QUANTITATIVE ALL-ATOM AND COARSE-GRAINED MOLECULAR DYNAMICS SIMULATION STUDIES OF DNA

BY

CHRISTOPHER MICHAEL MAFFEO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Klaus J. Schulten, Chair
Associate Professor Aleksei Aksimentiev, Director of Research
Professor Taekjip Ha
Professor John D. Stack

Abstract

The remarkable molecule that encodes genetic information for all life on earth—DNA—is a polymer with unusual physical properties. The mechanical and electrostatic properties of DNA are utilized extensively by cells in the replication, regular maintenance, and expression of their genetic material. This can be illustrated by considering the journey of a typical gene regulating protein, the *lac* repressor, which recognizes a particular gene and prevents its expression. First, the large electrostatic charge density of DNA provides an energetic track that guides the repressor's search for its target binding site. Next, as the protein moves along the DNA, it attempts to deform the DNA. The repressor is only able to form an active complex with DNA that has the right sequence-dependent flexibility. Finally, the repressor is believed to form a very small DNA loop that prevents the gene from being expressed. The stability of the loop can be expected to depend sensitively on the global flexibility of DNA. Thus, the key to understanding some of the most important cellular processes lies in understanding the physical properties of DNA. Single-molecule experiments allow direct observation of the behavior of individual DNA molecules, but act on length and timescales that are often too large and fast to observe underlying DNA and DNA–protein dynamics. Acting on length and timescales that complement single-molecule experiments, molecular dynamics simulations provide a high-resolution glimpse into the mechanics of a biomolecular world. Here, several simulation studies are presented, each of which quantified one or more properties of DNA. Specifically, the repulsive forces between parallel duplex DNA molecules were measured; the short-ranged, attractive end-to-end stacking energy was obtained; a single-stranded DNA model was developed that reproduced experimental measurements of its extension upon applied force; and finally the nature of single-stranded DNA binding to a single-stranded DNA binding protein was investigated. These works represent important steps towards larger simulations of more biologically complete DNA–protein systems.

For Sidney.

Acknowledgments

This project would not have been possible without the support of many people. Many thanks to my adviser, Aleksei Aksimentiev, who has guided me through the years. Thanks to my committee members, Prof. Taekjip Ha, Prof. Klaus Schulten and Prof. John Stack for their guidance and support. Thanks to the Center for the Physics of Living Cells for providing fruitful collaborations and financial and academic support over the years. Thanks to my collaborators in the groups of Prof. Ralf Seidel, Prof. Gero Wedemann, Prof. Taekjip Ha, and Prof. Yann Chemla. For his rigor and devotion to teaching, I thank my undergraduate adviser, Francesc Roig. To my sweetheart, thank you for your patience, unwavering love, understanding and encouragement. And finally, to my family and friends, thank you all for your kindness, love and support.

This work was supported by the National Institutes of Health (R01-HG003713, PHS5 P41-RR05969, R01-HG007406, R01-GM065367, and P41-GM104601), the National Science Foundation (PHY-0822613 and DMR-0955959), the Petroleum Research Fund (48352-G6), and the TeraGrid (MCA05S028).

Table of Contents

Chapter 1

Background

1.1 DNA is more than just a sequence

After water and oxygen, DNA is, very likely, the most famous molecule of life known to humankind. This is not surprising, as we all know that an eye-catching, double-helical molecule of DNA carries instructions to manufacture and assemble all the components of a living organism. The wealth of information encoded in a DNA molecule often overshadows its unique physical properties, which facilitate its role as the carrier of genomic information [2–4]. For example, the hybridization of a DNA strand with a complementary strand, which occurs through interlocked hydrogen bonds, provides a nearby redundant source of information in case of spontaneous mutation. More impressively, the proteins that recognize DNA mismatches are believed to detect these genetic errors by sensing the flexibility of the DNA molecule.

Despite a very large number of theoretical and experimental studies, the behavior of DNA in very many contexts remains highly debated. With the advent of massively parallel supercomputers it has become possible to characterize DNA properties directly, through all-atom molecular dynamics (MD) simulations. Because the all-atom MD method explicitly describes the trajectory of every atom in the system with femtosecond resolution, it has the potential to give unparalleled insight into an experimental system. The primary use of the MD method is to suggest a physically plausible explanation or justification of an experimental measurement by animating an equivalent system *in silico*. Equipped with a physically correct description of interatomic interactions and adequate computational power, the MD method should be able to predict the physical behavior of any biological system. Despite ever-increasing availability of massive parallel computing platforms, making quantitative predictions using MD remains challenging, in part due to imperfections of the inter-atom interaction models.

Before we proceed, let's review the basic chemical structure of DNA, Fig. 1.1. A molecule of DNA is a polymer made up of many DNA nucleotides linearly arranged into a polymer chain. Single-stranded DNA (ssDNA) is made of one such chain, whereas in double-stranded DNA (dsDNA), two ssDNA molecules are arranged into a DNA double helix through non-covalent interactions. The basic unit of DNA structure—a DNA nucleotide—has three major groups: backbone, sugar and base. The backbone is negatively charged under physiological conditions and has a direction ($5'$-to-$3'$) determined by the order of the atoms forming the backbone. The sugar group links the backbone to the base. The chemical difference between DNA and RNA is the presence of an extra hydroxyl (OH) moiety in the sugar group, which strongly alters the properties of

the molecule. The DNA base carries genetic information and typically comes in the one following four types: adenine (A), cytosine (C), guanine (G) and thymine (T). The complementary hydrogen bond paring of A with T and G with C governs the nucleotide sequence-specific assembly of two single strands into a double helix. However, chemical modification of DNA bases are common. The most abundant variation is the addition of a methyl group to cytosine, transforming it into a methylated cytosine—a carrier of epigenetic information (derived from the history of an organism). Modifications are possible, including hydroxymethylation and other derivatives [5]. Under physiological conditions, each DNA nucleotide carries a charge of one electron. A single DNA strand is much more flexible than a double helix. The following discussion implicitly assumes that DNA is in an aqueous environment at pH 7.0.

The rest of this Chapter includes a brief historical review of the field of DNA simulation and a review some key DNA properties. The second Chapter discusses, in mostly general terms, the simulation methods used throughout this thesis. Finally, the remaining Chapters present simulation studies that quantitatively investigated the properties of DNA.

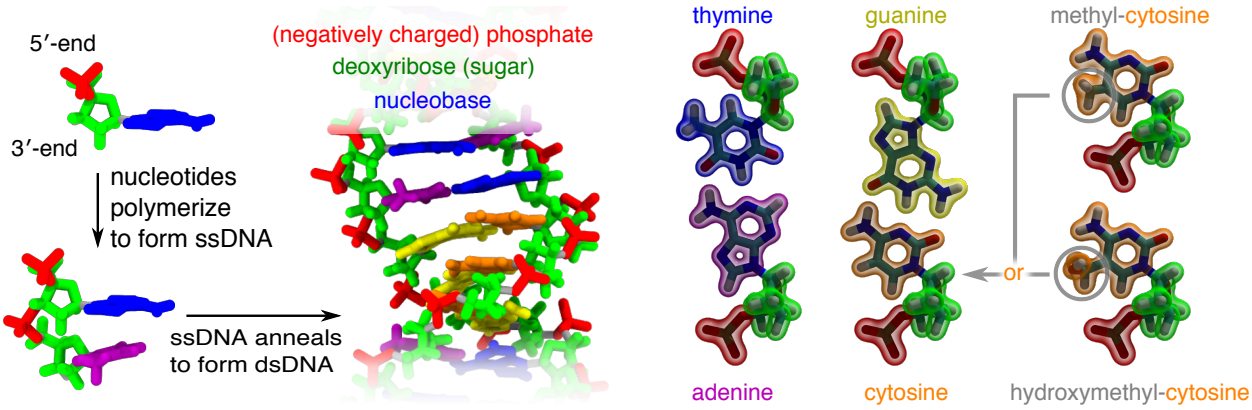

Figure 1.1: Chemical model of DNA. DNA is a polymer composed of nucleotides, each having a negatively charged phosphate, a deoxyribose sugar ring and one of the four nucleobases: adenine, thymine, guanine, cytosine. Two single DNA strands form a double helix held together through non-covalent interactions. In addition to the four types of DNA nucleotides, chemical modification of DNA occurs frequently and includes methylation and hydroxymethylation.

1.2 Historical perspective of DNA modeling and simulation

The year 1952 witnessed the publication of the first solid evidence that DNA is the genetic material [6]. Just one year later in 1953, the basepaired double-helical structure of DNA was proposed [7]. In that same year, Hermann Staudinger received a Nobel Prize for his work in 1922 demonstrating that polymers are composed of long chains of covalently bonded atoms, and Paul Flory—later a Nobel laureate for his work

in theoretical polymer chemistry—published his seminal book "Principles of Polymer Chemistry". Thus, a theoretical framework for discussing polymers was available for the study of DNA, including the widely employed freely-jointed and worm-like chain models [8–10], see Sec. 1.3. No time was wasted; in that same year, the worm-like chain (WLC) model was used to analyze the angular dependence of the intensity of scattered light to suggest a surprisingly accurate measure of the flexibility of DNA [11, 12]. Other early theoretical studies of DNA focused on the thermodynamic transition of denaturation using a variety of approaches [13–16]. One perhaps iconic approach was an adaptation of the one-dimensional Ising model (originally developed to describe magnetization), whereby the likelihood of a given basepair's unpairing depends on the paired status of its nearest neighbors in a DNA helix [13].

By 1975, it was known that DNA was a rather rigid polymer, yet it could form compact structures like chromatin [17], sparking a debate about kinked versus smoothly bent DNA. In the earliest atomic-modeling studies that employed computation, models of kinked and bent DNA were produced by performing least-squares searches of toy potentials linking rigid groups of atoms [18, 19]. Soon after, Levitt borrowed atomic interaction potentials from protein and RNA refinement studies to relax the full set of atomic coordinates of a DNA molecule that was bent and twisted by varying amounts [20]. Although both solvent and electrostatics were absent in the calculations and the calculations probed only the local energy minima, the study correctly determined the 10.5 basepairs-per-turn pitch of a DNA helix in solution, departing from the 10 basepairs-per-turn pitch observed in the solid-state Watson-Crick structure.

The first room-temperature computations involving DNA were performed by Clementi and Corongiu in 1979 and 1980 [21, 22]. These Monte Carlo (MC) simulations, performed using a model optimized to reproduce energies of *ab initio* calculations, revealed the structure of water molecules around fixed DNA helices and nucleotides. In the MC method, atomic coordinates were propagated according to Boltzmann statistics, so equilibrium, but not dynamical, properties of the system could be studied. In the beginning of 1983, Levitt, borrowing again the methods developed for the study of proteins, reported an MD simulation that showed asymmetric bending and twisting motions of duplex DNA in the absence of solvent [23]. This was followed only a few months later by a similar MD report by the Karplus group [24]. Finally in 1985, the Kollman group performed the first MD simulation of a DNA fragment in electrolyte solution [25].

Beveridge *et al.* comprehensively reviewed the following decade of MD studies of DNA [26, 27]. At that time, simulations were rapidly approaching the nanosecond timescale [28]. Nevertheless, most simulations from that era described DNA using implicit solvent and truncated electrostatics. In 1995, it was noted that the outcome of a simulation could qualitatively depend on the method used to describe the dielectric environment of a charged polymer [29]. The Kollman group (1995) demonstrated that simulations of DNA

using the particle mesh Ewald (PME) method, which efficiently calculates long-range electrostatic interactions in Fourier space, are more accurate than the simulations performed using truncated electrostatics [30]. Although accurate representation of electrostatics is essential for modeling of a highly charged molecule such as DNA, the community was at first slow to adopt the particle mesh Ewald method due to the high computational cost [31]. At that time, several "second-generation" all-atom force fields with explicit solvent were released, including AMBER-94 and CHARMM22 [32]. Around the same time, National Science Foundation supercomputers became more accessible to researchers and the major MD codes AMBER and CHARMM were parallelized.

Prompted by these methodological advances, longer and more accurate studies of DNA structure and dynamics were performed. Of particular note, Cheatham and Kollman (1996) observed a spontaneous transition of DNA from A-form to a more stable B-form during the first multi-nanosecond simulation of explicitly solvated DNA, indicating good accuracy for the AMBER force field [33]. Shortly thereafter Young, Ravishanker and Beveridge reported the first 5-nanosecond trajectory of B-DNA [34], which revealed substantial fluctuation of the DNA structure, in agreement with with x-ray crystallography and NMR. The latter study was one among several that found that MD simulations employing the CHARMM or AMBER force field could reproduce the crystal and NMR structures [31]—an important test of the force fields. The sequence-specificity of DNA structures began to be investigated including comparisons of twist, roll, and tilt obtained for different basepair stacks [35]. More dramatically, the intrinsic curvature of A-tract DNA was observed [34] (see Sec. 1.3 for details). Back in 1986, it had been recognized from crystal structures that water may considerably affect the conformation of DNA (A-form vs. B-form) [36]. Feig and Pettitt (1998) reviewed molecular dynamics simulations that investigated the structure and properties of water surrounding DNA [37]. Another question of outstanding importance was the location and dynamics of ions, which were believed to bind DNA tightly [38]. A number of studies investigated whether and where the ions would bind to DNA [39] and began to characterize the ion atmosphere [40–42]. Around the same time, the generalized Born method [43] was introduced to implicit solvent simulations with AMBER [44] and CHARMM [45] parameters for more accurate estimation of the electrostatic solvation energy.

From 1995 to 2000, the era of quantitative MD simulations of DNA began to emerge, starting with free energy perturbation simulations of ligand binding [31, 35]. The structural and intramolecular character of DNA was investigated through free energy simulations of base pairing, stacking, and DNA stretching [31]. In particular, MacKerell and Lee (1999) used umbrella sampling MD simulations and atomic force microscopy to study stretch-induced melting of short DNA fragments [46]. Near-quantitative agreement was obtained between simulation and experiment, indicating good overall performance of the CHARMM22 force field. It

must be cautioned that the AFM results were not in complete agreement with the seminal study of stretching long (48.5 kbp) λ-phage DNA using an optical trap [47], perhaps because different DNA constructs were employed [46]. Nevertheless, the simulation represented the first, to our knowledge, quantitative comparison of mechanical DNA properties observed in experiment and in simulation.

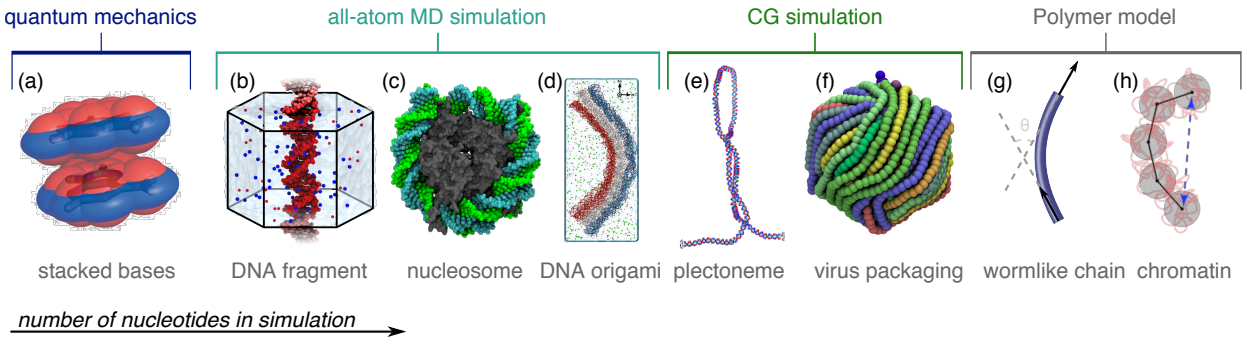

Figure 1.2: Nucleic acid systems that represent the range of scales amenable to various computational and theoretical methods. From left to right, generally increasing in number of nucleotides: (a) Base-stacking interactions can be studied through quantum mechanics calculations, and reveal asymmetric van der Waals radii [48]. The all-atom MD method with explicit or implicit solvent provides a balance between computational speed and descriptive detail. This method can be used to study systems ranging from a few nucleotides to (b) several turns of DNA [49] to (c) hundreds nucleotides as in the nucleosome [50]. (d) A large multi-layer, curved DNA origami construct described by MD simulation with explicit solvent [51]. Coarse-grained simulations can describe DNA with a wide range of detail, from near atomic (e.g. to describe (e) a plectoneme [52]) to many nucleotides per site (e.g. to describe (f) the packaging of a virus [53]). The two dominant polymer models of DNA are (g) the wormlike chain (WLC) model [10], and the freely jointed chain (FJC) model [8, 9]. The WLC model describes DNA under tension better than FJC, but FJC is convenient for very coarse descriptions of DNA, (h) such as at the level of chromatin [54]. Panels adapted with permission from (a) Ref. 55, (b) Ref. 49, (d) Ref. 51, (e) Ref. 52, (f) Ref. 53, and (h) Ref. 54.

In the new millennium, there have been a few general reviews of DNA simulations [56–58]. However, the field has grown such an extent that there have been very many special-topic reviews on a wide range of subjects. For example, the base stacking interactions were reviewed from a quantum mechanics (QM) perspective [55]. Another recent review that included QM and MD descriptions of DNA focused on the backbone rather than bases [48]. Work continued toward understanding the relationship between DNA and the solvent [59, 60], however, it was only recently found that the standard parameterization of cation–phosphate interaction had considerable problems [61]. Advances in computing permitted the mechanical properties of DNA to be examined in more detail [62–65]. Finally, there were many investigations of DNA association with other molecules, including small molecules such as anti-cancer drugs [66, 67], sequence-specific DNA binding proteins [66, 68], non-specific DNA binding proteins such as the nucleosome [69] and synthetic structures such as carbon nanotubes [70] and silicon nitride nanopores [71].

In recent years, MD simulations of DNA have moved toward large systems and long durations. The

first microsecond simulation of B-DNA was performed by the Orozco group in 2007 [72], which revealed the limitations of the AMBER-99 force field and a new parameterization of the all-atom model that addressed the problem. Meanwhile, a number of impressive developments have happened in the related field of RNA systems simulations. An all-atom model of one of the smallest viruses—the satellite tobacco mosaic virus—was simulated for ~ 50 ns, which included 60 copies of the coat protein, a 1063-nucleotide single-stranded RNA molecule and the electrolyte solution, about one million atoms in total [73]. Even larger RNA systems have since been simulated. Atomic simulations of the ribosome, an enormous protein synthesis factory predominantly composed of RNA, were recently reviewed by Sanbomatsu [74]. The simulations examined the conformational changes in the ribosome, the effects of point mutations and quantified the kinetics and free energy barriers of conformational transitions using a 3 million atom model and the aggregate simulation time of ~ 2 μs.

Large-scale MD simulations of DNA systems have only recently matched the scale of the largest RNA simulations, perhaps because of the lack of atomic-scale structures of large DNA assemblies. Matching the setup of DNA array experiments of Rau and Parsegian [75], Yoo and Aksimentiev simulated the structure, dynamics, ionic atmosphere and intermolecular forces of a DNA array [61]. The outcome of these simulations revealed the limitations of the standard ion-DNA interactions and a method to fine-tune the parameters to achieve quantitative agreement with experiments. One of the current frontiers of DNA simulations are DNA origami and related self-assembled nanostructures [76–78]. The first atomic-resolution simulation of DNA origami was reported very recently [51], revealing the *situ* structure of several DNA origami designs as well as their local and global mechanical properties.

The above discussion has focused predominantly on fully-atomic simulation of DNA. By employing a less detailed, "coarse-grained" (CG) model the timescale accessible to simulations of DNA can be significantly extended [79–81]. In the past five years, many CG models of DNA have been developed to reproduce various properties of double-stranded DNA [82–92]. Most available CG models employ a few interaction sites to represent each nucleotide, but CG models can span an enormous range of scales, from the near-atomic [91] to the meso and macroscopic [53]. An excellent review of available CG DNA models can be found in the recent article from the de Pablo group [82].

Parameters describing the interactions of coarse-grained models are usually obtained by fitting against experimental data, such as the nearest neighbor DNA melting parameters [93], or by reproducing structural parameters from atomic resolution simulations. In analogy to the atomic force fields, which are optimized against a mix of finer quantum calculations and experimental data, it is our belief that the most accurate general purpose CG models of DNA will be obtained through a combination of all-atom simulations and

single-molecule experiment. A recent application of such an approach is a CG model of single-stranded thymine homopolymers specifically developed to match both the all-atom simulations and single-molecule fluorescence resonance energy transfer measurements [94].

Finally, the coarsest biologically relevant description of DNA ignores its physical form altogether, representing each nucleotide with one of four letters, A-C-G-T. In the field of genomics, computers play an instrumental role in organizing, sorting, search and comparing the genomes of different species and individuals. Since this field lies largely outside of our expertise, we merely mention computational genomics as an important and exciting area of computational modeling the biological function of DNA.

1.3 Duplex DNA mechanics

Mechanical deformation of double-stranded DNA (dsDNA) occurs during many gene regulation processes. Although the studies presented in this thesis did not directly probe the mechanical properties of DNA, we review the mechanical properties of duplex DNA because of their important biological implications.

At length scales greatly exceeding the helical pitch of dsDNA, polymer models can provide an adequate description of dsDNA's equilibrium conformations. The wormlike chain (WLC) model mathematically represents a polymer of length l by a continuous set of unit-length tangent vectors $\mathbf{t}(s)$ parameterized by the distance along the contour $s \in (0, l)$. The polymer is assumed to behave like a cylindrically symmetric beam with a mechanical rigidity described by an elastic modulus. Thus local bending of the polymer requires an energy du proportional to $|\frac{\partial \mathbf{t}}{\partial s}|^2$. In thermodynamic equilibrium $\frac{\partial \mathbf{t}(s)}{\partial s}$ should sample from the Boltzmann distribution and it can be shown by integrating over all possible polymer configurations that the expected value for $\langle \mathbf{t}(s) \cdot \mathbf{t}(0) \rangle = e^{-s/L_p}$ [10]. Thus, the tangent vector of a polymer is expected to be correlated for a characteristic contour length of L_p, the persistence length. For DNA, a range of values for L_p have been obtained, centered around 45–50 nm by fitting the WLC model to a wide variety of experimental measurements [95]. Finally, it is worth noting that the WLC model has been extended to incorporate elastic stretching and twisting [96–98].

The local bending of dsDNA is highly dependent on the DNA sequence. For example, dsDNA is more likely to bend at the AT step. The network of hydrogen bonds formed by water molecules on DNA surface varies in accord with the DNA sequence. For a poly(dA)·poly(dT) fragment, a water "spine" is formed in the minor groove of the dsDNA, bridging the N3 and O2 atoms in neighboring basepairs. When a sequence contains 4–6 dA nucleotides repeated with helical periodicity, the DNA curves on average toward the minor groove by $\sim$10–20° per turn. These so-called A-tracts have been recently reviewed in outstanding detail [99].

The molecular mechanisms that underlie A-tract curvature are a challenge to determine because there are many possible causes, including solvent interactions, ion-binding, anomalous stacking or base-pairing. It is tempting to attempt to describe A-tract curvature using a nearest neighbor model, such as the model that accurately describes the free energy of DNA hybridization as a sum of energies associated with pairs of stacked basepairs [93]. However, it was shown using MD simulation that the conformation of a CG basepair stack depended on the sequence context in the DNA [100]. Thus, it appears that the sequence-dependent attributes of DNA mechanics cannot be fully described by a nearest neighbor model, at least if we assume that the MD result was not due to artifacts in the force field.

1.3.1 Twisting dsDNA

A DNA molecule can be bent, stretched along its helical axis, or twisted about this axis. Measurements of the extension of a double-stranded DNA molecule under low applied force (< 10 pN) probes the mechanical bending of DNA [101]. At higher force (but still below the overstretching transition), the same measurements can characterize elastic stretching along the helical axis [101]. Such experiments are usually performed using optical traps which allow the DNA to relax torsionally. In biological cells, torsional stress in DNA was shown to regulate gene expression [102, 103].

Usually, the term flexibility refers to a polymer's ability to bend, which is quantified by its persistence length. For a generic dsDNA molecule, the persistence length is $\sim$45-50 nm [64, 95] or about 20 times its diameter (2.5 nm), which makes DNA fairly stiff for a natural polymer. In direct analogy to the persistence length, one can define a torsional persistence length for dsDNA that represents the characteristic length scale along which a DNA molecule's twist is correlated. This length is similar to the persistence length, but varies considerably in the literature from 36–75 nm [95, 104, 105].

Analysis of the crystal structures suggests that the intrinsic twist of a DNA helix depends on its nucleotide sequence and can range from 31° for an AG step to 40° for a TA step [106, 107]. Here we follow the usual convention that the first letter is on the 5′-end of one strand, and we note that some sequences are "palindromes" (e.g. AA with TT). The standard deviation of the twist is similar for different basepair steps, ranging from 3.9° for an AA step to 6.5° for a CA step, suggesting a mild dependence of the torsional persistence length on the DNA sequence. The sequence-dependent average twist and twist moduli were obtained from MD simulations of DNA fragments [104] using the AMBER force field [108] without the parmbsc0 refinement. The results of the MD simulations were in overall agreement with the results of the crystal structure survey [107] with some discrepancies, primarily for GG steps. CG steps were found to be the most conformationally flexible, and AT steps the least. Overall, the study demonstrated significant

sequence-specific coupling between various coordinates used to describe basepair steps.

A more recent simulation study employed the newer AMBER-99 force field and a much larger, systematic set of sequences to examine the flexibility of specific base-steps [109]. Despite reporting on overall flexibility, which included contributions from both twist and bending, the results of the study were consistent with previous observations, assessing the TA, CG and CA (but not AT, GC, and AC) steps as the most flexible. The authors emphasized a correlation between the flexibility of a step and the absence of water molecules in the minor groove that bridge the base pairs forming the step.

Simulations of longer DNA molecules have also been performed. Mazur has studied many DNA properties through simulations employing an all-atom MD method that permits the integration timestep to be increased from 1-2 to 10 fs by treating groups of DNA atoms rigidly [110]. That method enabled long-timescale ($>$ 160 ns) simulations of short ($<$ 2 turns) DNA fragments that revealed a torsional persistence length of 120 nm for AMBER-98 [111]. In contrast however, Noy and Golestanian [112] more recently found the torsional persistence length to be 80 nm from simulations of relatively long (3-5 turns) random sequence DNA molecules using the AMBER-99 force field with the parmbsc0 correction. The disagreement resulted in a debate between the two groups [113, 114], which may partially be attributed to the rather different simulation protocols employed.

When DNA is twisted enough, linear elastic models fail to describe its properties. This was demonstrated in MD simulations where a three-turn DNA molecule was twisted by -0.22 to $+0.39$ turns [115]. Initially, the DNA had uniform twist, but by the end of a 10-ns simulation it was slightly undertwisted or extremely overtwised; the DNA conformation included a mixture of B-DNA and locally denatured DNA accommodating the overall twist. The MD simulations also indicated that counterions were able to access the minor groove more easily in overwound than underwound DNA.

1.3.2 DNA looping

Because cells can suppress expression of genes by looping DNA, the behavior of DNA in small loops has long been of interest. Under thermal fluctuations or applied force, DNA bends away from its equilibrium curvature. For moderate bending, the wormlike chain model (WLC) describes this bending very well with a 45–50 nm persistence length. Early simulations were not long enough to allow accurate determination of the persistence length of DNA. However, the studies from Mazur [111] and from Noy and Golestanian [112] reported persistence lengths of 65-85 nm for AMBER-98 and 43–51 nm for AMBER-99bsc0, respectively.

The likelihood of DNA having complementary single-stranded overhangs to form a loop [116], or to cyclize, depends on both the bending and twisting properties of DNA. As a function of DNA length, this

likelihood has a peak around 400 bp, where the DNA is neither too long (so that the ends cannot find each other), nor too short (so that the DNA is unable to bend into a loop). This assay has been used to unambiguously distinguish intrinsic sequence-dependent curvature from isotropic flexibility [116]. For DNA molecules shorter than 500 bp, the likelihood of DNA cyclization was found to fluctuate periodically with respect to the length of the DNA fragment [64].

Under high curvature, the WLC model must break down, however, there is much disagreement with respect to the critical curvature where the model loses its validity. It was experimentally observed that very short DNA fragments ($\sim$100 bp) could form circles more readily than predicted by the elastic models of DNA [105], suggesting strong bending or kinking. Similar results were observed from atomic force microscope images [2] and from the interference of x-ray scattering off gold nanoparticles tethered to either end of a short DNA molecule [117]. Other experiments rigorously defended the validity of the WLC model for short DNA [64, 65, 118, 119]. The controversy continues as DNA was seen to be highly flexible in recent melting experiments of curved DNA [120] and in single-molecule experiments where the looping of 100-bp DNA fragments was monitored by fluorescence [4]. It is interesting to note that most experimental studies that reported sharp bending of DNA employed either divalent electrolytes or high concentrations of monovalent electrolyte [2, 4, 120]. In the nucleus, counterions with valence > 2 strongly screen electrostatic interactions, so studies of DNA bending at elevated concentrations of monovalent electrolyte may still be biologically relevant.

1.4 Duplex DNA electrostatics

Besides the base pairing and stacking interactions, the most dominant attribute of a DNA molecule is its extreme negative charge; each nucleotide carries a surplus electron charge. In biology, a DNA molecule is surrounded by counterions that screen its electrostatic field. For small electrolytes, the Poisson-Boltzmann (PB) equation provides an approximate description of electrostatics.

The PB equation is easily derived. First, the density ρ_i of ion species i is assumed to be Boltzmann distributed with respect to the charge q_i times the average electrostatic potential $\phi(\mathbf{r})$ in the system, $\rho_i(\mathbf{r}) = q_i\rho_{0,i}\mathrm{e}^{-q_i\phi(\mathbf{r})/k_\mathrm{B}T}$, where $\rho_{0,i}$ is the bulk concentration of ion species i and $k_\mathrm{B}T$ is the thermal energy. The average electrostatic potential is obtained from the average charge density $\rho(\mathbf{r})$ obtained from the sum of the fixed charge density $\rho_f(\mathbf{r})$ and the average mobile charge densities using Poisson's equation of classical electrostatics, $\nabla^2\phi(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon\epsilon_0}$, where ϵ is the dielectric constant of the medium and ϵ_0 is the permittivity of

free space. By combining these, the PB equation is obtained,

$$\nabla^2\phi(\mathbf{r}) = -\frac{1}{\epsilon\epsilon_0}\left(\rho_f(\mathbf{r}) + \sum_i q_i\rho_{0,i}\mathrm{e}^{-q_i\phi(\mathbf{r})/k_\mathrm{B}T}\right). \tag{1.1}$$

The PB formalism presents a nonlinear equation that is difficult to solve analytically [121]. For potentials with $|q_i\phi| \ll k_\mathrm{B}T$, one may linearize the equation by expanding the exponentials so that

$$\nabla^2\phi(\mathbf{r}) - \left(\sum_i \frac{\rho_{0,i}q_i^2}{\epsilon\epsilon_0 k_\mathrm{B}T}\right)\phi(\mathbf{r}) = -\frac{1}{\epsilon\epsilon_0}\left(\rho_f(\mathbf{r}) - \sum_i q_i\rho_{0,i}\right). \tag{1.2}$$

For the typical case of an electrolyte solution with cation and anions both of valence z and bulk concentration ρ_0, this expression can be simplified $\nabla^2\phi(\mathbf{r}) + \frac{2\rho_0 z^2e^2}{\epsilon\epsilon_0 k_\mathrm{B}T}\phi(\mathbf{r}) = -\frac{1}{\epsilon\epsilon_0}\rho_f(\mathbf{r})$. For a fixed point particle with charge q, the above equation has the solution of the Yukawa potential $\phi_q(r) = \frac{q}{4\pi\epsilon\epsilon_0 r}\mathrm{e}^{-r/\lambda_\mathrm{D}}$, where the Debye length $\lambda_\mathrm{D} = \sqrt{\frac{k_\mathrm{B}T\epsilon\epsilon_0}{2z^2e^2\rho_0}}$ is the characteristic electrostatic screening length of the solution [121]. In monovalent electrolyte near physiological concentration, the Debye length is about 1 nm. This is considerably longer than the linear charge density of a DNA molecule, which indicates that the validity of linearizing the PB equation is dubious when applied to DNA.

Chapter 2

Methodology

2.1 All-atom molecular dynamics simulation

The all-atom molecular dynamics (MD) method provides the most detailed and complete microscopic description of biomolecules out of all currently available methods—theoretical, computational and experimental. This expressive detail can give unparalleled insight into the physical mechanisms underlying biomolecular behavior. The historically increasing availability of massive parallel computing platforms has stimulated a shift in all-atom simulations towards long-duration simulations and ensembles of simulations, providing improved statistics that permit quantitative observations. Nevertheless, such observations remain challenging to obtain as thermal noise is very pronounced in atomic-scale systems.

2.1.1 General formulation of the MD method

$$U(\mathbf{r}^N) = U_{\text{bonded}}(\mathbf{r}^N) + U_{\text{nonbonded}}(\mathbf{r}^N) \tag{2.1}$$

In MD simulations, molecules are treated as collections of point particles that each represent an atom. Time is discretized and advances in increments referred to as the "timestep". At a given timestep, all of the instantaneous positions and velocities of the atoms are known. From the positions of the particles, the MD software calculates the force on each atom. With this information, Newton's equation ($\mathbf{F} = m\mathbf{a}$) is integrated, advancing the system by one timestep.

The force on each atom is determined from the full atomic coordinates using a force field, which is a set of equations and parameters that describe the system's Hamiltonian. The most popular force fields for MD simulation describe biomolecules as collections of atoms connected by bonded potentials applied to bonds (two-body interactions), angles (three-body interactions) and dihedral angles (four-body interactions). Harmonic potentials are generally applied to the bonds, angles and improper dihedral angles, whereas a periodic potential is applied to each dihedral angle.

Atoms further apart than two or three bonds (or in different molecules) additionally interact through nonbonded Coulomb and van der Waals (Lennard-Jones) potentials. Each atom is associated with a Lennard-Jones radius and well-depth usually optimized to reproduce properties of liquid systems for model compounds. Chemical groups have a well-defined (integer) total charge, but the way this charge is distributed among the comprising atoms depends on the force field.

Intramolecular and intermolecular nonbonded forces dominate the tertiary structure of a molecule and molecular binding, respectively. Thus, accurate description of the nonbonded forces is essential. The full treatment of non-bonded interactions requires computation of N^2 pair interaction, where N is the number of atoms in a system, which does not scale well to large systems. The number of computations can be reduced to order N by only considering interactions between atoms within some cutoff distance, usually 8–12 Å. For many years this approach was widely practiced, but it was shown roughly 20 years ago by Cheatham *et al.* that B-form DNA conformations were stable on a ten-nanosecond timescale only when long-range electrostatic interactions were considered [30]. Rather than calculating the pair interaction of each particle in real space, the particle mesh Ewald (PME) method was used to efficiently compute both short- and long-range electrostatic interactions the crystal system (requiring periodic boundary conditions). The PME method is currently standard for the treatment of electrostatics in all-atom MD simulations. Optimized versions of the Ewald method [123] permit highly parallelized MD simulations to be efficiently performed (order $N \log N$).

2.1.2 Atomic force fields

Presently, several biomolecular force fields exist. The force fields can be categorized into types based on whether all the atoms are explicitly treated or not. All-atom force fields, which include CHARMM [124], AMBER [108, 125], and OPLS-AA [126, 127], treat all atoms explicitly. In the united-atom force fields (*e.g.* GROMOS [128, 129]), some nonpolar hydrogens are neglected. Additionally, force fields that represent atomic polarization are under development [130, 131], but have not been applied widely and were not used here. For the in-depth review of various force fields, interested readers are referred to a comprehensive review by Mackerell [132].

Although there exist several all-atom force fields, the CHARMM [133] and AMBER [108] force fields are the two most popular choices for MD simulations of DNA, and these are the force fields that have been employed in the studies described herein. More often than not, simulations today greatly exceed the duration and complexity of the simulations that were originally used to develop and validate the force fields. Therefore, the force fields have received frequent updates to keep up with the opportunities offered by modern supercomputing systems [58, 134–138]. Currently, the best practice for MD simulations of DNA includes the application of the AMBER ff99bsc0 [136] or CHARMM36 [137] force field with the recent NBFIX corrections for ion–ion and ion–phosphate interactions from Yoo and Aksimentiev [61]. Simulations of ssDNA might be best performed using the CHARMM36 force field, as the AMBER parameterization was found to enforce the helical conformation of dsDNA on single DNA strands [139].

2.1.3 Thermodynamic ensemble

Except where specified, all MD simulations reported in this thesis were performed using the MD program NAMD [140]. In most MD software, including NAMD, it is not possible to simulate a system with a changing number of particles. Unless a thermostat is employed, the simulation will be performed in the NVE ensemble—that is, with a constant number of particles, volume and energy. If a thermostat is employed, a barostat may optionally be used to adjust the volume by small amounts to maintain a given value of the pressure. Hence, the thermodynamic ensemble will either be NVT or NPT, having a constant temperature and a constant volume or pressure.

The Langevin thermostat, employed in the studies described in this thesis, is commonly applied to control temperature in an MD simulation. With the Langevin thermostat applied, each thermostatted particle is subject to a drag force to remove excess kinetic energy from the system and a random force that injects kinetic energy into the system. The drag coefficient and magnitude of the random force are coupled such that the target temperature is obtained on average. All NPT simulations presented in this these employed the Langevin Nosé-Hoover barostat.

2.2 Coarse-grained molecular dynamics simulation

Atomic simulation can provide the most detailed description of large biomolecules, such as DNA and proteins, out of any available method. However, that detail comes at a large computational cost that often prohibits observation of biologically significant dynamics and events. By reducing the number of atoms present in a simulation system, one can increase the timestep and perform fewer calculations each timestep. A model in which the atomic coordinates are replaced with a smaller set of collective coordinates is generally referred to as a coarse-grained (CG) model. The CG beads, which each represent multiple atoms, still interact through pair potentials that must be parametrized. Parametrization can be classified into two approaches: "top-down", whereby relevant experimental data and intuition guide hand-tuning of parameters affecting ascribed potential function; and "bottom-up", whereby potentials are optimized against a more detailed description such as all-atom MD. The CG studies presented herein employed a "bottom-up" approach to parametrization using the iterative Boltzmann inversion procedure, followed by refinement against experimental results.

2.2.1 Iterative Boltzmann Inversion

Under the iterative Boltzmann inversion (IBI) procedure, one or more all-atom simulation trajectories are mapped into a coarse-grained representation. A set of "target" bonded and non-bonded distributions is

extracted from the CG-mapped all-atom trajectories. These distributions are transformed into CG potentials via Boltzmann inversion, e.g. $u(x) = -k_{\mathrm{B}}T\ln(\rho_0(x))$, where x is some degree of freedom, and $\rho_0(x)$ is the distribution of x. If x is an isolated degree of freedom, $\rho_{\mathrm{CG}}(x)$ obtained in a CG simulation employing $u(x)$ would exactly equal the target distribution $\rho(x)$ at all values of x. Typically, $\rho_{\mathrm{CG}}(x)$ will be too small (or too large) at a given value of x, indicating that the potential $u(x)$ should be decreased (or increased) at x. Thus, IBI prescribes a scheme for updating a potential by $\Delta u(x) = -k_{\mathrm{B}}T\ln\frac{\rho_{\mathrm{CG}}(x)}{\rho_0(x)}$ after a CG simulation is performed and ρ_{CG} calculated. This process is repeated until CG potentials are obtained causing the CG distribution to converge to the target distribution.

2.3 Free energy methods

Perhaps the greatest disadvantage of the MD method is that simulations are costly and are currently limited to the microsecond time scale—a duration insufficient to observe statistically significant numbers of most biologically relevant processes. Very often, a researcher is interested in the free energy difference between conformational states of a system as well as the free energy landscape that the system must traverse to transition between states. Provided a collective coordinate x (one calculated from the underlying coordinates of the system) that describes when the system is in these conformational states, the free energy along the coordinate is the potential of the mean force (PMF) $W(x)$. By definition,

$$W(x) = W(x^\star) - k_B T \log\left[\frac{\langle\rho(x)\rangle}{\langle\rho(x^\star)\rangle}\right], \tag{2.2}$$

$\langle\rho(x)\rangle$ is the average distribution function and $x^\star$ and $W(x^\star)$ are arbitrary constants [141]. The PMF can be calculated from brute-force all-atom simulations simply by observing the fraction of time x dwells at a particular value, and building a histogram to estimate $\langle\rho(x)\rangle$. In practice, such simulations do not efficiently sample x and are, therefore, too computationally demanding to enjoy regular use. Fortunately a host of techniques have been developed for the purpose of calculating the PMF. Interested readers are directed to recent comprehensive reviews on this subject [142, 143].

One of the most established methods for obtaining the PMF is the umbrella sampling method, which enforces uniform sampling along one or more collective coordinates. This is achieved by harmonically restraining x about regularly distributed values in an ensemble of equilibrium simulations [144]. The effect of the restraining potentials can be removed and data from multiple simulations combined to construct the potential of mean force (PMF) along the collective coordinate by using the Weighted Histogram Analysis Method (WHAM) [145]. This method is widely considered a gold standard against which other PMF-

producing methods are compared, though the simulations are generally recognized as being rather costly to perform.

A similar method, free energy perturbation (FEP), allows one to estimate a change to the Hamiltonian governing the system dynamics by considering the change in small discrete steps [146,147]. The FEP method is, in principle, very flexible and can be used to find the PMF. In practice, the umbrella sampling method is easier for finding the PMF, but FEP can be applied to problems beyond the scope of umbrella sampling. For example, using FEP one can account for free energy associated with abstract changes to the system such as the creation and destruction of atoms or the application of new terms in the systems Hamiltonian. Such non-physical procedures must carefully consider a complete thermodynamic cycle in order to remain physically meaningful. However, the technique is extremely flexible and can be used to investigate and answer questions about the physical origin of phenomena that would be very difficult to answer using other techniques.

Other equilibrium methods for finding the PMF exist [148, 149], but it is also possible to estimate the PMF from non-equilibrium simulations [150–156]. For example, during a steered MD (SMD) simulation, one end of a spring is tethered to an atom and the other end of the spring is pulled at a constant velocity. The force applied on the atom is recorded, allowing one to estimate the PMF using the Jarzynski equality [157], which averages the work over a large ensemble of simulations.

Chapter 3

Interaction between parallel DNA duplexes

3.1 Interaction between DNA duplexes in monovalent electrolyte

The charge density of a DNA molecule is a fundamental property that governs its biological function by influencing DNA folding, packaging [158], pairing [159], and interactions with other biological macromolecules [160]. In order to develop meaningful quantitative models describing such systems and processes, a precise knowledge of the interaction between two DNA molecules, which is mostly of electrostatic origin, is mandatory. Counterions in the electrolyte surrounding a DNA screen its phosphate groups. The Poisson-Boltzmann theory describes the electrostatics of elecrolyte systems, see Sec. 1.4 for an introduction. Under this theory, the multibody problem of an atmosphere of interacting ions is made tractable by allowing the ions to feel one another only through the average electrostatic potential. Further, the finite size of ions is ordinarily neglected. Although imperfect, the Poisson-Boltzmann equation and its linearized variant are used widely because they provide a reasonable and accessible description of electrostatics in a medium relevant for biomolecules.

When modeling DNA at a meso- or macroscopic level, it is common to describe electrostatics using a potential obtained from Poisson-Boltzmann using a heuristic charge-reduction factor (CRF) that accounts for the shortcomings of the description. Often the CRF is taken to have a value of ~ 0.75 [161, 162], consistent with the effective charge determined from measurements of the electrophoretic mobility of DNA [163]. Recently, however, we demonstrated through combination of experiments, analytical analysis and atomic that the CRF appropriate for the description of DNA–DNA interactions in a plectoneme was 0.42, not the much larger factor of ~ 0.75 obtained from electrophoresis.

3.2 Plectoneme experiment

The experimental assay of our collaborators is shown in Fig. 3.1A. One end of a 1.8–11-kbp DNA duplex is tethered to a surface, the other end to a bead in a magnetic trap. The magnetic field applies a constant tension to the DNA and also turns the bead to induce supercoiling while the extension is monitored. Initially, the supercoiling induces only a small change in the DNA extension L. However, when a critical amount of supercoiling is reached, the extension begins to drop at a constant rate with the number of turns N, see Fig. 3.1B. The drop in extension is due to the growth of a plectoneme (see DNA structure in Fig. 3.1A)

that contributes negligibly to the extension but absorbs the twist introduced by turning the bead. The growth of the plectoneme depends on the applied tension F and ion concentration. Because the two halves of the DNA molecule are in relatively close proximity within the plectoneme, the growth of the plectoneme can provide a very sensitive measure of the electrostatic repulsion between DNA molecules.

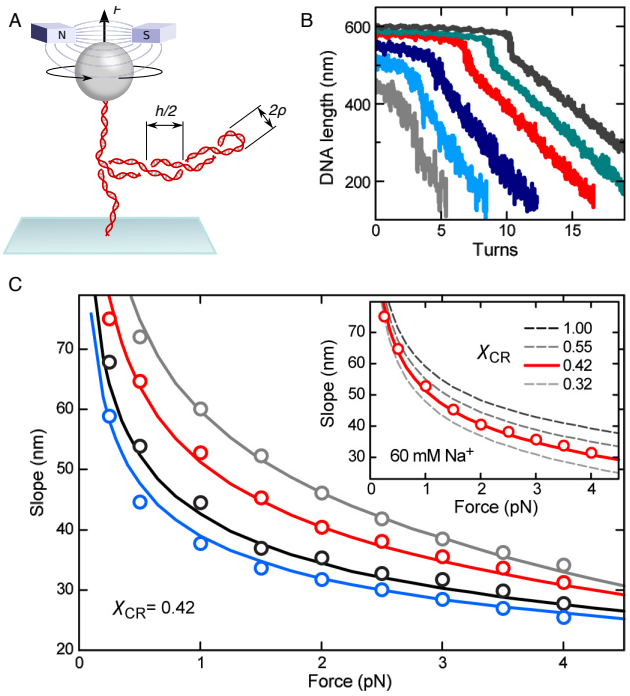

Figure 3.1: Dependence of DNA supercoiling on force and salt concentration. (A) Experimental setup and superhelix parameters. (B) DNA supercoiling curves recorded in buffer containing 170 mM Na^+ at stretching forces of 0.25, 0.5, 1.0, 2.0, 3.0 and 4.0 pN (grey, light blue, dark blue, red, green and dark gray lines). Continuous twisting was carried out at 0.5 Hz. Data was taken at 300 Hz and smoothed to 20 Hz. (C) Slopes after buckling as a function of force obtained from the supercoiling curves for Na^+ concentrations of 30, 60, 170 and 320 mM (grey, red, black and blue circles) together with the prediction from the theoretical model (solid lines) calculated for a charge adaptation factor χ_{CR} of 0.42. Inset: slopes for 60 mM Na^+ (red circles) together with theoretical predictions for different values of χ_{CR} (lines). Best agreement is found for χ_{CR}= 0.42 for all ionic strengths considered. Only slight variations of χ_{CR} to 0.32 and 0.55 lead to a significant under- or overestimation, respectively, of the slopes.

The energy E_{tot} required to add a turn to the plectoneme can be decomposed into

$$E_{tot} = E_{force} + E_{elas.} + E_{elec.}, \tag{3.1}$$

where E_{force} represents the cost of shortening the DNA under applied tension F, $E_{elas.}$ the energy required to bend the DNA in the plectoneme, and $E_{elec.}$ the DNA–DNA interaction energy. We can trivially write $E_{force} = F\frac{dL}{dN}$. The plectoneme is assumed to have a tension-dependent geometry with average radius ρ and length h per turn. Fluctuations are neglected here; it was shown that inclusion of fluctuations did not

significantly affect the results of the present analysis [49]. Thus $\frac{dL}{dN} = \frac{(2\pi\rho)^2+h^2}{h}$. The DNA length added to the plectoneme by each turn of the bead is $\frac{dL}{dN} = \frac{(2\pi\rho)^2+h^2}{h}$ [162]. The DNA in the plectoneme is curved, which—applying an elastic model for DNA bending [162]—requires a bending energy of

$$E_{\text{elas.}} = \frac{dL}{dN}\frac{1}{2}pk_{\text{B}}T\left[\frac{2\pi\rho}{(2\pi\rho)^2+h^2}\right]^2, \tag{3.2}$$

where p is the bending persistence length (50 nm) [95] and $k_{\text{B}}T$ is the thermal energy.

Although DNA is a highly charged molecule, the structure of a plectoneme causes two portions of the DNA double helix to be in close proximity. Therefore, one expects the electrostatic energy per plectonemic turn to be significant. This energy can be approximated as the interaction energy between two cylindrical rods of uniform charge density under the linearized Poisson-Boltzmann or Debye-Hückel (DH) theory, see Sec. 1.4. The DH theory linearizes the Poisson-Boltzmann equation, which is convenient because the interaction energy between the two halves of the plectoneme can then be decomposed into pairwise contributions from infinitesimal segments of each strand. Under DH, two monovalent point charges q_1 and q_2 at distance r interact with energy $u(r) = k_{\text{B}}Tl_{\text{B}}\frac{q_1q_2}{e^2}\frac{\mathrm{e}^{-r/\lambda_D}}{r}$, where λ_D is the ion concentration-dependent Debye length ($\sim$1 nm in 100 mM solvent), and l_{B} is the Bjerrum length—the distance at which the Coulomb energy between monovalent charges e is $k_{\text{B}}T$ in water, about 0.7 nm at room temperature.

The exponential screening of electrostatic interactions allows the problem to be further simplified by approximating the plectoneme as an infinite structure so that the electrostatic energy per turn of the plectoneme is independent of the overall length of the plectoneme. Consequently, the electric field due to one half of the plectoneme is uniform along the contour length of the other plectoneme. Then the electrostatic energy per plectoneme turn is

$$E_{\text{elec.}} = k_{\text{B}}Tl_{\text{B}}\int_{s_1}\int_{s_2}\frac{\mathrm{e}^{-r/\lambda_D}}{r}\frac{dq_1dq_2}{e^2}, \tag{3.3}$$

where the charge of an infinitesimal section of contour length ds_i is $dq_i = \xi_{\text{eff}}ds_i$ and ξ_{eff} is the effective charge density of the DNA. The integral over s_2 spans only half the DNA added to the plectoneme, $\frac{1}{2}\frac{dL}{dN}$. Dropping the remaining subscript,

$$E_{\text{elec.}} = \frac{1}{2}\frac{dL}{dN}k_{\text{B}}Tl_{\text{B}}\left(\frac{\xi_{\text{eff}}}{e}\right)^2\int_s\frac{\mathrm{e}^{-r/\lambda_D}}{r}ds. \tag{3.4}$$

The effective charge density can be written $\xi_{\text{eff}} = \xi\,\chi_{\text{CR}}\,\chi_{\text{PB}}$, where ξ is the density of charge on DNA $2e/0.34$ nm, the charge reduction factor χ_{CR} is left as a tunable parameter. The ion concentration-dependent factor χ_{PB} fits the electrostatic potential due to a DH line charge to that from a Poisson-Boltzmann cylinder

of uniform charge density $2e$/nm and radius 1.2 nm (see Fig. S2 of Supporting Information of Ref. 49 for details). For a given applied force and ion concentration, this integral was evaluated numerically and the overall energy per turn E_{tot} was minimized with respect to the plectoneme parameters ρ and h, allowing theoretical calculation of $\frac{dL}{dN}$. By using a significantly reduced DNA charge with $\chi_{\mathrm{CR}} \sim 0.42$, excellent agreement to the experimental slopes $\frac{dL}{dN}$ was obtained across a very large range of applied tensions (0.25–4 pN) and sodium concentrations (30–320 mM).

The most commonly employed CRF of ~ 0.75 was obtained from a best fit to the electrophoresis measurements of Schellman and Stigter [163]. This CRF is commonly used in the description of DNA–DNA interactions although the physics of these interactions is considerably different from the physics of electrophoretic transport [163]. Previous experiments that quantitatively investigated the DNA–DNA interactions out of the condensed phase also failed to find a CRF of 0.42 [161, 164]. The success of the present experiment is likely because force was used to confine the DNA in close proximity, thereby probing the repulsion between strands in a region of phase-space that is rarely sampled in more conventional experiments.

3.3 Verification of reduced charge factor by all-atom MD simulation

In order to obtain a microscopic verification for χ_{CR}, all-atom molecular dynamics simulations [165] were employed to obtain the force between parallel DNA molecules at different salt concentrations. We obtained good agreement with the force calculated according to Eq. 3.1 using $\chi_{\mathrm{CR}} = 0.42$ and very poor agreement with $\chi_{\mathrm{CR}} = 1.0$, see Fig. 3.2.

To test whether the value found for χ_{CR} is specific to DNA–DNA interactions or is a universal constant for DNA electrostatics, we obtained the ion distributions around isolated double-stranded DNA molecules. The radial ion distribution extended further from the DNA than expected from PB theory with $\chi_{\mathrm{CR}} = 0.42$ and approached the distribution for $\chi_{\mathrm{CR}} = 1.0$ at low ionic strength, see Figs. 3.3 and 3.4A. Thus, the small value for χ_{CR} is specific to DNA–DNA interactions.

At elevated ion concentrations, the mean electrostatic potential around the DNA, obtained using the PMEPot plugin of VMD [166,167], was too weak to create the ion distribution as observed in simulation and as predicted by PB theory, see Fig. 3.4 B,C. This suggests that ion distributions are not only determined by electrostatics, but that other effects such as correlations in the ion clouds, the noncontinuum nature of the dielectric surrounding, and ion exclusion can have a significant influence. This in turn is likely to cause the low DNA–DNA interaction forces. Additionally, deviations from the homogeneously charged rod model,

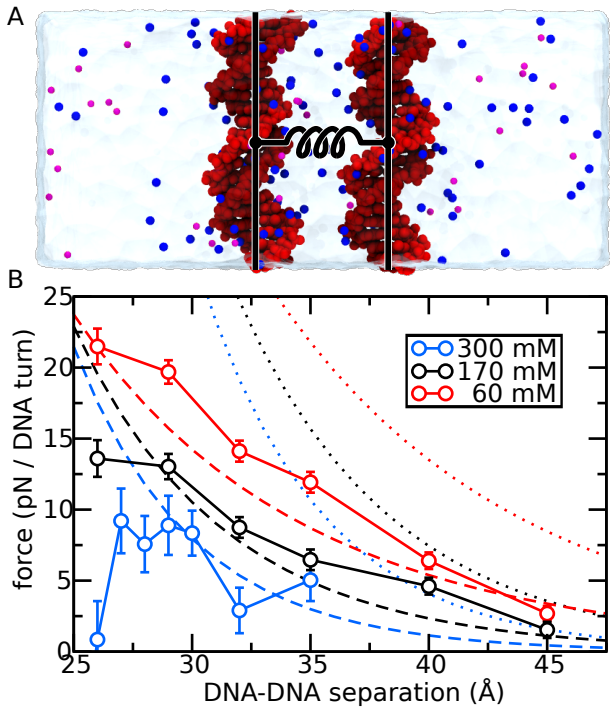

Figure 3.2: All-atom molecular dynamics simulations of the effective force between double-stranded DNA. (A) All-atom model used to determine the force. The DNA atoms are depicted as red spheres; the counter and co ions are depicted as blue and purple spheres, respectively; the water is shown as a semi-transparent molecular surface. The distance between the DNA molecules was restrained by a harmonic potential, schematically depicted as a spring. (B) The simulated mean force between the DNA. Data from MD simulations (circles, solid lines) are shown alongside the theoretical predictions using a charge adaptation factor χ_{CR} of 0.42 (dashed lines) and 1.0 (dotted lines) at 60 (red) and 170 mM (black) bulk ion concentrations. Data for 300 mM (blue) bulk ion concentrations is reproduced from previous work [165]. The relative azimuthal orientation of the DNA did not appear to affect the obtained force within error. The atomic structure of the interacting DNA and nearby solvent cause the force to drop at 300 mM as the separation between the DNA surfaces approaches the atomic scale [165].

such as strongly localized charges at the phosphates and counterions entering the DNA grooves, may reduce the interaction forces.

3.4 Conclusion

By combining single-molecule experiments, theoretical considerations, and all-atom simulations, we have shown that, within a cylinder approximation, DNA–DNA interactions can be described only by a significantly reduced DNA charge. Furthermore, we have provided a theory that accurately describes DNA supercoils over a broad range of tension and ionic strength. Supercoiling under tension is a unique way to bring two DNA molecules into close proximity in the absence of interfering surfaces. In contrast to previous topological investigations of long DNA [161, 164, 168], our force-based experiments allow a more

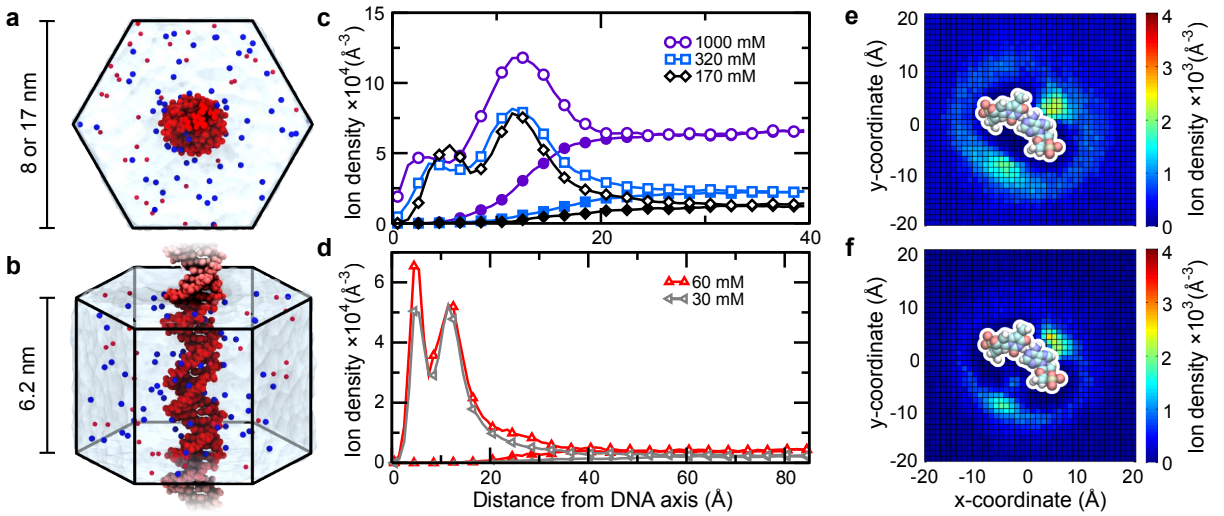

Figure 3.3: The distribution of monovalent ions around double-stranded DNA. (a,b) All-atom model used to find the ion distribution around DNA in MD simulations. The DNA atoms are depicted as red spheres; the counter and co ions are depicted as blue and red spheres, respectively; the water is shown as a semi-transparent molecular surface. (c,d) The radial ion distribution around DNA. Counter-ion (open symbols) and co-ion (filled symbols) distributions are shown for 1000, 320, 170 (c), 60, and 30 (d) mM NaCl concentrations. (e,f) The 2D counter-ion density around DNA. A z-dependent rotation was applied in the xy-plane to counter the helical pitch of canonical DNA; the counter-ion density was subsequently averaged along the z-axis. Data is shown for 320 (e) and 30 (f) mM bulk ion concentrations. A typical basepair is shown after the transformation was applied, indicating the approximate position of the DNA. The transformation has the effect of radially smearing the ion distribution, making it appear as though the ion concentration near the adenine is larger than that near the DNA phosphates (which are located further from the origin). In actuality, more counter ions can be found in close proximity of a DNA phosphate than an adenine at a given time.

reliable quantification of DNA–DNA interactions since much closer proximities are achieved (see Supporting Information of Ref. 49). The surprisingly small DNA–DNA interactions result from a complex interplay of a highly charged and structured molecule with solvent molecules and ions. The simple effective interaction potential will be an important contribution for quantitative models of complex biomolecular systems which cannot be treated with atomic detail such as DNA packaging in chromatin and viruses and possibly also for protein–DNA interactions. The findings reveal that particular caution is necessary when applying effective charge parameters obtained from experiments that probe a different physics such as electrophoresis [163].

3.5 Simulation methods

All MD simulations were performed using the parmbsc0 refinement of the AMBER parm99 force field [136, 169], the TIP3P water model [170], standard parameters for ions [171], periodic boundary conditions, and 1–2–4-fs multiple time stepping [172]. Van der Waals and short-range electrostatic potentials were calculated

using a smooth (10–12 Å) cutoff; the particle mesh Ewald method was used to compute long-range electrostatics using a 1.0 Å-spaced grid. The temperature was kept constant by applying Langevin forces [173] to all non-hydrogen atoms; the Langevin damping constant was set to 0.1 ps^{-1}. Simulations were performed in the NPT ensemble using Nosé-Hoover Langevin piston pressure control [174] at 1 bar. At the time that this research was performed, the NBFIX correction to ions was unavailable [61]. Due to the relatively low ion concentrations and the use of monovalent electrolyte, we expect that the MD results would not be qualitatively changed by the updated parameters.

For the force calculations, a previously described method was followed [165]. Two copies of $(dA{\cdot}dT)_{20}$ were placed 26, 29, 32, 35, 40, or 45 Å apart in a box containing 60 or 170 mM NaCl electrolyte. The size of the simulation system was chosen to ensure that the force due to periodic images would be <2% of the force between the DNA in the same unit cell, assuming an exponentially decaying force. The distance between the DNA molecules was harmonically restrained with a force constant of 3000 pN/Å while the force was recorded. Three copies of the system at each DNA separation and ion concentration were simulated. The first nanosecond of the force data was removed. The data was subsequently concatenated and averaged in 200-ps blocks to remove correlations affecting the standard error of the mean. The relative azimuthal orientation of the DNA did not appear to affect the obtained force within error.

To find the ion distribution around double-stranded DNA, a two-turn poly(dA·dT) molecule was immersed in NaCl electrolyte. The DNA was placed so that its helical axis coincided with the z-axis of the simulation cell. Bonds were placed across the periodic boundary to make an effectively infinite DNA molecule. Hexagonal boundary conditions were employed in the xy-plane with 40 Å between periodic images for systems containing 1000, 320, or 170 mM electrolyte, and 85 Å for systems containing 60 or 30 mM electrolyte, Fig. 3.3 A,B. The resulting average ion distributions are show in Fig. 3.3.

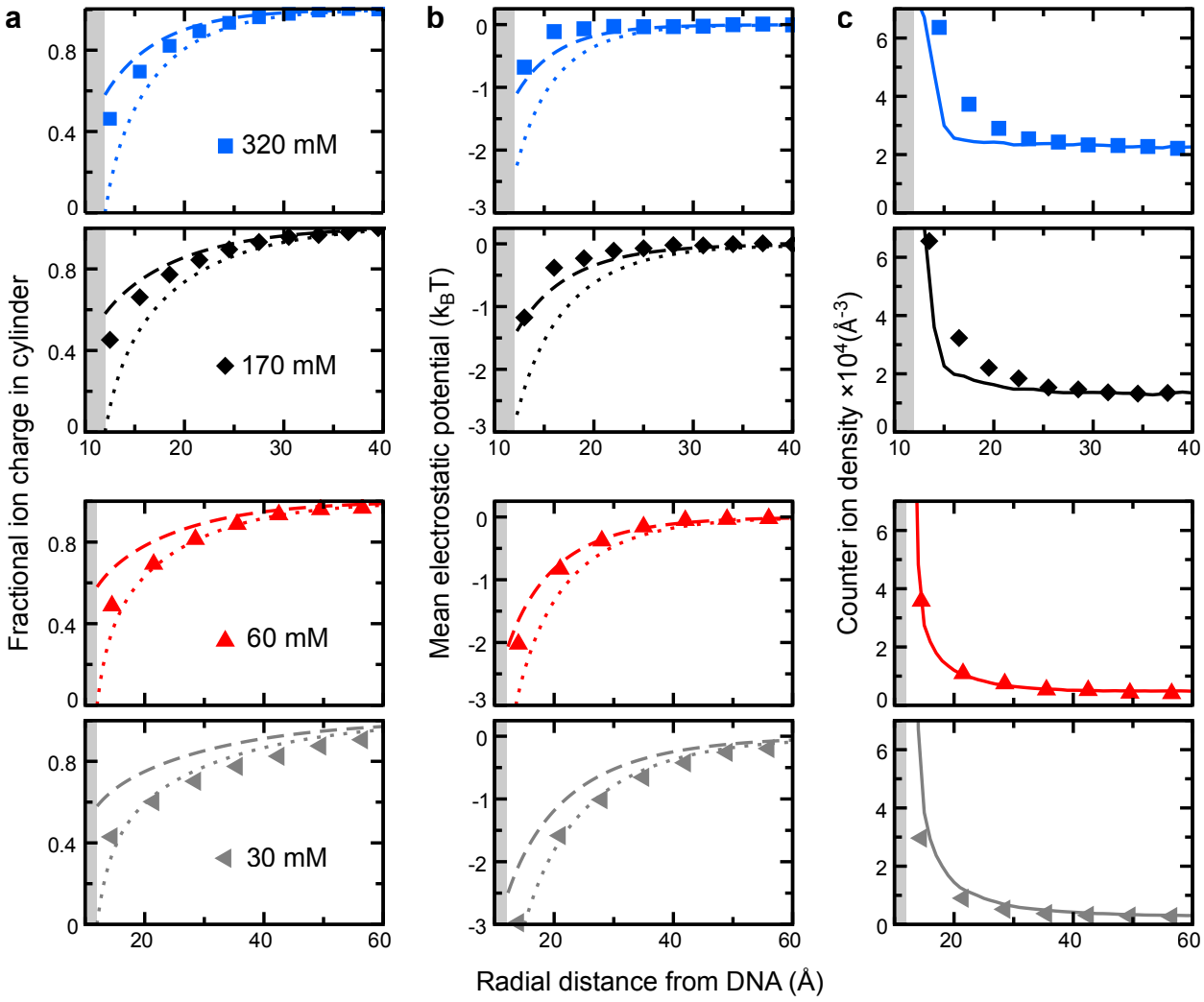

Figure 3.4: Charge distribution and potential around double-stranded DNA. Data resulting from MD simulations described in Fig. 3.3 are plotted as solid symbols. Solutions to the non-linear Poisson-Boltzmann equation for a cylinder of 1.2 nm radius within a hexagonal unit cell (of same size as used in the simulations) are shown for $\chi_{CR} = 1.0$ (dotted lines) and $\chi_{CR} = 0.42$ (dashed lines) at each ion concentration. The cylinder boundary is represented by a gray box. (a) The fraction of the total ionic charge contained within virtual cylinders of increasing radii around the DNA nearly matches the PB solution for $\chi_{CR} = 1.0$ at low ion concentration. At higher ion concentration, the enclosed charge is somewhere between the PB predictions for $\chi_{CR} = 0.42$ and $\chi_{CR} = 1.0$. (b) The mean electrostatic potential around the DNA similarly matches the prediction for $\chi_{CR} = 1.0$ at low ion concentration, but falls to zero even more rapidly than predicted with $\chi_{CR} = 0.42$ at high ion concentration. The instantaneous 3D electrostatic potential was calculated by averaging over the trajectory in 5 ps intervals. Subsequently the z-axis and azimuthal angles were averaged out to obtain the data shown. Using Boltzmann weights to perform the spatial average was found to affect the results only very weakly. (c) Comparing the counter ion distribution expected from a Boltzmann weight of the potential from the MD simulations in b (solid line) to the directly observed counter ion density (symbols) reveals that the mean electrostatics fails to describe the counter ion density at high ion concentration.

Chapter 4

End-to-end attraction of duplex DNA

4.1 Introduction

Self-assembly properties of nucleic acids are vital to the basic functions of a biological cell and have been extensively exploited in biotechnology. DNA hybridization—self-assembly of complementary sequence single-stranded DNA (ssDNA) into a double helix—is a central biotechnological process [176], used, among others, in platforms for DNA detection [177], programmable assembly of DNA nano-structures [76,178], directional transport of cargo [179], molecular computing [180], and nanofabrication [181, 182]. Another process of outstanding importance is DNA condensation, where counterions transform electrostatic repulsion between naked DNA molecules into attraction, facilitating packaging of double stranded DNA (dsDNA) in cell nuclei and viral capsids [183, 184].

Recently, an entirely different type of DNA self-assembly was discovered: spontaneous end-to-end aggregation of short duplex DNA fragments into rod-like structures [185]. When water was evaporated from solution containing a high concentration of short (6–20 bp) DNA fragments, liquid crystal phases were observed. Since the DNA fragments were nearly as wide as they were long, the observation of axial ordering could only be explained if the fragments formed rod-like supramolecules, suggesting end-to-end aggregation. Further experimental evidence of end-to-end association was obtained from the analysis of small angle x-ray scattering data from a system containing short DNA fragments and a divalent electrolyte [186, 187]. The second virial coefficient extracted from these data was shown to be positive for DNA fragments capped with a short hairpin (indicating overall repulsion) and negative for DNA fragments without such caps (indicating overall attraction). It was concluded that end-to-end attraction was large enough to overcome electrostatic repulsion in a divalent electrolyte. While the side-by-side force between long DNA molecules has been the subject of many experimental [75, 188, 189] and theoretical [190, 191] studies, little is known about the conditions and microscopic mechanism of DNA association end-to-end. Furthermore, the effects of end-to-end attraction of duplex DNA in biological and technological processes are entirely unexplored.

Whereas traditional single-molecule experiments have provided extensive information about DNA hybridization and side-by-side interactions [49, 75], applying these tools to study end-to-end assembly is extremely difficult, as a DNA duplex's cross section is just 5 nm^2 and the effective concentration of DNA ends in a solution amenable to single-molecule studies is typically small. The all-atom molecular dynamics (MD) method is well suited for the study of systems that share the length scale of short dsDNA molecules

and can be used to probe the atomic origin of intermolecular forces [192]. Here, we use the MD method to characterize end-to-end association of duplex DNA in unprecedented detail, elucidating the microscopic mechanism of spontaneous association, its free energy costs and the kinetic rates. At the end of the chapter, we discuss the relationship between our findings and the pertinent experimental observations.

4.2 Collapse of aligned dsDNA

Spontaneous end-to-end association of duplex DNA was observed in the simulations of two $(\mathrm{dA}{\cdot}\mathrm{dT})_{10}$ fragments constrained to diffuse along a common axis. Figure 4.1 a illustrates the initial state of a typical simulation system comprising DNA in an anisotropic volume of 100 mM NaCl electrolyte. Figure 4.1 b plots the distance between the DNA fragments versus time for two simulation systems differed by the termination of the DNA's 5′-ends. The DNA fragments were observed to freely diffuse along the common axis until the end-to-end distance fell below ≈ 8 Å, whereupon the fragments rapidly collapsed into an end-to-end bound complex. The relative azimuthal orientation of the DNA fragments ϕ continued to change after the collapse, switching between several metastable orientations, Fig. 4.1 c. We define the relative azimuthal angle ϕ as the angle between the projections of the vectors connecting the O5′ and O3′ atoms of the terminal basepairs into the plane normal to the common DNA axis (see Sec. 4.9).

In the final conformation adopted by the blunt-ended fragments, the 5′-to-3′ direction of the backbone was discontinuous at the end-to-end junction. In the case of the 5′-phosphorylated fragments, the 5′-to-3′ direction of the backbone was continuous at the end-to-end junction as though the backbone of a continuous 20 bp B-DNA had been cut.

To determine the statistical significance of the above observations, we performed 17 additional simulations for each system type, different by the relative azimuthal orientation of the DNA fragments at the onset of the simulation: $\phi_{t=0} = i \times 20^\circ$, where $i = 1, \ldots, 17$. In all simulations, we observed collapse of the DNA fragments into an end-to-end bound complex. Figure 4.1 d plots the the relative azimuthal orientation at the time of collapse against the time to collapse. The collapse of blunt-ended fragments occurs irrespective of their azimuthal orientation, whereas formation of a 5′-phosphorylated end-to-end assembly did not occur with ϕ between 90 and 230°.

After collapse, ϕ continues to evolve, reaching the states characterized in Fig. 4.1 e. The clustering of ϕ

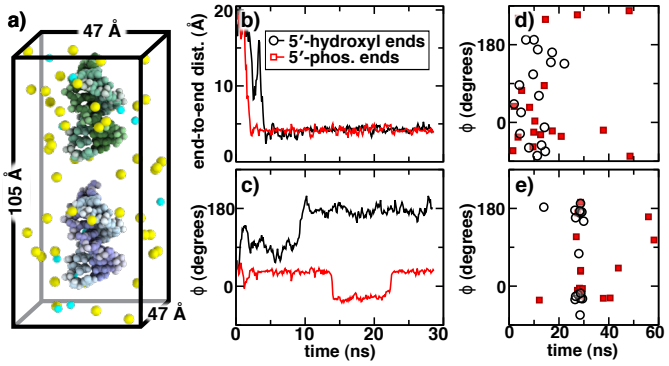

Figure 4.1: Collapse of aligned dsDNA. (a) Simulation system containing axially aligned duplex DNA. Each DNA fragment is free to move along and rotate about the common axis. The DNA duplexes (blue and green) are shown in van der Waals representation; sodium and chloride ions are shown as yellow and cyan spheres; water is not depicted. (b,c) The end-to-end distance (b) and the relative azimuthal angle ϕ (c) of two duplex DNA in representative simulations of the end-to-end collapse. Data from the same pair of simulations are plotted in (b) and (c). (d,e) Scatter plot showing the relative azimuthal angle ϕ at the time of collapse (d), and at the end of simulation (e). One data point is shown for each of 36 simulations of blunt-ended (black circles) and 5′-phosphorylated (red squares) dsDNA fragments in 100 mM NaCl electrolyte.

values around -20, 36 and 180° indicates the three preferred binding states. At $\phi \approx 36°$, the conformation of the bound complex is similar to that of a continuous B-form DNA. At $\phi \approx -20°$, the backbones of the terminal basepairs overlap slightly (Fig. 4.2 b). At $\phi \approx 180°$, the 5′-ends of the fragments are in direct contact (Fig. 4.2 c). The conformations of the end-to-end assemblies with $\phi = -20$ and 36° are similar, differing primarily in the order in which the 5′- and 3′-termini are lapped and in their base stacking geometry. Thus, relative to the coordinates of two consecutive basepairs in a canonical B-DNA helix, the terminal basepairs forming an end-to-end junction have a time-averaged root mean squared deviation of 2.1 and 0.9 Å for the $\phi = -20$ and 36° conformations, respectively. For reference, basepairs in the middle of one of the DNA fragments had a root mean squared deviation of 0.7 Å.

The preference for these three orientations suggests a hydrophobic origin of the attractive force, as such conformations reduce exposure of the hydrophobic bases to water. The relative orientations of the blunt-ended DNA are nearly equally split between the -20 and 180° states. More than 50% of the 5′-phosphorylated fragments formed the $-20°$ state, 35% formed either the 36° or 180° state (three systems each), and 12% formed the state with $\phi \approx 100°$. We attribute such preferential alignment of the 5′-phosphorylated fragments to the electrostatic repulsion between the terminal phosphate groups, which apparently renders the 180° orientation energetically less favorable than the $-20°$ one. The free-energy difference between these bound states is discussed below.

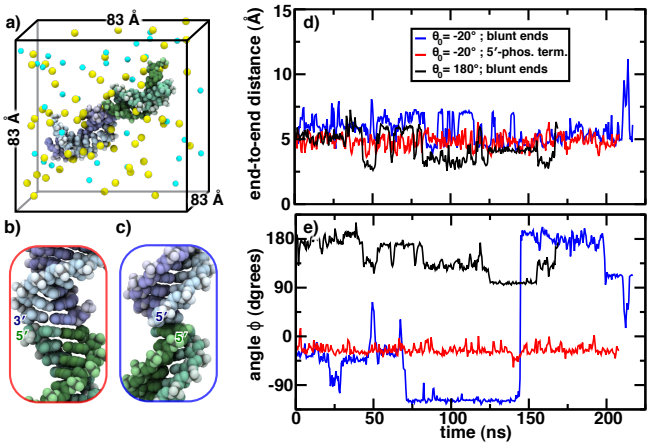

Figure 4.2: Stability of the end-to-end DNA assembly (a) Simulation system containing two spatially unrestrained dsDNA fragments in 100 mM NaCl electrolyte. Both fragments are free to rotate and move about the simulation box. The system is drawn as in Fig. 4.1 a. (b,c) End-to-end junction of 5′-phosphorylated dsDNA with ϕ=$-20°$ (b) and blunt-ended dsDNA with ϕ=180° (c). (d,e) End-to-end distance (d) and relative azimuthal angle, ϕ, (e) of the DNA fragments in three unrestrained MD trajectories.

4.3 Stability of the end-to-end complex

DNA fragments initially forming a bound state were simulated in the absence of any restraints using the isotropic system shown in Fig. 4.2 a. Three systems were constructed: one containing 5′-phosphorylated DNA bound with $\phi = -20°$(Fig. 4.2 b) and two containing blunt-ended DNA fragments bound at $\phi = 180°$ (Fig. 4.2 c) and $\phi = -20°$. All three systems contained 100 mM NaCl electrolyte.

The plot of the end-to-end distance, Fig. 4.2 d, indicates that all three assemblies remained bound for the entire duration of the simulations (>200 ns). The standard deviation of the end-to-end distance in the 5′-phosphorylated system was 0.68 Å, twice less than that of the blunt-ended systems. The greater stability of the 5′-phosphorylated complex may be due to hydrogen bonds that were observed between the 5′-terminal phosphate of one DNA fragment and the 3′-terminal hydroxyl of the other fragment. The plot of the relative azimuthal angle reveals that the 5′-phosphorylated system maintained the same stable conformation at $\phi \approx -20°$, depicted in Fig. 4.2 b. Starting from a similar conformation, the blunt-ended complex underwent two sudden rotations at 70 and 140 ns that brought ϕ from $-20°$ to $-110°$ and to 180°; the relative orientation continued to evolve after that.

In the simulation of the 180° blunt-ended complex, the terminal basepair of one of the fragments ruptured after 33 ns. During the next few nanoseconds, Watson-Crick pairs within that fragment stochastically broke and re-annealed, propagating the unpaired base towards the opposite end of the DNA fragment, and slipping the entire DNA strand with respect to the other by one basepair. Despite the slippage, the DNA fragments remained stably bound.

4.4 Mechanics of end-to-end dissociation

The lifetime of a bound complex sharply decreases when an external force is applied to disrupt it [193]. Thus, under a constant force of 150 pN directed along the common axis of the DNA fragments, the assembly remained intact during a 50-ns simulation but dissociated within 5 ns under a 200 pN force. These simulations used the same initial conditions as the SMD simulations described below in Sec. 4.4.2. The phosphorous atoms of one DNA fragment were harmonically restrained ($k_s = 445$ pN/nm) about their initial coordinates. Tilting of the terminal basepairs, which we associate with shearing of the end-to-end assembly, was suppressed by the restraints; the force needed to rupture the assembly would likely have been lower in the absence of these restraints.

To determine the dissociation pathway, the rupture force and the mechanical work required to dissociate the assemblies, the DNA fragments were subject to the force of a harmonic spring whose equilibrium-extension length was increased at a constant rate. Spring-driven rupture of this sort has been used extensively in the study of proteins [194, 195]. Here, the spring was tethered to different sites. In all cases, rupture was observed to occur by sliding of one terminal basepair relative to the other. Although the three rupture protocols yielded different dependencies of the force on the separation distance, the average work performed was 9.4 ± 1.5 kcal/mol, independent of the rupture protocol. The typical rupture forces varied between 100 and 200 pN and were considerably lower when the springs were tethered to each fragment's center of mass (CoM). Inclusion of short overhangs of complimentary sequence ssDNA at the ends of the fragments increased the work require to rupture the end-to-end assemblies. Details of these simulations are provided below.

4.4.1 Axial stretching of the end-to-end assembly

Force induced dissociation of the end-to-end DNA assembly was simulated using the steered molecular dynamics (SMD) method [151]. Figure 4.3 a schematically illustrates the process. Each DNA fragment was attached to one end of a virtual spring. The ends of the spring were anchored to the center of mass (CoM) of either the phosphorous atoms of the DNA fragments (6 simulations) or the CoM of the closest DNA ends (4 simulations). The rest length of the spring was increased at a constant rate of 0.4 or 0.2 Å/ns until the rupture of the assembly occurred. The work performed during each of these simulations was recorded and combined using the Jarzynski's equality [157] to estimate the potential of mean force (PMF) for the process considered:

$$e^{-\beta \Delta F} = \left\langle e^{-\beta W} \right\rangle \tag{4.1}$$

The angle brackets denote an ensemble average; β is $1/k_BT$; ΔF is the change in free energy when the system is brought from one state to another; W is the work done during the change of state. No clear relationship between the pulling rate and the work performed was observed. We note that these simulations may not have sampled rare trajectories required for accurate estimation of the PMF using Jarzynski's equality [157].

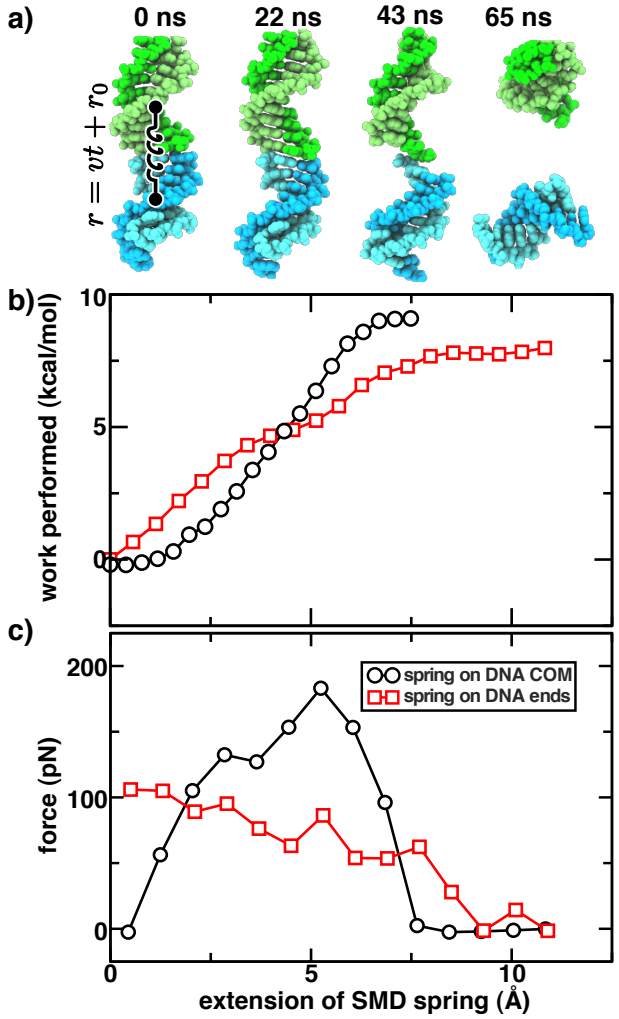

Figure 4.3: Axial stretching of the end-to-end DNA assembly. (a) Snapshots illustrating the conformations of DNA fragments in a typical simulation of axial stretching performed using a spring tethered to the CoM of the fragments. The DNA fragments are drawn using green and blue van der Waals spheres. (b) Average work performed during the stretching simulations. The force of the virtual spring was applied either to the centers of mass (black circles; 6 simulations) or to the ends (red squares; 4 simulations) of the DNA fragments. (c) Force applied by the spring during the axial stretching simulations, averaged in blocks (roughly 1 block/Å extension) using data from multiple simulation trajectories.

The free energy difference between bound and unbound states was $\sim$ 8.5 kcal/mol regardless of the choice for the spring anchors, Fig. 4.3 b. However, the variation of the applied force with the rest length of the spring (Fig. 4.3 c) depended strongly on the simulation protocol employed, reflecting differences in the

rupture pathway. Figure 4.3 a depicts the rupture process of the end-to-end assembly by a spring tethered to the CoM of the DNA: the DNA fragments stretch until the terminal basepairs tilt and the ends separate by shearing. When the spring force was applied to the ends of the fragments, the rupture occurred by shearing but without significant deformation of the DNA.

4.4.2 Transverse shearing of the end-to-end assembly

In the axial stretching simulations, the rupture of the end-to-end assembly occurred by shearing of the terminal basepairs regardless of where the spring was anchored. We carried out a set of SMD simulations designed to produce shearing deformation: pulling one DNA fragment with respect to the other in a direction perpendicular to the common axis of the DNA assembly. Prior to production simulation, an end-to-end complex of two DNA fragments was equilibrated in the presence of restraints that aligned its symmetry axis with the z axis of the simulation cell. In the resulting configuration, the major groove of the end-to-end junction nearly faced the y axis of our coordinate system, while the minor groove nearly faced the $-y$ axis. During production simulations, one end of a spring ($k_s = 4000$ pN/nm) was tethered to the terminal basepair of one DNA fragment while the other end of the spring was pulled in the $-x$, $-y$, $+x$, or $+y$ direction at a rate of 0.2 Å/ns. All phosphorous atoms of the other DNA fragment were harmonically restrained ($k_s = 13,900$ pN/nm per atom) about their initial coordinates.

Figure 4.4 a depicts a typical MD trajectory resulting from the SMD simulations of transverse shearing. The work performed during each of the four simulations is shown in Fig. 4.4 b. By the time of rupture, which is indicated by an open circle, the work reached $\sim$10 kcal/mol. These simulations, which did not mechanically deform the DNA, confirm the surprisingly high energy required to rupture the assembly during the SMD simulations of axial stretching.

4.4.3 Rupture of end-to-end assemblies containing complementary ssDNA overhangs

Five systems containing 20 DNA basepairs were built to determine the effect of complementary overhangs on the stability of the end-to-end DNA complex. The first system was a continuous 20-basepair (two-turn) helical dsDNA fragment immersed in 0.1 M aqueous solution of NaCl. The other four systems were built from the first system by cleaving the dsDNA into a pair of equal-length 5′-phosphorylated dsDNA fragments containing 0-, 1-, 2- or 4-nucleotide, complementary ssDNA tails, Fig. 4.5 a. Following usual minimization and equilibration protocols, the DNA assembly was axially stretched using the SMD method described in Sec. 4.4.1. The force of the virtual spring was applied to the CoM of the DNA fragments.

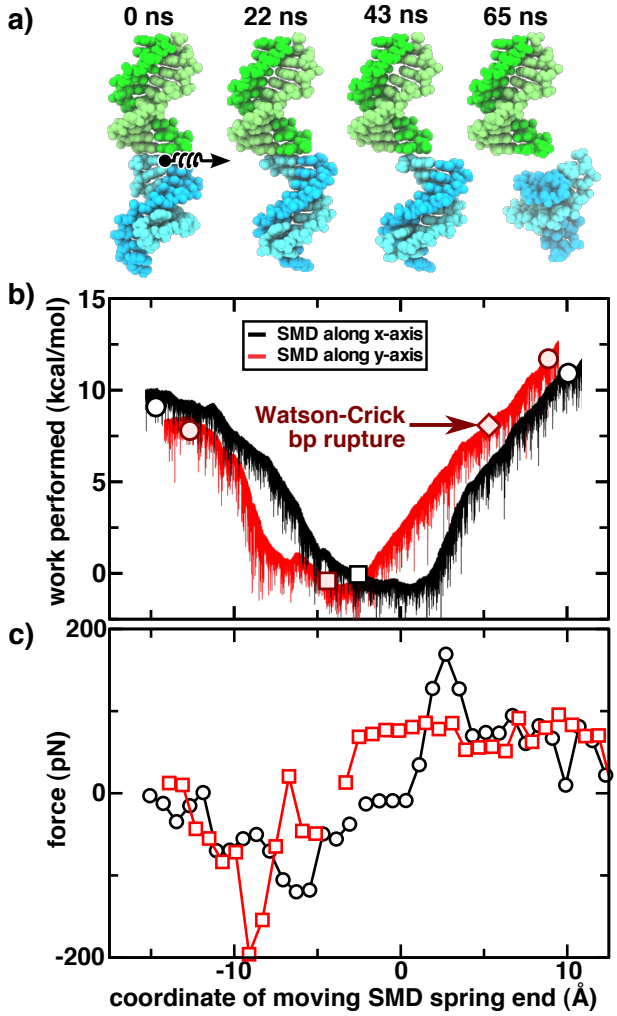

Figure 4.4: Transverse shearing of the end-to-end assembly. (a) The rupture process. The phosphorous atoms of one DNA fragment (top; green) were harmonically restrained about their initial coordinates while the other DNA fragment (bottom; blue) was pulled along the $-x$ axis. The DNA atoms are depicted as in Fig. 4.3 a. (b) Work required to shear the DNA ends. Plotted is the work performed by the SMD spring as a function of the coordinate of the moving end of the spring along the x (black) or y (red) axis. The spring force was applied to the CoM of the terminal basepair of the moving fragment. The symbols denote the state of the system at the beginning of the simulations (squares) and at the moment of rupture (circles). In one simulation, the terminal basepair tethered to the SMD spring ruptured; this moment is indicated by a diamond. (c) The force of the spring applied to the DNA fragment, averaged in $\sim$1-Å-extension blocks.

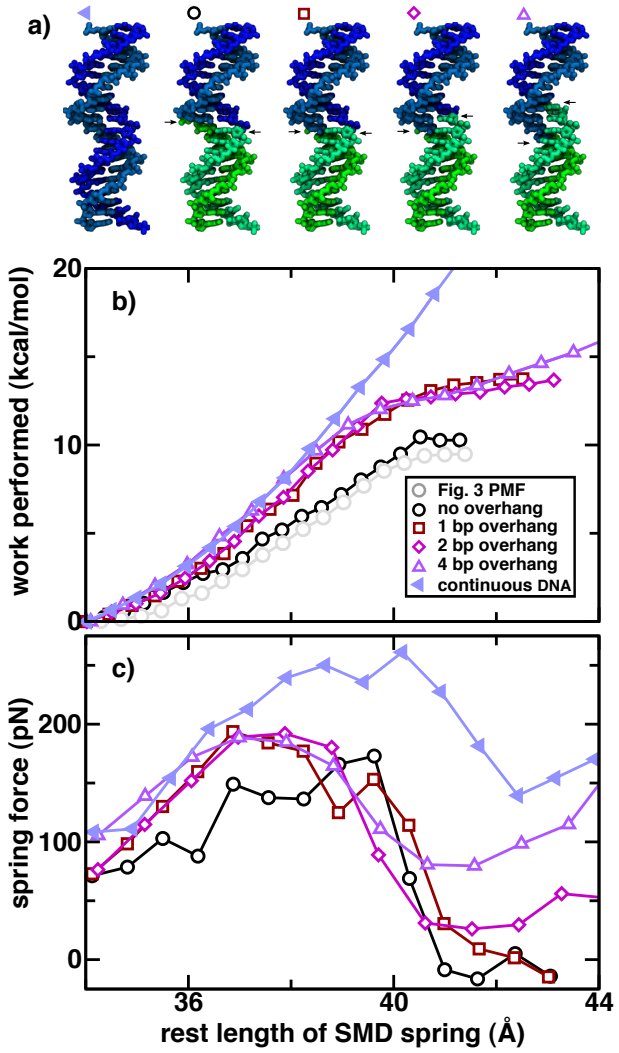

Figure 4.5: Axial stretching of end-to-end assemblies having complementary ssDNA overhangs. (a) Five DNA systems were explored in this set of simulations. In addition to a continuous 20-basepair dsDNA helix (closed left-facing blue triangles), the systems contained two disjoint DNA fragments with 0 (black circles), 1 (red squares), 2 (magenta diamonds), or 4 (purple triangles) complementary ssDNA overhangs. The arrows indicate the positions of nicks in the DNA assemblies. (b) The average work performed by the spring during axial stretching. For comparison, grey circles reproduce the results of axial stretching shown in Fig. 4.3 b (grey circles). (c) The force of the spring averaged in 1-Å blocks of spring extension.

Figure 4.5 b shows the average work performed, estimated for each system from two independent SMD trajectories. The average work profile in the absence of ssDNA overhangs (black circles) agrees well with the result of our previous simulation of a similar system (grey circles), which began with a different conformation of the DNA fragments. The work profiles for stretching two blunt-ended DNA fragments begins to diverge from that of a continuous dsDNA fragment (closed left-facing blue triangles) as the basepairs forming the junction in the former begin to shear. Rupturing the end-to-end complex of two DNA fragments having 1-nucleotide overhangs (red squares) requires $\sim$ 4.5 kcal/mol more work than rupturing blunt-ended fragments; the presence of ssDNA overhangs suppresses the shearing of basepairs at the end-to-end junction at low forces. In the case of the fragments having 2- or 4-nucleotide overhangs, the average work profiles (magenta diamonds and purple triangles) initially follows that of a continuous dsDNA but diverges from it at larger extensions once base stacking in the end-to-end assemblies ruptures, leaving stretched ssDNA tails that maintain contact over the end-to-end junction.

From these simulations, we conclude that the presence of sticky overhangs makes the end-to-end assembly similar to a continuous duplex DNA at load forces smaller than that required to overstretch a dsDNA helix. In comparison to a complex of two blunt-ended DNA fragments, rupture of the fragments having ssDNA overhangs requires additional work, which, in the case of single nucleotide overhangs, we found to be in good agreement with the free energy required to rupture an isolated AT base [196].

4.5 Potential of mean force of axially aligned DNA duplexes

To improve our estimates of the force and free energy of the end-to-end interaction, 100 variants of the two-DNA fragment system, Fig. 4.1 a, were simulated at a constant end-to-end distance enforced by a harmonic spring potential in 100 mM NaCl electrolyte. Each simulation explored a unique combination of the end-to-end distance and the azimuthal angle and lasted 18 ns (aggregate simulation time was 1.86 μs). The DNA fragments were kept aligned by weak harmonic restrains (see Sec. 4.9.6) that allowed the terminal basepairs to shear.

Fig. 4.6 shows the dependence of the effective end-to-end force on the end-to-end distance and the potential of mean force (PMF) reconstructed from this set of simulations by the weighted histogram analysis method [141, 144]. The PMF can be thought of as the change of free energy along a chosen coordinate. The force sharply increases with the end-to-end distance between $\sim$3.5 Å—the distance between consecutive basepairs in a DNA helix—and $\sim$6.5 Å, the separation allowing water molecules to penetrate the volume between the ends of the fragments. The force rapidly decreases as the end-to-end distance exceeds 6.5 Å

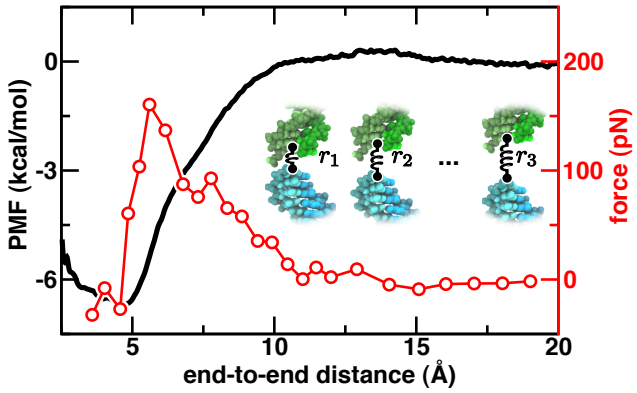

Figure 4.6: Representative dependence of the effective force (red) and the free energy (black) of two axially aligned DNA fragments on the end-to-end distance. The data result from 100 independent simulations of two 5′-phosphorylated fragments, the end-to-end distance of which was maintained at a specified value by a harmonic spring. Additional restrains were applied to maintain the axial alignment. The strength of the restraints was found to affect the values but not the general shape of both curves (see text). The image in the background illustrates the simulation method. The DNA fragments were immersed in 100 mM NaCl electrolyte, and were free to rotate about their helical axes.

and becomes slightly repulsive after ∼13 Å (2–10 pN). Thus, the end-to-end force has a large but very short-range attractive component caused by the hydrophobic effect and a much smaller long-range repulsive component that originates from screened electrostatic interactions between the DNA fragments [49].

A variation of the above protocol (described below in detail) was used to calculate the PMF for DNA fragments different by their terminal chemistry and relative azimuthal orientation and for several concentrations of the surrounding electrolyte. Table 4.1 lists the change in the depth of the PMF minima ($\Delta(\min[\mathrm{PMF}] - \mathrm{PMF}(\infty))$ relative to its value for the 5′-phosphorylated fragments at $\phi = 36°$ orientation in 100 mM NaCl. These calculations (detailed in Fig. S4 of the Supporting Information of Ref. 175) demonstrate that increasing the electrolyte concentration from 0.1 to 1 M has negligible effect on the PMF. Among the three most likely orientations that 5′-phosphorylated fragments form, the $\phi = -20°$ state has the lowest free energy, in good agreement with the $\phi = -22°$ angle frequently observed in the crystal structures of poly(dA·dT) oligonucleotides [197]. The 5′-phosphorylated fragments exhibit deeper minima than the blunt-ended fragments. For the latter, the conformation of $\phi = 180°$ is preferred over $\phi = 36°$. These variations in the depth of the PMF are consistent with the occupancy of bound states observed in our simulations of spontaneous collapse of aligned DNA fragments (see Fig. 4.1 d).

Table 4.1: Relative free energy change, $\Delta\Delta$G, upon formation of the end-to-end complex. Here, the free energy change ΔG is approximated by the minimum of the end-to-end PMF obtained from umbrella sampling simulations, Fig. S4 d of the Supporting Information of Ref. 175. The values of ΔG are given relative to the value measured for the system containing 5′-phosphorylated fragments end-joined in the 36° orientation at 100 mM NaCl. In each simulation, the relative azimuthal orientation of the DNA fragments was enforced using harmonic restraints (see Sec. 4.9.6). The application of such restraints introduced a bias to the estimates of ΔG. As all simulations employed the same restraints, this bias was assumed to cancel out in the calculation of $\Delta\Delta G$.

5′ chem.	**ion conc. (M)**	ϕ	**$\Delta\Delta$G (kcal/mol)**
phos.	0.1	36°	0
phos.	1.0	36°	0.4
phos.	1.0	−20°	−1.8
phos.	1.0	180°	0.1
hydrox.	1.0	36°	2.3
hydrox.	1.0	180°	1.2

4.6 Standard binding free energy of end-to-end association

To obtain a robust estimate of the free-energy of end-to-end association, we computed the standard free energy of binding G_{bind} using a variation of the method described previously [146]. For a system of two DNA fragments, G_{bind} determines the fraction of time the fragments form a bound state, which can be, in principle, observed directly in an all-atom MD simulation. However, the dissociation rate of the end-to-end assembly was found to exceed 200 ns (see Fig. 4.2 d) and therefore using such a brute force approach was not possible. Equivalently, G_{bind} can be obtained from the logarithm of the equilibrium binding constant, which is the ratio of the kinetic rates of end joining and rupture (k_{on} and k_{off}, respectively).

For our calculations of G_{bind}, we considered a process consisting of the following four steps (see Fig. S5 in the Supporting Information of Ref. 175). First, we evaluated the free energy cost of enforcing axial alignment restraints on a pair infinitely separated DNA fragments. Second, we computed the approximate cost of bringing a pair of DNA fragments from infinity to the maximum-separation state considered in our atomic simulations (CoM–CoM distance of 52 Å). Third, we determined the free energy change of bringing two axially aligned DNA fragments into an end-to-end bound complex. Finally, we evaluated the cost of releasing the axial alignment restraints from the end-to-end assembly. The sum of these terms yielded the free energy change upon formation of the end-to-end DNA complex $G_{\text{bind}} = -6.3 \pm 1$ kcal/mol for a DNA concentration of 1 M in 120 mM NaCl. Since a pair of DNA ends have only 1 binding configuration, we can express the standard binding free energy in terms of DNA ends, so that $G_{\text{bind}}^{\text{ends}} = -5.4$ kcal/mol. A complete description of the methods used in this section is provided in Sec. 3.1 of the Supporting Information of Ref. 175.

The above value for G_{bind} represents our best effort to quantify the strength of the end-to-end interaction.

In general, G_{bind} may depend on the DNA sequence. Although the question of sequence dependence is enticing, investigations of the sequence dependence of the end-to-end interaction have been deferred to future studies.

4.6.1 Estimate of rate of end-to-end dissociation

The escape rate for a Brownian particle trapped in a potential well can be estimated by finding the mean first passage time, τ, from the Smoluchowski equation, which describes the motion of a Brownian particle in a potential. It can be shown that for a particle with a 1-dimensional diffusion coefficient D [198],

$$\tau = \frac{1}{D} \int_{r_{\min}}^{r_{\max}} dy \, e^{W(y)/k_B T} \int_0^y dz \, e^{-W(z)/k_B T}, \tag{4.2}$$

where $r_{\min}$ is the minimum of the PMF, $W(r)$, and $r_{\max}$ is the height of the barrier. Since $r_{\max}$ is an inflection point, the rate of crossing, k_{off}, is 1/2 the rate of arrival $1/\tau$. This approach was used previously to obtain k_{off} from D and W, which in turn were obtained through MD simulations [199].

The value of k_{off} is sensitive to the location of the peak PMF, $r_{\max}$. Estimating that $D \sim 25$ Å^2/ns from trajectories of the end-to-end distance obtained in Sec. 4.2, the range of reasonable choices for the location of the maximum (10–15 Å) gives $k_{\text{off}}^{-1} \sim$ 170,000–860,000 ns. We estimate that $r_{\max}$ is at 12 Å, yielding $k_{\text{off}}^{-1} \sim$ 480,000 ns. However, the use of axial alignment restraints in our calculations of the PMF may have affected the above estimates of k_{off} by reducing the pathways available to the DNA fragments for escape. Thus, we expect the above numbers represent an upper bound estimate for k_{off}^{-1}.

4.7 Spontaneous assembly of long end-to-end aggregates

Our results until this point described the interaction of two DNA fragments in isolation. To simulate multi-fragment aggregation, 458 DNA fragments, each 10 bp in length, were placed in a cube of 100 mM NaCl solution (23.8 nm on each side) making the system shown in Fig. 4.7 a. During a 260-ns MD simulation, the DNA fragments diffused about their initial positions and interacted with their neighbors to form aggregates up to 11 DNA fragments (110 basepairs) in length. The DNA fragments that formed the longest 10 aggregates are shown at the beginning and end of the simulation, Fig. 4.7 b,c. The number of aggregates of a given length at three different instances of the MD trajectory is shown in Fig. 4.7 d. The plot reveals rapid growth of end-to-end aggregates and roughly exponential distribution of the lengths of the aggregates. The number of aggregates did not reach a steady state by the end of the simulation, see Fig. 4.8.

We model the process of end-to-end aggregation using a simplified reversible step-growth polymerization

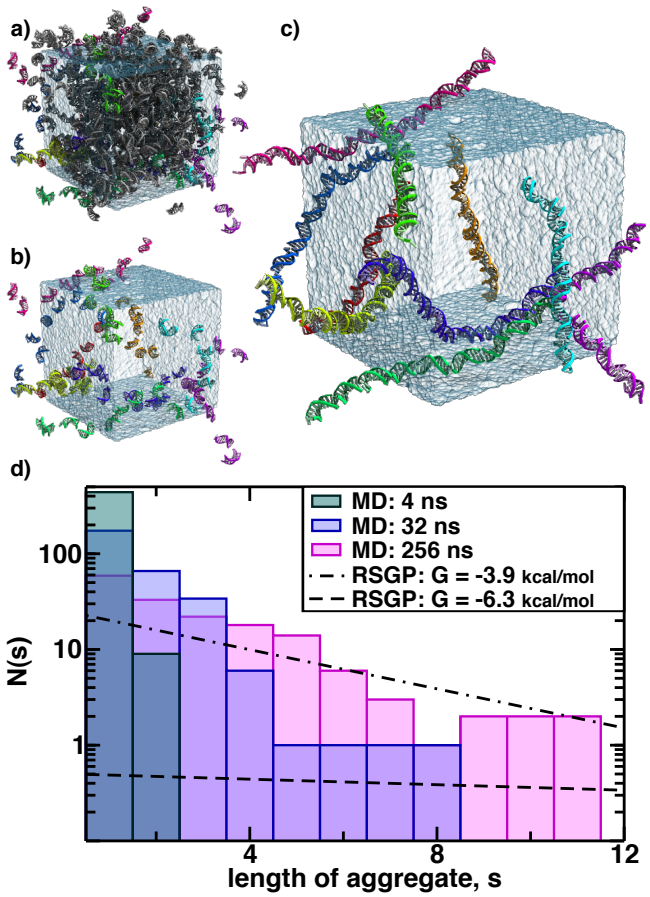

Figure 4.7: Spontaneous aggregation of duplex DNA into long rod-like structures. (a) A system containing 458 duplex DNA fragments placed at random. NaCl solution is shown as a semi-transparent molecular surface. (b-c) Initial (b) and final (c) conformations of the DNA fragments that composed the longest 10 aggregates at the end of a 260 ns MD simulation. DNA fragments forming each aggregate are shown in a different color. (d) The instantaneous number of aggregates, $N(s)$, formed by s DNA fragments in the MD simulation. The lines show the equilibrium distribution of the aggregate length according to the reverse step-growth polymerization model [200] for specified values of G_{bind}.

model, which has an analytical time-dependent solution that relates the mean aggregation number, $\langle N \rangle$ to G_{bind} as a function of DNA concentration, c, [200]. Here, we refer to a "monomer" as a single 10 bp DNA fragment and a "polymer" as a chain of end-to-end bound DNA fragments. For analysis purposes, a pair of DNA ends were considered bound if atoms from their terminal basepairs were well-formed and included atoms within 3.75 Å of one another. For a basepair to be considered well-formed we required that the distance between adenine N1 and thymine N3 atoms was 2.7–3.3 Å, averaged over 1 ns. Additionally, transient encounters and ruptures were eliminated by requiring DNA ends to be in contact for at least 2 ns before counting a binding event and by requiring bound DNA ends to be out of contact for at least 3 ns before counting a rupture. Despite these measures, a handful of false ruptures had to be visually examined before being disregarded for subsequent analysis.

Under the approximation that polymerization and depolymerization rates are independent of chain length, the time-dependent concentration of polymers starting from a solution of monomers can be described analytically [200]. Following the original derivation but under the simplification that no condensate product is produced by polymerization, we obtain solutions for the concentration, c_n, of polymers of length n: $c_n = x(t)y^{n-1}(t)$, with $x(t) = \lambda^2(t)/c_0$, $y(t) = 1 - \lambda(t)/c_0$, initial concentration of monomers, c_0, and the total concentration of polymers, λ. In turn, λ satisfies the 2$^{\text{nd}}$ order rate equation $\frac{d\lambda}{dt} = -k_{\text{on}}\lambda^2 + k_{\text{off}}(c_0 - \lambda)$. At equilibrium, $\lambda = \frac{1}{2K_{\text{eq}}}\left[\sqrt{1 + 4K_{\text{eq}}c_0} - 1\right]$, and, c_n can be expressed as an exponential function that decays with respect to n.

The reversible step-growth polymerization model yields a relation between the mean aggregation number, $\langle N \rangle$, and the equilibrium binding constant, K_{eq}, as a function of DNA concentration, c,

$$\langle N \rangle = \left[\ln \frac{1 + 2K_{\text{eq}}c - (1 + 4K_{\text{eq}}c)^{\frac{1}{2}}}{2K_{\text{eq}}c} \right]^{-1}. \tag{4.3}$$

Using our best estimate of the binding free energy $G_{\text{bind}} = -6.3$ kcal/mol, we obtain $\langle N \rangle = 39$ for the mean aggregation number at equilibrium for this system. The longest aggregate observed in the simulation was 11 DNA fragments, indicating that the simulation was still far from equilibrium.

The above simulation partially mimics the experimental assay of the Clark and Bellini groups [185], which involved very dense fluids of short DNA fragments. A transition from an isotropic phase to a nematic liquid crystal phase was found for 10 bp DNA fragments at $c = 875$ mg/ml ($\sim$150 mM). Through a combination of calibration experiments and theory, the authors estimated $\langle N \rangle = 9$ at the isotropic–nematic phase transition. Under the model of reversible step-growth polymerization, $\langle N \rangle = 9$ at $c = 150$ mM corresponds to $G_{\text{bind}} = -3.87$ kcal/mol. By contrast, using our prediction of $G_{\text{bind}} = -6.3$ kcal/mol yields

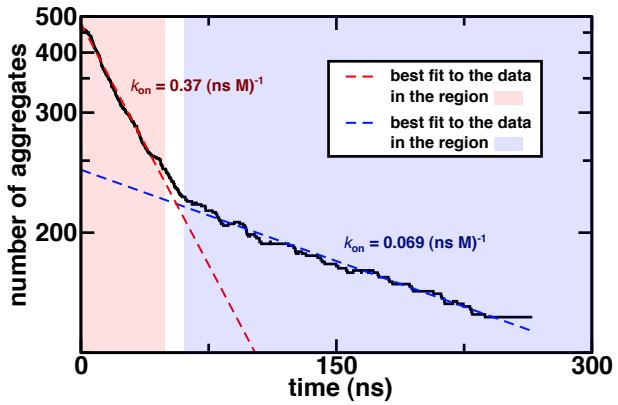

Figure 4.8: The number of end-to-end aggregates (including monomers) during a simulation starting with 458 DNA monomers in a cube of solvent. The reversible step-growth polymerization model predicts 2$^{\text{nd}}$ order kinetics for the total number of aggregates. We plot the best fit of this data in two regions: during the first 49.5 ns (red) and after 60.9 ns (blue). The fits provided estimates of k_{on} for the end-to-end association of two polymers.

$\langle N \rangle = 64$ at a DNA concentration of 150 mM.

4.7.1 Estimate of rates of end-to-end association and dissociation

In the reversible step-growth polymerization model, the total number of associated molecules is described by a second-order kinetic equation. Since unbinding of DNA fragments is negligible in our system, k_{on} can be extracted by fitting $N(t) = \frac{N_0}{1+tc_0k_{\text{on}}}$ to the data, where $N(t)$ is the total number of end-to-end bound DNA fragments at time t, $N_0 = N(0)$, and c_0 is the initial DNA concentration. In our simulation of spontaneous aggregation, the association rate reduced from $k_{\text{on}} = 0.37$ (nsM)$^{-1}$ observed within the first 50 ns to 0.069 (nsM)$^{-1}$ for the 60–250-ns interval, see Fig. 4.8. The change in association rate occurred as longer end-to-end aggregates formed (> 3 duplex fragments) and the rotational and translational diffusive motions of shorter aggregates slowed.

The aggregation simulation provides a lower bound estimate for the dissociation rate k_{off}. During the course of the simulation, 307 DNA ends were bound for 217 ns on average and unambiguous unbinding was observed for only one end-to-end associated complex. A few events were observed in which partial unbinding occurred via rupture of the Watson-Crick basepair of terminal nucleotides, as well as one instance where a bound DNA fragment was transferred to an unbound fragment; we neglect these events for the subsequent analysis. A Poisson distribution yields the probability that exactly one end-to-end bound DNA pair would rupture during the simulation given a value of k_{off}. For $k_{\text{off}} < 1/596,000$ ns^{-1}, it is likely that more than one rupture would have occurred. For $k_{\text{off}} > 1/19,000$ ns^{-1}, it is likely that no rupture would have occurred. Thus, the probability that exactly 1 rupture occurred is $> 10\%$ for $k_{\text{off}}^{-1} = 19,000$–$596,000$ ns. The greatest

probability of exactly 1 rupture occurring is 37% and is for $k_{\text{off}}^{-1} = 67,000$ ns. The ranges estimated for k_{on} and k_{off} suggest $-G_{\text{bind}}$ in between 4.4 and 7.6 kcal/mol, with a most probable range of 5.2–6.2 kcal/mol (k_{on} =0.069–0.37 (nsM)$^{-1}$ and $k_{\text{off}}^{-1} = 67,000$ ns).

The rate of dissociation of an end-to-end assembly, k_{off}, can also be estimated from the PMF and diffusion coefficient, D, by computing the mean first passage time [198, 199]. From the simulations performed in Sec. 4.2, we estimate $D \sim 25$ Å^2/ns. Due to uncertainty in the location of the barrier peak, we obtain the range $k_{\text{off}}^{-1} \sim$ 170,000–860,000 ns, with a best estimate of $k_{\text{off}}^{-1} \sim$ 480,000 ns. The use of axial alignment restraints in our calculations of the PMF limits pathways ordinarily available to rupturing DNA, so these values likely represent an upper bound for k_{off}^{-1}.

The mean first passage time can also be used to estimate k_{on}. Assuming that end-to-end binding is limited by translational diffusion, the upper bound estimate for k_{on} is $4\pi D R_0 \sim 7$ (ns M)$^{-1}$, where $R_0 = 37$ Å is the CoM distance between a pair of DNA fragments at which binding is expected to occur. The true value of k_{on} should be smaller because the DNA fragments must be axially aligned for binding to occur. Furthermore, long-range electrostatic repulsion may reduce the value of k_{on}. Our estimates of G_{bind} and k_{off} suggest a range for k_{on} of $\sim$ 0.03–0.16 (ns M)$^{-1}$.

4.8 Discussion

Stability of a dsDNA molecule can be conveniently described as a sum of base stacking and basepairing energies contributed by individual nucleotides that comprise the molecule. It is tempting to conceptually equate the base stacking interactions within a continuous molecule with the base stacking interactions that drive the assembly of two disjoint DNA duplexes. Thus, the unified nearest neighbor parameters, which can predict the energy for DNA hybridization based on DNA melting data, suggest $G_{\text{bind}} = -16.94 + 2 \times 6.94 = -3.06$ kcal/mol [93] for the association of two 10 bp poly(dA·dT) DNA fragments into a continuous 20-bp molecule. Such a simple calculation may be, however, flawed as additional conformational flexibility afforded by the lack of phosphodiester bonds at the end-to-end interface should allow the base-stacking geometry to be optimized, magnifying the base-stacking contribution to the free energy. Accordingly, the average interaction energy between adjacent basepairs with $\phi = -20°$ was measured to be 2 kcal/mol lower in bases forming the end-to-end junction than in bases in the middle of one of the DNA fragments (see Fig. 4.9). This finding is in good agreement with a survey of crystallographic structures that found DNA fragments terminating with AT basepairs to stack with $\phi = -20°$ [197]. We note that the stacking geometry may depend on the sequence of the DNA termini.

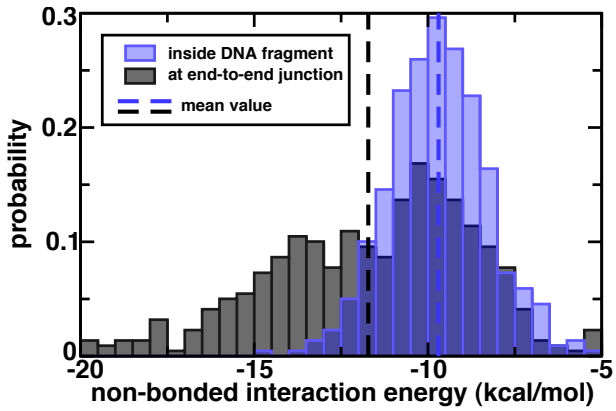

Figure 4.9: Distribution of the van der Waals and Coulomb interaction energy between two adjacent DNA basepairs forming an end-to-end junction (grey) and buried in the middle of one of the DNA fragments forming an end-to-end assembly (blue). The distributions were obtained from analysis of a trajectory of a end-to-end assembly formed from two 5′-phosphorylated 10-bp DNA fragments. The assembly was free to translate and tumble and had $\phi = -20°$ at the onset of simulation (red traces in Fig. 4.2 d,e). Only inter-nucleobase interactions were considered; the deoxyribose and nucleic backbone were excluded.

The base stacking free energy has been experimentally quantified by introducing a dangling nucleotide or a nick (a cut in the backbone of one strand) to a DNA molecule and observing the change in melting temperature [201–203], and by observing the mobility of a nicked DNA molecule relative to intact DNA and DNA with a gap [204, 205]. These experiments provide estimates for the base stacking free energy between -0.65 and -2.0 kcal/mol for stacks formed by thymine and adenine. However, the extraction of the free energy values is indirect with these methods.

We provided the first direct estimate of the standard binding free energy of end-to-end association of DNA fragments, obtaining $G_{\rm bind} = -6.3 \pm 1$ kcal/mol. The simulations reported here are similar to the liquid crystal condensation experiments of the Clark and Bellini groups [185], which estimated ~ -3.8 kcal/mol for the end-to-end free energy, in reasonable agreement with our estimate. The small-angle x-ray scattering experiments of the Pollack group demonstrated that the end-to-end interaction dominates over electrostatic repulsion in a divalent electrolyte [186, 187], which, unfortunately, is not sufficient to estimate the standard free energy of end-to-end binding. We note that the value of $G_{\rm bind}$ obtained in this study is larger than values reported in experiments. Having employed multiple methods, we are confident that the range of values obtained for $G_{\rm bind}$ accurately reflects the standard binding free energy within the limits of the molecular force field used in our study. Nevertheless, we cannot rule out the possibility that the present MD force field somewhat exaggerates the interactions driving end-to-end self-assembly of duplex DNA.

It is interesting to note that hydrophobic interactions between DNA bases and inorganic materials can be significantly stronger than the stacking energies observed in biochemical assays. AFM experiments indicate

that single-stranded DNA adheres to graphite with a free energy of -4.9 or -6.8 kcal/mol per nucleotide for cytosine and guanine, respectively [206]. For comparison, a computational investigation of hydrophobic interactions revealed a free energy of ~ -55 kcal/mol for the adhesion of a pair of 11×12 Å^2 graphene sheets, or about 10 times the energy per unit surface area [207]. Interactions of similar strength were found to promote DNA–fullerene and ssDNA–carbon nanotube association in all-atom MD simulations [208, 209].

Given the relatively large free energy of the end-to-end interaction, we pose the following question: why has end-to-end association only recently been observed? In most biological and nanotechnological settings, the concentration of DNA ends is too low for end-to-end association to be statistically significant, and the lifetime of end-to-end interactions ($\sim$ 70 μs) is somewhat shorter than the temporal resolution of many experimental techniques. To illustrate this point, we plot in Fig. 4.10 the fraction of bound DNA ends versus the concentration of free DNA ends in chemical equilibrium. The concentration of DNA ends is relatively high in DNA cyclization assays that measure the fraction of DNA molecules bent into a circle. The J-factor, which represents the concentration of one end in the proximity of the other, has a maximum value of $\sim 10^{-4}$ mM [210], which two orders of magnitude too small to promote cyclization of a significant fraction of blunt-ended DNA. The introduction of sticky ends increases the interaction energy by a few kcal/mol so that the DNA cyclization can be observed.

The large standard binding energy of end-to-end association implies that end-to-end interactions will be important in systems containing a high concentration of DNA ends. For instance, a remarkable new method of creating patterned, self-assembled structures out of DNA—termed DNA origami [76]—introduces many nicks along a path of DNA, which may enhance the stability of the resultant pattern. We speculate that broadened awareness of end-to-end association will influence the development of nanotechnologies where the end-to-end interaction can be used advantageously, such as with DNA origami, or can pose a limitation, such as for DNA microarrays. In cells, double-stranded DNA breakage, which poses a mortal threat, results in nearby blunt or sticky DNA ends. During non-homologous end-joining—the dominant repair pathway for such breaks in multicellular eukaryotes [211, 212]—the ruptured DNA ends are held together by proteins such as the Ku heterodimer or DNA-PK, or by nucleosome interactions until damaged DNA can be removed and ligation of the DNA backbone occurs [211]. Since DNA attracts end-to-end, it is not necessary that this complex, whose microscopic structure is not yet known, hold the DNA ends in strict alignment. It is sufficient for the ends to be held proximally so that the effective concentration of DNA ends is large enough to promote end-to-end association (Fig. 4.10 purple), whereupon ligation may occur. The free energy of dsDNA end-to-end association found in this study suggests that placing DNA ends in proximity of 3 nm or less will produce an end-to-end associated state with a probability of $\sim$ 95%.

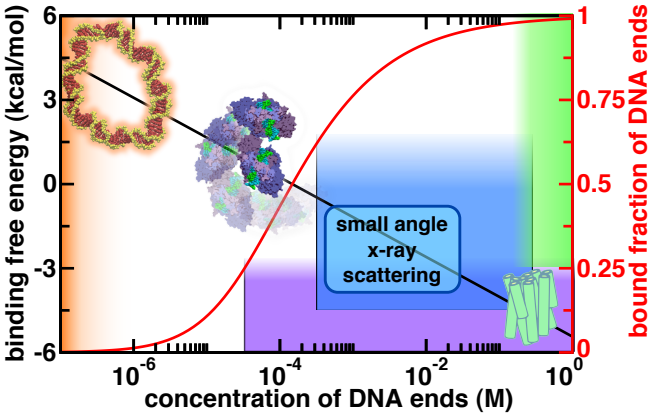

Figure 4.10: The effect of end-to-end attraction in different DNA systems. The binding free energy (black) and the fraction of bound DNA ends (red) are plotted against the reference concentration of DNA ends. Background images schematically illustrate four DNA systems in which the end-to-end attraction may or may not play a role. From top left to bottom right: the maximum local concentration of DNA ends (i.e. the J-factor [213]) is about two orders of magnitude too low to induce an observable fraction of blunt-ended DNA circles of any length (orange); translation and rotational confinement of DNA ends at dsDNA breakage (e.g. by the protein Ku (PDB:1JEY)) will promote binding of the DNA ends, which likely aids repair of DNA during non-homologous end joining [211] (purple); the structure factor obtained from small-angle x-ray scattering experiments of short DNA duplexes in a divalent electrolyte reveals end-to-end attraction [186] (blue); at very high DNA concentrations, long DNA aggregates form and align in liquid crystal phases [185] (green).

4.9 Simulation methods

All MD simulations were performed using the program NAMD [140], the parmbsc0 refinement of the AMBER-parm99 force field [136, 169], the TIP3P model of water [170], standard parameters for ions [171], periodic boundary conditions, particle-mesh Ewald (PME) full electrostatics with a PME grid density of about 1 Å per grid point. Except where specified, van der Waals and short-range electrostatic energies were calculated using a smooth (10–12 Å) cutoff, and integration was performed using 1–2–4 fs multiple timestepping [140]. The temperature was held constant using a Langevin thermostat [140] applied to all non-hydrogen atoms; the Langevin damping constant was set to 0.1 ps^{-1}. For simulations in the NPT ensemble, constant pressure was maintained at 1 bar using the Nosé-Hoover Langevin piston pressure control [174].

Each simulation reported in this study used one of the following three system types: elongated along the z-axis to minimize the amount of solvent around 2 DNA fragments ($\sim$24,000 atoms, Fig. 4.1 a); isotropic to allow 2 DNA fragments to tumble freely ($\sim$56,000 atoms, Fig. 4.2 a); and large and isotropic to allow unbiased interaction between 458 DNA fragments ($\sim$1.4 M atoms, Fig. 4.7 a). The DNA sequence was poly(dA·dT) in all systems. Counter-ions were added to each system to neutralize the DNA charge prior to the addition of a number of ions corresponding to the reported molarity (100 mM, except where specified) of NaCl electrolyte. Steric clashes that were introduced during the assembly of each system were removed from

each system through minimization using a conjugate gradient method [214]. Equilibration was performed in the NPT ensemble, and subsequent production simulations were performed in the NVT ensemble, except where specified.

4.9.1 Collapse of aligned dsDNA

Systems were built using the anisotropic unit cell. The dimensions of the system were chosen to provide a minimum of 2 nm distance between the surfaces of the DNA fragments across the periodic boundary, which should accommodate a majority of screening counterions for NaCl solutions of 100 mM or greater concentration (Debye length $\leq$ 1 nm). Steric clashes were removed through 3,000 minimization steps. Each system was subsequently equilibrated for 65 ps with the DNA backbone atoms harmonically restrained to their initial positions. Axial alignment of the DNA fragments was enforced by harmonically restraining each phosphorous atom of the DNA to the surface of an 11-Å-radius cylinder (with spring constants of 139 pN/nm per atom). The DNA fragments could translate along the long axis of the cylinder and rotate azimuthally. The starting conformation was characterized by a 20.5 Å end-to-end separation, which we define as the distance between the centers of mass of the nearest terminal basepairs, taking the periodic boundary condition into account. The end-to-end distance was projected along the common DNA axis, which we take as the elongated axis of the simulation system. The relative azimuthal angle ϕ of the terminal basepairs was defined as the angle between the projections of the vectors connecting the O5′ and O3′ atoms of the terminal basepairs into the plane normal to the common DNA axis. For two consecutive basepairs in a B-DNA helix, $\phi \approx 36°$, depending on the sequence.

4.9.2 Stability of the end-to-end complex

Systems were built by placing collapsed end-to-end DNA assemblies in an isotropic volume of 100 mM NaCl electrolyte solvent, Fig. 4.2 a. Steric clashes with the solvent were removed through minimization with the DNA backbone restrained. Subsequent simulation was performed in the NPT ensemble.

4.9.3 Mechanics of end-to-end dissociation

The coordinates of the system containing 5′-phosphorylated DNA fragments in a 100 mM electrolyte were taken after 140 ns of simulation in Sec. 4.3 to provide the initial conformation for simulations of rupture of the DNA fragments. The NVT ensemble was used during these simulations; Langevin thermostat was applied to water oxygens and ions. A harmonic spring of k_{s}=4000 pN/nm was used to produce the rupture,

by increasing its rest length at a rate of 0.4 or 0.2 Å/ns. The work performed at both rates were in good agreement.

4.9.4 Axial stretching of the end-to-end assembly

Force induced dissociation of the end-to-end DNA assembly was simulated using the steered molecular dynamics (SMD) method [151]. Figure 4.3 a schematically illustrates the process. Each DNA fragment was attached to one end of a virtual spring. The ends of the spring were anchored to the center of mass (CoM) of either the phosphorous atoms of the DNA fragments (6 simulations) or the CoM of the closest DNA ends (4 simulations). The rest length of the spring was increased at a constant rate of 0.4 or 0.2 Å/ns until the rupture of the assembly occurred. The work performed during each of these simulations was recorded and combined using the Jarzynski's equality [157] to estimate the potential of mean force (PMF) for the process considered:

$$e^{-\beta\,\Delta F} = \left\langle e^{-\beta\,W} \right\rangle \tag{4.4}$$

The angle brackets denote an ensemble average; β is $1/\mathrm{k_B T}$; ΔF is the change in free energy when the system is brought from one state to another; W is the work done during the change of state. No clear relationship between the pulling rate and the work performed was observed. We note that these simulations may not have sampled rare trajectories required for accurate estimation of the PMF using Jarzynski's equality [157].

4.9.5 Rupture of end-to-end assemblies containing complementary ssDNA overhangs

Five systems containing 20 DNA basepairs were built to determine the effect of complementary overhangs on the stability of the end-to-end DNA complex. The first system was a continuous 20-basepair (two-turn) helical dsDNA fragment immersed in 0.1 M aqueous solution of NaCl. The other four systems were built from the first system by cleaving the dsDNA into a pair of equal-length 5′-phosphorylated dsDNA fragments containing 0-, 1-, 2- or 4-nucleotide, complementary ssDNA tails, Fig. 4.5 a. Following usual minimization and equilibration protocols, the DNA assembly was axially stretched using the SMD method described in Sec. 4.4.1. The force of the virtual spring was applied to the CoM of the DNA fragments.

4.9.6 Potential of mean force of axially aligned DNA duplexes

Umbrella sampling simulations were performed using the anisotropic systems (Fig. 4.1 a) and two simulation protocols different by the method used to set up initial systems and the alignment restraints. Both protocols

enforced the end-to-end distance r using a harmonic spring of k_s=4000 pN/nm for $3.5 < r < 12$ Å in 0.5-Å intervals and k_s=1000 pN/nm for $13 < r < 19$ Å in 1.0-Å intervals. The first protocol was used to provided the estimate of the PMF plotted in Fig. 4.6. The initial conformation for each simulation was obtained by placing the DNA fragments a specified distance r apart at one of the four ϕ = 0, 90, 180, and 270° (four simulations for each r). The systems were equilibrated for 2 ns before data accumulation during production simulations lasting $\sim$16 ns. In the second protocol, which we used to compute the relative binding free energies, 4.1, the initial conformations were generated iteratively by shifting the minimum of the restraining potential in steps and followed by 0.5-ns equilibration, starting from the final frames obtained in the simulations of **Collapse of aligned dsDNA**. Subsequently, each system was equilibrated for at least 2.5 ns before accumulation of data during production simulations lasting 7.5–15 ns. Axial alignment was maintained as described for **Collapse of aligned dsDNA**, using k_s= 13.9 and 139 pN/nm for the first and second protocol, respectively. In the second protocol, a torque pointing along the common DNA axis was distributed among the phosphorous atoms of each DNA molecule to restrain ϕ about -20°, 36°, or 180°, with a spring constant of 219.4 pN nm/rad^2, which roughly corresponds to an 8° RMSF.

4.9.7 Spontaneous assembly of long end-to-end aggregates

The system depicted in Fig. 4.7 a was assembled through the sequential placement of 458 DNA fragments in a cubic volume (250 Å on each side). To place a DNA fragment, trial positions and orientations were randomly selected until the DNA coordinates did not clash with any previously placed fragments. During the first 50,000 ps of equilibration, the system shrank to its equilibrium size of 238 Å on each side. To improve computational efficiency, a 7–8 Å cutoff was used along with 2–2–6 fs time stepping scheme.

Chapter 5

A coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment

5.1 Introduction

It has become apparent that physical properties of DNA can play a fundamental role in its biological function and can determine the utility of DNA for nanotechnological applications [215–217]. Although the genetic code is stored in double-stranded form, it is the single-stranded form of DNA that plays active roles in central biological processes such as transcription, replication and DNA repair [218–221]. The emerging field of DNA nanotechnology exploits self-assembly of single-stranded DNA to create novel nanostructures [76, 222, 223] that can, for example, programatically transport cargo [224], arrange carbon nanotubes in a transistor-like device [178] and provide an arena for competition between different motor proteins [225]. Thus, single-stranded DNA (ssDNA) is ubiquitous in biology and (bio)nanotechnology, yet much less is known about its properties in comparison to double-stranded DNA (dsDNA).

Computer simulations can provide detailed insights into the structure, dynamics and energetics of a biological or nanotechnological system [51, 58, 226]. In this regard, all-atom molecular dynamics (MD) simulations that explicitly represent every atom of the system can offer the most detailed account of the system's inner workings [139, 227]. However, all-atom MD simulations are currently limited to timescales that are short in comparison to the relaxation timescales of even quite small ssDNA molecules [228]. By describing DNA using a less detailed, "coarse-grained" model the timescale accessible to simulation can be significantly expanded [79–81]. However, existing coarse-grained models of DNA have been foremost optimized to reproduce the properties of the double-stranded form [82–92].

5.2 Model architecture and parametrization

We have developed a simple, coarse-grained (CG) model of ssDNA that represents each nucleotide using two interaction sites, B (base) and P (backbone/phosphate) beads, shown schematically in Fig. 5.1. The interactions between beads are described through interaction potentials tabulated to accurately reproduce the conformations of ssDNA observed in all-atom simulations. The bonded potentials, which describe chemical links between the beads, includes pairwise interaction of each B bead with the two neighboring P beads, four three-bead angle terms, and three four-bead dihedral angle terms. The three non-bonded potentials describe interactions between one P and one B bead, two P beads, and two B beads using 1–3 exclusions:

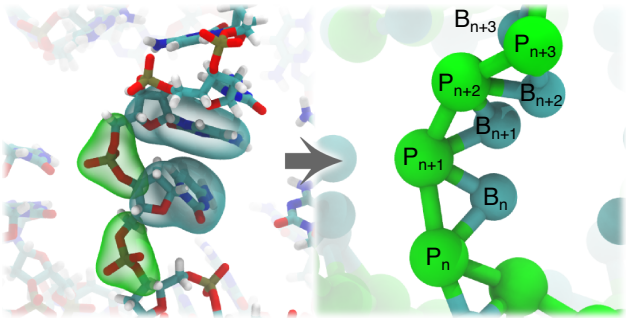

Figure 5.1: Scheme used to map atoms onto coarse-grained beads. The left panel shows a portion of an all-atom model of ssDNA. The backbone/phosphate and sugar/base atoms of two nucleotides are enclosed by green or cyan semitransparent surfaces, respectively. The enclosed groups of atoms are mapped onto the P (backbone/phosphate) and B (sugar/base) beads, which are shown in the right panel as green and cyan spheres, respectively.

the non-bonded potential is zero between beads separated by one or two bonds. The solvent surrounding DNA is modeled implicitly through a Langevin thermostat, see Sec. 2.1.3, and ion concentration-dependent non-bonded potentials. Figures 5.3, 5.4, 5.5 and 5.6 illustrate the CG potentials of our model.

5.2.1 All-atom simulations of ssDNA

To obtain CG potentials that are consistent with the all-atom model, we performed all-atom MD simulations of a dT_{60} strand submerged in a 80×80×80 $Å^3$ volume of 100 and 1000 mM NaCl electrolytes. The reported molarity was determined by counting the total number of chloride ions and water molecules in the system, not taking into account 59 cations added to neutralize the charge of the DNA strand. An ensemble of 21 simulations was performed for both 100 and 1000 mM systems, providing an aggregate simulation time of 6.2 and 3.4 μs, respectively. The initial conformations for the ensemble simulations were randomly chosen from a 1.3-μs trajectory of a dC_{60} molecule; the cytosine bases were mutated into thymines using the psfgen package. We chose not to present a CG model derived from the all-atom simulations of dC_{60} because of the following possible artifact of the all-atom model. For poly(dC), we observed hydrogen bonds formed by the amine groups of the cytosine bases and the phosphate groups of the neighboring nucleotides about 5% of the time. We have recently found that the CHARMM amine–phosphate interaction is overestimated [61]. Such persistent hydrogen bonding promoted relatively closed P_n–P_{n+1}–P_{n+2} angles (peak around 120°), which was not observed in poly(dT).

The resulting all-atom trajectories were converted into our CG representation (P and B beads). For a given nucleotide, the P bead represented the O5T′, O5′, P, O1P, O2P, and C5′ atoms of that nucleotide and the C3′ and O3′ atoms of the adjacent nucleotide such that the bead was roughly centered on the phosphorous group of the DNA backbone. The remaining atoms of the nucleotide were mapped onto the

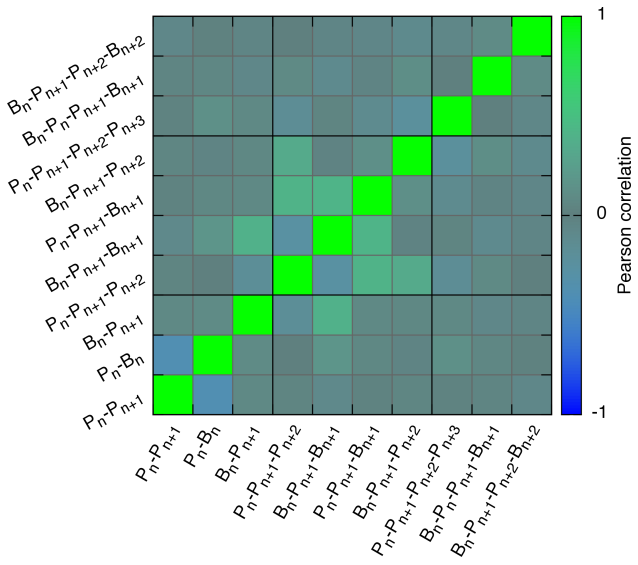

Figure 5.2: Pearson correlation for CG-mapped bonds, angles and dihedrals obtained from all-atom simulation. A value of 1 indicates perfect correlation, whereas a value of -1 indicates perfect anti-correlation.

B bead. For both types of beads, the conversion procedure was done by computing the center of mass of the respective groups of atoms; hydrogen atoms were neglected during the conversion procedure. The bonds, angles and dihedrals selected for our parameterization were fairly independent from one another (see Fig. 5.2), suggesting that our all-atom system can be approximated well by a CG model that applies potentials to these degrees of freedom.

5.2.2 Bonded interactions and potentials

An initial guess for each CG potential was obtained via Boltzmann inversion of the corresponding distribution extracted from the CG-mapped all-atom trajectory. The effect of ion concentration was taken into account by introducing two sets of CG potentials (for 100 and 1000 mM electrolytes) parametrized using the corresponding all-atom trajectories. Using a CG system identical in composition and size to the all-atom one, the bonded—but not non-bonded—potentials were refined by performing thirty iterations of the iterative Boltzmann inversion (IBI) procedure. In IBI, the CG potentials are iteratively adjusted until the distributions obtained from CG simulations converge to the target all-atom distributions, see Section 2.2.1.

Upon completion of the IBI refinement of bonded interactions, the distributions of CG beads corresponding to the bonded interactions (i.e., the distribution of bond lengths, angles and dihedrals) were in excellent agreement with equivalent distributions obtained by CG-mapping the reference all-atom simulations. Fig. 5.3 A-C show the distributions of the distances between P and B beads within the same nucleotide (P_n– B_n), adjacent P beads (P_n–P_{n+1}), and the B and P beads of adjacent nucleotides (B_n–P_{n+1}). The sub-

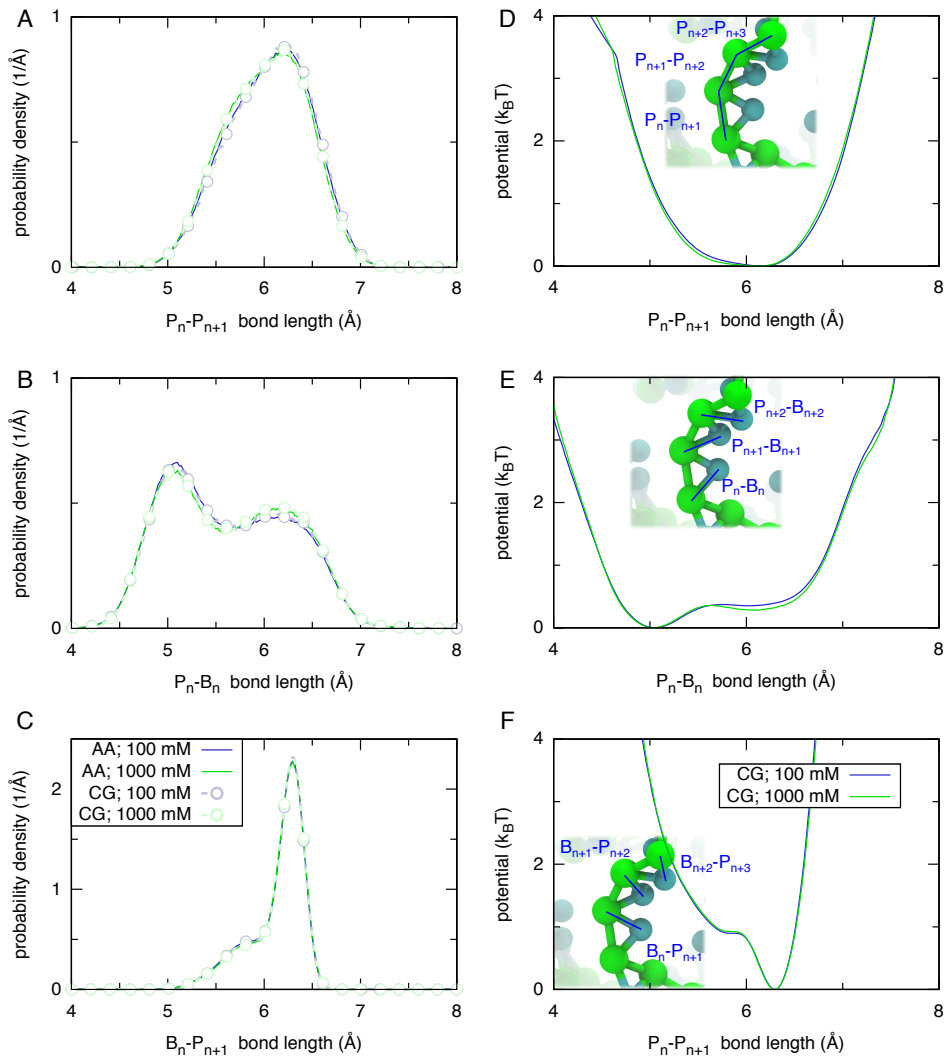

Figure 5.3: Parameterization of bonded interactions. (A-C) Distributions of bond lengths in all-atom and coarse-grained simulations of a dT_{60} molecule at 0.02 Å resolution. The solid lines depict the distributions obtained from all-atom simulations in 100 (blue) and 1000 (green) mM NaCl electrolytes. The open circles connected by dashed lines depict the distributions obtained from CG simulations at the 100 (light-blue) and 1000 (light-green) mM conditions. (D-F) The final CG bond length potentials corresponding to the 100 (blue) and 1000 (green) mM conditions. The inset of each panel highlights three bonds of the type characterized in that row of the figure. The image shows a CG representation of ssDNA molecule oriented to have its 5′-to-3′ direction pointing up in the figure.

script refers to the nucleotide number that increases in the 5′-to-3′ direction, see Fig. 5.1. The distributions obtained from the all-atom and CG models are in perfect agreement and are characterized by narrow (~ 1 Å width) peaks and a small skew. The distributions of the P_n–B_n and B_n–P_{n+1} bonds differ significantly, suggesting that a structurally precise CG model must apply different potentials to describe these bonds. In particular, the B_n bead is found, on average, closer to the P_n bead than to the P_{n+1} bead; the P_n–B_n distribution has broader peaks than the P_{n+1}–B_n one. By preserving the asymmetry of P_n–B_n and P_{n+1}–B_n bonds, the 5′-to-3′ directionality of the DNA strand is incorporated in our two-beads-per-nucleotide CG model.

Fig. 5.4 A-D shows the distributions of the P_n–P_{n+1}–P_{n+2}, B_n–P_{n+1}–B_{n+1}, P_n–P_{n+1}–B_{n+1}, and B_n–P_{n+1}–P_{n+2} angles considered in our model. The distributions are broad, spanning roughly 120°, and are poorly fit by single gaussians. The angle formed by three consecutive backbone beads (P_n–P_{n+1}–P_{n+2}) is likely to be around 140° regardless of the ion concentration, but with less likelihood at higher ion concentrations (presumably because greater electrostatic screening allows more "closed" conformations). Of the remaining angle distributions, only the angle between two consecutive bases (B_n–P_{n+1}–B_{n+1}) is significantly affected by the ion concentration. At 1000 mM, the likelihood of the B_n–P_{n+1}–B_{n+1} angle to be between 40 and 60° is reduced, compared to the 100 mM case, but enhanced in the 60 to 90° interval. The B_n–P_{n+1}–B_{n+1} angle appears to be anti-correlated with the backbone angle (P_n–P_{n+1}–P_{n+2}), Fig. 5.2, suggesting that the ion concentration dependence of the former could derive from the ion concentration dependence of the latter.

Finally, the distributions of dihedral angels P_n–P_{n+1}–P_{n+2}–P_{n+3}, B_n–P_n–P_{n+1}–B_{n+1}, and B_n–P_{n+1}–P_{n+2}–B_{n+2} are shown in Fig. 5.5A-C. In 100 mM solvent, the backbone dihedral (Fig. 5.5A) shows a weak preference for positive values over negative values, indicating a slight right-handed chirality along the 5′-to-3′ direction. In 1000 mM solvent, this chirality nearly vanishes, and a preference for backbone dihedral angles around zero appears. The B_n–P_n–P_{n+1}–B_{n+1} dihedral angle has peaks around 0 and 180° (Fig. 5.5B), corresponding to the DNA conformations where neighboring bases are on the same side and on opposite sides of the DNA backbone. By contrast the B_n–P_{n+1}–P_{n+2}–B_{n+2} dihedral angle has a single peak around zero (Fig. 5.5C), indicating a preference for bases that are separated by a single base in between to lie on the same side of the DNA backbone.

5.2.3 Non-bonded interactions and potentials

The 1–3 excluded radial pair distribution functions (PDFs) describing non-bonded interactions of two P, two B, and B and P type beads were obtained by first converting the all-atom MD trajectories into CG

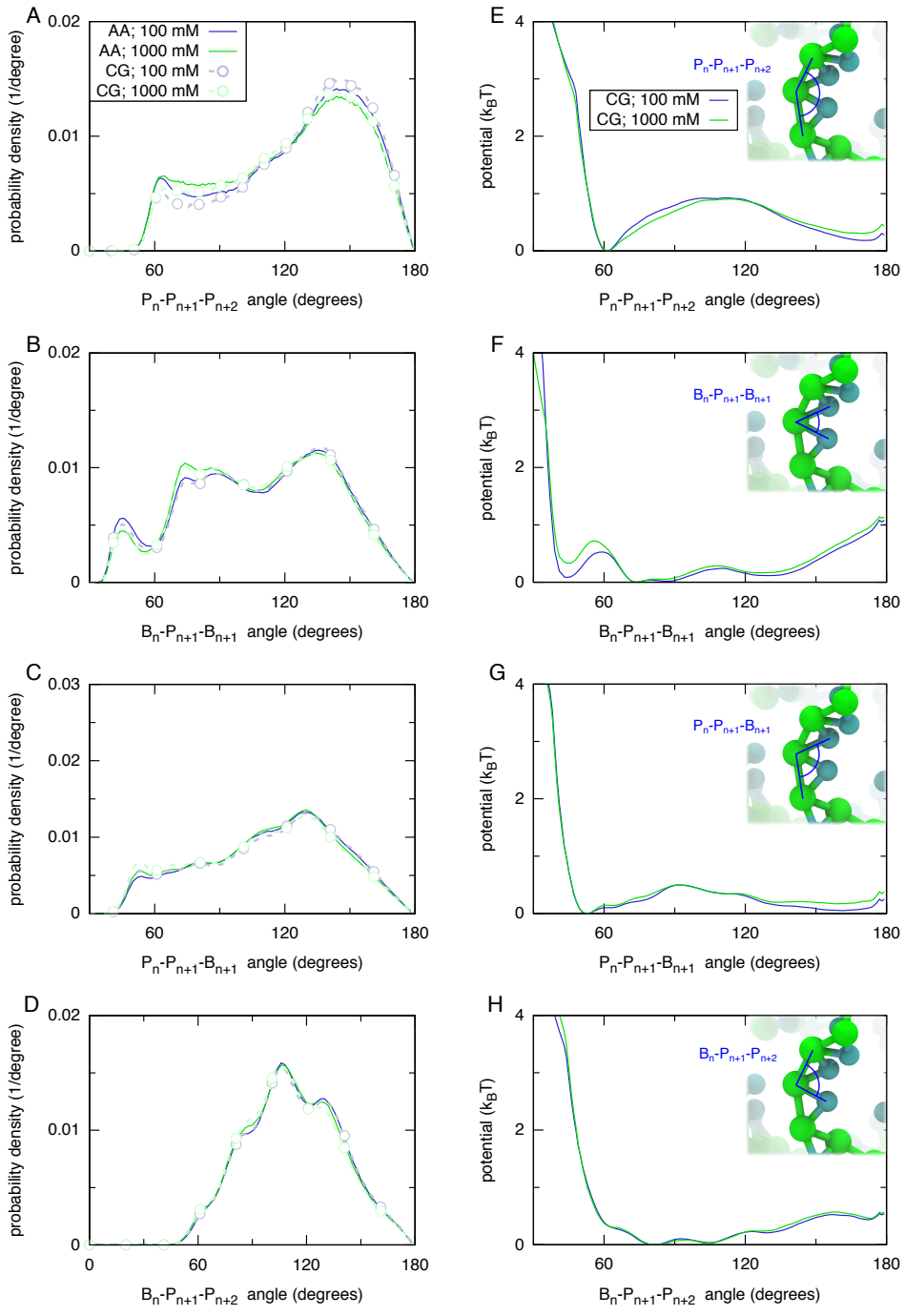

Figure 5.4: The distributions of angles (A-D) in all-atom and coarse-grained simulations of a dT_{60} molecule and the corresponding CG potentials (E-H) at 1 degree resolution. The inset of each panel in (E-H) illustrates the angle characterized in that row of the figure. Colors and inset images are as in Fig. 5.3.

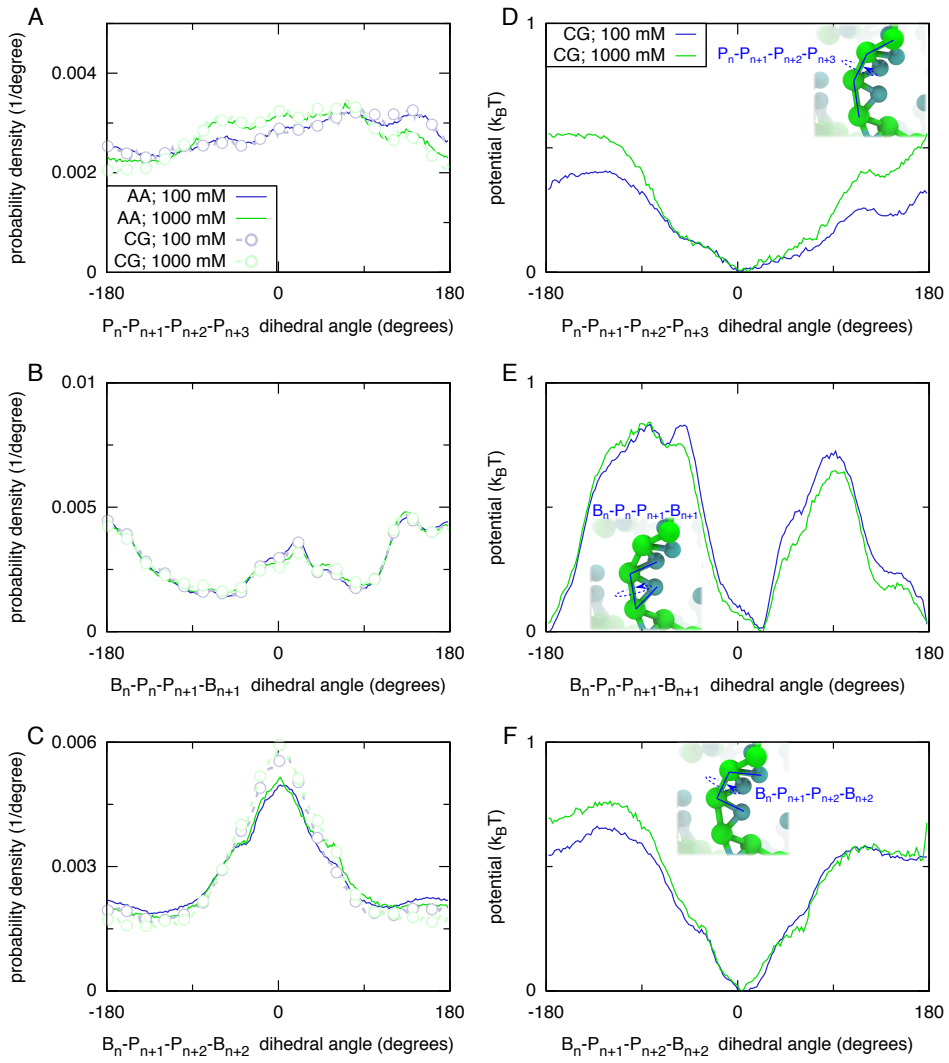

Figure 5.5: The distributions of dihedral angles (A-C) in all-atom and CG simulations of a dT_{60} molecule and the corresponding CG potentials (D-F) at 2 degree resolution. The inset of each panel in (D-F) illustrates the dihedral angle characterized in that row of the figure. Colors and inset images are as in Fig. 5.3.

representations (CG mapping) and computing the unnormalized PDF as $g_0(r) = \sum_{i,j} \frac{1}{4\pi r^2} \langle \delta(r_{i,j} - r) \rangle$. Here, 1–3 excluded means that the sum runs over only pairs of CG beads i, j that are more than two bonds apart. The angle brackets indicate averaging over MD trajectories. The normalized PDF was then produced using $g(r) = \frac{g_0(r)}{\int g_0(r')dr'}$. In practice, the PDF was calculated by binning the inter-bead distances at 0.05 Å resolution.

Non-bonded interaction potentials were refined through the IBI procedure (see Sec. 2.2.1) applied to the 1–3 excluded pair distribution functions of the P and B beads; the bonded potentials were kept fixed. The resulting CG model produced PDFs that were in excellent agreement with equivalent distributions obtained by CG-mapping the reference all-atom simulations; the agreement between bonded distributions was maintained. As shown in Fig. 5.6A-C, the PDF density for all bead types is nearly zero at small distances due to steric repulsion. The B–B bead distribution increases rapidly at around 4-Å separation due to stacking interactions between non-consecutive bases.

5.2.4 Refinement of non-bonded potentials against experimentally measured radius of gyration data

Having completed the "bottom-up" stage of parametrization, the radius of gyration—a representation of the size of a molecule—was determined for a dT_{60} molecule using the 100 and 1000 mM parametrizations of the CG model. The radius of gyration was found to be 89% and 75% of the experimentally measured values $R_{g,exp}$ (44.6 and 38.2 Å) [229]. The experimentally observed radius of gyration for a DNA molecule of a given length is taken to be the value given by the power law obtained by Sim *et al.* that best fits the experimental data [229]. This disagreement likely originates from the imperfections of the all-atom model but may also be caused by the finite size of the all-atom reference system.

Thus, we further refined the non-bonded potentials of our CG model until agreement with experimentally measured radii of gyrations was reached for the 100 mM and 1000 mM models in 25 and 14 steps, respectively. For this "top-down" refinement, the non-bonded potential describing the interaction of the P beads was systematically altered by adding or subtracting a Yukawa potential of the appropriate Debye-length value, see Sec. 1.4.

During each step of refinement, 12 copies of a CG dT_{60} molecule were each simulated in a cubic box 1000 Å on a side (preventing all interaction between periodic images) for at least an equivalent (see Sec. 5.3.2) of $\sim$1280 ns. At the end of the simulation, the average radius of gyration $R_{g,CG}$ was computed. The non-bonded P–P potential was adjusted by adding a scaled Yukawa potential $-k_BT \frac{R_{g,CG} - R_{g,exp}}{R_{g,exp}} \frac{1 \text{ nm}}{r} e^{-r/\lambda_{DH}}$, where r was the distance separating the beads and λ_{DH} was the Debye length at the appropriate ion concentration.

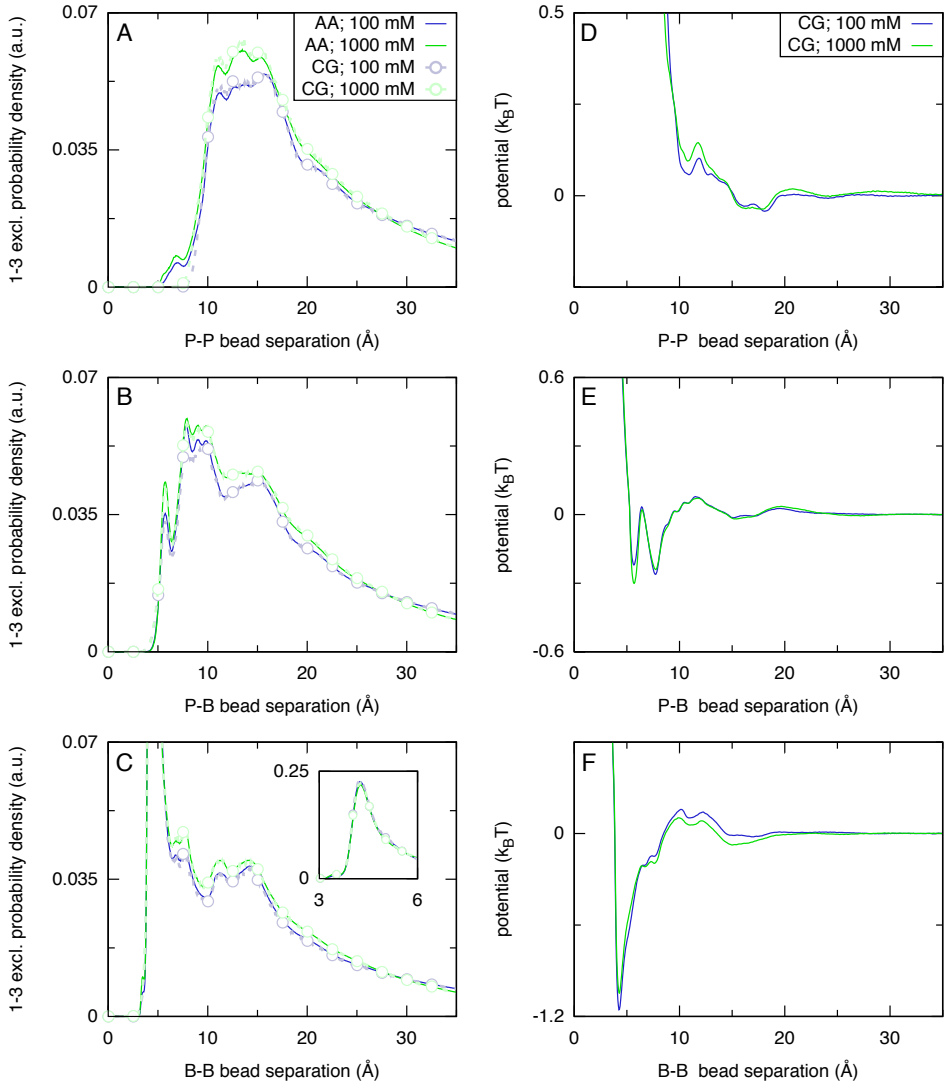

Figure 5.6: The normalized 1–3 excluded radial pair distribution functions (PDFs) (A-C) in all-atom and CG simulations of a dT_{60} molecule and the corresponding CG potentials (D-F) at 0.05 Å resolution. The inset in (C) shows the peaks of the PDFs at close separation for B–B beads. Colors are as in Fig. 5.3.

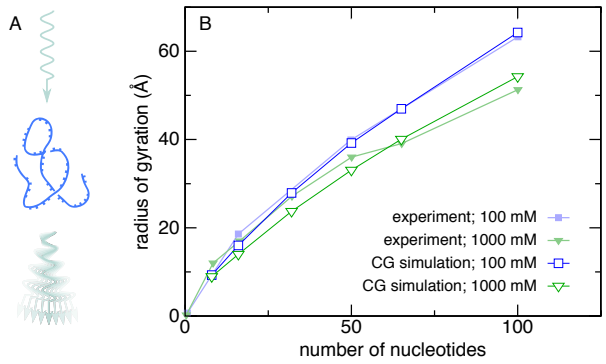

Figure 5.7: The radius of gyration, R_g, of a thymine homopolymer versus the homopolymer length. (A) Illustration of the experimental method, small-angle x-ray scattering, used to measure R_g. (B) Radius of gyration as a function of the polymer length. Good agreement was obtained between experimental measurements (open symbols) and coarse-grained simulation (filled symbols) under NaCl concentrations of 100-125 mM (blue squares) and 1000-1025 mM (green triangles). Experimental data was extracted from Ref 229.

The final potentials are shown in Fig. 5.6D-F. Upon completion of this procedure, the radius of gyration obtained in CG simulations agreed well with that measured in experiment across a broad range of polymer lengths and at both ion concentrations, see Fig. 5.7.

5.3 Results

5.3.1 CG simulations of the force-extension dependence

Without any further refinement of the model, the simulated force-extension dependence of a dT_{200} molecule was in excellent agreement with the experimentally measured dependence of one strand of λ-phage DNA (48,500 nts) under high applied force and similar ionic conditions [47], see Fig. 5.8. Similar simulations were performed using the "top-down" CG models of de Pablo (3SPN.2) [82] and Louis (base-average oxDNA) [230] groups, see the end of this section for details of these simulations. Our model, which was optimized specifically for ssDNA, performs extremely well for dT_{200} if compared to the 3SPN.2 model. Comparison with the oxDNA model isn't entirely possible because that model was parametrized for 500 mM monovalent electrolyte. In the high force regime (above 20 pN), where electrolyte conditions are not expected to influence the extension of ssDNA [231], the oxDNA model fits the experimental data well.

At forces below $\sim$ 10 pN, secondary structures form in λ-phage DNA, reducing its extension if compared to poly(dT). In one experimental study, glyoxal was used to chemically denature λ-phage DNA, allowing the authors of that study to probe the low-force extension of ssDNA in the absence of secondary structure formation [231]. Unfortunately, the denaturation process may have introduced chemical crosslinks so that

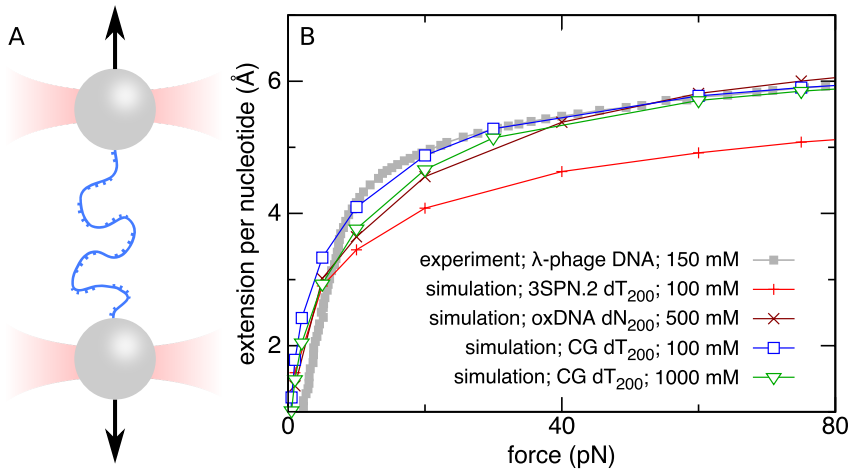

Figure 5.8: Simulated and measured force-extension dependence of ssDNA. (A) Illustration of the experimental method used to apply the force and measure the extension. Double-stranded λ-phage DNA (48.5 kbp) was caught between two beads in a dual optical trap. Melting and washing off the complementary DNA strand allowed the force-extension curve of the remaining strand to be determined. (B) Force-extension curve of ssDNA obtained through CG simulations and experiment. Data obtained using our CG model (open symbols) are in good agreement with experiment [47] (filled symbols) at high force and both NaCl concentrations. Quantitative comparison at low force is not possible because of the secondary structure of λ-phage DNA that shortens DNA extension in experiment. For comparison, we present force-extension dependence of dT_{200} in 100 mM electrolyte obtained using the 3SPN.2 model [82] and of 200 base-average nucleotides in 500 mM electrolyte obtained using the oxDNA model [230].

the absolute length of the ssDNA molecule was unknown, precluding comparison of absolute extension per nucleotide. Nevertheless, after dividing the extension values by the extension at 20 pN for each measurement, the simulated extension of dT_{200} was in good agreement with the experimentally measured extension across two orders of magnitude of the applied force, see Fig. 5.9. The 3SPN.2 and oxDNA models also agree well with the rescaled force-extension curves.

The simulations of 3SPN.2 DNA were performed using LAMMPS [232]. The simulations of oxDNA were performed using the oxDNA software [230]. One simulation was performed for each data point reported in Figs. 5.8,5.9. Forces of the same magnitude but opposite directions were applied to the ends of the DNA molecule, stretching the latter. Projection of the end-to-end distance on the axis defined by the direction of the applied forces was recorded. The average extension was obtained by discarding the initial 50 ns of the trajectory required to reach a steady state and averaging over the remaining frames of the trajectory, which typically covered at least 1,000 ns. For most data points shown in Figs. 5.8,5.9, the error bars are smaller than the symbols.

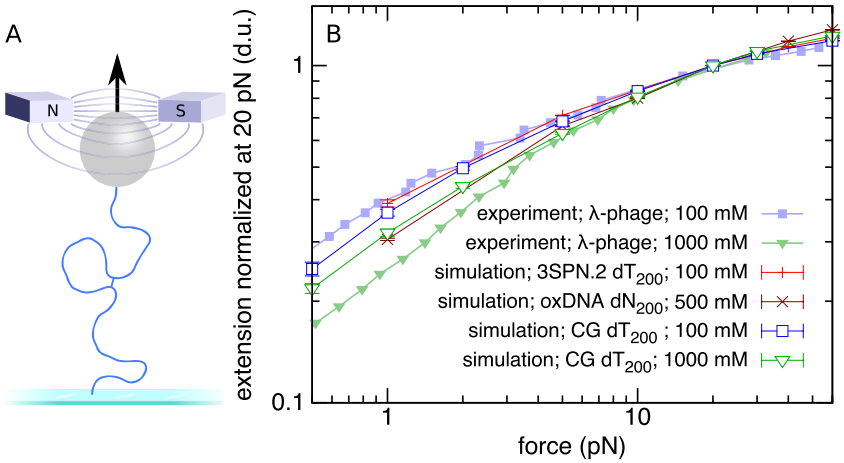

Figure 5.9: Simulated and measured force-extension dependence of ssDNA. (A) Illustration of the experimental method used to apply the force and measure the extension. Chemically denatured λ-phage DNA was stretched between a glass slide and a magnetic bead. The denaturant, which prevents formation of secondary structure, also introduced crosslinks between parts of the DNA, which made determination of the absolute extension not possible. (B) Force-extension curve of ssDNA obtained through CG simulations and experiment. Good agreement was observed between experimental measurements [231] (filled symbols) and coarse-grained simulations performed using our model (open symbols), the 3SPN.2 model [82] in 100 mM electrolyte and the oxDNA model [230] in 500 mM electrolyte.

5.3.2 The time scale of CG simulations

Coarse-graining, in general, involves smoothing out many degrees of freedom. As a result, processes tend to occur faster in a CG simulation than prescribed by its nominal time steps [233]. In our particular model, the effect of solvent is represented solely through the Langevin thermostat that adds independent random forces and viscous drag forces to each CG bead. Hence, the polymers described by our model exhibit Rouse-like dynamics [234]. For example, the diffusion coefficient of a molecule scales as $1/N$, where N is the number of nucleotides. Experimentally, the diffusion of single-stranded DNA molecules is observed to scale as $1/\sqrt{N}$ [235], consistent with the Rouse-Zimm polymer model, which considers hydrodynamic interactions between beads [236]. Unfortunately, hydrodynamic interactions are long-range, and rather expensive to calculate, even approximately [237]. Support for calculating hydrodynamic interactions is lacking in many popular molecular dynamics codes. Accordingly, hydrodynamic interactions are not present in our model, and a chain-length-dependent deviation from experimental timescales can be expected.

Many of the simulations in this study were performed using a dT_{60} molecule. The most important timescales with respect to sampling are the internal dynamics of the molecule. Experimentally, the timescale of end-to-end collisions for a 20 nt DNA fragment was measured to be 800 ns and was found to scale as $N^{3.5}$ with the length of the fragment [228]. From our simulations, the timescale of end-to-end collision for a dT_{20}

molecule was estimated to be range between 2 and 20 ns, depending on the definition of what constitutes a collision event (the end-to-end distance range of 15–25 Å). Furthermore, we found the collision timescale to scale roughly as $N^{2.8}$ for molecules ranging from 8 to 100 nt. Thus, for a dT_{60} molecule, the end-to-end collision dynamics should be enhanced by a factor of roughly 80. We take this factor into account when we report the CG simulation times.

5.3.3 Stretching of dT_{14} using combination of optical tweezers and smFRET

The use of long, mixed-sequence DNA molecules in previous experimental studies of ssDNA elasticity [47,231] complicates direct comparison with the simulation data. To validate our CG model for very short, chemically unmodified DNA fragments we turned to advanced single molecule techniques. Specifically, we used fluorescence resonance energy transfer (FRET) detection to measure the extension of a dT_{14} molecule under tension applied by an optical trap [238], see Fig. 5.10A. In this assay, the DNA construct was immobilized on a polyethylene glycol (PEG) coated glass slide at one end using the biotin–neutravidin interaction. The other end of the construct was connected to a micron polysterene bead via a λ-phage DNA linker. The preparation of the DNA construct is described in detail in Section 5.5.5. The bead was optically trapped, putting the DNA construct under tension. A pair of dyes (Cy3 and Cy5) was attached at the two ends of the dT_{14} fragment to provide a FRET signal that effectively allowed the end-to-end distance to be monitored as a function of the applied force.

Fig. 5.10B shows the FRET vs. force curves obtained at two different salt conditions as described in Section 5.5.5. As the tension increases, the FRET signal decreases, indicating extension of the dT_{14} fragment of the construct. In the low-force regime, the FRET values depend on the ionic conditions but converge in the high-force regime (> 10 pN). The low-force FRET is larger under high NaCl concentration, implying greater compaction of ssDNA caused by stronger electrostatic screening. This observation is consistent with our earlier work [239] and with the observed shrinking of ssDNA under high salt conditions [229]. At high force, the FRET curves converge as the extension of the polymer approaches its contour length. The FRET efficiency computed from our CG simulations of dT_{14} under tension is in good agreement with the experimental FRET traces at 100 and 1000 mM NaCl, see Fig. 5.10B.

5.3.4 Preliminary dsDNA model

A toy model of double-stranded DNA was constructed from the CG ssDNA model by adding a set of harmonic potentials between beads involved in base-pairing. Specifically, restrained bonds included the base paring term P_i–P_j ($r_0 = 7.8$ Å; $k = 10$ kcal/mol Å^2), and cross-stacking terms P_i–P_{j-1} ($r_0 = 8.1$ Å) and

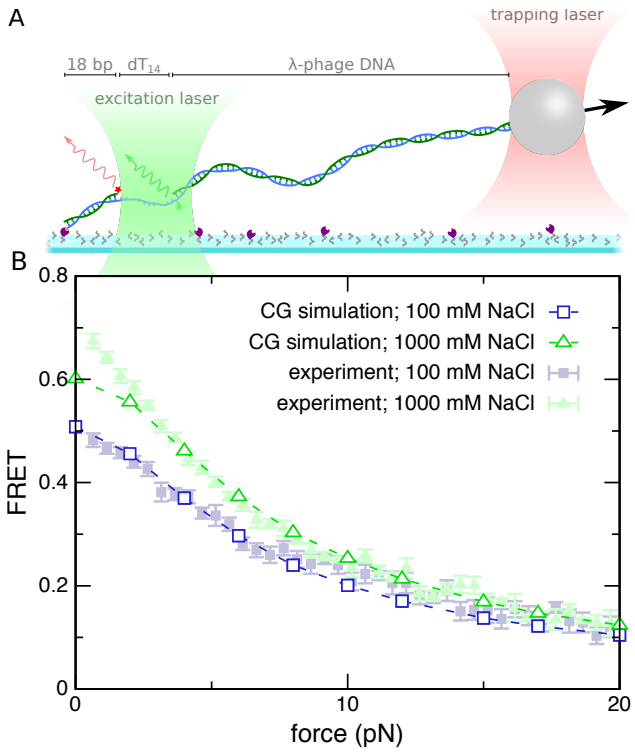

Figure 5.10: Force-extension dependence of dT_{14}. (A) Illustration of the experimental method used to simultaneously stretch DNA and measure the FRET signal. A green laser excites the Cy3 donor dye; some energy is non-radiatively transferred to the Cy5 acceptor dye. The amount of energy transferred, the FRET efficiency, is related to the distance between the dyes. An optical trap applies tension. (B) FRET efficiency vs force observed in experiment (filled symbols) and calculated from the CG simulations of dT_{14} under tension (open symbols). The following expression was used to compute FRET based on the distance r between the terminal P beads: $\langle\ 1/(1+(\frac{r+\delta}{R_0})^6)\ \rangle$, where $R_0 = 60$ Å is the Förster distance and $\delta = 22$ Å is a constant factor associated with the physical dimensions of the dyes. The angle brackets represent an average over the simulation trajectory. The agreement between simulation and experiment was good at 100 (blue squares) and 1000 mM (green triangles).

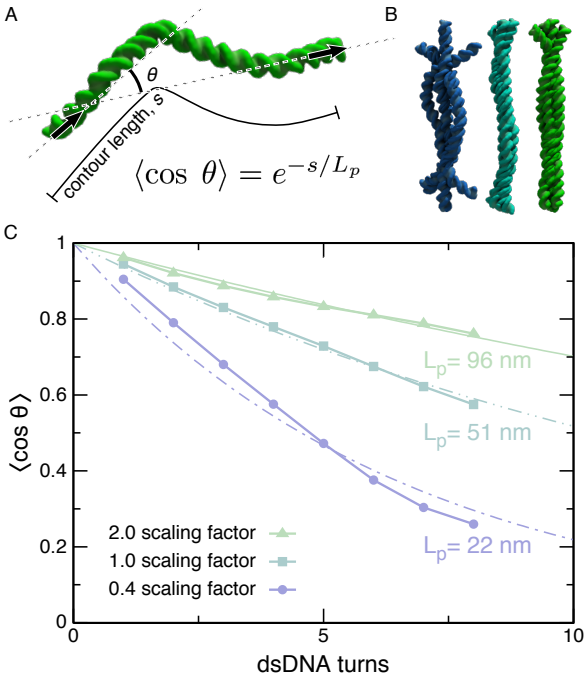

Figure 5.11: Custom flexibility CG model of dsDNA. (A) Definition of the persistence length L_p. The persistence length is a measure of a polymer's flexibility that can be directly derived from a simulation trajectory. (B) Representative conformations of a 100-bp dsDNA fragment having different parameterization of inter-strand interactions (a = 0.4; blue, 1.0; teal and 2.0; green). In our dsDNA model, parameter a defines the strength of inter-strand interactions. Individual DNA strands are described using our CG model of ssDNA. (C) Angular correlation of two DNA fragments versus distance between the fragments. An exponential fit reveals the persistence length. Colors are as in (B). The flexibility of dsDNA can be controlled by changing the scaling factor a.

P_i–P_{j+1} ($r_0 = 8.0$ Å) with $k = a$ kcal/mol Å^2, where r_0 is the rest length, k is the spring constant, a is the scaling factor used to tune the persistence length of the model, j is taken to be the pair of nucleotide i, and the subscripts increase along the 5′-to-3′ direction. Restrained angles included backbone restraints P_i–P_{i+1}–P_{i+2} and P_j–P_{j+1}–P_{j+2} ($\theta_0 = 150°$), additional intra-strand restraints P_i–P_{i+1}–B_{i+1} and P_j–P_{j+1}–B_{j+1} ($\theta_0 = 87°$), and inter-strand restraints P_i–B_i–B_j and P_j–B_j–B_i ($\theta_0 = 162°$), where θ_0 is the rest angle of the restraint and all angle restraints employed the spring constant $k = 90a$ kcal/mol degree^2. The rest lengths of all potentials were obtained from canonical poly(dT)·poly(dA). By changing the scaling factor a, the double-stranded DNA could be made stiffer or more flexible, see Fig. 5.11. Using a scaling factor of $a = 1$ was found to produce dsDNA with approximately the correct persistence length.

5.4 Conclusion

From a practical perspective, our CG model permits microsecond-per-day simulations of hundreds of nucleotides on a single processor core. One should note that dynamics is usually enhanced in coarse-grained simulations compared to all-atom, for example, due to smoothing of the free energy landscape [233]. For dT_{60} using our model, each CG nanosecond corresponds to ~ 80 real-world nanoseconds. However, we found that the enhancement for a DNA molecule depends on its length, see Section 5.3.4 for details. The only non-standard features of an MD code required to perform CG simulations using our model are tabulated bonded and non-bonded potentials. It must also be possible to apply bonds by bead index rather than bead type as the P_n–B_n and B_n–P_{n+1} potentials differ.

The most significant limitations of our model are three-fold. First, the model is currently limited to simulations of poly(dT). Second, the base beads are spheres that lack orientation that may be important for accurate modeling of base-pairing and base-stacking in duplex DNA. However, anisotropic base–base interactions require additional computation and reduce portability of the model. Finally, our model lacks a description of hydrodynamic interactions, which makes interpretation of kinetic information difficult, see Section 5.3.2 for details.

Despite its simplicity, our CG model provides a structurally accurate portrait of a poly(dT) molecule across a wide range of polymer lengths, applied tensions and ion concentrations. This makes our model immediately suitable for CG studies of ssDNA systems where sequence-specific effects, such as strong adenosine stacking [240], can be neglected. For example, we have already used a preliminary version of our model to study the effect of local heating on the process of ssDNA transport through a solid-state nanopore [241].

With only two-sites per nucleotide, extensions to the model can be easily made. For example, it was trivial to create a toy model of double-stranded DNA from the CG ssDNA model by adding a set of harmonic potentials between beads involved in base-pairing, see Section 5.3.4 for details. Other extensions are also possible, for example by representing proteins using a grid-based potential, one may study the interaction between a CG DNA molecule and a DNA-binding protein. Our simple, computationally efficient, yet accurate model of ssDNA is a step toward a complete physical model of the DNA processing machinery of a living cell.

5.5 Methods

5.5.1 All atom simulations of ssDNA

Our all-atom MD simulations were performed using the program NAMD [140], the CHARMM36 force field [137, 242–244] the TIP3P model of water [170], standard parameters for ions [171], with NBFIX corrections applied to ion–nucleic acid interactions [61]. Van der Waals energies and short-range electrostatics were calculated using a smooth (7–8 Å) cutoff. Periodic boundary conditions were employed and long-range electrostatics were calculated using the particle-mesh Ewald [245] (PME) method over a 1.2 Å per grid PME array. Integration was performed using 2–2–6 fs multiple timestepping [172]. To enable 2-fs timestepping for bonded interactions, water bonds (and angles) and non-water covalent bonds with hydrogens were held rigid using the SETTLE [246] and RATTLE [247] algorithms, respectively. Steric clashes that were introduced during the assembly of the systems were removed through minimization using a conjugate gradient method. Subsequent simulations were performed in the NPT ensemble. A temperature of 291 K was maintained by applying Langevin forces [173] to all non-hydrogen atoms (1 ps^{-1} damping coefficient). A pressure of 1 bar was maintained by Nosé-Hoover Langevin piston pressure control [174].

5.5.2 CG simulations of ssDNA

All CG simulations were carried out using a custom version of the MD program NAMD [140]. The custom version of the code allowed the use of tabulated potentials for the description of bond, angle, and dihedral interactions (resolutions of 0.02 Å, 1°, and 2°, respectively). Piecewise cubic (Catmull-Rom) interpolation was used to calculate the forces due to such potentials at each timestep. Non-bonded interactions were calculated using tabulated non-bonded potentials (0.05 Å resolution) and a smooth (34–35 Å) cutoff. Stochastic forces from the solvent were mimicked via a Langevin thermostat set to a temperature of 291 K and a damping coefficient of 1.24 ps^{-1}. The mass of each CG bead was set to the mass of the atoms it was designed to represent (the mapping procedure is described in the main text): 160.1 and 181.1 Da for the P and B beads, respectively. Integration was performed using a 20-fs timestep.

5.5.3 Refinement of bonded potentials

An initial guess for each CG potential was obtained through Boltzmann inversion of the corresponding all-atom (target) distribution followed by a triangular smoothing filter that was applied over 25, 11, and 5 points for bonds, angles and dihedrals, respectively, to mitigate the effects of noise.

The target distributions for bonds and angles had regions of zero values. Noise in the distributions

bordering these empty regions can pose a significant impediment to the convergence of the refinement of the CG potentials. Therefore, regions where the density fell below 10^{-3} Å^{-1} or 10^{-5} degree^{-1} were replaced by a repulsive half-harmonic potential $u_{\text{wall}}(x) = U_0(x - x_{\text{wall}})^2 \times m_{\text{wall}}(x - x_{\text{wall}}) + U_{\text{wall}}$, where x_{wall} was the position where the density fell below the threshold, and the gradient and potential at x_{wall} were given by m_{wall} and U_{wall}, respectively. U_0 was k_BT Å^{-2} for bonds and $\frac{k_BT}{180}$ degree^{-2} for angles.

Each IBI iteration was performed as follows. Within each iteration, six CG simulations of dT_{60} were performed in parallel to obtain six trajectories of approximately 1600 ns each (taking into account the factor of 80 speedup described below). Periodic boundary conditions were applied in the CG simulations according to the dimensions of the all-atom model. While keeping the non-bonded potentials fixed, bonded potentials were updated as $U_{n+1} = U_n - \alpha k_BT \log(\rho_{\text{cg}}(x)/\rho_{\text{target}}(x))$, using the bonded distributions $\rho_{\text{cg}}(x)$ extracted from the latest ensemble of simulations. The scaling factor α was set to 0.25 for angle and 0.5 for bond and dihedral potentials to ensure gradual convergence. The $\log(\rho_{\text{cg}}(x)/\rho_{\text{target}}(x))$ term was smoothed before it was added to the previous-generation potential U_n using the same smoothing filters as described above. The repulsive half-harmonic potentials at the boundaries of zero-density regions were updated. The final conformations of the dT_{60} molecules from the ensemble of simulations were used as initial conformations for the next iteration. After thirty such iterations, the bonded potentials (Figures 5.3D-F, 5.4E-H, 5.5D-F) produced CG distributions (not depicted) in very close agreement with the all-atom distributions. The CG distributions shown in Figures 5.3, 5.4, 5.5 were taken from the final model, after refinement of both bonded and non-bonded potentials.

5.5.4 Refinement of non-bonded potentials against all-atom MD data

The initial non-bonded potentials were taken to be flat except in the regions where the normalized PDF was below 0.02 Å^{-1}. In these regions, the potential was taken to be $u_{\text{wall}}(x) = U_0(x - x_{\text{wall}})^2 \times m_{\text{wall}}(x - x_{\text{wall}}) + U_{\text{wall}}$. This approach contrasts with the usual IBI procedure, in which Boltzmann inversion of the radial distribution function provides the initial potential. We found, however, that using a flat initial potential was essential to ensure fast convergence of the iterative refinement. The difference in approach likely originates from the history of IBI that was commonly used as a tool to study polymer melts [248]. Here, we adapt this approach to obtain potentials for a polymer in a dilute solution regime.

Iterative refinement of non-bonded potentials was performed exactly as in the case of bonded potentials, except that the bonded potentials were held fixed while the non-bonded potentials were varied. The (intra-chain) 1–3 excluded radial pair distribution functions (PDF), obtained for P–P, B–B, and B–P beads provided the density needed to update each non-bonded potential according to the IBI formula. Each up-

date was smoothed by a triangular filter (10 points) before being added to the previous generation potential. Parameters of the harmonic wall that approximated steric repulsion at small distances were updated in each iteration.

For IBI refinement of the 100 mM model, all three non-bonded interactions were updated simultaneously. The scaling factor α was set to 0.3 for the first 44 iterations but was reduced to 0.05 for the final 24 iterations. Each iteration included 24 replicas; each replica was simulated for an equivalent of 1280 ns.

For the IBI refinement of the 1000 mM model, a slightly different procedure was adopted in an attempt to hasten convergence. With a value of 0.5 for the scaling factor, the B–B interaction was first refined for 14 iterations while all other potentials were held fixed. Then the B–P interaction was refined for 15 iterations while all other potentials (including B–B) were held fixed. Next the P–P interaction was refined for 5 iterations with the scaling factor set to 0.5, and then for 10 iterations with the scaling factor reduced to 0.1. Finally, all three non-bonded interactions were simultaneously refined for 120 iterations with the reduced scaling factor. Each iteration included 12 replicas; each replica was simulated for an equivalent of ~ 2500 ns.

At the end of these simulations, the non-bonded potentials (not depicted) provided CG PDFs (Fig. 5.6A-C) that agreed well with the all-atom PDFs. Good agreement between all-atom and CG distributions of bonds, angels and dihedrals was maintained.

5.5.5 Preparation and execution of fleezers measurement

The DNA construct was made from two DNA oligos 5′–/5Phos/GGG CGG CGA CCT T /iAmMC6T/T TTT TTT TTT TTT GCC TCG CTG CCG TCG CCA and 5′–TGG CGA CGG CAG CGA GGC /3Cy5Sp/ (IDT DNA). The first oligo was labeled with Cy3-NHS (GE Healthcare) according to Roy *et al.* [249] The two oligos were annealed by mixing at 1:1 molar ratio of 10 μM in T50 buffer (10 mM Tris-HCl pH 8, 50 mM NaCl) and heating to 90° C for 5 min followed by slow cooling over 4 hours. After that, the DNA construct was annealed to λ-phage DNA (NEB) and an oligonucleotide containing digoxigenin. First, we added 1 μl of 5 M NaCl and 1 μl of 100 mM $MgCl_2$ electrolytes to 40 μl λ-phage DNA stock of ~ 16 nM (NEB) in an Eppendorf tube. The mixture was heated to 80° C for 10 min, and then placed on ice for 5 min. The DNA constructs were added to a final concentration of 8 nM and BSA was added to a final concentration of 0.1 mg/ml. The tube was covered in foil and the mixture was incubated in the dark with rotation at room temperature for 2-3 hours. DIG oligo 5′–AGG TCG CCG CCC TTT/digoxigenin/ (IDT DNA) was added to a final concentration of 0.2 μM and then incubated with rotation at room temperature for 1-2 hours. This sample was stored at $-20°$ C until assembly on a microscope slide for single-molecule

data acquisition.

The DNA sample was assembled on a glass surface coated with polyethyleneglycol (mixture of mPEG-SVA and Biotin-PEG-SVA, Laysan Bio) according to Roy *et al.* [249] Before adding the DNA sample, the surface was incubated with neutravidin at 0.25 mg/ml for 5 minutes and then blocking buffer (10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mg/ml BSA (New England Biolabs), 1 mg/ml tRNA (Ambion)) for 1 hour. The DNA construct was then incubated on the surface at the concentration of 20 pM in T50 buffer (10 mM Tris-HCl pH 8, 50 mM NaCl) for 10 min. Next, anti-dioxigen-coated 1 μm polystyrene beads (Polysciences) diluted in T50 buffer were added to the imaging chamber for about 30 minutes to allow attachment of beads to the free end of each tether. Finally, the imaging buffer (20 mM Tris-HCl pH 8, 0.5 mg/ml BSA (NEB), 0.1% v/v Tween-20 (Sigma), 0.5% w/v D-Glucose (Sigma), 165 U/ml glucose oxidase (Sigma), 2170 U/ml catalase (Roche), 3 mM Trolox (Sigma)) and NaCl electrolyte of 10 mM or 100 mM or 1 M was added for data acquisition. Single-molecule data acquisition was performed according to Ref. 238. The full description of the setup can be found in our recent review [250]. The DNA stretching experiment was performed by moving the stage in steps at the average loading rate of 455 nm/s. The confocal excitation was scanned following the stage movement. Fluorescence emission was detected for 20 ms after each movement step of the stage. All experiments were carried at room temperature (22° C).

Chapter 6

A coarse-grained model of single-stranded DNA binding protein derived from atomistic simulation and tempered by experiment

6.1 Introduction

A complex of proteins is responsible for DNA replication at a site known as the replication fork. In the replication fork, a protein called DNA helicase separates the DNA duplex into complementary single strands. A DNA polymerase is associated with each strand, traversing it and synthesizing the complementary daughter strand. Because DNA synthesis occurs in the 5′-to-3′ direction, one of the polymerase proteins works in the same direction as the helicase (leading) and the other polymerase works in the opposite direction (lagging), "back-stitching" about every 100 (eukaryote) or 1000 nts (prokaryote). Thus, a considerable amount of single-stranded DNA (ssDNA) can be exposed during the replication procedure. This can pose a problem for high-fidelity DNA replication, as single-stranded DNA is vulnerable to protein degradation [251, 252]. Moreover, processive replication can stall because the single strands emerging from the helicase reaneal [251, 252]. Finally, hairpin structures can form in the self-complementary regions of the lagging strand that will later cause the polymerase to stall, or worse for the cell, be skipped [251, 252]. To protect against all of these effects, all cells have proteins that bind and stabilize ssDNA.

In bacteria, these proteins are known as single-stranded DNA binding proteins (SSBs). SSB binds long ssDNA molecules with such high affinity that it has been difficult to measure the absolute binding affinity quantitatively [253]. The extremely high affinity of SSB for DNA is at apparent odds with the requirements of rapid replication; in the model organism, *E. coli*, a replication fork can proceed at an astonishing rate (> 500 bp/s) [254]. This information provokes the following questions: how does SSB move out of the way of other components of the replication machinery, and how is SSB removed from the shrinking ssDNA of a lagging (Okazaki) fragment?

Recent single-molecule experiments demonstrated the ability for SSB to diffuse along a short stretch of ssDNA. The step size (3 nt) and step rate (60 s^{-1} at 37° C) were estimated for SSB diffusing along ssDNA [255]. Thus, a diffusion coefficient of $D = 270$ nt^2/s was estimated [255]. Using the Einstein-Smoluchowski relation $\langle F \rangle = \frac{v_d k_B T}{D}$, this rate of diffusion is slow enough that a large average force $\langle F \rangle$ (~ 20 pN) for each SSB molecule would need to be supplied to sustain a drift velocity v_d of 500 nt/s. Thus, the displacement of SSB during DNA replication remains an unresolved mystery of the cell.

6.2 Atomic simulations of SSB and ssDNA

An x-ray structure containing two $\sim$ 30 basepair ssDNA fragments bound to SSB was recently published [256]. The structure depicts a homotetramer with four ssDNA binding motifs known as OB-folds that hold the DNA in place through an apparent mix of base-stacking and electrostatic interactions. The

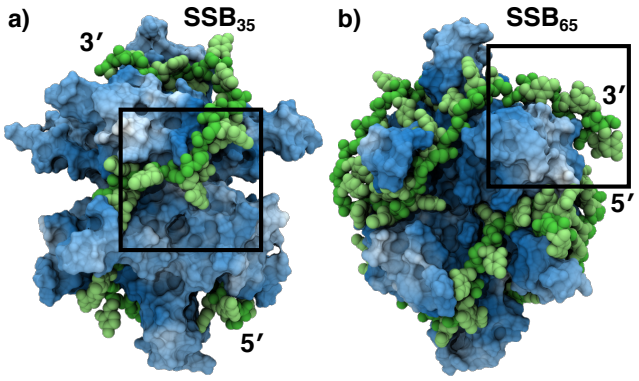

Figure 6.1: Models of SSB bound to ssDNA. The subscript denotes the number of nucleotides occluded by the binding mode. In SSB_{35}, the DNA wraps once so that the DNA ends emerge on opposite sides of the protein. In SSB_{65}, the DNA ends emerge near one another. The DNA is depicted with green van der Waals spheres. The protein is depicted as a white and blue molecular surface colored by the electrostatic potential solved by APBS, Blue and white respectively correspond to positive and negative potentials. The black squares indicate regions that highlight the dynamic character of ssDNA–SSB interactions in the snapshots of Fig. 6.2.

crystallographers connected and extended the resolved DNA fragments to provide models of SSB_{35} and SSB_{65}—the ion-concentration-dependent complexes formed by SSB and a number of nucleotides indicated by the subscript. These models, shown in Fig. 6.1 a and b, were placed in a cubic 120 mM KCl electrolyte 98 Å on each side to provide. The resulting simulation systems were used to perform equilibrium simulations of the SSB complex and in simulations where an external force was used to rupture the complex.

6.2.1 DNA is mobile on SSB in solution

Each system was equilibrated in the NPT ensemble for 60 ns using a 1-fs timestep and 10-Å cutoff. The simulations were continued using a 2-fs timestep with a shorter 8-Å cutoff for an additional 750 and 630 ns for SSB_{35} and SSB_{65}, respectively.

The equilibration of both models revealed extensive DNA dynamics on a comparatively static protein surface. The ends of the DNA were observed to come unbound from the SSB surface before adopting a new conformation and rebinding, see second row of snapshots in Fig. 6.2. Additionally, lateral translation of some nucleotides was observed, see first row of snapshots in Fig. 6.2.

To quantify the mobility of the DNA, the root mean square deviation (RMSD) was calculated for the DNA backbone and for the backbone of the SSB α-helices and β-sheets, Fig. 6.2 B,D. The DNA RMSD was substantially larger than the protein RMSD, especially for the SSB_{65} system, where the end of the DNA peeled away from its initial binding site. We also calculated the root mean square fluctuations (RMSF) of the center of mass of each DNA nucleotide, Fig. 6.2 A,C. The RMSF was broken into radial and transverse

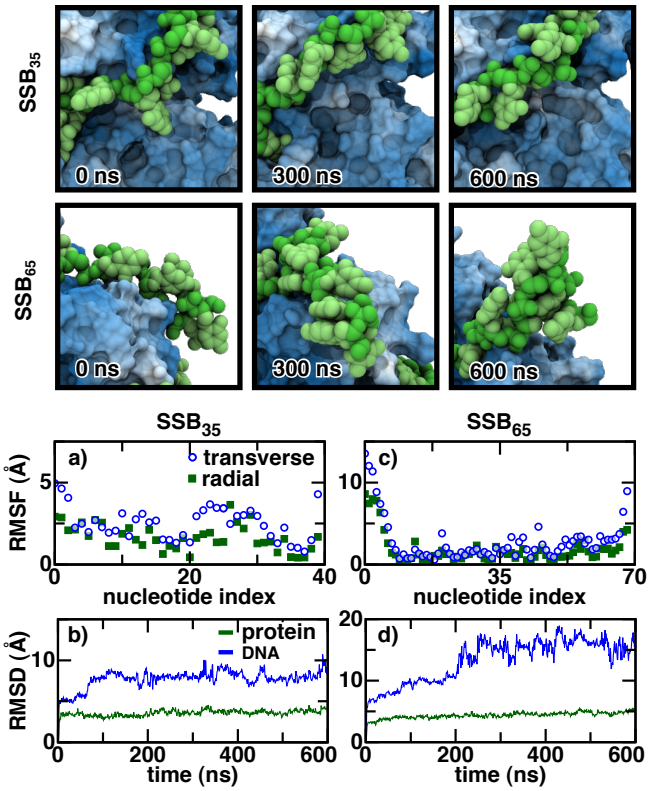

Figure 6.2: DNA bound to SSB is dynamic. During unbiased equilibration simulations, nucleotides were frequently observed performing local rearrangements (first row). Unbinding and subsequent rebinding of the DNA ends was also observed (second row). Colors are as in Fig. 6.1. (A,C) Per nucleotide root mean square fluctuation (RMSF) of SSB_{35} (A) and SSB_{65} (C). The root mean square fluctuation of each nucleotide was obtained relative to the protein backbone atoms initially within 10 Å of any of the nucleotide atoms. With respect to the center of the SSB, the RMSF was broken into radial (closed green squares) and transverse (open blue circles) components. (B,D) Root mean square deviations (RMSDs) of the DNA (blue) and protein (green) from their initial coordinates for SSB_{35} (B) and SSB_{65} (D). Only atoms in the backbone of the DNA and the backbone of protein α-helices and β-sheets contributed to the RMSDs.

coordinates with respect to the center of the SSB protein. The RMSF is highest near the DNA ends, indicating the greatest mobility at these sites.

The extensive DNA dynamics observed on the 100-ns timescale suggests that the majority of DNA bound to SSB is easily repositioned. This observation is consistent with portraits of SSB that depict a protein that can diffuse rapidly along ssDNA to allow local redistribution of SSB. Much of the DNA dynamics was observed in nucleotides that were resolved in the x-ray structure, suggesting that most of the dynamic character of the SSB–ssDNA interactions is representative of SSB in solution. However, some of the DNA nucleotides remained in their initial binding pockets for the complete duration of the simulation, and the timescale for unbinding of such nucleotides remains unclear.

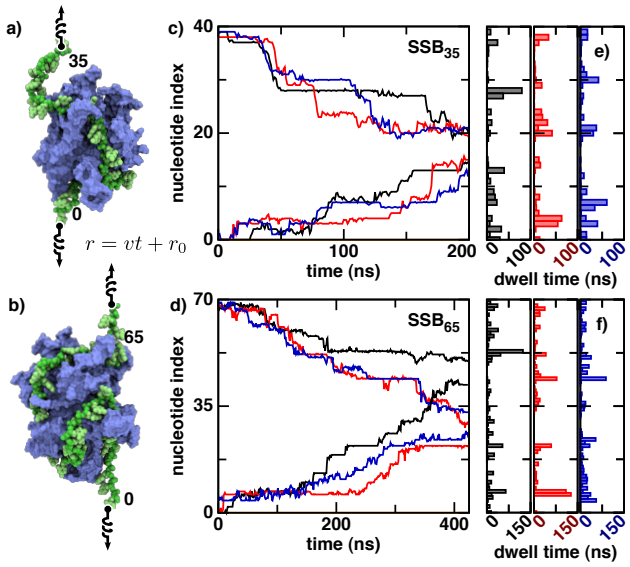

Figure 6.3: DNA was removed from each ends of SSB until the protein was bound only to a small, linear stretch of ssDNA. (a,b) Depiction of the rupture protocol. Each end of a DNA strand wrapped around SSB_{35} (a) or SSB_{65} (b) was tethered to one end of a spring. The other end of each spring was pulled away from SSB along the z-axis at a steady rate of 1 Å/ns. (c,d) Timeseries of the index of the final DNA nucleotide bound to SSB at each end for SSB_{35} (c) and SSB_{65} (d). A nucleotide was assumed bound to SSB if any of its initial contacts with SSB were present. A contact is taken to be any pair of non-hydrogen atoms within 3 Å of one another. Colors of each trace correspond to those in Fig. 6.4. The simulations were performed after roughly 100 (black), 300 (blue) and 500 (red) ns unbiased equilibration. (e,f) Average time a nucleotide spent as the terminal bound nucleotide prior to rupture for SSB_{35} (e) and SSB_{65} (f).

6.2.2 Unwrapping SSB

Rupture of SSB_{35} and SSB_{65} was achieved by pulling each end of the DNA in opposite directions. First, the complex was rotated so that the vector connecting the center of mass (CoM) of the terminal nucleotides lay along the z-axis. One end of a virtual one-dimensional spring was tethered to the CoM of each terminal nucleotide, Fig. 6.4 a, b. The free ends of the springs were pulled in opposite directions at a rate of 1 Å/ns. When a terminal nucleotide fell more than 55 Å away from the center of the SSB along the z-axis, the simulation was stopped and the nucleotide was removed from the system along with a randomly selected potassium ion. The spring was reattached to the nascent terminal nucleotide so that the previous tension was maintained. With the free end of the spring held fixed, the system was equilibrated for 19 ps in the NPT ensemble while the non-hydrogen protein and DNA atoms were harmonically restrained. Pulling of the DNA was resumed in the NVT ensemble. The pulling procedure was completed 3 times for each system, using snapshots of the SSB–ssDNA complex after different amounts of simulation, see Sec. 6.2.1.

Rupture from one side or the other stalled at roughly six distinct locations. Visual examination of initial binding sites of stalled nucleotides revealed that all stall locations on the protein were arginine rich

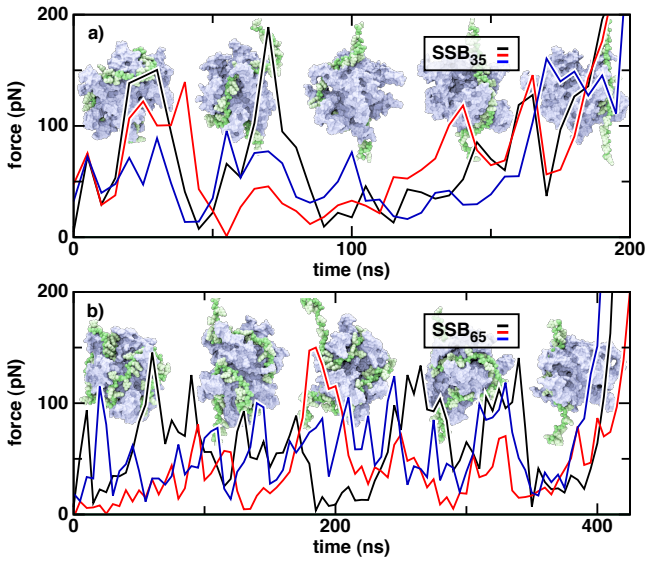

Figure 6.4: Removal of ssDNA from SSB. (a, b) The force was recorded as DNA was removed from SSB_{35} (a) and SSB_{65} (b). The three curves correspond to simulations performed after the system was equilibrated without application of force for different durations. In the background of each panel, the SSB–DNA complex is shown at uniform intervals of during the rupture process.

and included the OB-folds that are known to bind DNA with high affinity. Although the local arginine concentration is greatest at the OB-folds, the stalls at other locations were not necessarily shorter in duration. Since no force was applied directly to SSB, if rupture on one end of the DNA stalled, it could proceed along the other end until that end also stalled. When both ends stalled, tension in the DNA would build until the DNA was released from one of the ends.

The force exhibited large peaks in all simulations, Fig. 6.4. These peaks occurred when rupture stalled at both ends of the DNA. Unlike the distribution of stall sites, the peaks of the force were not distributed with any discernible pattern or regularity. This is likely because the precise unwrapping path taken by the DNA varied in each simulation. The force could also depend on the orientation of the SSB. Finally, the variability of the force could be due to intrinsic differences in the instantaneous conformation of bound DNA. In fact, the average applied force decreased with increasing duration of the equilibration prior to rupture. Overall, the average forces are much larger than those observed in experiment [257]

We continued one of our DNA stretching simulations for 150 ns so that the DNA was held extremely taught. The contacts between the DNA and protein shifted, and the protein rolled slightly but did not dissociate from the over-stretched DNA. Notably, contact with the OB-fold was rupture for 50 ns, during which time contact with the loops L45 and L23 was observed. This observation is qualitatively consistent with a recent result from a fleezers study performed by the Ha group that demonstrated bound ssDNA could be reversibly removed from SSB at the slight cost of 0.1–0.2 k_BT per nucleotide, but complete removal of

DNA involved an 8 k_BT barrier [257].

6.3 Development of a coarse-grained model of SSB

Our atomic simulations demonstrated clearly that access to longer timescales was needed to simulate realistic removal of DNA from SSB. Thus, we developed a coarse-grained (CG) representation of SSB to interact with our CG ssDNA model introduced in Chapter 5. This ssDNA model represents each nucleotide with two beads and was built using the iterative Boltzmann inversion method to match structural distributions from all-atom simulations before refining the model to reproduce the experimentally measured radius of gyration. Our model faithfully reproduces the dependence of extension of a DNA molecule on applied tension [47,231], although it was not specifically parameterized to do so. The success of our CG ssDNA model led us to apply the same approach to SSB. Thus, we developed a model for the interaction between the CG ssDNA and a rigid SSB molecule by using the iterative Boltzmann inversion method to match the distribution of short ssDNA fragments around SSB in an all-atom simulation.

6.3.1 Refinement of CG SSB against all-atom simulation

To parameterize the CG SSB–ssDNA interaction we constructed an atomic model containing thirty-one dT_5 fragments distributed around an SSB molecule in a small cubic volume of 100 mM NaCl electrolyte, 10.6 nm on each side, see Fig. 6.5 A. All-atom molecular dynamics simulations were performed in the NVT (constant number of particles, volume and temperature) using five replicas of the system for $\sim$300 ns per replica; the initial coordinates of each replica were obtained by mutating dC_5 into dT_5 at regular intervals in a previously performed 700 ns simulation. During the simulations, the backbone atoms of residues forming α-helices and β-sheets were harmonically restrained about their initial positions (1 kcal/mol Å^2 spring constant), as the DNA fragments were adsorbed onto and desorbed from the SSB surface.

At the end of the simulations, the DNA atoms were mapped into a coarse-grained representation as described in Section 5.2.1, see Fig. 6.5 B. From the CG representation of the all-atom trajectories, the three-dimensional, position-dependent density was computed for each type of DNA bead with a grid spacing of 2 Å using VMD [167], see Fig. 6.5 C. We refer to these densities as the "target" densities against which the CG SSB–ssDNA interaction potentials are optimized. The SSB protein is a homotetramer with invariance after 180° rotations about any of its three principal axes. We averaged the target distributions over these rotations.

A CG simulation system was built with the same nucleic acid composition as the all-atom system.

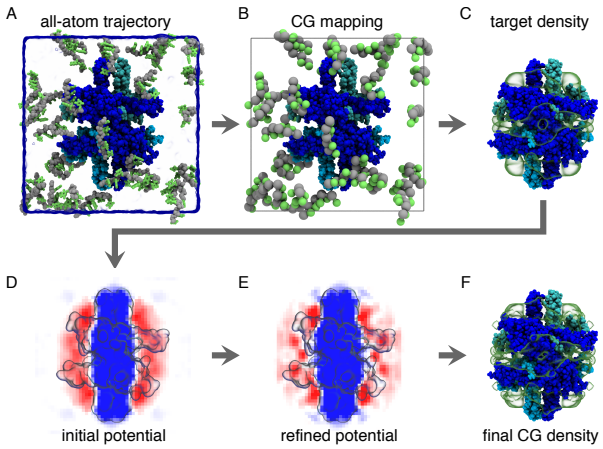

Figure 6.5: Parametrization of SSB–ssDNA potential from all-atom simulation. (A) All-atom MD simulations of SSB surrounded by small ssDNA fragments are performed. The SSB protein was restrained about its initial coordinates. (B) The DNA fragments in the all-atom trajectory are mapped into a CG representation. (C) The target density of each CG ssDNA bead type is extracted from the mapped trajectory. (D) Boltzmann inversion of each target density provides an initial estimate for the CG SSB–ssDNA interaction potential. (E) Via the iterative Boltzmann inversion method, each CG potential is refined until the resulting CG density (F) matches the target density. The potentials and densities shown are for the P ssDNA beads.

Boltzmann inversion of one-tenth of the target densities provided an initial estimate of the potential for each type of bead, see Fig. 6.5 D. The division by 10 roughly accounted for the potential being applied to each bead of a CG DNA fragment. The average value at the edge of the potential grid was subtracted everywhere from the potential.

Subsequent CG simulations applied these potentials using the GridForces [258] feature of NAMD [140] and employed the custom version of NAMD described in Chapter 5. The resulting CG trajectory provided CG density maps for each DNA bead type, which were computed using the same protocol used to obtain the all-atom derived target densities. The ratios of the CG densities to the target densities provided corrections to each SSB potential, according to

$$\Delta u = -a k_{\mathrm{B}} T \ln \frac{\rho_{\mathrm{target}} + \rho_0}{\rho_{\mathrm{CG}} + \rho_0},$$

where $a = 0.1$ was a scaling factor to ensure gradual convergence, and $\rho_0 = 10^{-6}$ beads/Å^3 protected against numerical instability. The correction to the potential was spatially smoothed using a Gaussian filter with 2-Å width. After adding the corresponding correction, each potential was set to 20 $k_{\mathrm{B}}T$ at positions where the corresponding target bead density was less than 10^{-6} beads/Å^3. All software for manipulating the density and potential grids was developed in-house using C++.

In total, 225 iterations were performed. In a typical iteration, 6 replicas of the system each provided

a 30-ns trajectory. Note that a coarse-grained nanosecond corresponds to considerably more real-world nanoseconds due to smoothing of potentials, see Section 5.3.2 for a discussion. By the end of these iterations, CG potentials, Fig. 6.5 E, were obtained that produced coarse-grained densities, Fig. 6.5 F, that were well converged to the target densities.

6.3.2 Refinement of model against experimental results

The binding affinity between an SSB and a dT_8 molecule has been experimentally determined to be $K_{eq} = k_{on}/k_{off} = 20,000\ M^{-1}$ [259], where k_{on} and k_{off} are the kinetic rates of association and dissociation. In general, the binding affinity can be related to the fraction of molecules that are bound at equilibrium using the equation. $\frac{d[SSB \cdot dT_8]}{dt} = k_{on}[SSB][dT_8] - k_{off}[SSB \cdot dT_8]$. At equilibrium, this reduces to

$$k_{on}[SSB][dT_8] = k_{off}[SSB \cdot dT_8].$$

For a system consisting of one dT_8 and one SSB molecule in a volume V, the time-averaged value of the two sides of the above equation must be equal. If the two molecules are bound a fraction f of the time, then we obtain

$$k_{on}\frac{1}{V}\frac{1}{V}(1-f) = k_{off}\frac{1}{V}f,$$

which simplifies to

$$K_{eq} = \frac{f}{1-f}V.$$

It is worth noting that this differs from the expression $K_{eq} = \frac{f}{(1-f)^2}\frac{1}{c}$ that one obtains when considering a system that has many molecules present with total concentration c of each molecular species. The suitability of the former expression for describing the present system of one SSB and one dT_8 molecule was verified by performing simulations of a simpler system across a range of volumes.

The experimentally determined value of K_{eq} then corresponds to a cubic volume 32 nm on each side that contains one SSB and one dT_8 molecule that are bound 50% of the time. For comparison to experiment, we effectively [1] performed 1000 simulations of a dT_8 molecule interacting with our SSB potential in such a volume, see Fig. 6.6 A. The simulations each lasted $\sim$1500 ns. The fraction of bound nucleotides f was estimated by averaging the fraction of bound DNA fragments at each frame of the trajectory during the last 200 ns of simulation. A DNA fragment was considered bound if any of its beads were within a 4-nm radius sphere centered on the SSB potential. We found that the SSB potential, refined against all-atom results, vastly overestimated the affinity of SSB for dT_8, see black curve in Fig. 6.6 B. The overestimation of the

[1] this was accomplished by including multiple non-interacting dT_8 molecules in the same simulation

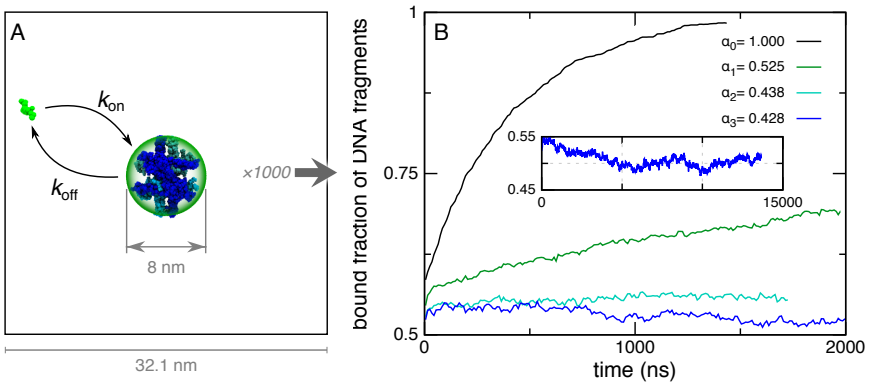

Figure 6.6: Refinement of all-atom results against experimental measurements. (A) Simulations were performed of 1000 copies of a dT_8 fragment interacting with the SSB potentials with the negative portion of the SSB potentials scaled by α_i, see main text for a detailed description. (B) The resulting fraction f of DNA fragments bound to SSB as a function of time. At a given moment, a DNA fragment was considered bound if any of its beads were inside the green sphere around the SSB molecule in (A). For this system, the available experimental data indicates that f should be 0.5.

binding affinity likely arises from errors in the all-atom force field, but could be due to sampling error in the all-atom simulations.

Accordingly, the scaling factor $\alpha_1 = \alpha_0[1-(f-0.5)]$ was applied to the negative values of each potential, reducing the depth of the potential wells. We scaled only negative values of the SSB potentials because the positive regions of the potentials grid were largely due to steric repulsion. Since, only the negative values of the noise at the edge of the system were scaled, the average value at the edge of the each potential grid was subtracted uniformly. Finally, at points where the target density was very low (10^{-6} beads/Å^3), the potential was set to 20 k_BT to ensure that the DNA beads would not be able to enter the core of the SSB molecule. Simulations were subsequently performed using the new SSB potentials to obtain a new estimate for f. This process was repeated iteratively with the scaling factor adjusted by $\alpha_i = \alpha_{i-1}[1-(f-0.5)]$. After only three iterations, the fraction f of DNA fragments binding to the SSB was well converged to the experimentally measured value of 0.5, see Fig. 6.6 B.

6.3.3 Motion of the SSB model

Upon completion of the refinement procedure, the potential map was able to faithfully reproduce the experimental determined binding fraction of dT_8 fragments, and therefore the standard binding free energy. Presuming that this per-nucleotide interaction potential is suitable for longer DNA fragments, it is possible to simulate interactions between a long ssDNA fragment and SSB. However, in such a simulation, the SSB would be fixed both rotationally and translationally, providing a poor mimic of a real SSB molecule. Therefore, we extended the MD simulation package NAMD to allow grids to be associated with a "rigid body"

having mass and a moment of inertia. At each timestep, the force applied to each ssDNA bead by the grid is inverted to obtain the force applied by the bead on the SSB; the corresponding torque is also obtained. A Langevin force and torque [260] are added to the total force and torque to mimic the effect of the solvent. Finally, these forces and torques are used to perform integration of the rigid body's equation of motion to update position and orientation of the SSB molecule using a symplectic algorithm [261]. The mass and the moments of inertia assigned to the CG SSB molecule were calculated from the all-atom model of SSB.

6.3.4 Kinetics of SSB–ssDNA interactions

When one coarse-grains a system, the interaction potentials are smoothed spatially, resulting in faster kinetics. The kinetics of ssDNA diffusion and end-to-end collisions and of SSB diffusion were found to be 5–50 times faster in the coarse-grained simulation than those obtained from experiment. However, the dissociation rate of an 8-nucleotide ssDNA fragment was 3000 times greater in CG simulations than measured experimentally [259].

6.4 Extension of dT_{70}·SSB under tension

Our experimental collaborators in the Chemla group have studied the free energy landscape of SSB binding to DNA using the experimental assay depicted in the inset of Fig. 6.7 [262]. The extension of a DNA construct under tension is measured before (X_{ssDNA}) and after (X_{SSB}) SSB binds. The difference between these extensions, $X_{\mathrm{ssDNA}} - X_{\mathrm{SSB}}$, indicates how much the DNA construct shrinks due to SSB binding, see Fig. 6.7. At low tensions, both extension X_{ssDNA} and X_{SSB} approach zero, and the difference is small. Under a moderate tension ($\sim$5 pN), the DNA is stretched against an entropic force. The SSB-bound DNA molecule exhibits less extension, presumably because some of the DNA is wrapped by the SSB and is unavailable for stretching. At slightly higher forces, the SSB gradually releases its DNA. At even higher forces, the SSB molecule dissociates from the DNA and $X_{\mathrm{ssDNA}} - X_{\mathrm{SSB}}$ approaches zero.

Analogous CG simulations were performed using our ssDNA and SSB models. In each simulation a dT_{70} molecule was stretched by applying opposite constant forces to the P beads at each end of the molecule. The extension of the molecules in simulations without SSB converged rapidly to a (dynamic) steady state within about 20 ns. At the onset of simulations containing SSB, the protein was placed proximal to the DNA. The SSB molecule rapidly bound to the DNA regardless of the applied tension. Dissociation was not observed, even at the highest applied tension (20 pN). Within 50 ns of the onset of each simulation, the extension of the DNA qualitatively reached a (dynamic) steady-state. Using the experimentally tempered CG potentials,

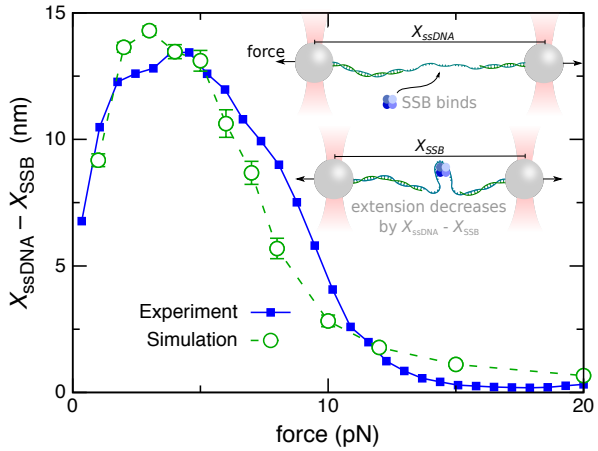

Figure 6.7: Comparison between CG simulation and experiment of the average change in extension upon SSB dissociation from a dT_{70} molecule under variable tension. In the experimental assay, schematically illustrated in the inset, a long double-stranded DNA molecule was stretched between two optically trapped beads that reported on the extension of the molecule. SSB was captured on the single-stranded DNA at low applied tension. The force on the DNA was slowly increased as the extension was recorded until the SSB molecule dissociated. The process was repeated in the absence of DNA molecules. The difference in the extension of SSB-free and SSB-bound DNA (filled blue squares) was extracted after averaging many traces. Complementary CG simulations were performed of a dT_{70} molecule under constant tension in the presence and absence of SSB. The force was applied in opposite directions along a fixed axis. The extension of the DNA molecule along the same axis was averaged and the difference between SSB-free and SSB-bound extensions was obtained at each force (open green circles). Experimental data was provided by the Chemla group [262]

we obtained good emergent agreement with the experimental measurements of $X_{\mathrm{ssDNA}} - X_{\mathrm{SSB}}$ across the full range of forces, see Fig. 6.7.

6.5 Conclusion

We employed a combination of atomically detailed MD simulations and coarse-grained modeling study the dynamics of complexes formed from SSB and ssDNA. The atomically detailed MD simulations, demonstrated rearrangements of ssDNA on the SSB surface occurring on the 100-nanosecond timescale. This is, perhaps, surprising because the timescale of SSB diffusion is closer to seconds [255]. However, for the SSB to diffuse, the rearrangements observed in our atomic simulations would need to propagate around the SSB surface. The rupture pathway taken by SSB when both ends were pulled in atomic simulation was apparently stochastic. Rupture would proceed from one end of the DNA until an arginine-rich "hot spot" was reached, whereupon rupture would, temporarily stall. The forces obtained atomic simulations were roughly one to two orders of magnitude larger than the forces obtained in equivalent experiments because the simulations had to be performed many orders of magnitude more quickly to be computationally tractable.

To obtain an equilibrium measure of the extension of an ssDNA molecule bound to SSB under variable tension, we developed a coarse-grained model of SSB to interact with the coarse-grained DNA model developed in Chapter 5. The CG model of SSB was obtained by tuning a grid-based potential for each type of ssDNA bead until the three-dimensional density of the beads matched equivalent densities obtained from all-atom simulation. The resulting CG model displayed an exaggerated affinity for short ssDNA fragments that we suspect derives from errors in the all-atom force field. After refining the CG model against the experimentally determined binding affinity for dT_8, our model was able to accurately replicate the extension of a dT_{70} molecule bound to SSB across a range of applied tensions.

6.6 Simulation methods

All MD simulations were performed using the program NAMD [140], the TIP3P model of water [170], standard parameters for ions [171], periodic boundary conditions, particle-mesh Ewald (PME) full electrostatics with a PME grid density of about 1 Å per grid point.

The temperature was held constant using a Langevin thermostat [140] applied to all non-hydrogen atoms; the Langevin damping constant was set to 0.1 ps^{-1}. For simulations in the NPT ensemble, constant pressure was maintained at 1 bar using the Nosé-Hoover Langevin piston pressure control [174].

6.6.1 Atomic simulations of SSB and long ssDNA fragments

In the simulations of SSB_{35} and SSB_{65}, the CHARMM27 force field [243, 244] was employed with van der Waals and short-range electrostatic energies were calculated using a smooth (10–12 Å) cutoff, and integration was performed using 1–2–4 fs multiple timestepping [140].

6.6.2 Atomic simulations of SSB in a solution of short DNA fragments

In the simulations of SSB surrounded by a solution of dT_5 fragments, the newer CHARMM36 force field [137, 242–244] was employed with NBFIX corrections applied to ion–nucleic acid interactions and to the arginine and lysine amine–nucleic acid interactions [61]. Van der Waals and short-range electrostatic energies were calculated using a smooth (7–8 Å) cutoff. Integration was performed using 2–2–6 fs multiple timestepping [172]. To enable 2-fs timestepping for bonded interactions, water bonds (and angles) and non-water covalent bonds with hydrogens were held rigid using the SETTLE [246] and RATTLE [247] algorithms, respectively.

References

[1] Christopher Maffeo, Jejoong Yoo, Jeffrey Comer, David Wells, Binquan Luan, and Aleksei Aksimentiev. Close encounters with DNA. *J. Phys.: Condens. Matter*, 2014. Accepted.

[2] Paul a Wiggins, Thijn van der Heijden, Fernando Moreno-Herrero, Andrew Spakowitz, Rob Phillips, Jonathan Widom, Cees Dekker, and Philip C Nelson. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotech.*, 1(2):137–141, 2006.

[3] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236):362–U129, 3 2009.

[4] Reza Vafabakhsh and Taekjip Ha. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science*, 337(6098):1097–1101, 2012.

[5] Shinsuke Ito, Li Shen, Qing Dai, Susan C. Wu, Leonard B. Collins, James A. Swenberg, Chuan He, and Yi Zhang. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 333(6047):1300–3, 9 2011.

[6] A. D. Hershey and Martha Chase. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.*, 36(1):39–56, 1952.

[7] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acids. *Nature*, 171:737–738, 1953.

[8] Werner Kuhn. Beziehungen zwischen Molekülgröße, statistischer Molekülgestalt und elastischen Eigenschaften hochpolymerer Stoffe. *Colloid & Polymer Science*, 76(3):258–271, 1936.

[9] H. A. Kramers. The behavior of macromolecules in inhomogeneous flow. *J. Chem. Phys.*, 14(7):415, 1946.

[10] O. Kratky and G. Porod. Röntgenuntersuchung gelöster Fadenmoleküle. *Recl. Trav. Chim. Pay. B.*, 68(12):1106–1122, September 1949.

[11] Von Anton Peterlin. Lichtzerstreuung an ziemlich gestreckten Fadenmolekülen. *Makromol. Chem.*, 9(1):244–268, February 1953.

[12] Paul Doty. The physical chemistry of deoxyribonucleic acids. *J. Cell. & Comparat. Physiol.*, 49(S1):27–57, May 1957.

[13] Terrell L. Hill. Generalization of the onedimensional Ising model applicable to helix transitions in nucleic acids and proteins. *J. Chem. Phys.*, 30(2):383–387, February 1959.

[14] J. H. Gibbs and E. a. DiMarzio. Statistical mechanics of helix–coil transitions in biological macromolecules. *J. Chem. Phys.*, 30(1):271, 1959.

[15] Bruno H. Zimm. Theory of "melting" of the helical form in double chains of the DNA type. *J. Chem. Phys.*, 33(5):1349–1356, 1960.

[16] P G de Gennes. Some conformation problems for long macromolecules. *Rep. Progr. Phys.*, 32(1):187–205, January 1969.

[17] F. H. C. Crick and A. Klug. Kinky helix. *Nature*, 255(5509):530–533, 1975.

[18] H M Sobell, C C Tsai, S C Jain, and S G Gilbert. Visualization of drug-nucleic acid interactions at atomic resolution. III. Unifying structural concepts in understanding drug–DNA interactions and their broader implications in understanding protein–DNA interactions. *J. Mol. Biol.*, 114(3):333–365, August 1977.

[19] J L Sussman and E N Trifonov. Possibility of nonkinked packing of DNA in chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, 75(1):103–107, January 1978.

[20] M Levitt. How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc. Natl. Acad. Sci. U.S.A.*, 75(2):640–644, February 1978.

[21] E. Clementi and G. Corongiu. Interaction of water with DNA single and double helix in the B conformation. *Int. J. Quantum Chem.*, 16(4):897–915, October 1979.

[22] Enrico Clementi and Giorgina Corongiu. A theoretical study on the water structure for nucleic acids bases and base pairs in solution at $T = 300$ K. *J. Chem. Phys.*, 72(7):3979–3992, 1980.

[23] M. Levitt. Computer simulation of DNA double-helix dynamics. *Cold Spring Harbor Symp. Quant. Biol.*, 47:251–262, January 1983.

[24] B. Tidor, K. K. Irikura, B. R. Brooks, and M. Karplus. Dynamics of DNA oligomers. *J. Biomol. Struct. Dyn.*, 1:231–252, 1983.

[25] G. L. Seibel, U. C. Singh, and P. A. Kollman. A molecular dynamics simulation of double-helical B-DNA including counterions and water. *Proc. Natl. Acad. Sci. U.S.A.*, 82:6537–6540, 1985.

[26] DL Beveridge, S Swaminathan, G Ravishanker, JM Withka, J Srinivasan, C Prevost, S Louise-May, DR Langley, FM DiCapua, and PH Bolton. Molecular dyanmics simulation on hydration, structure and motions of DNA oligomers. In Eric Westhof, editor, *Water and biological macromolecules*, chapter 6, page 165. CRC Press, 1993.

[27] B. Jayaram and D. Beveridge. Modelling DNA in aqueous solutions – theoretical and computer simulation studies on the ion atmosphere of DNA. *Annu. Rev. Biophys. Biomol. Struct.*, 25:367–394, 1996.

[28] D. L. Beveridge and G. Ravishanker. Molecular dynamics studies of DNA. *Curr. Opin. Struct. Biol.*, 4:246–255, 1994.

[29] C F Anderson and M T Record. Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.*, 46:657–700, January 1995.

[30] T. E. III Cheatham, J. L. Miller, T. Fox, T A Darden, and P A Kollman. Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.*, 117(14):4193–4194, April 1995.

[31] T E Cheatham and Peter A Kollman. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.*, 51:435–471, January 2000.

[32] D L Beveridge and K J McConnell. Nucleic acids: theory and computer simulation, Y2K. *Curr. Opin. Struct. Biol.*, 10(2):182–196, April 2000.

[33] T. E. Cheatham and P. A. Kollman. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.*, 259:434–444, 1996.

[34] M. A. Young, G. Ravishanker, and D. L. Beveridge. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motion and solvation. *Biophys. J.*, 73:2313–2336, 1997.

[35] Thomas E Cheatham and Matthew A Young. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers*, 56(4):232–256, 2000.

[36] W Saenger, W N Hunter, and O Kennard. DNA conformation is determined by economics in the hydration of phosphate groups. *Nature*, 324(6095):385–388, 1986.

[37] Michael Feig and B Montgomery Pettitt. A molecular simulation picture of DNA hydration around A- and B-DNA. *Biopolymers*, 48(4):199–209, January 1998.

[38] G. S. Manning. Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. *J. Chem. Phys.*, 51:924, 1969.

[39] M. A. Young, B. Jayaram, and D. L. Beveridge. Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.*, 119:59–69, 1997.

[40] Alexander P. Lyubartsev and Lars Nordenskiöld. Monte carlo simulation study of ion distribution and osmotic pressure in hexagonally oriented DNA. *J. Phys. Chem.*, 99(25):10373–10382, 1995.

[41] Michael Feig and B. Montgomery Pettitt. Sodium and chlorine ions as part of the DNA solvation shell. *Biophys. J.*, 77(4):1769 – 1781, 1999.

[42] B. Jayaram and D. L. Beveridge. Modeling DNA in aqueous solutions: Theoretical and computer simulation studies on the ion atmosphere of DNA. *Annu. Rev. Biophys. Biomol. Struct.*, 25(1):367–394, 1996.

[43] W. Clark Still, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[44] B Jayaram, D Sprous, and D L Beveridge. Solvation free energy of biomacromolecules: Parameters for a modified generalized Born model consistent with the AMBER force field. *J. Phys. Chem. B*, 102(47):9571–9576, November 1998.

[45] Brian N. Dominy and Charles L. Brooks, III. Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B*, 103:3765–3773, 1999.

[46] A D MacKerell and G U Lee. Structure, force, and energy of a double-stranded DNA oligonucleotide under tensile loads. *Eur. Biophys. J.*, 28(5):415–426, January 1999.

[47] Steven B. Smith, Yujia Cui, and Carlos Bustamante. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271:795–799, 1996.

[48] Jií Šponer, Arnošt Mládek, Judit E Šponer, Daniel Svozil, Marie Zgarbová, Pavel Banáš, Petr Jurečka, and Michal Otyepka. The DNA and RNA sugar-phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies. *Phys. Chem. Chem. Phys.*, 14(44):15257–15277, November 2012.

[49] Christopher Maffeo, Robert Schöpflin, Hergen Brutzer, René Stehr, Aleksei Aksimentiev, Gero Wedemann, and Ralf Seidel. DNA–DNA interactions in tight supercoils are described by a small effective charge density. *Phys. Rev. Lett.*, 105(15):158101, 2010.

[50] Mithun Biswas, Karine Voltz, Jeremy C. Smith, and Jörg Langowski. Role of histone tails in structural stability of the nucleosome. *PLoS Comput. Biol.*, 7(12):e1002279, 12 2011.

[51] Jejoong Yoo and Aleksei Aksimentiev. In situ structure and dynamics of DNA origami determined through molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 110(50):20099–20104, 2013.

[52] Jonathan P K Doye, Thomas E Ouldridge, Ard a Louis, Flavio Romano, Petr Šulc, Christian Matek, Benedict E K Snodin, Lorenzo Rovigatti, John S Schreck, Ryan M Harrison, and William P J Smith. Coarse-graining DNA for simulations of DNA nanotechnology. *Phys. Chem. Chem. Phys.*, 15(47):20395–20414, December 2013.

[53] Geoffrey C Rollins, Anton S Petrov, and Stephen C Harvey. The role of DNA twist in the packaging of viral genomes. *Biophys. J.*, 94(5):L38–40, March 2008.

[54] Mariliis Tark-Dame, Roel van Driel, and Dieter W Heermann. Chromatin folding–from biology to polymer models and back. *J. Cell Sci.*, 124(Pt 6):839–845, March 2011.

[55] Jií Šponer, Judit E Šponer, Arnošt Mládek, Petr Jurečka, Pavel Banáš, and Michal Otyepka. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers*, 99(12):978–988, December 2013.

[56] Jan Norberg and Lennart Nilsson. Molecular dynamics applied to nucleic acids. *Acc. Chem. Res.*, 35(6):465–472, June 2002.

[57] Cheatham TE. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.*, 14(3):360–7, 2004.

[58] Alberto Pérez, F Javier Luque, and Modesto Orozco. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.*, 45(2):196–205, February 2012.

[59] Martin Egli. DNA-cation interactions: Quo vadis? *Chem. Biol.*, 9(3):277–286, March 2002.

[60] Francesca Mocci and Aatto Laaksonen. Insight into nucleic acid counterion interactions from inside molecular dynamics simulations is "worth its salt". *Soft Matter*, 8(36):9268, 2012.

[61] Jejoong Yoo and Aleksei Aksimentiev. Improved parametrization of Li^+, Na^+, K^+, and Mg^{2+} ions for all-atom molecular dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett.*, 3(1):45–50, 2012.

[62] D L Beveridge, Surjit B Dixit, Gabriela Barreiro, and Kelly M Thayer. Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers*, 73(3):380–403, February 2004.

[63] Chantal Prévost, Masayuki Takahashi, and Richard Lavery. Deforming DNA: from physics to biology. *ChemPhysChem*, 10(9-10):1399–1404, July 2009.

[64] Justin P Peters and L James Maher. DNA curvature and flexibility *in vitro* and *in vivo*. *Quart. Rev. Biophys.*, 43(1):23–63, February 2010.

[65] Alexander Vologodskii and Maxim D Frank-Kamenetskii. Strong bending of the DNA double helix. *Nucl. Acids Res.*, 41(14):6785–6792, August 2013.

[66] Emmanuel Giudice and Richard Lavery. Simulations of nucleic acids and their complexes. *Acc. Chem. Res.*, 35(6):350–357, June 2002.

[67] Katrin Spiegel and Alessandra Magistrato. Modeling anticancer drug–DNA interactions via mixed QM/MM molecular dynamics simulations. *Organic & biomolecular chemistry*, 4(13):2507–2517, July 2006.

[68] Hao Wang and Charles A Laughton. Molecular modelling methods for prediction of sequence-selectivity in DNA recognition. *Methods (San Diego, Calif.)*, 42(2):196–203, June 2007.

[69] Shantanu Sharma, Feng Ding, and Nikolay V Dokholyan. Multiscale modeling of nucleosome dynamics. *Biophys. J.*, 92(5):1457–1470, March 2007.

[70] Mu Gao and Klaus Schulten. Integrin activation in vivo and in silico. *Structure*, 12:2096–2098, 2004.

[71] Aleksij Aksimentiev, Jiunn Benjamin Heng, Gregory Timp, and Klaus Schulten. Microscopic kinetics of DNA translocation through synthetic nanopores. *Biophys. J.*, 87:2086–2097, 2004.

[72] Alberto Pérez, F. Javier Luque, and Modesto Orozco. Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.*, 129(47):14739–45, 2007.

[73] Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson, and Klaus Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14:437–449, 2006.

[74] Karissa Y Sanbonmatsu. Computational studies of molecular machines: the ribosome. *Curr. Opin. Struct. Biol.*, 22(2):168–174, April 2012.

[75] D. C. Rau, B. Lee, and V. A. Parsegian. Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: Hydration forces between parallel DNA double helices. *Proc. Natl. Acad. Sci. U.S.A.*, 81(9):2621–2625, 1984.

[76] P.W.K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.

[77] Hendrik Dietz, Shawn M. Douglas, and William M. Shih. Folding DNA into twisted and curved nanoscale shapes. *Science*, 325(5941):725–30, 2009.

[78] Yonggang Ke, Shawn M. Douglas, Minghui Liu, Jaswinder Sharma, Anchi Cheng, Albert Leung, Yan Liu, William M. Shih, and Hao Yan. Multilayer DNA origami packed on a square lattice. *J. Am. Chem. Soc.*, 131(43):15903–8, 2009.

[79] Marissa G Saunders and Gregory A Voth. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.*, 42:73–93, January 2013.

[80] Davit A. Potoyan, Alexey Savelyev, and Garegin A. Papoian. Recent successes in coarse-grained modeling of DNA. *WIREs Comput Mol Sci*, 3(1):69–83, January 2013.

[81] Jonathan P K Doye, Thomas E Ouldridge, Ard A Louis, Flavio Romano, Petr Sulc, Christian Matek, Benedict E K Snodin, Lorenzo Rovigatti, John S Schreck, Ryan M Harrison, and William P J Smith. Coarse-graining DNA for simulations of DNA nanotechnology. *Phys. Chem. Chem. Phys.*, 15(47):20395–20414, December 2013.

[82] Daniel M Hinckley, Gordon S Freeman, Jonathan K Whitmer, and Juan J de Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.*, 139(14):144903, October 2013.

[83] Petr Šulc, Flavio Romano, Thomas E Ouldridge, Lorenzo Rovigatti, Jonathan P K Doye, and Ard A Louis. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.*, 137(13):135101, October 2012.

[84] Pablo M. De Biase, Carlos J. F. Solano, Suren Markosyan, Luke Czapla, and Sergei Yu Noskov. BROMOC-D: Brownian dynamics/Monte-Carlo program suite to study ion and DNA permeation in nanopores. *J. Chem. Theory Comput.*, 8(7):2540–2551, 7 2012.

[85] Alexey Savelyev and Garegin A Papoian. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 107(47):20340–20345, November 2010.

[86] Yi He, Maciej Maciejczyk, Stanisław Ołdziej, Harold A. Scheraga, and Adam Liwo. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys. Rev. Lett.*, 110(9):098101, February 2013.

[87] Chia Wei Hsu, Maria Fyta, Greg Lakatos, Simone Melchionna, and Efthimios Kaxiras. *Ab initio* determination of coarse-grained interactions in double-stranded DNA. *J. Chem. Phys.*, 137(10):105102, September 2012.

[88] Alex Morriss-Andrews, Joerg Rottler, and Steven S Plotkin. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J. Chem. Phys.*, 132(3):035105, January 2010.

[89] Pablo D. Dans, Ari Zeida, Matías R. Machado, and Sergio Pantano. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J. Chem. Theory Comput.*, 6(5):1711–1725, May 2010.

[90] Maciej Maciejczyk, Aleksandar Spasic, Adam Liwo, and Harold A Scheraga. Coarse-grained model of nucleic acid bases. *J. Comput. Chem.*, 31(8):1644–1655, June 2010.

[91] Srinivasa M Gopal, Shayantani Mukherjee, Yi-Ming Cheng, and Michael Feig. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins: Struct., Func., Bioinf.*, 78(5):1266–1281, April 2010.

[92] Robert C DeMille, Thomas E Cheatham, and Valeria Molinero. A coarse-grained model of DNA with explicit solvation by water and ions. *J. Phys. Chem. B*, 115(1):132–142, January 2011.

[93] John SantaLucia Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 95(4):1460–1465, 1998.

[94] Christopher Maffeo, Thuy T. M. Ngo, Taekjip Ha, and Aleksei Aksimentiev. A coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment. *J. Chem. Theory Comput.*, 2014. Article ASAP.

[95] P. J. Hagerman. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.*, 17:265–286, 1988.

[96] F. B. Fuller. The writhing number of a space curve. *Proc. Natl. Acad. Sci. U.S.A.*, 68:815–819, 1971.

[97] Mary D. Barkley and Bruno H. Zimm. Theory of twisting and bending of chain macromolecules; analysis of the fluorescence depolarization of DNA. *J. Chem. Phys.*, 70(6):2991–3007, 1979.

[98] Jiro Shimada and Hiromi Yamakawa. Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules*, 17(4):689–698, July 1984.

[99] Tali E Haran and Udayan Mohanty. The unique structure of A-tracts and intrinsic DNA bending. *Quart. Rev. Biophys.*, 42(1):41–81, February 2009.

[100] David L. Beveridge, Gabriela Barreiro, K. Suzie Byun, David A. Case, III Thomas E. Cheatham, Surjit B. Dixit, Emmanuel Giudice, Filip Lankas, Richard Lavery, John H. Maddocks, Roman Osman, Eleanore Seibert, Heinz Sklenar, Gautier Stoll, Kelly M. Thayer, Péter Varnai, and Matthew A. Young. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. i. research design and results on d(c_pg) steps. *Biophys. J.*, 87:3799–3813, 2004.

[101] C Bustamante, S B Smith, J Liphardt, and D Smith. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.*, 10(3):279–285, 2000.

[102] Fedor Kouzine, Suzanne Sanford, Zichrini Elisha-Feil, and David Levens. The functional response of upstream DNA to dynamic supercoiling *in vivo*. *Nat. Struct. Mol. Biol.*, 15(2):146–154, February 2008.

[103] Tracy A Brooks and Laurence H Hurley. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer*, 9(12):849–861, December 2009.

[104] Filip Lankas, Jirí Sponer, Jörg Langowski, and Thomas E Cheatham. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.*, 85(5):2872–2883, November 2003.

[105] Timothy E Cloutier and Jonathan Widom. Spontaneous sharp bending of double-stranded DNA. *Mol. Cell*, 14(3):355–362, May 2004.

[106] A. A. Gorin, V. B. Zhurkin, and W. K. Olson. B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.*, 247:34–48, 1995.

[107] Wilma K. Olson, Andrey A. Gorin, Xiang-Jun Lu, Lynette M. Hock, and Victor B. Zhurkin. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.*, 95:11163–11168, September 1998.

[108] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[109] Yoshiteru Yonetani and Hidetoshi Kono. Sequence dependencies of DNA deformability and hydration in the minor groove. *Biophys. J.*, 97(4):1138–1147, August 2009.

[110] Alexey K. Mazur. Symplectic integration of closed chain rigid body dynamics with internal coordinate equations of motion. *J. Chem. Phys.*, 111(4):1407–1414, 1999.

[111] Alexey K. Mazur. Local elasticity of strained DNA studied by all-atom simulations. *Phys. Rev. E*, 84(2):021903, August 2011.

[112] Agnes Noy and Ramin Golestanian. Length scale dependence of DNA mechanical properties. *Phys. Rev. Lett.*, 109(22):228101, November 2012.

[113] Alexey K. Mazur. Comment on Length scale dependence of DNA mechanical properties. *Phys. Rev. Lett.*, 111(17):179801, October 2013.

[114] Agnes Noy and Ramin Golestanian. Noy and Golestanian reply:. *Phys. Rev. Lett.*, 111(17):179802, October 2013.

[115] Graham L Randall, Lynn Zechiedrich, and B Montgomery Pettitt. In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucl. Acids Res.*, 37(16):5568–5577, September 2009.

[116] J D Kahn and D M Crothers. Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. U.S.A.*, 89(14):6343–6347, July 1992.

[117] Rebecca S Mathew-Fenn, Rhiju Das, and Pehr a B Harbury. Remeasuring the double helix. *Science*, 322(5900):446–449, October 2008.

[118] Quan Du, Chaim Smith, Nahum Shiffeldrim, Maria Vologodskaia, and Alexander Vologodskii. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. U.S.A.*, 102(15):5397–5402, April 2005.

[119] Alexander J Mastroianni, David a Sivak, Phillip L Geissler, and a Paul Alivisatos. Probing the conformational distributions of subpersistence length DNA. *Biophys. J.*, 97(5):1408–1417, September 2009.

[120] Hao Qu, Yong Wang, Chiao-Yu Tseng, and Giovanni Zocchi. Critical torque for kink formation in double-stranded DNA. *Phys. Rev. X*, 1(2):021008, November 2011.

[121] F Fogolari, A Brigo, and H Molinari. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Magn. Reson.*, 15(6):377–392, 2002.

[122] Christopher Maffeo, Swati Bhattacharya, Jejoong Yoo, David Wells, and Aleksei Aksimentiev. Modeling and simulation of ion channels. *Chem. Rev.*, 112(12):6250–6284, 2012.

[123] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–92, 1993.

[124] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[125] Paul K. Weiner and Peter A. Kollman. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.*, 2(3):287–303, 1981.

[126] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[127] George A. Kaminski, Richard A. Friesner, Julian Tirado-Rives, and William L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, 105(28):6474–6487, 2001.

[128] Jan Hermans, Herman J. C. Berendsen, Wilfred F. Van Gunsteren, and Johan P. M. Postma. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23(8):1513–1518, 1984.

[129] T. M. Schmeing, R. M. Voorhees, A. C. Kelley, and V. Ramakrishnan. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat. Struct. Mol. Biol.*, 18:432–436, 2011.

[130] Jay W. Ponder, Chuanjie Wu, Pengyu Ren, Vijay S. Pande, John D. Chodera, Michael J. Schnieders, Imran Haque, David L. Mobley, Daniel S. Lambrecht, Robert A. DiStasio, Martin Head-Gordon, Gary N. I. Clark, Margaret E. Johnson, and Teresa Head-Gordon. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B*, 114(8):2549–64, 3 2010.

[131] W. Jiang, D.J. Hardy, J.C. Phillips, A.D. MacKerell Jr, K. Schulten, and B. Roux. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J. Phys. Chem. Lett.*, 2:87–92, 2011.

[132] Roderick Mackinnon. Potassium channels and the atomic basis of selective ion conduction (nobel lecture). *Angew. Chem. Int. Ed.*, 43:4265–4277, 2004.

[133] N. Foloppe and A. D. MacKerrell Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, 21:86–104, 2000.

[134] Thomas E. Cheatham, Piotr Cieplak, and Peter A. Kollman. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, 16(4):845–862, 1999. PMID: 10217454.

[135] Péter Várnai, Dragana Djuranovic, Richard Lavery, and Brigitte Hartmann. α/γ transitions in the B-DNA backbone. *Nucleic Acids Research*, 30(24):5398–5406, 2002.

[136] A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.*, 92:3817–3829, 2007.

[137] Katarina Hart, Nicolas Foloppe, Christopher M. Baker, Elizabeth J. Denning, Lennart Nilsson, and Alexander D. MacKerell. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.*, 8(1):348–362, 2012.

[138] Thomas E. Cheatham and David A. Case. Twenty-five years of nucleic acid simulations. *Biopolymers*, 99(12):969–77, 12 2013.

[139] Andrew T Guy, Thomas J Piggot, and Syma Khalid. Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophys. J.*, 103(5):1028–1036, September 2012.

[140] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26:1781–1802, 2005.

[141] Benoit Roux. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.*, 91:275–282, 1995.

[142] Peter Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.*, 93:2395–2417, 1993.

[143] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *J. Phys. Chem. B*, 114(32):10235–10253, 2010.

[144] GM Torrie and JP Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comp. Phys.*, 23(2):187–199, 1977.

[145] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13:1011–1021, 1992.

[146] Hyung-June Woo and Benoit Roux. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 102(19):6825–6830, 2005.

[147] U. C. Singh, Frank K. Brown, Paul A. Bash, and Peter A. Kollman. An approach to the application of free energy perturbation methods using molecular dynamics: applications to the transformations of $CH_3OH \rightarrow CH_3CH_3$, $H_3O^+ \rightarrow NH_4^+$, glycine $\rightarrow$ alanine, and alanine $\rightarrow$ phenylalanine in aqueous solution and to $H_3O^+(H_2O)_3 \rightarrow NH_4^+(H_2O)_3$ in the gas phase. *J. Am. Chem. Soc.*, 109(6):1607–1614, 1987.

[148] A. Laio and F.L. Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Progr. Phys.*, 71:126601, 2008.

[149] D. Rodriguez-Gomez, E. Darve, and A. Pohorille. Assessing the efficiency of free energy calculation methods. *J. Chem. Phys.*, 120:3563–3578, 2004.

[150] Helmut Grubmüller, Berthold Heymann, and Paul Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271:997–999, 1996.

[151] Barry Isralewitz, Sergei Izrailev, and Klaus Schulten. Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations. *Biophys. J.*, 73:2972–2979, 1997.

[152] Sergei Izrailev, Sergey Stepaniants, Manel Balsera, Yoshi Oono, and Klaus Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72:1568–1581, 1997.

[153] Sergey Stepaniants, Sergei Izrailev, and Klaus Schulten. Extraction of lipids from phospholipid membranes by steered molecular dynamics. *J. Mol. Mod.*, 3:473–475, 1997.

[154] Siewert-Jan Marrink, Oliver Berger, Peter Tieleman, and Fritz Jähnig. Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics simulations. *Biophys. J.*, 74:931–943, 1998.

[155] Dorina Kosztin, Sergei Izrailev, and Klaus Schulten. Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. *Biophys. J.*, 76:188–197, 1999.

[156] Gerhard Hummer and Attila Szabo. Free energy surfaces from single-molecule force spectroscopy. *Acc. Chem. Res.*, 38(7):504–513, 1 2005.

[157] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[158] Davide Marenduzzo, Enzo Orlandini, Andrzej Stasiak, De Witt Sumners, Luca Tubiana, and Cristian Micheletti. DNA–DNA interactions in bacteriophage capsids are responsible for the observed DNA knotting. *Proc. Natl. Acad. Sci. U.S.A.*, 106(52):22269–22274, December 2009.

[159] AA Kornyshev and S. Leikin. Sequence recognition in the pairing of DNA duplexes. *Phys. Rev. Lett.*, 86(16):3666–3669, 2001.

[160] Y Kao-Huang, A Revzin, A P Butler, P O'Conner, D W Noble, and P H von Hippel. Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound *escherichia coli lac* repressor *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.*, 74(10):4228–4232, October 1977.

[161] K Klenin, H Merlitz, and J Langowski. A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.*, 74(2):780–788, February 1998.

[162] John F. Marko and Eric D. Siggia. Statistical mechanics of supercoiled DNA. *Phys. Rev. E*, 52(3):2912–2938, 1995.

[163] J. A. Schellman and D. Stigter. Electrical double layer, zeta potential, and electrophoretic charge of double-stranded DNA. *Biopolymers*, 16:1415, 1977.

[164] M Hammermann, C Steinmaier, H Merlitz, U Kapp, W Waldeck, G Chirico, and J Langowski. Salt effects on the structure and internal dynamics of superhelical DNAs studied by light scattering and Brownian dynamics. *Biophys. J.*, 73(5):2674–2687, November 1997.

[165] B. Luan and A. Aksimentiev. DNA attraction in monovalent and divalent electrolytes. *J. Am. Chem. Soc.*, 130:15754–15755, 2008.

[166] Aleksij Aksimentiev and Klaus Schulten. Imaging alpha-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability and the electrostatic potential map. *Biophys. J.*, 88:3745–3761, 2005.

[167] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[168] Alexander Vologodskii and N Cozzarelli. Modeling of long-range electrostatic interactions in DNA. *Biopolymers*, 35(3):289–296, March 1995.

[169] T. E. Cheatham, P. Cieplak, and P. A. Kollman. A modified version of the Cornell et al. force field with improved sugar pucker phase and helical repeat. *J. Biomol. Struct. Dyn.*, 16:845, 1999.

[170] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[171] D. Beglov and B. Roux. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.*, 100:9050–9063, 1994.

[172] P. F. Batcho, D. A. Case, and T. Schlick. Optimized particle-mesh Ewald/multiple-time step integration for molecular dynamics simulations. *J. Chem. Phys.*, 115(9):4003–4018, 2001.

[173] Axel T. Brünger. *X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR*. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992.

[174] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101(5):4177–4189, 1994.

[175] Christopher Maffeo, Binquan Luan, and Aleksei Aksimentiev. End-to-end attraction of duplex DNA. *Nucl. Acids Res.*, 40(9):3812–3821, 2012.

[176] J. Bath and A. J. Turberfield. DNA nanomachines. *Nature Nanotech.*, 2:275–284, 2007.

[177] Arno Pihlak, Göran Baurén, Ellef Hersoug, Peter Lönnerberg, Ats Metsis, and Sten Linnarsson. Rapid genome sequencing with short universal tiling probes. *Nature Biotech.*, 26(6):676–684, 2008.

[178] Hareem T Maune, Si-Ping Han, Robert D Barish, Marc Bockrath, William A Goddard Iii, Paul W K Rothemund, and Erik Winfree. Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nature Nanotech.*, 5(1):61–66, 2010.

[179] Jong-Shik Shin and Niles a Pierce. A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc.*, 126(35):10834–10835, 2004.

[180] Y Benenson, T Paz-Elizur, R Adar, E Keinan, Z Livneh, and E Shapiro. Programmable and autonomous computing machine made of biomolecules. *Nature*, 414(6862):430–434, 2001.

[181] Dmytro Nykypanchuk, Mathew M Maye, Daniel van der Lelie, and Oleg Gang. DNA-guided crystallization of colloidal nanoparticles. *Nature*, 451(31):549–552, 2008.

[182] Sung Yong Park, Abigail K R Lytton-Jean, Byeongdu Lee, Steven Weigand, George C Schatz, and Chad A Mirkin. DNA-programmable nanoparticle crystallization. *Nature*, 451(31):553–556, 2008.

[183] Hin Hark Gan and Tamar Schlick. Chromatin ionic atmosphere analyzed by a mesoscale electrostatic approach. *Biophys. J.*, 99(8):2587–2596, 2010.

[184] D. N. Fuller, J. P. Rickgauer, P. J. Jardine, S. Grimes, D. L. Anderson, and D. E. Smith. Ionic effects on viral DNA packaging and portal motor function in bacteriophage φ29. *Proc. Natl. Acad. Sci. U.S.A.*, 104(27):11245–11250, 2007.

[185] M. Nakata, G. Zanchetta, B.D. Chapman, C.D. Jones, J.O. Cross, R. Pindak, T. Bellini, and N.A. Clark. End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes. *Science*, 318:1276, 2007.

[186] X. Qiu, K. Andresen, L.W. Kwok, J.S. Lamb, H.Y. Park, and L. Pollack. Inter-DNA attraction mediated by divalent counterions. *Phys. Rev. Lett.*, 99:38104, 2007.

[187] Li Li, Suzette a Pabit, Jessica S Lamb, Hye Yoon Park, and Lois Pollack. Closing the lid on DNA end-to-end stacking interactions. *Appl. Phys. Lett.*, 92(22):223901–1–223901–3, 2008.

[188] V.A. Bloomfield. DNA condensation. *Curr. Opin. Struct. Biol.*, 6(3):334–341, 1996.

[189] Gerard C L Wong and Lois Pollack. Electrostatics of strongly charged biological polymers: ion-mediated interactions and self-organization in nucleic acids and proteins. *Annu. Rev. Phys. Chem.*, 61:171–189, 2010.

[190] A. Yu. Grosberg, T. T. Nguyen, and B. I. Shklovskii. Colloquium: The physics of charge inversion in chemical and biological systems. *Rev. Mod. Phys.*, 74(2):329–345, 2002.

[191] Alexei A. Kornyshev. Physics of DNA: unravelling hidden abilities encoded in the structure of 'the most important molecule'. *Phys. Chem. Chem. Phys.*, 12(39):12352–12378, 2010.

[192] A.D. MacKerell Jr and L. Nilsson. Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.*, 18(2):194–199, 2008.

[193] O.K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):108101, 2006.

[194] Albert Galera-Prat, Angel Gómez-Sicilia, Andres F Oberhauser, Marek Cieplak, and Mariano Carrión-Vázquez. Understanding biology by stretching proteins: recent progress. *Curr. Opin. Struct. Biol.*, 20(1):63–69, 2010.

[195] Barry Isralewitz, Jerome Baudry, Justin Gullingsrud, Dorina Kosztin, and Klaus Schulten. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics and Modeling*, 19:13–25, 2001.

[196] E. Stofer, C. Chipot, and R. Lavery. Free energy calculations of watson-crick base pairing in aqueous solution. *J. Am. Chem. Soc.*, 121(41):9503–9508, 1999.

[197] Lourdes Campos, Núria Valls, Lourdes Urpí, Catherine Gouyette, Trinidad Sanmartín, Michael Richter, Elida Alechaga, Alicia Santaolalla, Roberto Baldini, Marc Creixell, Ruth Ciurans, Petr Skokan, Joan Pous, and Juan a Subirana. Overview of the structure of all-AT oligonucleotides: organization in helices and packing interactions. *Biophys. J.*, 91(3):892–903, 2006.

[198] Robert Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, New York, 2001.

[199] Jeff Wereszczynski and Ioan Andricioaei. Free energy calculations reveal rotating-ratchet mechanism for DNA supercoil relaxation by topoisomerase IB and its inhibition. *Biophys. J.*, 99(3):869–878, 2010.

[200] Anil Kumar and Ashok Khanna. Analytical solution of reversible step growth polymerization in semibatch reactors. *J. Appl. Polymer Sci.*, 41(910):2077–2094, 1990.

[201] Salvatore Bommarito, Nicolas Peyret, and John SantaLucia Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucl. Acids Res.*, 28(9):1929–1934, 2000.

[202] Vadim A Vasiliskov, Dmitry V Prokopenko, Andrei Mirzabekov, and D. Parallel multiplex thermodynamic analysis of coaxial base stacking in DNA duplexes by oligodeoxyribonucleotide microchips. *Nucl. Acids Res.*, 29(11):2303–2313, 2001.

[203] Dmitrii V Pyshnyi, Eugenii L Goldberg, and Eugenia M Ivanova. Efficiency of coaxial stacking depends on the DNA duplex structure. *J. Biomol. Struct. Dyn.*, 21(3):459–468, 2003.

[204] Ekaterina Protozanova, Peter Yakovchuk, and Maxim D Frank-Kamenetskii. Stacked–unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.*, 342(3):775–785, 2004.

[205] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucl. Acids Res.*, 34(2):564–574, 2006.

[206] Suresh Manohar, Amber R Mantz, Kevin E Bancroft, Chung-Yuen Hui, Anand Jagota, and Dmitri V Vezenov. Peeling single-stranded DNA from graphite surface to determine oligonucleotide binding energy by force spectroscopy. *Nano Lett.*, 8(12):4365–4372, 2008.

[207] Niharendu Choudhury and B Montgomery Pettitt. On the mechanism of hydrophobic association of nanoscopic solutes. *J. Am. Chem. Soc.*, 127(10):3556–3567, 2005.

[208] Xiongce Zhao. Interaction of C_{60} derivatives and ssDNA from simulations. *J. Phys. Chem. C*, 112(24):8898–8906, 2008.

[209] Robert R Johnson, A T Charlie Johnson, and Michael L Klein. Probing the structure of DNA–carbon nanotube hybrids with molecular dynamics. *Nano Lett.*, 8(1):69–75, 2008.

[210] David Shore, Jorg Langowski, and Robert L. Baldwin. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U.S.A.*, 78(8):4833–4837, 1981.

[211] Michael R Lieber, Yunmei Ma, Ulrich Pannicke, and Klaus Schwarz. Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.*, 4(9):712–720, 2003.

[212] Eric Weterings and Dik C van Gent. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair*, 3(11):1425–1435, 2004.

[213] D M Crothers, J Drak, J D Kahn, and S D Levene. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Meth. Enzym.*, 212(1981):3–29, 1992.

[214] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64:1045–1097, 1992.

[215] Pasquale De Santis and Anita Scipioni. Sequence-dependent collective properties of DNAs and their role in biological systems. *Phys. Life Rev.*, 10(1):41–67, March 2013.

[216] Rodolfo Ghirlando and Gary Felsenfeld. Physical chemistry of nucleic acids and their complexes. *Biopolymers*, 99(12):910–915, December 2013.

[217] David Yu Zhang and Georg Seelig. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chem.*, 3(2):103–113, February 2011.

[218] P E Pestryakov and O I Lavrik. Mechanisms of single-stranded DNA-binding protein functioning in cellular DNA metabolism. *Biochemistry (Moscow)*, 73(13):1388–1404, December 2008.

[219] Shou Waga and Bruce Stillman. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem*, 67:721–751, January 1998.

[220] Lee Zou and Stephen J Elledge. Sensing DNA damage through ATRIP recognition of RPA–ssDNA complexes. *Science*, 300(5625):1542–1548, June 2003.

[221] John W Chase and K R Williams. Single-stranded DNA binding proteins required for DNA replication. *Annu. Rev. Biochem*, 55:103–136, January 1986.

[222] Dongran Han, Suchetan Pal, Jeanette Nangreave, Zhengtao Deng, Yan Liu, and Hao Yan. DNA origami with complex curvatures in three-dimensional space. *Science*, 332(6027):342–6, 2011.

[223] Nadrian C. Seeman. Nanomaterials based on DNA. *Annu. Rev. Biochem*, 79:65–87, 2010.

[224] Hongzhou Gu, Jie Chao, Shou-Jun J. Xiao, and Nadrian C. Seeman. A proximity-based programmable DNA nanoscale assembly line. *Nature*, 465(7295):202–5, 2010.

[225] N. D. Derr, B. S. Goodman, R. Jungmann, A. E. Leschziner, W. M. Shih, and S. L. Reck-Peterson. Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science*, 338(6107):662–5, 11 2012.

[226] Ron O Dror, Robert M Dirks, J P Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.*, 41:429–452, January 2012.

[227] Charles A. Laughton and Sarah A. Harris. The atomistic simulation of DNA. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(4):590–600, July 2011.

[228] Takanori Uzawa, Takashi Isoshima, Yoshihiro Ito, Koichiro Ishimori, Dmitrii E Makarov, and Kevin W Plaxco. Sequence and temperature dependence of the end-to-end collision dynamics of single-stranded DNA. *Biophys. J.*, 104(11):2485–2492, July 2013.

[229] Adelene Sim, Jan Lipfert, Daniel Herschlag, and Sebastian Doniach. Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E*, 86(2):1–5, 2012.

[230] Thomas E Ouldridge, Ard A Louis, and Jonathan P K Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, 134(8):085101, February 2011.

[231] O. Saleh, D. McIntosh, P. Pincus, and N. Ribeck. Nonlinear low-force elasticity of single-stranded DNA molecules. *Phys. Rev. Lett.*, 102(6):068301, 2009.

[232] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117(1):1–19, March 1995.

[233] Robert Zwanzig. Diffusion in a rough potential. *Proc. Natl. Acad. Sci. U.S.A.*, 85(7):2029–2030, April 1988.

[234] Prince E. Rouse. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.*, 21(7):1272–1280, 1953.

[235] A. E. Nkodo, J. M. Garnier, B. Tinland, H. Ren, C. Desruissaux, L. C. McCormick, G. Drouin, and G. W. Slater. Diffusion coefficient of DNA molecules during free solution electrophoresis. *Electrophoresis*, 22:2424, 2001.

[236] Bruno H. Zimm. Dynamics of polymer molecules in dilute solution: Viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.*, 24(2):269, 1956.

[237] Tri T Pham, Ulf D Schiller, J Ravi Prakash, and Burkhard Dünweg. Implicit and explicit solvent models for the simulation of a single polymer chain in solution: Lattice Boltzmann versus Brownian dynamics. *J. Chem. Phys.*, 131(16):164114, October 2009.

[238] Sungchul Hohng, Ruobo Zhou, Michelle K. Nahas, Jin Yu, Klaus Schulten, David M. J. Lilley, and Taekjip Ha. Fluorescence-force spectroscopy maps two-dimensional reaction landscape of the Holliday junction. *Science*, 318(5848):279–83, 2007.

[239] M C Murphy, Ivan Rasnik, Wei Cheng, Timothy M Lohman, and Taekjip Ha. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, 86(4):2530–2537, April 2004.

[240] Dustin B. McIntosh, Gina Duggan, Quentin Gouil, and Omar A. Saleh. Sequence-Dependent Elasticity and Electrostatics of Single-Stranded DNA: Signatures of Base-Stacking. *Biophys. J.*, 106(3):659–666, February 2014.

[241] Maxim Belkin, Christopher Maffeo, David B. Wells, and Aleksei Aksimentiev. Stretching and controlled motion of single-stranded DNA in locally heated solid-state nanopores. *ACS Nano*, 7(8):6816–6824, 2013.

[242] Elizabeth J Denning, U Deva Priyakumar, Lennart Nilsson, and Alexander D Mackerell. Impact of 2′-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.*, 32(9):1929–1943, July 2011.

[243] A. D. MacKerell Jr. and N. K. Banavali. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.*, 21(2):105–120, 2000.

[244] N Foloppe and A D MacKerrell Jr. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, 21:86–104, 2000.

[245] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.

[246] Shuichi Miyamoto and Peter A. Kollman. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water molecules. *J. Comput. Chem.*, 13(8):952–962, 1992.

[247] H.C. Andersen. RATTLE: A "velocity" version of the SHAKE algorithm for Molecular Dynamics calculations. *J. Comp. Phys.*, 52(1):24–34, 1983.

[248] Dirk Reith, Mathias Pütz, and Florian Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.*, 24:1624–1636, 2003.

[249] Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule FRET. *Nature Meth.*, 5(6):507–516, June 2008.

[250] Ruobo Zhou, Michael Schlierf, and Taekjip Ha. Force–fluorescence spectroscopy at the single-molecule level. In *Methods in Enzymology*, volume 475, chapter 6, pages 405–426. Elsevier Inc., 1 edition, January 2010.

[251] A. Bochkarev and E. Bochkareva. From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. *Curr. Opin. Struct. Biol.*, 14(1):36–42, 2004.

[252] W.A. Rosche, A. Jaworski, S. Kang, S.F. Kramer, J.E. Larson, D.P. Geidroc, R.D. Wells, and R.R. Sinden. Single-stranded DNA-binding protein enhances the stability of CTG triplet repeats in Escherichia coli. *J. Bacteriol.*, 178(16):5042–5044, 1996.

[253] R R Meyer and P S Laine. The single-stranded DNA-binding protein of *escherichia coli*. *Microbiol. Rev.*, 54(4):342–380, December 1990.

[254] C A Wu, E L Zechner, and K J Marians. Coordinated leading- and lagging-strand synthesis at the *escherichia coli* DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. *J. Biol. Chem.*, 267(6):4030–4044, February 1992.

[255] Rahul Roy, Alexander G. Kozlov, Timothy M. Lohman, and Taekjip Ha. SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. *Nature*, 2009. available online.

[256] S. Raghunathan, A.G. Kozlov, T.M. Lohman, and G. Waksman. Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat. Struct. Mol. Biol.*, 7(8):648–652, 2000.

[257] Ruobo Zhou, Alexander G Kozlov, Rahul Roy, Jichuan Zhang, Sergey Korolev, Timothy M Lohman, and Taekjip Ha. SSB functions as a sliding platform that migrates on DNA via reptation. *Cell*, 146(2):222–232, July 2011.

[258] David B. Wells, Volha Abramkina, and Aleksei Aksimentiev. Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics. *J. Chem. Phys.*, 127:125101, 2007.

[259] G Krauss, H Sindermann, U Schomburg, and G Maass. *Escherichia coli* single-strand deoxyribonucleic acid binding protein: stability, specificity, and kinetics of complexes with oligonucleotides and deoxyribonucleic acid. *Biochemistry*, 20(18):5346–5352, September 1981.

[260] Dan Gordon, Matthew Hoyles, and Shin-Ho Chung. Algorithm for rigid-body Brownian dynamics. *Phys. Rev. E*, 80(6):066703, 2009.

[261] Andreas Dullweber, Benedict Leimkuhler, and Robert McLachlan. Symplectic splitting methods for rigid body molecular dynamics. *J. Chem. Phys.*, 107(15):5840–5851, 1997.

[262] Sukrit Suksombat, Rustem Khafizov, and Yann Chemla. personal communication, 2014.