag{2.2.16}$$

where l is the width of each "well" of the potential, where the potential returns to zero at positions 0, l, $2l$, etc., and each well has a depth of E_i (a negative quantity) halfway between these points. The exponent $2n$ sets the shape of this periodic potential, with larger n leading to steeper walls of the potential wells. A periodic potential is obtained by setting

all E_i to the same value E.

We now obtain the relationship between D and potential well depth E for periodic potentials, including the continuous periodic potential above, with n = 1, and 1,000, as well as the periodic potential with alternating square wells and square barriers of the same width using Eq. 2.2.8. The results are presented in Fig. 2.2.5. Note that for shallow potential wells ($< 3k_BT$), D decreases slowly as the well depth increases and then drops exponentially with increasing energy well depth at depths greater than around $5k_BT$. Taking the continuous well wall steepness towards infinite steepness by setting n to infinity, the periodic potential ceases to impede diffusing particles, as it becomes flat with infinitely narrow, delta-function, peaks. This illustrates that the shape of the 1D potential along DNA strongly affects the diffusivity.

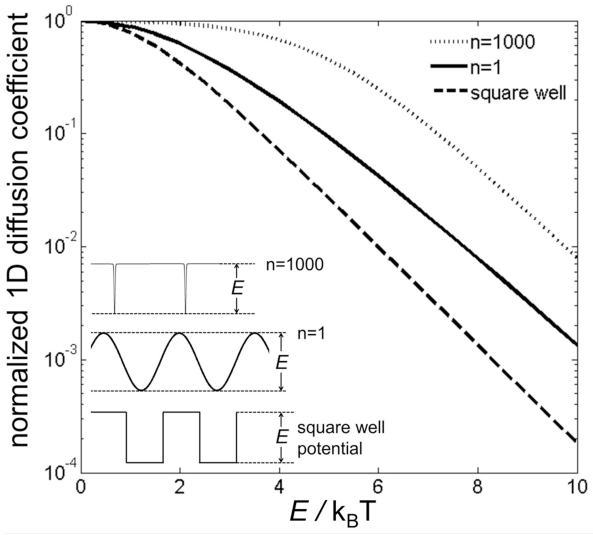

Figure 2.2.5: D vs. periodic potential well depth. The *dashed line* is for the square-well potential, and the *solid line* is for the continuous sinusoidal potential, Eq. 2.2.16 with n = 1, while the *dotted line* is the sinusoidal potential with n =1000

Two-state model

To investigate the diffusion of proteins that might use two domains to bind DNA or switch

between two conformations, we propose a two-state model similar to that suggested by Slutsky and Mirny, in which a protein can switch between a "searching" (S) state and "recognition" (R) state on any site. However, we do not allow dissociation of the protein from the DNA. For both states, we assume the site energies to be Gaussian distributed:

$$E_i^S = N\left(\mu^S, \sigma^S\right), \quad E_i^R = N\left(\mu^R, \sigma^R\right) \tag{2.2.17}$$

We assume the protein can slide along the DNA in both S and R states (as p53 does) with the S state allowing for much faster diffusion than the R state does ($D^S \gg D^R$, $\sigma^S \ll \sigma^R$). Both the sliding along DNA and the conformational switching are controlled by activation barriers. The diffusive properties of the two-state model can be characterized by a sliding transition rate in the S state $\omega_{i,i\pm1}^S$, a sliding transition rate in the R state $\omega_{i,i\pm1}^R$, a transition rate from the S to the R state ω_i^{SR}, and a transition rate from the R to the S state ω_i^{RS}. These rates are given by:

$$\begin{aligned}
\omega_{i,i\pm1}^S &= \omega_0 exp\left[-\left(\widetilde{E}_a^S - \widetilde{E}_i^S\right)\right] \\
\omega_{i,i\pm1}^R &= \omega_0 exp\left[-\left(\widetilde{E}_a^R - \widetilde{E}_i^R\right)\right] \\
\omega_i^{SR} &= \omega_0 exp\left[-\left(\widetilde{E}_a^C - \widetilde{E}_i^S\right)\right] \\
\omega_i^{RS} &= \omega_0 exp\left[-\left(\widetilde{E}_a^C - \widetilde{E}_i^R\right)\right]
\end{aligned} \tag{2.2.18}$$

E_a^S,E_a^R,and E_a^C are activation barriers (normalized by k_BT) for diffusing from one site to the next in the "searching" state, for diffusing in the "recognition" state, and for conformational transitioning from the search to the recognition state, respectively. $\widetilde{E}_i^S$ $\left(\widetilde{E}_i^S \equiv E_i^S/k_BT\right)$ and $\widetilde{E}_i^R$ $\left(\widetilde{E}_i^R \equiv E_i^R/k_BT\right)$ are Gaussian-distributed site-specific (normalized) energies in the "searching" state and in "recognition" state, respectively, where, without loss of generality, the S state can be taken to have zero average energy $\widetilde{\mu}^S = 0$. This is equivalent to

interpreting $\widetilde{\mu}^R$ as the energy difference between the recognition state and the searching state. We use the same prefactors for all the rates above, but note that we can absorb differences in prefactors into the apparent activation energies, since we are not considering effects of varying temperature here. Note that the ratios of forward to reverse rates of transition between any two microstates (i to $i \pm 1$ in either S or R configuration and S to R or R to S at any site i) are given by Boltzmann functions of the energy difference between the two states. Thus, detailed balance is satisfied and at a long time the protein will sample an equilibrium distribution of states. Since the site-specific free energies $\widetilde{E}_i^S$ and $\widetilde{E}_i^R$ are each Gaussian distributed without threshold about a mean energy $\widetilde{\mu}^S$ and $\widetilde{\mu}^R$, $\widetilde{E}_i^S$ and $\widetilde{E}_i^R$ can be higher than the activation barriers. In that case, the transition rate of translocation from site i to site $i \pm 1$ or the transition rate from the R state to the S state at site i will be overestimated. However, if the activation barrier is large compared to the standard deviations, such events should be very rare, and have little effect on the diffusion coefficient.

To obtain the relation between D and the roughness of the energy landscape in the R state, a kinetic Monte Carlo simulation was performed using the Bortz-Kalos-Lebowitz (BKL) algorithm [74]. The diffusion coefficient was obtained by taking the slope of mean square displacement vs. time.

Since the 1D diffusion coefficient for the R state is much lower than that for the S state, in other words, the S state energy roughness is much less than the R state energy roughness $\widetilde{\sigma}^S \ll \widetilde{\sigma}^R$, we need the protein to prefer the S state if we wish a fast search. Therefore, since we are taking the mean energy of the search state to be zero, the mean energy of the recognition state (relative to the searching state) $\widetilde{\mu}^R$ should be positive, especially for large $\widetilde{\sigma}^R$. So, although in one-state Models I and II, we chose a zero mean energy for the

Gaussian-distributed energy, for this model we vary $\widetilde{\mu}^R$ from 0 to 4. Although the protein thermodynamically prefers the S state away from the target site, the two-state model assumes that at the target site the free energy in the recognition state is at its minimum, and at the target the free energy is lower in the recognition state than in the search state. The conformation transition activation barrier $\widetilde{E}_a^C$ determines the frequency of state switching. Obviously, a small barrier will lead to a fast switching between these two states and thereby facilitate the sliding motion of the protein along DNA. However, a small $\widetilde{E}_a^C$ will also result in a small average residence time at the target site τ_T, which will allow fast reaction only when the reaction rate constant Ω is much larger than $\omega_{T,T\pm1}^R + \omega_T^{RS}$, where T is the target site number. Otherwise, the fast switching between R and S states will block any speed up of the targeting process, since the target site, once found, will be abandoned before a reaction can occur. Similar to the one-state analysis, as shown in the Appendix A, the reaction rate $r(x=0)$ is given by:

$$\frac{c_\infty^S}{r} = \frac{1}{2\sqrt{k_{off}D}} + \frac{1}{\omega_T^{SR} exp\left(-\widetilde{E}_T^S\right)} + \frac{1}{\Omega exp\left(-\widetilde{E}_T^R\right)} \tag{2.2.19}$$

where $\widetilde{E}_T^S$, $\widetilde{E}_T^R$, $\widetilde{c}_\infty^S$ are target energy in "searching state", target energy in "recognition state", and far field protein concentration in "searching state", respectively. From Eq. 2.2.19, when $\Omega \ll \omega_T^{SR} exp\left(\widetilde{E}_T^R - \widetilde{E}_T^S\right) = \omega_T^{RS}$ and $\Omega \ll \sqrt{k_{off}D} exp\left(\widetilde{E}_T^R\right)$, the protein targeting is *reaction limited*, and the reaction rate is:

$$r = c_\infty^S exp\left(-\widetilde{E}_T^R\right)\Omega \tag{2.2.20}$$

Note that at the target site the free energy in the recognition state is low, while the free

energy of the search state is relatively insensitive to protein location along the DNA. Thus, although away from the target site the searching-state free energy is generally lower than the recognition state energy, at the target site, we can assume that $\widetilde{E}_T^S - \widetilde{E}_T^R > 0$, and the reaction rate is *sped up* a factor of $exp\left(-\widetilde{E}_T^R\right)$, relative to the reaction rate for a flat free energy landscape in the recognition site. The reason is that in the reaction-limited regime, all protein states are sampled in proportion to their equilibrium distribution before the protein is able to finally react. Since the protein in the recognition state at the target site has a low free energy, it is thermodynamically favored; that is the protein is *stable* at the target site and this increases the net reaction rate. To account for the concentration of protein at the target site in the recognition state, relative to the concentration away from the target, we need an additional factor, analogous to that given by Eq. 2.2.6 for the one-state model. Since there are both recognition and search states, the concentration of protein in the recognition state at the target is:

$$\begin{aligned} c_T^R &= \frac{exp\left(-\widetilde{E}_T^R\right)}{\frac{1}{n}\sum_i^n exp\left(-\widetilde{E}_i^S\right) + \frac{1}{n}\sum_i^n exp\left(-\widetilde{E}_i^R\right)} c_\infty \\ &= \frac{exp\left(-\widetilde{E}_T^R\right)}{exp\left[\frac{1}{2}\left(\widetilde{\sigma}^S\right)^2\right] + exp\left[-\widetilde{\mu}^R + \frac{1}{2}\left(\widetilde{\sigma}^R\right)^2\right]} c_\infty \\ &\approx \frac{exp\left(-\widetilde{E}_T^R\right)}{1 + exp\left[-\widetilde{\mu}^R + \frac{1}{2}\left(\widetilde{\sigma}^R\right)^2\right]} c_\infty \end{aligned} \tag{2.2.21}$$

where we have again used analytical results for the expected values of the log-normal distributions given by the above averages of exponentials of $\widetilde{E}_i^R$ and $\widetilde{E}_i^S$, which are Gaussian distributed quantities. Also, in the above, we remember that we have as a reference energy taken the average energy in the searching state μ^S to be zero, and since we assume that

the energy roughness in the searching state is low, we have neglected the factor involving σ^S in the final equation. We have an overall reaction rate of

$$r = c_\infty \frac{exp\left(-\widetilde{E}_T^R\right)\Omega}{1 + exp\left[-\widetilde{\mu}^R + \frac{1}{2}\left(\widetilde{\sigma}^R\right)^2\right]} \tag{2.2.22}$$

Note c_∞ is the far-field protein concentration in both "searching" and "recognition" states. This result shows that *there is no enhancement in reaction rate to be gained by having a low value of* E_T^S, i.e., having a local minimum in free energy at the target site in the searching state. Such a local minimum would enhance the concentration of protein there, but would correspondingly diminish the ratio of concentration in the recognition state at the target, relative to that in the searching state, and so produce no gain in reaction rate in the reaction limited regime. The reason is that in the reaction-limited regime, only the equilibrium concentration at the target in the recognition state matters. In their two-state model, Slutsky and Mirny [65] suggested that a low free energy in the searching state at the target might enhance the overall reaction rate, but the argument just given indicates that this will not be effective in the reaction-limited regime. In the diffusion-limited regime, it will of course also be ineffective, since in that case the far-field diffusivity controls the reaction rate, and not the concentration at the target, in either the recognition or searching state.

Returning to Eq. 2.2.19, when $\omega_T^{RS} \ll \Omega$ and $\omega_T^{SR} \ll \sqrt{k_{off}D}exp\left(\widetilde{E}_T^S\right)$, from Eq. 2.2.19, the targeting process is *conformational-switch limited*, and the reaction rate is:

$$r = c_\infty^S exp\left(-\widetilde{E}_T^S\right)\omega_T^{SR} \tag{2.2.23}$$

This regime is entered only if $\omega_T^{RS}/\Omega \ll 1$, but the preference of the protein for the recog-

nition state to the searching state at the target requires that $\omega_T^{RS} < \omega_T^{SR}$, although away from the target site, as mentioned above, the protein prefers the search state. Thus, it is possible that this regime could lead to a speed-up of targeting, if $\omega_T^{SR} > \Omega \gg \omega_T^{RS}$. In this case, the protein is trapped at the target site long enough to react there, even if the rate of reaction is not fast. Hence, the overall rate of reaction might be enhanced.

Finally, when $\sqrt{k_{off}D} \ll \omega_T^{SR} exp\left(-\widetilde{E}_T^S\right)$ and $\sqrt{k_{off}D} \ll \Omega exp\left(-\widetilde{E}_T^R\right)$, the targeting process is *diffusion limited*,and the reaction rate is:

$$r = \frac{2c_\infty^S D}{\sqrt{D/k_{off}}} \tag{2.2.24}$$

This formula is the same as for the one-state model in the diffusion-limited regime, namely Eq. 2.2.5. However, the diffusion coefficient D in the above is the average diffusion coefficient away from the binding site, and this diffusion coefficient is dominated by diffusion in the search state, which is fast, since we have assumed that away from the target site, the free energy of the search state is lower than that of the recognition state; i.e., $\widetilde{E}_i^S - \widetilde{E}_i^R < 0 (i \neq T)$.

Therefore, the apparent 1D diffusion coefficient in the two-state model can be approximated by:

$$D \approx D^S f^S \approx D^S \left(1 - f^R\right) \tag{2.2.25}$$

where D^S is the one-dimensional diffusion coefficient of the protein in "searching state", and f^S and f^R are the fraction of time or the probability that protein stays in "searching" and "recognition" states, respectively.

At equilibrium, the probability of residing in the search state can be computed by $f^S =$

$\frac{\sum_{i=1}^{n} exp\left(-\widetilde{E}_i^S\right)}{\sum_{i=1}^{n} exp\left(-\widetilde{E}_i^S\right)+\sum_{i=1}^{n} exp\left(-\widetilde{E}_i^R\right)} = \frac{\left\langle exp\left(-\widetilde{E}_i^S\right)\right\rangle}{\left\langle exp\left(-\widetilde{E}_i^S\right)\right\rangle+\left\langle exp\left(-\widetilde{E}_i^R\right)\right\rangle} = \frac{exp\left[-\widetilde{\mu}^S+\frac{1}{2}\left(\widetilde{\sigma}^S\right)^2\right]}{exp\left[-\widetilde{\mu}^S+\frac{1}{2}(\widetilde{\sigma}^S)^2\right]+exp\left[-\widetilde{\mu}^R+\frac{1}{2}(\widetilde{\sigma}^R)^2\right]}$, where n is the number of sites, and both $\left\langle exp\left(-\widetilde{E}_i^S\right)\right\rangle$ and $\left\langle exp\left(-\widetilde{E}_i^R\right)\right\rangle$ are expected values of log-normal distributions. By taking $\widetilde{\mu}^S = 0$, $\widetilde{\sigma}^S = 0$ (very small roughness of S state), we have $f = \frac{1}{1+exp\left[-\widetilde{\mu}^R+\frac{1}{2}(\widetilde{\sigma}^R)^2\right]}$. Assuming the 1D diffusion in the R state contributes little to the overall 1D diffusion, the 1D diffusion coefficient can be approximated by:

$$D \approx D^S \cdot f^S = \frac{D^S}{1 + exp\left[-\widetilde{\mu}^R + \frac{1}{2}\left(\widetilde{\sigma}^R\right)^2\right]} \tag{2.2.26}$$

When the activation energy for transitioning from the search to the recognition state is much lower than that for diffusing in the search state, $E_a^C \ll E_a^S$, the protein can switch its conformation frequently before translocating to the next base pair. So, the protein will usually go directly to the recognition state whenever the recognition state energy E_i^R is lower that the searching state energy E_i^S. Then introducing the searching state will have a similar effect as introducing a low threshold to the Gaussian random potential, low enough that the protein usually takes advantage of it to diffuse from one site to another. This can also be seen by examining the free-energy landscape at the bottom of Figure 2.2.2 for the two-state model, where the searching state free energy acts like a low-energy cut-off for the free-energy landscape of a one-state model. Thus, in this limit, the diffusion is along a flat landscape with occasional deep traps, which in the two-state model are the few locations where the recognition state has lower free energy than the searching state. Note that when $\sigma(\sigma^R)$ becomes large, Eq. 2.2.11 reduces to Eq. 2.2.13, which gives a value twice as high as given by Eq. 2.2.26, where the constant ν in Eq. 2.2.11 now is the threshold energy or average search state energy relative to the average energy of the recognition state, or $\nu = \mu^S - \mu^R = -\mu^R$.

The denominator in Eq. 2.2.26 is the same factor that slows down targeting in the reaction-limited regime discussed above; see Eq. 2.2.22. The relation between D and the roughness of the free-energy landscape in the recognition state as well as the relation between f^R and the free-energy roughness in the recognition state can also be obtained by the BKL algorithm and are compared with the theoretical result from Eq. 2.2.26 in Fig. 2.2.6a, where the average free energy in the "recognition" state is taken to be $\mu^R = 2k_BT$.

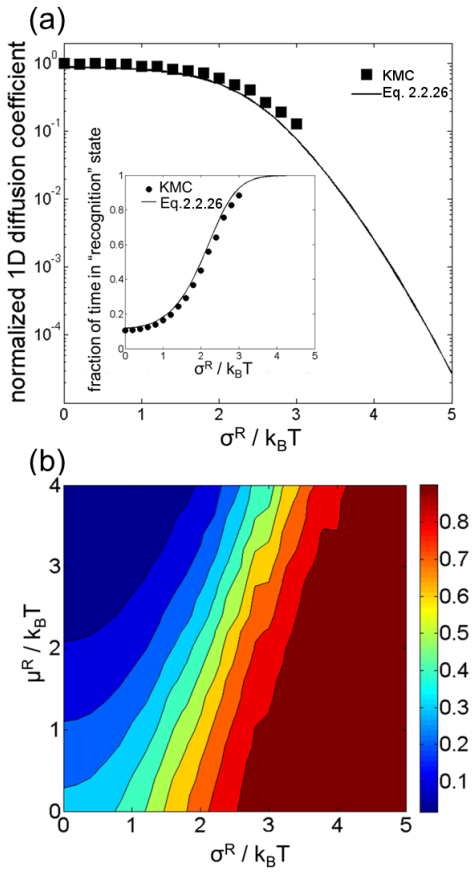

Figure 2.2.6: (a) 1D diffusion coefficient and fraction of time that protein is in the "recognition" state vs. energy landscape roughness in the "recognition" state based on Eq. 2.2.26 *lines* and KMC simulation results (*squares* and *circles*). (b) Colors indicate the fraction of time that protein is in the "recognition" state as a function of both energy roughness and average energy of the "recognition" state

The dependence of the fraction of time in the recognition state f^R on average energy μ^R and roughness of recognition state is plotted in Fig. 2.2.6b. This fraction is higher for larger roughness or lower average recognition state energy. As a result, to speed up the protein sliding, we need the protein to spend a high fraction of time in the searching state and a low fraction of time in the recognition state, and so the average energy of the recognition state should be high relative to the searching state.

Both Eq. 2.2.26 and KMC simulations demonstrate that this two-state model can facilitate protein targeting process in the *diffusion limited* regime, since it can increase the overall 1D diffusion coefficient dramatically, especially when the average energy of the recognition state is high, which results in a faster reaction rate (Eq. 2.2.24). However, as μ^R increases, the frequency of protein conformational switch from S to R state will decrease. In other words, for high average energy of the recognition state, the conformational switch will become the speed-limiting step.

2.3 Applications and discussion

For most DNA-binding proteins (DBPs), the sequence-dependent free energy (E_i) and the roughness of the 1D disordered free-energy landscape along DNA are inaccessible. However, it is noteworthy that the one-dimensional diffusion coefficients of proteins with long average sliding length ($\langle L_{sliding} \rangle$), such as T7 RNA polymerase (T7 RNAP,$\langle L_{sliding} \rangle \gg$ 1,000 bp) and hOgg1 ($\langle L_{sliding} \rangle$ ~200 bp) are of the order $10^{-13} m^2/s$, which is close to Schurr's [75] estimate $D \approx 5 \times 10^{-13} m^2/s$ of the effective 1D diffusivity of a protein following the helical path marked by the major groove of the DNA, accounting only for viscous drag of moving along this path and ignoring the random potential. Thus, for some

proteins, there may be little effect of the potential on the diffusivity. Either the roughness of the potential is very small ($\sigma \leq 0.25 k_B T$, based on Eq. 2.1.1); or there is a mechanism to reduce the effective roughness of the potential, such as an energy threshold or the searching-state mechanism discussed above. Note that for a flat potential with very small roughness ($\sigma \leq 0.25 k_B T$), the concentration or the occupancy probability of the protein along the DNA is roughly sequence-independent. In other words, the enhancement of the effective reaction rate r resulted from accumulation of DBP's on the target site will be very limited. However, by introducing the threshold, even for a rough landscape the protein can maintain its relatively high probability at the target site without slowing down its diffusion significantly.

To illustrate how our model can be used to determine 1D diffusion coefficients and targeting rates, we consider the free-energy landscape model of Barbi [67, 68]. This model uses the number of hydrogen bonds (HB) formed at each position of the protein along the DNA, to set the depth of the site-specific potential well. Thus, in Barbi's model, the well depth for a given site is simply estimated by:

$$E_i = -N_i \varepsilon \tag{2.3.1}$$

where N_i is the number of the hydrogen bonds and ε is the strength of the hydrogen bonds. We can then obtain the 1D potential landscape of the T7 RNAP-DNA system by plugging Eq. 2.3.1 into Eq. 2.2.16. Since the actual well shape of the 1D potential along DNA is unknown, we set n in Eq. 2.2.16 to 1, which guarantees that the protein feels a smooth continuous force along this potential landscape for simplicity. This method assumes that both electrostatic and hydrophobic forces have a negligible effect on the roughness. This

is a crude assumption, but allows us to illustrate how to compute long-range 1D diffusivity from any model that gives a free-energy landscape for DNA-binding proteins.

Using the base sequence for T7 DNA, we thus obtain the number of hydrogen bonds (HB) for each site, where a "site" consists of the 5-bp DNA sequence that T7 RNAP binds to at any one instant when moving along the DNA. The number of hydrogen bonds for each site was obtained for a sequence of 3,000 base pairs (T7 genome, bases 17,001 to 20,000 [76]) using Barbi's method. Barbi's method provides a 5×4 matrix that allows the total number of hydrogen bonds to be computed for any ordered sequence of five bases. The T7 RNAP was modeled by an over-damped Brownian particle 5 nm in diameter. To obtain the 1D diffusivity of the protein, we perform a Brownian dynamics simulation using 30,000 non-interacting particles that are initially evenly distributed over the 1,000 base pairs in the center of the sequence of 3,000 base pairs. In Barbi's model, the number of hydrogen bonds N_i for each 5-bp site can be well fitted by a Gaussian distribution with standard deviation $\sigma_{HB} \approx 2$. (Since this can yield negative values for N_i, in our simulations, any N_i below 0 is set to 0.) Here we vary the HB strength ε up to $1.5k_BT$,a range that includes the strength typical of a water-mediated hydrogen bond, i.e., $0.8k_BT$ [77]. The standard deviation of the distribution of energies, E_i , is then found to be $\sigma = \sigma_{HB}\varepsilon \approx 2\varepsilon$, which varies from 0 to 3 k_BT. The values of D for different potential roughness were obtained by taking the slopes of plots of mean square displacement versus time from the simulation and averaging over all 30,000 particles. For comparison, the diffusion coefficients over this potential were also obtained numerically using the MFPT, computed from Eq. 2.2.7.

The resulting relation between D and roughness is plotted in Fig. 2.3.1a. For roughness as high as $\sigma = 2k_BT$, the effect of this random potential on one dimensional diffusion coefficient is less than a factor of 10, which is roughly consistent with the measurement of

D of the single molecule experiments on T7 RNAP or hOgg1. If the potential well is made narrower by increasing n in Eq. 2.2.16, the disordered potential will have even less effect on D.

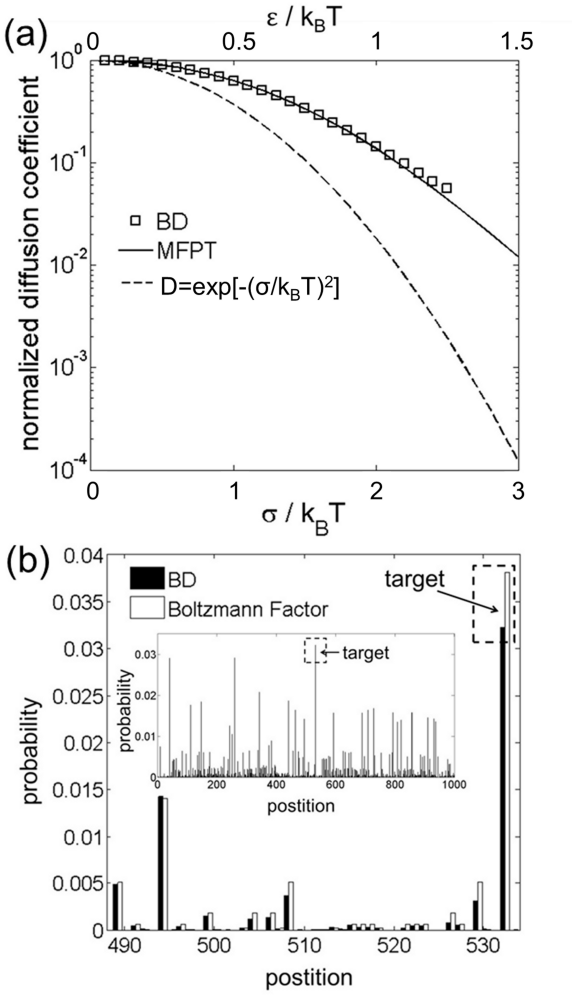

Figure 2.3.1: (a) Normalized diffusivity D/D_0 vs. roughness σ of the continuous random potential Eq. 2.2.16 (Model III). *Squares* are obtained by Brownian dynamics (BD) simulations, and the *line* is from the MFPT. (b) Histogram of particle distribution over 50 potential wells, where the black bars are obtained by BD simulations, and white bars from Boltzmann weights with $\sigma = 2.4k_BT$. The inset is the histogram of particle distribution over the central 1,000 base pairs from a set of 3,000 base pairs for seven obtained by BD simulation over with $\sigma = 2.4k_BT$

The final locations after 10^7 time steps simulation of the 30,000 non-interacting particles were used to generate the equilibrium number density distribution over the potential well, which is shown in Fig. 2.3.1b, and found to match that from the Boltzmann distribution, which indicates that equilibrium was achieved in this run. As a result, our proposed continuous random potential indicates that the sliding of T7 RNAP can be rapid without losing searching fidelity.

To test and improve this model, long DNA with repeated base pair sequences (e.g., T7 RNAP promoter sequences) could be used to compute the 1D diffusion coefficients of T7 RNAP on different periodic potential landscapes, and the results compared to experimental measurements of the 1D diffusion coefficient.

2.4 Conclusions

Using several models of protein/DNA diffusion, and multiple methods of solving those models, we have explored conditions under which proteins might find and react quickly with their target sites on DNA. A key consideration is whether the targeting is reaction limited or diffusion limited. In the former case, thermodynamic considerations dominate, and the stability or residence time of the target binding is paramount. The targeting rate is given by the reaction rate constant at the target site times the concentration or probability of the protein to be at this site and a deep potential well at the target site is needed so that the protein has a high thermodynamic probability to be at the target site long enough to react. If this is achieved by creating a very rough free energy landscape, then the 1D diffusivity can be greatly slowed, possibly pushing the process towards the diffusion-limited regime. The targeting in the diffusion-limited regime is set by the concentration or proba-

bility of the protein far from the reaction site, times the average diffusivity of the protein, divided by the diffusion distance, which is the square root of the ratio of diffusivity to the dissociation constant. We showed that for diffusion distances of order 100 basepairs, using a diffusion coefficient averaged over the whole landscape gives good accuracy for the targeting rate, unless the landscape roughness exceeds around 3 k_BT. Fast targeting in the diffusion-limited regime is enhanced by a high diffusivity, which requires an *effectively* shallow free energy landscape. This can be achieved even if the bare landscape is quite rough, in a number of ways explored here. Firstly, the shape of the landscape matters, and if barriers are narrow and sharply peaked, then they are effectively less rough. Secondly, if proteins can "hop"-i.e., briefly escape into solution and rapidly re-bind at a nearby site, then diffusion is much faster. Finally, if the protein can change conformation and undergo rapid diffusion in a conformation in which it does not recognize the DNA, then if the switching between searching and recognition states is fast compared to diffusion, the search efficiency is improved. Both "hopping" and introduction of a "searching" mode can under such circumstances be modeled simply by putting a maximum threshold into an otherwise Gaussian random landscape. We have shown that multiple theoretical methods, including first passage time methods, and the Festa-d' Agliano formula, allow us to compute analytically the diffusivity of such models. We have confirmed our analytical expressions using Brownian dynamics and Kinetic Monte Carlo methods. Further progress in understanding this complex problem will probably require more detailed understanding of the actual free energy landscape experienced by real proteins. However, the theoretical results presented here should help guide and interpret future experimental work.

CHAPTER III

DNA condensation by partially acetylated PAMAM dendrimers: effects of dendrimer charge density on complex formation

The ability of poly(amido amine) (or PAMAM) dendrimers to condense semiflexible dsDNA and penetrate cell membranes gives them great potential in gene therapy and drug delivery but their high positive surface charge makes them cytotoxic. Here, we describe the effects of partial neutralization by acetylation on DNA condensation using light scattering, circular dichroism, and single molecule imaging of dendrimer-DNA complexes combed onto surfaces and tethered to those surfaces under flow. We find that DNA can be condensed by generation-five (G5) dendrimers even when the surface charges are more than 65% neutralized, but that such dendrimers bind negligibly when an end-tethered DNA is stretched in flow. We also find that when fully charged dendrimers are introduced by flow to end-tethered DNA, all DNA molecules become equally highly coated with dendrimers at a rate that becomes very fast at high dendrimer concentration, and that dendrimers

remain bound during subsequent flow of dendrimer-free buffer. These results suggest that the presence of dendrimer-free DNA coexisting with dendrimer-bound DNA after bulk mixing of the two in solution may result from diffusion-limited irreversible dendrimer-DNA binding, rather than, or in addition to, the previously proposed cooperative binding mechanism of dendrimers to DNA.

3.1 Introduction

Macroions such as cationic dendrimers, polylysine, poly(ethylene imine), and surfactants can be used to condense semiflexible DNA, and so serve as potential substitutes for viral vectors in gene delivery [78–82]. A large number of studies on DNA condensation by poly(amidoamine) or PAMAM dendrimers have been performed because of its great potential as a gene carrier. The interaction between PAMAM dendrimer and DNA is mainly electrostatic, as is the case with histones and DNA, and dendrimers of generation 5 or 6 have diameters roughly comparable to those of histones. Therefore, PAMAM dendrimers also provide an interesting model to study some facets of chromatin formation. To determine how the interaction between dendrimers and DNA affects complex formation, the dendrimer size or generation [83], salt concentration [84], pH [85], and surface groups [86] have been varied, and the morphologies and local structures of the resulting DNA-dendrimer complexes have been resolved using cryo-TEM [87] and small angle X-ray scattering [85]. The binding between fully charged dendrimers and DNA has been shown to be irreversible using optical tweezers to pull on dendrimer-bound DNA without dislodging the bound dendrimers [13].

Although PAMAM dendrimers can bend and condense semiflexible DNA, fully charged den-

drimers are toxic to living cells because of their highly charged surfaces. Thus, reducing dendrimer surface charge density by PEGylation [88] or acetylation can be very important for the application of dendrimers *in vivo*. Besides, varying the surface charge density of nanoparticles might be used to improve understanding of chromatin formation, since dendrimers with the same size and charge density as histones can be used to compact DNA and explore whether some aspects of chromatin formation can thereby be induced. Various models of polyelectrolyte condensation by macroions have [89, 90] been proposed, which can also be validated if the surface charge density of the dendrimer can be accurately controlled. In addition, weakly charged nanoparticles may melt rather than bend dsDNA [91], which is also of great interest.

In this work, we systematically investigate the interactions between DNA and dendrimers acetylated to various extents using dynamic light scattering, fluorescence spectroscopy, circular dichroism, and fluorescence microscopy imaging. The aim is to help find the appropriate charge density of PAMAM dendrimer for condensing DNA with less cytotoxicity and to help test and clarify the proposed cooperative binding mechanism [78, 92] of the DNA-dendrimer interaction.

3.2 Results and discussion

3.2.1 Dynamic Light Scattering

DNA and PAMAM dendrimers (generation 5) were acetylated to various extents, mixed in 10 mM NaBr solutions, and studied by dynamic light scattering. According to our titration experiments (data not shown here), the number of charges on non-acetylated G5 den-

drimers is 114, instead of 128, the number of positive charges on a perfectly synthesized dendrimer [93]. Therefore, assuming that defects that lead to the reduced charge occur randomly, the standard deviation of number of charges on a non-acetylated dendrimer can be approximated by:$Var[N_k] = 1 - N_k \left(\frac{N_k}{N}\right)$ where N is the maximum number of charges possible, which equals 128, and N_k is the measured mean value of 114. If the acetylation of dendrimer terminal groups is also taken to be random, the standard deviation of the number of charges on an individual acetylated dendrimer can also be approximated by the equation above. Thus, the numbers of positive charges, and their standard deviations computed by propagation of errors using both sources of variation mentioned above, on dendrimers acetylated to 0%, 15%, 30%, 50%, 65%, and 85%, are 114 ± 3.5, 97 ± 5.2, 80 ± 6.0, 57 ± 6.5, 40 ± 6.2, 17 ± 4.0, respectively. The relaxation time distributions of pure dsDNA as well as DNA-dendrimer complexes were obtained from autocorrelation functions of scattering light intensities using analysis with CONTIN 2DP. Selected relaxation time distributions of DNA-dendrimer complexes are shown in Figure 3.2.1 with scattering angle θ fixed at 50°. The apparent hydrodynamic radii of the DNA-dendrimer complexes were computed using the peaks of the relaxation time distributions along with the Einstein-Stokes equation, and plotted in Figure 3.2.2. Note that r_{charge} here and in the following discussion is defined as the ratio of the total number of positive charges on all dendrimers to the total number of negative charges on all DNA molecules in the solution, that is, $r_{charge} = [NH_3^+]/[PO_4^-]$. All the primary amine groups (pKa = 9.0 [94] $\sim$ 10.77 [95]) were assumed to be protonated in pH 7$\sim$ 8 in this study. This assumption was supported by titration experiments and the observation that solutions become cloudy when r_{charge} defined above is close to unity. In this thesis, we limit ourselves to the cases with r_{charge} less than 1 to study DNA-dendrimer complexes before phase separation takes place.

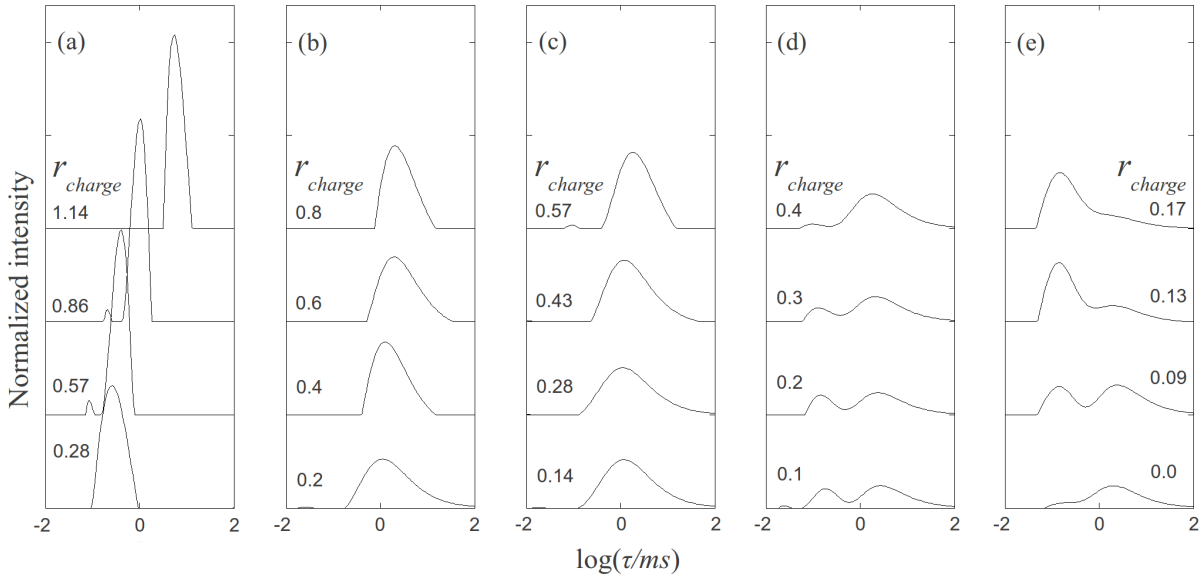

Figure 3.2.1: Relaxation time distributions of PAMAM dendrimer/DNA samples measured at scattering angle $\theta = 50°$. (a) non-acetylated G5/DNA. (b) 30% acetylated G5/DNA. (c) 50% acetylated G5/DNA. (d) 65% acetylated G5/DNA. (e) 85% acetylated G5/DNA

The relaxation time distribution of 0.15 mg/mL pure salmon sperm DNA (2000 $\pm$ 500 bp) in 10 mM NaBr solution is plotted in Figure 3.2.1e, where r_{charge} is 0. The two peaks of the DNA relaxation time distribution correspond to the internal (shorter τ) and translational (longer τ) modes of DNA. The apparent hydrodynamic radius of DNA computed based on the translational peak is 110 nm, in agreement with previous studies [78, 83]. DNA concentrations for all dynamic light scattering measurements were fixed at 0.15 mg/mL, while the dendrimer concentration was varied to obtain the specified r_{charge}. The apparent hydrodynamic radius of the G5 PAMAM dendrimer was determined to be around 3 nm by DLS, indicating that aggregation of dendrimer in the absence of DNA was negligible.

As shown in Figure 3.2.1a, non-acetylated G5 dendrimer-DNA complexes give sharp peaks, indicating that the complexes size distribution is relatively narrow. Comparing the positions of these peaks to those of the pure DNA, we can conclude that the former, which

correspond to faster relaxation, and therefore smaller objects, represent condensed DNA-dendrimer complexes. It should be noted that values of r_{charge} of 0.28, 0.57, 0.86, and 1.14 correspond to r_{molar} values of 10, 20, 30, and 40, where r_{molar} is defined by [dendrimer]/[DNA]. The decay time distributions of the acetylated denrimer-DNA complexes with the same molar ratios (r_{molar}) are plotted in Figure 3.2.1b-e, except for the pure DNA curve at the bottom of Figure 3.2.1e. For the 30% acetylated dendrimer (Figure 3.2.1b) and 50% acetylated dendrimer (Figure 3.2.1c), the decay time distributions are much broader than for the non-acetylated dendrimer in Figure 3.2.1a, probably because the number of positive charges on the acetylated dendrimers varies somewhat from dendrimer to dendrimer, as estimated in the standard deviations in numbers of charges given above. For the dendrimers with the highest acetylation ratios (65%, 85%), the relaxation time distributions exhibit two peaks, the shorter of which is presumably contributed by internal motion within the DNA complex, while the slower mode corresponds to translational motion of the complexes. However, the slower mode for 85% acetylated dendrimer-DNA is almost the same as for pure DNA, indicating that the weakly charged dendrimer ($\sim$ 15 charges per dendrimer) was not able to condense the DNA significantly. This result agrees with a previous study of DNA/Poly-L-lysine [80], which showed that a transition of the DNA-polycation complex conformation occurs when the number of charges on the polycation is decreased. Highly charged polycations can bend and condense dsDNA while weakly charged polycation only attach to it without producing compaction.

The relaxation time distributions of acetylated dendrimer-DNA complexes with higher molar ratios (above 40) are not presented here, since for these cases a small fraction of the most highly charged dendrimers from the polydisperse charge distribution can condense the DNA even when the average charge per dendrimer is low and this small fraction dom-

inates the dynamic light scattering signals.

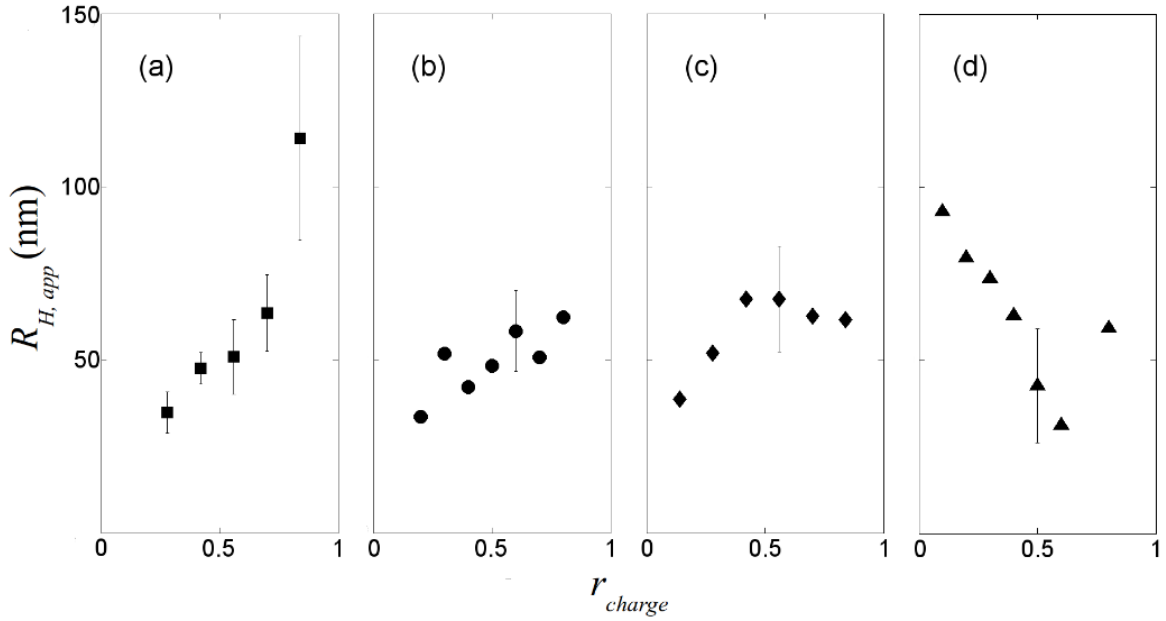

Figure 3.2.2: Apparent hydrodynamic radii of dendrimer/DNA complexes measured at $\theta = 50^{\circ}$. (a) non-acetylated G5/DNA. (b) 30% acetylated G5/DNA. (c) 50% acetylated G5/DNA. (d) 65% acetylated G5/DNA. Some typical error bars are given

The apparent hydrodynamic radii of the DNA-dendrimer complexes are plotted in Figure 3.2.2. For highly charged dendrimers (0%, 30%, 50% acetylated), the hydrodynamic radii of the complexes are around 50 nm. Therefore, this level of acetylation of primary amine groups on the PAMAM dendrimer does not significantly reduce its ability to condense DNA, despite the lessening of the dendrimer charge, which is needed for condensation [90]. Since reduction of the charge density on the dendrimer can reduce the cytotoxicity of dendrimer significantly [88], partial acetylation of the primary amine groups of PAMAM dendrimer might have significant potential in gene therapy. For highly charged dendrimers, as rcharge increases, the apparent hydrodynamic radius of the complexes also increases (Figure 3.2.2a), which is likely the result of formation of large complexes containing more than

one DNA chain. However, for weakly charged dendrimers (65% acetylated), the complex radius decreases when r_{charge} is increased up to 0.5, as the weakly charged dendrimer at low concentration binds to DNA without compaction, but compaction apparently increases with increasing numbers of bound dendrimers per DNA molecule. Based on our dynamic light scattering experiments, to condense DNA in the gene delivery process, PAMAM dendrimers with around half of their primary amine groups acetylated might be a good choice. To determine whether the polydispersity of charges on acetylated dendrimer will affect the gene delivery, transcription experiments should be performed.

3.2.2 Steady-State Fluorescence Spectroscopy

To determine the fraction of free DNA in the DNA-dendrimer mixture, steady-state fluorescence spectroscopy experiments were carried out 10 min after mixing DNA-dendrimer complex solutions with nucleic acid stain GelStar®. The excitation wavelength was fixed at 493 nm, and emission light intensity was recorded at 527 nm. The normalized emission light intensity of dendrimer-DNA complexes for various dendrimer concentrations are plotted in Figure 3.2.3.

GelStar® was assumed to bind only to free DNA, that is to portions of the DNA not blocked by dendrimer [78]. After reacting GelStar® with DNA, the emission light intensity increased dramatically although the emission from free GelStar® was negligible. Previous studies [78, 83] assumed that GelStar® is not able to react with dendrimer-bound DNA. This is validated by our non-acetylated G5-dendrimer results, for which the normalized emission light intensity is roughly equal to $(1 - r_{charge})$; see Figure 3.2.3. Therefore, GelStar® only binds to the free DNA segments which have not been neutralized by den-

drimer, since the interaction between DNA and GelStar® is also dominated by electrostatic force. However, the relation $I/I_0 = (1 - r_{charge})$ does not hold for DNA-acetylated dendrimers. This is apparently because the binding affinity of acetylated dendrimers to DNA decreases as the acetylation ratio increases. One implication is that the DNA segments condensed by dendrimer with fewer charges might still be accessible to proteins.

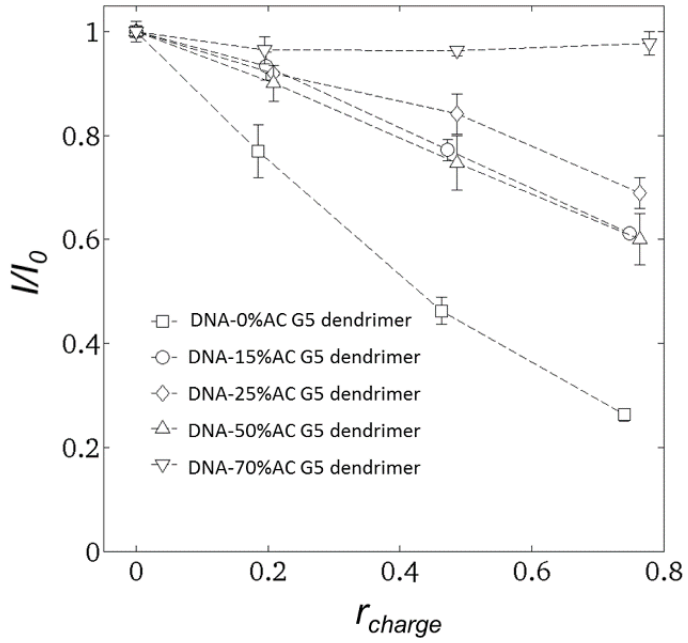

Figure 3.2.3: Emission light intensity (527 nm) versus rcharge for DNA condensed by dendrimers with various acetylation ratios. DNA-0% acetylated dendrimer (*squares*); DNA-15% acetylated dendrimer (*circles*); DNA-25% acetylated dendrimer (*diamonds*); DNA-50% acetylated dendrimer (*up triangles*); DNA-70% acetylated dendrimer (*down triangles*)

3.2.3 Circular Dichroism Spectroscopy

Using circular dichroism (CD) spectroscopy, we explore the DNA conformations within DNA-dendrimer complexes. The DNA concentrations were fixed, with r_{charge} also fixed to

0.2 for various dendrimers. The CD spectra for different complexes are shown in Figure 3.2.4. The negative peak (245 nm) and positive peak (275 nm) of free DNA are consistent with CD spectra of B-form dsDNA [96, 97]. CD spectra of DNA-dendrimer complexes exhibit the same shape as the free B-form dsDNA. The shift of peaks is negligible. Therefore, at low charge ratio (r_{charge}), the DNA in DNA-acetylated dendrimer complexes remains in the classical B-form. In other words, dendrimers and acetylated dendrimers bend DNA without disrupting the local helical structure of DNA.

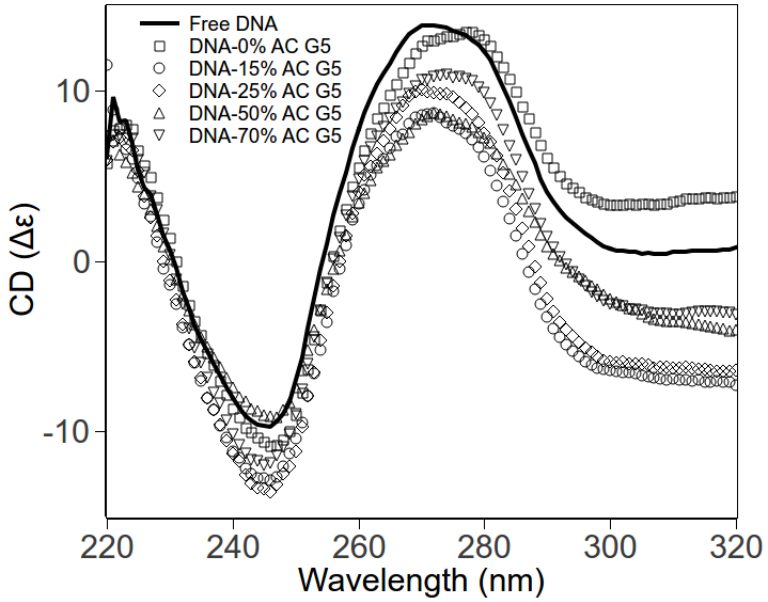

Figure 3.2.4: CD spectra of pure DNA and DNA-PAMAM dendrimer complexes in HEPES (pH 7.3). Free DNA (*solid line*); DNA-0% acetylated dendrimer (*squares*); DNA-15% acetylated dendrimer (*circles*); DNA-25% acetylated dendrimer (*diamonds*); DNA-50% acetylated dendrimer (*up triangles*); DNA-70% acetylated dendrimer (*down triangles*)

3.2.4 Molecular Combing Assay

Linearized λ-DNA has a hydrophobic 12-base overhang on each end, which allows it to stick and anchor at each end to a polystyrene-coated cover glass [15, 98]. When the PS coated cover glass is pulled out from DNA solution, the YOYO-1 stained free DNA molecules are aligned on the cover glass surface by the high air-water surface tension, and then are visualized by a fluorescence microscope with blue excitation (Figure 3.2.5b). We also fluorescently labeled dendrimers with amine-reactive tetramethylrhodamine isothiocyanate (TRITC) dye molecules. TRITC-labeled PAMAM dendrimer-DNA complexes were visualized by sequentially imaging DNA and dendrimer molecules using blue and green excitation, respectively. The overlays of DNA and dendrimer images are shown in Figure 3.2.5c. Since the λ-DNA molecules were condensed by dendrimer, even the high surface tension (up to $4.0 \times 10^{-10} N$) is not able to stretch the DNA out to its full contour length (16 μm). Thus, we are able to image the condensed form of λ-DNA directly. Here, we only present images of dendrimer-DNA complexes at low molecular ratio (100 dendrimers per DNA) where the dendrimer-DNA complexes are somewhat condensed but do not form dense globular particles, and a few complexes stick to the PS surface so that we can image them. For dendrimer-DNA complexes with higher molecular ratio (for example, around 1,000 dendrimer/DNA), the complexes do not stick to the surface, possibly because the two hydrophobic ends of the λ-DNA at these ratios are trapped within the denser complex particle, preventing the DNA from binding to the PS surface.

In the dendrimer-DNA complexes, as shown in Figure 3.2.5c, the dendrimer molecules concentrate along some regions of the DNA, while other regions remain dendrimer free. This is consistent with a previous study [78] in which dendrimer-bound DNA was shown to coexist with dendrimer-free DNA. These earlier results were interpreted as evidence

that the dendrimer molecules bind to the DNA in a “cooperative” manner, in which dendrimers have higher binding affinity to DNA with dendrimers already attached to it than to bare DNA. In what follows, we suggest that instead of, or along side of, cooperative binding, that irreversible diffusion-limited binding might help account for these results. In any event, the dendrimer-DNA complexes, once formed, are strong enough to resist unraveling during combing, implying a resistance to a force of around 500 pN per complex [98].

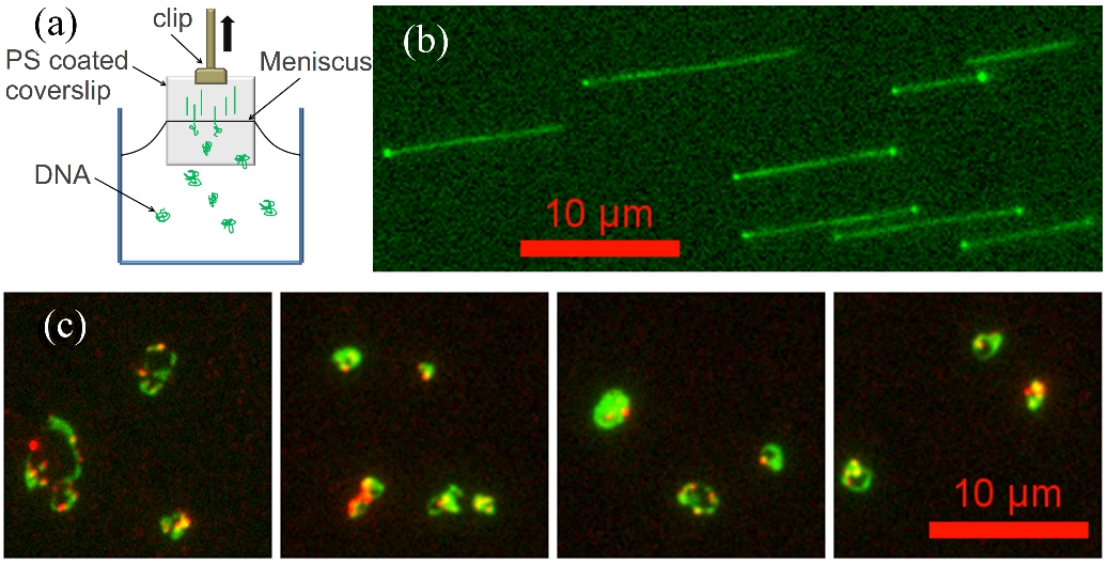

Figure 3.2.5: (a) Experimental setup for molecular combing. (b) Immobilized, aligned, YOYO-1 stained λ-DNA on PS surface. (c) λ-DNA/G5 PAMAM dendrimer (TRITC labeled) complexes deposited on PS coated cover glass surface (*green*: λ-DNA, *red*: PAMAM dendrimer)

We also incubated λ-DNA molecules aligned on the PS surface with a TRTIC-labeled G5 dendrimer solution, expecting dendrimers might slide one dimensionally along the DNA deposited on the surface, since a previous study [66] revealed that electrostatic forces were sufficient to confine the charged nanoparticle in one dimension. However, rather than experiencing 1D diffusion along DNA as some DNA-binding proteins do, G5 den-

drimers either stick to the DNA or the PS-coated surface or performed simple 3D Brownian diffusion (data not shown). The failure to undergo 1D diffusion along DNA might be due to the high binding energy [90] of the irreversible interaction of charged dendrimers with DNA [13].

3.2.5 Flow Stretching Assay

After biotinylated λ-DNA with one end attached to a neutravidin monolayer was deposited on the flow channel surface, the YOYO-labeled DNA was stretched by a shear flow. Based on parabolic flow assumption, at the flow rate imposed, the shear rate at the free end of the DNA, which based on the DNA coil size is estimated to be about 1 μm above the cover glass surface, is around 300 s^{-1}, which is able to stretch the DNA almost to its full contour length.

A TRITC-labeled G5 dendrimer (non-acetylated) solution was then introduced into the flow cell to interact with the tethered λ-DNA. The binding process was visualized and recorded by fluorescence microscopy using green excitation. We observed that the dendrimer bound and decorated the entire λ-DNA molecule, as opposed to the partial binding we observed in the molecular combing assay. When the λ-DNA is fully stretched out, each segment of the DNA apparently has almost the same probability of binding a dendrimer molecule. Apparently because the binding is irreversible and the well-mixed dendrimer solution is continuously injected into the flow channel, the dendrimers can cover the whole DNA chain evenly. After formation of the dendrimer-DNA complex, we used TE buffer to wash out the flow channel. No significant fluorescence decrease was observed on the tethered DNA until photocleavage or neutravidin detachment occurred, usually around 5 $\sim$ 10

min after illumination, which confirmed that dendrimer attached to DNA irreversibly.

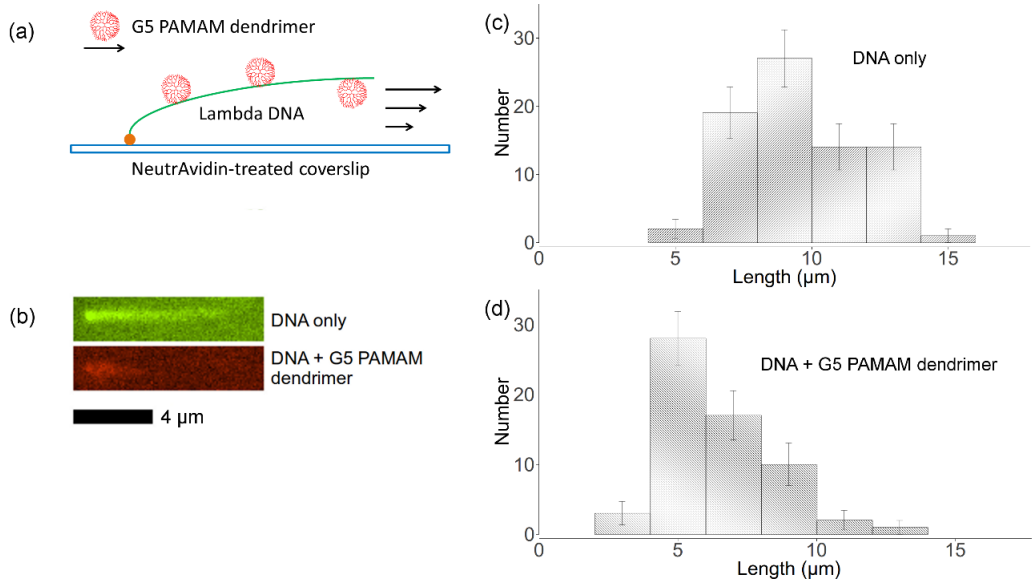

Figure 3.2.6: (a) Experimental setup for imaging dendrimer binding to flow-stretched DNA with one end tethered to the surface. (b) *Green*: tethered YOYO-labeled λ-DNA in flow; *red*: tethered TRITC-labeled dendrimer-DNA complex in flow. (c) Distribution of free λ-DNA molecules lengths. (d) Distribution of dendrimer-DNA complex lengths

Figure 3.2.6c,d show the length distributions of both free DNA and dendrimer-DNA complexes. 77 λ-DNA lengths as well as 62 dendrimer-DNA complexes lengths were recorded for these distributions. The error bar was approximated using equation:$Var[N_k]=N_k\left(1-\frac{N_k}{N}\right)$, where N_k is the number of data points of k_{th} bin and N is the total number of data points. The lengths distributions are wide, which is mainly due to the photocleavage of $lambda$-DNA, which shortens some of the DNA molecules before imaging is complete. However, these two histograms indicate that flow-stretched tethered DNA molecules are condensed by G5 dendrimer bound to the DNA chains, agreeing with optical tweezer experiments [13].

The interaction between flow-stretched λ-DNA and TRITC-labeled G5 PAMAM dendrimers

with 50% primary amine groups acetylated were studied using the same procedures. Because the neutravidin was physically attached to the cover glass, it could easily detach the surface under high drag force. Thus, we only imaged the DNA in flow for less than 30 min. Within this time, binding of 50% acetylated dendrimer to DNA was not observed. This indicates a significantly reduced binding affinity upon partial acetylation of the dendrimer, in agreement with our dynamic light scattering results as well as fluorescence spectroscopy results discussed above.

3.2.6 Cooperative Binding vs. Diffusion Limited Reaction

The differences in DNA-dendrimer complex conformations between the molecular combing assay and the flow stretching assay suggest an explanation to the observed apparent "cooperative binding" of dendrimers to DNA. "Cooperative binding" was proposed based on the coexistence of free DNA and compacted DNA-dendrimer complexes [78,92] formed in bulk solution. However, no coexistence of free DNA with complexes was observed in our flow-stretching assay. It is possible that the "cooperative binding" is much stronger in coiled DNA than in stretched DNA, perhaps because of coupling between DNA bending and dendrimer binding. However, another possibility is that kinetics limitation plays a role. According to our imaging results, this irreversible DNA-dendrimer binding reaction can be very fast. For example, the tethered DNA was covered by dendrimer in just minutes when we used 1 nM dendrimer solution, while it is covered by dendrimer instantly within our temporal resolution, which is a few seconds when using a higher concentration of dendrimer (>20 nM). Thus, the formation of DNA-dendrimer complexes is likely diffusion limited. If so, when r_{charge} is less than unity, coexistence of free DNA and DNA-dendrimer

complexes in bulk experiments might be simply due to the presence of free DNA that had not yet encountered free dendrimer before the free dendrimer was exhausted from the solution. Distinguishing between cooperative and simply irreversible binding would require very slow and careful mixing of dendrimers with DNA, to ensure that dendrimers can access every segment of DNA with equal probability before irreversible reaction occurs.

3.3 Experimental

3.3.1 DNA Preparation

Salmon sperm DNA (2,000 bp, 10.0 mg/mL in TE buffer) was purchased from Invitrogen (Grand Island, NY, USA) and used without further purification. Before using it, the ratio of absorbance at 260 nm to that at 280 nm was checked by a GeneQuant spectrophotometer, and found to be above 1.8, which is required to ensure that protein contamination was negligible. Lambda phage DNA (48.5 kbp) was purchased from New England Biolabs (Ipswich, MA, USA), and heated at 65 °C for 10 min followed by quick cooling to restore the molecules to linear form. For lambda DNA staining, the intercalating dye YOYO-1 (Molecular Probes, Grand Island, NY, USA) was diluted to 100 nM in Tris-EDTA buffer (pH 8.0) and mixed with lambda DNA at a staining ratio of 1 dye per 20 bp for both combing and the flow-stretching assay. Stained λ-DNA (3 nM) was ligated to a biotinylated 12mer (5'-agg tcg ccg ccc-biotin, 30 nM, Operon) using T4 DNA ligase (New England Biolabs) for flow stretching.

3.3.2 PAMAM Dendrimer Preparation

Generation 5 PAMAM dendrimer was purchased from Dendritech and purified by dialysis against water as described elsewhere [93, 99]. The mixtures of acetic anhydride and PAMAM dendrimer were prepared in anhydrous methanol. The reactions between acetic anhydride and dendrimer were carried out in a glass flask at room temperature for 24 h. The ratio between acetic anhydride and dendrimer was varied to synthesize PAMAM dendrimers with various degrees of acetylation. The standard deviation of the average acetylation ratio was about 5%, based on titration experiments. The reaction products were dialyzed first against PBS buffer (pH 8.0) and then against deionized water over night. The purified acetylated PAMAM dendrimer samples were lyophilized and stored at -20 °C. More details about dendrimer acetylation can be found elsewhere [100]. TRITC labeled PAMAM dendrimer was prepared in methanol and purified by 10 KD MWCO dialysis and ultra-filtration. Purified TRITC labeled dendrimer was stored at -20 °C and kept away from light.

3.3.3 Dynamic Light Scattering

Salmon sperm DNA and PAMAM dendrimer with various degrees of acetylation were prepared in 10 mM NaBr solutions and filtered by 0.2 μm Minisart filters (Sartorius, New York, NY, USA). The mixtures of DNA and dendrimer were prepared by adding 500 μL dendrimer solutions into equal volume of DNA solutions. The concentrations of DNA in all mixtures were fixed at 0.15 mg/mL. Dynamic light scattering measurements were carried out at 25 °C after at least 3 h-reaction. Then the pH of mixtures were determined to be $7 \sim 8$, where the primary amine groups of the PAMAM dendrimer are protonated. More details

about DNA dendrimer complex formation can be found elsewhere [78]. All dynamic light scattering experiments were conducted on an ALV (Langen, Hessen, Germany) compact goniometer system. The wavelength of incident laser light was $\lambda = 488$ nm (Innova 70C argon ion laser, Coherent Inc., Santa Clara, CA, USA). Scattered light was collected by dual avalanche photodiode detectors, which were in the transmission mode and then sent to a multi tau correlator (ALV-5000E). The time-averaged normalized intensity autocorrelation function was constructed by cross correlating the signal. The hydrodynamic radius of the complex particle was calculated from the decay times using CONTIN 2DP analysis together with the Stokes-Einstein equation. Light scattering angle was fixed at 50° for all measurements.

3.3.4 Fluorescence Spectroscopy

Steady-state fluorescence emission measurements were carried out on Fluoromax-2 fluorimeter using a 10 $\times$ 10 mm quartz cuvette (Starna, Atascadero, CA, USA). The excitation wavelength was set to 493 nm. Emission spectra were recorded between 400 nm and 700 nm in 1 nm increments. Integration time was set to 0.1 s. For each sample, 3 scans were accumulated and averaged. Salmon sperm DNA-dendrimer complexes solutions were incubated with equal amount of GelStar® nucleic acid stain for more than 10 min before measurement, and DNA final concentrations were fixed at 2 μg/mL.

3.3.5 Circular Dichroism Spectroscopy

All circular dichroism (CD) measurements were conducted on an Aviv model 202 circular dichroism spectrometer at 25 °C using a 10 mm quartz cell. Spectra were recorded

between 220 nm and 320 nm with 1 nm increments and 1 s averaging time. The bandwidth was set to 1 nm. Five scans were averaged for each sample. Salmon sperm DNA and PAMAM dendrimer with various degrees of acetylation were incubated together to form complexes in 20 mM HEPES buffer for at least 3 h before CD measurements. DNA concentrations in the final solutions were fixed at 65.0 μg/mL. All CD measurements were baseline corrected by subtracting the blank from the recorded spectra.

3.3.6 Molecular Combing Assay

For molecular combing, YOYO-1 stained λ-DNA was diluted into 2 pM in TE buffer (pH 8.0). λ-DNA was mixed with TRITC labeled PAMAM dendrimer to form complexes. For DNA-dendrimer complexes solutions, the ratio between dendrimer and DNA was set to 100. 100 mg/mL polystyrene (MW 10,000, Sigma, St. Louis, MO, USA) solutions were prepared in toluene and spin-coated onto a cover glass (25 $\times$ 25 mm, Corning, No. 1) at 3,000 rpm for 30 s. Then the PS-coated cover glasses were each dipped into either DNA solution or DNA-dendrimer complexes solutions for 3 min and then pulled out at a constant rate (200 μm/s) using a linear stage (Zaber, T-LLS series, Vancouver, BC, Canada), as shown in Figure 3.2.5a.

3.3.7 Flow Stretching Assay

Cover glasses (25 $\times$ 25 mm, Corning, No. 1) and glass slides (Fisher, Pittsburgh, PA, USA) were cleaned by nitric acid and hydrochloric acid and then rinsed by deionized water. Then the cover glass was attached to the slide using double-sided tape to form the bottom wall of a flow channel (length 20 mm, width 8 mm and height 80 μm). The flow channel

was incubated first by 5 mg/mL neutravidin (Invitrogen) for 1 h and then by 2 mg/mL α-casein (Sigma) for 0.5 h. Biotinylated λ-DNA (3 nM) was injected into the flow channel and reacted with the neutravidin attached to the flow channel surface for 5 min. Then the flow channel was washed by TE buffer (pH 8.0) for 1 min to expel the free DNA. 1 nM TRITC labeled PAMAM dendrimer in TE buffer was then introduced to flow channel at a constant rate (10 mL/h) using a syringe pump (KD Scientific 100 series) to stretch tethered λ-DNA for 20 min which allowed the dendrimer to bind to the λ-DNA, which is shown in Figure 3.2.6a.

3.3.8 Fluorescence Microscopy

YOYO-1-stained λ-DNA and TRITC-labeled PAMAM dendrimer were visualized by a Nikon TE2000-U inverted fluorescence microscope with a front-illuminated CCD camera (Cascade 512F, Roper Scientific, Tucson, AZ, USA). A dual-band excitation and emission FITC-TRITC filter with X-CITE® 120 series lamp were used to illuminate the specimens. The integration time for each frame was set to 200 ms. The camera exposure time and filter wheel shutter were controlled by MetaMorph 7.0.

3.4 Conclusions

In conclusion, PAMAM dendrimers with a fraction of their primary amine groups acetylated have were used to condense DNA and the complexes were characterized using light scattering, circular dichroism, and optical fluorescence microscopy both after combing complexes formed in bulk solution onto a hydrophobic surface, and during complex formation onto

flow-stretched DNA tethered onto a surface. Using dynamic light scattering, we confirmed that unless a high fraction (>70%) of primary groups was neutralized the acetylated dendrimer was able to compact DNA even with reduced charge, and therefore presumably with reduced cytotoxicity. The bound DNA retains its B-form, despite the condensation, and condensed form is strong enough to resist the forces created by molecular combing, which are on the order of 500 pN [98]. We imaged this dendrimer-DNA binding process in solution and in real time for the first time. When DNA tethered to a surface was exposed to dendrimer introduced continuously by flow, we observed all DNA molecules gradually load up with dendrimers, which were not washed out by subsequent dendrimer-free solutions. We did not observe co-existence of dendrimer-loaded DNA with dendrimer-free DNA, and at high concentrations of dendrimer, the DNA became loaded with dendrimers almost instantly. Thus, taken together, our single-molecule experimental results indicate that the coexistence of DNA/dendrimer complexes with free DNA observed in bulk solution [78, 92] might not be due to cooperative binding of dendrimer to DNA, but rather to diffusion-limited and irreversible reaction between dendrimer and DNA that can occur when dendrimer and DNA are mixed in the bulk.

CHAPTER IV

Monte-Carlo simulations of PAMAM dendrimer-DNA interactions

We use Monte Carlo simulations to determine the influence of poly(amido amine) (PAMAM) dendrimer size and charge on its interactions with double-stranded DNA conformation and interaction strength. To achieve a compromise between simulation speed and molecular detail, we combine the coarse-grained DNA model of de Pablo et al. [35] which resolves each DNA base using three beads —and thereby retains the double-helix structure —with a dendrimer model with resolution similar to that of the DNA. The resulting predictions of the effects of dendrimer generation, dendrimer surface charge density, and salt concentration on dendrimer-DNA complexes are in agreement with both experiments and all-atom MD simulations. The model predicts that DNA wraps a fully charged G5 or G6 dendrimer at low salt concentration (10 mM) similarly to a histone octamer, and for the G5 dendrimer, DNA super helices with both handednesses occur. At salt concentrations above 50 mM, or when a high fraction of dendrimer surface charges are neutralized by acetylation, DNA adheres but does not compactly wrap the dendrimer, in agreement with

experimental findings. We are also able to simulate pairs of dendrimers binding to the same DNA strand. Thus, our mesoscale simulation not only elucidates dendrimer-DNA interactions, but also provides a methodology for efficiently simulating chromatin formation and other cationic macroion-DNA complexes.

4.1 Introduction

Extensive experimental studies, e.g., cryo-TEM [83, 87], small angle X-ray scattering [85], and dynamic light scattering [78, 101], have recently been performed to investigate the size and morphology of complexes of cationic poly(amido amine) (PAMAM) dendrimer with double-stranded DNA. This interest is driven by the ability of this dendrimer to condense dsDNA, opening up possible applications in gene delivery [102, 103]. Since the strongest interaction between cationic dendrimers and DNA is electrostatic, the size as well as the morphology of the complex are highly dependent on salt concentration [104], pH [85], and dendrimer surface groups [86]. Therefore, a dendrimer-DNA model that captures the electrostatic interaction correctly, and with appropriate resolution, would stimulate deeper understanding of the physical properties of the complexes and help in the design of novel dendrimers for gene delivery. In addition, dendrimers of generation 5 or larger are roughly comparable in size to histone octamers and so can model DNA-histone interactions, and help in understanding DNA-protein interactions more generally. However, accurate all-atom molecular dynamics (AAMD) simulations can only be applied routinely to fairly small dendrimer-DNA systems [105, 106], since AAMD simulations are computationally expensive. On the other hand, widely used coarse-grained (CG) models [107–110], which use bead-spring or bead-rod chains to model dsDNA as an effectively

single-strand linear polyelectrolyte, can simulate binding and wrapping of DNA around a dendrimer. However since the diameter of dsDNA is almost the same as the radius of the dendrimer, the double-helix structure of DNA might be critical for local arrangement of the DNA along the dendrimer surface and thereby strongly affect the structure of the dendrimer-DNA complex. To the best of our knowledge, only the recent work of Cao et al. [111], which employs Savelyev and Papoian's double stranded coarse-grained DNA model [112] discusses how a double helix polyelectrolyte can be captured by a charged nanosphere, thus providing a simplified model of DNA condensation by a macroion. Note that in Savelyev and Papoian's DNA model, DNA melting and sequence-dependent bending were not allowed and a hard sphere with hundreds of charges evenly distributed over its surface was used for the macroion. Such a model, while useful especially for DNA condensation by hard, uniform nanoparticles, does not describe details of local DNA structures adsorbed to flexible nanoparticles such as dendrimers, and cannot describe local DNA melting. To study DNA local bending and melting caused by strong attractions between a dendrimer and DNA, a finer-scale, yet still coarse-grained, DNA model is necessary, one that captures the double helix structure of dsDNA, allowing DNA double strand super-coiling, melting and re-association. Such a model, called the "3SPN" coarse-grained model, was developed by de Pablo and co-workers [26, 35, 113, 114], in which a single bead represents each of the phosphate, sugar, and bases, reducing the number of degrees of freedom significantly relative to atomistic models. Moreover, this "3SPN" model can generate accurate DNA sequence-dependent melting curves as well as an accurate persistence length for double stranded DNA. Thus, the 3SPN model makes it easily affordable to study the interactions of a relatively long strand of dsDNA (e.g., 144 bp) with dendrimers of various generations without sacrificing the DNA local double helical structure. Another

main advantage of the 3SPN model is that the electrostatic force is treated using the Debye-Hückel theory which can easily be adapted (as we do here) to compute the electrostatic interactions with other molecules, in particular, charged dendrimers. However, since the 3SPN model is based on the classical B-form dsDNA structure, Hogsteen interactions are not included in this DNA model.

It is worth mentioning that a CG model that incorporates both this 3SPN DNA model and a CG dendrimer model with similar resolution has not heretofore been published. Thus, here, we combine the charged bead-spring dendrimer model of Muthukumar and coworkers [107] with the 3SPN DNA model by choosing an appropriate repulsive force between DNA and dendrimer beads that leads to a good match to the free energy of dendrimer-DNA binding found in all-atom MD simulations [106]. As will be discussed below, the combination of Muthukumar and de Pablo models does not significantly disrupt the local double helix structure of the DNA.

Using this coarse-grained model, which we call the 3SPN-WM (Welch, Muthukumar [107]) model, we seek to study system-atically DNA-dendrimer complex formation as a function of dendrimer generation, dendrimer surface charge density, and salt concentration, and to validate our coarse-grained model by comparing these simulation results to our previous experimental work [101]. Our coarse-gained model could also be extended to describe other DNA-macroion complexes, such as the DNA-histone complex, which is of great biological importance. Compared to recent coarse-grained models of DNA-histone complexes [115–117], our new coarse-grained model presented here should give more detailed information regarding sequence-dependent bending of DNA for relatively long DNA sequences (>146 bp). Hence, our dendrimer-DNA (3SPN-WM) model is not only of interest in its own right, but is a good starting point for developing a new CG model

of the DNA-histone complex which can bridge the gap between an all-atom model and a simpler bead-spring model of this complex [115]. Such a CG model can help in the design of dendrimers with optimal surface charge density, which might increase transfection efficiency.

4.2 Simulation model and method

4.2.1 Dendrimer model

In this thesis, a PAMAM dendrimer with an ethylenediamine core was modeled by a bead-spring, so-called "united-atom" model [107]. An ethylenediamine (EDA) core PAMAM dendrimer has 33% more terminal groups than the same-generation ammonia-core PAMAM dendrimer simulated by Welch and Muthukumar [107], and thus is less flexible than the ammonia-core dendrimer. For an ethylenediamine-core dendrimer, the relationship between the number of dendrimer beads (N_D) in the CG model and the dendrimer generation number (G) is:

$$N_D = N_{interior} + N_{terminal} = (2^{G+2} - 2) + 2^{G+2} \tag{4.2.1}$$

where $N_{interior}$ and $N_{terminal}$ are the number of interior and terminal beads, respectively. Fig.4.2.1 shows how a generation 1 (G1) dendrimer is modeled, where the interior and terminal beads are rendered in magenta and cyan, respectively. The mass of each bead is 113 (amu). (Assignment of specific atoms to CG beads of each color in Fig.4.2.1 was not attempted, and bending and torsional potentials were not included in the dendrimer model,

although they are present in the DNA model of de Pablo and coworkers; see Eq.4.2.7-4.2.14) The CG model of an ammonia-core dendrimer by Welch and Muthukumar contains only a single bead at the center, attached to three branches, rather than the pair of beads attached to four branches for the ethylenediamine-core dendrimer depicted in Fig.4.2.1 Apart from this difference, the CG model used here is identical to that of Welch and Muthukumar [107]. Here, we include only three forces: bond stretching, excluded

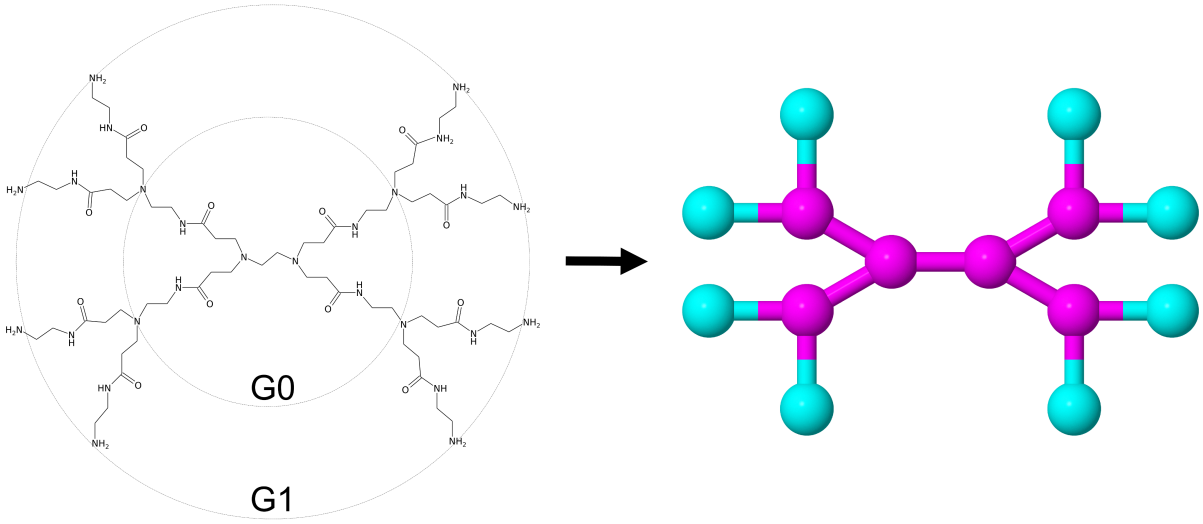

Figure 4.2.1: Depiction of the bead-spring model for a generation 1 (G1) PAMAM dendrimer. This, and other molecular depictions, were generated using Jmol: an open-source Java viewer for chemical structures.

volume, and Coulombic. As described in ref. [107], we used the finitely extensible nonlinear elastic (FENE) force to model the bond stretching energy (Eq.4.2.2).

$$\frac{U_{FENE}}{k_B T} = -KR^2 \sum_{i=1}^{N_D} log\left[1 - \left(\frac{l_i - l_0}{R}\right)^2\right] \tag{4.2.2}$$

where U_{FENE} is the total bond stretching energy for the dendrimer; K, set to $20.0/l_0^2$, is the spring constant; and l_i is bond length of the i^{th} bond. Also $l_0 = (l_{max} + l_{min})/2$, where l_{max}

and l_{min} are the maximum and minimum bond lengths respectively, and $R = l_{max} - l_{min}$. In all our simulations, l_{max} was set to $1.0l_B$, and l_{min} to $0.4l_B$, where l_B is the Bjerrum length (Eq.4.2.5) which is 7.1Å at 25 °C in water which are the same values as used by Welch and Muthukumar [107]. N_D is the number of dendrimer beads.

The electrostatic repulsion force between the charged dendrimer terminal beads is computed using the Debye-Hückel potential [107]: (Eq.4.2.3-4.2.5) explicitly demonstrate that the electrostatic interaction terms from de Pablo et al. and from Welch and Muthukumar are identical. The second term in Eq.4.2.3 was used in the work of de Pablo et al. [35], while the third term was used in Welch and Muthukumar [107])

$$\frac{U_{elec}}{k_B T} = \sum_{i<j}^{N_T} \frac{q_i q_j e_c^2 e^{-\kappa r_{ij}}}{4\pi\epsilon_0 \epsilon(T,C) r_{ij} k_B T} = l_B \sum_{i<j}^{N_T} \frac{q_i q_j e^{-\kappa r_{ij}}}{r_{ij}} \tag{4.2.3}$$

where U_{elec} is the electrostatic energy and κ is inverse Debye length, which is defined by:

$$\kappa^{-1} = \sqrt{\frac{\varepsilon_0 \varepsilon(T,C) k_B T}{2 N_A e_c^2 I}} \tag{4.2.4}$$

The Bjerrum length l_B is given by:

$$l_B = \frac{e_c^2}{4\pi\varepsilon_0 \varepsilon(T,C) k_B T} \tag{4.2.5}$$

Here r_{ij} is the distance between the i^{th} and j^{th} beads and q_i and q_j are the number of charges on the i^{th} and j^{th} beads, respectively. N_A is Avogadro's number; ε_0 is the vacuum permittivity, and $\varepsilon(T,C)$ is the temperature (T) and salt-concentration (C)-dependent dielectric constant. Since all the simulations were carried out at 25 °C, the dielectric constant only depends on salt concentration. The definition of $\varepsilon(T,C)$ can be found somewhere

else [118]. e_c and I are the elementary charge and ionic strength, respectively. N_T is the number of charged terminal groups on the dendrimer. To avoid overlapping of dendrimer beads, the excluded volume potentials between all dendrimer beads were computed using the Morse potential [107], as shown in Eq.4.2.6:

$$\frac{U_{Morse}}{k_BT} = \frac{\sigma}{k_BT}\sum_{i<j}^{N_D}\left[\left(e^{-\alpha(r_{ij}-d)}-1\right)^2-1\right] \tag{4.2.6}$$

where $\sigma = k_BT/0.7$; $\alpha^{-1} = l_B/24$ and $d = 0.8l_B$. Again, N_D is the number of dendrimer beads. For $r_{ij} > 9$Å the excluded volume potential was cut off.

4.2.2 DNA model (3SPN.1 force field)

We used the coarse-grained model of dsDNA developed by de Pablo and coworkers [35, 113] (3SPN.1 force field) to simulate the DNA. For completeness, we list below all equations used. The constants used in these equations and further details about this force field can be found in the original papers [35,113]. The bonded interactions of the 3SPN.1 force field are given by:

$$\frac{U_{bond}}{k_BT} = \beta\sum_{i=1}^{n_{bond}}\left[k_1\left(d_i-d_{i0}\right)^2+k_2\left(d_i-d_{i0}\right)^4\right] \tag{4.2.7}$$

$$\frac{U_{bend}}{k_BT} = \beta\sum_{i=1}^{n_{bend}}\frac{k_\theta}{2}\left(\theta_i-\theta_{i0}\right)^2 \tag{4.2.8}$$

$$\frac{U_{tors}}{k_BT} = \beta\sum_{i=1}^{n_{tors}}k_\phi\left[1-cos\left(\phi_i-\phi_{i0}\right)\right] \tag{4.2.9}$$

Here U_{bond}, U_{bend}, and U_{tors} represent bond stretch energy, bending energy, and dihedral energy respectively, with $\beta = 1/k_BT$. The nonbonded, pairwise interactions between DNA

beads are given by:

$$\frac{U_{stck}}{k_BT} = \beta \sum_{i=1}^{n_{stck}} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{4.2.10}$$

$$\frac{U_{base}}{k_BT} = \beta \sum_{i=1}^{n_{base}} 4\varepsilon_{bi} \left[5 \left(\frac{\sigma_{bi}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{bi}}{r_{ij}} \right)^{10} \right] \tag{4.2.11}$$

$$\frac{U_{nnat}}{k_BT} = \beta \sum_{i<j}^{n_{nnat}} \begin{cases} 4\varepsilon \left[\left(\frac{\sigma_0}{r_{ij}} \right)^{12} - \left(\frac{\sigma_0}{r_{ij}} \right)^{6} \right] + \varepsilon, & r_{ij} < r_{coff} \\ 0, & r_{ij} \geq r_{coff} \end{cases} \tag{4.2.12}$$

$$\frac{U_{elec}}{k_BT} = \sum_{i<j}^{n_{elec}} \frac{q_i q_j e_c^2 e^{-\kappa r_{ij}}}{4\pi\epsilon_0 \epsilon(T,C) r_{ij} k_B T} = l_B \sum_{i<j}^{n_{elec}} \frac{q_i q_j e^{-\kappa r_{ij}}}{r_{ij}} \tag{4.2.13}$$

$$\frac{U_{solv}}{k_BT} = \beta \sum_{i<j}^{n_{solv}} \varepsilon_S \left[1 - e^{-\alpha'(r_{ij} - r_S)} \right]^2 - \varepsilon_S \tag{4.2.14}$$

Both the base-stacking effects (U_{stck}), and the intra-strand interaction and hydrogen-bonding interactions between DNA complementary base pairs (U_{base}) contribute to the stiffness of the DNA double-strand helix. A complementary base-pair is considered to be hydrogen-bonded only when the distance between those two base beads is less than ($\sigma_{bi} + 2.0$Å). The excluded volume potential between DNA beads was included in the nonnative contact potential U_{nnat}. The electrostatic repulsions between DNA phosphate beads were calculated from a Debye-Hückel potential, as is the case with charged dendrimer beads; i.e., Eq.4.2.13 is the same as Eq.4.2.3. The solvent-induced energy U_{solv} was used to capture electrostatic correlation effects along the DNA [119]. A complete description of how to determine the parameters and DNA related potentials can be found in Sambriski et al [35].

4.2.3 Dendrimer-DNA interaction

Since the attractive force between DNA and the PAMAM dendrimer is dominated by electrostatic interactions [105], for simplicity, we only considered two different forces between DNA and dendrimer beads, namely the electrostatic and excluded volume forces. We again used the Debye-Hückel potential to model the electrostatic attraction between DNA and dendrimer beads (Eq.4.2.15). Compared to the dendrimer and DNA macroions, the salt ions and water molecules are small enough that an implicit model for their effect hopefully provides an adequate description.

$$\frac{U_{elec}}{k_B T} = \sum_{i=1}^{N_T} \sum_{j=1}^{N_P} \frac{q_i q_j e_c^2 e^{-\kappa r_{ij}}}{4\pi\epsilon_0 \epsilon(T,C) r_{ij} k_B T} = l_B \sum_{i=1}^{N_T} \sum_{j=1}^{N_P} \frac{q_i q_j e^{-\kappa r_{ij}}}{r_{ij}} \tag{4.2.15}$$

Note N_P here is the number of charged DNA beads, that is, the number of phosphate beads. Eq.4.2.15 is the same as Eq.4.2.3 and Eq.4.2.13, except that it is summed over only the cross-interactions between a DNA bead with a dendrimer bead. The excluded volume potential between PAMAM dendrimer beads and DNA beads was modeled by a truncated Lennard-Jones potential:

$$\frac{U_{excluded}}{k_B T} = \beta \sum_{i=1}^{N_D} \sum_{j=1}^{N_{DNA}} \begin{cases} 4\varepsilon \left[\left(\frac{\sigma'}{r_{ij}} \right)^{12} - \left(\frac{\sigma'}{r_{ij}} \right)^{6} \right] + \varepsilon, & r_{ij} < r_{cutoff} \\ 0, & r_{ij} \geq r_{cutoff} \end{cases} \tag{4.2.16}$$

where $\sigma' = 2^{-1/6} r_{cutoff}$, where ε in the above takes the same value as assigned to this symbol in Eq.4.2.10 and Eq.4.2.12. To determine r_{cutoff}, we compared the potential of mean force (PMF) computed from our CG model for a generation 3 (G3) PAMAM dendrimer interacting with a 24 base pair dsDNA (Fig.4.3.3a) with the corresponding

PMF for the interaction of these molecules from an all-atom molecular dynamics simulation. 12 The DNA base pair sequence used in these simulations is CGCGAATTCGCGCGC-GAATTCGCG [106, 120].

4.2.4 Simulation methods

We employed a Monte-Carlo simulation method in a canonical ensemble using the classical Metropolis sampling method to simulate the CG DNA-PAMAM dendrimer interaction. The trial displacements of each bead Δx, Δy, Δz were randomly distributed uniformly within the range (-0.35Å,0.35Å). The acceptance probability of our simulation was around 33%. Note we use an implicit solvent model, and periodic boundary conditions were not used. For our umbrella sampling simulations, the reaction coordinate was defined as the distance between the dendrimer center of mass and the center of mass of the two base pairs in the middle of the 24 bp DNA. A harmonic potential with a force constant of 0.25 kcal mol^{-1} Å^{-2} was applied along the reaction coordinate over each of 90 successive windows in 1 Å increments. Force constants of 0.5 kcal mol^{-1} Å^{-2} were also used for limited runs, and showed similar results to those obtained with a constant of 0.25 kcal mol^{-1} Å^{-2}. For each window, the system was equilibrated for 3×10^6 Monte Carlo steps (MCS), where a single step consists of an attempt to move a single bead. Larger force constants (e.g., above 1 kcal mol^{-1} Å^{-2}) resulted in poorer overlap of the histograms, which caused bigger errors for the specific simulation time used here. The potential of mean force (PMF) was obtained using the weighted histogram analysis method (WHAM) [121].

4.3 Results and discussion

4.3.1 PAMAM dendrimer simulations

As described above, we use the dendrimer model of Welch and Muthukumar [107] to simulate a PAMAM dendrimer with an ethylenediamine (EDA) core. To test the suitability of the model, we report here the radii of gyration (R_g) of EDA-core dendrimers of generations G3 through G6, and for G5 dendrimers acetylated to different extents. The largest dendrimer that we simulate (G6) is represented by 510 beads.

Dendrimer generation and salt concentration effects. As shown in 4.3.1, G3-G6 PAMAM dendrimers show a 12% (G3) to 21% (G6) decrease in the radius of gyration (R_g) as the salt concentration increases from 10 to 1000 mM. As expected, dendrimers of higher generation shrink more as salt concentration increases, and repulsive forces of charged terminal groups are reduced by ion screening. Table below compares values of R g from our simulations with those from small angle X-ray scattering (SAXS) experiments and from molecular dynamics simulations. The good agreement shows that the CG model of Welch and Muthukumar can be used to obtain the correct size of the EDA-cored PAMAM dendrimers considered here, as well as the ammonia-core dendrimers considered by Welch and Muthukumar [107].

Note that increasing salt concentration has similar effects on dendrimer size as does increasing pH, since at high pH only a fraction of the dendrimer terminal amine groups is protonated, which weakens the repulsion between these terminal groups relative to those with fully charged surface groups at low pH. However, the molecular dynamics simulations of a G4 dendrimer by Liu et al. [122] show only a 4.9% decrease in R_g when the pH increases from low (5) to high (>10).

Table Radii of gyration R_g (in Å) for EDA-core PAMAM dendrimers[a]

Dendrimer generation	No. of beads	No. of charges	R_g (1M[Na^+]) (this work)	R_g Maiti et al. [123, 124]	R_g Lee et al. [125]	R_g SAXS [126]
G3	62	32	13.94±0.60	11.23 (Ref. [123])		15.8
G4	126	64	17.19±0.41	17.01±0.10 (Ref. [124])		17.1
G5	254	128	21.05±0.52	22.19±0.14 (Ref. [124])	25.1± 0.1	24.1
G5-90%	254	13.7±4.0	20.78±0.37		21.1± 0.1	
G6	510	256	25.96±1.06	27.28±0.39 (Ref. [124])		26.3

[a] Debye length in a 1M [Na^+] solution is 3Å.The R_g values from the work of Maiti et al. [123, 124] were obtained at high salt concentration with Debye length equal to 4.3Å. The R_g values from Lee et al. [125] were obtained by MD simulations with no salt. SAXS experiments [126] were carried out in methanol.

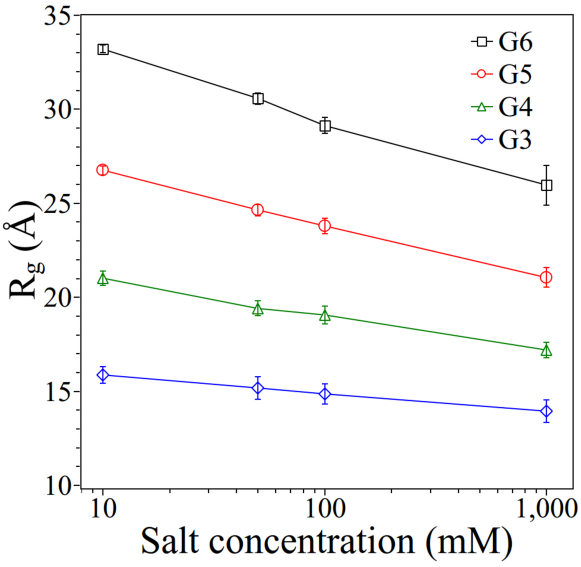

Figure 4.3.1: Radii of gyration of PAMAM dendrimers as a function of salt concentration. (The error bars are standard deviations).

Acetylation effects. The surface charge density of a dendrimer affects its ability to condense DNA as well as its cyto-toxicity. For example, a dendrimer with a higher surface

charge density has a higher transfection efficiency [88], but could be more cytotoxic as it can generate pores on the cell membrane. Thus, adjusting the dendrimer charge density by surface modifications such as PEGylation or acetylation is of great interest. Since we adopted a coarse-grained dendrimer model and the acetyl functional group is small compared to the size of dendrimer CG beads, we simulate the acetylated dendrimer simply by randomly assigning zero charges to some of the terminal beads of the dendrimer. The R_g of the resulting acetylated dendrimers are plotted against percentage acetylation in Fig.4.3.2. For each degree of acetylation, three runs at 10 mM salt concentration were averaged together. For a low degree of acetylation (<50%), R_g decreases almost linearly. For higher degrees of acetylation, a near-plateau in dendrimer size is eventually attained. A fully neutralized G5 dendrimer has a 22% smaller radius of gyration than a non-acetylated dendrimer, which is a somewhat greater reduction than shown by corresponding atomistic MD simulations ($\sim$16%) [125].

4.3.2 DNA simulations

We verified our code by simulating a 38 bp dsDNA and comparing its persistence lengths at different salt concentrations to the results of de Pablo and coworkers [35], using the same method described in their paper. The persistence length in that paper was defined by $\langle \bar{u}(0) \cdot \bar{u}(s) \rangle = e^{-s/l_p}$, where $\bar{u}(s)$ is the unit tangent vector at relative position s along the DNA, and l_p is the persistence length. Since different persistence length formulas give different results for macromolecular chains [127], we only compare our persistence length results to those of de Pablo et al. Our DNA sequence is chosen to be the same as that used by Nandy and Maiti [105] with CG content of DNA equals to 0.659. The

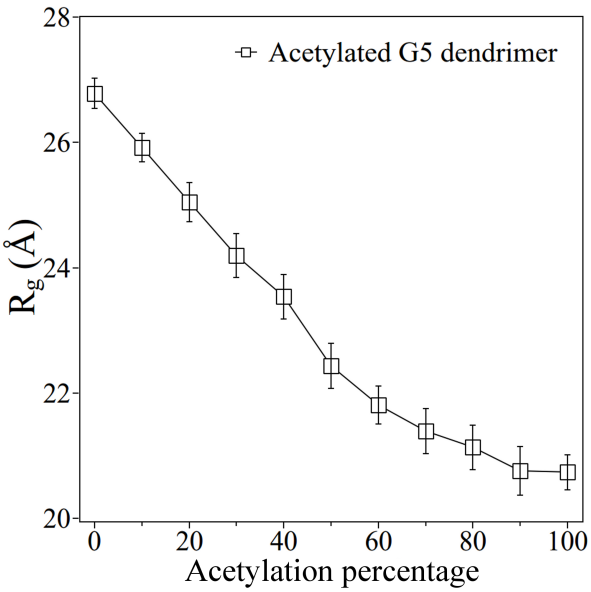

Figure 4.3.2: Radius of gyration of PAMAM dendrimer as a function of acetylation percentage. (The error bars are standard deviations).

persistence lengths of this specific DNA obtanined by our simulations are 43±3 nm, 39±6 nm, 74±2 nm for 150 mM, 100 mM, 10 mM salt concentrations, respectively, which agree with the results of de Pablo and coworkers. The increase in persistence length (l_p) of DNA when salt concentration is reduced to 10 mM is caused by the stronger repulsions between phosphate beads of DNA. Because obtaining l_p using Monte-Carlo (MC) simulations is computationally expensive, we do not obtain l_p for longer DNA using our MC code.

4.3.3 Potential of mean force (PMF) of dendrimer-DNA binding

The free energy change or potential of mean force (PMF) of a G3 PAMAM dendrimer binding to a 24 bp DNA molecule was obtained by umbrella sampling and the results are plotted in Fig.4.3.3b. The PMFs were used for parameterization of the dendrimer-DNA

interaction. Specifically, we compared the PMFs from our CG model with various excluded volume cut-off distances (defined in Eq.4.2.16) to results from all-atom MD (AAMD) simulations [106]. We then chose a cut-off distance giving good agreement to the PMF from the AAMD simulations for the G3-DNA interaction, and used this for higher generation dendrimers as well, since the dendrimer terminal groups were the same and were represented by the same bead type regardless of the generation. However, to confirm that this potential is adequate for other generations of dendrimer, structural parameters such as the radius of gyration of the DNA-dendrimer complex should also be compared to results from the AAMD simulations. As shown in Fig.4.3.3b, when $r_{cutoff} = 10$Å, the minimum of the PMF curve is located between 20Å and 30Å, and free energy change of binding reaction is -10 kcal mol^{-1}, which agree well with AAMD results (-10.9$\sim$13.5 kcal mol^{-1}) [106]. Therefore, here we use $r_{cutoff} = 10$Å for all our simulations. The large difference between the curves in Fig.4.3.3b indicates that the binding free energy is very sensitive to this cut-off, r_{cutoff}. This sensitivity arises from the strong distance dependence of the electrostatic interactions between dendrimer and DNA beads. The simple cut-off scheme is chosen, despite the sensitivity to the cut-off parameter, because our coarse-grained model cannot capture the atomic-level interactions that determine the actual local structure. Hence we use a cut-off tuned to ensure that at least we obtain the same PMF as is found in the all-atom simulations, so that model correctly the net attractive interaction between DNA and dendrimer.

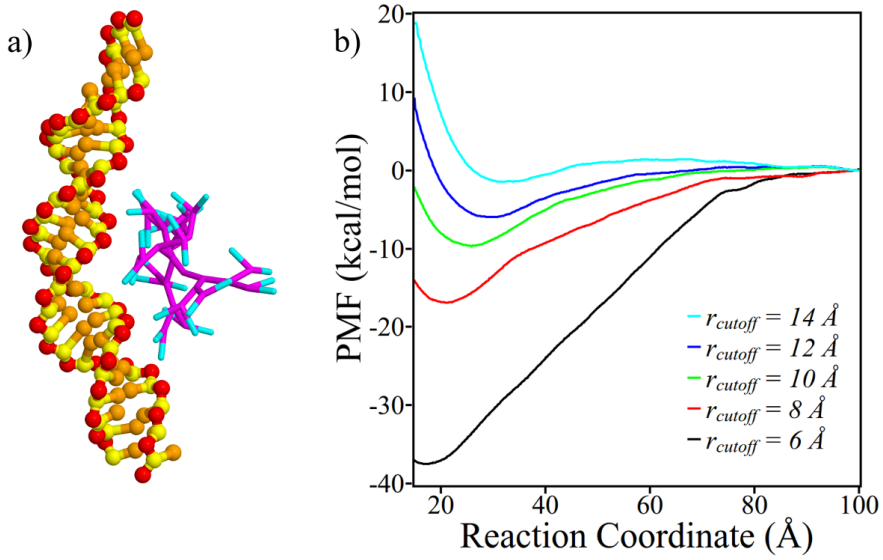

Figure 4.3.3: (a) Image of G3 PAMAM dendrimer binding a 24 bp DNA. (b) Corresponding potential of mean force (PMF) curves for different cutoff distances (r_{cutoff}).

4.3.4 PAMAM dendrimer-DNA complex formation

Snapshots of a 144 bp dsDNA binding to a G6 dendrimer are shown in Fig.4.3.4. Since all our simulations were carried out using a Monte-Carlo simulation method, this binding process does not necessarily show the actual dynamics but only one path to the final structure. A G6 dendrimer is of similar size and charge density to a histone octamer. Therefore, just as histones do not form the classical nucleosome structures with DNA at salt concentrations as low as 10 mM, a G6 dendrimer attracts DNA too strongly for an equilibrium complex structure to be achieved. The dendrimer-DNA complex shown in Fig.4.3.4.c is only one of the multiple metastable structures we obtained in our simulations (others not shown) after simulations up to 1.6×10^7 Monte Carlo steps (MCS). While the structure in Fig.4.3.4.c is a right-handed helix, others have different handedness, or disordered structures. These metastable structures are sufficiently long-lived that their properties, such

as R_g, could be averaged over a long enough time to obtain meaningful results. Since dendrimer-DNA complex formation might also be kinetic-controlled in experiments [78], the metastable structures determined here by simulation might provide physically realistic structural information of the complexes. In the simulation depicted in Fig.4.3.4, we used the same 144 DNA base-pair sequence given elsewhere [35].

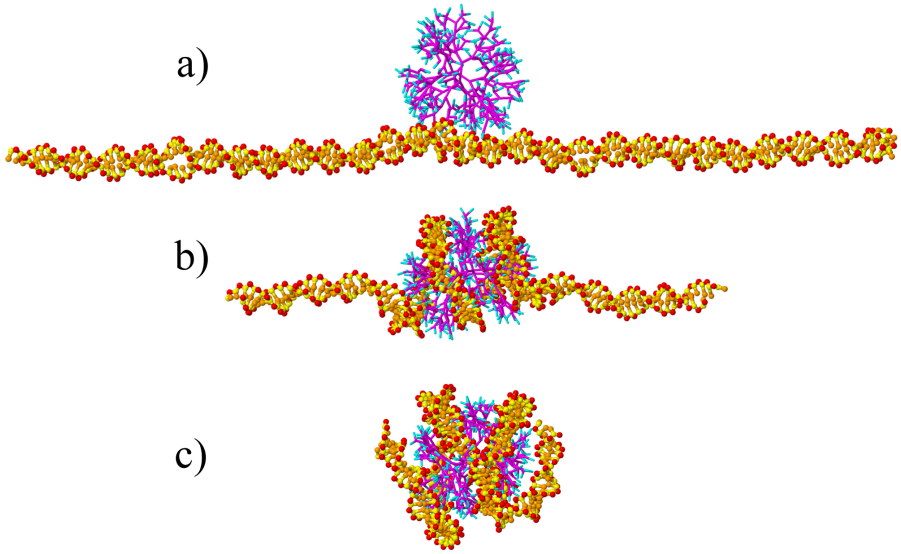

Figure 4.3.4: Snapshots of 144bp dsDND-G6 dendrimer interaction at 10mM salt concentration: (a) after 6×10^4 MCS, (b) after 6×10^6 MCS, (c) after 1.6×10^7 MCS.

Effects of dendrimer generation and salt concentration. To further validate our CG model, we investigated the condensation of a 38 bp DNA by G3, G4, and G5 dendrimers at various salt concentrations and compared the results to those of AAMD simulations 11 directly. We define r_{charge} by Eq.4.3.1:

$$r_{charge} = \frac{\text{number of positive charges on dendrimer}}{\text{number of negative charges on DNA}} \tag{4.3.1}$$

Note that r_{charge} is less than unity for G3-38 bp DNA and G4-38 bp DNA complexes but larger than unity for the G5-38 bp DNA complex. We also simulated a G5-72 bp DNA complex, for which $r_{charge}<1$. The base-pair sequence of the 72 bp DNA is the same as the central 72 base-pairs of the 144 bp DNA discussed above. We limited ourselves primarily to $r_{charge}<1$ to study DNA condensation by dendrimers rather than decoration of DNA by dendrimers. The effects of dendrimer generation and salt concentration on complex formation are summarized in Fig.4.3.5-4.3.9.

As shown in Fig.4.3.5, a DNA molecule wraps around a dendrimer tightly at low salt

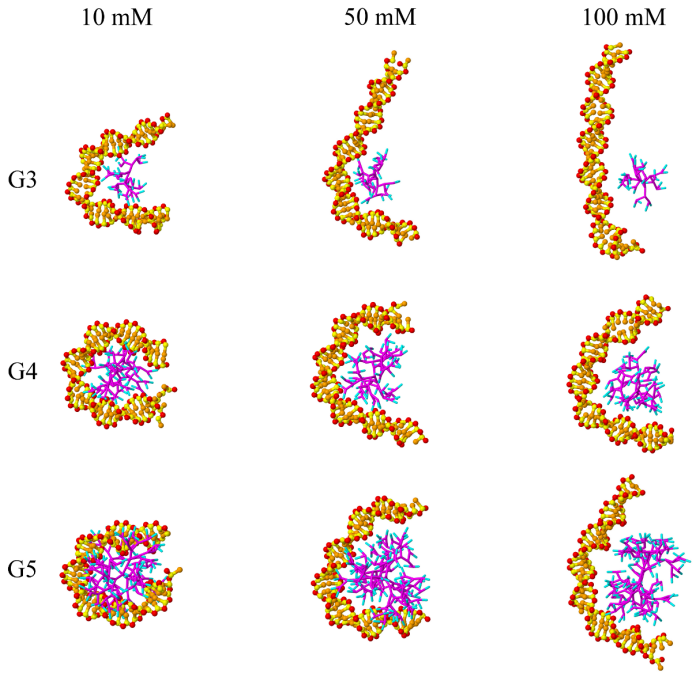

Figure 4.3.5: 38 bp DNA condensed by G3, G4, and G5 dendrimers at salt concentrations of 10, 50, and 100 mM.

concentration (10 mM), but is much straighter at high salt concentration (100 mM), due to the stronger ion screening effects at higher salt concentration. On the other hand, for a given salt concentration the DNA compacts more tightly with a higher generation den-

drimer. Therefore, a high generation dendrimer at high salt concentrations behaves like a lower generation dendrimer at lower salt concentration, in agreement with observations by cryo-TEM [104]. (Unlike the larger dendrimer-DNA complex (G6-144 bp DNA), the specific structures in Fig.4.3.5 can be achieved in repeat runs.)

Radii of gyration (R_g) of complexes and of the dendrimers and DNA molecules within those complexes are plotted in Fig.4.3.6. Error bars in Fig.4.3.6 are the standard deviations of R_g from 500 sampling points where bead coordinates were recorded every 1×10^4 MCS after 5×10^6 MCS equilibration. Our R_g results for G4-G5 dendrimer-38 bp DNA complexes at 10 mM salt concentration are similar to the AAMD results [105], which further validates our CG model of the dendrimer-DNA interaction. AAMD simulation results for longer DNA-dendrimer complexes are unavailable. Note that the R_g values of these complexes are very sensitive to the salt concentration. The R_g of the dendrimer alone increases monotonically when salt concentration is reduced because the electrostatic repulsion within the dendrimer molecule becomes stronger. But the R_g of the DNA in the complex and the R_g of complex itself both decrease monotonically with decreasing salt concentration because of increased attraction between DNA and dendrimer at the lower salt concentration. And for 10 mM salt, even a G3 dendrimer is able to condense a 38 bp DNA significantly.

The fractions of DNA phosphate beads adsorbed to the dendrimer surfaces for different complexes are plotted in Fig.4.3.7a. An "adsorbed" DNA phosphate bead is defined as one whose distance to the nearest dendrimer bead is less than $l_B + r_{cutoff} = 17.1$Å. As shown in Fig.4.3.7a, the adsorption in all four complexes decreases monotonically with increasing salt concentration, as expected. To quantify the arrangement of adsorbed DNA segments

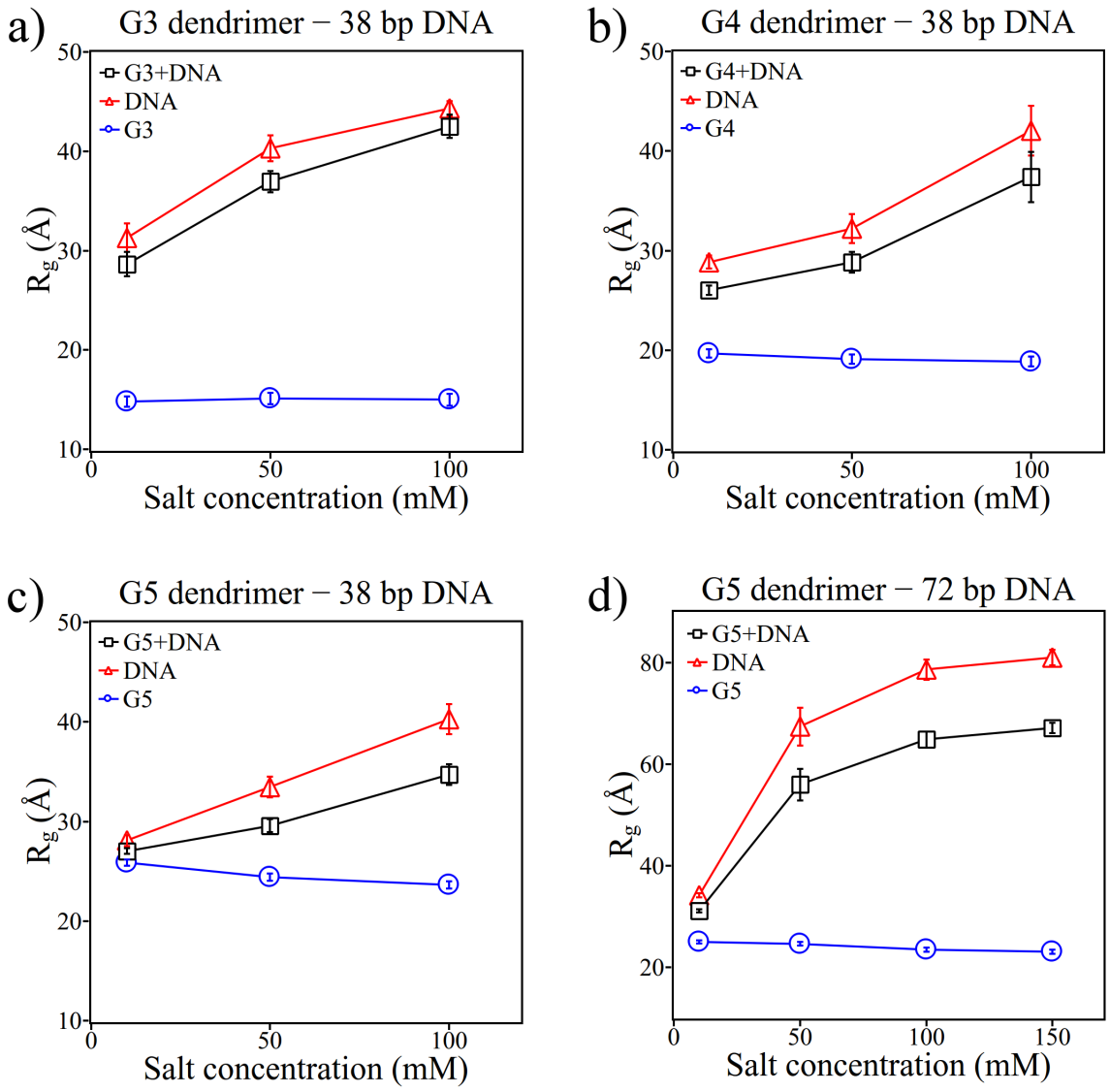

Figure 4.3.6: R_g of complexes (black) as well as of dendrimers (blue) and DNA molecules (red) within those complexes as functions of salt concentrations. (a) G3-38 bp DNA, (b) G4-38bp DNA, (c) G5-38 bp DNA,(d) G5-72 bp DNA. (The error bars are standard deviations).

on the dendrimer surface, we use a curvature order parameter η [111, 128] defined by:

$$\eta = \frac{|\sum_{\langle i \rangle} \bar{r}_{i,i+1} \times \bar{r}_{i+1,i+2}|}{N_d} \tag{4.3.2}$$

where $\bar{r}_{i,i+1} = \frac{r'_{i,i+1}}{|r'_{i,i+1}|}$, $r'_{i,i+1} = r'_{i+1} - r'_i$, $r'_i = (r_i + r_{i+n_{base}})/2$, N_d is the number of adsorbed DNA phosphate bead pairs, n_{base} is the total number of base pairs of the DNA, r_i is the i^{th} phosphate bead position, and $r_{i+n_{base}}$ is the phosphate bead position of the complementary base. Note that Eq.4.3.2 only accounts for the cross product of adjacent vectors $r_{i,i+1}$ and $r_{i+1,i+2}$ for which the i^{th}, $(i+1)^{th}$, $(i+2)^{th}$ base pairs were all adsorbed onto the dendrimer (the i^{th} base pair was considered as "adsorbed" when either i^{th} or $(i + n_{base})^{th}$ phosphate bead was adsorbed onto the dendrimer). The order parameter defined in Eq.4.3.2 is zero when the DNA conformation is straight, and when it is randomly coiled. The curvature order parameter could be unity if the backbone of the DNA were to bend an angle of 90° at each basepair, always in the same direction, forming a tight helix. This order parameter was originally designed for a single-strand helix-forming molecule, [128] and when applied to each strand of a straight dsDNA molecule, taking the position vectors to be phosphates, yields an order parameter of $\eta = 0.4271$. For straight B-form dsDNA with position vectors taken to be the midpoints between the paired phosphate beads as assumed in Eq.4.3.2, the order parameter is nearly zero ($\eta = 0.0076494$). (For dsDNA, the order parameter defined in Eq.4.3.2 is used to describe supercoiling of double strand DNA.)

Disordered complexes, where the DNA chain is adsorbed to the dendrimer surface randomly, produce small values of the order parameter η (e.g., <0.05 (ref. [111]). Our results (Fig.4.3.7b) shows that the degree of DNA order decreases with decreasing salt concentration, due to more severe bending of DNA at low salt concentrations. The error bars in Fig.4.3.7 are standard deviations over 500 sampling points, which were taken every 1×10^4 MCS after equilibration. The standard deviation in the order parameter is much larger than of the adsorption fraction because there are many microstates with similar adsorption fraction but quite different order parameter, since the system energy is mainly

controlled by the electrostatic attraction. The standard error of the mean is much smaller than the standard deviations shown by the error bars in Fig.4.3.7.

Since a dendrimer can induce global bending of DNA double strands through strong elec-

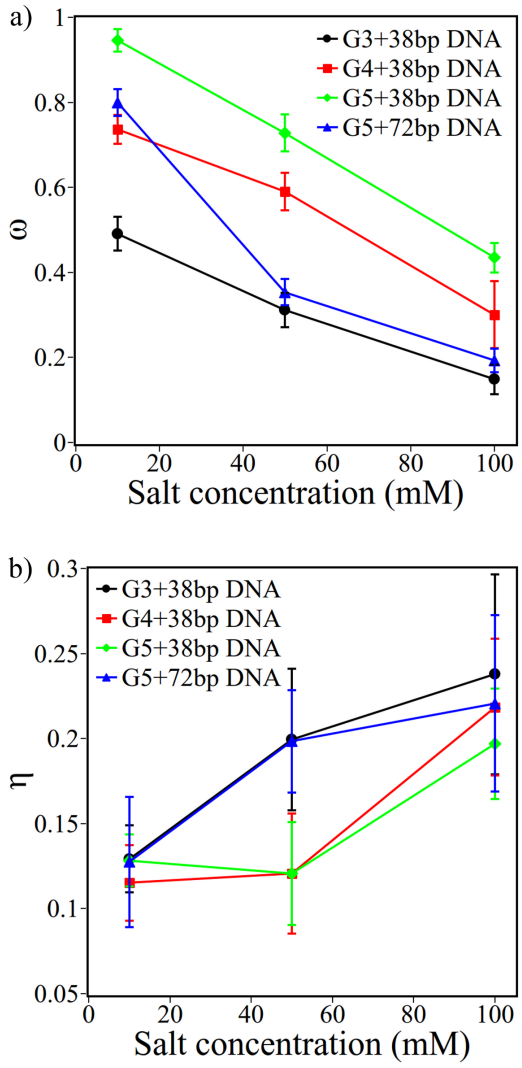

Figure 4.3.7: (a) Fraction ω of adsorbed DNA phosphate beads and (b) order parameter η in dendrimer-DNA complexes as functions of salt concentration. The error bars represent standard deviations taken over 500 sampling points of the simulation.

trostatic interactions, it can also alter the local structure of DNA or even disrupt the DNA

complementary base-pairs. To quantify this effect, we measured the DNA base-pair denaturing ratios $(1-\Phi)$ over 500 sampling points, which were taken every 1×10^4 MCS after equilibration. We define denaturation, or hydrogen-bond breakage, as occurring when the distance between the two complementary base beads becomes larger than $(\sigma + 2.0\text{Å})$. The results are presented in Fig. 4.3.8, where the error bars are the standard errors of those 500 sampling points. The denaturation fraction for a 38 bp DNA double-strand is larger than for a 72 bp DNA, because the melting temperature of a short DNA molecule is lower than that of a long one, and all our simulations were carried out at the same temperature. More complementary base-pairs denature in the DNA molecules condensed by the dendrimer than in free DNAs because of bending induced by the dendrimer. Note that all these denaturation ratios were around 0.1. Therefore, although the bending or distortion introduced by dendrimers bound to the DNA might cause local base-pair melting, this DNA denaturing is not strong for salt concentrations ranging from 10 mM to 100 mM. This finding perfectly agrees with our previous circular dichroism (CD) experiments [101] showing that DNA molecules remain in the classical B-form in dendrimer-DNA complexes. Unlike some other synthetic nanoparticles [91], no strong hydrophobic force exists between dendrimer and DNA bases. Therefore, PAMAM dendrimer can condense the dsDNA effectively without disrupting the local complementary base-pair structure, as confirmed by our simulations.

Dendrimer surface charge density effects. One of our objectives in developing this CG model of dendrimer-DNA complexation is to investigate how the surface charge density of a dendrimer affects the condensed dsDNA structure with improved resolution over earlier results that used more coarse-grained models [107–109]. By varying pH or salt concentration, the interaction between DNA and the dendrimer can be changed significantly.

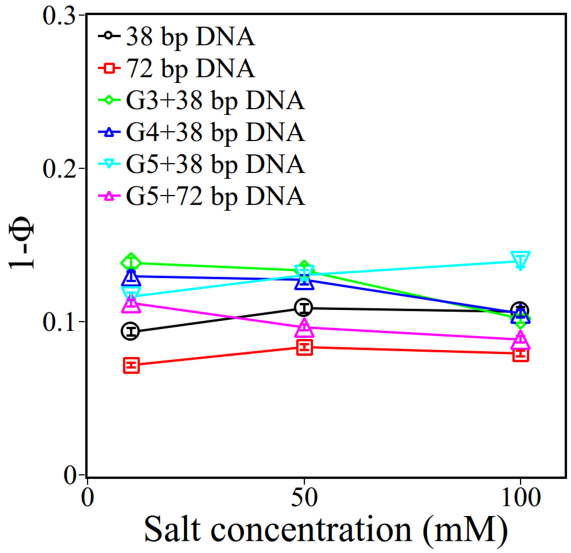

Figure 4.3.8: Fraction of base-pair denaturation of free DNAs or DNAs condensed by dendrimer molecules as a function of salt concentration.The small error bars shown are the standard errors of the mean.

However, changing pH or salt concentration might also change DNA structure, which is not captured in DNA models that do not include base-pairing, such as effectively single-strand polyelectrolyte models. More importantly, to condense or compact DNA molecules for gene delivery, the range of pH and of salt concentration is constrained to avoid cell damage. Thus, modifying the surface charge density of a dendrimer, or more generally, any DNA carrier, through acetylation or PEGylation, can be a better method of controlling macroion-DNA interaction, than adjusting pH or salt concentration.

Here we only consider acetylation. But as discussed recently [129], it is easy to extend our model to study PEGylated dendrimer-DNA interactions. As discussed above, here we account for acetylation of a dendrimer terminal group by simply randomly "switching off" the dendrimer terminal bead charges with probability corresponding to the fraction of neutralized charges. The radii of gyration (R_g), adsorption ratios (ω), and order parameters

(eta) were obtained by averaging over three independent runs for each acetylation ratio. Because charges of dendrimer terminal beads were randomly set to zero, there were different numbers of charges on the dendrimer in each of the three simulations. This allows us to account for dispersion in properties of the complexes (e.g., R_g) due to non-uniformity in dendrimer charge, which is present in experimentally synthesized dendrimers.

Average properties as well as standard deviations of acetylated G5 complexes at 10 mM salt concentration are plotted in Fig.4.3.9. As shown in Fig.4.3.9a, the R_g of the complex increases with the acetylation fraction, as a result of decreasing dendrimer-DNA attraction. However, the R_g of the complex increases very slowly when the acetylation is less than 40%, but it increases rapidly above about 50%. This R_g vs. acetylation curve explains well our dynamic light scattering results [101] that only when the acetylation is less than 50% is there a clear one-peak relaxation time distribution of the dendrimer-DNA complex. When the acetylation of the dendrimer was larger than 50%, two peaks in the relaxation time distribution were observed, similar to that of free DNA, with no dendrimer present at all. The two peaks represent the internal motion and translational motion of the free DNA or of the complexes. The persistence of these two peaks for DNA interacting with highly acetylated dendrimers implies that, rather than strongly condense DNA, weakly charged G5 dendrimers only slightly bend DNA double strands, as predicted by our CG model, thereby allowing DNA to retain its internal motion.

As discussed above, in 10 mM salt, adsorbed DNA phosphate beads neutralize the positive surface charges of the dendrimer. As a result, the fraction of DNA adsorption decreases almost linearly as dendrimer acetylation increases, as shown in Fig.4.3.9b.

As shown in Fig.4.3.9c, the order parameter η for dendrimers acetylated to different extents remains higher than 0.05, which indicates that 72 bp dsDNA curves around the

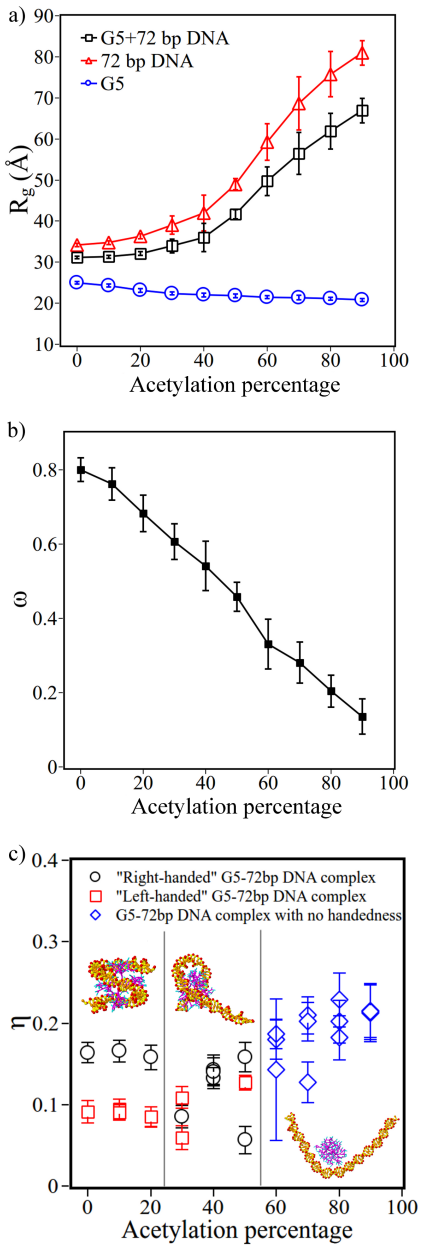

Figure 4.3.9: (a) Radius of gyration of DNA, dendrimers, and complexes vs. acetylation fraction at 10mM salt. (b) Adsorption fractions of DNA phosphate beads onto dendrimer surface and (c) order parameters of DNA in dendrimer-DNA complexes. (The error bars are standard deviations.

dendrimer. The order parameter η increases as acetylation increases from 30% to 90%, but is constant for the less acetylated dendrimers, for which we report two values of the order parameter in Fig.4.3.9c, to be discussed in the next section.

Handedness of DNA-dendrimer complexes. Complexes between DNA and any high generation (>5) dendrimer are unlikely to achieve the global free energy minimum because of strong electrostatic attraction, which makes it difficult for the complexes to explore configurational space and escape local free energy minima. However, for a fully charged G5 dendrimer interacting with a 72 bp DNA, after equilibration at 10 mM salt concentration, we observe that the complex eventually relaxes to either a "left-handed" (Fig.4.3.10a) or a "right-handed" (Fig.4.3.10b) superhelix. Although in these complexes, DNA lengths are only half that in a nucleosome, the complex structures seemed to be the similar to those in a nucleosome. However, the DNA supercoiling in a nucleosome is left-handed, as resolved by Luger et al. [130], although recent research has revealed that the right-handed nucleosome can sometimes exist both *in vitro* and *in vivo* [131]. The mechanism of chirality selection in the nucleosome is still not fully understood, especially since simulations of nucleosome assembly from free DNA and histones remains very challenging. Our CG simulations of dendrimer-DNA complexes suggest that macroions with no chirality, such as dendrimers, are not able to compact DNA into superhelices with a specific handedness, although the wrapping energy might not be completely symmetric, because of the chirality of DNA. In fact, we find that the order parameter η of the left-handed wrapping is 0.09$\pm$0.01, while η for the right-handed complex is 0.16 $\pm$ 0.01 (errorsare standard deviations) for dendrimers acetylated to 20% or less, as shown in Fig.4.3.9c. When the acetylation ratio is high (>50%), the dendrimer adheres poorly to DNA, so that the complexes have no handedness. At intermediate acetylation (between 20% and 50%), the handedness of the

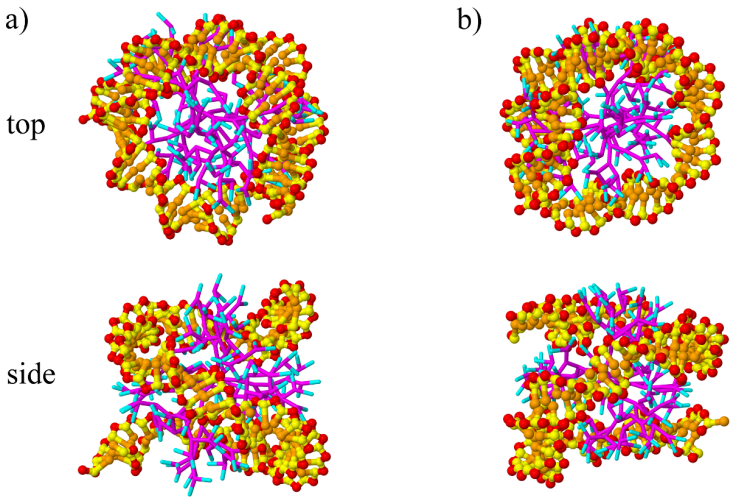

Figure 4.3.10: (a) "Left-handed" G5 dendrimer-72bp DNA complex. (b) "Right-handed" G5 non-acytylated dendrimer-72bp DNA complex at 10.

DNA wrapping can be discerned, but the wrapping is incomplete and fluctuating, leading to a fluctuating order parameter, as indicated by the results in Fig.4.3.9c.

Multiple dendrimers on one DNA. Since no periodic boundary conditions are applied in any of our simulations and so we have no box size limitations, we can study spatially extended structures, so as two dendrimers interacting with a single DNA molecule to assess dendrimer-dendrimer interactions along the DNA. To compare results for dendrimers of different generation, we choose three systems with the same r_{charge}, as shown in Fig.4.3.11. The radius of gyration of a G3 dendrimer is only about 1.6 nm, less than the diameter of dsDNA, which is about 2.2 nm. Therefore, in a G3-DNA complex, the DNA molecules tend to remain linear (see Fig.4.3.11a) because of the repulsion between DNA segments as well as the repulsion between the two G3 dendrimers. In a G4-DNA complex, on the other hand, the DNA can form a zig-zag structure (see Fig.4.3.11b), and remains in the same plane. Since the R_g of the G5 dendrimer is larger than the diameter of DNA, the DNA can cover a larger fraction of the G5 surface than of the G3 surface. Thus, a nucleosome-like

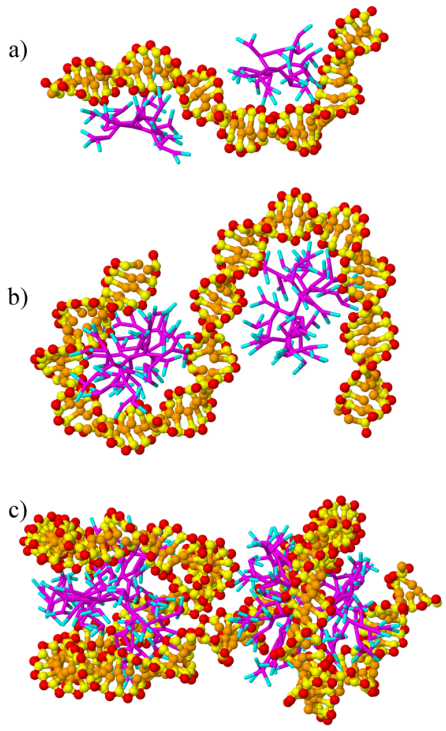

Figure 4.3.11: (a) Two Ge dendrimer on a 36 bp DNA; (b) two G4 dendrimers on a 72bp DNA; and (c) two G5 dendrimers on a 144bp DNA.

complex can form, as shown for the G5 complex in Fig.4.3.11c.

These results show the differences in the local structure of complexes between dendrimers of different generation for the same values of r_{charge}. These results are a starting point to help explain the differences in structure of the much larger complexes formed experimentally when DNA is condensed by dendrimers of different generation at the same r_{charge}, as observed under cryo-TEM [83]. In particular, for large dendrimers, such as G6, large globular aggregates of size 100 nm are observed, while for smaller G4 and G2 dendrimers, a transition occurs to folded bundles and eventually toroids. The structures we see in simulations might be the precursor sub-structures that assemble into the larger experimental structures, with the more linear and planar structures seen for G3 and G4 in Fig.4.3.11a

and b assembling into bundles or toroids, while the globular structures formed by G6 in Fig.4.3.11c might condense into large globular aggregates. Of course, this is speculation at this point, but the results presented here at least point the direction for further work to determine the hierarchy of structures formed by DNA-dendrimer complexes, and their possible analogies to the hierarchies present in DNA-histone structures. A good future step would be to design a coarse-grained model that predicts correctly the two-dendrimer-DNA structures shown in Fig.4.3.11, but is cheap enough to simulate complexes formed by much longer DNA interacting with many dendrimers.

4.4 Conclusions and future work

We combined the 3SPN.1 course-grained (CG) DNA model of de Pablo [35] with the CG dendrimer model of Welch and Muthukumar [107], to define a hybrid CG model that can be used to simulate PAMAM dendrimer-DNA complexes much more rapidly than possible for atomistic simulations, but with resolution of the double helical structure of DNA. Using this CG model, we were able to investigate the effect of dendrimer generation, salt concentration, and dendrimer surface charge density on the structure of the complex. Our simulations confirm that at low salt concentration (10-100 mM) PAMAM dendrimers of various generation are able to condense dsDNA significantly without disrupting the local DNA double helix structure, in agreement with experimental studies [83]. But when the number of charges on the dendrimer are reduced to less than 10 by acetylation, at least for generation G5, the dendrimer no longer induces DNA wrapping, but merely attaches to the DNA. A similar transition from strong to weak DNA attachment has been observed as

the number of charges is reduced to below 15 or 10, for other condensing agents, including charged nanoparticles [91] and polyelectrolytes [80]. The structure of the complex depends strongly on both the charge and size of the dendrimer. A fully charged G5 dendrimer at low salt (10 mM) is wrapped by DNA as in the nucleosome, but with either left- and right-handed DNA superhelices. The simulation method is fast enough to simulate pairs of dendrimers binding to a single long DNA strand, which begins to mimic the smallest scales of chromatin-like structures. The use of accelerated Metropolis schemes using cluster moves, faster simulation methods, faster computers, and/or coarser-grained models tuned to capture the effects reported here, should open the door to simulating larger structures containing multiple dendrimers and longer pieces of DNA. This new CG model can also easily be extended to other interesting polyelectrolytes [80, 82, 91] as well as dendrimers with different surface modifications [129]. Using a coarse-grained protein model with appropriate resolution [30], and a new version of the 3SPN force field recently incorporated into LAMMPS [26], the model could be adapted to simulate the dynamics of protein-DNA binding, including the initial phases of chromatin formation. Since electrostatic interactions in our simulations are estimated using the Debye-Hückel theory, our model has limitations and likely becomes inaccurate at high salt concentrations. Nevertheless, a mesocale model with a resolved DNA double helix, combined with mesoscale models of large condensing agents could be very helpful to understand cationic macroion-DNAinteraction generally, as suggested by our simulations.

CHAPTER V

Summary and future work

149 years ago, Johann Mendel published his work on plant hybridization [132]. But it was not until 1900s that his great work was rediscovered. And the real breakthrough took place in 1953, when Watson and Crick revealed the structure of DNA [133], after Schrödinger published his famous book "What is Life?" in 1944. Of course, this discovery of structure of the molecule which carries hereditary information could only occur after X-ray diffraction technique was ready to investigate this molecule. No wonder why there is a over 40 years debate [53] on how the DNA-binding protein finds its target on DNA through facilitated diffusion after Riggs et al. published "The lac represser-operator interaction III. Kinetic studies" in 1970, since studying protein kinetics not only requires high spatial resolution but also requires high temporal resolution. How protein finds its target and how protein regulates gene expression will finally be resolved after and only after a "3D super high spatial & temporal resolution microscopy" is released. Before that, we have to develop some simple models to explain why protein is so "smart" in finding its target and in packaging DNA into regular chromosome structure. On one hand, we proposed a "high energy barrier-cutoff" model and demonstrated that this simple model is equivalent to "two-state" model [58], which implies that protein might not be that "smart" that the

protein targeting speed does not exceed diffusion limit actually. On the other hand, we were studying a "not that smart but protein-like" molecule/macroion, the PAMAM dendrimer, both experimentally and theoretically. We hope to find out what is the key factor for protein function by comparing the different performances of protein and dendrimer when interacting with DNAs. In this thesis, we have demonstrated that some progress have been made in both directions. The specific shape and charge distribution is vital for protein to function in a proper way.

To confirm this conclusion, it might be interesting to apply DiSCO method to several proteins, using hundreds of point charges to represent the protein, and perform the similar simulations as described in chapter 4. If the simulation results agree with experimental results, for example, if we could obtain the same chromatin structures as obtained by cryo-EM [19], we can verify our coarse-grained model and conclude that electrostatic interaction dominates the binding affinity between the specific proteins and DNA. At this point, it is still too early to include all interactions between DNA and protein, such as hydrogen-bonding, hydrophobic force, in this coarse-grained model, as the point charges do not represent real atoms or residues on protein. Therefore, in the near future, we will focus on using DiSCO method to construct an irregular charged suface to model a charged DNA-binding protein.

APPENDICES

APPENDIX A
Reaction rate derivation

Effective reaction rate

At the target site, the protein finds a local free energy minimum, and so its concentration there will be higher than elsewhere according to the ratio of Boltzmann factors for the target site relative to the average over all sites. That is,

$$c_T = \frac{exp\left(-\frac{E_T}{k_BT}\right)}{\frac{1}{n}\sum_{i=1}^{n} exp\left(-\frac{E_i}{k_BT}\right)} c_\infty = \frac{exp\left(-\frac{E_T}{k_BT}\right)}{exp\left[-\frac{\mu}{k_BT} + \frac{1}{2}\left(\frac{\sigma}{k_BT}\right)^2\right]} c_\infty = c_\infty exp\left[\frac{\mu - E_T}{k_BT} - \frac{1}{2}\left(\frac{\sigma}{k_BT}\right)^2\right] \tag{6.1.1}$$

where c_T and c_∞ are the concentrations of proteins on the target site and the average concentration far away from target site respectively; E_i is the free energy of binding of the protein at site i ($i = 1, 2, 3, ...$) which we take to be Gaussian distributed with average μ and standard deviation σ, $E_i = N(\mu, \sigma)$; E_T is the target site energy; n is the total number of base pairs. Note that $\frac{1}{n}\sum_{i=1}^{n} exp(-E_i/k_BT)$ is the expected value of a log-normal distribution, which is given by $exp[-\mu/k_BT + (\sigma/k_BT)^2/2]$. Thus, we can identify the effective reaction rate constant in the reaction-limited regime as Ω^{eff}:

$$\Omega^{eff} = \Omega exp\left[\frac{\mu - E_T}{k_BT} - \frac{1}{2}\left(\frac{\sigma}{k_BT}\right)^2\right] \tag{6.1.2}$$

Variance of the D_n/D_0

We can estimate the landscape-dependent variance of D_n, the diffusion coefficient averaged over the diffusion distance $n = \sqrt{D/k_{off}}$. Considering n Gaussian distributed energy sites ($E_i = N(0, \sigma)$, $i = 1, 2, ..., n$) with variance s^2:

$$s^2 = \left(\frac{1}{n-1}\right)\left(\sum_{i=1}^{n} E_i^2 - n\overline{E_n}^2\right) \tag{6.1.3}$$

where $\overline{E_n}$ is the energy averaged over these n sites. The diffusion coefficient D_n averaged over these n sites can be approximated by $D_n = D_0 exp(-s^2)$. Note:

$$(n-1)s^2/\sigma^2 = \sum_{i=1}^{n}(E_i/\sigma)^2 - \left(\sqrt{n\overline{E_n}}/\sigma\right)^2 = \chi^2(n) - \chi^2(1) \tag{6.1.4}$$

Since E_i/σ and $\sqrt{n\overline{E_n}}/\sigma$ are both distributed according to a standard normal distribution. χ^2 is the Chi-squared distribution. Therefore, the variance of the D_n/D_0 can be estimated by:

$$Var\left[-ln\left(\frac{D_n}{D_0}\right)\right] = Var(s^2) = \left(\frac{\sigma^2}{n-1}\right)^2 Var\left[\chi^2(n) - \chi^2(1)\right] \approx 2\sigma^4/n \quad (n \gg 1) \tag{6.1.5}$$

So, when the diffusion distance is about 100 ($n \approx 100$, typical value for DBP's), the variance of D_n over this diffusion distance will be significant for σ larger than $3k_BT$.

Reaction rate for two-state model

Assuming $\omega^R_{T,T\pm1}$ is very small and setting c^S and c^R to be the concentrations of the protein in the searching and recognition states respectively, at steady state, from a flux balance of protein entering and leaving the recognition state at the target site T, we have:

$$\omega_T^{SR} c_T^S - \Omega c_T^R - \omega_T^{RS} c_T^R = 0 \tag{6.1.6}$$

where c_T^S and c_T^R are protein concentrations at the target site in the "S" and "R" state, respectively. Eq. 7.1.6 includes the flux from the search to the recognition state, the reaction rate, and the flux back from recognition to the search state, with negligible diffusion in the recognition state. A flux balance into and out of the searching state at the target site, which includes the diffusion rate, and has no contribution from reaction, which only occurs in the recognition state, is:

$$-2D\frac{\partial c^S}{\partial x}\Big|_{x=T} - \omega_T^{SR}c_T^S + \omega_T^{RS}c_T^R = 0 \tag{6.1.7}$$

Using Eq. 7.1.6 and Eq. 7.1.7 to eliminate c_T^R, we obtain at target site $T = 0$:

$$-2D\frac{\partial c^S}{\partial x}\Big|_{x=T} - \left(\frac{\Omega\omega_T^{SR}}{\Omega + \omega_T^{RS}}\right)c_T^S = 0 \tag{6.1.8}$$

Solving Eq. 2.2.1, subject to the boundary conditions Eq. 7.1.8 as well as $c^S(x \to \infty) = c_\infty^S = (k_{on}/k_{off})c_{bulk}$, and note $c_T^S = c^S(x=0)e^{-E_T^S/k_BT}$, the concentration of protein in the searching state along the DNA is:

$$c^S(x) = c_\infty^S - c_\infty^S\left[\frac{1}{1 + 2\sqrt{k_{off}/D}\,(1/\omega_T^{SR} + 1/\Omega\cdot\,\omega_T^{RS}/\omega_T^{SR})\,exp\,(E_T^S/k_BT)}\right]exp\left(x\sqrt{k_{off}/D}\right) \tag{6.1.9}$$

Since at the target site, we have $\omega_T^{RS} = \omega_T^{SR}exp\left[(E_T^R - E_T^S)/k_BT\right]$, where E_T^R and E_T^S are the target energy in "R" and "S" state respectively, the reaction rate ($x = 0$) is:

$$r = \frac{c_\infty^S}{\frac{1}{2\sqrt{k_{off}D}} + \frac{1}{\omega_T^{SR}exp\left(-\frac{E_T^S}{k_BT}\right)} + \frac{1}{\Omega exp\left(-\frac{E_T^R}{k_BT}\right)}} \tag{6.1.10}$$

Brownian dynamics simulation

The positions of T7 RNAP particles are obtained by solving the following stochastic differential equation:

$$x^{(n+1)} = x^{(n)} + \frac{\Delta t}{6\pi\eta r_{protein}} F^{(n)} + \sqrt{\frac{2k_B T \Delta t}{6\pi\eta r_{protein}}}\xi^{(n)} \tag{6.1.11}$$

where $x^{(n)}$ is the position of the protein at n time step; $F^{(n)}$ is the overall force on protein particle at time step n ($F^{(n)} = -dE^{(n)}/dx^{(n)}$); the timestep Δt is taken to be 1ps; $r_{protein}$ is the radius of the protein (2.5 nm), the viscosity η is $0.00089Pa \cdot s$; and ξ is a Gaussian random number with zero mean and unit variance .

APPENDIX B
Combed DNA-PAMAM dendrimer interaction

λ-DNA molecules were stretched out and aligned by high air-water interface tension ($\approx$400 pN). They can be immobilized on hydrophobic polystyrene surface because both the 12 bp overhangs in the two ends and melted base-pairs which are exposed to the surface are hydrophobic. In reaction condition, PAMAM dendrimer is highly charged (112 positive charges per particle), its surface is very hydrophilic. Therefore, the only force that exists in this system is the electrostatic force theoretically. However, the PAMAM dendrimer is a flexible macromolecule, especially compared to the DBPs. It is still possible that the hydrophobic internal core of the dendrimer gets exposed to the polystyrene surface and gets stuck to the hydrophobic surface. So, the experiments have to be carried out very carefully to prevent or at least reduce this surface effect. Based on "charged sphere on string" model [90], the free energy change for a single DNA to wrap around the dendrimer will far exceeds the thermal energy, which means once the dendrimer associates with DNA it is impossible to dissociate. And this has been confirmed by the optical tweezers experiment [13]. For our experiment, the DNA keeps a stretched conformation, part of the DNA sequence may even change to a Z-conformation. Hence, the real interaction between the PAMAM dendrimer and combed DNA cannot be predicted by the analytical model [90] precisely. And only when the dendrimer gets very close to the DNA (usually within 1 nm), it can feel the strong electrostatic force. At single-molecule level, both the DNA and den-

drimer concentration are very low. So, the probability for dendrimer to diffuse closed to DNA within Debye length is small.

Some events that the Brownian motions of dendrimers were suppressed by electrostatic forces exerted from DNA molecule have been observed. The duration of interaction is usually less than 5 seconds. One example is shown in Fig.6.1.1. The mean-square displacement in x and y directions were obtained using Eq.6.1.12 [18]:

$$\begin{aligned} MSD(i\Delta t)_x &= \sum_{j=1}^{N} \frac{(x_{i+j}-x_j)^2}{N-i} \\ MSD(i\Delta t)_y &= \sum_{j=1}^{N} \frac{(y_{i+j}-y_j)^2}{N-i} \end{aligned} \tag{6.1.12}$$

The error bar of MSD vs. time plot was estimated by Eq.6.1.13 [18]:

$$\sigma_i^2 = \frac{(2D_1 i\Delta t)^2(2i^2+1)}{3i(N-i+1)} \tag{6.1.13}$$

Since the events of dendrimer sliding on combed DNA are so rare, we are not able to confirm if dsDNA can suppresse the 3D diffusion of PAMAM dendrimer to 1D diffusion.

Some TRITC labeled PAMAM dendrimer got stuck onto the fluorescent DNA. The PAMAM dendrimers were imaged with a device that many polystyrene strips were deposited on its suface. Due to the strong autofluorescence of the polystyrene strips (see Fig.6.1.2), we were not able to see if the PAMAM dendrimer can slide on hydrophobic polystyrene strips.

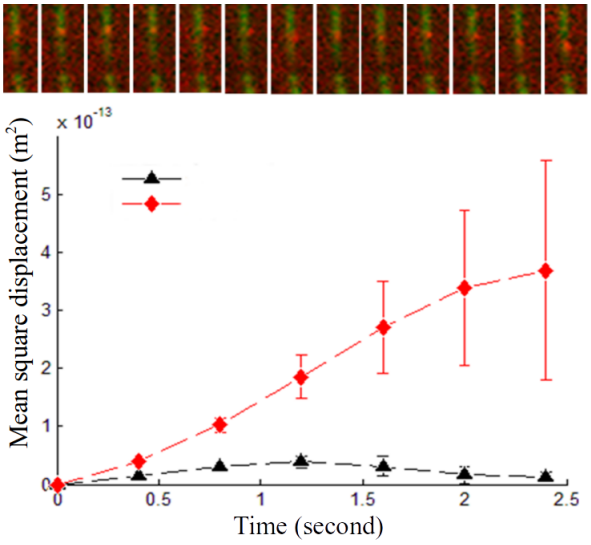

Figure 6.1.1: 1D diffusion of TRTIC stained PAMAM dendrimer (Generation 5) along immobilized DNA on polystyrene-coated surface (top); the mean square displacements (MSD) of the dendrimer vs. time in x (perpendicular to DNA sequence) and y (along DNA) directions (bottom).

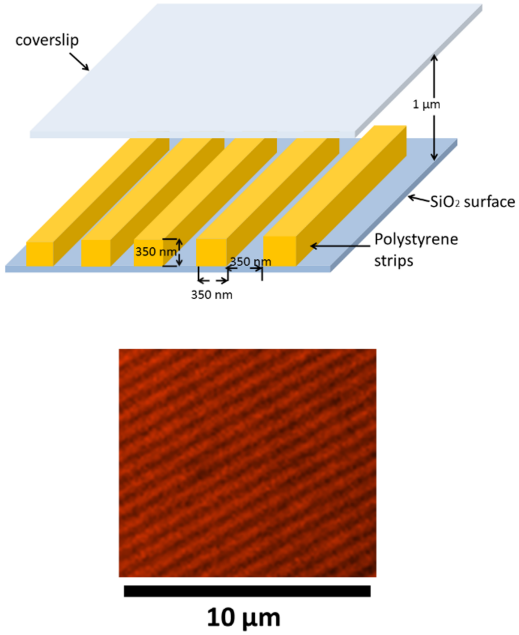

Figure 6.1.2: polystyrene strips were deposited on silicon dioxide surface by lithography method (done by Ashwin Pandy, J. Guo group) (top), the red fluorescent signals (bottom) came from auto fluorescence polystyrene strips.

APPENDIX C
Interaction between nanoparticles and flow stretching DNA

Large rhodamine B labeled & CTAB-protected polystyrene nanoparticles were synthesized in ultrasound bath by vaporizing chloroform in a two-phase mixture (oil phase: chloroform, polystyrene, rhodamine B; water phase, CTAB solution) to perform the single-molecule imaging experiment to study the size and charge effects on DNA-nanoparticle interaction.

The size of this nanoparticle ranges from 200nm to 500nm. The number of charges per particle which is estimated from the zeta potential is about 30,000. And the charge density per area of this polystyrene nanoparticle is closed to fully charged G5 PAMAM dendrimer (zeta potential = 50≈60mV). Since the diameter of this polystyrene particle is much larger than the Kuhn length of ds-DNA (50nm), it is easy for the DNA to bind and wrap around this large particle which has been observed in our experiment (shown in Fig.6.1.3 left).

The strong electrostatic forces between PS particles and DNA make this binding process irreversible. And the unevenly distributed charges on the surface which result in a very rough potential will prevent the sliding of PS particles on DNA. In our experiment, the large PS nanoparticles are either attracted by DNA and surface or washed out by high speed laminar flow. No one-dimensional diffusion of PS particle along DNA has been observed. This result may imply that anionic DNA is not able to suppress the Brownian

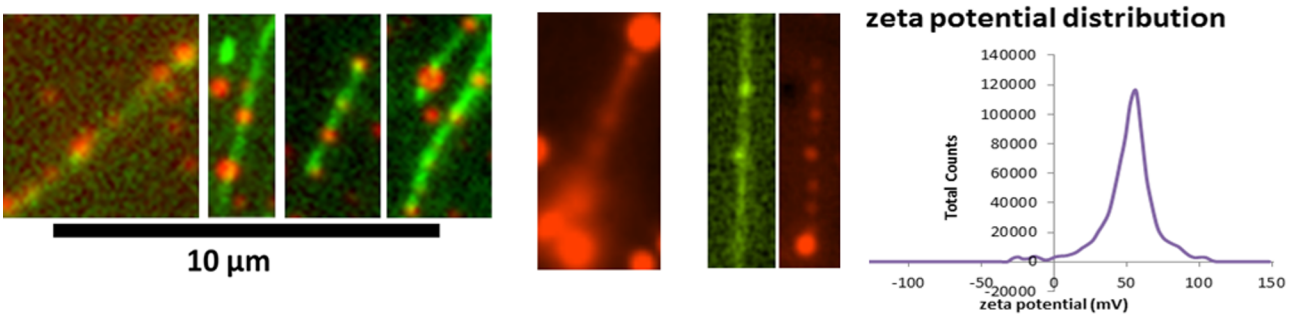

Figure 6.2.3: positively charged polystyrene nanoparticles bind to flow stretched DNA, red signals come from rhodamine B encapsulated by polystyrene nanoparticles, green signals are from the YOYO-1 stained λ-DNA (left); Zeta-potential distribution of the rhodamine B labeled CTAB-PS nanoparticle (right).

motion of an object with dissimilar size.

Bibliography

Bibliography

[1] Y. H. Cai and H. Huang, "Advances in the study of protein-DNA interaction," *Amino Acids*, vol. 43, no. 3, pp. 1141–1146, 2012.

[2] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, no. 3, pp. 561–563, 1970.

[3] M.-f. O. Donohue, B. Revett, and A. P. Ii, "Interaction of the histone (H3-H4) 2 tetramer of the nucleosome with positively supercoiled DNA minicircles : Potential flipping of the protein from a left- to a right-handed superhelical form," *Proc. Natl. Acad. Sci.*, vol. 93, pp. 7588–7593, 1996.

[4] C. Yang, E. Bolotin, T. Jiang, F. M. Sladek, and E. Martinez, "Prevalence of the Initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters," *Gene*, vol. 389, pp. 52–65, 2006.

[5] L. M. Hellman and M. G. Fried, "Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions," *Nat. Protoc.*, vol. 2, no. 8, pp. 1849–1861, 2007.

[6] K. D. Connaghan-Jones, A. D. Moody, and D. L. Bain, "Quantitative DNase footprint titration: a tool for analyzing the energetics of protein-DNA interactions," *Nat. Protoc.*, vol. 3, no. 5, pp. 900–914, 2008.

[7] J. Jahanmir, B. Haggar, and J. Hayes, "The scanning probe microscope," *Scanning microscopy*, vol. 6, pp. 625–660, September 1992.

[8] X. D. Hoa, a. G. Kirk, and M. Tabrizian, "Towards integrated and sensitive surface plasmon resonance biosensors: A review of recent progress," *Biosens. Bioelectron.*, vol. 23, no. 2, pp. 151–160, 2007.

[9] J. H. Kim, V. R. Dukkipati, S. W. Pang, and R. G. Larson, "Stretching and immobilization of DNA for studies of proteinâDNA interactions at the single-molecule level," *Nanoscale Res. Lett.*, vol. 2, no. 4, pp. 185–201, 2007.

[10] N. Walter, C. Huang, A. Manzo, and M. Sobhy, "Do-it-yourself guide: how to use the modern single-molecule toolkit," *Nat. Methods*, vol. 5, no. 6, pp. 475–489, 2008.

[11] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, "Multicolor super-resolution imaging with photo-switchable fluorescent probes," *Science*, vol. 317, no. 5845, pp. 1749–1753, 2007.

[12] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[13] F. Ritort, S. Mihardja, S. Smith, and C. Bustamante, "Condensation transition in DNA-polyaminoamide dendrimer fibers studied using optical tweezers," *Phys. Rev. Lett.*, vol. 96, p. 118301, 2006.

[14] T. Lionnet, S. Joubaud, R. Lavery, D. Bensimon, and V. Croquette, "Wringing Out DNA," *Phys. Rev. Lett.*, vol. 96, no. 17, pp. 1–4, 2006.

[15] A. Bensimon, A. Simon, A. Chiffaudel, and V. Croquette, "Alignment and sensitive detection of DNA by a moving interface," *Science*, vol. 265, pp. 2096–2098, 1994.

[16] T. Fazio, M.-L. Visnapuu, S. Wind, and E. C. Greene, "DNA curtains and nanoscale curtain rods: high-throughput tools for single molecule imaging," *Langmuir*, vol. 24, no. 18, pp. 10524–10531, 2006.

[17] A. Tafvizi, F. Huang, A. R. Fersht, L. a. Mirny, and A. M. van Oijen, "A single-molecule characterization of p53 search on DNA," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 2, pp. 563–568, 2011.

[18] J. H. Kim and R. G. Larson, "Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules," *Nucleic Acids Res.*, vol. 35, no. 11, pp. 3848–3858, 2007.

[19] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-m. Xu, and P. Zhu, "Cryo-EM study of the chromatin fiber tetranucleosomal units," *Science*, vol. 344, pp. 376–380, 2014.

[20] R. Elder and A. Jayaraman, "Coarse-grained simulation studies of effects of polycation architecture on structure of the polycation and polycation-polyanion complexes," *Macromolecules*, vol. 45, pp. 8083–8096, 2012.

[21] R. Esfand and D. a. Tomalia, "Poly(amidoamine) (PAMAM) dendrimers: from biomimicry to drug delivery and biomedical applications.," *DDT*, vol. 6, no. 8, pp. 427–436, 2001.

[22] a. G. Cherstvy, "Critical polyelectrolyte adsorption under confinement: planar slit, cylindrical pore, and spherical cavity," *Biopolymers*, vol. 97, no. 5, pp. 311–317, 2012.

[23] R. Winkler and A. Cherstvy, "Strong and weak polyelectrolyte adsorption onto oppositely charged curved surfaces," *Adv. Polym. Sci.*, pp. 1–56, 2013.

[24] S. de Carvalho, R. Metzler, and A. Cherstvy, "Critical adsorption of polyelectrolytes onto charged Janus nanospheres," *Phys. Chem. Chem. Phys.*, vol. 16, no. 29, pp. 15539–15550, 2014.

[25] A. D. MacKerell and L. Nilsson, "Molecular dynamics simulations of nucleic acid-protein complexes," *Curr. Opin. Struct. Biol.*, vol. 18, no. 2, pp. 194–199, 2008.

[26] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo, "An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization," *J. Chem. Phys.*, vol. 139, no. 14, p. 144903, 2013.

[27] D. M. Hinckley, J. P. Lequieu, and J. J. De Pablo, "Coarse-grained modeling of DNA oligomer hybridization: Length, sequence, and salt effects," *J. Chem. Phys.*, vol. 141, no. 3, 2014.

[28] G. Freeman, J. Lequieu, and D. Hinckley, "DNA shape dominates sequence affinity in nucleosome formation," *Phys. Rev. Lett.*, vol. 113, no. 16, p. 168101, 2014.

[29] T. Terakawa and S. Takada, "RESPAC: Method to determine partial charges in coarse-grained protein model and its application to DNA-binding proteins," *J. Chem. Theory. Comput.*, vol. 10, no. 2, pp. 711–721, 2014.

[30] Q. Zhang, D. a. Beard, and T. Schlick, "Constructing irregular surfaces to enclose macromolecular complexes for mesoscale modeling using the discrete surface charge optimization (DISCO) algorithm," *J. Comput. Chem.*, vol. 24, no. 16, pp. 2063–2074, 2003.

[31] C. M. Wood, J. M. Nicholson, S. J. Lambert, L. Chantalat, C. D. Reynolds, and J. P. Baldwin, "High-resolution structure of the native histone octamer.," *Acta Crystallogr. F-Struct. Biol. Cryst. Commun.*, vol. 61, pp. 541–545, 2005.

[32] G. Cheetham and D. Jeruzalmi *Nature*.

[33] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. a. Baker, "PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations," *Nucleic Acids Res.*, vol. 35, no. SUPPL.2, pp. 522–525, 2007.

[34] N. a. Baker, D. Sept, S. Joseph, M. J. Holst, and J. a. McCammon, "Electrostatics of nanosystems: application to microtubules and the ribosome," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 18, pp. 10037–10041, 2001.

[35] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, "A mesoscale model of DNA and its renaturation," *Biophys. J.*, vol. 96, no. 5, pp. 1675–1690, 2009.

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

[37] P. H. von Hippel and O. G. Berg, "Facilitated target location in biological systems," *J. Biol. Chem.*, vol. 264, no. 2, pp. 675–678, 1989.

[38] B. Van Den Broek, M. A. Lomholt, S. J. Kalisch, R. Metzler, and G. J. L. Wuite, "How DNA coiling enhances target localization," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, no. 41, pp. 15738–15742, 2008.

[39] A. Jeltsch, C. Wenz, F. Stahl, and A. Pingoud, "Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme," *EMBO J.*, vol. 15, no. 18, pp. 5104–5111, 1996.

[40] D. M. Gowers, G. G. Wilson, and S. E. Halford, "Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 44, pp. 15883–15888, 2005.

[41] 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," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 15, pp. 5752–5757, 2006.

[42] J. D. Schonhoft and J. T. Stivers, "Timing facilitated site transfer of an enzyme on DNA," *Nat. Chem. Biol.*, vol. 8, no. 2, pp. 205–210, 2012.

[43] Y. Wang, R. Austin, and E. Cox, "Single molecule measurements of repressor Protein 1D Diffusion on DNA," *Phys. Rev. Lett.*, vol. 97, p. 048302, 2006.

[44] P. Hammar, P. Leroy, and A. Mahmutovic, "The lac repressor displays facilitated Diffusion in Living Cells," *Science*, vol. 336, pp. 1595–1598, 2012.

[45] J. Elf and G. Li, "Probing transcription factor dynamics at the single-molecule level in a living cell," *Science*, vol. 316, pp. 1191–1194, 2007.

[46] M. A. Lomholt, B. van den Broek, S.-M. J. Kalisch, G. J. L. Wuite, and R. Metzler, "Facilitated diffusion with DNA coiling," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 20, pp. 8204–8208, 2009.

[47] C. A. Brackley, M. E. Cates, and D. Marenduzzo, "Facilitated Diffusion on Mobile DNA: Configurational Traps and Sequence Heterogeneity," *Phys. Rev. Lett.*, vol. 109, p. 168103, 2012.

[48] M. Bauer and R. Metzler, "In vivo facilitated diffusion model," *PLOS ONE*, vol. 8, no. 1, p. e53956, 2013.

[49] S. E. Halford and J. F. Marko, "How do site-specific DNA-binding proteins find their targets," *Nucleic Acids Res.*, vol. 32, no. 10, pp. 3040–3052, 2004.

[50] A. G. Cherstvy, A. B. Kolomeisky, and A. A. Kornyshev, "Protein-DNA interactions: reaching and recognizing the targets," *J. Phys. Chem. B*, vol. 112, no. 15, pp. 4741–4750, 2008.

[51] A.-M. Florescu and M. Joyeux, "Description of nonspecific DNA-protein interaction and facilitated diffusion with a dynamical model," *J. Chem. Phys.*, vol. 130, p. 015103, 2009.

[52] A.-M. Florescu and M. Joyeux, "Comparison of kinetic and dynamical models of DNA-protein interaction and facilitated diffusion," *J. Phys. Chem. A*, vol. 114, no. 36, pp. 9662–9672, 2010.

[53] S. E. Halford, "An end to 40 years of mistakes in DNA-protein association kinetics," *Biochem. Soc. Trans.*, vol. 37, pp. 343–348, 2009.

[54] A. Riggs and S. Bourgeois, "The lac represser-operator interaction III. Kinetic studies," *J. Mol. Biol.*, vol. 53, pp. 401–417, 1970.

[55] 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 Res.*, vol. 36, no. 12, pp. 4118–4127, 2008.

[56] O. G. Berg and P. H. von Hippel, "Selection of DNA binding sites by regulatory proteins 1: statistical-mechanical theory and application to operators and promoters," *J. Mol. Biol.*, vol. 193, pp. 723–743, 1987.

[57] D. Fields, Y. He, A. Al-Uzri, and G. Stormo, "Quantitative specificity of the Mnt repressor," *J. Mol. Biol.*, vol. 271, pp. 178–194, 1997.

[58] M. Slutsky and L. A. Mirny, "Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential," *Biophys. J.*, vol. 87, no. 6, pp. 4021–4035, 2004.

[59] S. Denisov and W. Horsthemke, "Mean first-passage time for an overdamped particle in a disordered force field," *Phys. Rev. E*, vol. 62, no. 3, pp. 3311–3321, 2000.

[60] M. Sheinman, O. Bénichou, R. Voituriez, and Y. Kafri, "First-passage time distribution for a random walker on a random forcing energy landscape," *J. Stat. Mech. Theory Exp.*, p. P09005, 2010.

[61] J. Hirth and J. Lothe, *Theory of Dislocations*. McGraw-Hill, New York, 1968.

[62] J. Mathé, H. Visram, V. Viasnoff, Y. Rabin, and A. Meller, "Nanopore unzipping of individual DNA hairpin molecules," *Biophys. J.*, vol. 87, no. 5, pp. 3205–3212, 2004.

[63] B. Jun and D. Weaver, "One-dimensional potential barrier model of protein folding with intermediates," *J. Chem. Phys.*, vol. 116, no. 1, pp. 418–426, 2002.

[64] R. Zwanzig, "Diffusion in a rough potential," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 85, pp. 2029–2030, 1988.

[65] 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," *J. Phys. A: Math. Theor.*, vol. 42, p. 434013, 2009.

[66] I. Minoura, E. Katayama, K. Sekimoto, and E. Muto, "One-dimensional Brownian motion of charged nanoparticles along microtubules: a model system for weak binding interactions," *Biophys. J.*, vol. 98, no. 8, pp. 1589–1597, 2010.

[67] M. Barbi, C. Place, V. Popkov, and M. Salerno, "A model of sequence-dependent protein diffusion along DNA," *J. Biol. Phys.*, vol. 30, no. 3, pp. 203–226, 2004.

[68] M. Barbi, C. Place, V. Popkov, and M. Salerno, "Base-sequence-dependent sliding of proteins on DNA," *Phys. Rev. E*, vol. 70, no. 4, pp. 1–6, 2004.

[69] O. Bénichou, Y. Kafri, M. Sheinman, and R. Voituriez, "Searching Fast for a Target on DNA without Falling to Traps," *Phys. Rev. Lett.*, vol. 103, p. 138102, 2009.

[70] L. Hu, A. Grosberg, and R. Bruinsma, "Are DNA transcription factor proteins Maxwellian Demons," *Biophys. J.*, vol. 95, no. 3, pp. 1151–1156, 2008.

[71] D. A. Gorokhov and G. Blatter, "Diffusion and creep of a particle in a random potential," *Phys. Rev. B*, vol. 58, no. 1, pp. 213–217, 1998.

[72] R. Festa and E. d' Agliano, "Diffusion coefficient for a Brownian particle in a periodic field of force: I. Large friction limit," *Phys. A. Stat. and Theor. Phys.*, vol. 90, no. 2, pp. 229–244, 1978.

[73] C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, and R. Kaptein, "Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes," *Science*, vol. 305, pp. 386–389, 2004.

[74] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, "A new algorithm for Monte Carlo simulation of Ising spin systems," *J. Comp. Phys.*, vol. 18, pp. 10–18, 1975.

[75] J. M. Schurr, "The one-dimensional diffusion coefficient of proteins absorbed on DNA hydrodynamic considerations," *Biophys. Chemist.*, vol. 9, pp. 413–414, 1975.

[76] J. J. Dunn and F. W. Studier, "Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements," *J. Mol. Biol.*, vol. 166, no. 4, pp. 477–535, 1983.

[77] Y. Deng, J. Glimm, and Y. Wang, "Prediction of protein binding to DNA in the presence of water-mediated hydrogen bonds," *J. Mol. Model.*, vol. 5, pp. 125–133, 1999.

[78] M.-L. Orberg, K. Schillén, and T. Nylander, "Dynamic light scattering and fluorescence study of the interaction between double-stranded DNA and poly(amido amine) dendrimers," *Biomacromolecules*, vol. 8, no. 5, pp. 1557–1563, 2007.

[79] E. Froehlich, J. S. Mandeville, C. M. Weinert, L. Kreplak, and H. a. Tajmir-Riahi, "Bundling and aggregation of DNA by cationic dendrimers," *Biomacromolecules*, vol. 12, no. 2, pp. 511–517, 2011.

[80] T. Akitaya, A. Seno, T. Nakai, N. Hazemoto, S. Murata, and K. Yoshikawa, "Weak interaction induces an ON/OFF switch, whereas strong interaction causes gradual change: folding transition of a long duplex DNA chain by poly-L-lysine," *Biomacromolecules*, vol. 8, no. 1, pp. 273–278, 2007.

[81] I. Y. Perevyazko, M. Bauer, G. M. Pavlov, S. Hoeppener, S. Schubert, D. Fischer, and U. S. Schubert, "Polyelectrolyte complexes of DNA and linear PEI: formation, composition and properties," *Langmuir*, vol. 28, no. 46, pp. 16167–16176, 2012.

[82] K. Schille and T. Nylander, "DNA Compaction by cationic surfactant in solution and at polystyrene particle solution interfaces: a dynamic light scattering study," *Phys. Chem. Chem. Phys.*, vol. 6, pp. 1603–1607, 2004.

[83] M.-L. Ainalem, A. M. Carnerup, J. Janiak, V. Alfredsson, T. Nylander, and K. Schillén, "Condensing DNA with poly(amido amine) dendrimers of different generations: means of controlling aggregate morphology," *Soft Matter*, vol. 5, no. 11, pp. 2310–2320, 2009.

[84] M.-L. Ainalem and T. Nylander, "DNA condensation using cationic dendrimers-morphology and supramolecular structure of formed aggregates," *Soft Matter*, vol. 7, no. 10, pp. 4577–4594, 2011.

[85] R. Dootz, A. C. Toma, and T. Pfohl, "PAMAM6 dendrimers and DNA: pH dependent "beads-on-a-string" behavior revealed by small angle X-ray scattering," *Soft Matter*, vol. 7, no. 18, pp. 8343–8351, 2011.

[86] X. Wang, Y. He, J. Wu, C. Gao, and Y. Xu, "Synthesis and evaluation of phenylalanine-modified hyperbranched poly (amido amine) s as promising gene carriers," *Biomacromolecules*, vol. 11, pp. 245–251, 2009.

[87] A. M. Carnerup, M.-L. Ainalem, V. Alfredsson, and T. Nylander, "Watching DNA condensation induced by poly(amido amine) dendrimers with time-resolved cryo-TEM," *Langmuir*, vol. 25, no. 21, pp. 12466–12470, 2009.

[88] K. Fant, E. E. Rner, and A. Jenkins, "Effects of PEGylation and acetylation of PAMAM dendrimers on DNA binding, cytotoxicity and in vitro transfection efficiency," *Mol. Pharm.*, vol. 7, no. 5, pp. 1734–1746, 2010.

[89] H. Boroudjerdi and R. R. Netz, "Interactions between polyelectrolyte-macroion complexes," *Europhys. Lett.)*, vol. 64, pp. 413–419, 2003.

[90] K. Qamhieh, T. Nylander, and M.-L. Ainalem, "Analytical model study of dendrimer/DNA complexes," *Biomacromolecules*, vol. 10, no. 7, pp. 1720–1726, 2009.

[91] J. G. Railsback, A. Singh, R. C. Pearce, T. E. McKnight, R. Collazo, Z. Sitar, Y. G. Yingling, and A. V. Melechko, "Weakly charged cationic nanoparticles induce DNA bending and strand separation," *Adv. Materials*, vol. 24, no. 31, pp. 4261–4265, 2012.

[92] K. Fant, E. K. Esbjörner, P. Lincoln, and B. Nordén, "DNA condensation by PAMAM dendrimers: self-assembly characteristics and effect on transcription," *Biochemistry*, vol. 47, no. 6, pp. 1732–1740, 2008.

[93] S. K. Choi, P. Leroueil, M. Li, and A. Desai, "Specificity and negative cooperativity in dendrimer-oxime drug complexation," *Macromolecules*, vol. 44, pp. 4026–4029, 2011.

[94] D. Cakara, J. Kleimann, and M. Borkovec, "Microscopic protonation equilibria of poly(amidoamine) dendrimers from macroscopic titrations," *Macromolecules*, vol. 36, no. 11, pp. 4201–4207, 2003.

[95] M. S. Diallo, S. Christie, P. Swaminathan, L. Balogh, X. Shi, W. Um, C. Papelis, W. a. Goddard, and J. H. Johnson, "Dendritic chelating agents. 1. Cu(II) binding to ethylene diamine core poly(amidoamine) dendrimers in aqueous solutions," *Langmuir*, vol. 20, no. 7, pp. 2640–2651, 2004.

[96] M. Vorlícková, "Conformational transitions of alternating purine-pyrimidine DNAs in perchlorate ethanol solutions," *Biophys. J.*, vol. 69, pp. 2033–2043, 1995.

[97] J. Kypr and M. Vorlícková, "Circular dichroism spectroscopy reveals invariant conformation of guanine runs in DNA.," *Biopolymers*, vol. 67, pp. 275–277, 2002.

[98] J. H. Kim, W.-X. Shi, and R. G. Larson, "Methods of stretching DNA molecules using flow fields," *Langmuir*, vol. 23, pp. 755–764, 2007.

[99] S. K. Choi, T. P. Thomas, P. Leroueil, A. Kotlyar, A. F. L. Van Der Spek, and J. R. Baker, "Specific and cooperative interactions between Oximes and PAMAM Dendrimers As Demonstrated by 1H NMR Study," *J. Phys. Chem. B*, vol. 116, pp. 10387–10397, 2012.

[100] I. Majoros, B. Keszler, and S. Woehler, "Acetylation of poly (amidoamine) dendrimers," *Macromolecules*, vol. 36, pp. 5526–5529, 2003.

[101] S. Yu, M.-h. Li, S. K. Choi, J. R. Baker, and R. G. Larson, "DNA Ccondensation by partially acetylated poly(amido amine) dendrimers: effects of dendrimer charge density on complex formation," *Molecules (Basel, Switzerland)*, pp. 10707–10720, 2013.

[102] K. Wada, H. Arima, T. Tsutsumi, Y. Chihara, K. Hattori, F. Hirayama, and K. Uekama, "Improvement of gene delivery mediated by mannosylated dendrimer/alpha-cyclodextrin conjugates," *J. Control. Release*, vol. 104, no. 2, pp. 397–413, 2005.

[103] H. L. Crampton and E. E. Simanek, "Dendrimers as drug delivery vehicles: non-covalent interactions of bioactive compounds with dendrimers," *Polym. Int.*, vol. 496, pp. 489–496, 2007.

[104] A. M. Carnerup, M.-L. Ainalem, V. Alfredsson, and T. Nylander, "Condensation of DNA using poly(amido amine) dendrimers: effect of salt concentration on aggregate morphology," *Soft Matter*, vol. 7, no. 2, pp. 760–768, 2011.

[105] B. Nandy and P. K. Maiti, "DNA compaction by a dendrimer," *J. Phys. Chem. B.*, vol. 115, no. 2, pp. 217–230, 2011.

[106] M. Mills, B. Orr, M. M. Banaszak Holl, and I. Andricioaei, "Microscopic basis for the mesoscopic extensibility of dendrimer-compacted DNA," *Biophys. J.*, vol. 98, no. 5, pp. 834–842, 2010.

[107] P. Welch and M. Muthukumar, "Dendrimer-polyelectrolyte complexation: A model guest-host system," *Macromolecules*, vol. 2, no. 1, pp. 6159–6167, 2000.

[108] S. V. Lyulin, A. A. Darinskii, and A. V. Lyulin, "Computer simulation of complexes of dendrimers with linear polyelectrolytes," *Macromolecules*, vol. 38, no. 9, pp. 3990–3998, 2005.

[109] S. V. Larin, A. A. Darinskii, A. V. Lyulin, and S. V. Lyulin, "Linker formation in an overcharged complex of two dendrimers and linear polyelectrolyte.," *J. Phys. Chem. B.*, vol. 114, no. 8, pp. 2910–2919, 2010.

[110] T. Lewis, G. Pandav, A. Omar, and V. Ganesan, "Complexation between weakly basic dendrimers and linear polyelectrolytes: effects of grafts, chain stiffness, and pOH," *Soft Matter*, vol. 9, no. 29, pp. 6955–6969, 2013.

[111] Q. Cao, C. Zuo, Y. Ma, L. Li, and Z. Zhang, "Interaction of double-stranded DNA with a nanosphere: a coarse-grained molecular dynamics simulation study," *Soft Matter*, vol. 7, no. 2, pp. 506–514, 2011.

[112] A. Savelyev and G. Papoian, "Molecular renormalization group coarse-graining of polymer chains: Application to double-stranded DNA," *Biophys. J.*, vol. 96, no. 10, pp. 4044–4052, 2009.

[113] T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, "A coarse grain model for DNA," *J. Chem. Phys.*, vol. 126, no. 8, p. 084901, 2007.

[114] V. Ortiz and J. de Pablo, "Molecular origins of DNA flexibility: sequence effects on conformational and mechanical properties," *Phys. Rev. Lett.*, vol. 106, no. 23, pp. 3–6, 2011.

[115] I. V. Dobrovolskaia and G. Arya, "Dynamics of forced nucleosome unraveling and role of nonuniform histone-DNA interactions," *Biophys. J.*, vol. 103, no. 5, pp. 989–998, 2012.

[116] T. Schlick and O. Perisić, "Mesoscale simulations of two nucleosome-repeat length oligonucleosomes," *Phys. Chem. Chem. Phys.*, vol. 11, no. 45, pp. 10729–10737, 2009.

[117] G. Arya and T. Schlick, "A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments.," *J. Phys. Chem. A.*, vol. 113, no. 16, pp. 4045–4059, 2009.

[118] A. Stogryn, "Equations for calculating the dielectric constant of saline water," *IEEE Trans. Microw. Theory Tech.*, vol. 19, no. 5, pp. 733–736, 1970.

[119] D. C. Rau and V. a. Parsegian, "Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces," *Biophys. J.*, vol. 61, no. 1, pp. 246–259, 1992.

[120] H. Drew and R. Wing, "Structure of a B-DNA dodecamer: conformation and dynamics," *Proc. Nati. Acad. Sci.*, vol. 78, no. 4, pp. 2179–2183, 1981.

[121] S. Kumar, D. Bouzida, R. Swendsen, P. Kollman, and J. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *J. Comput. Chem.*, vol. 13, no. 8, pp. 1011–1021, 1992.

[122] Y. Liu, V. S. Bryantsev, M. S. Diallo, and W. A. Goddard, "PAMAM dendrimers undergo pH responsive conformational changes without swelling," *J. Am. Chem. Soc.*, vol. 131, no. 8, pp. 2798–2799, 2009.

[123] P. K. Maiti, T. Cagin, G. Wang, and W. A. Goddard, "Structure of PAMAM dendrimers: generations 1 through 11," *Macromolecules*, vol. 37, no. 16, pp. 6236–6254, 2004.

[124] P. K. Maiti, T. Cagin, S.-T. Lin, and W. A. Goddard, "Effect of solvent and pH on the structure of PAMAM dendrimers," *Macromolecules*, vol. 38, no. 3, pp. 979–991, 2005.

[125] H. Lee, J. R. Baker, and R. G. Larson, "Molecular dynamics studies of the size, shape, and internal structure of 0% and 90% acetylated fifth-generation polyamidoamine dendrimers in water and methanol," *J Phys. Chem. B.*, vol. 110, no. 9, pp. 4014–4019, 2006.

[126] T. Prosa and B. Bauer, "A SAXS study of the internal structure of dendritic polymer systems," *J. Polym. Sci. Part B*, vol. 35, pp. 2913–2924, 1997.

[127] P. E. Theodorakis, H.-P. Hsu, W. Paul, and K. Binder, "Computer simulation of bottle-brush polymers with flexible backbone: good solvent versus theta solvent conditions.," *J. Chem. Phys.*, vol. 135, no. 16, p. 164903, 2011.

[128] T. Sakaue, K. Yoshikawa, S. Yoshimura, and K. Takeyasu, "Histone core slips along DNA and prefers positioning at the chain end," *Phys. Rev. Lett.*, vol. 87, no. 7, p. 078105, 2001.

[129] T. Zhou and S. B. Chen, "Monte Carlo simulations of dendrimer-polymer conjugates," *Macromolecules*, vol. 38, no. 20, pp. 8554–8561, 2005.

[130] K. Luger, A. Mader, and R. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, pp. 251–260, 1997.

[131] C. Lavelle, "Left is right, right is wrong," *EMBO Rep.*, vol. 10, no. 11, pp. 1185–1186, 2009.

[132] G. Mendel, "Experiments on plant hybridization," *Verhandlungen des naturforschenden Vereines in Brünn*.

[133] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid," *Nature*.