SIMULATION STUDIES OF DNA AT THE NANOSCALE: INTERACTIONS WITH PROTEINS, POLYCATIONS, AND SURFACES

by

ROBERT M. ELDER

B.S., Oregon State University, 2009

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Chemical and Biological Engineering

2013

This thesis entitled:

Simulation studies of DNA at the nanoscale: Interactions with proteins, polycations, and surfaces

written by Robert M. Elder

has been approved for the Department of Chemical and Biological Engineering

---

Arthi Jayaraman, Ph.D.

Mark Borden, Ph.D.

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Abstract

Elder, Robert Mitchell (Ph.D., Chemical and Biological Engineering)

Simulation studies of DNA at the nanoscale: Interactions with proteins, polycations, and surfaces

Thesis directed by Professor Arthi Jayaraman

Understanding the nanoscale interactions of DNA, a multifunctional biopolymer with sequence-dependent properties, with other biological and synthetic substrates and molecules is essential to advancing these technologies. This doctoral thesis research is aimed at understanding the thermodynamics and molecular-level structure when DNA interacts with proteins, polycations, and functionalized surfaces.

First, we investigate the ability of a DNA damage recognition protein (HMGB1a) to bind to anti-cancer drug-induced DNA damage, seeking to explain how HMGB1a differentiates between the drugs *in vivo*. Using atomistic molecular dynamics simulations, we show that the structure of the drug-DNA molecule exhibits drug- and base sequence-dependence that explains some of the experimentally observed differential recognition of the drugs in various sequence contexts. Then, we show how steric hindrance from the drug decreases the deformability of the drug-DNA molecule, which decreases recognition by the protein, a concept that can be applied to rational drug design.

Second, we study how polycation architecture and chemistry affect polycation-DNA binding so as to design optimal polycations for high efficiency gene (DNA) delivery. Using a multiscale computational approach involving atomistic and coarse-grained simulations, we examine how rearranging polylysine from a linear to a grafted architecture, and several aspects of the grafted architecture, affect polycation-DNA binding and the structure of polycation-DNA complexes. Next, going beyond lysine we

examine how oligopeptide chemistry and sequence in the grafted architecture affects polycation-DNA binding and find that strategic placement of hydrophobic peptides might be used to tailor binding strength.

Third, we study the adsorption and conformations of single-stranded DNA (an amphiphilic biopolymer) on model hydrophilic and hydrophobic surfaces. Short ssDNA oligomers adsorb to both surfaces with similar strength, with the strength of adsorption to the hydrophobic surface depending on the composition of the DNA strands, i.e. purine or pyrimidine bases. Additionally, DNA-surface and DNA-water interactions near the surfaces govern the adsorption. For longer ssDNA oligomers, the effects of surface chemistry and temperature on ssDNA conformations are rather small, but either the hydrophilic surface or increased temperature favor slightly more compact conformations due to energetic and entropic effects, respectively.

For my parents.

Acknowledgements

First and foremost, I thank my research advisor Dr. Arthi Jayaraman, to whom I am profoundly indebted for her guidance, wisdom, and unwavering positivity in matters scientific, professional, and personal. I also thank the members of my thesis committee for their encouragement, helpful comments, and challenging questions: Drs. Mark Borden, Charles Musgrave, Ted Randolph, and Dan Schwartz.

I thank the National Science Foundation (NSF-DMR 1206894) for financial support and supercomputing resources.

I am grateful to the Jayaraman group members whom I have had the pleasure of working alongside for creating a stimulating and entertaining environment: postdoctoral fellows Dr. Eric Jankowski and Dr. Nitish Nair; graduate students Arezou Seifpour, Mohamed Seyam, Tyler Martin, Hilary Marsh, Charles Starbird, Alex VanFosson, Daniel Johnson, and Carla Estridge; and undergraduate students Xiao Ba, Brandon Lin, and Anna McLeland.

I also thank Dr. Todd Emrick, Dr. Sangram Parelkar, and Rachel Letteri for many insightful and interesting discussions.

As she would expect as *quid pro quo*, and as she deserves, I thank my friend Arezou Seifpour an additional time for sharing her good humor, sage advice, and psychotherapy services, and for faithfully recording the many statements of questionable profundity that I have made over the past few years. I also thank my other good friends who have made my time in graduate school a memorable one.

Finally, I thank my grandparents, aunts, uncles, cousins, and, not least, my parents (all of them) for their love and support, and for giving me the gift of curiosity.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Nucleic acids, like deoxyribonucleic acid (DNA), possess complex structural, dynamic, mechanical, and energetic properties that dictate their interactions with other molecules and control their function. The complexity of these properties arises from the fact that DNA is a multifunctional polymer. Single-stranded DNA (ssDNA) is an amphiphilic, negatively-charged polyelectrolyte with sequence-specific interactions that allow the self-assembly of ssDNA into complex nanoscale structures, such as the well-known double-stranded DNA (dsDNA) double helix. These multifunctional properties and self-assembly behaviors make DNA a promising basis for biomaterials, in addition to the widely recognized importance of DNA in biomedical applications. However, harnessing the complex properties of DNA for these applications requires a nuanced understanding of how DNA interacts with other molecules. The research described in this thesis aims to elucidate the thermodynamics and molecular-level structural details of the interactions of DNA with three relevant classes of biomolecules and substrates: proteins, polycations, and hydrophilic and hydrophobic functionalized surfaces.

While experimental methods have had great success in understanding the interactions of DNA with these classes of molecules, certain types of information are currently difficult to obtain with experimental tools, such as the molecular-level details of the structure, dynamics, and energetics. To complement ongoing experimental efforts with these details, we adopt a molecular simulation approach that allows us to study the specific interactions (e.g. hydrogen bonds, hydrophobic aggregation) between DNA and proteins, polycations, and surfaces. The additional knowledge obtained using simulation methods allows us to explain experimental observations, and, owing to the relatively low expense and

high speed of simulations compared to experiments, allows us to explore future avenues of experimental inquiry and suggest which avenues may be the most fruitful.

In the first section of this thesis (chapters 2 and 3), we focus on our studies of protein-DNA interactions. If one considers DNA as the information encoding mechanism of life, then proteins are analogously the machinery that decodes and maintains the integrity of this information. Proteins maintain the integrity of DNA by recognizing and repairing DNA damage caused by various environmental agents (e.g., ionizing radiation, oxidizing chemicals), which is crucial for the prevention of diseases such as cancer. To better understand the mechanisms by which DNA repair occurs, we examine the interactions between DNA and a protein involved in DNA damage recognition. In particular, in chapters 2 and 3 we study the molecular-level interactions that enable this protein to recognize DNA damage caused by a class of chemotherapy agents, with the aim of providing guidelines for the rational improvement of these anticancer drugs.

In chapter 2, we describe a simulation study of a damage recognition protein, HMGB1a, interacting with DNA damaged by two related platinum-based anticancer drugs. These platinum-based anticancer drugs, cisplatin and oxaliplatin, covalently bind to adjacent guanine bases in DNA to form intra-strand adducts. Differential recognition of drug-DNA adducts by the damage recognition protein HMGB1a is related to differences in the efficacy of these drugs in tumors. Additionally, the DNA sequence context of the adduct (i.e. the bases flanking the adduct) has a marked effect on the drug-DNA binding affinity of HMGB1a. We perform atomistic molecular dynamics simulations of DNA with a cisplatin- or oxaliplatin-DNA adduct in four sequence contexts (AGGC, CGGA, TGGA, TGGT) in the absence and presence of HMGB1a. The structure of HMGB1a-bound drug-DNA molecules is independent of sequence context and drug identity, suggesting that differential recognition cannot be explained by the protein-bound drug-DNA structure. However, we find drug- and sequence-dependent differences in the static structure and conformational dynamics of the drug-DNA molecules in the absence of the protein, and these differences explain some, but not all, experimental observations of differential binding affinity of HMGB1a. Since certain structural features (i.e. minor groove width and helical bend)

of all drug-DNA molecules in the absence of HMGB1a are considerably different from the protein-bound drug-DNA structures, we suggest that HMGB1a must actively deform the DNA during binding. This motivated us to calculate the energetic cost of deforming the structure of the drug-DNA molecules, as we hypothesized that energetic cost to deform the drug-DNA molecule to be an additional factor affecting the ability of HMGB1a to recognize drug-DNA adducts in various sequence contexts.

In chapter 3, we describe our study of the deformability of the drug-DNA molecules, which examines the hypothesis put forth in the conclusion of chapter 2. In this study, we use atomistic molecular dynamics coupled with biasing techniques and string methods for estimating the free energy change of deforming the drug-DNA molecules from the unbound structure (i.e. without HMGB1a) to the protein-bound structure. We link the differential binding affinity of HMGB1a to the free energy of deforming (i.e. bending and minor groove opening) the drug-DNA molecule during HMGB1a binding. Specifically, the minimal binding affinity of HMGB1a for oxaliplatin in the TGGA context is explained by its larger deformation free energy compared to cisplatin-DNA molecules or oxaliplatin in other DNA sequence contexts. A mechanistic explanation for the higher deformation free energy involving the molecular-level details of the drug-DNA molecules is also described. Specifically, methyl groups on neighboring thymine bases in the TGGA context crowd the minor groove and sterically hinder the motion of the diaminocyclohexane ring of oxaliplatin, leading to reduced DNA deformability and lower HMGB1a binding affinity. This finding suggests that similar platinum-based drugs designed to maximize steric hindrance will have lower HMGB1a binding affinity and improved efficacy in some cancers.

In the next section (chapters 4 through 6), we shift our focus to our studies of polycation-DNA interactions, which are motivated by improving gene therapy technologies. Gene therapy aims to deliver therapeutic nucleic acids (DNA or siRNA) into diseased cells to provoke a beneficial change in cell behavior. Nucleic acids cannot enter cells without the aid of a delivery agent, and cationic polymers, or polycations, are a promising class of chemical agents that enable gene delivery. However, the effectiveness of delivery depends on many characteristics of the polycation, such as chemistry (i.e. the chemical nature of the cationic groups, and other moieties such as hydrophobic groups) and architecture

(i.e. the arrangement of the chemical groups within the polycation). We study how the architecture and chemistry of a particularly effective class of polycations affect molecular-level polycation-DNA interactions, and in turn how polycation architecture and chemistry may be used to design more effective polycationic gene delivery agents.

In chapter 4, we describe our study of how polycation architecture affects polycation-DNA interactions. Polycations with varying architectures have been synthesized and used in DNA transfection with varying levels of success, but it is unclear how differences in architecture lead to differences in transfection efficiency. In this study, we connect poly-L-lysine (PLL) architecture to DNA-binding strength, and in turn transfection efficiency, since experiments have shown that a grafted oligolysine architecture [e.g. poly(cyclooctene-*g*-oligolysine)] exhibits higher transfection efficiency than linear PLL. We use atomistic molecular dynamics simulations and free energy calculations to study structural and thermodynamic aspects of polycation-DNA binding of linear PLL and grafted oligolysines with varying graft lengths. Structurally, linear PLL binds in a concerted manner, while each oligolysine graft binds independently of its neighbors in the grafted architecture. The presence of a hydrophobic backbone in the grafted architecture weakens polycation-DNA binding compared to linear PLL by restricting the motion of the individual grafts. The binding free energy of the grafted polycation varies non-monotonically with the graft length due to non-monotonic trends in entropic contributions from ion dissociation. The strength of polycation-DNA binding is strongest for the linear architecture and the grafted architecture at the largest (Poly5) and smallest (Poly2) graft lengths, and is weaker for intermediate graft lengths (Poly3 and Poly4). These trends qualitatively match experimental results in the stability of polycation-DNA complexes and suggest an explanation for the higher transfection efficiency of the weaker-binding polycations (Poly3 and Poly4) grafted oligolysines compared to the stronger-binding polycations (Poly5, Poly2 and linear PLL).

In chapter 5, we continue our study of the effects of polycation architecture on polycation-DNA binding, but we examine structural features of polycation-DNA complexes (polyplexes) at larger length and time scales than in chapter 4. To study these larger polyplexes, we develop coarse-grained models

and conduct Langevin dynamics simulations of lysine-based polycations isolated in solution and forming polyplexes with DNA-like polyanions. We first characterize the structure and flexibility of linear polylysine and oligolysines grafted to a backbone in the absence of a polyanion, and then characterize the structure of complexes formed by these polycations with polyanions of varying flexibility. We find that increasing oligolysine graft length and decreasing graft spacing both increase the size and rigidity of the grafted oligolysine polycations, although they remain less rigid than semiflexible linear polylysine. We also examine the effects of counterions on the polycation structure, and we find that increasing ionic strength or counterion valency both reduce polycation size and eliminate many architecture-dependent effects by screening charge-charge repulsion. Considering the properties of polyplexes, the effects of polycation architecture on polyplex size and flexibility are dependent on the ratio of polycation positive charge to polyanion negative charge (i.e. the charge ratio) and on the flexibility of the polyanion. Polyplex surface charge increases with increasing linear charge density of the grafted polycation, which can be achieved by increasing the graft length or decreasing the graft spacing. Because polyplexes must have a positive surface charge to enter the cell by interacting with negatively charged proteoglycans on the cell surface, modifying the polycation linear charge density may provide a way to modify the interactions of polyplexes with cells, and hence improve transfection efficiency.

In chapter 6, we expand our atomistic study of the interactions between oligopeptide-grafted polycations and DNA by examining the effect of the chemistry and sequence of the oligopeptide grafts. Taking inspiration from nature, we incorporate a nuclear localization sequence (NLS), a cationic oligopeptide that targets molecules for nuclear entry, into the grafted polycation architecture. We use atomistic molecular dynamics simulations to study the effect of peptide chemistry and sequence on the DNA-binding behavior of NLS-based grafted polycations by systematically mutating the residues in the grafts, which are based on the SV40 NLS (peptide sequence PKKKRKV). Replacing arginine (R) with lysine (K) reduces binding strength by eliminating arginine-DNA interactions, but placing R in a less hindered location (e.g., farther from the grafting point to the polycation backbone) has surprisingly little effect on polycation-DNA binding strength. Changing the positions of the hydrophobic proline (P) and

valine (V) residues relative to the polycation backbone changes hydrophobic aggregation within the polycation prior to DNA binding and, consequently, changes the conformational entropy loss that occurs upon polycation-DNA binding. Since conformational entropy loss affects the free energy of binding, the positions of P and V in the grafts affect DNA binding affinity. Therefore, strategically incorporating hydrophobic residues into polycations may provide a route to tune polycation-DNA binding affinity and, in turn, transfection efficiency.

In the third section of this thesis (chapters 7 and 8), we study how nucleic acids interact with functionalized surfaces, which is relevant in many biomedical and biomaterials contexts. For example, a polycationic drug formulation for gene delivery will likely be stored in a container prior to administration, and interactions between the DNA and the walls of the container must not destabilize the drug formulation. Other examples of the relevance of surface-DNA interactions include DNA microarrays, which are a high-throughput technique for simultaneously determining the sequence of many DNA strands, and the self-assembly of DNA-functionalized nanoparticles; in these technologies, DNA hybridization occurs near the surface of a microarray or nanoparticle. Similarly, the self-assembly of DNA strands into predesigned nanostructures (i.e. "DNA origami") might be aided by using a functionalized surface as a substrate to stabilize an existing structure after assembly in solution, or to guide the bottom-up self-assembly of structures directly on the surface. In all of these technologies, understanding the interactions of nucleic acid molecules with surfaces might be used to improve performance. To help develop such an understanding, we study the molecular-level interactions that govern the adsorption and stability of DNA molecules near surfaces with varying chemistry.

In chapter 7, we describe our studies of the adsorption of short single-stranded DNA oligomers to model hydrophilic and hydrophobic self-assembled monolayers (SAMs). We use atomistic molecular dynamics simulations and a technique to enhance conformational sampling (i.e. umbrella sampling) to observe ssDNA oligomers adsorbing to the model SAMs to elucidate the molecular-level interactions between the DNA, the SAM, and water molecules. Because ssDNA is an amphiphilic molecule possessing both hydrophilic and hydrophobic character, the ssDNA oligomers adsorb to both hydrophilic

and hydrophobic SAMs with similar strength. When the ssDNA is in the vicinity of the hydrophilic SAM, it forms attractive hydrogen bonds with the surface species, while facing a repulsive force from the dense, strongly adsorbed water molecules at the surface. When the ssDNA is in the vicinity of the hydrophobic SAM, it experiences weak hydrophobic attraction to the surface, as expected, but also experiences a surprising repulsive force associated with a low-density region of water near the hydrophobic surface. This study reveals details of ssDNA adsorption that may also be relevant to understanding the adsorption of other amphiphilic molecules to surfaces of varying hydrophobicity.

In chapter 8, we extend the previous study by examining the molecular-level interactions of longer ssDNA oligomers than those in chapter 7 with the same model hydrophilic and hydrophobic SAMs. By combining atomistic molecular dynamics simulations with multiple techniques to enhance conformational sampling, we determine the effects of surface chemistry and temperature on the conformations of the ssDNA oligomers. We find that the surface chemistry does not strongly affect the exposure of the relatively hydrophobic nucleobases or the hydrophilic phosphate backbone. Likewise, the surfaces do not strongly affect the preferred size of the ssDNA compared to bulk solution, although the hydrophilic surface does favor slightly more compact DNA conformations than the hydrophobic surface. In more compact conformations, the negative charge of the DNA is more concentrated, and therefore the interactions between the DNA and the hydrophilic surface are stronger, which consequently favors smaller ssDNA sizes. Regardless of the presence or chemistry of a surface, increasing the temperature makes it less unfavorable for the ssDNA to assume both compact and extended conformations, but the energetic cost of assuming a compact conformation is reduced to a greater extent than the cost of assuming an extended conformation. The reason for this difference appears to be the entropically favorable release of DNA-bound water molecules upon assuming a compact conformation. This study is a first step toward understanding how molecular-level DNA-surface interactions affect the behavior of DNA molecules.

In chapter 9, we conclude by summarizing our findings on the interactions of DNA with proteins, polycations, and surfaces. We also describe the limitations and challenges of the molecular simulation

approaches we have employed, and we discuss future directions for each of these studies. In our studies of protein interactions with drug-DNA molecules, despite the limitation in quantitative accuracy that is inherent to the approximate atomistic models we have used, we have shown that steric hindrance between the drug and the DNA may be a viable route for improving efficacy of platinum-based anticancer drugs. Future work could further explore the role of steric hindrance in the success of next-generation anticancer drugs. In our studies of polycation-DNA interactions, we have shown how polycation architecture affects polycation-DNA binding at multiple length scales, and how polycation chemistry affects binding at the atomistic level. The next step in this work is to develop a coarse-grained model to examine how polycation chemistry, in particular hydrophobicity, affects the structure of full-sized polyplexes. Additionally, the effect of hydrophobic aggregation on the strength of polycation-DNA binding could be further explored by introducing other hydrophobic peptides into the grafted polycation architecture. In our studies of surface-DNA interactions, future directions include modifications to the surface chemistry and surface grafting density, both to explore the fundamental effects of these changes and to more closely mimic experimental conditions. Finally, overcoming limitations in the speed of conformational sampling by using newly developed enhanced sampling techniques will allow more efficient study of the effect of surfaces and temperature on the stability of more complicated DNA structures, such as ssDNA hairpins and dsDNA double helices.

Chapter 2

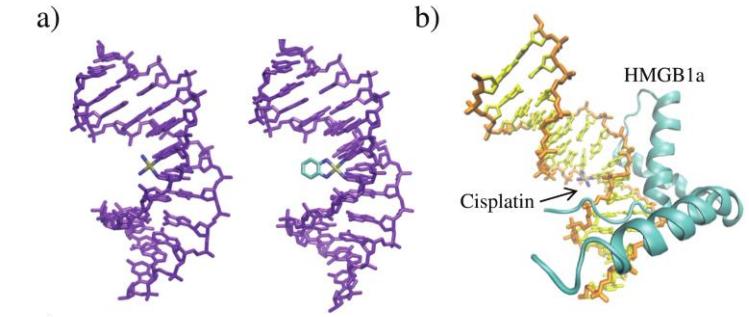
Role of structure and dynamics of DNA with cisplatin and oxaliplatin adducts in various sequence contexts on the DNA binding of HMGB1a

Adapted from: *Mol. Sim.* **38** 793 (2012)

2.1 Introduction

Cis-diamminedichloroplatinum(II) (cisplatin) is one of the most commonly used anti-cancer drugs (Figure 2.1), especially for treatment of testicular, ovarian, head, neck and non-small cell lung cancer.[1-4] DNA is the primary biological target of cisplatin and the cytotoxic mode of action is the formation of cisplatin-DNA adducts. Upon entering the cell, cisplatin binds covalently to DNA at the N7 atom of purine bases, most commonly forming 1,2-*intra*strand adducts with adjacent guanines [1]. The rigid square planar coordination complex of cisplatin induces significant structural distortions in DNA. For example, depending on the bases flanking the adduct (i.e. sequence context), bending toward the major groove by 20°-80°, widening and flattening of the minor groove, and unwinding of the helix by 10°-80° can occur [1-7]. Despite its wide use to treat certain cancers, cisplatin suffers from high tumor resistance and mutagenicity. A good drug is one that has high cytotoxicity, low tumor resistance and low mutagenicity. To circumvent the limitations of cisplatin next generation of platinum-drugs have been developed, of which oxaliplatin (*cis*-1,2-diaminocyclohexaneoxalato platinum (II), Figure 2.1) has been quite successful [7-12]. Although oxaliplatin has a bulky aromatic ring not present in the formula of cisplatin (Figure 2.1), cisplatin and oxaliplatin have been shown to induce similar global structural distortions to the DNA, they have also been shown to exhibit differential cytotoxicity, mutagenicity and tumor resistance [1-6, 10, 12-15]. The differential performance of these drugs has been linked to how various proteins, including high-

mobility group (HMG) box proteins [16-21], mismatch repair (MMR) proteins [22-25], and translesion polymerases [26-29] discriminate between the drug-DNA adducts. For example, HMG proteins bound to the drug-DNA adducts may shield the DNA from nucleotide excision repair (NER), activate signaling pathways leading to cancer cell death, or hijack transcription factors needed for cell division, all of which are important to maintain high cytotoxity and low tumor resistance [21, 25, 30-33].



c)

5’	C1	C2	T3	C4	**A/C/T5**	**G6**	**G7**	**C/A/T8**	C9	T10	C11	C12	3’
3’	G24	G23	A22	G21	T/G/A20	C19	C18	G/T/A17	G16	A15	G14	G13	5’

Figure 2.1 (a) Structures of cisplatin (left) and oxaliplatin (right) bound to DNA (PDB ID: 2NPW, 1PGC); (b) Rendering of HMGB1a bound to cisplatin-DNA (PDB ID: 1CKT); and (c) nomenclature of sequence context. In (c), the platinated guanine bases (G6 and G7) are underlined, and the sequence context (AGGC, CGGA, TGGA, TGGT) is shown in bold text.

Repair proteins with high mobility group (HMG) domains have been shown to bind with differential affinity to cisplatin- and oxaliplatin-DNA adducts [6, 16-20, 34]. Although the exact mechanism connecting HMG protein binding to drug-resistance is not known, understanding how HMG proteins recognize drug-DNA adducts is an important step toward designing new drugs that defy resistance mechanisms. In this regard, particular attention has been paid to HMGB1a, the A-domain of the high-mobility group protein HMGB1, as a model for the binding of other HMG proteins to drug-DNA adducts. Past experimental work has shown that the binding of HMGB1a to drug-damaged DNA is driven by the pre-bent structure and the intercalation of a phenylalanine residue into the ‘hydrophobic notch’ created between the damaged base pairs [16, 35]. It has also been shown that the affinity of HMGB1a for the drug-DNA adducts strongly depends on both the identity of the drug (i.e. cisplatin or oxaliplatin) and on the bases flanking the damaged bases (the *sequence context*) [11, 36-38]. In two studies on protein

binding affinity for cisplatin-DNA adducts, Lippard and coworkers systematically tested the effect of the 16 possible sequence contexts on HMGB1a binding affinity [36, 37], finding that the base 3' of the cisplatin-DNA adduct (i.e. Y in XGGY where the cisplatin is bound to GG) significantly impacts binding affinity with the relation A > T > G > C. These studies also show that the base 5' of the adduct has a significant impact when the 3' base is either G or C with the relation A > C ≈ G > T. Interestingly, experiments to determine the structural distortions introduced by cisplatin show that the global bending and unwinding are independent of sequence context, suggesting that *local* rather than global distortions may be responsible for differential binding affinity [39]. Additional experiments by Lippard and coworkers demonstrate that HMGB1a has a greater affinity for cisplatin-DNA (Cp-DNA) adducts than for oxaliplatin-DNA (Ox-DNA) adducts in both the AGGC and TGGA sequence contexts, where the central GG pair is bound to the drug [38]. Most interestingly, their work shows that, while cisplatin adducts are equally well-recognized in both contexts, Ox-TGGA (oxaliplatin in the TGGA context) is completely unrecognized by HMGB1a [38]. Furthermore, they show that Ox-TGGA is also unrecognized by two other proteins that otherwise show affinity for the drug-DNA adducts [38]. Experimental work by Brabec and coworkers [11] confirms that the affinity of HMGB1a is greater for cisplatin-DNA adducts than oxaliplatin-DNA adducts in all contexts tested; however they find that the binding affinity is approximately equal in the AGGC and CGGA contexts and greater in the TGGT context for both drugs, which does not match the trends seen by Lippard and coworkers in Ref. [36, 37].

To test if varying binding affinities could be related to structures of the drug-DNA molecules, experimental structures have been determined for cisplatin- and oxaliplatin-damaged DNA in several contexts: Cp-AGGC [40], Ox-AGGC [41], Cp-TGGT [2, 4, 42, 43], Ox-TGGT (PDB ID: 2K0T), as well as Cp-TGGA bound by HMGB1a [16]. Comparison of the experimental structures of Cp-DNA and Ox-DNA in the AGGC context reveals significant structural differences between the drug-DNA adducts, such as differences in the twisting and helical axis bend [40]. Likewise, the structures of Cp-DNA in the AGGC and TGGT contexts show substantial sequence-dependent structural differences, with the NMR solution structures of Cp-AGGC and Cp-TGGT having helical axis bends of 34° and 74°, respectively [4,

40]. However, comparison of x-ray crystal structures and NMR solution structures reveals striking differences: for instance, the x-ray crystal structure of Cp-TGGT has a helical axis bend of 25° while the NMR solution structure has a bend of 74° [4, 42]. This disparity could be attributable to stabilization of high-energy conformations by crystal packing effects (e.g. intermolecular hydrogen bonding), emphasizing the fact that the limitations of each method must be considered carefully when comparing results from different sources [42].

Computer simulations have also sought to elucidate the basis for both drug- and sequence-dependent recognition of drug-DNA adducts by HMGB1a. The earliest efforts in this regard were the development of a molecular dynamics force field to describe the structure of cisplatin-DNA and oxaliplatin-DNA adducts [8, 44, 45]. Increasing computational power has allowed both cisplatin and oxaliplatin to be studied in several sequence contexts. Molecular dynamics simulations of Cp-CGGA have been performed, revealing the expected increase in minor groove width and increased conformational freedom near the drug-DNA adduct [6]. Hybrid quantum mechanics/molecular mechanics (QM/MM) have been applied to Cp-TGGT and HMGB1a-bound Cp-TGGA, showing a high helical axis bend when the DNA is bound by the HMG protein [46]. Unique structural and dynamic differences between cisplatin- and oxaliplatin-DNA adducts have been revealed through the use of molecular dynamics: the structures of the drugs show significant differences from undamaged DNA and from each other, including hydrogen bonds with adjacent bases that cause distinct structural distortions [7]. Finally, recent work by Dokholyan and coworkers has compared the structures and conformational dynamics (i.e. the time spent in different structural states) of Cp-DNA and Ox-DNA in two sequence contexts (AGGC and TGGA) that are recognized differently by HMGB1a [15, 36-38]. Dokholyan and coworkers found that, while the average structures of Cp-DNA and Ox-DNA in the AGGC and TGGA contexts were mostly similar, dynamic states unique to Cp-DNA could explain the differential recognition of the drug-DNA adducts. The dynamic states sampled by Cp-DNA are not seen in Ox-DNA because the bulky cyclohexane moiety of oxaliplatin may prevent Ox-DNA from exploring these states [15]. To the best of our knowledge such studies of structure and conformational dynamics of drug-DNA adducts have only

been studied in AGGC and TGGA contexts [15] and not in other sequence contexts. As described above, experimental studies have quantified HMGB1a binding affinity for cisplatin- and oxaliplatin-DNA adducts in various sequence contexts, and the experimental trends in binding affinity of the two drugs are non-trivial [36-38]. This motivates a comprehensive study of cisplatin- and oxaliplatin-DNA adducts in multiple sequence contexts both in the presence and absence of HMGB1a to form a *complete* understanding of the sequence-and drug- dependent differences in structure and conformational dynamics of drug-DNA adducts to show if they can explain the basis of differential recognition of the various drugs in different sequence contexts by HMG proteins.

In this paper we conduct atomistic molecular dynamics simulations of double-stranded DNA with intrastrand adducts formed by cisplatin and oxaliplatin in the unbound state and *bound to* the protein HMGB1a. We study the drug-adducts in four sequence contexts (AGGC, CGGA, TGGA, and TGGT); these sequences were chosen to elucidate the structural effects of A, T and C bases on both 5' and 3' side of the drug-DNA adducts. We characterize the structure and conformational dynamics of these systems to determine how these features relate to the binding affinity of HMGB1a for the drug-DNA adducts.

2.2 Methods

2.2.1 Simulation methods

We performed unrestrained atomistic molecular dynamics simulations using NAMD 2.7 [47]. We simulated B-DNA (without any drugs), and two drug-damaged DNA systems: cisplatin (Cp-DNA) and oxaliplatin (Ox-DNA) in the absence and presence of bound HMGB1a. We used the Amber *parm99* force field with the *SB* (protein) and *bsc0* (nucleic acid) modifications to parameterize DNA, protein, explicit TIP3P-model water molecules, and sodium counterions. Force field parameters and partial charges for the platinum-based drugs are available in literature [7, 8, 45]. Three modifications were made to the existing parameters for the platinum-based drugs. First, the van der Waals parameters of the platinum atom were changed from those recommended by Scheeff *et al.* to more physically realistic values: a well depth (ε) of 0.2 and radius (σ) of 1.7. These values correspond to those of Bondi and have been recommended in

literature [48, 49]. Second, the van der Waals parameters of the ammine nitrogen atoms were updated to *parm99* values for consistence with the rest of the force field. Third, we found that the improper torsions used to keep the platinum atom in the plane of each guanine base (denoted as CK-NB1/NB2-CB-PT in Ref. [8, 45] produced values of G6-G7 roll that were substantially higher than the experimental structures, and consequently we reduced the value of the force constant from 10 kcal/mol to 7.5 kcal/mol after performing some exploratory simulations. Using a lower value of this force constant more satisfactorily reproduces the roll angle of the experimental structures. Details of the procedure we employed to calibrate this value are available in section 2.6.2.

For systems in the absence of protein, referred to as *unbound-drug-DNA* systems, we used available NMR solution structures of Cp-DNA and Ox-DNA (PDB ID: 2NPW; PDB ID: 1PGC) to develop starting structures for both drugs in four sequence contexts by manually replacing the base pairs flanking the platinated guanines with base pairs from B-DNA using VMD 1.8.7 [50]. The four sequences studied here are: 5′-d(CCTC**A**<u>**GG**</u>**C**CTCC)-3′, 5′-d(CCTC**C**<u>**GG**</u>**A**CTCC)-3′, 5′-d(CCTC**T**<u>**GG**</u>**A**CTCC)-3′, 5′-d(CCTC**T**<u>**GG**</u>**T**CTCC)-3′, where the sequence context is indicated by bolded text and the platinated guanines are underlined (Figure 2.1). As in the experimental structures, the drugs were covalently bound to the N7 atoms of the central adjacent guanines (G6 and G7). For systems in which the drug-DNA complex is bound by HMGB1a, referred to as *bound-drug-DNA* systems for simplicity, we used the experimental structure of HMGB1a bound to Cp-TGGA (PDB ID: 1CKT) and the same method as above to build starting structures of drug-DNA (in the same four contexts) bound by HMGB1a. In our simulations of bound-drug-DNA systems we use a 16bp DNA instead of 12bp used in the unbound-drug-DNA, because the experimental structure of HMGB1a bound drug DNA has a 16bp. The sequence of this 16bp nucleic acid was identical to the 12bp DNA used in the drug-DNA (i.e. in the absence protein) simulations except for the addition of 2 G-C base pairs on each end of the strand: for example, the AGGC context in the protein-simulations reads 5′-d(CCCCTC**A**<u>**GG**</u>**C**CTCCCC)-3′. We note that, although the number of base pairs is different in the bound-drug-DNA systems, we retain the same nomenclature as in the drug-DNA for base numbering for clarity. For example, although the platinated guanines are actually

G8 and G9 in the bound-drug-DNA systems, we still refer to these bases as G6 and G7. Undamaged, canonical B-DNA was generated using Nucleic Acid Builder (*nab*) in AmberTools 1.4. In all cases where a molecule was manually edited (e.g. mutation of a base pair), the edited structure was subjected to 500 steps of conjugate gradient minimization with NAMD 2.7 to eliminate any extremely unfavorable geometry before being used for simulation. Counterions (Na^+) were added to neutralize the system charge. Systems were solvated with TIP3P water molecules with a minimum distance of 10 Å from the solute to the edge of the box; in each case, approximately 5000 to 5500 water molecules were added.

Prior to production simulation, each system was equilibrated with NAMD 2.7 using the protocol described in section 2.6.1. Subsequently, production simulations were performed: for unbound drug-DNA systems we performed five unrestrained, fully solvated molecular dynamics simulations of 20 ns each for Cp-DNA, Ox-DNA, and undamaged B-DNA; for bound-drug-DNA systems, we performed only three molecular dynamics simulations due to increased computational expense of protein and DNA systems as compared to DNA only systems. Snapshots were recorded every 2 ps, and the trajectories were judged to have equilibrated after 2 ns, at which point the RMSD with respect to the initial structure had reached a stable value, resulting in a total of 45000 snapshots for each drug-DNA system and 27000 snapshots for each HMGB1a-drug-DNA system. Production MD was performed in the NPT ensemble, using Langevin dynamics to maintain a temperature of 300 K and a pressure of 1 atm. The SHAKE algorithm was used to restrain the lengths of all bonds to hydrogen, and a timestep of 2 fs was used. Each production run used the same minimized and relaxed structure as starting coordinates, which were generated using the equilibration protocol in section 2.6.1, but the initial velocities were randomized during the heating procedure at the beginning of production simulation.

2.2.2 Analysis methods

We use Curves+ and VMD for the various structural analyses and to generate snapshots of the simulations [51]. Intra- and inter- base pair rotational and translational parameters backbone angles, sugar pucker, width and depth of the major and minor grooves, and the total helical axis bend, which describe the local and global structure of DNA, were calculated using the Curves+ program on simulation

trajectories and experimental structures. In all cases, the first and last base pairs in the strand were excluded from analysis because of unrealistic distortions caused by end effects. Intra- and inter-base pair rotational and translational helical parameters fully describe the relative orientations of adjacent bases and base pairs. For example, the *roll* angle quantifies the dihedral angle between two adjacent base pairs. Backbone angles describe the conformation of the DNA phosphate backbone. All of these structural parameters are relevant to predicting protein-DNA interactions: for instance, numerous structure-specific proteins recognize bending of the helical axis toward the major groove and the concomitant widening and flatting of the minor groove [4, 5, 10]. Backbone angles and sugar pucker are relevant because they help explain the connections between related structural distortions. Histograms were assembled by averaging the values obtained from 3 (for protein-drug-DNA systems) or 5 (for drug-DNA in the absence of protein) independent trials conducted with each system. The frequency distributions are normalized to 100%, and the error bars indicate the standard error of the mean of the multiple trials.

Even though we calculate all the structural parameters (listed in section 2.6.3), in this paper we report the parameters considered important for the binding affinity of HMGB1a for drug-DNA complexes: G6-G7 roll, total helical axis bend, minor groove width at the G7-C18 base pair, and G7-X8 twist, where "X" indicates any nucleotide. G6-G7 roll is important because higher values of roll could allow a phenylalanine residue of HMGB1a to intercalate more easily into the 'hydrophobic notch' at the platinated site (i.e. the G6-G7 base pair step). Helical axis bend is important because HMG proteins are known to recognize bent DNA. The minor groove width is important because in the experimental structures of HMG-Cp-DNA the minor groove is widened and one of the helices of HMG protein placed along the minor groove. Twist is important because bending induces slight unwinding (lower twist) of the DNA helix. We also highlight in supplementary information (section 2.6.9 and 2.6.10) particularly significant structural differences in other parameters, if any, to see if these differences could affect the binding of drug-DNA adducts.

To validate our approach and force field parameters we first compared our structural results from our simulations to known structures from the protein data bank: unbound-Cp-AGGC (PDB ID: 2NPW),

unbound-Ox-AGGC (PDB ID: 1PGC), HMGB1a bound-Cp-TGGA (PDB ID: 1CKT), and unbound Cp-TGGT (PDB ID: 3LPV). Figure 2.10 (see supplementary section 2.6) shows that the distributions of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 base pair, and G7-X8 twist from our simulations are generally in good quantitative agreement with the experimental structures. This validates our approach and force field parameters for these drug-DNA systems in the absence and presence of HMGB1a.

We compare all of the structural and dynamics results of Cp-DNA and Ox-DNA in the presence and absence of HMGB1a to experimental HMGB1a binding affinity studies in Ref. [11, 36-38]. We choose these studies because they provide the most comprehensive set of data on HMGb1a binding affinity to Cp-DNA and Ox-DNA in various sequence contexts. To account for differing DNA and protein concentrations between the HMGB1a binding affinity experiments in Ref. [36-38] we have calculated the dissociation constant (K_D, nM) using the method described in Ref. [15] (see Table 2.5 in supplementary section 2.6 for K_D values). We note that there are some apparent inconsistencies in the binding affinity results of Ref. [36, 38], in that the latter shows that Cp-AGGC and Cp-TGGA are recognized equally by HMGB1a, while the former shows that Cp-TGGA is recognized substantially better than Cp-AGGC. Likewise, Ref. [11] shows that Cp-AGGC and Cp-CGGA are bound with approximately equal affinity, whereas Ref. [36] shows that HMGB1a has greater affinity for Cp-CGGA than for Cp-AGGC. The DNA sequences in these experimental studies are not identical *outside* of the immediate sequence context (i.e. XGGY), and it is possible that the differing DNA sequences are responsible for the discrepancies in binding affinity [11, 36-38]. One might expect that the difference in sequence may be particularly significant because the sequence used in Ref. [11] introduces a T-A base pair 3’ of the XGGY context while the sequence in Ref. [36-38] have a C-G base pair 3’ of the XGGY. T-A base pairs, having only two hydrogen bonds compared to the three hydrogen bonds in C-G pairs, are known to explore a wider range of conformations, and therefore this change in sequence could alter the structure [52]. To test this hypothesis, we simulated oxaliplatin in the AGGC and TGGA contexts in the 5’-d(CCTCXGGYCTCC)-3’ sequence, the structure of which has been determined experimentally, and in

the 5'-d(CCTTCTXGGYTCTTC)-3' sequence, which has been used in binding affinity experiments. We found the structural differences between the two sequences to be minor, as shown in Figure 2.11 (supplementary section 2.6), although it is possible that the difference in sequence changes other properties of the DNA that are relevant to binding affinity, such as deformability. We have used the 5'-d(CCTCXGGYCTCC)-3' sequence throughout this paper.

2.3 Results

2.3.1 Structures of cisplatin-DNA and oxaliplatin-DNA bound by HMGB1a

First, we discuss the effects of drug adducts in different sequence contexts on the structure of cisplatin-DNA and oxaliplatin-DNA when bound to the HMGB1a protein. Characterizing the bound drug-DNA structure can explain if there are inherent differences in the structure of the bound DNA between the various drugs and sequence contexts that could favor one drug/sequence over others, leading to differential recognition by the protein.

The first row of Figure 2.2 shows the schematic for G6-G7 roll, helical bend, minor groove width at the G7-C18 base pair and G7-X8 twist, followed by the frequency distributions of each of the four structural parameters for HMGB1a-bound cisplatin-DNA (second row) and oxaliplatin-DNA (third row) in the four sequence contexts: AGGC (black line with circles), CGGA (red line with squares), TGGA (green line with upward triangle), and TGGT (blue line with downward triangle). All of the structural parameters shown in Figure 2.2 show the same distribution irrespective of the drug or the sequence context. For both drugs in all four contexts, the mean G6-G7 roll (Figure 2.2a and 2.2e) is approximately 55°, and the mean total helical axis bend (Figure 2.2b and 2.2f) is approximately 50°. The minor groove width at the G7-C18 base pair (Figure 2.2c and 2.2g) has a narrower distribution with mean at 10-12 Å. The minor groove widths at base pairs that have greater contact with the protein (i.e. those 3' of the drug-DNA adduct: X8-N17, C9-G16, and so on) are wider (Figure 2.12 in supplementary section 2.6) than those 5' of the drug adduct. The twist at G7-X8 exhibits a distribution with mean of 27° and deviation of 5-7°. There are a few minor differences in the other structural parameters listed in section 2.6.3 (data not

shown), but most of the differences lie well within the error bars. Since the protein-bound drug-DNA structure is essentially independent of sequence context and identity of the drug we conclude that the *bound* drug-DNA structures cannot explain the differences in the protein recognition event and binding process. This motivates us to investigate the structure of the drug-DNA molecule *not bound* to HMGB1a, discussed next.

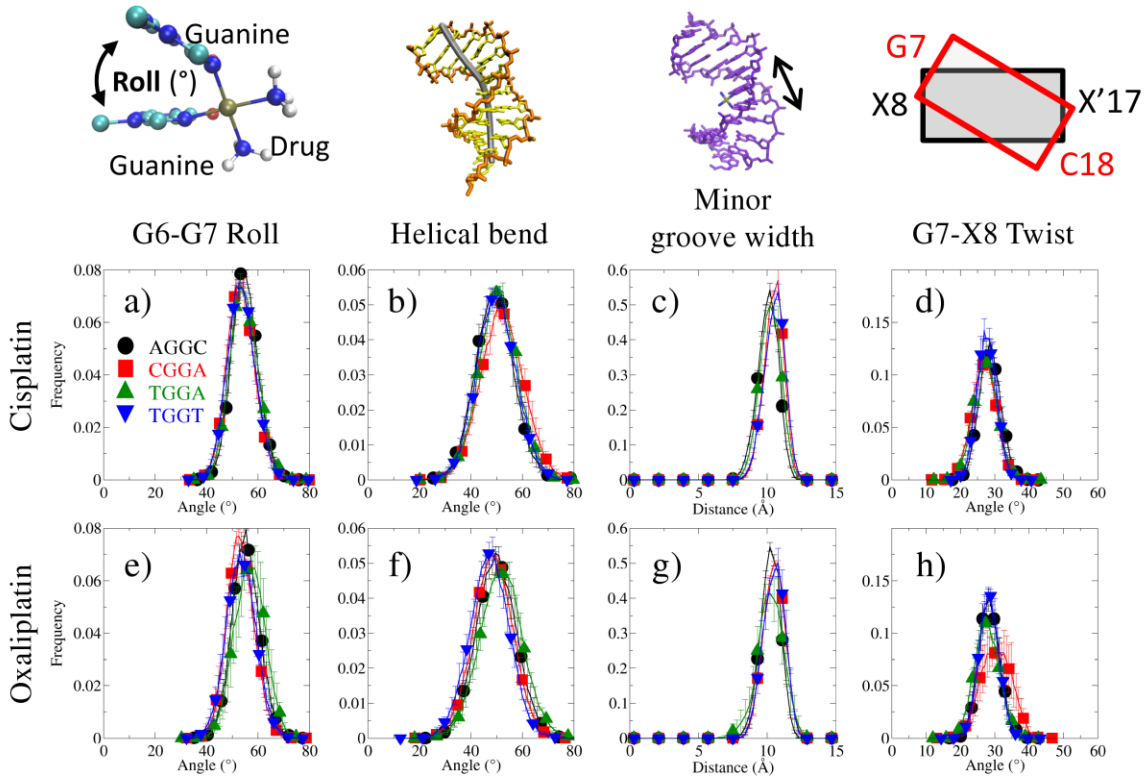


Figure 2.2 Ensemble average distributions of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 base pair, and the G7-X8 twist for B-DNA, Cp-DNA, and Ox-DNA for the AGGC, CGGA, TGGA, and TGGT sequence contexts bound by HMGB1a.

2.3.2 Structures of B-DNA, cisplatin-DNA, and oxaliplatin-DNA not bound by HMGB1a

For undamaged B-DNA (Figure 2.3a-d), the G6-G7 roll and helical bend of all four sequences is identical, but there are small sequence-dependent differences in the minor groove width at the G7-C18 base pair and in the G7-X8 twist. The AGGC context has a slightly narrower minor groove than other sequences, as evidenced by a slight shift of the distribution (black-circle line in Figure 2.3c) to lower values of minor groove widths. The CGGA (red-square line) and TGGA (green-upward triangle line)

contexts have slightly higher twists than other sequences (Figure 2.3d). The structural differences between the different sequence contexts of the central G6 and G7 bases can be traced to the chemical nature (pyrimidine or purine) of the X5 and X8 bases (XN refers to any base at the N^{th} position) flanking G6 and G7. For example, for the X5-G6 slide (see Figure 2.13 in supplementary section 2.6) we observe that TGGA and TGGT are identical while AGGC and CGGA are different because TGGA and TGGT share the same base pair at the X5 position while the other two sequences have different base pairs. For the X8-C9 slide (see Figure 2.13 in supplementary section 2.6) we observe that CGGA and TGGA are identical while the other two sequences differ, which is again due to the identity of the base (A in this case) in the X8 position. These structural differences seen in B-DNA caused by sequence context are minor compared to those seen in cisplatin-DNA (Cp-DNA) and oxaliplatin-DNA (Ox-DNA) shown in the second and third rows in Figure 2.3 and discussed next.

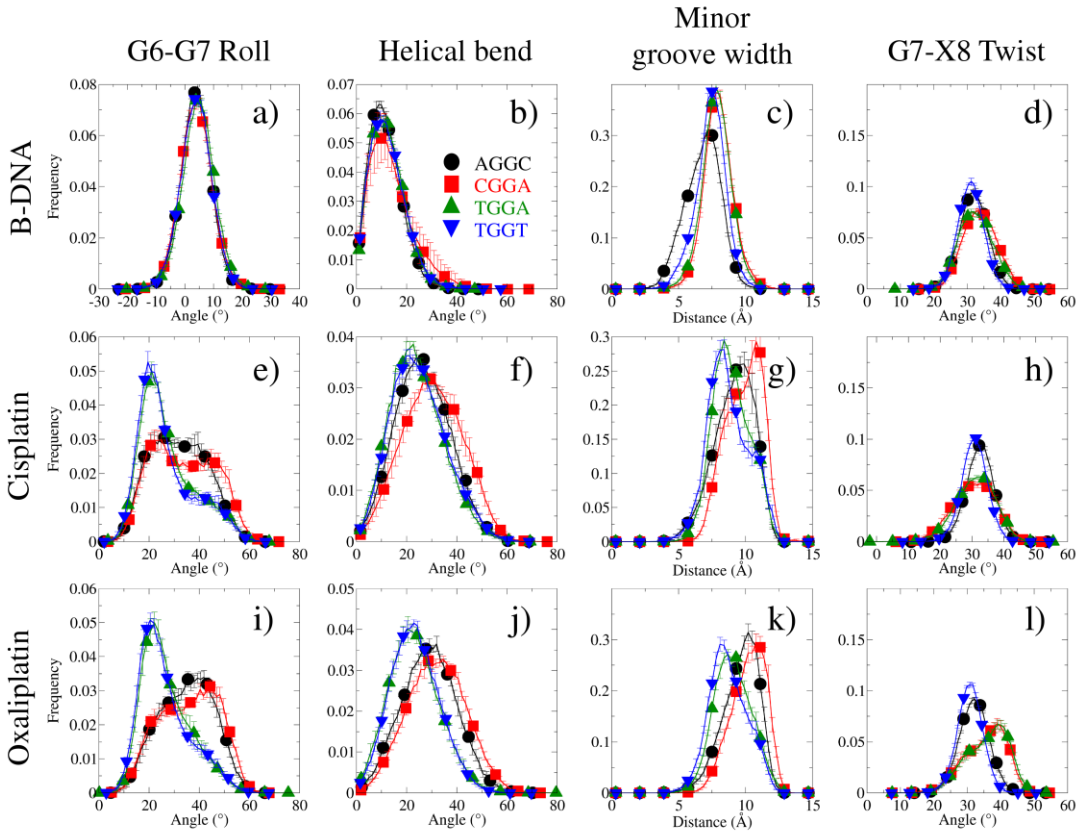


Figure 2.3 Ensemble average distributions of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 base pair, and the G7-X8 twist for Cp-DNA and Ox-DNA for the AGGC, CGGA, TGGA, and TGGT sequence contexts.

For Cp-DNA the distributions of G6-G7 roll (Figure 2.3e), helical axis bend (Figure 2.3f), and minor groove width at the G7-C18 base pair (Figure 2.3g) for Cp-AGGC (black circles) and Cp-CGGA (red squares) are similar and the distributions for Cp-TGGA (green upward triangle) and Cp-TGGT (blue downward triangle) are similar. Cp-AGGC and Cp-CGGA explore conformations with higher G6-G7 roll (ranging from 10° to 60°) and minor groove width (ranging from 5-13 Å) than Cp-TGGA and Cp-TGGT, with Cp-CGGA (red squares) reaching higher values of G6-G7 roll and minor groove width at G7-C18 than the other three sequences. Cp-CGGA (red squares) also samples higher values of total helical bend than the other three sequences. For the G7-X8 twist (Figure 2.3h), we observe that while the mean value of about 30° is the same in all four contexts, the sequences with an adenine at the X8 position (i.e. CGGA, TGGA) have a slightly wider distribution of twist angles than the AGGC and TGGT. And unlike other parameters the G7-X8 twist distributions of TGGA and TGGT do not overlap, but rather, AGGC matches TGGT and CGGA matches TGGA. This pattern is because the A8 bases of CGGA and TGGA match and the pyrimidine (C8, T8) bases of the AGGC and TGGT sequences match. Since higher G6-G7 roll angle, higher helical bend angle, wider minor groove width and lower G7-X8 twist are more favorable for binding HMGB1a protein, it appears that the presence of a T immediately 5' of the Cp-DNA adduct (i.e. TGGA and TGGT) results in less favorable structure for HMGB1 binding than other base pairs (i.e. AGGC and CGGA).

For Ox-DNA adducts (Figures 3i-3k), we observe that the trends in G6-G7 roll, helical bend and minor groove width at G7-C18 are similar to those observed for Cp-DNA adducts. For G6-G7 roll (Figure 2.3i), helical axis bend (Figure 2.3j), and minor groove width (Figure 2.3k), the distributions of Ox-AGGC and Ox-CGGA are similar although Ox-CGGA reaches slightly higher values than Ox-AGGC, while the distributions of Ox-TGGA and Ox-TGGT are identical. The helical bend distributions of Ox-AGGC (black circles) and Ox-CGGA (red squares) are both substantially higher than the distributions of Ox-TGGA and Ox-TGGT, whereas for Cp-DNA only Cp-CGGA had greater helical bend than the other sequences. We also observe a much different pattern in the G7-X8 twist (Figure 2.3l) distributions with

Ox-DNA than with Cp-DNA: Ox-CGGA and Ox-TGGA are significantly over-twisted as compared to Ox-AGGC and Ox-TGGT, whereas the mean twist angle at G7-X8 was essentially the same for all sequences for Cp-DNA. Higher twist could negatively impact binding of HMGB1a by hindering interactions of the DNA with the protein. To summarize, some of the structural distortions of Ox-DNA have similar trends as Cp-DNA, with a T base immediately 5' of the oxaliplatin adduct (i.e. TGGA and TGGT) introducing structural effects that are less favorable for binding. However, the combination of the Ox-DNA adduct and the A base 3' of the adduct (i.e. CGGA and TGGA) produces unique conformations, such as higher twist angle at G7-X8 that are different from those seen in Cp-DNA in the same sequence context.

It is evident in Figure 2.3 that the G6-G7 roll, the total helical axis bend, and the minor groove width are all substantially greater for drug-DNA than for undamaged B-DNA, and the G7-X8 twist shows non-trivial distortions. Some of the other structural parameters (e.g. slide) are also affected by the presence of the drug in the different sequence contexts, and these differences are shown in the supplementary information (see Figures 2.14 and 2.15 in supplementary section 2.6).

2.3.3 Comparison of the structures of B-DNA, cisplatin-DNA, and oxaliplatin-DNA in the unbound states to the HMGB1a-bound states

To show the differences between unbound undamaged B-DNA, unbound drug-DNA and bound drug-DNA, in Figure 2.4a-d we plot the distributions of G6-G7 roll of undamaged B-DNA, unbound-drug-DNA and bound-drug-DNA systems. Qualitatively, Figure 2.4a-d demonstrates that in all four sequence contexts the unbound drug-DNA structures (red circles for Cp-DNA and blue circles for Ox-DNA) are intermediate between healthy B-DNA (black line) and the HMGB1a-drug-DNA complexes (red squares for Cp-DNA and blue squares for Ox-DNA). Additionally, in AGGC (Figure 2.4a) and CGGA (Figure 2.4c) sequence contexts, both unbound-Cp-DNA and unbound-Ox-DNA sample G6-G7 roll values that are closer to the final bound state than in the TGGA (Figure 2.4d) and TGGT (Figure 2.4b) contexts. Distributions of helical axis bend and minor groove width follow a similar trend, with AGGC and CGGA

being closer to the bound structure than TGGA and TGGT (not shown). However, the G7-X8 twist (Figure 2.4e-h) shows a different trend from the other structural parameters. The distributions of G7-X8 twist for unbound Cp-DNA (all contexts), Ox-AGGC, and Ox-TGGT are identical to the corresponding B-DNA distributions and somewhat higher than in the bound state. However, Ox-CGGA and Ox-TGGA exhibit twist that is substantially higher than either the B-DNA or HMGB1a-bound drug-DNA structures. Ox-CGGA and Ox-TGGA are significantly more twisted than Ox-AGGC and Ox-TGGT, owing to the identity of the base pair in the 8th position of each sequence (i.e. A8 of CGGA and TGGA),while all the sequences have the same average G7-X8 twist for cisplatin-DNA irrespective the base pair in the 8th position of the sequence context. The over-twisting of Ox-CGGA and Ox-TGGA has been linked in section 2.3.4 to a specific hydrogen bond seen only in case of Ox-CGGA and Ox-TGGA, and not in Cp-CGGA and Cp-TGGA.

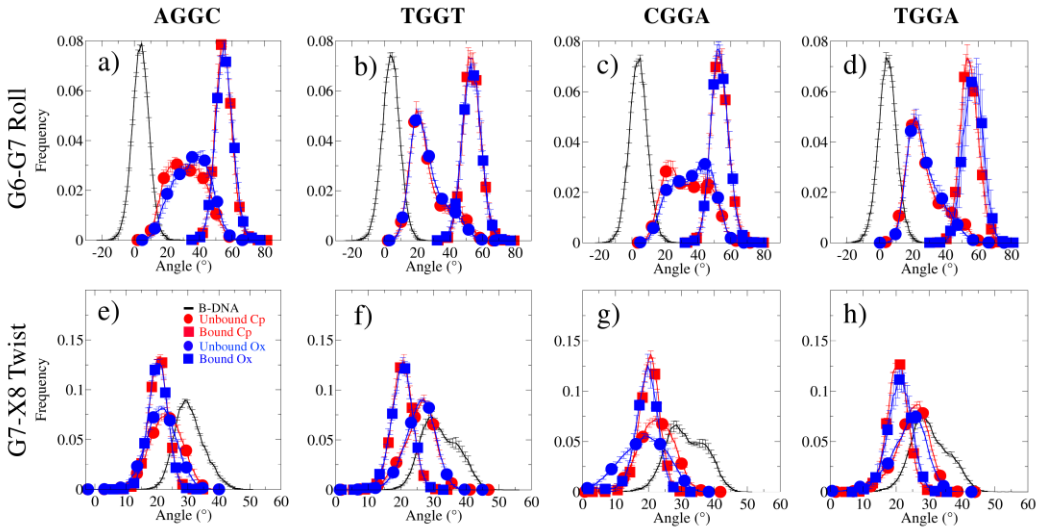


Figure 2.4 Ensemble average frequency distributions of G6-G7 roll and G7-X8 twist for B-DNA, Cp-DNA, and Ox-DNA in the (a, e) AGGC, (b, f) TGGT, (c, g) CGGA, and (d, h) TGGA sequence contexts either free in solution (*unbound*) or bound by HMGB1a (*bound*). Note that the black line representing B-DNA is obscured by the unbound-drug-DNA data in (e-h). Since high G6-G7 roll (closer to 60°) and low helical twist (closer to 30°) both facilitate binding of HMGB1a, a pattern emerges: stated succinctly, Ox-AGGC has high roll and low twist, Ox-CGGA has high roll but high twist, Ox-TGGT has low roll but low twist, but Ox-TGGA has low roll (~20°) *and* high twist (~40°). Of the sequences, only Ox-TGGA has both structural characteristics that could negatively impact binding, which could account for the poor recognition of this sequence by HMGB1a.

Since the structural distributions of the unbound-drug-DNA systems in certain sequence contexts are similar to the distributions of bound-drug-DNA systems, the drug-DNA adducts could be recognized

better in those sequence contexts. However, previous work has shown that the ensemble average distributions can disguise *rare and dynamic states* that may be important for binding affinity [7, 15]. These rare, dynamic states can be associated with specific patterns of hydrogen bonding between the drug and adjacent base pairs, and as such we next discuss the hydrogen bonding patterns and associated structural changes of cisplatin- and oxaliplatin-DNA adducts in all four sequence contexts.

2.3.4 States with hydrogen bonds between the drug and adjacent DNA bases in the unbound drug-DNA

We find that both drugs form unique hydrogen bonds with adjacent bases (Figure 2.5), and that the location and frequency of each hydrogen bond depend on the sequence context. In all sequence contexts, the NH_3 of cisplatin and NH_2 oxaliplatin can act as donors to form a hydrogen bond with the O6 atom of G7 acting as acceptor; this atom is in close proximity to the drugs, so it is not surprising that this hydrogen bond is so common. In addition, both drugs frequently form another hydrogen bond to an adjacent base depending on the sequence context and the drug. Cp-AGGC and Ox-AGGC form hydrogen bonds with the N7 atom of A5; Cp-CGGA and Cp-TGGA form hydrogen bonds to the N7 atom of A8, while Ox-CGGA and Ox-TGGA form hydrogen bonds to the O4 atom of T17; and both Cp-TGGT and

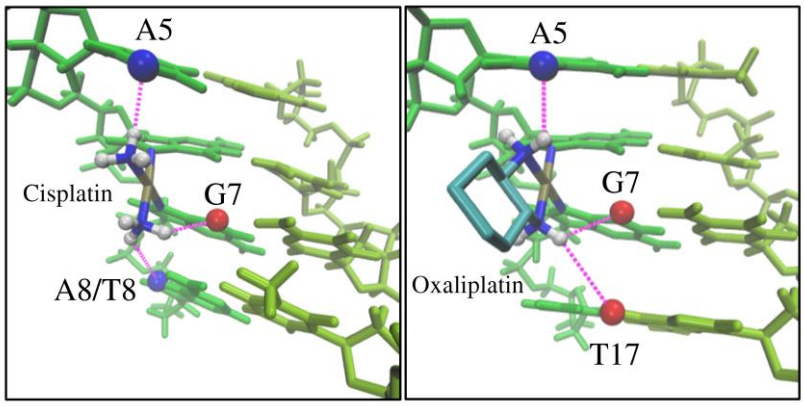


Figure 2.5 Schematic of the location of various hydrogen bonds between the drug and the neighboring bases. The two strands of DNA are shown in two shades of green, and atoms that accept hydrogen bonds from the drugs are shown as red (oxygen) and blue (nitrogen) spheres. All hydrogen atoms are hidden except for those of the ammine groups of cisplatin and amine group of oxaliplatin. Although only a single context is shown, the available hydrogen bonds depend on the sequence. Only the central 4 base pairs are shown and all hydrogen atoms on the DNA are hidden for clarity.

Ox-TGGT form hydrogen bonds with the O4 atom of T8. Figure 2.5 schematically shows the location of these hydrogen bonds.

These hydrogen bonding partners for Cp-DNA and Ox-DNA are in agreement with previous studies, although we note that the previous work studied only the AGGC and TGGA sequence contexts [7, 15]. The difference in hydrogen bonding partners between cisplatin and oxaliplatin in the CGGA and TGGA sequence context is linked to the over-twisting of Ox-DNA at the G7-X8 base pair step, with over-twisting bringing oxaliplatin within reach of the O4 atom of T17. The difference in twist that enables these unique hydrogen bonds is difficult to discern visually in simulation snapshots, but is apparent in the frequency distribution of G7-X8 twist (Figure 2.4). We cannot conclusively say if the higher G7-X8 twist is a result of the hydrogen bond between the amines of oxaliplatin and the O4 atom of T17 or if the hydrogen bond is the result of the higher twist angle at G7-X8 base pair.

Table 2.1 and Table 2.2 show the frequency of these individual hydrogen bonds (see Table 2.1A and 2.2A) in the unbound drug-DNA systems, as well as the frequency of the unique hydrogen bonding states classified by the combination of hydrogen bonds that are active (denoted by +) or inactive (denoted by –) (see Table 2.1B for Cp-DNA and Table 2.2B for Ox-DNA). We find that the G7 hydrogen bond is quite common for Cp-DNA and Ox-DNA, forming approximately 40-44% and 77-85% of the time, respectively, while the hydrogen bonds to flanking bases form somewhat less frequently, which is in quantitative agreement with previous studies [7, 15]. The increased frequency of formation of the G7 hydrogen bond by Ox-DNA may be due to the structural constraint imposed by the cyclohexane ring on the donor amine hydrogen atoms of oxaliplatin. Unlike cisplatin, the hydrogen atoms of the secondary amine groups in oxaliplatin must maintain their orientation with respect to adjacent atoms. Specifically, in oxaliplatin the equatorial amine hydrogen of the amine *cis* to G7 is constrained to point almost directly at the O6 atom of G7, as shown in Figure 2.16 (supplementary section 2.6). Cisplatin, lacking this constraint, forms a hydrogen bond with G7 less frequently. Some of the hydrogen bonding states are rare, particularly for the CGGA and TGGA contexts. For example, the G7+/A8+ state occurs 3.8% of the time for Cp-CGGA, and 2.9% of time for Cp-TGGA (Table 2.1B). Additionally, for Ox-DNA in all four

contexts, the hydrogen bonding state in which the G7 hydrogen bond is absent and only the flanking-base hydrogen bond is present (i.e. **G7–/X+**) is visited less frequently (Table 2.2B). For example, in the AGGC context the **G7–/A5+** state occurs only 7.1% of the time for Ox-DNA in contrast to Cp-DNA where it occurs 29.1% of the time. The **G7–/X+** states are visited less frequently by Ox-DNA adducts as compared to Cp-DNA because the G7 hydrogen bond is much more common for Ox-DNA (approximately 80% frequency) than Cp-DNA (approximately 40% frequency), while the flanking-base hydrogen bond frequency is similar for both drugs, and consequently states *without* the G7 hydrogen bond are less common for Ox-DNA than Cp-DNA.

Table 2.1 Frequency of hydrogen bonds and hydrogen bonding states for Cp-DNA in the AGGC, CGGA, TGGA, and TGGT contexts.

A

	G7	A5	A8	T8
AGGC	44.0	52.4	–	–
CGGA	41.2	–	11.0	–
TGGA	41.3	–	9.0	–
TGGT	40.8	–	–	51.0

B

G7	A5/A8/T8	AGGC	CGGA	TGGA	TGGT
+	+	23.1	3.8	2.9	17.0
+	–	21.0	37.4	38.4	23.7
–	+	29.1	7.3	6.1	34.0
–	–	26.8	51.5	52.6	25.2

Table 2.2 Frequency of hydrogen bonds and hydrogen bonding states for Ox-DNA in the AGGC, CGGA, TGGA, and TGGT contexts.

A

	G7	A5	T17	T8
AGGC	83.0	35.2	–	–
CGGA	77.3	–	19.9	–
TGGA	81.2	–	23.4	–
TGGT	85.5	–	–	57.4

B

G7	A5/T17/T8	AGGC	CGGA	TGGA	TGGT
+	+	34.3	14.5	17.9	47.5
+	–	48.7	62.8	63.2	38.0
–	+	7.1	5.4	5.5	9.8
–	–	10.0	17.3	13.4	4.7

In addition to the frequency, we have also examined the residence time of the drug-DNA molecule in each state. The drug-DNA molecule spends long, continuous periods of time in the most frequent states, generally only residing in the rarer states for a few picoseconds, but it can also sample the less common states for nearly continuous 1-2 ns periods (see Figure 2.17 in supplementary section 2.6 for Ox-AGGC and Ox-TGGA).

Next, we characterize the structure of the drug-DNA in these hydrogen bonding states to see if there are unique favorable or unfavorable structural features in these states that are disguised in the ensemble averages. Figure 2.6 shows the values of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 base pair, and G7-X8 twist for *Cp-DNA* in all four contexts.

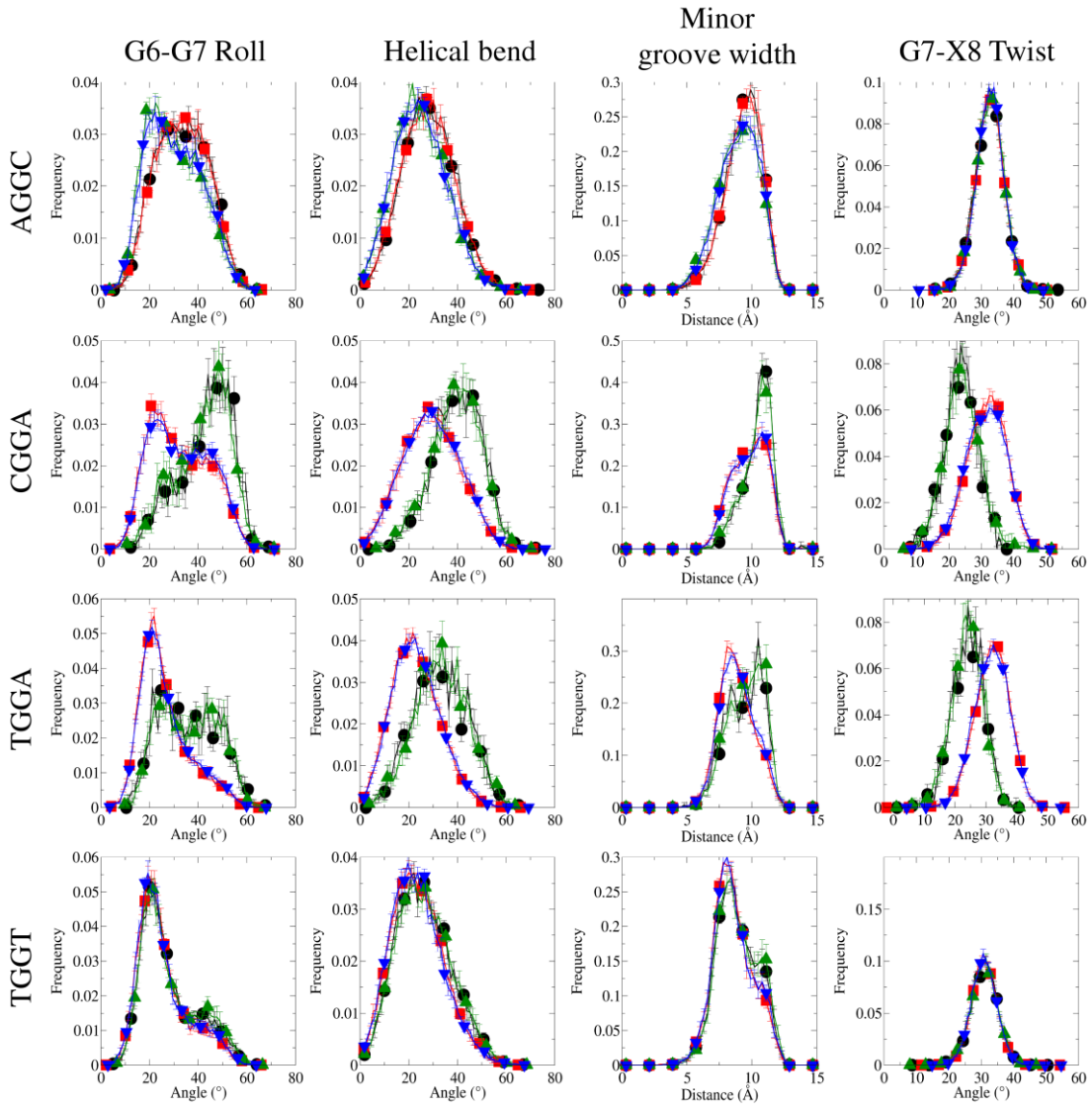


Figure 2.6 Frequency distributions of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 base pair, and the G7-X8 twist for **Cp-DNA** in the AGGC, CGGA, TGGA, and TGGT sequence contexts in the possible hydrogen bonding states. Legend: black circles denote G7+ X+, red squares denote G7+ X–, green upward triangles denote G7– X+, and blue downward triangles denote G7– X–, where "X" refers to the hydrogen bond that is forming with an adjacent base.

For Cp-AGGC, the states in which a hydrogen bond is formed with G7 (black-circle and red-square) are associated with a *slightly higher* G6-G7 roll, helical axis bend, and minor groove width than the other states (green-upward-triangle and blue-downward-triangle). The G7-X8 twist angle distribution is identical for all four hydrogen bonding states. For Cp-CGGA, the states in which there is a hydrogen bond with the flanking base A8 (black-circle and green-upward-triangle) show *much higher* G6-G7 roll, helical axis bend, and minor groove width and *much lower* G7-A8 twist than the states without this

hydrogen bond (red-square and blue-downward-triangle). Cp-TGGA behaves similarly to the CGGA context, which is not surprising given that Cp-DNA can form the same hydrogen bond with A8 in both contexts. For Cp-TGGT, the structural differences between the hydrogen bonding states are insignificant, although the slight shoulder in the G6-G7 roll at around 45° is associated with the formation of a hydrogen bond with T8. In general, the CGGA and TGGA contexts display more significant differences between hydrogen bonding states than do AGGC or TGGT due to the rare hydrogen bond with the flanking A8 base.

Figure 2.7 shows the values of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 base pair, and G7-X8 twist for *Ox-DNA* in all four contexts in various hydrogen bonding states. For Ox-AGGC, hydrogen bonding states have little effect on the structure of Ox-DNA. For Ox-CGGA, the states in which a hydrogen bond is formed with T17 (black circle and green upward triangle) display higher G6-G7 roll, helical axis bend, and minor groove width than the state without T17 hydrogen bond (black-circle line). This data conclusively links the formation of the T17 hydrogen bond is associated with an *increased* G7-A8 twist. However, it is not clear if the hydrogen bond with T17 substantially stabilizes the high-energy, over-twisted state, or if the over-twisted state merely enables the formation of this hydrogen bond. The hydrogen bonding behavior of Ox-TGGA is similar to Ox-CGGA, including higher G6-G7 roll, helical axis bend, minor groove width, and G7-A8 twist when the bond with T17 forms. For Ox-TGGT, as with Ox-AGGC, the structural differences associated with the hydrogen bonding states are minimal, although as with Cp-TGGT a slight shoulder in the roll around 45° may be associated with the T8 hydrogen bond.

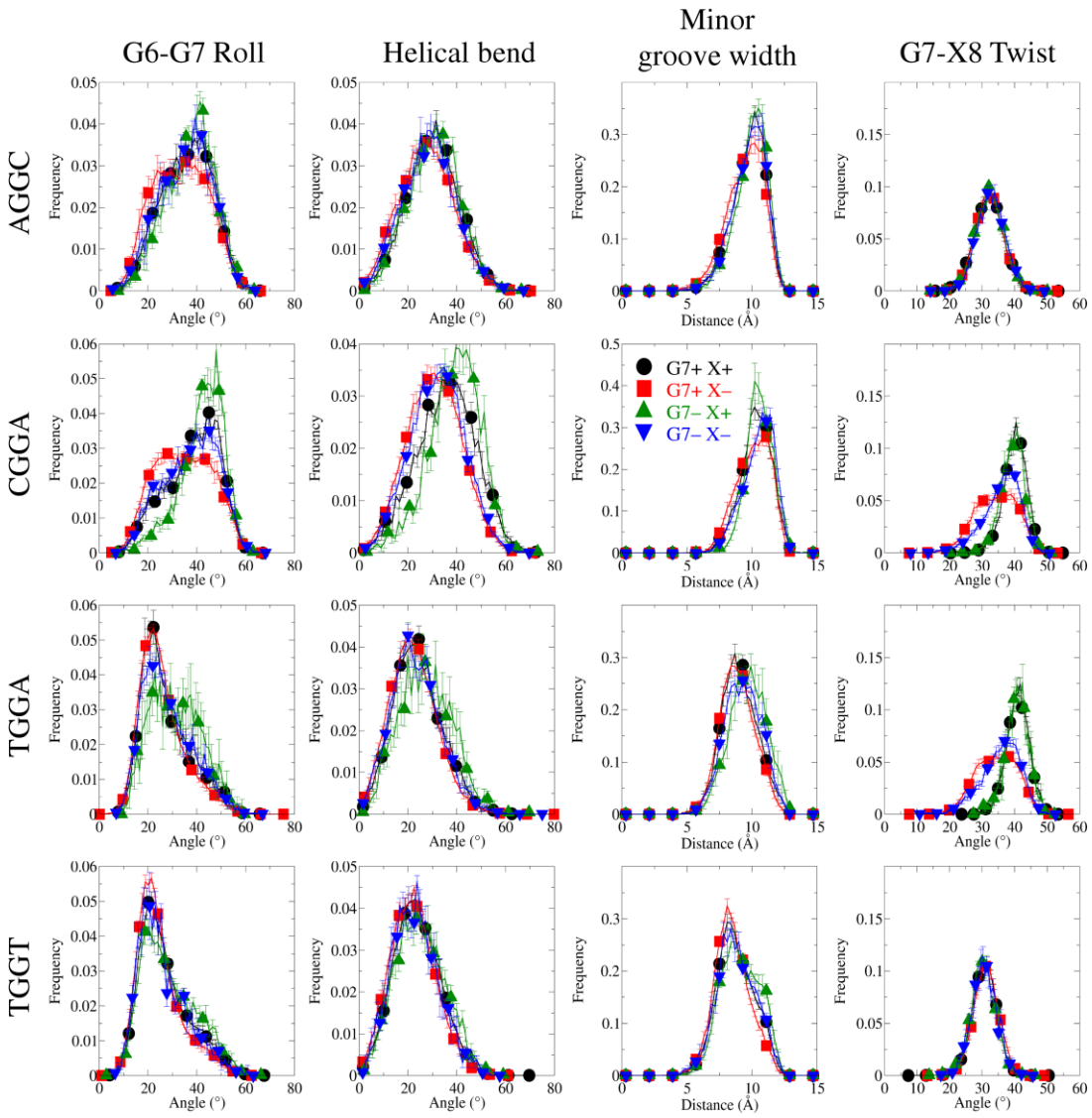


Figure 2.7 Frequency distributions of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 base pair, and the G7-X8 twist for **Ox-DNA** in the AGGC, CGGA, TGGA, and TGGT sequence contexts in the possible hydrogen bonding states. Legend: black circles denote G7+ X+, red squares denote G7+ X–, green upward triangles denote G7– X+, and blue downward triangles denote G7– X–, where "X" refers to the hydrogen bond that is forming with an adjacent base.

In Figure 2.8, we compare the hydrogen bonding states in Cp-DNA (red circles) and Ox-DNA (blue squares) that exhibit the *highest* G6-G7 roll, *highest* helical bend, *largest* minor groove width at G7-C18, and *lowest* G7-X8 twist; in other words, the structures *most favorable for binding* by HMGB1a. The states that exhibit these features are listed in Figure 2.8 caption. In the AGGC sequence context, Ox-DNA explores states with slightly higher roll, helical bend, and minor groove width than Cp-DNA, but the two drugs have the same distribution of twist. In the CGGA context, Cp-DNA exhibits slightly higher roll and

minor groove width than Ox-DNA, equal helical bend distributions as Ox-DNA, and significantly lower twist than Ox-DNA.

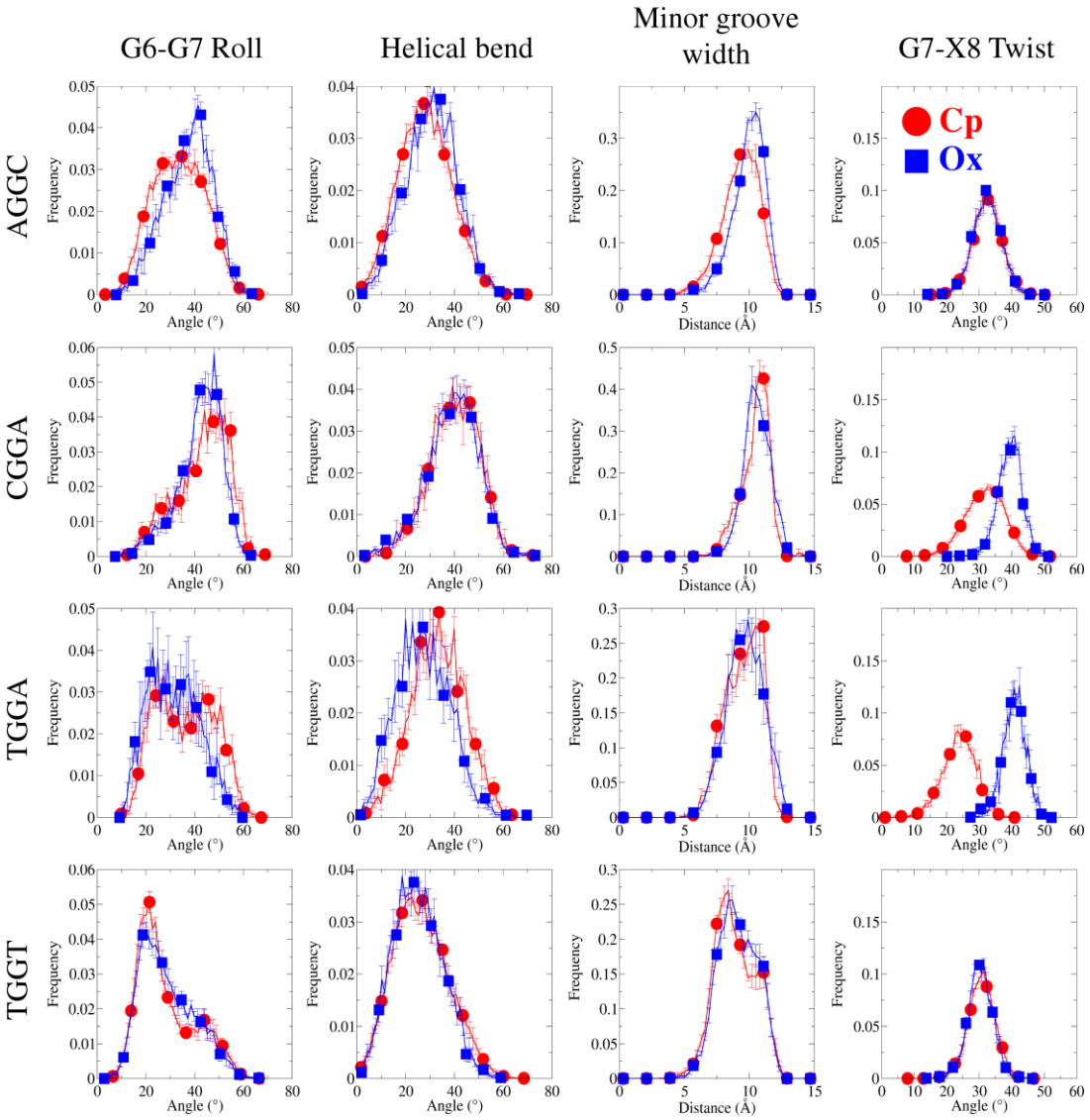


Figure 2.8 Hydrogen bonding states producing the *highest* values of G6-G7 roll, helical axis bend and minor groove width (most favorable for HMGB1a binding), and *lowest* G7-X8 twist (most unfavorable for HMGB1a binding) for both cisplatin and oxaliplatin adducts. For Cp-DNA, the hydrogen bonding state shown varies by sequence: for AGGC, **G7+ X–** for all parameters; for CGGA, **G7+ X+** for all parameters except for twist which is **G7- X+**; for TGGA, **G7– X+** for all parameters except for twist which is **G7- X+**; and for TGGT, **G7– X+** for all parameters. For Ox-DNA in all sequences the **G7– X+** state is shown for all parameters, except for Ox-CGGA and TGGA where the **G7+X-** is shown for twist. In cases where the distribution is similar between two states, the smoother distribution is shown for clarity.

The TGGA context exhibits similar trends as the CGGA context, with Cp-DNA having a slightly higher roll, helical bend, and minor groove width and significantly lower twist than Ox-DNA. In the TGGT

context, Cp-DNA shows two peaks in the distributions of G6-G7 roll and minor groove, although the value and frequency of occurrence of the higher peak is the same as Ox-DNA, while the helical bend and twist distributions are the same for both drugs. In general, values of the parameters considered most relevant to binding affinity are more favorable for Cp-DNA in the CGGA and TGGA contexts, approximately equal for the drugs in the TGGT context, and more favorable for Ox-DNA in the AGGC context.

Finally, we quantify how binding by HMGB1a alters the frequency of hydrogen bonding states shown in Table 2.3 and Table 2.4, and how this changes the structural differences between hydrogen bonding states. For bound-Cp-DNA (Table 2.3A) in all sequence contexts except CGGA, the bond to the adjacent base (i.e. A5, A8, or T8) is more frequent than in the unbound state (Table 2.1A), while the frequency of the hydrogen bond to G7 remains about the same as the unbound state (Table 2.1A). In the CGGA context, the hydrogen bonding states of bound-Cp-CGGA show some minor differences (Figure 2.18 in supplementary section 2.6). Like the unbound-Cp-CGGA, in the bound-Cp-CGGA the states in which a hydrogen bond forms with the adjacent A8, the G6-G7 roll and helical axis bend are slightly higher than other hydrogen bonding states. However, the differences between the various states are reduced in the bound state as compared to the unbound state (Figure 2.18 in supplementary section 2.6). In the bound state Cp-DNA in AGGC, TGGT and TGGA sequence contexts show no significant differences between the structures of the various hydrogen bonding states (Figure 2.19 in supplementary section 2.6). For bound Ox-DNA in the AGGC and TGGT sequence contexts, the hydrogen bond to the adjacent base becomes more common going from unbound to bound state, whereas for Ox-CGGA and Ox-TGGA, the hydrogen bond to T17 is essentially eliminated because the over-twisted state associated with this hydrogen bond is eliminated when the protein binds (Figure 2.2h). For all sequences of bound-Ox-DNA, the frequency of the hydrogen bond to G7 is unchanged, and there are no significant differences between the structures of the hydrogen binding states (Figure 2.19 in supplementary section 2.6). The overall lack of structural differences induced by the various hydrogen bonding states when

bound by HMGB1a further reinforces the fact that differences in the *unbound*-drug-DNA molecules is critical to the differential recognition of the drug-DNA adducts by HMGB1a.

Table 2.3 Frequency of hydrogen bonds and hydrogen bonding states for Cp-DNA *bound by HMGB1a* in the AGGC, CGGA, TGGA, and TGGT contexts.

A

	G7	A5	A8	T8
AGGC	49.6	62.1	–	–
CGGA	42.9	–	9.2	–
TGGA	40.0	–	26.6	–
TGGT	38.4	–	–	71.0

B

G7	A5/A8/T8	AGGC	CGGA	TGGA	TGGT
+	+	30.8	3.1	9.0	23.2
+	–	18.8	39.7	31.0	15.1
–	+	31.3	6.0	17.5	47.8
–	–	19.1	51.1	42.4	13.8

Table 2.4 Frequency of hydrogen bonds and hydrogen bonding states for Ox-DNA *bound by HMGB1a* in the AGGC, CGGA, TGGA, and TGGT contexts.

A

	G7	A5	T17	T8
AGGC	90.5	43.5	–	–
CGGA	84.7	–	4.7	–
TGGA	88.0	–	0.0	–
TGGT	81.0	–	–	75.6

B

G7	A5/T17/T8	AGGC	CGGA	TGGA	TGGT
+	+	39.4	4.0	0	59.2
+	–	51.1	80.7	100	21.9
–	+	4.1	0.7	0	16.4
–	–	5.4	14.6	0	2.6

2.4 Discussion

We have shown that *HMGB1a-bound* drug-DNA adducts do not show significant structural differences with varying sequence context or drug chemistry, either in their ensemble average distributions or in the different hydrogen bonded states (also termed *conformational dynamics* in this paper), and therefore we expect that the differential recognition of Cp-DNA and Ox-DNA must arise from differences in the unbound state of the drug-DNA molecule. Having quantified the differences in structure and conformational dynamics between cisplatin-DNA adducts and oxaliplatin-DNA adducts in the AGGC,

CGGA, TGGA, and TGGT contexts in the absence of the protein, we examine whether these differences could explain trends in binding affinity that have been observed in previous experimental studies.

Experimental trends for Cp-DNA show that HMGB1a binds to the CGGA and TGGA contexts most effectively and approximately equally, followed by the TGGT context, and finally the AGGC context (see Table 2.5 in supplementary section 2.6) [36-38]. In our simulations, we observe that while CGGA and TGGA explore different ensemble average distributions of G6-G7 roll, helical bend, and minor groove width, TGGA explores rare states whose structural features closely resemble that of CGGA, which could explain why HMGB1a exhibits *similar* affinity for cisplatin in the CGGA and TGGA contexts. Cp-TGGT and Cp-TGGA have similar ensemble average distributions of G6-G7 roll, helical bend, and minor groove width, but Cp-TGGT *lacks* the structural features of the rare states of Cp-TGGA, which could explain why HMGB1a has lower binding affinity for Cp-TGGT. However, while experimentally Cp-CGGA is recognized by HMGB1a much better than Cp-AGGC, our simulations show that Cp-AGGC and Cp-CGGA have similar ensemble average distributions and explore similar dynamic conformations. Therefore, these structural results explain only some of the trends in the sequence-dependent binding affinity of HMGB1a for Cp-DNA adducts.

Although experimental data for HMGB1a binding affinity for oxaliplatin-DNA adducts is limited, the AGGC, CGGA, TGGA and TGGT contexts are available for comparison [11, 38]. Experiments show that HMGB1a exhibits approximately equal affinity for Ox-AGGC, Ox-CGGA and Ox-TGGT, whereas Ox-TGGA is essentially unrecognized by HMGB1a. Our simulations show that Ox-AGGC and Ox-CGGA explore similar values of G6-G7 roll, helical bend, and minor groove width, which could explain the equal affinity of HMGB1a for these two contexts. However, there are no apparent structural or dynamic patterns to explain why Ox-TGGT is recognized as well as Ox-AGGC and Ox-CGGA. Although Ox-TGGA and Ox-TGGT have similar structures, with lower roll, helical bend, and minor groove width than Ox-AGGC and Ox-CGGA, Ox-TGGA shows over-twisting (up to an average of 40°, compared to only 30°) at the G7-A8 base pair step as compared to Ox-TGGT. This over-twisting is also present in Ox-CGGA. Since high G6-G7 roll and low helical twist both facilitate binding of HMGB1a, a pattern

emerges: stated succinctly, Ox-AGGC has high roll and low twist, Ox-CGGA has high roll but high twist, Ox-TGGT has low roll but low twist, but Ox-TGGA has low roll *and* high twist. Of the sequences, only Ox-TGGA has both structural characteristics that could negatively impact binding. However, it is unclear how these structural differences could *completely* preclude binding of HMGB1a to Ox-TGGA.

Experimental results have shown that cisplatin is recognized more readily than oxaliplatin by HMGB1a in all sequence contexts that have been studied [11, 38]. Comparing our simulation results for cisplatin-DNA and oxaliplatin-DNA adducts, the TGGA and TGGT sequence contexts have essentially the same ensemble average distributions of G6-G7 roll, helical axis bend, and minor groove width for both drugs, but Ox-TGGA displays much higher twist than Cp-TGGA, Cp-TGGT, or Ox-TGGT, which suggests a possible reason for the lower recognition of Ox-TGGA (versus Cp-TGGA) but does not explain the lower recognition of Ox-TGGT (versus Cp-TGGT). In the AGGC and CGGA contexts, Ox-DNA exhibits somewhat higher ensemble average values of G6-G7 roll, helical bend, and minor groove width than Cp-DNA (Figure 2.2), but Cp-DNA can dynamically reach higher values of these parameters than Ox-DNA, which provides a possible explanation for the greater affinity of HMGB1a for cisplatin adducts in these two contexts (Figure 2.8).

We conclude that these differences in structure and conformational dynamics between Cp-DNA and Ox-DNA rationalize *many, but not all* of the observed experimental trends, and therefore it seems likely that there are additional factors that must contribute to protein binding affinity besides the structure of the unbound-drug-DNA complex. One additional factor that could contribute to differential recognition and binding affinity is the thermodynamic pathway between the unbound-drug-DNA and the HMGB1a-bound drug-DNA states: some paths might contain larger energetic barriers in some drugs/sequence contexts than in others, and these differing energetic barriers could arise from differences in DNA flexibility. In undamaged B-DNA, DNA flexibility has been shown to be sequence-dependent, with A-T base pairs generally being more flexible than G-C base pairs, although the *extent and direction* of flexibility depend on the order of the base pairs. For example, A-tract DNA (consecutive sequences of 4-6 A-T base pairs) can induce a local bend in DNA, and, examining one strand in a dsDNA helix from 5’ to

3', the direction of the induced bend depends on whether the adenine or the thymine of the A-T pairs is in the strand [53]. It has also been shown that the flexibility of pyrimidine-purine (Y-R) steps is greater than purine-pyrimidine (R-Y), and that Y-R steps can act as 'hinges' that allow local bending, with the location of these hinges altering the location of bending induced during protein-DNA binding [54]. We note that the sequences we study in our drug-DNA systems each have 1 Y-R step and 1 R-Y step, although in different positions. Experimentally, it has been suggested, but not proven, that sequence-specific flexibility may play a large role in cisplatin-DNA binding affinity, with the higher flexibility of A-T versus G-C base pairs possibly related to higher binding affinity [37]. Additionally, differences in the structure of the drug could affect flexibility. How the presence of oxaliplatin affects helical axis flexibility has not been studied, although it has been suggested that the bulky cyclohexane ring of oxaliplatin hinders the additional bending of the helical axis required for HMGB1a to bind to the DNA [38]. Given the lower binding affinity of HMGB1a for oxaliplatin than for cisplatin, it seems likely that studies of sequence-specific and drug-specific DNA flexibility, combined with structural differences and conformational dynamics presented here, together will help rationalize the protein recognition patterns of Cp-DNA and Ox-DNA adducts. Work elucidating sequence-specific flexibility of drug-DNA molecules will be the focus of a future publication.

In conclusion, the atomistic molecular dynamics simulations studies of cisplatin- and oxaliplatin-DNA adducts presented here represent the most comprehensive computational study to date of the effects of DNA sequence context on the structure and dynamics of these drug-DNA adducts and in turn, binding affinity of HMGB1a repair protein. This work provides a molecular-level insight to the structural differences between these two drug-DNA adducts in various sequence contexts and the importance of the unbound state (drug-DNA molecule in the absence of HMGB1a protein) over the bound state (presence of HMGB1a protein) in understanding differential HMGB1a binding affinity. Additionally, this work suggests that the protein must actively deform the DNA during binding by widening the minor groove and increasing the helical bend, and the thermodynamic pathway between the unbound and bound states

could be an additional factor in the binding affinity of HMGB1a for drug-DNA adducts in various sequence contexts.

2.5 References

[1] A. M. J. Fichtingerschepman, J. L. Vanderveer, J. H. J. Denhartog, P. H. M. Lohman, and J. Reedijk, Biochemistry 24 (1985) 707.
[2] P. M. Takahara, A. C. Rosenzweig, C. A. Frederick, and S. J. Lippard, Nature 377 (1995) 649.
[3] D. M. J. Lilley, Journal of Biological Inorganic Chemistry 1 (1996) 189.
[4] A. Gelasco and S. J. Lippard, Biochemistry 37 (1998) 9230.
[5] E. R. Jamieson and S. J. Lippard, Chemical Reviews 99 (1999) 2467.
[6] M. A. Elizondo-Riojas and J. Kozelka, Journal of Molecular Biology 314 (2001) 1227.
[7] S. Sharma, P. Gong, B. Temple, D. Bhattacharyya, N. V. Dokholyan, and S. G. Chaney, Journal of Molecular Biology 373 (2007) 1123.
[8] E. D. Scheeff, J. M. Briggs, and S. B. Howell, Molecular Pharmacology 56 (1999) 633.
[9] J. M. Woynarowski, S. Faivre, M. C. S. Herzig, B. Arnett, W. G. Chapman, A. V. Trevino, E. Raymond, S. G. Chaney, A. Vaisman, M. Varchenko, and P. E. Juniewicz, Molecular Pharmacology 58 (2000) 920.
[10] S. G. Chaney, S. L. Campbell, E. Bassett, and Y. B. Wu, Critical Reviews in Oncology Hematology 53 (2005) 3.
[11] J. Malina, O. Novakova, M. Vojtiskova, G. Natile, and V. Brabec, Biophysical Journal 93 (2007) 3950.
[12] P. Tyagi, P. Gahlot, and R. Kakkar, Polyhedron 27 (2008) 3567.
[13] S. Teletchea, T. Skauge, E. Sletten, and J. Kozelka, Chemistry- A European Journal 15 (2009) 12320.
[14] D. Wang and S. J. Lippard, Nature Reviews Drug Discovery 4 (2005) 307.
[15] S. Ramachandran, B. R. Temple, S. G. Chaney, and N. V. Dokholyan, Nucleic Acids Research 37 (2009) 2434.
[16] U. M. Ohndorf, M. A. Rould, Q. He, C. O. Pabo, and S. J. Lippard, Nature 399 (1999) 708.
[17] E. A. Pasheva, I. Ugrinova, N. C. Spassovska, and I. G. Pashev, International Journal of Biochemistry and Cell Biology 34 (2002) 87.
[18] H. Arioka, K. Nishio, T. Ishida, H. Fukumoto, K. Fukuoka, T. Nomoto, H. Kurokawa, H. Yokote, S. Abe, and N. Saijo, Japanese Journal of Cancer Research 90 (1999) 108.
[19] J. C. Huang, D. B. Zamble, J. T. Reardon, S. J. Lippard, and A. Sancar, Proceedings of the National Academy of Sciences of the United States of America 91 (1994) 10394.
[20] Y. Mikata, Q. He, and S. J. Lippard, Biochemistry 40 (2001) 7533.
[21] B. Marples, H. Adomat, P. C. Billings, N. P. Farrell, C. J. Koch, and K. A. Skov, Anti-Cancer Drug Design 9 (1994) 389.
[22] M. Magdalena, S. Pawel, N. M. Ewa, M. Marcin, M. Piotr, R. Wojciech, K. Helena, Z. Maciej, and S. Marek, Ginekologia Polska 79 (2008) 826.
[23] J. A. Mello, S. Acharya, R. Fishel, and J. M. Essigmann, Chemistry & Biology (Cambridge, MA, United States) 3 (1996) 579.
[24] F. R. Salsbury, J. E. Clodfelter, M. B. Gentry, T. Hollis, and K. D. Scarpinato, Nucleic Acids Research 34 (2006) 2173.
[25] K. Wozniak and J. Blasiak, Acta Biochimica Polonica 49 (2002) 583.
[26] A. Vaisman, S. E. Lim, S. M. Patrick, W. C. Copeland, D. C. Hinkle, J. J. Turchi, and S. G. Chaney, Biochemistry 38 (1999) 11026.
[27] A. Vaisman, C. Masutani, F. Hanaoka, and S. G. Chaney, Biochemistry 39 (2000) 4575.

[28] E. Bassett, N. M. King, M. F. Bryant, S. Hector, L. Pendyala, S. G. Chaney, and M. Cordeiro-Stone, Cancer Research 64 (2004) 6469.
[29] S. Chijiwa, C. Masutani, F. Hanaoka, S. Iwai, and I. Kuraoka, Carcinogenesis 31 (2010) 388.
[30] Y. W. Jung and S. J. Lippard, Chemical Reviews 107 (2007) 1387.
[31] S. S. Lange and K. M. Vasquez, Molecular Carcinogenesis 48 (2009) 571.
[32] M. Stros, Biochimica et Biophysica Acta, Gene Regulatory Mechanisms 1799 (2010) 101.
[33] S. Park and S. J. Lippard, Biochemistry 50 (2011) 2567.
[34] J. Kasparkova, O. Delalande, M. Stros, M. A. Elizondo-Riojas, M. Vojtiskova, J. Kozelka, and V. Brabec, Biochemistry 42 (2003) 1234.
[35] Q. He, U. M. Ohndorf, and S. J. Lippard, Biochemistry 39 (2000) 14426.
[36] S. U. Dunham and S. J. Lippard, Biochemistry 36 (1997) 11428.
[37] S. M. Cohen, Y. Mikata, Q. He, and S. J. Lippard, Biochemistry 39 (2000) 11771.
[38] M. Wei, S. M. Cohen, A. P. Silverman, and S. J. Lippard, Journal of Biological Chemistry 276 (2001) 38774.
[39] K. Stehlikova, H. Kostrhunova, J. Kasparkova, and V. Brabec, Nucleic Acids Research 30 (2002) 2894.
[40] Y. B. Wu, D. Bhattacharyya, C. L. King, I. Baskerville-Abraham, S. H. Huh, G. Boysen, J. A. Swenberg, B. Temple, S. L. Campbell, and S. G. Chaney, Biochemistry 46 (2007) 6477.
[41] Y. B. Wu, P. Pradhan, J. Havener, G. Boysen, J. A. Swenberg, S. L. Campbell, and S. G. Chaney, Journal of Molecular Biology 341 (2004) 1251.
[42] R. C. Todd and S. J. Lippard, Journal of Inorganic Biochemistry 104 (2010) 902.
[43] D. Yang, S. S. G. E. van Boom, J. Reedijk, J. H. van Boom, and A. H. J. Wang, Biochemistry 34 (1995) 12912.
[44] J. Kozelka, R. Savinelli, G. Berthier, J. P. Flament, and R. Lavery, Journal of computational chemistry 14 (1993) 45.
[45] S. Yao and J. P. Plastaras, Inorganic Chemistry 33 (1994) 6061.
[46] K. Spiegel, U. Rothlisberger, and P. Carloni, Journal of Physical Chemistry B 108 (2004) 2699.
[47] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of Computational Chemistry 26 (2005) 1781.
[48] A. Bondi, Journal of Physical Chemistry 68 (1964) 441.
[49] T. W. Hambley, Inorganic Chemistry 37 (1998) 3767.
[50] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics and Modelling 14 (1996) 33.
[51] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, Nucleic Acids Research 37 (2009) 5917.
[52] Z. S. Juo, T. K. Chiu, P. M. Leiberman, I. Baikalov, a. J. Berk, and R. E. Dickerson, Journal of molecular biology 261 (1996) 239.
[53] R. Stefl, H. Wu, S. Ravindranathan, V. Sklenár, and J. Feigon, Proceedings of the National Academy of Sciences of the United States of America 101 (2004) 1177.
[54] J. A. Parkinson, Y. Chen, P. del Socorro Murdoch, Z. Guo, S. J. Berners-Price, T. Brown, and P. J. Sadler, Chemistry- A European Journal 6 (2000) 3636.

2.6 Supplementary information

2.6.1 Equilibration procedure

First, in constant volume and temperature conditions (NVT), the water and ions were minimized with the conjugate gradient algorithm for 10000 steps, with the DNA subjected to a 500 kcal/mol-A^2 harmonic restraining potential. Second, in NVT conditions, with the DNA restraining potential reduced to 100 kcal/mol-A^2, the system was heated from 0 to 300 K over 20 ps. Third, in constant pressure and temperature conditions (NPT) and with the restraint reduced to 50 kcal/mol-A^2 the system was relaxed for another 20 ps. Fourth, in NVT, the restraining potential was reduced in three steps (50, 10, and 5 kcal/mol-A^2) and minimized for 2000 steps each. Fifth, in NVT with the same 5 kcal/mol-A^2 restraint the system was heated from 10 to 300 K over 20 ps. Sixth, in NPT, the restraint was reduced from 5 to 1 to 0.1 to 0 kcal/mol-A^2 over 20 ps. Seventh, in NVT with no restraints, the system was heated from 10 to 300 K over 20 ps. Finally, the system was heated from 100 to 300 K over 20 ps in NPT at the beginning of each production run.

2.6.2 Calibration of improper torsions designed to keep the platinum atom in the plane of the guanine bases

In the original force fields for cisplatin and oxaliplatin, two improper torsion terms were included to keep the platinum atom in the plane of the two platinated guanine bases and prevent the bases from puckering (see Figure 2.9a and 2.9b) [8, 45]. The specification of these improper torsions was CK-NB1/NB2-CB-PT. This is inconsistent with the conventions of the current Amber force field, which requires the central atom of an improper torsion (in this case, NB1 and NB2) to be the 3rd atom in the list. If these parameters are given in the force field modification file in the original form, they will not be applied by *leap*, and therefore the order of the atoms must be changed to CK-CB-NB1/NB2-PT.

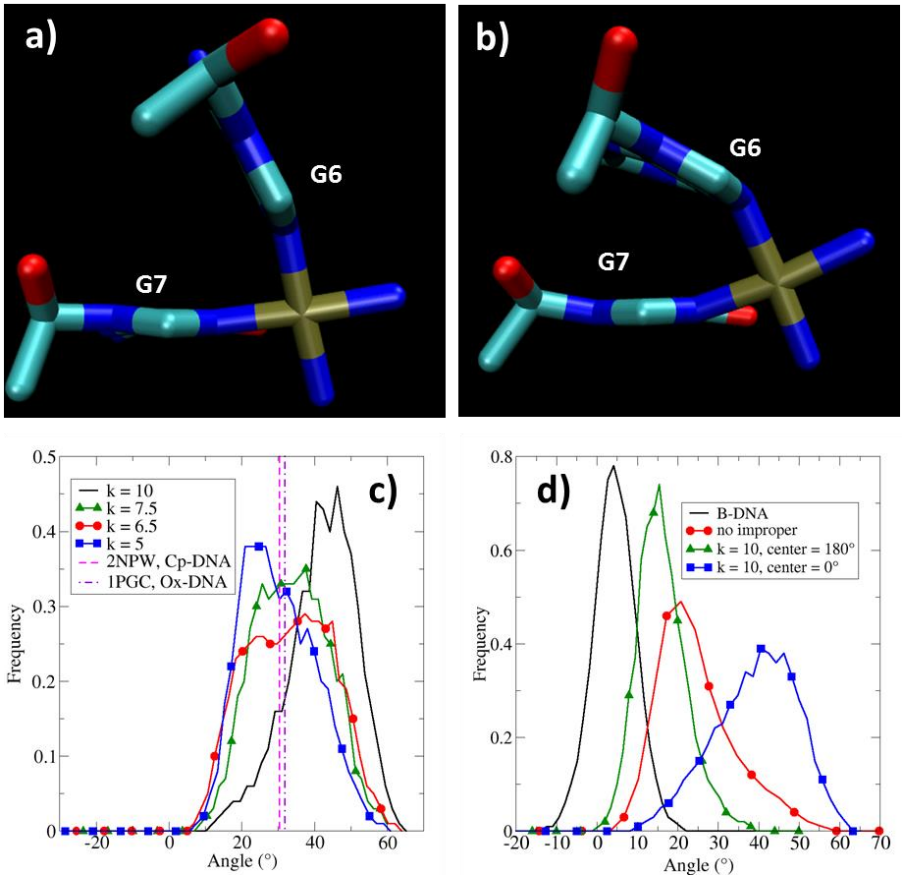


Figure 2.9 Panels (a) and (b) are representative snapshots of the G6-G7 roll before and after calibration of the cisplatin force field. Panel (c) shows the distributions of the G6-G7 roll for several values of the force constant of the CK-CB-NB1/NB2-PT improper torsion. Panel (d) shows the effect of changing the center of the improper torsion.

Also note that the original specification was for a periodicity of 1 and center of 180°, in contrast to most other improper torsions, which have a periodicity of 2 and a center of 180°. When the order of the atoms is changed, it changes the definition of 0° and 180°, such that a center of 180° actually pushes the platinum atom out of the base plane (Figure 2.9d). Changing the center to 0° corrects this problem (Figure 2.9d). With a periodicity of 2, this problem does not occur because 0° and 180° are equivalent. Additionally, we observed that the G6-G7 roll in the absence of this improper torsion was considerably lower than the experimental values, indicating that this parameter is important for reproducing the structure.

However, after reversing the atom order we found that the force constant (10 kcal/mol) of this improper was too high and produced values of G6-G7 roll that were considerably higher than those found in experimental structures (Figure 2.9a and 2.9d). Consequently, we performed some exploratory studies and found that a force constant of 7.5 kcal/mol results in structures that more closely match the experimental structure (Figure 2.9c and Figure 2.9b).

We note that the force fields for cisplatin and oxaliplatin are parameterized to best replicate the X-ray crystal structure of the TGGT context and thus may be less accurate when applied to other sequence contexts [2, 8, 45]. However, we believe that the resulting structures are quantitatively accurate because none of the force field parameters directly depend on the identity of the flanking bases.

2.6.3 List of all DNA structural parameters

In this work we calculated and examined trends in all of the following structural parameters. For brevity and due to their higher relevance to protein binding we report in the main manuscript only the trends in roll, helical bend, minor groove width and twist. For a more complete description of these parameters, see the documentation for Curves+ [51].

- Intra base pair
    - Translations: Shear, stretch, stagger
    - Rotations: Buckle, propeller, opening
- Inter base pair
    - Translations: Shift, slide, rise
    - Rotations: Tilt, roll, twist
- Relative to helical axis
    - Translations: X-displacement, Y-displacement, Z-displacement
    - Rotations: Inclination, tip
    - Total helical axis bend
- Grooves:
    - Major groove width
    - Major groove depth
    - Minor groove width
    - Minor groove depth
- Backbone angles:
    - Alpha, beta, chi, delta, epsilon, gamma, zeta

2.6.4 Agreement of our simulation structures with experimental structures

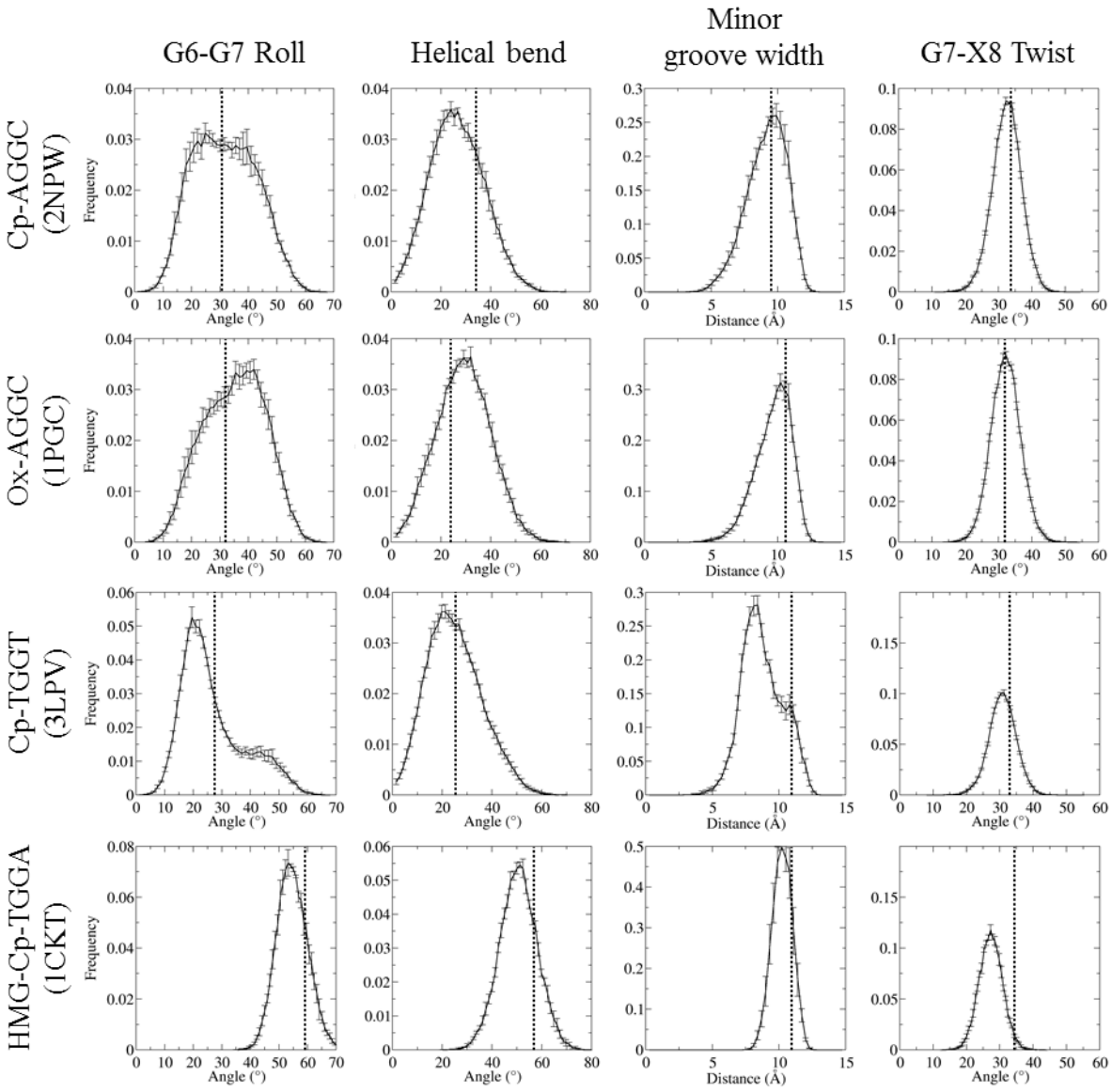


Figure 2.10 Our simulated structures of cisplatin-DNA and oxaliplatin-DNA show agreement with the *four different experimental* structures (2NPW.pdb, 1PGC.pdb, 3LPV.pdb and 1CKT.pdb). Dotted vertical lines correspond to values from experimental structures. The RMSD values of Cp-AGGC (PDB: 2NPW), Ox-AGGC (PDB: 1PGC), Cp-TGGT (PDB: 3LPV) are 3.7 Å, 3.8 Å, and 3.8 Å respectively. The distributions of G6-G7 roll, helical axis bend, minor groove width at the G7-C18 basepair, and G7-X8 twist from our simulations are generally in good quantitative agreement with the experimental structures. The structure of HMGB1a-Cp-TGGA (PDB: 1CKT) also shows agreement, with an RMSD of 2.8 Å. The agreement of the HMG-Cp-TGGA simulation distributions with experiment is not as good as the drug-DNA systems, although most parameters are somewhat close to the experimental values. We also note that the crystal structure available for this system (PDB: 1CKT) is only of 2.5 Å resolution, which could negatively impact the agreement between simulation and experiment.

2.6.5 Calculation of dissociation constants for HMGB1a binding to drug-DNA to quantify relative binding affinities

Table 2.5 Concentrations of DNA ($[DNA]_{total}$) and protein ($[HMGB1a]_{total}$), values of the fraction of DNA bound by HMGB1a (θ), and the corresponding values of the dissociation constant (K_D), which are calculated with the method described by Ramachandran *et al*. [15] Values of θ are estimated from the papers listed in the Source column (numbers correspond to Ref. #)

Sequence	Source	$[DNA]_{total}$ (nM)	$[HMGB1a]_{total}$ (nM)	θ (estimated)	K_D (nM)
Cp-AGGA	[36]	5	5	0.405	**4**
Cp-AGGT	[36]	5	5	0.26	**11**
Cp-AGGG	[37]	5	7.6	0.32	**13**
Cp-AGGC	[36]	5	5	0.16	**22**
Cp-TGGA	[36]	5	5	0.43	**4**
Cp-TGGT	[36]	5	5	0.22	**14**
Cp-TGGG	[37]	5	7.6	0.125	**49**
Cp-TGGC	[36]	5	5	0.02	**240**
Cp-GGGA	[37]	5	7.6	0.36	**10**
Cp-GGGT	[37]	5	7.6	0.25	**19**
Cp-GGGG	[37]	5	7.6	0.15	**39**
Cp-GGGC	[37]	5	7.6	0.08	**83**
Cp-CGGA	[36]	5	5	0.49	**3**
Cp-CGGT	[36]	5	5	0.255	**11**
Cp-CGGG	[37]	5	7.6	0.155	**37**
Cp-CGGC	[36]	5	5	0.06	**74**
Ox-AGGC	[38]	12	30	0.145	**167**
Ox-CGGA	[11]	10	30	0.2	**112**
Ox-TGGA	[38]	17	30	0.01	**2953**
Ox-TGGT	[11]	10	30	0.22	**99**

The dissociation constant (K_D) is a measure of the binding affinity of HMGB1a for drug-DNA, with lower values of K_D indicating higher affinity. This set of studies shows that the base 3' of the cisplatin-DNA adduct significantly impacts binding affinity with the relation A > T > G > C. These studies also found that the base 5' of the adduct had a significant impact when the 3' base was either G or C (i.e. weakly-bound adducts), with the relation A > C ≈ G > T. Oxaliplatin-DNA adducts are less well-recognized than cisplatin-DNA adducts, but it is difficult to define a pattern of sequence-dependence given the small amount of available data.

2.6.6 Structures of drug-DNA independent of DNA sequence (bases outside the immediate sequence context)

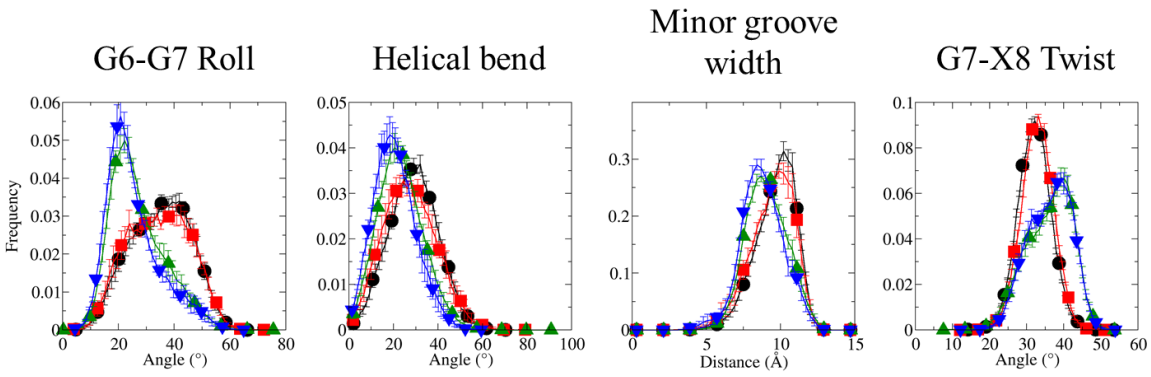


Figure 2.11 The G6-G7 roll, helical bend, minor groove width, and G7-X8 twist for Ox-AGGC and Ox-TGGA in two different DNA sequences: 'simulation sequence,' 5'-CCTC-XGGY-CTCC-3' (the sequence used elsewhere in this simulation study and in several experimental determinations of structure) and 'experimental sequence,' 5'-CCTTCT-XGGY-TCTTC-3' (the sequence used in the majority of the binding affinity experiments). We simulated these particular sequences (Ox-AGGC, Ox-TGGA) to understand what structural features could lead to the total lack of recognition of Ox-TGGA by HMGB1a compared to the relatively high recognition of Ox-AGGC [38]. We hypothesized that the bases flanking the immediate sequence context (i.e. XGGY) could significantly change the structure and explain the differential affinity of HMGB1a. One major difference between the simulation and experimental sequences is the base pairs flanking the immediate sequence context (i.e. C-XGGY-C vs. T-XGGY-T), which is significant because T-A base pairs generally explore a wider range of conformations [52]. However, we found the structural differences between the simulation and experimental structures to be minor. Legend: black circle, Ox-CCTC-AGGC-CTCC; red square, Ox-CCTTCT-AGGC-TCTTC; green upward triangle, Ox-CCTC-TGGA-CTCC; blue downward triangle, Ox-CCTTCT-TGGA-TCTTC.

2.6.7 Narrow distribution of minor groove widths in the presence of HMGB1a

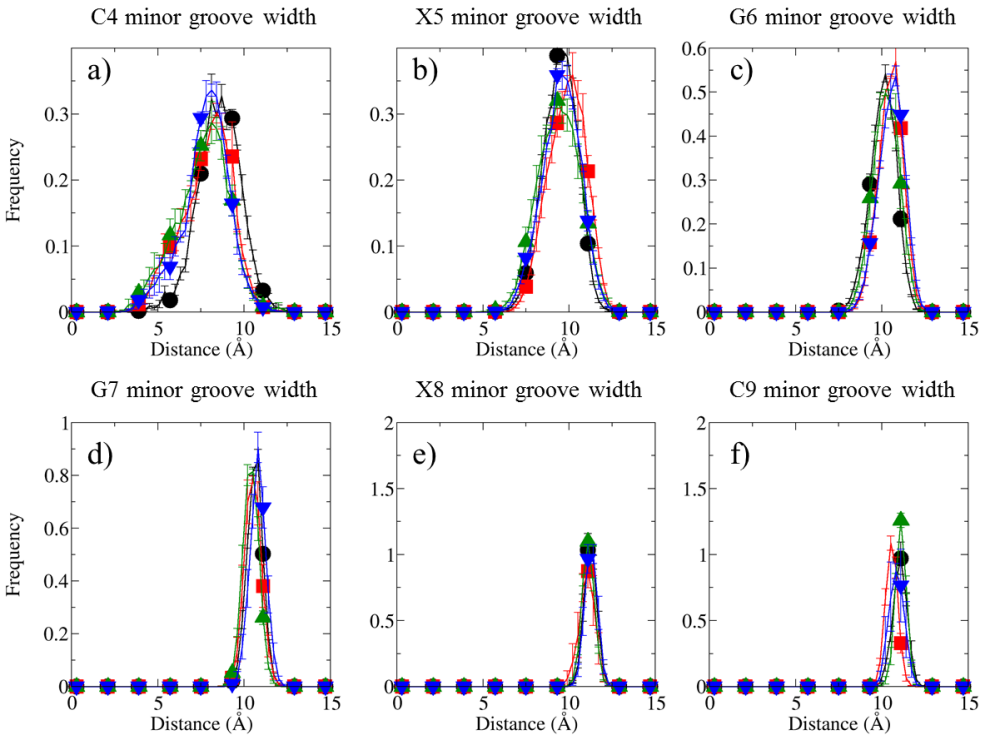


Figure 2.12 The minor groove width for HMGB1a-bound Cp-DNA at each of the central 6 base pairs of the DNA dodecamer, showing that the width has a narrow distribution in the region where the protein is bound (G6, G7, X8, and so on) and that the minor groove is wider near the base pairs with more contact with the protein (i.e. those 3' of the adduct). The results are presented for four sequence contexts AGGC (black circle), CGGA (red square), TGGA (green upward triangle) and TGGT (blue down triangle).

2.6.8 Sequence dependence of undamaged B-DNA structure

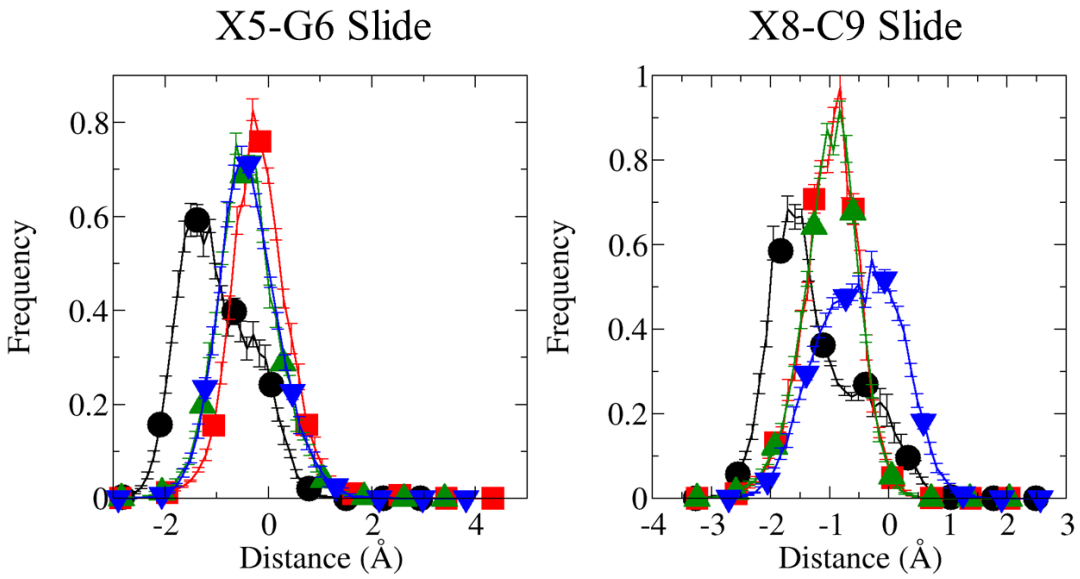


Figure 2.13 An example of the sequence-dependence of the undamaged B-DNA structure. We show on the left the slide between the X5 base pair and G6, one of the platinated guanines. Due to the identity of the X5 base, the AGGC (black circle) and CGGA (red square) sequence contexts are different from each other and from the other two contexts, while TGGA (green upward triangle) and TGGT (blue down triangle) are similar to each other. The image on the right shows the slide between the X8 base pair and the adjacent C9-G16 base pair. Again, the identity of the X8 base pair affects the slide: AGGC and TGGT differ from each other and the other sequences, while CGGA and TGGA are similar to each other.

2.6.9 Sequence-dependence of miscellaneous structural parameters: Cisplatin-DNA adducts

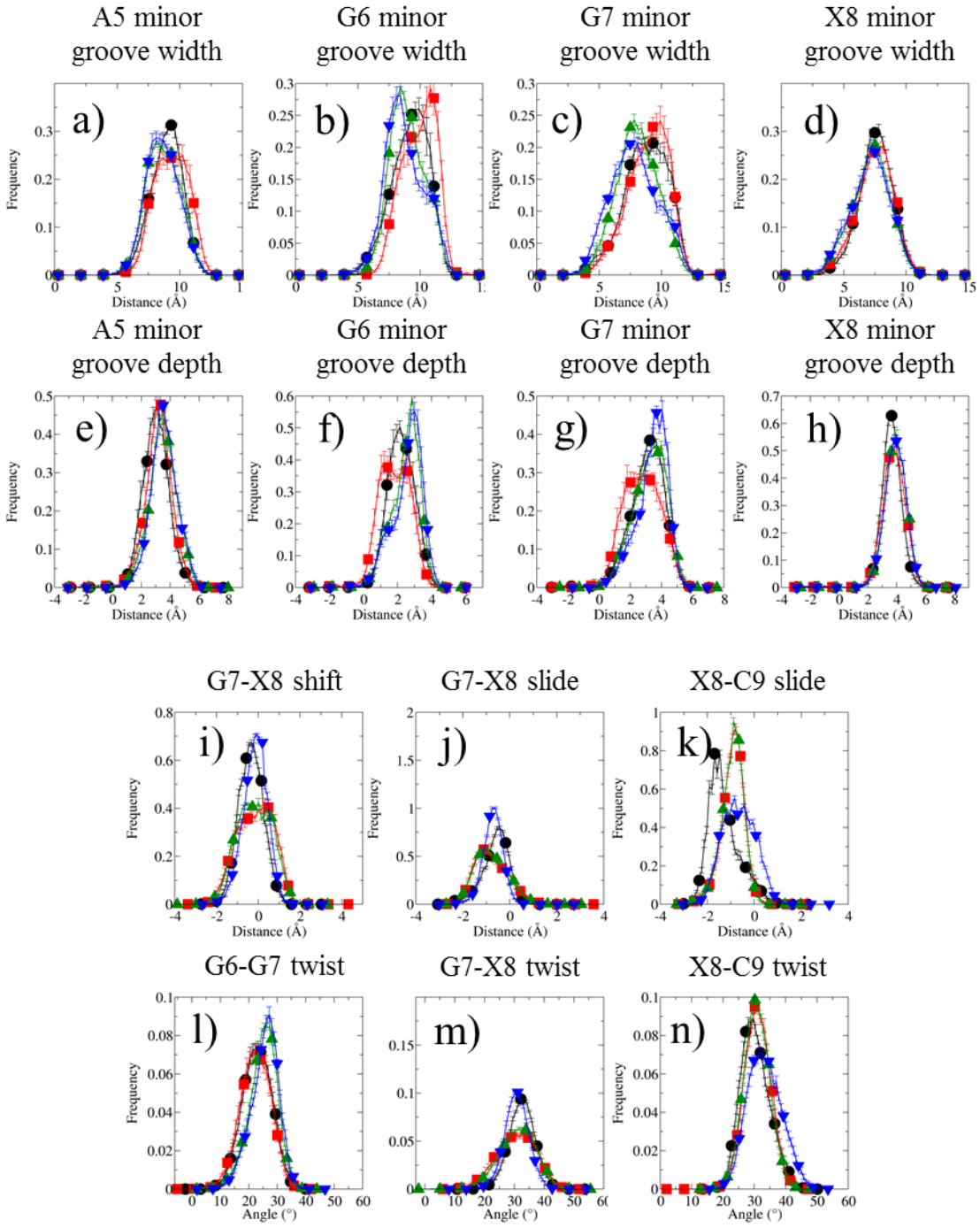


Figure 2.14 Sequence-dependence of various structural parameters for Cp-DNA. Only parameters with particularly apparent differences between sequences are shown. The widening and flattening of the minor groove at the site of the drug-DNA adduct are evident in (a-h). Some minor differences in slide, shift, and

twist near the adducts are shown in (i-n), although we reiterate that it is not likely that these structural differences are related to binding affinity. Backbone angles (e.g. α, γ) also show significant sequence-dependence, but we do not show them because changes in backbone angles are manifest in other structural parameters (e.g. a change in χ might be observed as a change in propeller twist). Legend: black circle, AGGC; red square, CGGA; green upward triangle, TGGA; blue downward triangle, TGGT.

2.6.10 Sequence-dependence of miscellaneous structural parameters: Oxaliplatin-DNA adducts

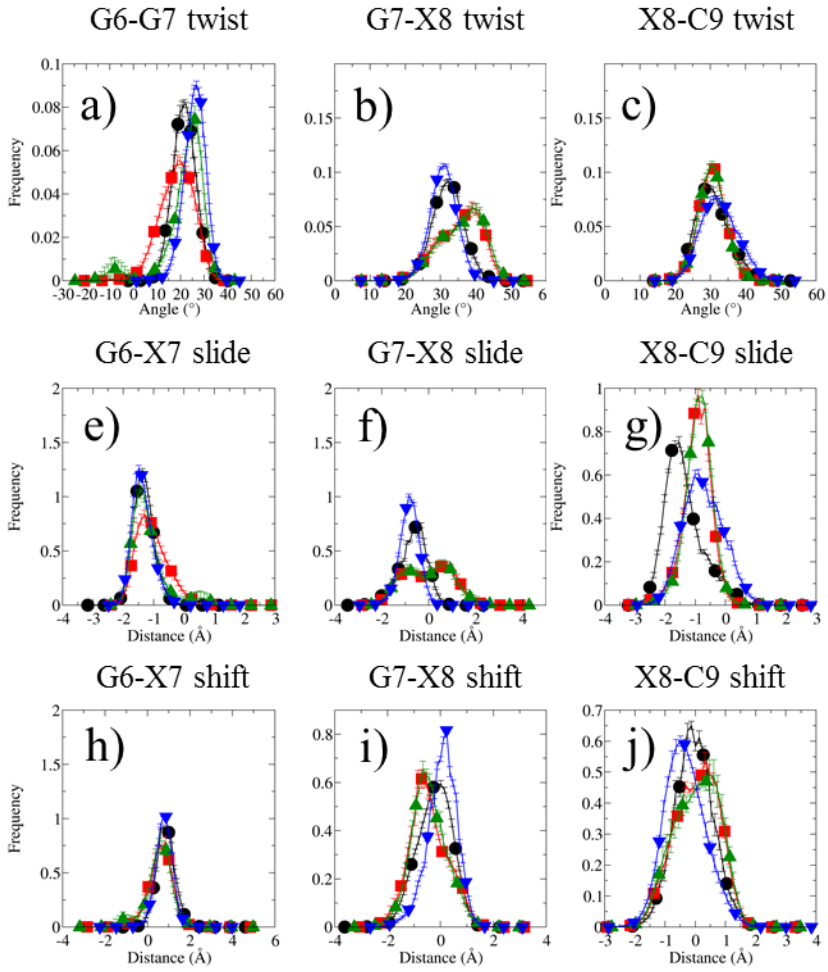


Figure 2.15 Sequence-dependence of various structural parameters for Ox-DNA. The minor groove width and depth are not shown because the plots are essentially the same as those in Figure 2.13 for Cp-DNA. Correlations between twist, slide, and shift are evident, especially at the G7-X8 base pair step (panels (b), (f), and (i)) where it is clear that higher values of twist are associated with higher values of slide and lower values of shift. The other base pair steps show the same trend but it is less pronounced. Legend: black circle, AGGC; red square, CGGA; green upward triangle, TGGA; blue downward triangle, TGGT.

2.6.11 Definition and calculation of hydrogen bond occupancy

Hydrogen bonding was measured with criteria of a distance lower than 3.5 Å between the donor and acceptor atoms and an angle greater than 120° between the donor, hydrogen, and acceptor atoms as in previous work [7]. Occupancy is defined as the frequency with which a hydrogen bond formed. The occupancy of canonical Watson-Crick hydrogen bonds was defined as the average occupancy of the 2 or 3 individual hydrogen bonds possible in AT and GC pairs, respectively. Hydrogen bonds also formed frequently between the Pt-ammines and adjacent bases: the ammine nitrogen atoms were tested for hydrogen bond formation with nearby acceptors, and hydrogen bonds with occupancy higher than 5% were judged to be significant. VMD 1.8.7 was used to analyze hydrogen bonding patterns.

2.6.12 Origin of high frequency of oxaliplatin-G7 hydrogen bond

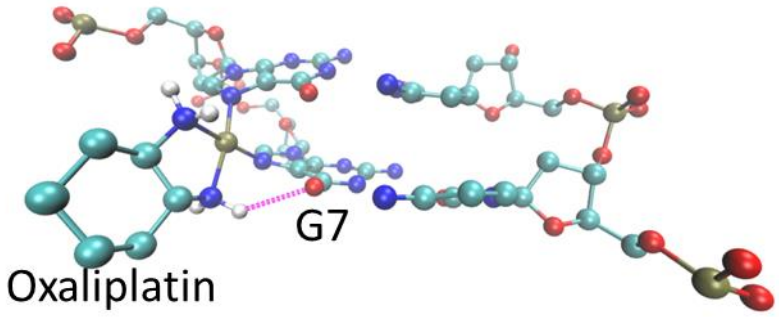


Figure 2.16 The hydrogen bond between oxaliplatin and the O6 atom of G7 is formed about twice as frequently compared to cisplatin-DNA, which is because the equatorial amine hydrogen atom of oxaliplatin is constrained to point toward the O6 atom.

2.6.12 Dynamics of hydrogen bonding states

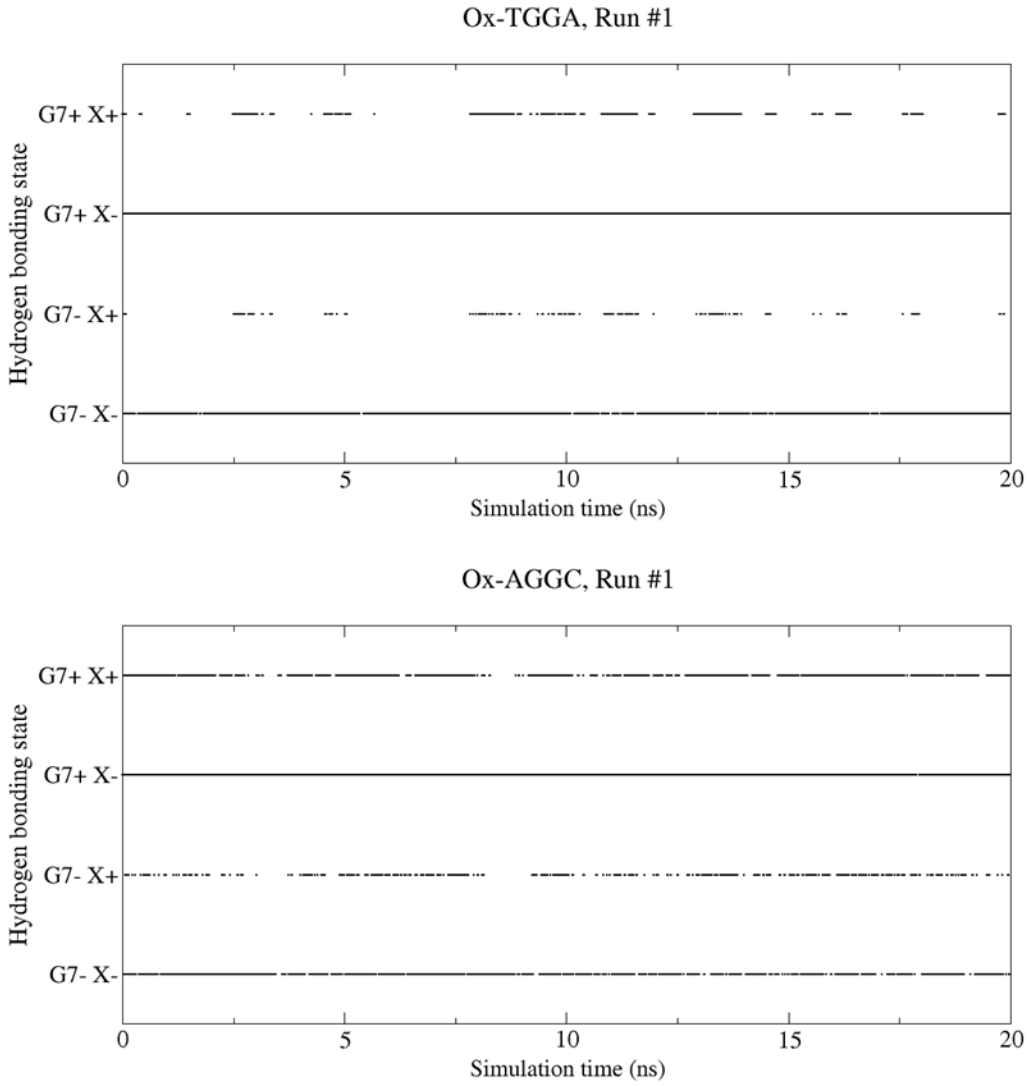


Figure 2.17 Two representative examples of the dynamics of hydrogen bonding states, demonstrating that the structure can rapidly interchange between states (evidenced by many, single isolated data points) but also that there are bursts of 1-2 ns during which a particular state forms more frequently. A continuous line of data points (e.g. G7+X-) occurs when that state is visited frequently.

2.6.13 HMGB1a-bound Cp-DNA in the CGGA context retains some unique hydrogen bonding states

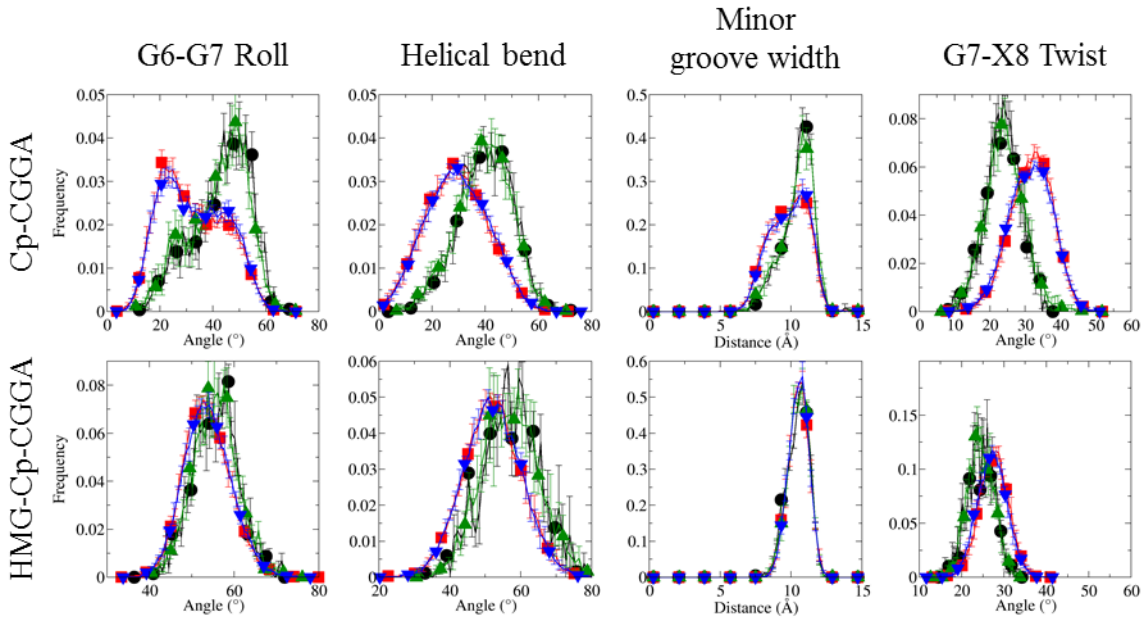


Figure 2.18 Structure of the hydrogen bonding states of unbound-Cp-CGGA (top) and HMGB1a bound-Cp-CGGA (bottom), demonstrating that a hydrogen bond between cisplatin and the adjacent adenine results in increased roll, helical bend, and minor groove width, and lower twist, in both the unbound and bound states. Legend: black circle, G7+ X+; red square, G7+ X–; green upward triangle, G7– X+; blue downward triangle, G7– X–.

2.6.14 Most bound-drug-DNA structures are unaffected by hydrogen bonding state

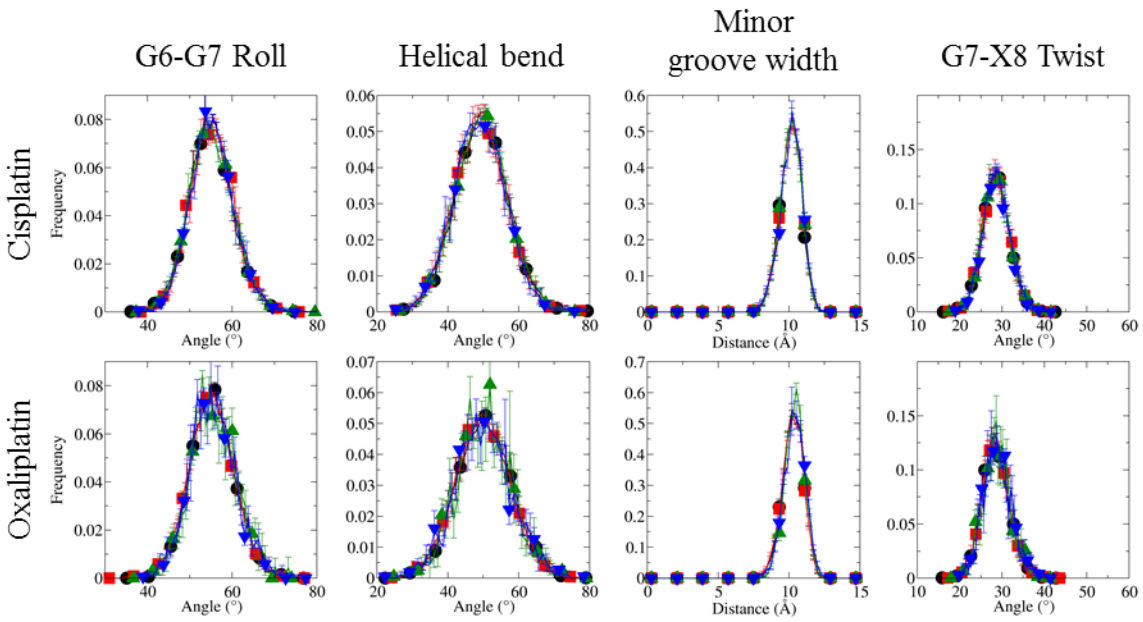


Figure 2.19 Structural distributions of the hydrogen bonding states of bound-Cp-AGGC and bound-Ox-AGGC as a representative example that systems other than bound-Cp-CGGA do not show any unique dynamic conformations in the HMGB1a-bound state. Legend: black circle, G7+ X+; red square, G7+ X–; green upward triangle, G7– X+; blue downward triangle, G7– X–.

Chapter 3

Sequence-specific recognition of cancer drug-DNA adducts by HMGB1a repair protein: The role of deformability

Adapted from: *Biophys. J.* **102** 2331 (2012)

3.1 Introduction

Platinum-based anticancer drugs, such as cisplatin (Cp) and oxaliplatin (Ox), are widely and successfully used to treat cancer [1, 2]. This class of drug damages DNA by covalently binding to adjacent purine bases to form drug-DNA adducts, which prevent replication proteins from accessing the genetic code and lead to cell death [1, 2]. Unfortunately, tumor cells can develop resistance to cisplatin, thereby reducing its effectiveness, but cisplatin-resistant tumor cells are often still sensitive to oxaliplatin despite the similar structure and mode of action of the two drugs [2, 3]. The causes of differential tolerance are not completely understood, but the cellular damage recognition and mismatch repair systems have been implicated [4]. Often, differential drug resistance arises because some of the repair proteins involved in these systems display greater binding affinity for cisplatin than for oxaliplatin [5, 6]. One particular damage recognition protein, the high mobility group protein B1 (HMGB1), has been studied in great detail because the A-domain of this protein (HMGB1a) is a model for other HMG-class DNA binding motifs, which are common in cellular repair mechanisms [7, 8].

Most experimental work has focused on consecutive guanines damaged by the platinum-based drugs, because these are the most common adducts [1]. The binding affinity of HMGB1a for these adducts depends on both the identity of the drug and the bases flanking the damaged guanines (i.e. the sequence context). While the binding affinity of HMGB1a for cisplatin adducts (Cp-DNA) has been studied in numerous sequence contexts, oxaliplatin adducts (Ox-DNA) have only been studied in a small

subset of the possible sequence contexts: AGGC, CGGA, TGGA, and TGGT. For cisplatin, it has been found that the binding affinity proceeds in this order (from highest to lowest affinity): CGGA = TGGA > TGGT > AGGC [5, 6, 9, 10]. For oxaliplatin, the binding affinity proceeds in a different order: TGGT > AGGC = CGGA >> TGGA [5, 6]. Most remarkably, oxaliplatin in the TGGA context (Ox-TGGA) is essentially unrecognized by HMGB1a and has a dissociation constant (K_D) similar to undamaged B-DNA [5]. However, we note that there is some disagreement in previous experimental studies: for instance, the K_D for Cp-TGGA estimated from Ref. [9] differs by an order of magnitude from that of Ref. [5]. Differences such as these qualitatively change the ordering of the binding affinities for cisplatin, which clearly complicates quantitative comparisons (see Table 3.3 in supplementary section 3.6 for additional details and a possible methodological reason for these discrepancies). In all cases of conflict, however, the discrepancies in binding affinity correspond to differences of less than the thermal energy between the two results, indicating that the resolution of the gel mobility shift assay used to obtain these data may be too low to accurately resolve the differences. In light of these conflicting results, we have sought to understand the most striking experimental result: the absolute lack of recognition of Ox-TGGA.

Our previous study focused on the structure and conformational dynamics of cisplatin and oxaliplatin adducts in different sequences [11]. First, we found that the ensemble average structure of the HMGB1a-*bound* drug-DNA was independent of both sequence and drug identity, confirming that the protein-bound drug-DNA structure cannot explain the differential recognition. Next, examining the drug-DNA molecules in the *absence of the protein* HMGB1a, we found that the ensemble average structural distributions only explain some of the differences in binding affinity. Ensemble averages do suggest a possible explanation for the poor recognition of Ox-TGGA. A large dihedral angle between the damaged guanine bases (also known as the roll angle) is considered favorable for binding because it more easily allows intercalation of a specific phenylalanine residue between the damaged bases, and low twist at the binding site is considered favorable because the HMGB1a-bound drug-DNA is unwound with respect to undamaged B-DNA. At least one of these two structural characteristics (i.e. roll or twist) is favorable for oxaliplatin in the AGGC, CGGA, and TGGT sequences, whereas *both* roll and twist are unfavorable for

Ox-TGGA. This provides a plausible case for the weak recognition of Ox-TGGA, but it is not entirely convincing given that the same structural argument does not explain the strong recognition of Ox-TGGT. In keeping with previous simulation studies [12-14], we observed conformational dynamics associated with transient hydrogen bonds that explained some other experimental trends. However, a number of experimental observations remained unexplained, not least of which was the particularly poor recognition of Ox-TGGA. It became apparent that many features, such as the overall bend angle and minor groove width of the drug-damaged DNA structure in the absence of the protein (termed unbound in this paper) is intermediate between the undamaged B-DNA and protein-drug-DNA (termed bound in this paper) structures (Figure 3.1a and 3.1b). This suggests that the protein must deform the drug-DNA structure substantially during binding (Figure 3.1c).

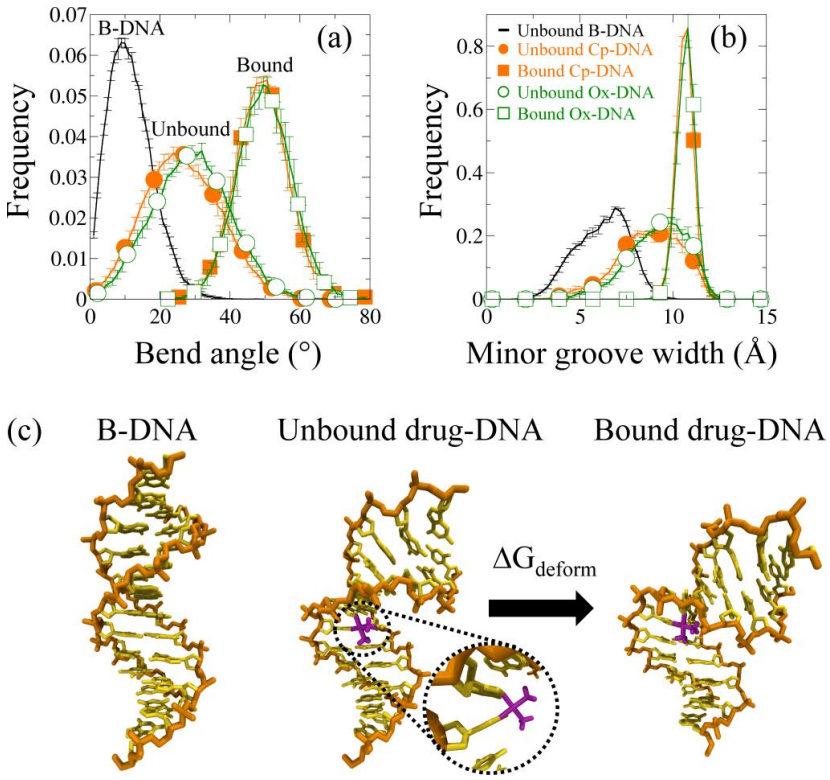


Figure 3.1 Frequency of the helical bend (a) and minor groove width (b) of the unbound drug-DNA structure are intermediate between that of B-DNA and drug-DNA when bound to HMGB1a. c) Single snapshot of structures of B-DNA, unbound drug-DNA, and drug-DNA when bound to HMGB1a protein. The unbound drug-DNA structure is substantially different from the B-DNA structure. This unbound drug-DNA structure is further deformed significantly during binding by HMGB1a. ΔG_{deform} corresponds to the free energy required to deform the unbound drug-DNA structure to the drug-DNA structure when bound to the HMGB1a, calculated in the absence of the protein.

Given that sequence-dependent curvature and flexibility are well-known properties of DNA [15, 16], we hypothesized that a high free energy of deformation of the Ox-TGGA molecule and/or high energy barrier in the kinetic pathway from unbound to bound state could explain the lack of recognition of Ox-TGGA. As such, we chose to employ computationally intensive biased MD simulations and free energy calculations to determine the free energy landscape of Cp-DNA and Ox-DNA in various sequence contexts as a function of reaction coordinates that represent the deformation states.

3.2 Methods

Figure 3.1c illustrates that ΔG_{deform} is the free energy needed to deform the drug-DNA from the unbound structure to the bound structure. In principle, one could calculate the overall free energy of HMGB1a binding to drug-DNA molecule using MMPBSA [17], free energy perturbation [18], or thermodynamic integration [19]; however, these methods are plagued by low accuracy in entropy calculations for such large molecules, or have difficulty in achieving sufficient sampling for large systems with large conformational changes between the initial and final states. Since our hypothesis is that the deformability of the DNA depends on the drug and the sequence context, we have focused purely on calculating ΔG_{deform}, which is a component of the overall free energy of binding, using atomistic molecular dynamics (MD) simulations combined with the umbrella sampling, followed by weighted histogram analysis method (WHAM) [20]. Details of our MD protocol are available in supplementary section 3.6.2, and details of the umbrella sampling procedure are presented below.

Determining the free energy landscape requires defining the configurational space, identifying reaction coordinate(s) that define the configurational space, and ensuring good sampling of structures at all values of the reaction coordinate(s), including those that the system rarely samples in traditional unbiased MD simulations [21-24]. Based on our knowledge of the system at hand (Figure 3.1a and 3.1b) we chose two reaction coordinates to define the configurational space: the minor groove width and the helical bend angle. These coordinates are chosen because our past work has shown that the most notable changes that occur during protein-drug-DNA binding are the helix bending and the minor groove

widening, although the exact order in which these changes occur is not known. Figure 3.2 qualitatively illustrates the definition of these reaction coordinates.

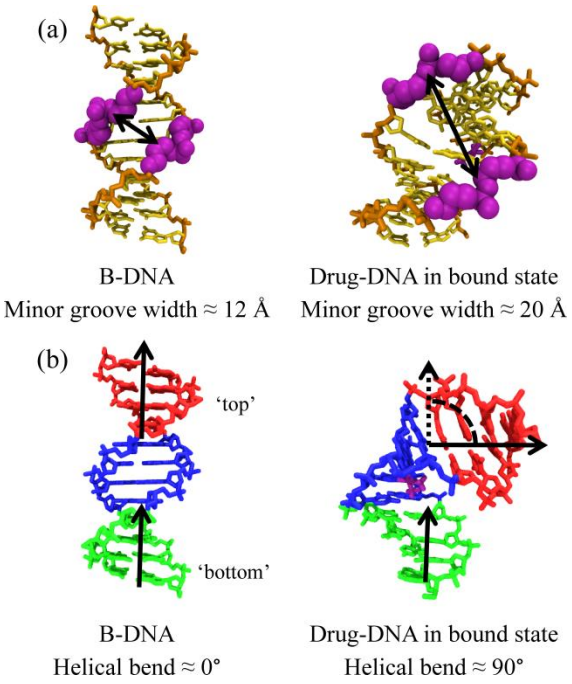


Figure 3.2 The reaction coordinates (a) minor groove width, defined as the distance between the centers-of-mass of the two groups of atoms shown here in van der Waals representation, and (b) the helical bend angle defined as the angle between the vector representing the 'top' 4 base pairs and the corresponding base pairs in an unbent reference structure. The 'bottom' 4 base pairs (b) are constrained to the same orientation as the corresponding base pairs in an unbent reference structure. The combination of the restraints on the top and bottom base pairs causes the DNA to bend.

The first reaction coordinate, minor groove width, is defined as the distance between the center of mass between two groups of backbone atoms, as shown in Figure 3.2a. One of these groups comprises the backbone atoms of one of the guanine bases linked to the drug, and the base 3' of that guanine, and the other group comprises the backbone atoms of the cytosine base paired with the other drug-bonded guanine and the base 3' of that cytosine. The selection of backbone atoms excludes hydrogen atoms. The constraint was applied with the *distance* collective variable in the NAMD 2.8 *colvars* module [25]. This reaction coordinate is the same as that described in Ref. [24]. The second reaction coordinate, the helical bend angle, is defined as the angle between the 'top' 4 base pairs of the DNA (red in Figure 3.2b) and the 'bottom' 4 base pairs (green in Figure 3.2 b). When applying the helical bend angle constraint, we maintain a 12 base pair canonical B-DNA structure aligned with the z-axis as the reference unbent state.

The top 4 base pairs are constrained to a certain angle with respect to this unbent reference structure by a harmonic potential, using the *tilt* collective variable, while the 'bottom' 4 base pairs are constrained to maintain the same orientation as their counterparts in the reference structure, using the *orientationAngle* collective variable. Mathematically, the projection (i.e. the scalar product) of the z-axis of the top 4 base pairs onto the z-axis of the reference structure is constrained to a certain value (e.g. 1 to maintain perfect alignment, 0.707 for an angle of 45°). The z-axis of the top 4 base pairs is determined after a best-fit rotation to fit the reference structure, but there is no rotational constraint actually applied to those base pairs. For the bottom 4 base pairs, the best-fit rotation that maintains the orientation with the reference structure is constrained to the unit quaternion (i.e. no rotation around any axis). We do not need to control the direction of the bend because the DNA is predisposed to bend toward the major groove, the same direction as when HMGB1a binds to DNA. To our knowledge, this definition of a bending reaction coordinate for umbrella sampling has not be reported in literature, but it is similar in nature to other DNA bending reaction coordinates [21, 22].

To enable sufficient sampling of all structural states, we employed the umbrella sampling methodology [26], which introduces a series of biasing potentials at discrete points along the reaction coordinate(s). The results from multiple simulations with differing biasing potentials, where each individual simulation is referred to as a window, are combined using the weighted histogram analysis method (WHAM) [20, 27, 28]. Initially, we performed 1-dimensional (1D) umbrella sampling with only the helical bend constraint. We confirmed that the 1D free energy profiles as a function of the helical bend angle qualitatively reproduce known trends in flexibility for A-tract DNA, and that the free energy profiles have essentially converged with the amount of sampling we have performed (see Figure 3.7 and Figure 3.8 in supplementary section 3.6 for details). For understanding differential binding affinity of HMGB1a to the various drug-DNA molecules the free energy profile as a function of helical bend angle alone is insufficient. This is because helical bend angle alone cannot describe the complete configurational space representing the deformation of the drug-DNA molecule, as during binding the protein not only bends the helix but also widens the minor groove. Therefore, subsequently we performed

2-dimensional (2D) umbrella sampling to determine the entire free energy surface in terms of the helical bend and the minor groove width. Additional details of both the 1D and 2D umbrella sampling procedures are in supplementary section 3.6.3.

We note the limitations with regards to the force field used in this work and its impact on these free energy calculations. First, a recent simulation study revealed that the persistence length of DNA as parameterized in the Amber force field is approximately 80 nm, rather than the typically cited value of 50 nm [29]. While the apparently greater stiffness of the DNA model may quantitatively affect our results, we believe our results are at least qualitatively accurate because: 1) that study (29) did not include the recent *bsc0* modification used in this work, which better represents the backbone torsion angles and in turn may alter the flexibility of the DNA model; and 2) although there are quantitative differences between the Amber force field and experimental measurements, "the differences appear within the range of variations observed with different experimental approaches and solvent conditions" [29]. Therefore, it seems likely that the Amber force field will produce qualitatively correct results regarding DNA deformability in this paper. Second, it is not known how the choice of a different force field to parameterize DNA (e.g. CHARMM) will affect the results of these free energy calculations. However, recent comparisons of the Amber and CHARMM force fields suggest they produce very similar results [30, 31]. It seems plausible that, despite quantitative differences between the various biomolecular force fields, our major qualitative findings would be preserved.

Once the free energy surfaces were obtained, we employed the finite-temperature string (FTS) method [32] to find the minimum free energy pathway (MFEP) between the "unbound" and "bound" structures on the free energy landscapes resulting from 2D umbrella sampling conducted in the absence of the protein. The original zero-temperature string (ZTS) method can be used to find the minimum free energy pathway on smooth free energy surfaces where the MFEP literally follows the lowest free energy values between two or more energy minima [33, 34]. The ZTS method places a series of points (the string) on the free energy surface and allows them to evolve according to the gradient of the free energy, analogous to an unbiased molecular dynamics simulation. However, on 'rough' free energy surfaces

where the roughness is on the order of the thermal energy, the literal MFEP may not be informative because the small-scale details (roughness) of the free energy surface are unimportant. For example, consider a wide shallow free energy 'tube' connecting two minima but with sinusoidal roughness of much smaller amplitude than the walls of the tube or the depth of the minima: qualitatively, the MFEP should be near the center of the tube, not the long, winding path that avoids the insignificant local maxima of the roughness. The FTS method solves this issue by augmenting the ZTS method with a random force with magnitude similar to the roughness of the surface, analogous to Langevin dynamics. This random force allows the points of the string to overcome minor energy barriers and reach a more reasonable MFEP on rough surfaces. We chose the FTS method to calculate MFEPs because the free energy surfaces we have calculated in this work are not entirely smooth. For our purposes, the initial and final points of the MFEP were fixed at the structural coordinates of the "unbound" and "bound" systems from our previous unbiased simulations of the unbound drug-DNA molecule and the HMGB1a-drug-DNA complex, respectively [11]. Additional details of our implementation of the FTS method are available in supplementary section 3.6.3.

Using previous experimental results for the binding affinity of HMGB1a for various DNA sequences, we calculated dissociation constants (K_D) using a method described in literature [13]. Subsequently, we used these K_D values to estimate the relative binding energy of HMGB1a for cisplatin versus oxaliplatin using the method detailed in supplementary section 3.6.4.

3.3 Results

3.3.1 Deformability using 1D reaction coordinate

First, we calculated the 1D free energy as a function of helical bend angle, and the results are presented in Figure 3.3. Most notably, we observe large differences between B-DNA (no drug) and the drug-DNA structures; clearly, the covalent binding of the drug inherently bends the DNA thereby shifting the minima in the free energy of bending to higher bend angles. The minima of the free energy curves coincide quantitatively with the average bend from unbiased drug-DNA simulations from our previous structural

studies[11]. The free energy profile is roughly quadratic for small perturbations but increases linearly for large deformations, in agreement with previous work that focused on the sequence-dependent deformability of B-DNA [21, 35]. There are no free energy barriers along the profile, which is reasonable because bending DNA in the absence of the protein is unfavorable. The presence of the protein during binding would provide additional favorable enthalpic interactions that would make the total binding free energy favorable. However, since here we are simply calculating the free energy cost for bending the drug-DNA molecule in the absence of the protein (ΔG_{deform}), the resulting shape of the free energy curve as a function of helical bend is reasonable.

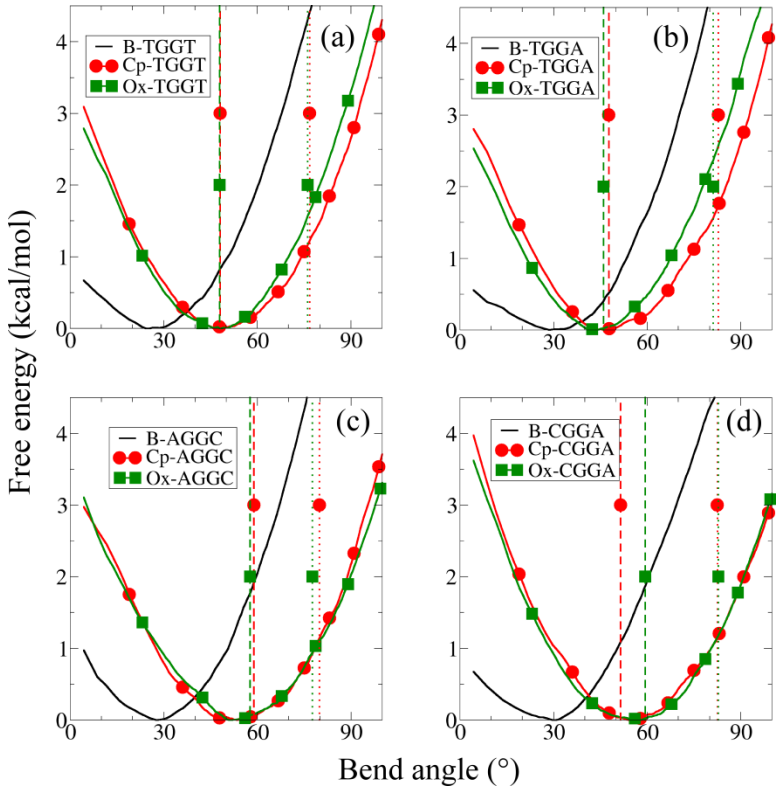


Figure 3.3 Free energy of bending (kcal/mol) for B-DNA, Cp-DNA and Ox-DNA in 4 sequence contexts: (a) TGGT, (b) TGGA, (c) AGGC and (d) CGGA. The dashed vertical lines correspond to the average helical bend angle of the drug-DNA structure from Ref. [11]; the dotted vertical lines correspond to the average helical bend angle of the structure of the drug-DNA when bound to HMGB1a protein. Symbols on these vertical lines denote either Cp (circles) or Ox (squares).

Table 3.1 shows the 1D bending free energy change between the unbound and bound states, $\Delta G_{deform,1D}$, and the difference in this free energy change between Cp-DNA and Ox-DNA, $\Delta\Delta G_{deform,1D}$. If $\Delta\Delta G_{deform,1D}$ is negative it simply implies that it is easier to bend Cp-DNA than Ox-DNA. Experimental

results suggest that HMGB1a has a more favorable binding affinity for Cp-DNA than Ox-DNA. While $\Delta\Delta G_{deform,1D}$ for the two drugs in the AGGC, CGGA, and TGGT contexts is too low to clearly distinguish the drugs from each other in these contexts, in the TGGA context we find that $\Delta\Delta G_{deform,1D}$ is much greater offering some evidence for the much weaker recognition of oxaliplatin in this context.

Sequence	$\Delta G_{deform,1D}$ Cp	$\Delta G_{deform,1D}$ Ox	$\Delta\Delta G_{deform,1D}$ (Cp-Ox)
AGGC	1.14	0.94	0.21
CGGA	1.14	1.18	-0.04
TGGA	1.73	2.37	-0.64
TGGT	1.27	1.57	-0.30

Table 3.1 Free energy of bending in kcal/mol ($\Delta G_{deform,1D}$) and the difference in bending free energy of cisplatin and oxaliplatin ($\Delta\Delta G_{deform,1D}$). The latter quantity is negative when it is easier to bend Cp-DNA than Ox-DNA. Note that $\Delta\Delta G_{deform,1D}$ is low for all sequences except for TGGA, in agreement with experimental results showing that HMGB1a easily distinguishes Ox-TGGA from Cp-TGGA and from other sequence contexts.

Next, we examine the effect of sequence context for Cp- and Ox-DNA adducts (Cp and Ox columns of Table 3.1). For Cp-DNA the variations in $\Delta G_{deform,1D}$ with sequence context do not correspond to the experimentally determined order in binding affinities, although we reiterate that the reported differences in experimentally observed binding affinity are relatively small and may not be discernible. For Ox-DNA, $\Delta G_{deform,1D}$ is greatest for the TGGA context, further explaining the weak recognition by HMGB1a. However, Ox-TGGT has the second highest $\Delta G_{deform,1D}$ and yet is known to have the highest binding affinity. Hence, we conclude that the 1D free energy profile is insufficient to explain all binding affinity trends and turn to the more comprehensive 2D umbrella sampling procedure.

3.3.2 Deformability using 2D reaction coordinate

Figure 3.4 shows frequency distributions from unbiased simulations and 2D free energy landscapes for Cp-TGGA and Ox-TGGA as a function of helical bend angle and minor groove width. The frequency distributions (Figure 3.4a and 3.4b) show the most common structure in the unbound state and the bound state. These distributions are used to identify the initial (I) and final (F) coordinates in the free energy landscapes. The free energy landscapes (Figure 3.4c and 3.4d) show that helical bend and minor groove width are correlated because the free energy increases more gradually when the two coordinates increase

in tandem, which is expected because we have also observed the positive correlation of helical bend and minor groove width in our structural studies [11].

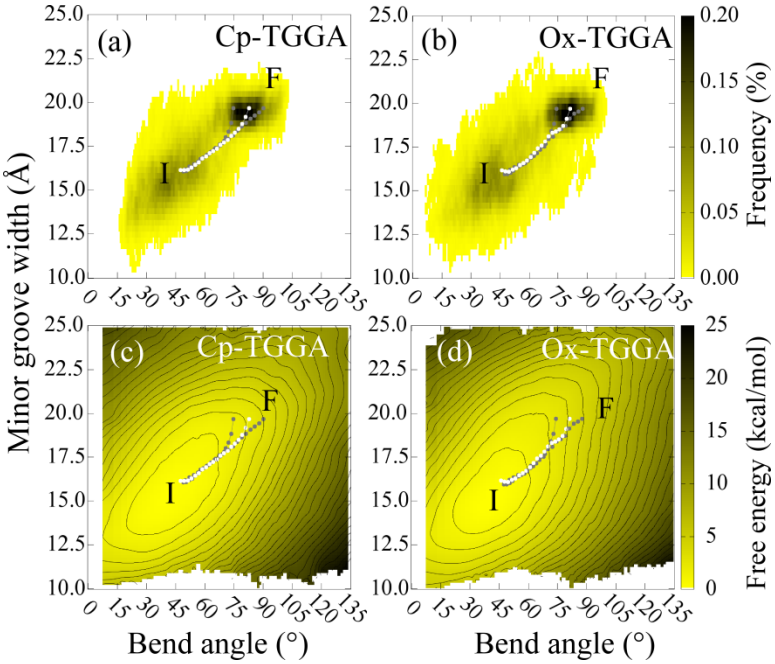


Figure 3.4 Frequency distributions of the structural parameters from unbiased simulations for (a) Cp- and (b) Ox-TGGA and 2D free energy landscapes of deformation for (c) Cp- and (d) Ox-TGGA. The frequency distributions (a) and (b) were calculated from the combined series of bend angle and minor groove width values visited by both the unbound and bound structures in unbiased simulations. The free energy landscapes (c) and (d) were calculated using umbrella sampling and WHAM, as described in the text. White and gray dotted lines are minimum free energy pathways (MFEPs) from the unbound structure (marked as I) to the bound structure (marked as F). The multiple MFEPs denote paths calculated from the mean values of the unbound state (I) to multiple endpoints in the bound state (F). The y-coordinate of each endpoint (F) is the mean value of the minor groove width. The x-coordinate of the endpoint (F) of the white path is the mean value of the bend angle, while the x-coordinates of the gray paths are the mean ± the standard deviation of the bend angle. Contour lines are spaced 1 kcal/mol apart. White regions indicate the absence of data.

As with the 1D bending free energy profiles, the 2D landscapes are relatively smooth and lack free energy barriers. We also find that the drugs have qualitatively different free energy landscapes, with oxaliplatin-DNA generally experiencing a greater increase in free energy with bend angle than cisplatin. Figure 3.9 and Figure 3.10 in supplementary section 3.6 show the free energy landscapes for all drug and sequence combinations, and Figure 3.11 in supplementary section 3.6 confirms that the 2D free energy landscapes have essentially converged with the amount of sampling we have performed. We note that the application of these two global structural constraints also results in many of the same local structural deformations

associated with the binding of HMGB1a, such as the increase in the roll angle between the damaged guanine bases, which suggests that this choice of reaction coordinates is reasonable.

Figure 3.4c and 3.4d also show multiple minimum free energy pathways (MFEPs) within each landscape. These multiple MFEPs correspond to the path with different final endpoints (marked as 'F' in Figure 3.4c and 3.4d), so as to assess the sensitivity of our findings to the structure of the bound state. The MFEPs in Figure 3.4c and 3.4d correspond to three different final structural states: the white pathway ends at the mean values of helical bend (x-coordinate) and minor groove width (y-coordinate), while the two gray pathways end at the mean plus-or-minus the standard deviation of the probability distribution of the helical bend (x-coordinate) and the mean of the minor groove width (y-coordinate). These structural coordinates of the helical bend angle and minor groove width were obtained from our previous study [11]. The initial point of the pathway (marked as 'I' in Figure 3.4) is the same in all cases because that region of the landscape is quite shallow, and consequently changes in the initial structure cause only small changes in the initial energy. We also calculated MFEPs ending at the mean plus-or-minus the standard deviation of the probability distribution of the minor groove width (y-coordinate) and mean of the bend angle (x-coordinate), but we found that the free energy values along these pathways, termed *free energy profiles*, did not differ significantly from the pathway ending at the mean structural values (i.e. the white pathways in Figure 3.4c and 3.4d). We have also found that the profiles are insensitive to small deviations in the specific pathway taken by calculating the free energy profile of a linear pathway between the initial and final states, which produces a free energy profile similar to that produced by the MFEP (Figure 3.12 in supplementary section 3.6). Consequently, we focus on the three free energy profiles along the MFEPs ending at the mean plus-or-minus the standard deviation of the helical bend (x-coordinate) and the mean minor groove width (y-coordinate).

Figure 3.5 shows the free energy values along the points of the minimum free energy pathways calculated for all drugs and sequence contexts (i.e. free energy profiles). The bold lines are for the MFEP ending at the mean structural coordinates (like the white pathways in Figure 3.4c and 3.4d) and the dotted lines are for the MFEPs ending at the mean plus-or-minus the standard deviation of the helical bend (like

the gray pathways in Figure 3.4c and 3.4d). For Cp-DNA, it is not possible to distinguish the sequence contexts because the standard deviation profiles largely overlap the mean profiles of the sequence contexts. However, the situation is quite different for Ox-DNA: although the standard deviation profiles of the AGGC, CGGA, and TGGT sequences essentially overlap each other, the TGGA sequence reaches significantly higher free energies throughout the pathway and does not overlap the other sequences. This provides the strongest evidence yet for the extremely low binding affinity of HMGB1a for Ox-TGGA.

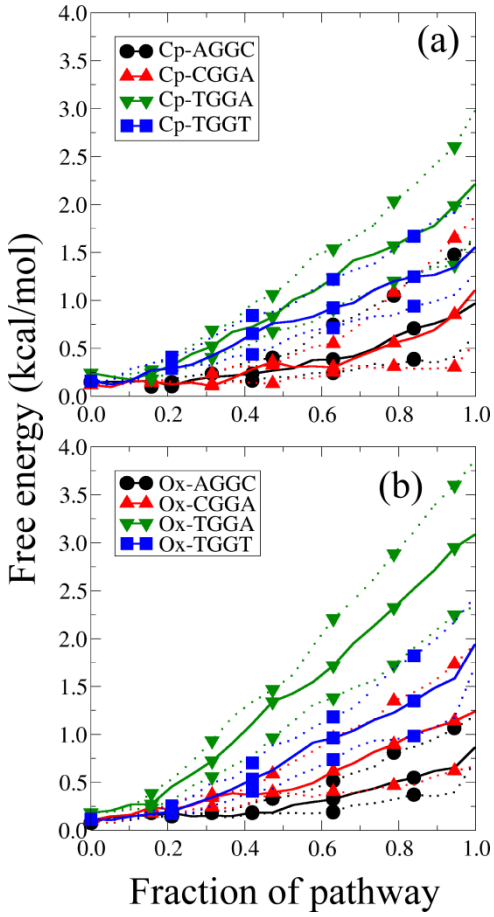


Figure 3.5 Free energy profiles along the MFEP of each drug and sequence combination. Bold lines correspond to MFEPs ending at the mean structure of the bound state (white pathways in Figure 3.4). Dotted lines correspond to the MFEPs ending at the mean bend angle plus-or-minus the standard deviation of the bend angle and at the mean minor groove width (gray pathways in Figure 3.4). The x-axis indicates the progress along the pathway, with the unbound state being 0.0 and the bound state being 1.0.

Table 3.2 shows the differences in 2D deformability free energy changes for cisplatin and oxaliplatin ($\Delta\Delta G_{deform,2D}$), alongside the overall free energy differences calculated from experimental binding affinity data ($\Delta\Delta G_{bind}$). We note that quantitative comparison between the simulation and

experimental data is not possible because the experimental data measures the binding free energy of HMGB1a (ΔG_{bind}) whereas the simulation methodology calculates only the free energy of deformation of the drug-DNA molecule (ΔG_{deform}) in the absence of the protein. However, qualitative comparisons are still valid. The differences in ΔG_{deform} between Cp-DNA and Ox-DNA are less than the thermal energy in the AGGC, CGGA, and TGGT contexts, and therefore it may not be possible to distinguish these contexts as discussed above. However, the TGGA context shows a much more significant discrepancy between cisplatin and oxaliplatin, in agreement with the experimental findings for $\Delta\Delta G_{bind}$.

Sequence (Source)	K_D ratio	$\Delta\Delta G_{bind}$ from K_D ratio	$\Delta\Delta G_{deform,2D}$
AGGC [6]	2.6	-0.56	0.10
CGGA [6]	3.2	-0.69	-0.13
TGGA [5]	55.1	-2.39	-0.87
TGGT [6]	1.8	-0.35	-0.39

Table 3.2 Relative free energy (kcal/mol) of HMGB1a binding to cisplatin versus oxaliplatin. Values are calculated from experimental binding affinity data ($\Delta\Delta G_{bind}$ from K_D ratio) and from 2D deformability umbrella sampling calculations ($\Delta\Delta G_{deform,2D}$). $\Delta\Delta G$ is the difference in 2D deformation energy between oxaliplatin and cisplatin, i.e. a negative value indicates deforming Cp-DNA is more favorable than deforming Ox-DNA.

The cause of this striking difference between Cp- and Ox-TGGA appears to be steric clashes between the bulky diaminocyclohexane ring of oxaliplatin and methyl groups on nearby thymine bases (the thymine of TGGA and thymine paired opposite the A of TGGA), as shown in Figure 3.6. Methyl groups on these thymine bases extend into the already crowded minor groove even in undamaged B-DNA. These methyl groups have previously been shown to alter the flexibility of both undamaged and drug-damaged DNA by interacting with each other in consecutive A-T base pairs or by sterically hindering a platinum-based drug with a bulky chemical group [36, 37]. The TGGA and TGGT contexts contain two such methyl groups, one on either side of the adduct, and these groups ‘pinch’ the diaminocyclohexane ring of oxaliplatin but do not interact closely with the smaller amine groups of cisplatin. This is visually evident when comparing the trajectories, and we have verified this observation by quantifying the distance between the platinum atom of the drugs and the adjacent methyl groups along

the MFEPs. The platinum atom was chosen as a measure of how closely each drug interacts with the methyl groups because both cisplatin and oxaliplatin contain this feature.

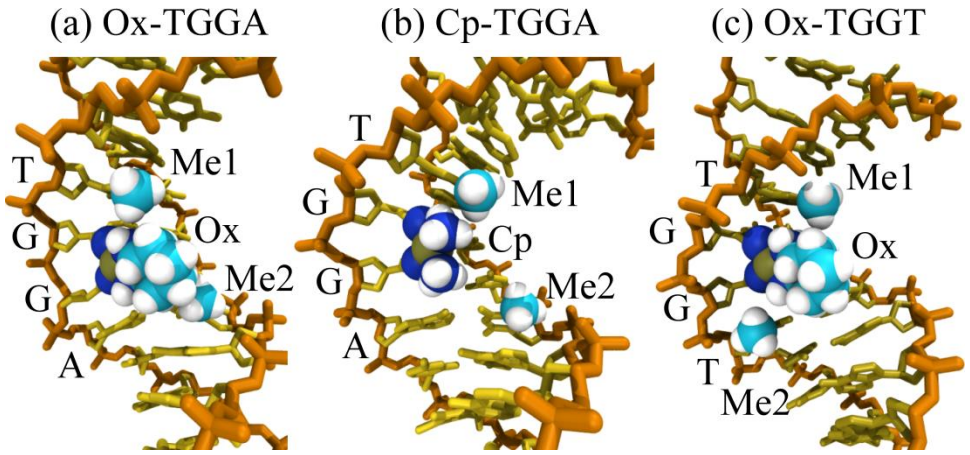


Figure 3.6 Position of thymine methyl groups (Me1 and Me2) relative to cisplatin (Cp) and oxaliplatin (Ox) in the TGGA and TGGT contexts. It is qualitatively evident that the bulky diaminocyclohexane moiety of Ox is 'pinched' between the two methyl groups in the TGGA context (a), which hinders deformation of Ox-TGGA, whereas the smaller amine groups of Cp do not interact strongly with the methyl groups (b). Oxaliplatin is less constrained by the two methyl groups in the TGGT context (c) than in the TGGA context (a).

In the TGGT context, both Cp and Ox adducts maintain a distance of approximately 5 Å from both methyl groups in both the unbound and bound states, suggesting that Cp-TGGT and Ox-TGGT present similar hindrance during binding (and deformation). In the TGGA context, Cp maintains distances of 5 Å and 9 Å from the two methyl groups in both the unbound and bound states; Ox maintains the same close contact of 5 Å with the methyl group immediately 5' of the adduct (Me1 in Figure 3.6) in the unbound and bound states, but the distance to the methyl group on the opposite DNA strand (Me2 in Figure 3.6) decreases from 9 Å to 5 Å between the unbound and bound states. This decrease in distance indicates that oxaliplatin interacts more closely with the methyl Me2 in the bound state, and this closer interaction causes steric hindrance that makes it more difficult to deform Ox-TGGA than Ox in other contexts. The increased difficulty of deformation is likely a major contributor to the lower binding affinity of HMGB1a for Ox-TGGA. In contrast to TGGA, the CGGA context possesses only the methyl group on the strand opposite the drug-DNA adduct (equivalent to Me2 in the TGGA context), and both drugs maintain a

platinum-methyl distance of approximately 9 Å, indicating lower steric hindrance during deformation of both Cp-CGGA and Ox-CGGA.

3.4 Conclusions

In the present work, we have calculated the free energy of deformation of DNA damaged by the cancer drugs cisplatin and oxaliplatin in several DNA sequence contexts. The results of this study suggest that the greater free energy penalty for deforming oxaliplatin in the TGGA context (versus cisplatin and other sequence contexts) is responsible for the extremely low binding affinity of HMGB1a for Ox-TGGA. This result stands in contrast to previous studies that have suggested structure is the key factor in differential binding affinity but is fully consistent with findings that deformability often plays a crucial role in protein binding affinity [37-39]. The reason for the sequence-dependent deformation free energy of oxaliplatin-DNA adducts appears to be steric hindrance by nearby thymine methyl groups, the locations of which depend on the sequence context. Recent efforts focused on steric hindrance between the drug and DNA [40-42], and our work in this paper further support the idea of future cancer drug design to focus on tuning repair protein binding by tailoring this type of steric hindrance with adjacent methyl groups.

3.5 References

[1] E. R. Jamieson and S. J. Lippard, Chemical Reviews 99 (1999) 2467.
[2] D. Wang and S. J. Lippard, Nature Reviews Drug Discovery 4 (2005) 307.
[3] J. M. Woynarowski, S. Faivre, M. C. S. Herzig, B. Arnett, W. G. Chapman, A. V. Trevino, E. Raymond, S. G. Chaney, A. Vaisman, M. Varchenko, and P. E. Juniewicz, Molecular Pharmacology 58 (2000) 920.
[4] S. G. Chaney, S. L. Campbell, E. Bassett, and Y. B. Wu, Critical Reviews in Oncology Hematology 53 (2005) 3.
[5] M. Wei, S. M. Cohen, A. P. Silverman, and S. J. Lippard, Journal of Biological Chemistry 276 (2001) 38774.
[6] J. Malina, O. Novakova, M. Vojtiskova, G. Natile, and V. Brabec, Biophysical Journal 93 (2007) 3950.
[7] J. C. Huang, D. B. Zamble, J. T. Reardon, S. J. Lippard, and A. Sancar, Proceedings of the National Academy of Sciences of the United States of America 91 (1994) 10394.
[8] D. S. Pilch, S. U. Dunham, E. R. Jamieson, S. J. Lippard, and K. J. Breslauer, Journal of Molecular Biology 296 (2000) 803.
[9] S. U. Dunham and S. J. Lippard, Biochemistry 36 (1997) 11428.
[10] S. M. Cohen, Y. Mikata, Q. He, and S. J. Lippard, Biochemistry 39 (2000) 11771.
[11] R. M. Elder and A. Jayaraman, Molecular Simulation 38 (2012) 793.
[12] S. Sharma, P. Gong, B. Temple, D. Bhattacharyya, N. V. Dokholyan, and S. G. Chaney, Journal of Molecular Biology 373 (2007) 1123.

[13] S. Ramachandran, B. R. Temple, S. G. Chaney, and N. V. Dokholyan, Nucleic Acids Research 37 (2009) 2434.
[14] D. Bhattacharyya, S. Ramachandran, S. Sharma, W. Pathmasiri, C. L. King, I. Baskerville-Abraham, G. Boysen, J. A. Swenberg, S. L. Campbell, N. V. Dokholyan, and S. G. Chaney, PLoS ONE 6 (2011) e23582.
[15] R. Stefl, H. Wu, S. Ravindranathan, V. Sklenár, and J. Feigon, Proceedings of the National Academy of Sciences of the United States of America 101 (2004) 1177.
[16] J. A. Parkinson, Y. Chen, P. del Socorro Murdoch, Z. Guo, S. J. Berners-Price, T. Brown, and P. J. Sadler, Chemistry- A European Journal 6 (2000) 3636.
[17] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case, Journal of the American Chemical Society 120 (1998) 9401.
[18] J. Aaqvist, Journal of Physical Chemistry 94 (1990) 8021.
[19] D. Rodriguez-Gomez, E. Darve, and A. Pohorille, Journal of chemical physics 120 (2004) 3563.
[20] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, Journal of Computational Chemistry 13 (1992) 1011.
[21] J. Curuksu, M. Zacharias, R. Lavery, and K. Zakrzewska, Nucleic Acids Research 37 (2009) 3766.
[22] J. Curuksu, K. Zakrzewska, and M. Zacharias, Nucleic Acids Research 36 (2008) 2268.
[23] K. Spiegel, A. Magistrato, P. Carloni, J. Reedijk, and M. L. Klein, Journal of Physical Chemistry B 111 (2007) 11873.
[24] M. Zacharias, Biophysical Journal 91 (2006) 882.
[25] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of Computational Chemistry 26 (2005) 1781.
[26] G. M. Torrie and J. P. Valleau, Journal of Computational Physics 23 (1977) 187.
[27] A. Grossfield, "WHAM: the weighted histogram analysis method," version 2.0.4, membrane.urmc.rochester.edu/content/wham.
[28] B. Roux, Computer Physics Communications 91 (1995) 275.
[29] A. K. Mazur, Biophysical Journal 91 (2006) 4507.
[30] A. Pérez, F. J. Luque, and M. Orozco, Accounts of Chemical Research 45 (2011) 196.
[31] Z. Liu, B. Ensing, and P. B. Moore, Journal of Chemical Theory and Computation 7 (2010) 402.
[32] E. Vanden-Eijnden and M. Venturoli, Journal of Chemical Physics 130 (2009) 194103.
[33] W. E, W. Ren, and E. Vanden-Eijnden, Physical Review B 66 (2002) 052301.
[34] W. E, W. Ren, and E. Vanden-Eijnden, Journal of Chemical Physics 126 (2007) 164103.
[35] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, Nature nanotechnology 1 (2006) 137.
[36] J. Malina, M. Voitiskova, V. Brabec, C. I. Diakos, and T. W. Hambley, Biochemical and Biophysical Research Communications 332 (2005) 1034.
[37] Z. S. Juo, T. K. Chiu, P. M. Leiberman, I. Baikalov, a. J. Berk, and R. E. Dickerson, Journal of molecular biology 261 (1996) 239.
[38] L. Czapla, J. P. Peters, E. M. Rueter, W. K. Olson, and L. J. Maher Iii, Journal of Molecular Biology 409 (2011) 278.
[39] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann, Annual review of biochemistry 79 (2010) 233.
[40] N. Margiotta, G. Natile, F. Capitelli, F. P. Fanizzi, A. Boccarelli, P. De Rinaldis, D. Giordano, and M. Coluccia, Journal of Inorganic Biochemistry 100 (2006) 1849.
[41] J. Kasparkova, T. Suchankova, A. Halamikova, L. Zerzankova, O. Vrana, N. Margiotta, G. Natile, and V. Brabec, Biochemical Pharmacology 79 (2010) 552.
[42] K. S. Lovejoy and S. J. Lippard, Dalton Transactions (2009) 10651.
[43] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics and Modelling 14 (1996) 33.

[44] T. J. Macke and D. A. Case, in Molecular Modeling of Nucleic Acids, Vol. 682, American Chemical Society, 1997, p. 379.
[45] A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco, Biophysical Journal 92 (2007) 3817.
[46] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins 65 (2006) 712.
[47] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, Journal of the American Chemical Society 117 (1995) 5179.
[48] E. D. Scheeff, J. M. Briggs, and S. B. Howell, Molecular Pharmacology 56 (1999) 633.
[49] S. Yao and J. P. Plastaras, Inorganic Chemistry 33 (1994) 6061.
[50] A. Bondi, Journal of Physical Chemistry 68 (1964) 441.
[51] T. W. Hambley, Inorganic chemistry 37 (1998) 3767.
[52] I. Brukner, M. Dlakic, A. Savic, S. Susic, S. Pongor, and D. Suck, Nucleic acids research 21 (1993) 1025.

3.6 Supplementary information

3.6.1 Discrepancies between experimental results for HMGB1a binding affinity for Cp-DNA adducts

Table 3.3 Several studies have been conducted on the binding affinity of the protein HMGB1a for cisplatin-DNA adducts in numerous sequence contexts, and these studies have produced some apparently inconsistent results. The total DNA concentration, total protein concentration, and fraction of bound DNA can be used to calculate a dissociation constant (K_D) using the method described in the supplementary information of Ref. [13]. The K_D values we have calculated from the gel-mobility shift assay data must be regarded as estimates because the values of fraction of bound DNA (θ) have been estimated visually from the sources noted in the table, and because the total protein concentration in Ref. [6] is not obvious it has been assumed 30 nM, the highest value apparently used in that work and consistent with Ref. [5]. The only apparent difference in protocol between these references is the length of time that HMGB1a is incubated with the drug-DNA prior to performing the gel-mobility shift assay: Refs. [6] and [5] incubate for 30 minutes whereas Refs. [9] and [10] incubate for 1 hour. We conjecture that differences in the kinetics of binding lead to differences in the apparent binding affinity, and indeed the binding affinity of the experiments using longer incubation time report similar or greater apparent binding affinity, which is consistent with a kinetic barrier affecting the results obtained with the shorter incubation time.

Sequence [source]	$[DNA]_{total}$ (nM)	$[Protein]_{total}$ (nM)	Fraction DNA bound (θ)	K_D (nM)
AGGC [6]	10	30	0.38	44
AGGC [5]	12	30	0.33	54
AGGC [9]	5	5	0.16	22
CGGA [6]	10	30	0.45	31
CGGA [9]	5	5	0.49	3
TGGA [5]	17	30	0.32	54
TGGA [9]	5	5	0.43	4
TGGT [6]	10	30	0.69	11
TGGT [9]	5	5	0.22	14

3.6.2 Equilibration procedure and details of simulation protocol

Our equilibration procedure is as follows. First, in constant volume and temperature conditions (NVT), the water and ions were minimized with the conjugate gradient algorithm for 10000 steps, with the DNA subjected to a 500 kcal/mol-A^2 harmonic restraining potential. Second, in NVT conditions, with the DNA restraining potential reduced to 100 kcal/mol-A^2, the system was heated from 0 to 300 K over 20 ps. Third, in constant pressure and temperature conditions (NPT) and with the restraint reduced to 50 kcal/mol-A^2 the system was relaxed for another 20 ps. Fourth, in NVT, the restraining potential was

reduced in three steps (50, 10, and 5 kcal/mol-A^2) and minimized for 2000 steps each. Fifth, in NVT with the same 5 kcal/mol-A^2 restraint the system was heated from 10 to 300 K over 20 ps. Sixth, in NPT, the restraint was reduced from 5 to 1 to 0.1 to 0 kcal/mol-A^2 over 20 ps. Seventh, in NVT with no restraints, the system was heated from 10 to 300 K over 20 ps. Finally, the system was heated from 100 to 300 K over 20 ps in NPT. This equilibration procedure was used prior to performing several short simulations to generate initial structures for umbrella sampling (described in supplementary section 3.6.3).

For systems in the absence of protein, referred to as *unbound* systems, we used available NMR solution structures of Cp-DNA and Ox-DNA (PDB ID: 2NPW; PDB ID: 1PGC) to develop starting structures for both drugs in four sequence contexts by manually replacing the base pairs flanking the platinated guanines with base pairs from B-DNA using VMD 1.9 [43]. The four sequences studied here are: 5′-d(CCTC**A*GG*C**CTCC)-3′, 5′-d(CCTC**C*GG*A**CTCC)-3′, 5′-d(CCTC**T*GG*A**CTCC)-3′, 5′-d(CCTC**T*GG*T**CTCC)-3′, where the sequence context is indicated by bolded text and the platinated guanines are italicized. As in the experimental structures, the drugs were covalently bound to the N7 atoms of the central adjacent guanines (G6 and G7). Undamaged, canonical B-DNA was generated using Nucleic Acid Builder (*nab*) in AmberTools 1.5 [44]. In all cases where a molecule was manually edited (e.g. mutation of a base pair), the edited structure was subjected to 500 steps of conjugate gradient minimization to eliminate any extremely unfavorable geometry before being used for simulation. Counterions (Na^+) were added to neutralize the system charge. Systems were solvated with TIP3P water molecules with a minimum distance of 10 Å from the solute to the edge of the box; in each case, approximately 5000 to 5500 water molecules were added. Production MD was performed in the NPT ensemble, using Langevin dynamics to maintain a temperature of 300 K and a pressure of 1 atm. The SHAKE algorithm was used to restrain the lengths of all bonds to hydrogen, and a time step of 2 fs was used. We used NAMD 2.9 for all simulations and the Amber ff99 force field (with the recent SB and bsc0 modifications) to parameterize nucleic acids, protein, TIP3P water molecules, and counterions [25, 45-47]. Force fields for cisplatin, oxaliplatin, and the damaged guanine bases are available in literature [48-51]; these are identical to those in our previous work [11].

3.6.3 Umbrella sampling procedure and details of the finite-temperature string method implementation

To obtain bent initial structures, we conducted a single 2 ns simulation for each system (B-, Cp-, and Ox-DNA) in each sequence in which the DNA was gradually bent from 0° to 135°; individual frames were extracted from this trajectory for use in each of 28 sampling windows in 5° increments from 0° to 135°. These were used as starting structures for 1D bending umbrella sampling simulations. These bent starting structures were also used to generate starting structures for 2D deformability umbrella sampling simulations: we performed another set of 2 ns simulations to change the minor groove width in 1 Å intervals over the nearly full range of 11 to 24 Å. Not all minor groove widths were simulated for all bend angles: several of the 24 Å windows were excluded because we found that sampling in these regions was already sufficient. The final structures of these groove-widening simulations were used as the starting structures for 2D deformability umbrella sampling simulations. Each umbrella sampling window was sampled for 10 ns, which we determined was sufficient by gradually increasing the amount per window until the free energy profiles had converged. Each 1D bending free energy profile requires 280 ns of simulation time and each 2D free energy surfaces requires approximately 3740 ns of simulation time.

We implemented the finite-temperature string (FTS) method as detailed in [32]. We used a temperature of 0.5, 20 points on the string, performed 50000 iterations with a time step of 0.01, and used a smoothing coefficient of 0.1. These particular parameters were determined to produce qualitatively reasonable minimum free energy pathways (MFEPs). In particular, the temperature of 0.5 was chosen by visual inspection of the 3D free energy surface to determine the approximate height of the small-scale roughness of the surface, followed by some experimentation to determine a reasonable value for the temperature. At lower temperatures, the pathways are not qualitatively smooth enough, while at higher temperatures numerical instabilities sometimes occur when the random force overwhelms the underlying potential surface, leading to unpredictable behavior. The initial guess for the string was a straight line

between the structural coordinates of the unbound and bound states determined from our previous unbiased simulations, and the initial and final points of the string were fixed at these values. Although the FTS method is not typically used with the endpoints fixed, it was necessary in this case because the initial and final points are not free energy minima separated by a free energy barrier. We note that the MFEP calculated *without* the endpoints fixed lies along the same pathway as *with* the endpoints fixed but expands to fill a longer portion of the landscape in the same diagonal direction as the MFEP *with* the endpoints fixed. The free energy gradient in the x- (helical bend) and y- (minor groove width) directions was estimated numerically using a central difference approximation at each point, except for points at the edges of the free energy surface, where either a forward or backward difference was used as appropriate. The code was implemented in MATLAB and modified from the code available at http://www.math.princeton.edu/string/code/ftsMueller.m.

3.6.4 Calculation of relative binding energies from dissociation constants (K_D)

After calculating the dissociation constants (K_D) of HMGB1a for various drug-DNA adducts using the method in the supplementary material of Ref. [13], we calculated the relative binding affinity using the following equations. States 1 and 2 can represent cisplatin and oxaliplatin or two different sequence contexts, so we can find the free energy difference for 'mutating' one drug to the other or one sequence to another ($\Delta\Delta G$).

$$\Delta G_1 = -RT \ln K_{D,1}$$

$$\Delta G_2 = -RT \ln K_{D,2}$$

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = -RT \ln \left[\frac{K_{D,2}}{K_{D,1}}\right] = \left(-1.985877 \times 10^{-3} \frac{kcal}{K\ mol}\right)(300\ K) \ln \left[\frac{K_{D,2}}{K_{D,1}}\right]$$

3.6.5 Bending free energy profiles of A-tract DNA with varying intrinsic stiffness and curvature using 1D umbrella sampling

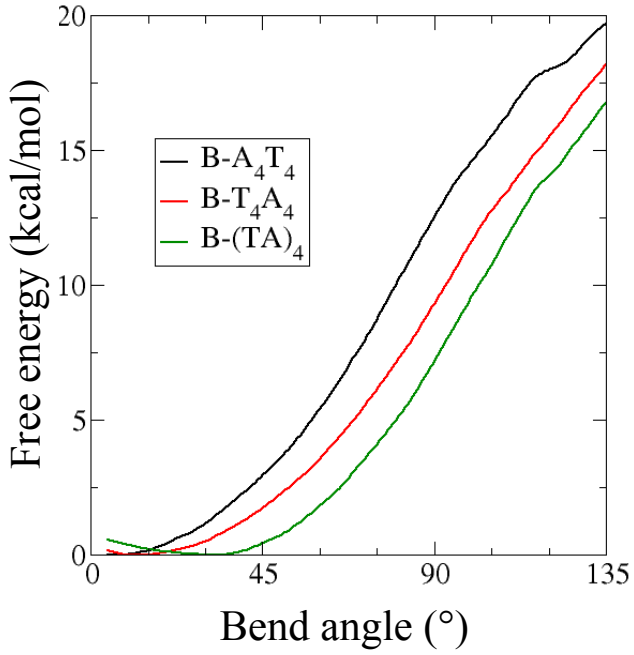


Figure 3.7 A-tract DNA, defined as 4-6 consecutive A-T base pairs, can show sequence-dependent curvature and stiffness. These adenine-rich sequences are often involved in DNA recognition, such as by the TATA-box binding protein, due to their intrinsic mechanical and structural properties [37, 39]. For this reason, we validated our helical bend reaction coordinate by performing umbrella sampling simulations on several A-tract sequences with known properties: A_4T_4 (5'-GCAAAATTTTGC-3'), which is intrinsically bent toward the minor groove, the opposite direction that DNA bending occurs with our constraint; T_4A_4 (5'-GCTTTTAAAAGC-3'), which is intrinsically straight; and $(TA)_4$ (5'-GCTATATATAGC-3'), which is intrinsically bent toward the major groove [21, 24, 52]. By showing that A_4T_4 is least deformable, that T_4A_4 has intermediate deformability, and that $(TA)_4$ is the most deformable, our method qualitatively captures this known trend. A sequence is considered more deformable if it takes less free energy to reach higher bend angles.

3.6.6 Reproducibility of 1D bending free energy profiles

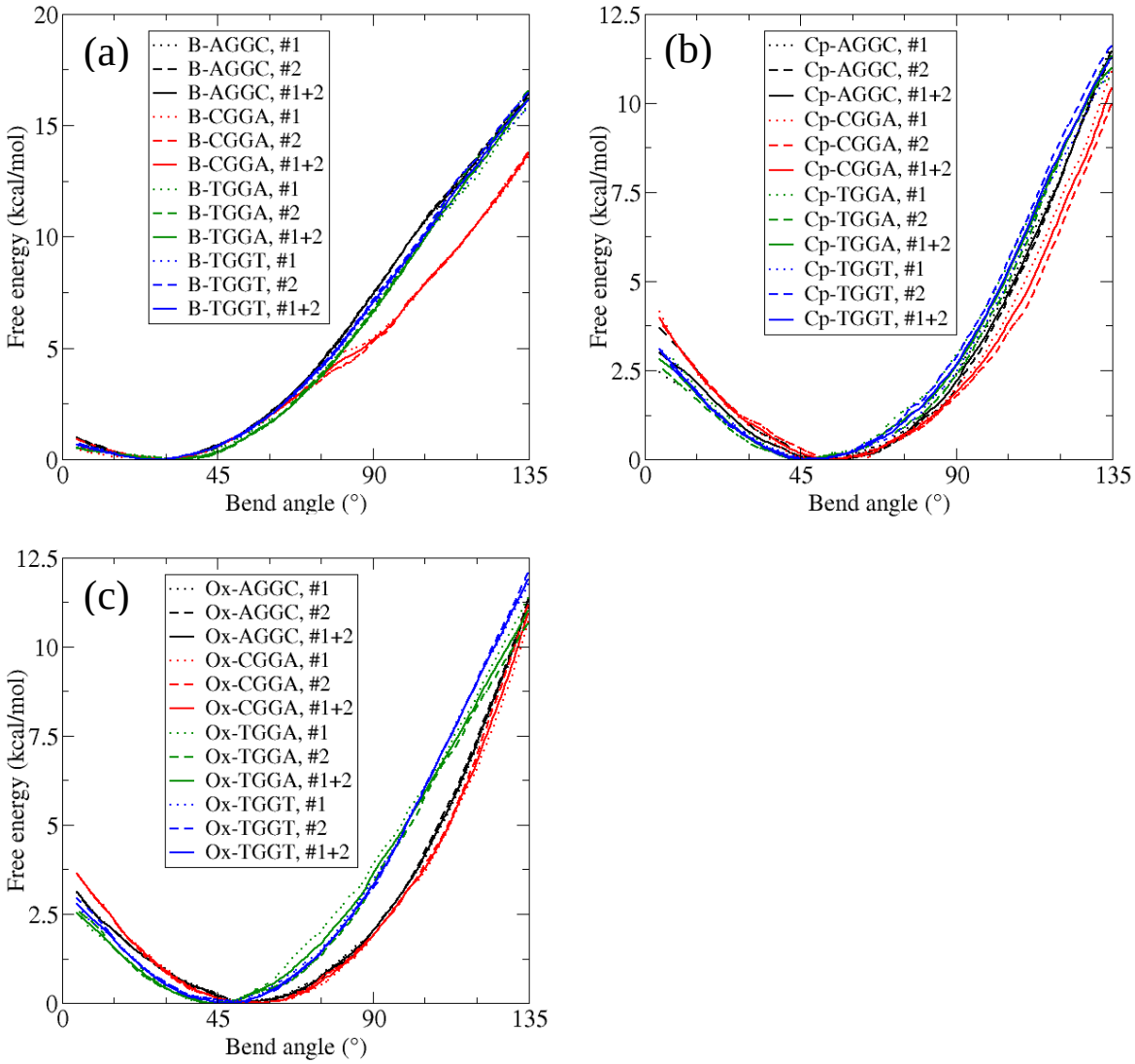


Figure 3.8 1D bending free energy profiles for B-DNA, Cp-DNA, and Ox-DNA in four sequence contexts. The lines denoted by "#1" and "#2" are from two independent sets of umbrella sampling simulations with 10 ns of sampling per window, and the lines denoted by "#1+2" is the free energy profile resulting from using the combined data from #1 and #2 in a WHAM calculation. The two independent trials have converged to yield nearly the same free energy profile.

3.6.7 2D deformability free energy landscapes for cisplatin-DNA adducts in four sequence contexts

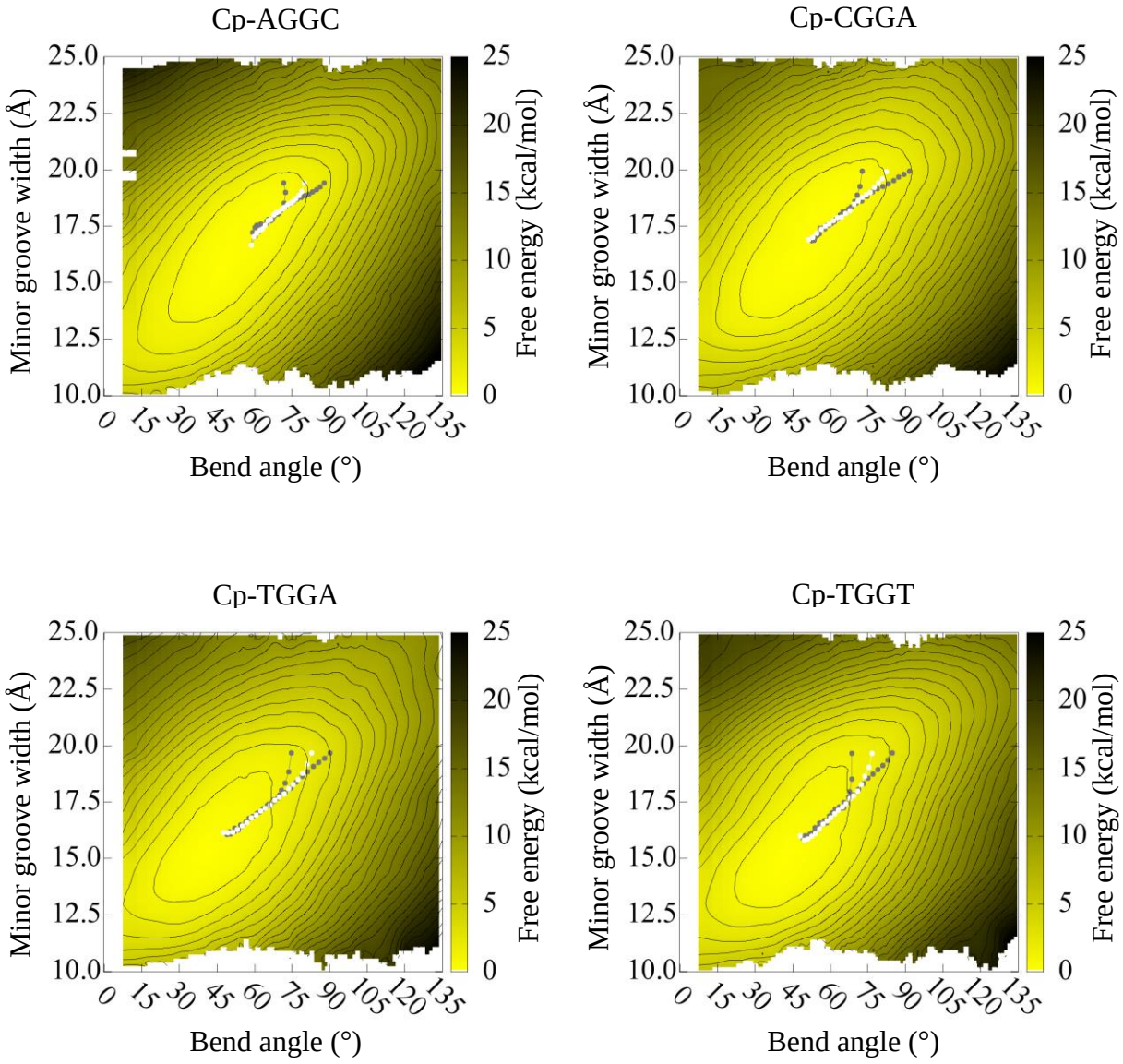


Figure 3.9 2D deformability free energy landscapes for Cp-DNA adducts in four sequence contexts, including the five MFEPs described in the main text, i.e. MFEPs ending at the mean plus-or-minus the standard deviation of the helical bend. Contour lines are separated by 1 kcal/mol. Free energy scale is in kcal/mol.

3.6.8 2D deformability free energy landscapes for oxaliplatin-DNA adducts in four sequence contexts

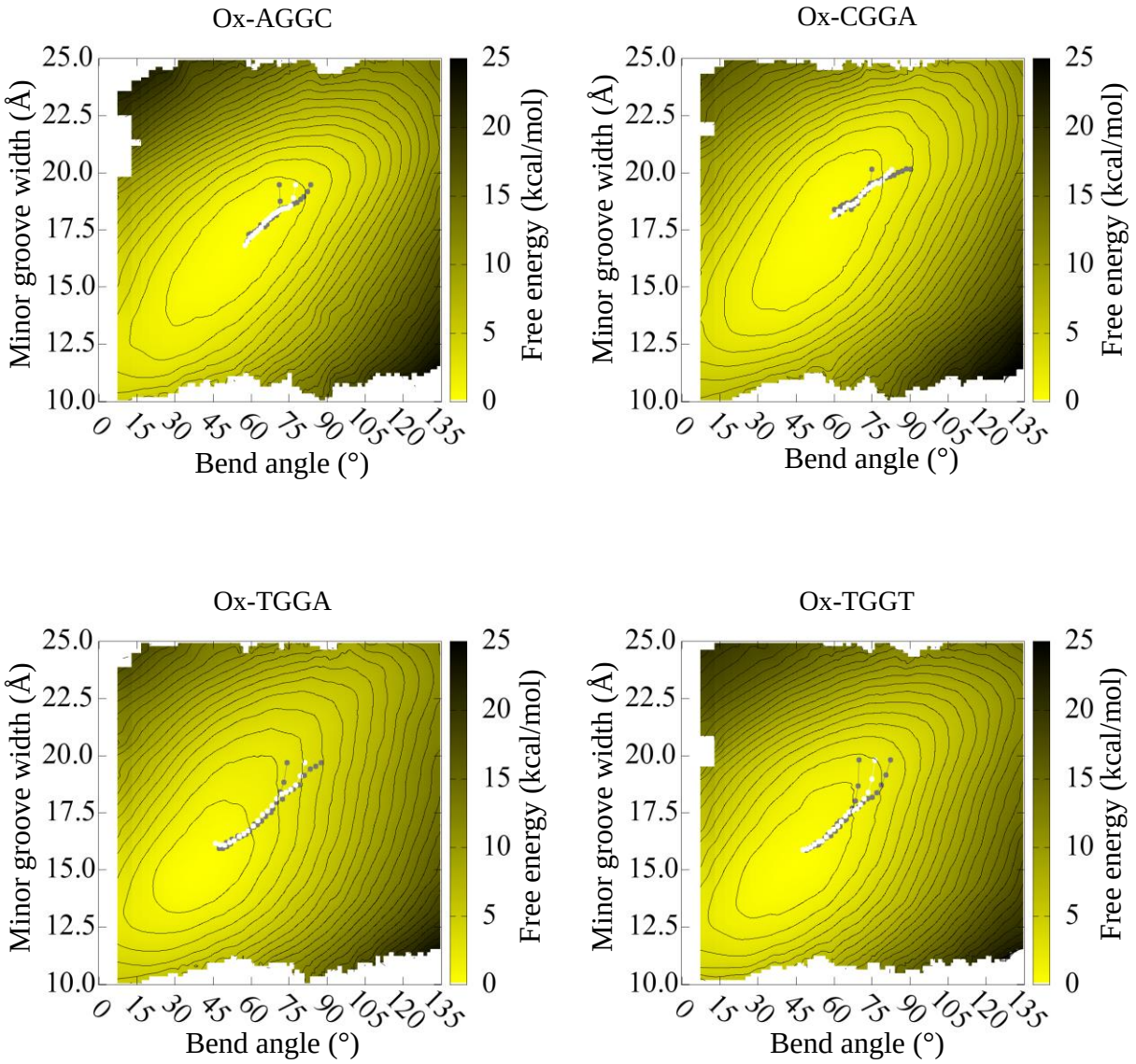


Figure 3.10 2D deformability free energy landscapes for Ox-DNA adducts in four sequence contexts, including the three MFEPs described in the main text, i.e. MFEPs ending at the mean plus-or-minus the standard deviation of the helical bend. Contour lines are separated by 1 kcal/mol. Free energy scale is in kcal/mol.

3.6.9 Reproducibility of 2D deformability free energy landscapes: Ox-TGGA

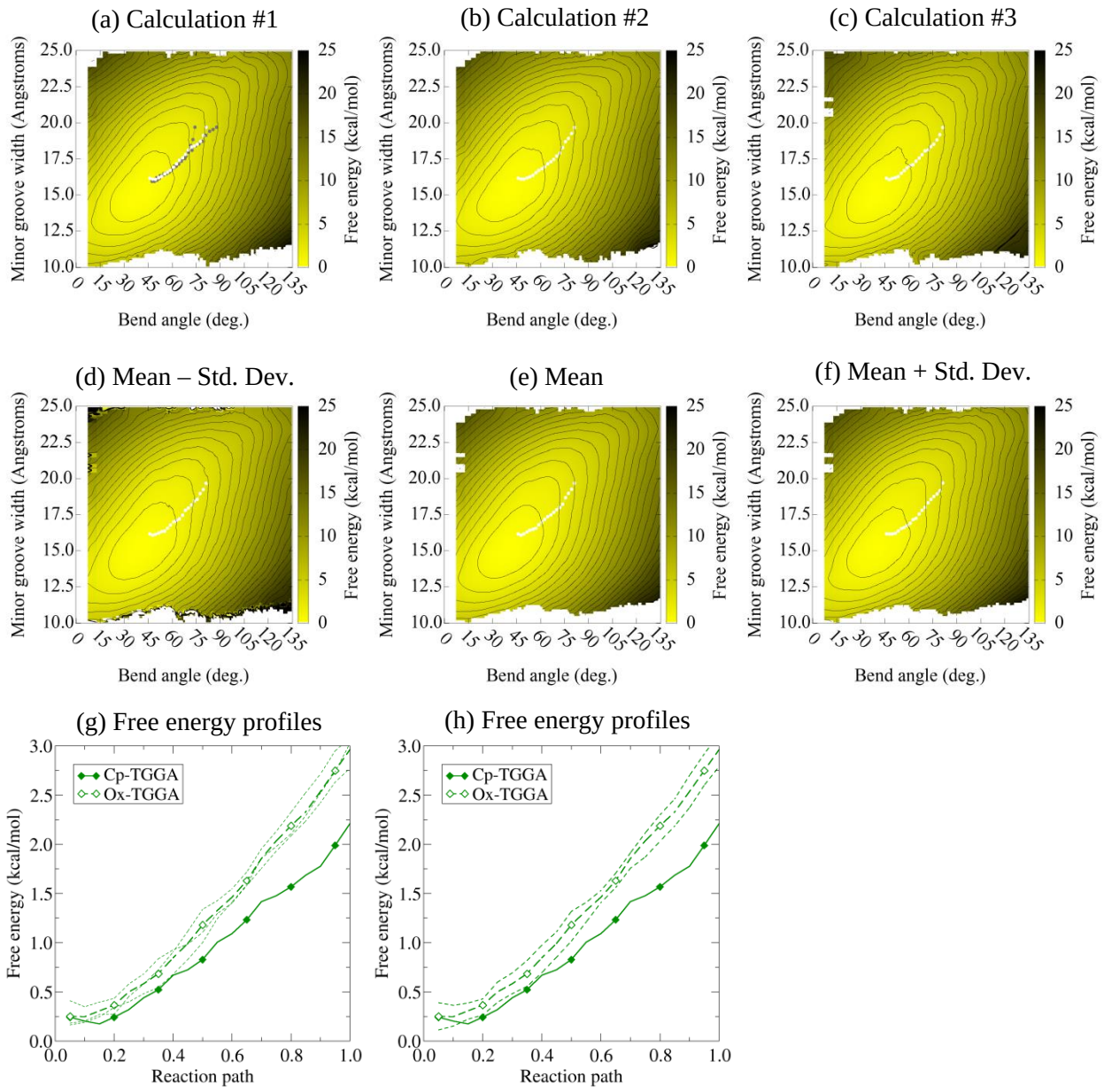


Figure 3.11 The deformation free energy landscapes shown in (a-c) are the result of three independent sets of umbrella sampling simulations conducted on Ox-TGGA, i.e. each helical bend and minor groove width window was simulated three times. Qualitatively, it is clear that these independent calculations have produced very similar free energy landscapes. The middle row of plots (e-f) shows the mean of these three independent calculations (e), and the mean ± the standard deviation of the three trials (d, f). In each case, the MFEP calculated on that surface is shown, demonstrating the similarity of the pathways. Finally, (g) and (h) show the free energy profiles along the MFEPs on the surfaces shown in (a-c, e) and (d-f), respectively, and Cp-TGGA is included for comparison. The low variation between the three independent free energy landscapes, and between the MFEPs calculated on the landscapes, is a good indication that each individual umbrella sampling calculation has a sufficient amount of sampling, even for other drug/sequence combinations. Contour lines are separated by 1 kcal/mol. Free energy scale is in kcal/mol.

3.6.10 Direct linear pathways on 2D free energy landscapes are similar to minimum free energy pathways

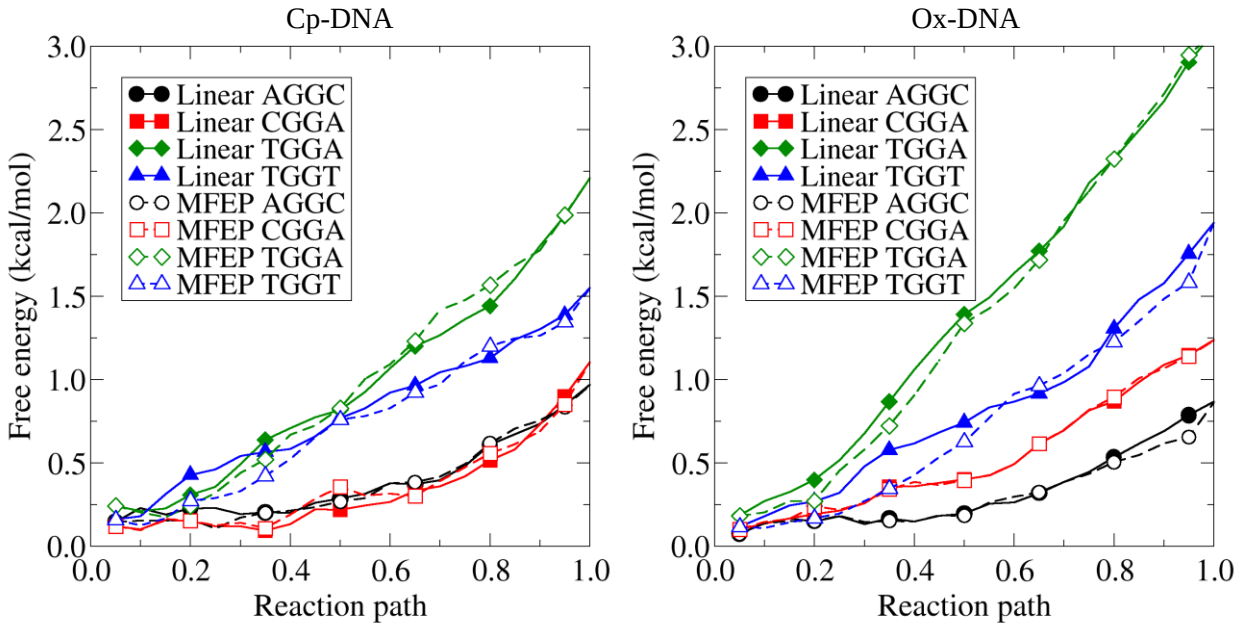


Figure 3.12 Comparison of the free energy profiles along the MFEPs calculated using the finite-temperature string (FTS) method and the free energy profile along a straight line between the initial and final structures. When compared to Figure 3.5, this demonstrates that the free energy profile along the pathway depends more strongly on the structure of the final (i.e. protein-bound) state than on the specific pathway along the free energy surface. This is sensible because the free energy surfaces are relatively smooth (although not perfectly so).

Chapter 4

Understanding the effect of polylysine architecture on DNA binding using molecular dynamics simulations

Adapted from: *Biomacromolecules* **12** 3870 (2011)

4.1 Introduction

Delivery of genetic material using a carrier material (v*ector)* for expression by the host cell is termed DNA *transfection*. Viruses are the most effective DNA transfection vectors because they are highly efficient in penetrating cells and integrating the exogenous gene into the host genome. But viral vectors are plagued with safety concerns, such as immunogenic responses leading to fatalities. Unlike viruses, non-viral vectors such as cationic polymers are non-pathogenic, but are limited by lower transfection efficiency and specificity [1]. Extensive efforts have been devoted to design and synthesize polycationic vectors that exhibit high transfection efficiency [2-5].

Polycations bind to polyanionic DNA to form complexes, termed *polyplexes*, having varying shapes (e.g. rods, toroids, spheres) and sizes (on the order of 20–200 nm) [6]. These polyplexes, having a net positive charge, can interact with negatively-charged proteoglycans in the cell membrane and undergo non-specific or caveolae-mediated endocytosis, which internalize the polyplexes in endosomes [7, 8]. These polyplexes must then escape endosomes to enter the nucleus, where DNA is then released to allow access by transcriptional proteins [7, 8]. Successful transfection requires polyplexes of appropriate size, surface charge, and physical and chemical stability [7, 8]. An optimal polycation would exhibit sufficiently tight binding to the DNA to protect the DNA from degradation by lysosomal hydrolytic enzymes and cytosolic nucleases, while being sufficiently loose to dissociate from the DNA within the

cell to allow transcription [7, 8]. Therefore, to design efficient polycationic vectors one must understand how polycations bind and unbind DNA, and in turn how that affects the characteristics of the polyplex.

Polycation chemistry (e.g. polyethyleneimine [2, 9], polylysine [3]), and architecture (e.g. dendrimers [10], grafted comb structures [4, 5]) along with the N/P ratio (i.e., the ratio of protonatable amines on the polycation to the negatively-charged phosphate groups in the DNA), pH, and ionic strength of the medium affect polycation-DNA binding [1, 7, 8]. Among linear polycations, poly-L-lysine (PLL) has been shown to condense DNA into compact polyplexes [5], but polyplexes formed from linear PLL have a net-positive charge, which causes the polyplexes to aggregate with negatively-charged serum proteins *in vivo* and hinders transfection [8]. Also, linear PLL has been shown to bind either too strongly or too loosely, depending on the medium, which leads to the lowered dissociation of the DNA from the polyplex or increased degradation of the DNA, respectively [4]. Emrick and coworkers recently synthesized polycations that have fixed oligolysine units (from 1 to 5 repeat units) pendent to a hydrophobic polyolefin backbone, revealing a striking effect of polylysine architecture on transfection [4, 5]. They found that graft-type oligolysine architectures [e.g. poly(cyclooctene-*g*-oligolysine)] exhibit significantly higher transfection efficiency than linear poly-L-lysine, despite having identical cationic functionality, and that the transfection efficiency varied non-monotonically with graft length. Here we use systematic molecular simulations to understand the effect of polycation architecture on the structural and energetic features of polycation-DNA binding.

Prior simulations have examined polycation-DNA binding using coarse-grained representations of DNA and polycations [11, 12]. Some atomistic studies of DNA, RNA, or siRNA complexed with various polycations have also been reported, including linear PLL [13, 14], linear PEI [13], spermine [15], and poly(amidoamine) (PAMAM) dendrimers [14]. These investigations quantified the energetics of binding, dissociation of counterions, number of polycation-DNA contacts, and structural changes, such as curvature or changes in groove width [14, 15].

Despite these prior experimental and theoretical reports, a clear relationship of how structure affects function is not yet established. With the aim of establishing such a structure-function relationship,

that can be applied broadly across macromolecular materials, we use atomistic molecular dynamics (MD) simulations to investigate effect of systematic variation in poly-L-lysine architecture [4, 5] on DNA binding. Specifically, we elucidate the effect of polycation architecture on the location and duration of polycation-DNA contacts, counterion displacement, and the free energy of binding. Together, the structural and thermodynamic effects of polycation-DNA binding describe a) the nature of binding (e.g. whether the amine groups of polylysine bind in a cooperative manner or in a competitive manner), and b) the entropic and energetic factors favoring or disfavoring polycation-DNA binding. The key result is that while the binding energy increases linearly with graft length, the entropy of binding displays non-monotonic trends, resulting in the total binding free energy being non-monotonic for the four graft lengths considered here. Although binding free energies have not been measured in experiments, polyplex stability experiments also show these non-monotonic trends [4, 5]. Our results also show that the binding free energy normalized to the number of polycation-DNA contacts generally increases with molecular weight and demonstrates increased cooperative behavior for longer grafts. Structurally, comparing linear poly-L-lysine to the grafted architecture, we observe that the binding mode of the linear polycation is concerted, with adjacent portions of the polycation strongly influencing each other, whereas the oligolysine grafts bind independently of each other. We also find that the hydrophobic backbone of poly(cyclooctene-*g*-oligolysine) restricts the motion of the individual grafts, constraining some grafts to interact with the solution rather than bind to the DNA, thus reducing the total binding free energy.

4.2 Methods

4.2.1 Simulation protocol

We performed atomistic molecular dynamics simulations of systems with an oligomeric polycation, a 16 basepair double stranded B-DNA with the sequence 5'-d(CGCGCGAATTCGCGCG)-3', explicit water molecules, and Na^+ and Cl^- counterions. We used the Amber *parm99* force field, including the *bsc0* and *SB* modifications, to parameterize the B-DNA strands, lysine, TIP3P-model water molecules, and Na^+ and Cl^- counterions and the general Amber force field (gaff) to parameterize the polycations (excluding

lysine) [16-19]. We used *nucleic acid builder* (Amber) to generate canonical B-DNA structures, and *Corina* (Molecular Networks) to generate structures of the poly(cyclooctene) (PCO) backbone of poly(cyclooctene-*g*-oligolysine) (*PolyN*). Supplementary section 4.6.1 provides details of how we used *antechamber* program (Amber) to assign force field parameters and calculate partial charges of PCO (Figure 4.7 and Tables 4.1 and 4.2 in supplementary section 4.6). We retained the charges for lysine in the Amber force field, as these values have been derived rigorously [20]. For each polycation-DNA system, we performed three unrestrained, explicitly solvated molecular dynamics simulation trials of 20 ns each using NAMD 2.7 [21], each trial beginning with different velocities but the same starting coordinates. Initially, the polycation and DNA were placed approximately 20 Å apart, aligned with the coordinate axes so as to decrease the size of the water box. Systems were solvated in a TIP3P water box with a distance of 10 Å between the solutes and the edge of the periodic box; this resulted in the addition of 7,000 to 18,000 water molecules. Na^+ and Cl^- counterions were introduced to neutralize the charge of the system; we attempted to maintain approximately the same total charge density in all systems. Temperature and pressure were controlled using Langevin dynamics. Electrostatic interactions were treated using the particle-mesh Ewald (PME) method. Each system was equilibrated using a standard protocol (see supplementary section 4.6.2). It is important to note the two limitations in the protocol employed. First, the force field parameters for the polycations are taken from a general force field and therefore are not optimized for these specific molecules. However, as the bond types and atom types used (e.g. carbon-hydrogen bonds, carbon-carbon single and double bonds) are common and well-understood, we expect that these parameters are good approximations. Second, the charges are derived with the semi-empirical AM1-BCC method [22]. This method is a parametric charge calculation scheme and hence is less accurate than *ab initio* quantum calculations. Considering that the AM1 method is optimized for organic molecules (especially those containing C, H, O, and N) and that the bond types (relevant for BCC, or *bond charge correction*) are common, this method will function sufficiently.

We have employed the ‘multi-trajectory’ approach in this work [23], wherein the bound and unbound states are simulated separately, in contrast to the ‘single-trajectory’ approach that is most

commonly used [24]. This was necessary to effectively sample the unbound states of polycation and DNA for free energy calculations. The single-trajectory approach is appropriate for systems where the conformations in the bound and unbound states are similar, which is not the case for our systems. Details of the free energy calculations are given in section 4.2. When simulating isolated polycation and DNA, we use the same protocol as above (supplementary section 4.6.2) and maintain the same water box volume and number of ions as used in the polycation-DNA systems.

4.2.2 Systems studied

We have conducted atomistic molecular dynamics (MD) simulations on two lysine-based polycation architectures – linear and grafted – shown in Figure 4.1.

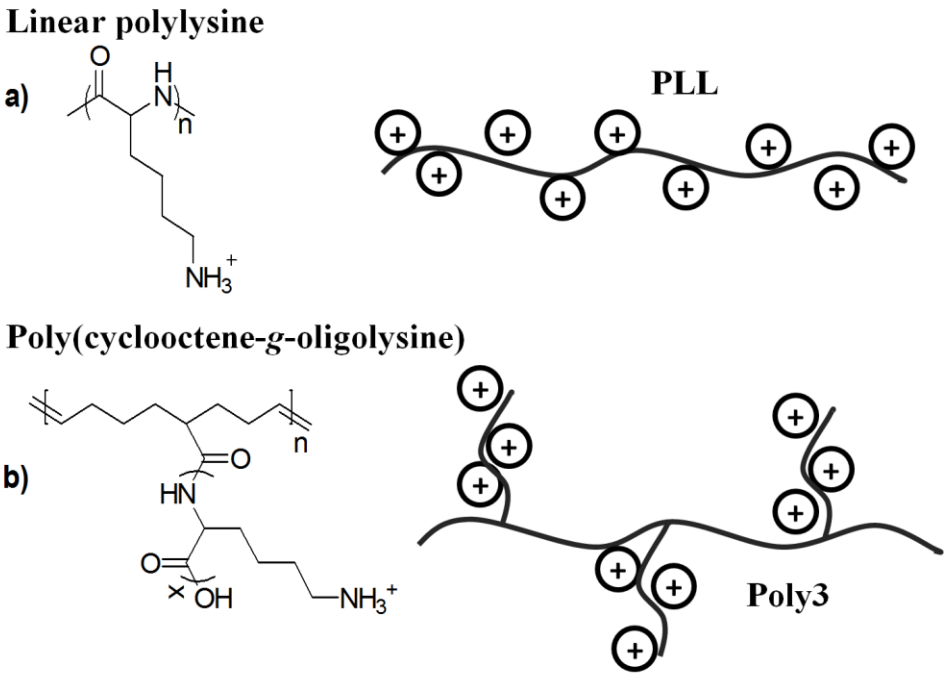


Figure 4.1 Chemical formulae and schematics of the chemistries and architectures in this paper. For Poly2 x=2, Poly3 (shown in the schematic) x=3, Poly4 x=4, and Poly5 x=5.

We have simulated long linear PLL, short linear PLL, and grafted PLL with varying numbers of lysine residues (8, 12, 16, and 20). For long linear PLL, all of the lysine residues are in a single polycation (1xPLLN). For short linear PLL, the lysine residues are divided into four oligomers (4xPLLN). In the grafted architecture, the lysine residues are divided among the 4 grafts (PolyNx4) placed on a

hydrophobic backbone. We have also simulated the grafted architecture with 8 repeat units (PolyNx8). Each of these systems includes a 16 base pair DNA oligomer with the sequence 5'-d(CGCGAATTCGCG)-3', and counterions to neutralize the charge of both polyelectrolytes. Table 4.3 (supplementary section 4.6) lists the details of all the systems studied here. We note that in these simulations, the N/P ratio ranges from 0.3 to 1.3. In experiments, the N/P ratios range from 1 to 5 or higher. In order to increase the N/P ratio in our simulations to match experimental values, we will need larger polycations or a higher number of short polycations, both of which require a larger simulation box and in turn a larger number of water molecules, making it unfeasible computationally. Therefore, we connect the simulated N/P ratios to experimentally relevant values, while staying within the bounds of what is computationally feasible (e.g. up to N/P of 1.5), by elucidating how binding free energy changes upon increasing N/P ratio.

4.2.3 Structural analysis

We observed several structural effects on the DNA as a result of polycation binding. To quantify these changes, we calculated the quantity and duration of contact sites between DNA and polycation, the curvature of the DNA backbone, and determined the change in the spatial distribution of counterions. Atomic contacts between polycation and DNA were measured using in-house scripts in VMD 1.8.7 [25]. These scripts calculate the number of polycation atoms within two distinct spatial regions: within 9 Å of the center of mass of each basepair (quantity G) and within 5 Å of the phosphate groups of each basepair (quantity B). These regions qualitatively define interaction of the polycation with the DNA grooves and backbone, respectively. The primary location of interaction between the polycation and each basepair is quantified by the 'binding density': the ratio of the polycation-backbone interactions to the total number of polycation-DNA interactions, i.e. B / (B + G). The density can range from 0 (interactions solely with grooves) to 1 (interactions solely with the DNA backbone).We represent the binding density as heat maps, which show the location and duration of polycation-DNA interactions: blue regions represent interaction with the grooves, yellow regions indicate interactions with the backbone, and white regions indicate no interactions with that basepair (i.e. both B and G are zero).

4.2.4 Free energy calculations

We calculated the polycation-DNA binding free energy (ΔG_{bind}) using the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) approach [24]. This methodology uses snapshots from an MD trajectory to calculate the average interaction energies based on the solute's internal energy calculated by molecular mechanics, and the solvation energy calculated using the Poisson-Boltzmann (PB) continuum solvent approach. The overall binding energy is the difference in energy between the polycation-DNA complex and the individual polycation and DNA:

$$\Delta G_{bind} = \Delta G_{complex} - (\Delta G_{polycation} + \Delta G_{DNA}) \quad (1)$$

where

$$\Delta G = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \quad (2)$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdW} \quad (3)$$

$$\Delta E_{internal} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{torsion} \quad (4)$$

$$\Delta G_{solv} = \Delta G_{nonpolar} + \Delta G_{PB} \quad (5)$$

ΔE_{MM} is the change in molecular mechanics energy, which consists of internal energy (bonds, angles, dihedrals), electrostatic, and van der Waals contributions. ΔG_{solv} is the solvation energy, which consists of nonpolar and electrostatic terms. Solvation energy calculations are performed at 150 mM ionic strength of monovalent counterions to mimic the effect of charge screening at physiological conditions. Using a consistent ionic strength for this calculation allows comparison between systems with different simulated ionic strength; the ionic strength of our systems ranges from 160 mM to 230 mM. The electrostatic contribution to the solvation energy, ΔG_{PB}, is calculated using the Poisson-Boltzmann (PB) method. The nonpolar contribution to solvation, $\Delta G_{nonpolar}$, is calculated by an empirical linear relationship using the solvent-accessible-volume (SAV):

$$G_{nonpolar} = p \cdot SAV + c \quad (6)$$

where the pressure coefficient p = 0.00378 kcal/mol-Å^3 and the constant c = -0.5692 kcal/mol [26]. Following the recommendations of Luo and coworkers for the use of their σ-decomposition scheme, we

use a surface probe radius of 0.557 Å and an effective water density of 1.129 times the bulk density of water [26]. We note that the solvation terms are actually *free energies*, in contrast with ΔE_{MM} and its constituent terms, which are simply energies, although we denote the sum of all the molecular mechanics and solvation energies with the term ΔE for convenience. $\Delta G_{nonpolar}$, ΔG_{PB}, and $\Delta E_{electrostatic}$ are calculated with the *pbsa* program in the Amber suite, while the other terms (i.e. those contributing to ΔE_{MM}) were calculated with the *sander* program in the Amber suite.

One of the entropic contributions to the binding energy is estimated using a harmonic-approximation normal mode analysis (NMA) using *nmode* (Amber). For these calculations, we minimize the structures of 20 snapshots from each trajectory to below a root-mean-square (RMS) gradient of 10^{-3}. By performing tests with the more stringent RMS gradients of 10^{-4}, 10^{-5} and 10^{-6}, we confirmed that the NMA entropy has already essentially converged at an RMS gradient of 10^{-3}. Normal mode analysis neglects two important contributions to the total entropy: translational entropy gain of counterions that dissociate during polycation-DNA binding, and conformational entropy loss of the polymers. The translational entropy contribution from freed counterions has been shown to drive the DNA-polycation condensation process [27, 28], and conformational entropy changes are known to drive some protein-binding events [29, 30]. Therefore, a normal mode analysis alone would provide a poor estimate of binding entropy.

To improve the entropy estimate, we also calculate the translational entropy gain of counterions and the conformational entropy change in the macromolecules (DNA and polycation). To approximate the translational entropy gain of counterions, we assume that the entropic contributions of counterions bound to either polyelectrolyte are negligible, and that the entropy only depends on the mobile counterions [31]:

$$\frac{-S}{k_B} = (1-\alpha_+)n_+ \ln[(1-\alpha_+)\phi_+] + (1-\alpha_-)n_- \ln[(1-\alpha_-)\phi_-] \qquad (7)$$

where S is the time-dependent entropy, k_B is the Boltzmann constant, α_i is the time-dependent fraction of bound ions, n_i is the total n`umber of ions, ϕ_i is the volume fraction of the counterions, and the subscript i refers to positive (Na^+) and negative (Cl^-) counterions [31]. The volume fraction, ϕ_i, is defined by

$$\phi_i = \frac{\left(\frac{4}{3}\pi R_{vdW,i}^3\right) n_i}{V_{box}} \tag{8}$$

where $R_{vdW,i}$ is the van der Waals radius of the counterion, as defined in the Amber force field, and V_{box} is the volume of the water box. Counterions are defined to be bound to the oppositely charged polymer species (Na^+ to DNA and Cl^- to polycation) if they are within the Bjerrum length of the charged groups of DNA or polycation. The Bjerrum length is the ratio of electrostatic attraction between unit charges and thermal energy, and in water at 300 K the Bjerrum length is approximately 7 Å. Each system has equilibrated after 4 ns, and the remaining 16ns of production simulation are used to calculate the average and standard error of the entropy. This method has been used to accurately calculate the entropy of ion dissociation during the binding of oppositely charged polyelectrolytes [31].

To approximate the conformational entropy loss of polycation-DNA binding, we use a recently-developed method that estimates the conformational entropy from the flexibility of dihedral angles [32]. The entropy of the k^{th} dihedral angle is given by:

$$\frac{-S_k}{k_B} = \int_0^{2\pi} P_k(q) \ln\left(P_k(q)\right) dq \tag{9}$$

where $P_k(q)$ is the von Mises kernel density estimate of the probability distribution of the k^{th} dihedral angle. The von Mises distribution function has the following form:

$$f(q, \mu, \kappa) = \frac{e^{\kappa \cos(q-\mu)}}{2\pi I_0(\kappa)} \tag{10}$$

where I_0 is the modified Bessel function of the first kind of order 0. The von Mises distribution is the circular (i.e. periodic) equivalent of the Gaussian function and is described by a mean value, μ, and a concentration parameter, κ, which is analogous to σ^{-2} of the Gaussian distribution. As κ increases, the density estimate generally becomes a better description of the actual distribution. After some calibration, we found that the entropy values converge around κ = 225 (see details in Table 4.4 in supplementary section 4.6), which we use in this work. The von Mises kernel density estimate, $P_k(q)$, averages multiple distributions to describe the actual density of each dihedral angle *k*:

$$P_k(q) = \frac{1}{N}\sum_{j=1}^{N} f(q, q_{k,j}, \kappa) \quad (11)$$

where N is the number of simulation snapshots and $q_{k,j}$ is the value of the dihedral angle k in snapshot j. The density estimate has greater magnitude at values that are more common in the dihedral angle distribution and lesser magnitude at values that are less frequently sampled by that dihedral angle. Simply put, the density estimate is a smooth representation of the dihedral angle distribution. Finally, the sum over all K dihedral angles yields the total entropy, S:

$$S = \sum_{k=1}^{K} (S_k - S_0) \quad (12)$$

where S_k is the conformational entropy of dihedral *k*, and the quantity S_0 is obtained when $P_k(q)$ is the von Mises kernel function, rather than the density estimate, with any value for μ (e.g. π, which we use here) and the same value for κ as in the calculation of S_k. This method has been used to successfully quantify the conformational entropy of proteins such as dialanine and the Villin Head Piece and to relate protein side chain entropy to NMR order parameters [32, 33].

As mentioned above, the 'single-trajectory' approach is typically used when calculating binding energy: a single simulation of the bound state is performed, and the coordinates of ligand, receptor, and complex are drawn from that simulation. This method is appropriate for systems where the conformations of the bound and unbound states are similar, such as when screening small drug molecules that do not significantly affect protein structure. In our systems, the bound and unbound states are significantly different: there are different numbers of bound counterions, and the conformations available to both polyelectrolytes are substantially different. For this reason, we chose to use the 'multi-trajectory' approach, wherein 3 simulations are run for each polycation-DNA system: one each for the DNA, polycation, and DNA-polycation complex. This allows us to *more accurately* calculate the ion dissociation and conformational entropy changes and hence the total entropic contribution to the free energy of binding. When performing DNA-only or polycation-only simulations, we kept the water box

volume and number of ions equal to the bound state simulations. We found that the ion dissociation entropy is insensitive to the volume, so the slight volume differences between the bound and unbound states, due to operating in the NPT ensemble, are of no concern.

Combining all of the above energetic and entropic contributions to the binding free energy provides a more complete description of the thermodynamics of binding. However, the calculated values of ΔG_{bind} will not be accurate quantitatively. Rather, our goal is to compare the qualitative trends in energy, entropy, and free energy of binding between different polycations.

4.3 Results

4.3.1 Effect of graft length on grafted oligolysine-DNA binding

In Figure 4.2, we present the binding modes from one simulation trial of 16 base pair (bp) double stranded dsDNA binding to 4 repeat units (n=4 in Figure 4.1) of grafted oligolysine for graft lengths of 2, 3, 4 and 5, represented as Poly2x4, Poly3x4, Poly4x4, Poly5x4, respectively.

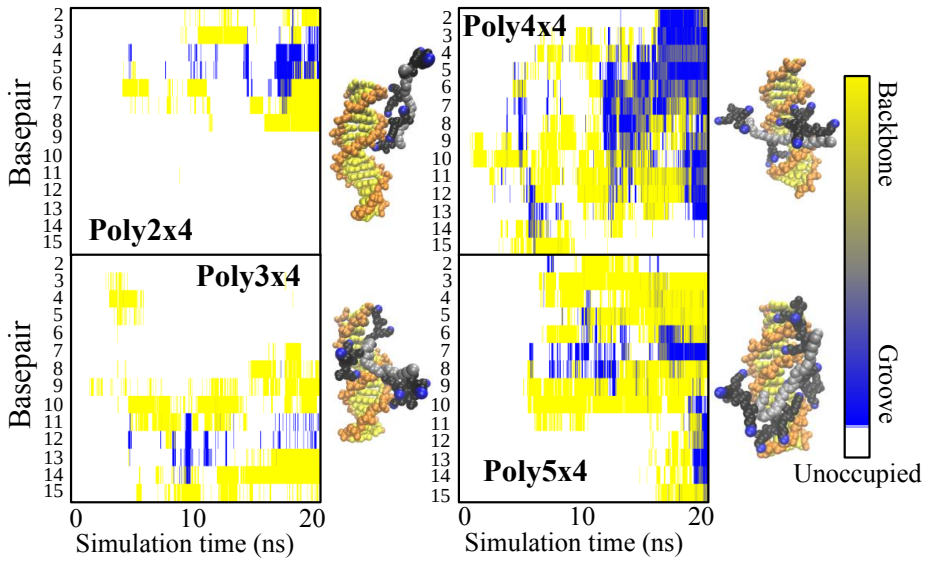


Figure 4.2 Binding maps from one simulation trial of 4-mer long poly(cyclooctene-*g*-oligolysine) with graft length 2 (Poly2x4), 3 (Poly3x4), 4 (Poly4x4), and 5 (Poly5x4) binding to 16 basepair dsDNA. Simulation snapshots of the DNA bound to the polycation are shown (water molecules and counterions are hidden for clarity).

We also present a single simulation snapshot of each polycation-DNA complex, with water molecules and counterions hidden for clarity. Figure 4.8 (supplementary section 4.6) shows the binding maps for all

three trials of Poly2x4, Poly3x4, Poly4x4, and Poly5x4. The binding maps show that the grafted oligolysine binds to the charged backbone (yellow regions) more than the uncharged minor and major grooves (blue regions), which is expected since the electrostatic interactions drive complex formation. We observe that as graft length increases from Poly2 to Poly5, the binding maps become more populated, suggesting that a greater number of contacts are formed, both on the backbone and within the grooves. As graft length increases, we observe a longer duration of contacts between the grafted oligolysines and the DNA backbone. In addition to the above binding mode maps, qualitative observation of the simulation trajectories suggest that grafted oligolysines do not display *one* specific mode of binding. Instead, each graft binds to the DNA independently of the other grafts, while being constrained to the grafting position along the flexible hydrophobic backbone. The independence of the grafts with the backbone constraint results in a *variety* of binding modes, such as compact blob-like structures of the polycation with the polycation wrapping around the DNA in case of the longer grafts, or extended conformations of the polycation where the grafts are stretching to reach the DNA in case of the shortest grafts. We also observe that in these modes some grafts are bound to DNA but other grafts are essentially free in solution, which affect the binding energies.

Figure 4.3 shows various relevant contributions to the total ΔE, $T\Delta S$ and ΔG as a function of the graft length. Note that Tables 4.6 and 4.7 (supplementary section 4.6) present all the contributions to the ΔE, $T\Delta S$ and ΔG as a function of the graft length for these 4-mer grafted oligolysines.

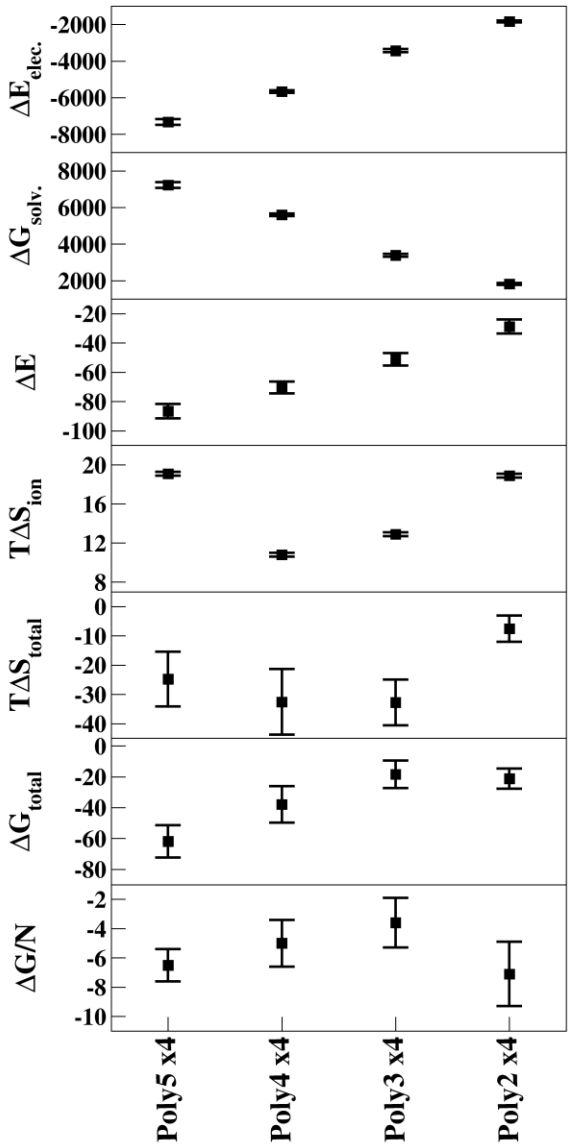


Figure 4.3 Contributions to the binding free energy for *four* repeat units of grafted architecture, PolyNx4, with varying graft length (N = 5 to 2) binding to 16bp DNA. Units are kcal/mol.

With increasing graft length from Poly2 to Poly5, the electrostatic contribution, ΔE_{elec}, increasingly favors DNA binding, as there are more positively-charged lysines that can bind to the phosphate backbone. Tables 4.6 and 4.7 (supplementary section 4.6) show the number of amine nitrogens *bound* to the DNA backbone, confirming that as graft length (the total number of amines) increases the number of amine nitrogens bound to DNA also increases. As graft length increases, the solvation free energy, ΔG_{solv}, decreases, and is positive in all cases (unfavorable for binding). The ΔG_{solv} consists of nonpolar and polar

contributions. The nonpolar contributions, ΔG_{cavity} and $\Delta G_{dispersion}$, depend non-monotonically on graft length (see Table 4.6 in supplementary section 4.6): Poly3x4, Poly4x4 and Poly5x4 yield approximately the same values of ΔG_{cavity} and $\Delta G_{dispersion}$, but the values of Poly2x4 are smaller in magnitude. The ΔG_{cavity}, which is the energy required to form a cavity in water large enough to enclose the solute, for all PolyNx4 is negative, indicating that the volume of the complex is less than the volume of the separate unbound constituents. Evidently, Poly3x4, Poly4x4 and Poly5x4 undergo a more significant decrease in volume upon binding than Poly2x4 does, suggesting that Poly2x4 binds relatively weakly and maintains greater mobility. The other nonpolar contribution, $\Delta G_{dispersion}$, is a measure of the favorable interactions between solute and solvent. The $\Delta G_{dispersion}$ for all PolyNx4 is positive, indicating that the solvent-accessible surface area decreases on binding, decreasing the area for favorable solvent-solute interactions. Again, the value of $\Delta G_{dispersion}$ is lower for Poly2x4 than for Poly3x4, Poly4x4 and Poly5x4, with the latter three having approximately the same value; this trend also suggests that Poly2x4 binds to DNA less tightly as compared to Poly3x4, Poly4x4 and Poly5x4. The polar contribution, ΔG_{PB}, increases with graft length from 2 to 5, because the number of charged groups increases with graft length. Due to the large magnitude of the polar contribution, ΔG_{PB} dominates the ΔG_{solv} term. The total ΔE term, which is the sum of all energetic contributions, becomes increasingly favorable for binding as graft length increases (Figure 4.3), and the ΔE term scales linearly with graft length (Figure 4.10 in supplementary section 4.6). We note that the ΔE term is missing the energetic contributions coming from explicit ions dissociating from DNA upon DNA-polycation complexation, although this is partially accounted for by the inclusion of the effects of ionic strength in the calculations.

The entropic contribution to the total binding free energy, $T\Delta S_{total}$, is assumed to be the sum of the terms from normal mode analysis, ion dissociation, and conformational entropy. The entropy values from normal mode analysis (translational, rotational and vibrational) are negative and approximately the same for all graft lengths suggesting restricted translational, rotational and vibrational motion upon binding for all graft lengths. Poly2x4 once again demonstrates the lowest magnitude for these normal mode analysis

terms (Table 4.6 in supplementary section 4.6), suggesting weaker binding to the DNA and hence less restricted motion than Poly3x4, Poly4x4 and Poly5x4. The ion dissociation entropy (Figure 4.3), $T\Delta S_{ion}$ is positive for all graft lengths, since counterions are displaced from the DNA backbone and gain translational entropy upon polycation-DNA complexation. $T\Delta S_{ion}$ also has a non-monotonic trend with increasing graft length, with Poly2x4 and Poly5x4 having the largest contributions and Poly3x4 and Poly4x4 having the smallest contributions. A large, positive $T\Delta S_{ion}$ is due to a larger number of ions being displaced during complexation and a larger free volume being available to these dissociated ions. Tables 4.8 and 4.9 (supplementary section 4.6) provide the ensemble average number and fraction of bound ions in Poly2-5x4 systems before and after complexation. The increasing number of charged amines with increasing graft length, increasing size of the PolyN molecule with increasing graft length, and the non-trivial trend in binding strength with increasing graft length each contribute to the non-monotonic trend in the difference in bound ions caused by complexation, and in turn the $T\Delta S_{ion}$ term. The entropic contribution of the solvent is included in the solvation binding free energy, ΔG_{solv}. The conformational entropy change upon polycation-DNA complex formation is generally negative and shows a non-monotonic trend with graft length (Table 4.6 in supplementary section 4.6). Conformational entropy decreases on binding because both polyelectrolytes are severely restricted in the bound state. Even though the conformational entropy of each simulation trial converges to a stable value within ~5 ns for all systems, there are large error bars on the average conformational entropy values (mean conformational entropy of many trials), because the polycation binds in a different conformation in each trial, leading to different conformational entropy in each trial. Summing these entropic contributions (normal mode analysis, ion dissociation, and conformational entropy) reveals that total entropy of binding is unfavorable.

These individual energy and entropy terms added together dictate the total free energy of binding, ΔG_{total}, as a function of graft length. We observe that the value of ΔG_{total} is negative, suggesting that complex formation is favored for all graft lengths. The ΔG_{total} is most favorable for Poly5x4 and least

favorable for Poly2x4 and Poly3x4. The non-monotonicity in ΔG_{total} with graft length stems from the non-monotonicity in the entropic terms. If we normalize the free energy of binding to the number of protonated amines bound to the DNA (an amine is defined as bound when it is within the Bjerrum length (7 Å) of a phosphate group) the mean free energy of binding per bound nitrogen is most favorable for Poly5x4 and Poly2x4 and least favorable for Poly3x4 and Poly4x4, albeit with large standard error.

In summary, structurally, each graft in the grafted oligolysines binds to the DNA independently of the other grafts, while being constrained to the grafting position along the flexible hydrophobic backbone. As the graft length increases, the contacts between the grafted oligolysines and the DNA backbone are longer in duration. Energetically, with increasing graft length, ΔE_{elec} increasingly favors DNA binding, solvation free energy, ΔG_{solv}, becomes less unfavorable for binding, and the total ΔE term becomes increasingly favorable for binding. Also, the ΔE term scales linearly with graft length. Entropically, values from normal mode analysis (translational, rotational and vibrational) are equally unfavorable for all graft lengths, ion dissociation entropy is favorable and varies non-monotonically with increasing graft length, and conformational entropy change is unfavorable and varies non-monotonically with graft length. In all, ΔG_{total} is most favorable for Poly5x4 and least favorable for Poly2x4 and Poly3x4.

4.3.2 Effect of increasing molecular weight on grafted oligolysine-DNA binding

To investigate the effect of repeat units on DNA binding, we next present results for 8-mers of grafted oligolysines (n = 8 in Figure 4.1) binding to 16bp DNA. This allows a simultaneous examination of the effect of increasing polycation molecular weight with increasing N/P ratio. Figure 4.4 (analogous to Figure 4.3) shows relevant contributions to the total ΔE, $T\Delta S$, and ΔG as a function of the graft length when 8-mer polycations bind to 16bp DNA.

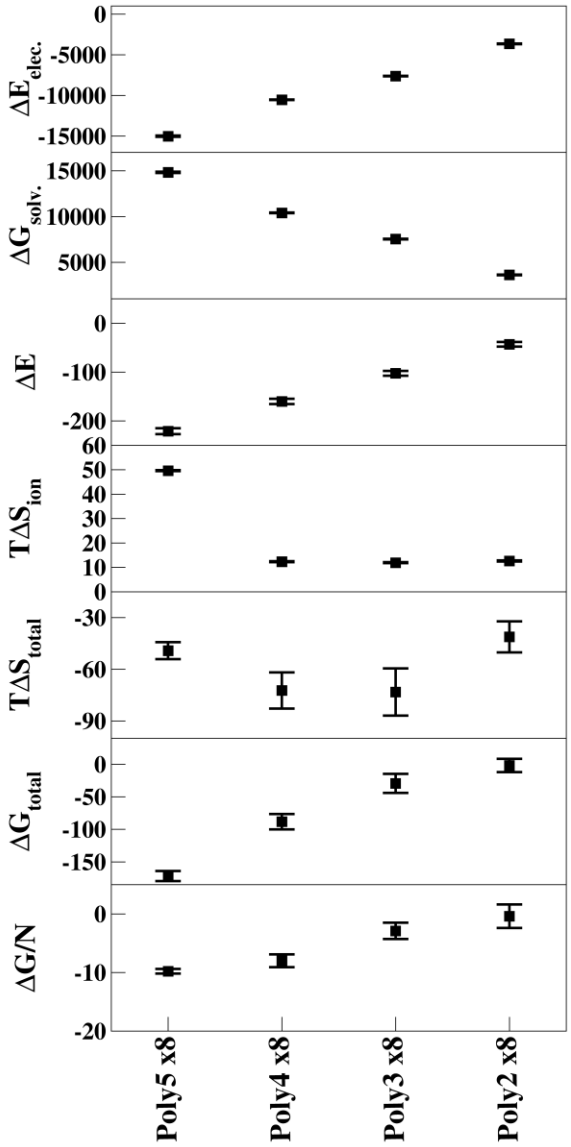


Figure 4.4 Contributions to the binding free energy for *eight* repeat units of grafted architecture, PolyNx8, with varying graft length (N = 5 to 2) binding to 16bp DNA. Units are kcal/mol.

For 8-mers (Figure 4.4), the ΔE terms ($\Delta E_{electrostatic}$ and ΔE_{total}) become more favorable for DNA binding with increasing graft length, and the total ΔE scales linearly with graft length, which is similar to the trends for 4-mers (Figure 4.3). However, the slope of the linear scaling between ΔE and graft length is higher for 8-mers (~7.4) than 4-mers (~4.8) (see Figure 4.10 in supplementary section 4.6), suggesting a cooperative effect in the longer polycations that increases with graft length. The slope of the linear scaling

between number of nitrogens bound and graft length is higher for 8-mers (~3.8) than 4-mers (~2.2), further confirming the cooperative effect associated with longer grafts.

Interestingly, while the trends in ΔG_{total} with increasing graft length are the same for 4-mers and 8-mers (Figure 4.3 and Figure 4.4, respectively), ΔG_{total} for the same structures differs significantly with repeat units. For Poly5, ΔG_{total} for 8-mers is approximately 2.8 times that of the 4-mers, for Poly4 2.3 times, and for Poly3 approximately 1.6 times. Surprisingly, Poly2 has a *lower* ΔG_{total} for 8-mers as compared to 4-mers. Since ΔE increases with the number of repeat units for all graft lengths, this unexpected result for Poly2 is likely entropic in origin. The $T\Delta S$ from the normal mode analysis increases in magnitude with the number of repeat units (Table 4.7 in supplementary section 4.6), by approximately the same amount for all graft lengths. The $T\Delta S_{ion}$ displays interesting trends when comparing 4-mers to 8-mers, decreasing for Poly2, remaining the same for Poly3 and Poly4, and increasing for Poly5. The trends are non-trivial, and best understood by calculating the fraction of bound and unbound ions before and after complexation. For all PolyN, as expected, before complexation the fraction of Cl^- ions bound to the polycation is higher for 8-mers than 4-mers (Tables 4.8 and 4.9 in supplementary section 4.6). After complexation, the fraction of ions (Na^+ and Cl^-) bound to the complex is slightly higher for 8-mers than 4-mers for Poly2 and Poly3, approximately equal for 8-mers and 4-mers for Poly4, and lower for 8-mers than 4-mers for Poly5. To examine whether these non-trivial trends in bound and unbound ions arise from the structure of the macromolecules (DNA and polycation) upon complexation, we calculated the curvature of DNA upon polycation complexation. Figures 4.11c and 4.11d (supplementary section 4.6) show that the 16bp DNA is essentially unbent in most cases, and maintains helical axis curvature of approximately 25°. Note that in the polyplex of a full-sized polycation with plasmid DNA (>1000 base pairs), the nucleic acid must have significant curvature, while in these 16bp segments the absence of curvature at this local length scale is not surprising. We hypothesized that bending the DNA might allow the polycation to bind to a greater portion of the DNA and thereby dissociate more counterions than would be possible without curvature. There are several instances in which PolyNx8 bends the DNA to

much higher curvatures than PolyNx4; however, since Poly2x8 introduces significant curvature in one trial, but the number of ions for this trial is not significantly different from the other two trials in which the DNA is essentially unbent, we do not think the introduction of local curvature significantly affects the dissociation of ions. Therefore, it remains unclear why Poly2 has lower ΔG_{total} for 8-mers as compared to 4-mers.Nonetheless, these studies confirm that the binding free energy ΔG_{total} becomes increasingly favorable (more negative) with increasing molecular weight or N/P ratio for graft lengths 5, 4 and 3. The increase in magnitude of favorable ΔG_{total} with increasing molecular weight or N/P ratio is larger as graft length increases.

4.3.3 Graft vs. Linear Architecture

Since experiments using lysine-containing polymers show that grafted architectures are better transfection agents than linear architectures, here we directly compare the structures at constant number of charged amines [4, 5]. Specifically, we compare DNA binding simulations of the grafted structures to that for a) one long linear polylysine of length *M* (denoted as 1xPLLM) with the same number of lysines as in 4-mers of grafted oligolysine PolyNx4, and b) 4 short linear lysines of length *N (*denoted as 4xPLLN). For example, we compare the Poly4x4 system to a 16mer linear lysine (1xPLL16) and 4 molecules of linear 4-mers (4xPLL4) lysine.

Structural information from binding maps (Figure 4.12 in supplementary section 4.6) and simulations show that the binding modes of the grafted oligolysine and linear polylysine systems are qualitatively different. Short linear polylysines exhibit mobility, and drift freely from grooves to backbone. Almost all parts of the DNA remain in contact with some portion of the short linear polylysines, and there are instances of long-lived (5-10 ns) contacts while the individual chains compete for favorable contacts. In contrast, long linear polylysine, with less conformational freedom, binds irreversibly, and in some cases for the entire duration of the simulation (20 ns), although individual contacts fluctuate on sub-nanosecond timescales. The two possible binding modes for long linear PLL are 1) a ‘zipping’ or axial pattern of binding, in which the polycation binds in parallel with the DNA helical

axis, and 2) a ‘winding’ or radial pattern, in which the polycation binds between the DNA backbones and wraps around the DNA through one of the grooves. In contrast, the graft architecture of PolyNx4 tends towards reversible binding, with the individual oligolysine grafts weakly bound and moving with relative freedom. As noted in section 4.3.1, as the graft length increases we observe each polycation-DNA contact to be longer in duration. Moreover, in contrast to short linear oligolysines, the grafted oligolysines are constrained by the hydrophobic backbone, forcing some of the grafts to face away from the DNA and remain unbound throughout the simulation.

The energetic and entropic contributions to the binding free energy for the grafted and linear architectures are shown in Figure 4.5 (and Table 4.10 in supplementary section 4.6). Considering systems with equal total number of charged amines, long and short linear polylysines have similar ΔE_{elec}, and the corresponding grafted architecture has a substantially lower magnitude electrostatic contribution to ΔE_{elec}. The less favorable $\Delta E_{elec.}$ for the grafted architecture is attributed to the hydrophobic backbone sterically hindering polycation-DNA electrostatic interactions. This is confirmed qualitatively through observation of the simulation trajectories, which show some of the oligolysine grafts are not bound to DNA, but are instead constrained by the backbone away from the DNA, and into solution. This constraint imposed by the backbone reduces the total ΔE of the grafted architecture.

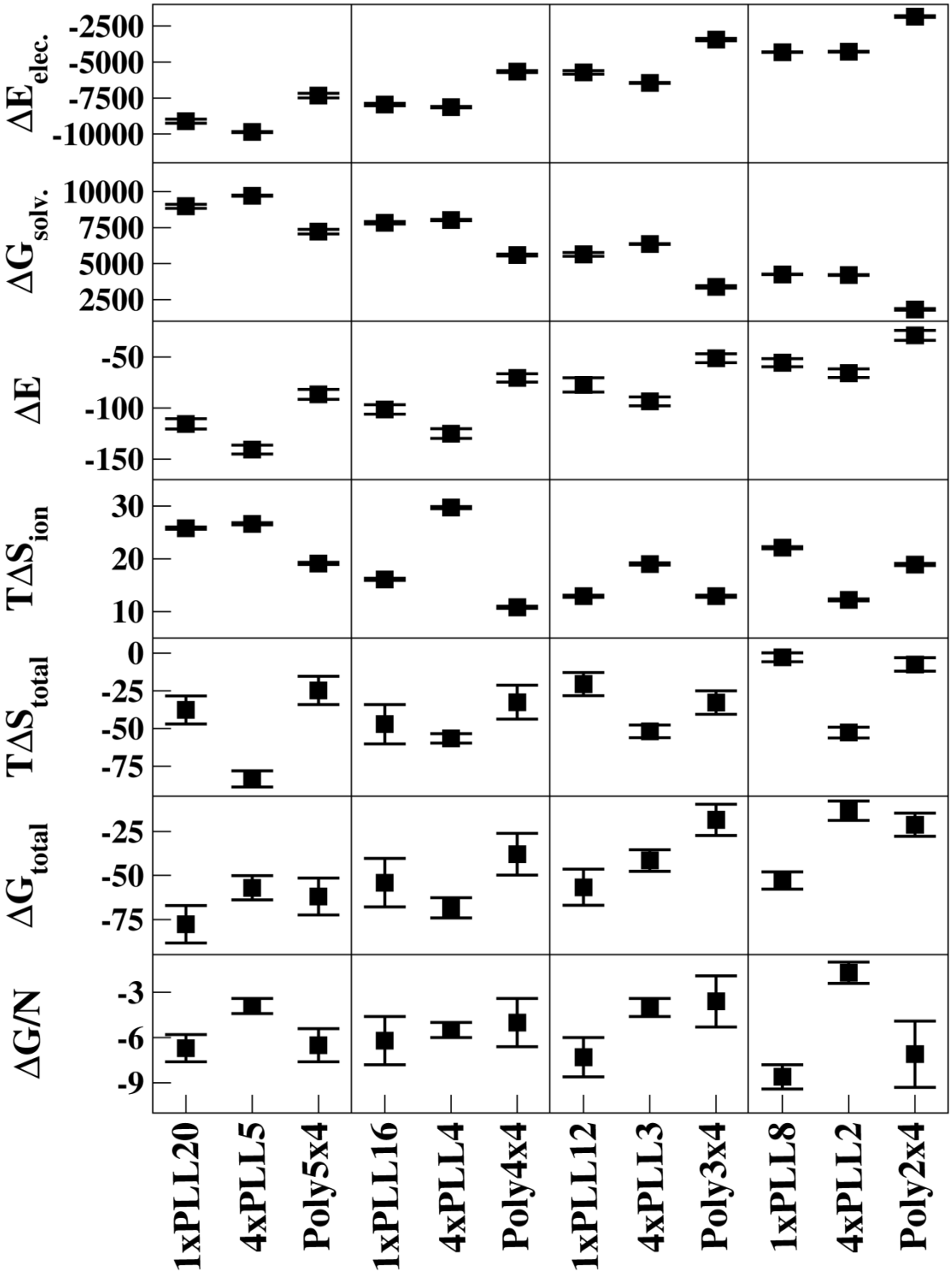


Figure 4.5 Contributions to the binding free energy for one long linear polylysine (1xPLLM), four molecules of short oligolysine (4xPLLN), and one oligomer with four repeat units of grafted oligolysines (PolyNx4) all binding to 16bp DNA. Units are kcal/mol.

Considering systems with equal total number of amines, the solvation free energy ΔG_{solv} is approximately the same for long and short PLL, but more unfavorable (positive and higher in magnitude) for the corresponding grafted architecture. The non-polar contribution to ΔG_{solv} includes ΔG_{cavity} and $\Delta G_{dispersion}$, both of which are generally the highest in magnitude for the short linear lysines, followed by long linear lysine, and finally the grafted counterpart, with the difference between the long linear PLL and grafted systems being higher for Poly5 than Poly2 (Table 4.10 in supplementary section 4.6). The polar contribution to ΔG_{solv}, $\Delta G_{PB,}$ is orders of magnitude larger than the non-polar contributions and is dependent on the number of charges and the charge density of the complex. Differing charge density for the linear and grafted architectures may cause differences in ΔG_{PB} and consequently in ΔG_{solv}. The lower charge density in the grafted architecture gives the systematically lower ΔG_{solv} values. As a result of these individual energetic contributions, the total ΔE (Figure 4.5) is highest in magnitude for the short oligolysines, followed by the long linear polylysine and then the grafted oligolysine. These data suggest that one long linear polylysine has less favorable ΔE than several shorter oligolysines, with the same total number of lysines. Thus, connecting many short oligolysines together into one single long polylysine makes DNA binding less energetically favorable than unconnected short oligolysines. In contrast, when the oligolysines are grafted on a hydrophobic backbone (PolyN), the ΔE is lower than the linear counterpart, in accord with the expectation that the hydrophobic backbone reduces the binding strength.

The trends in entropic term $T\Delta S_{ion}$ depend on both graft length and architecture: the long and short linear architectures, 1xPLL20 and 4xPLL5, have a more favorable $T\Delta S_{ion}$ than Poly5x4; for Poly3x4 and Poly4x4, the corresponding long linear PLL and the grafted architecture are roughly equal and *less* favorable than the short linear PLL; while for Poly2x4, the corresponding long linear PLL and the grafted architecture are equal and *more* favorable than the short linear PLL. These complex trends depend on the number of free ions before and after complexation (Tables 4.8 and 4.9 in supplementary section 4.6). In the absence of DNA, as the molecular weight of the polycation increases, the number and fraction of Cl^- ions bound to the polycation increases. In the DNA-only simulations, the number of bound ions is slightly different between different systems because of the differing total number of ions, although the fraction of

bound ions is approximately the same. Greater changes in the number of bound ions before and after complexation result in larger values of $T\Delta S_{ion}$. For example, Poly4x4 displaces only about 2 counterions for an entropic gain of 11 kcal/mol, while 4xPLL4 displaces nearly 6 counterions for a much greater entropic gain of approximately 30 kcal/mol. The total entropic contribution shows similar trends for all PolyN: $T\Delta S_{total}$ of the grafted architecture and long linear lysine have comparable values but lower magnitude than short linear lysine. For all systems, the total entropic contribution, $T\Delta S_{total}$, is negative and therefore disfavors binding.

In summary, for systems with different architectures but equal number of amines, the entropic and energetic contributions together make the ΔG_{total} is most favorable for one long PLL. The ΔG_{total} of the grafted architecture either falls between one long PLL and many short PLL as in the case of Poly2 and Poly5 or is the least favorable for binding as in the case of Poly3 and Poly4. Despite the less favorable energetic contribution for long linear PLL than short linear PLL, the increased entropic penalty for short linear PLL binding to DNA makes the total binding free energy most favorable for long linear PLL. Even though the entropic contributions of the grafted architecture are similar to those of long linear PLL, the energetic penalty arising from the hydrophobic backbone makes the overall binding free energy of the grafted architecture less favorable than linear PLL. This is an important element believed to contribute to the performance of these polymers as polyplexes with DNA in cell culture experiments.

4.3.4 Comparison with experiments

Effective polycationic transfection reagents should, through electrostatic binding, protect DNA from intracellular nucleases and destabilizing proteins, and upon reaching the nucleus should unbind from the DNA. Prior experimental studies show that oligolysine-grafted polymers are far more effective in transfection than linear polylysine, and in some cell lines a distinct effect of graft length was noted, where for example the longer grafts of Poly5 result in less effective transfection [4]. Interestingly, those experiments confirmed that all of the graft copolymer reagents capably deliver plasmids to the nucleus, leaving the ability for DNA to be released from the polyplex, i.e. the polycation unbinding from the DNA

as the key enabling step. Parelkar *et al*. [4] hypothesized that Poly5 and Poly2 formed stronger polyplexes than Poly3 and Poly4, due to the similarity in architecture (linear charge sequence) between Poly2 and PLL, and the length of grafts in Poly5 that tends towards PLL-like binding. In Figure 4.6, we show the normalized binding energy, $\Delta G_{total}/N$, for the grafted oligolysines, PolyN, and the corresponding long linear PLL systems.

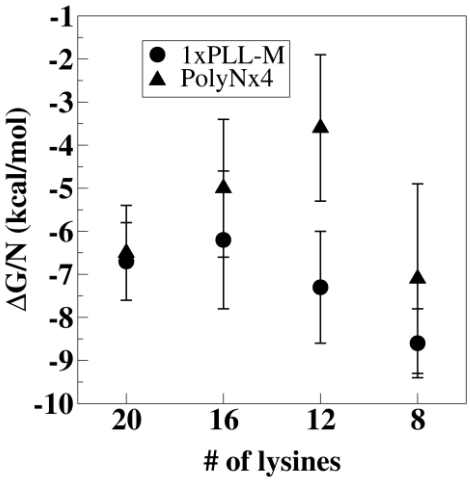


Figure 4.6 Comparison of normalized binding free energy, ΔG/N, versus number of lysines for linear, 1xPLLM, and grafted oligolysines, PolyNx4.

Considering Poly3x4, Poly4x4, and Poly5x4, we observe that with increasing graft length, $\Delta G_{total}/N$ for the PolyN systems approaches $\Delta G_{total}/N$ of the long linear PLL systems suggesting that with increasing graft length the DNA-binding behavior of the grafted architecture approaches that of the linear architecture. Although the $\Delta G_{total}/N$ values of Poly2x4 approach the linear counterpart 1xPLL8 (albeit with large error bars), we cannot conclusively say that Poly2 behaves PLL, since our results suggest that Poly2 binds the most weakly of all PolyN systems. Nevertheless, we can conclusively state that the success of Poly3 and Poly4 lies in their weaker DNA binding (both ΔG_{total} and $\Delta G_{total}/N$) than Poly5 and PLL.

While comparing these simulation results to experiments, we note certain limitations. First, atomistic simulations provide a high level of detail at the expense of limited scale, such that the largest feasibly studied system is smaller than a typical polyplex. To study polycation-DNA interactions

approaching the scale of experimental systems, coarse-grained models are needed, which sacrifice atomistic details while providing the necessary length scale. Nonetheless, these atomistic simulations provide critically important information for developing effective coarse-grained models, and are therefore a crucial first step in reaching larger scales [34-39]. Second, our approach focused on polycation and DNA binding, and did not capture the intracellular interactions between polycation, DNA, and various proteins. Intracellular polycation-DNA unbinding may be partially driven by competition with abundant intracellular proteins. In fact, the stability of polyplexes can be measured by their resistance to decomplexation by protein-rich media. It is interesting to note, however, that since the PicoGreen-intercalation and DNase-protection experiments involve unbinding of DNA [4], the trends from binding simulations described here could be extended to unbinding within the nucleus. Third, because the binding of each polycation is irreversible on the timescales in the simulation, it limits our ability to sample different binding modes. Devising methodology for temporarily weakening some polycation-DNA interactions could allow the complex to shift into different configurations, thus improving sampling.

4.4 Conclusions

In summary, we have used molecular dynamics simulations to examine the effect of lysine-based polymer architectures on the structure and energetics of DNA binding. Structurally, the grafted oligolysine architecture displays different binding modes as compared to linear polylysine, characterized by weaker, less permanent contacts with DNA. The hydrophobic backbone in the grafted architecture sterically hinders the motion of the grafts, and weakens binding. Thus, the grafted architecture exhibits less favorable binding energy than linear poly-L-lysine with the same number of charged amines, and the binding free energy varies non-monotonically with the graft length due to non-monotonic trends in entropic contributions. The binding free energy normalized to the number of bound amines is similar between the grafted and linear architectures at the largest (Poly5) and smallest (Poly2) graft length, and weaker for the intermediate graft lengths (Poly3 and Poly4). These trends correlate well with experimental results that show higher transfection efficiency for Poly3 and Poly4 graft structures relative

to Poly5, Poly2 and linear PLL. This study is a first step toward connecting molecular simulations and experiments of polycation architecture to DNA binding, and in turn DNA transfection efficiency.

4.5 References

[1] D. J. Glover, H. J. Lipps, and D. A. Jans, Nature Reviews Genetics 6 (2005) 299.
[2] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, and J. P. Behr, Proceedings of the National Academy of Sciences 92 (1995) 7297.
[3] E. Wagner, Proceedings of the National Academy of Sciences 88 (1991) 4255.
[4] S. S. Parelkar, D. Chan-Seng, and T. Emrick, Biomaterials 32 (2011) 2432.
[5] R. B. Breitenkamp and T. Emrick, Biomacromolecules 9 (2008) 2495.
[6] M. X. Tang and F. C. Szoka, Gene Therapy 4 (1997) 823.
[7] D. Lechardeur, A. S. Verkman, and G. L. Lukacs, Advanced Drug Delivery Reviews 57 (2005) 755.
[8] D. W. Pack, A. S. Hoffman, S. Pun, and P. S. Stayton, Nature Reviews Drug Discovery 4 (2005) 581.
[9] Y. Wang, M. Zheng, F. Meng, J. Zhang, R. Peng, and Z. Zhong, Biomacromolecules 12 (2011) 1032.
[10] J. Haensler and F. C. Szoka Jr, Bioconjugate Chemistry 4 (1993) 372.
[11] Z. Ou and M. Muthukumar, Journal of Chemical Physics 123 (2005) 074905.
[12] C. F. Narambuena, E. P. M. Leiva, M. Chávez-Páez, and E. Pérez, Polymer 51 (2010) 3293.
[13] J. Ziebarth and Y. M. Wang, Biophysical Journal 97 (2009) 1971.
[14] D. Ouyang, H. Zhang, H. S. Parekh, and S. C. Smith, Journal of Physical Chemistry B 114 (2010) 9231.
[15] G. M. Pavan, A. Danani, S. Pricl, and D. K. Smith, Journal of the American Chemical Society 131 (2009) 9686.
[16] T. E. Cheatham, P. Cieplak, and P. A. Kollman, Journal of Biomolecular Structure and Dynamics 16 (1999) 845.
[17] A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco, Biophysical Journal 92 (2007) 3817.
[18] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins 65 (2006) 712.
[19] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Journal of Computational Chemistry 25 (2004) 1157.
[20] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, Journal of the American Chemical Society 117 (1995) 5179.
[21] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of Computational Chemistry 26 (2005) 1781.
[22] A. Jakalian, D. B. Jack, and C. I. Bayly, Journal of Computational Chemistry 23 (2002) 1623.
[23] H. Gohlke and D. A. Case, Journal of Computational Chemistry 25 (2004) 238.
[24] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, Accounts of Chemical Research 33 (2000) 889.
[25] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics and Modelling 14 (1996) 33.
[26] C. Tan, Y.-H. Tan, and R. Luo, Journal of Physical Chemistry B 111 (2007) 12263.
[27] D. P. Mascotti and T. M. Lohman, Proceedings of the National Academy of Sciences of the United States of America 87 (1990) 3142.

[28] K. Wagner, D. Harries, S. May, V. Kahl, J. O. Rädler, and a. Ben-Shaul, Langmuir 16 (2000) 303.
[29] K. K. Frederick, M. S. Marlow, K. G. Valentine, and A. J. Wand, Nature 448 (2007) 325.
[30] M. S. Marlow, J. Dogan, K. K. Frederick, K. G. Valentine, and A. J. Wand, Nature Chemical Biology 6 (2010) 352.
[31] Z. Ou and M. Muthukumar, Journal of Chemical Physics 124 (2006) 154902.
[32] D. W. Li and R. Bruschweiler, Physical Review Letters 102 (2009) 118108.
[33] D. W. Li and R. Bruschweiler, Journal of the American Chemical Society 131 (2009) 7226.
[34] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, Biophysical Journal 96 (2009) 1675.
[35] S. M. Gopal, S. Mukherjee, Y.-M. Cheng, and M. Feig, Proteins 78 (2010) 1266.
[36] A. W. Stumpff-Kane, K. Maksimiak, M. S. Lee, and M. Feig, Proteins 70 (2008) 1345.
[37] W. G. Noid, J. W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, Journal of Chemical Physics 128 (2008)
[38] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, Biophysical Journal 92 (2007) 4289.
[39] A. Striolo, A. Jayaraman, J. Genzer, and C. K. Hall, Journal of Chemical Physics 123 (2005)
[40] D. Frenkel and B. Smit, Understanding Molecular Simulations: From Algorithms to Applications, Academic Press, 2001.

4.6 Supplementary information

4.6.1 Charge derivation protocol used to assign charges to poly(cyclooctene-g-oligolysine) and a listing of the charges used in this work

The *antechamber* program (a part of the Amber suite) [16-19] was used to calculate partial charges the poly(cyclooctene) (PCO) backbone of poly(cyclooctene-g-oligolysine) using the AM1-BCC method [22], a semi-empirical Hamiltonian with bond charge correction. We decided to retain the preexisting charges for lysine in the Amber force field, as these values have been rigorously tested. Integrating the lysine charges with the semi-empirical charges calculated with *antechamber* proceeded in multiple steps.

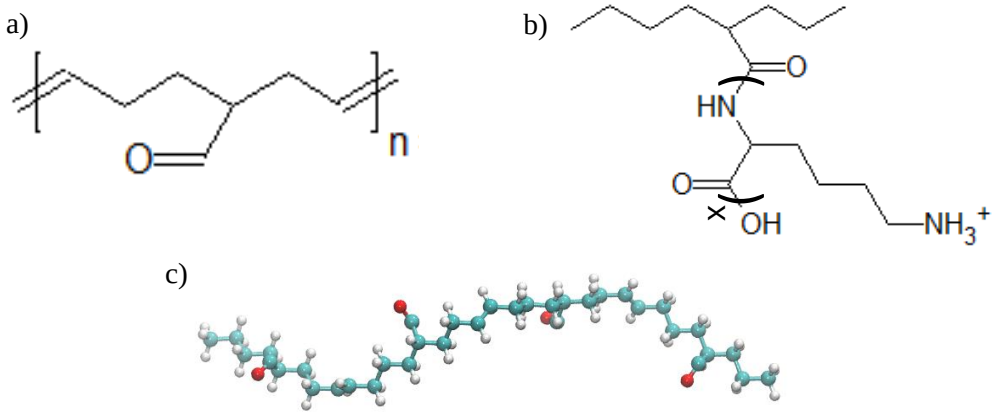


Figure 4.7 Chemical structures of poly(cyclooctene) backbone (a) and oligolysine-grafted octane (b) used to calculate partial charges of the PCO backbone of poly(cyclooctene-g-oligolysine). For our calculations, n = 4 and x = 5. (c) is a rendering of (a).

First, we calculated the partial charges of (**A**), the PCO backbone (Figure 4.7a) without poly-L-lysine grafts. (**A**) of octene monomers modified by the addition of an aldehyde group on the fourth carbon; four monomers were joined by double bonds. The purpose of the aldehyde groups was to provide a graft point for the oligolysine graft, after the removal of the hydrogen from the aldehyde groups. To use the graft point, the aldehyde hydrogen is removed and the aldehyde carbon is bonded to the first nitrogen atom in the oligolysine peptide backbone (Figure 4.7b). Second, we calculated the charges of (**B**), a single pentalysine-grafted octane (Figure 4.7b), and these charges were used to modify the charges calculated for (**A**). Specifically, to create the completed Poly-N structure, we took the following steps. First, the

charges of the oxygen and two carbons in the aldehyde group (including the carbon in the octene backbone), in (**A**) were changed to match those of (**B**) to account for the presence of the oligolysine. Second, the aldehyde hydrogen atom was removed and the charge was redistributed to the central aldehyde carbon. Third, because the replacement of our calculated lysine charges in (**A**) with the charges from the Amber force field resulted in a difference in the total charge of the molecule, we evenly redistributed the difference in total charge to the aldehyde oxygen and two aldehyde carbons (again, including the carbon in the octene backbone). The charge of the remaining PCO backbone atoms was unchanged from (**A**). Although this charge redistribution method is somewhat arbitrary, it resulted in changes of only 0.04 to 0.11 e^-. To create a PCO 8mer backbone for the PolyNx8 systems, two PCO 4mers (**A**) were bonded after the removal of a hydrogen atom from the terminal methyl groups, and the charges of the atoms of the two central monomers of the 8mer were changed to match the average charges of the atoms of the other non-terminal monomers. This resulted in a slight deviation in the total charge, which was evenly redistributed to all of the non-terminal monomer atoms (i.e. all atoms in cyclooctene monomers 2 to 7), which resulted in only small changes in the charges of individual atoms. The charges for lysine, as mentioned above, were taken from the Amber force field. We used residue type LYS for the non-terminal lysine residues and type CLYS for the terminal lysine residues. These charges are freely available and are not listed here.

Table 4.1 List of general Amber force field (*gaff*) atom types and partial charges for the PCO backbone with 4 cyclooctene monomers used for PolyNx4. Borders of table delineate the individual cyclooctene monomers.

#	Atom Name	gaff Atom Type	Charge (e-)	#	Atom Name	gaff Atom Type	Charge (e-)	#	Atom Name	gaff Atom Type	Charge (e-)
1	C1	c3	-0.0927	33	H34	hc	0.0498	65	H68	hc	0.0485
2	H2	hc	0.0336	34	C35	c3	-0.1465	66	H69	hc	0.0485
3	H3	hc	0.0379	35	H36	hc	0.0787	67	C70	c3	-0.0478
4	H4	hc	0.0326	36	C37	c	0.6405	68	H71	hc	0.0433
5	C5	c3	-0.0853	37	O39	o	-0.6355	69	H72	hc	0.0492
6	H6	hc	0.0469	38	C40	c3	-0.0725	70	C73	c2	-0.1665
7	H7	hc	0.0435	39	H41	hc	0.0435	71	H74	ha	0.1236
8	C8	c3	-0.0724	40	H42	hc	0.0496	72	C75	c2	-0.1711
9	H9	hc	0.042	41	C43	c3	-0.0813	73	H76	ha	0.1201
10	H10	hc	0.0495	42	H44	hc	0.0485	74	C77	c3	-0.0531
11	C11	c3	-0.1465	43	H45	hc	0.0487	75	H78	hc	0.0548
12	H12	hc	0.0793	44	C46	c3	-0.0479	76	H79	hc	0.0495
13	C13	c	0.6405	45	H47	hc	0.0433	77	C80	c3	-0.0688
14	O15	o	-0.6355	46	H48	hc	0.0491	78	H81	hc	0.0471
15	C16	c3	-0.0724	47	C49	c2	-0.1662	79	H82	hc	0.0497
16	H17	hc	0.0422	48	H50	ha	0.1234	80	C83	c3	-0.1465
17	H18	hc	0.0497	49	C51	c2	-0.1712	81	H84	hc	0.0783
18	C19	c3	-0.0816	50	H52	ha	0.1204	82	C85	c	0.6405
19	H20	hc	0.0487	51	C53	c3	-0.0529	83	O87	o	-0.6355
20	H21	hc	0.0486	52	H54	hc	0.0551	84	C88	c3	-0.0722
21	C22	c3	-0.048	53	H55	hc	0.0489	85	H89	hc	0.0431
22	H23	hc	0.043	54	C56	c3	-0.069	86	H90	hc	0.049
23	H24	hc	0.0494	55	H57	hc	0.0474	87	C91	c3	-0.0848
24	C25	c2	-0.1657	56	H58	hc	0.05	88	H92	hc	0.0474
25	H26	ha	0.1236	57	C59	c3	-0.1465	89	H93	hc	0.0445
26	C27	c2	-0.1719	58	H60	hc	0.0788	90	C94	c3	-0.0791
27	H28	ha	0.1202	59	C61	c	0.6405	91	H95	hc	0.0383
28	C29	c3	-0.0528	60	O63	o	-0.6355	92	H96	hc	0.0386
29	H30	hc	0.055	61	C64	c3	-0.0726	93	C97	c3	-0.0931
30	H31	hc	0.0488	62	H65	hc	0.0434	94	H98	hc	0.0338
31	C32	c3	-0.069	63	H66	hc	0.0497	95	H99	hc	0.0344
32	H33	hc	0.0475	64	C67	c3	-0.0814	96	H	hc	0.0351

Table 4.2 List of general Amber force field (*gaff*) atom types and partial charges for the PCO backbone with 8 cyclooctene monomers used for PolyNx8. Borders of table delineate the individual cyclooctene monomers.

#	Atom Name	gaff Atom Type	Charge (e-)	#	Atom Name	gaff Atom Type	Charge (e-)	#	Atom Name	gaff Atom Type	Charge (e-)
1	C1	c3	-0.0927	33	H34	hc	0.0498	65	H68	hc	0.0485
2	H2	hc	0.0336	34	C35	c3	-0.1465	66	H69	hc	0.0485
3	H3	hc	0.0379	35	H36	hc	0.0787	67	C70	c3	-0.0478
4	H4	hc	0.0326	36	C37	c	0.6359	68	H71	hc	0.0433
5	C5	c3	-0.0853	37	O39	o	-0.6355	69	H72	hc	0.0492
6	H6	hc	0.0469	38	C40	c3	-0.0725	70	C73	c2	-0.1665
7	H7	hc	0.0435	39	H41	hc	0.0435	71	H74	ha	0.1236
8	C8	c3	-0.0724	40	H42	hc	0.0496	72	C75	c2	-0.1711
9	H9	hc	0.042	41	C43	c3	-0.0813	73	H76	ha	0.1201
10	H10	hc	0.0495	42	H44	hc	0.0485	74	C77	c3	-0.0531
11	C11	c3	-0.1465	43	H45	hc	0.0487	75	H78	hc	0.0548
12	H12	hc	0.0793	44	C46	c3	-0.0479	76	H79	hc	0.0495
13	C13	c	0.6359	45	H47	hc	0.0433	77	C80	c3	-0.0688
14	O15	o	-0.6355	46	H48	hc	0.0491	78	H81	hc	0.0471
15	C16	c3	-0.0724	47	C49	c2	-0.1662	79	H82	hc	0.0497
16	H17	hc	0.0422	48	H50	ha	0.1234	80	C83	c3	-0.1465
17	H18	hc	0.0497	49	C51	c2	-0.1712	81	H84	hc	0.0783
18	C19	c3	-0.0816	50	H52	ha	0.1204	82	C85	c	0.6359
19	H20	hc	0.0487	51	C53	c3	-0.0529	83	O87	o	-0.6355
20	H21	hc	0.0486	52	H54	hc	0.0551	84	C88	c3	-0.0722
21	C22	c3	-0.048	53	H55	hc	0.0489	85	H89	hc	0.0431
22	H23	hc	0.043	54	C56	c3	-0.069	86	H90	hc	0.049
23	H24	hc	0.0494	55	H57	hc	0.0474	87	C91	c3	-0.0848
24	C25	c2	-0.1657	56	H58	hc	0.05	88	H92	hc	0.0474
25	H26	ha	0.1236	57	C59	c3	-0.1465	89	H93	hc	0.0445
26	C27	c2	-0.1719	58	H60	hc	0.0788	90	C94	c3	-0.0791
27	H28	ha	0.1202	59	C61	c	0.6359	91	H95	hc	0.0383
28	C29	c3	-0.0528	60	O63	o	-0.6355	92	H96	hc	0.0386
29	H30	hc	0.055	61	C64	c3	-0.0726	93	C97	c2	-0.04175
30	H31	hc	0.0488	62	H65	hc	0.0434	94	H99	ha	0.05195
31	C32	c3	-0.069	63	H66	hc	0.0497	95	C1	c2	-0.03895
32	H33	hc	0.0475	64	C67	c3	-0.0814	96	H4	ha	0.05155

Table 4.2 continued.

#	Atom Name	gaff Atom Type	Charge (e-)	#	Atom Name	gaff Atom Type	Charge (e-)	#	Atom Name	gaff Atom Type	Charge (e-)
97	C5	c3	-0.0853	129	O39	o	-0.6355	161	H72	hc	0.0492
98	H6	hc	0.0469	130	C40	c3	-0.0725	162	C73	c2	-0.1665
99	H7	hc	0.0435	131	H41	hc	0.0435	163	H74	ha	0.1236
100	C8	c3	-0.0724	132	H42	hc	0.0496	164	C75	c2	-0.1711
101	H9	hc	0.042	133	C43	c3	-0.0813	165	H76	ha	0.1201
102	H10	hc	0.0495	134	H44	hc	0.0485	166	C77	c3	-0.0531
103	C11	c3	-0.1465	135	H45	hc	0.0487	167	H78	hc	0.0548
104	H12	hc	0.0793	136	C46	c3	-0.0479	168	H79	hc	0.0495
105	C13	c	0.6359	137	H47	hc	0.0433	169	C80	c3	-0.0688
106	O15	o	-0.6355	138	H48	hc	0.0491	170	H81	hc	0.0471
107	C16	c3	-0.0724	139	C49	c2	-0.1662	171	H82	hc	0.0497
108	H17	hc	0.0422	140	H50	ha	0.1234	172	C83	c3	-0.1465
109	H18	hc	0.0497	141	C51	c2	-0.1712	173	H84	hc	0.0783
110	C19	c3	-0.0816	142	H52	ha	0.1204	174	C85	c	0.6359
111	H20	hc	0.0487	143	C53	c3	-0.0529	175	O87	o	-0.6355
112	H21	hc	0.0486	144	H54	hc	0.0551	176	C88	c3	-0.0722
113	C22	c3	-0.048	145	H55	hc	0.0489	177	H89	hc	0.0431
114	H23	hc	0.043	146	C56	c3	-0.069	178	H90	hc	0.049
115	H24	hc	0.0494	147	H57	hc	0.0474	179	C91	c3	-0.0848
116	C25	c2	-0.1657	148	H58	hc	0.05	180	H92	hc	0.0474
117	H26	ha	0.1236	149	C59	c3	-0.1465	181	H93	hc	0.0445
118	C27	c2	-0.1719	150	H60	hc	0.0788	182	C94	c3	-0.0791
119	H28	ha	0.1202	151	C61	c	0.6359	183	H95	hc	0.0383
120	C29	c3	-0.0528	152	O63	o	-0.6355	184	H96	hc	0.0386
121	H30	hc	0.055	153	C64	c3	-0.0726	185	C97	c3	-0.0931
122	H31	hc	0.0488	154	H65	hc	0.0434	186	H98	hc	0.0338
123	C32	c3	-0.069	155	H66	hc	0.0497	187	H99	hc	0.0344
124	H33	hc	0.0475	156	C67	c3	-0.0814	188	H	hc	0.0351
125	H34	hc	0.0498	157	H68	hc	0.0485				
126	C35	c3	-0.1465	158	H69	hc	0.0485				
127	H36	hc	0.0787	159	C70	c3	-0.0478				
128	C37	c	0.6359	160	H71	hc	0.0433				

4.6.2 Detailed description of simulation protocol

First, in constant volume and temperature conditions (NVT ensemble), the water and ions were minimized with the conjugate gradient algorithm for 10000 steps, with the DNA subjected to a 500 kcal/mol-A^2 harmonic restraining potential. Second, in the NVT ensemble, with the DNA restraining potential reduced to 100 kcal/mol-A^2, the system was heated from 0 to 300 K over 20 ps. Third, in constant pressure and temperature conditions (NPT ensemble) and with the restraint reduced to 50 kcal/mol-A^2 the system was relaxed for another 20 ps. Fourth, in the NVT ensemble, the restraining potential was reduced in three steps (50, 10, and 5 kcal/mol-A^2) and minimized for 2000 steps each. Fifth, in the NVT ensemble with the same 5 kcal/mol-A^2 restraint the system was heated from 10 to 300 K over 20 ps. Sixth, in the NPT ensemble, the restraint was reduced from 5 to 1 to 0.1 to 0 kcal/mol-A^2 over 20 ps. Seventh, in the NVT ensemble with no restraints, the system was heated from 10 to 300 K over 20 ps. Finally, the system was heated from 100 to 300 K over 20 ps in the NPT ensemble at the beginning of each production run. Langevin dynamics was used to control the temperature (damping coefficient of 1 ps^{-1}) and pressure (piston period of 100 fs and piston decay constant of 50 fs) at 300 K and 1 atm. The SHAKE [40] algorithm was used to constrain all bonds involving hydrogen, and a time step of 2 fs was used. Electrostatic interactions were treated with the particle-mesh Ewald (PME) summation method [40], with a tolerance of 1e-6 and interpolation order of 4. The non-bonded cutoff was 9.0 Å and the non-bonded list was updated every 10 steps. Snapshots were recorded every 2 ps.

4.6.3 List of systems studied

Table 4.3 Specification of the systems studied in this work, including name, description, N/P ratio, and the number of counterions. Nomenclature: NxPLLM refers to a system including N linear lysine with M-mers; PolyMxN refers to a system including a single poly(cyclooctene-*g*-oligolysine) molecule with grafts of length M and N repeat units. All systems include a 16 basepair double-stranded DNA oligomer. Molecular weight (MW) per repeat unit given is for non-terminal residues, even in the case where there are only terminal residues (i.e. PLL2).

System Name	Description	MW per repeat unit (g/mol)	N/P Ratio	# Na^+	# Cl^-
Linear Poly-L-Lysine					
4xPLL2	4 chains of PLL 2mers	130	0.3	30	8
4xPLL3	4 chains of PLL 3mers	130	0.4	30	12
4xPLL4	4 chains of PLL 4-mers	130	0.5	30	16
4xPLL5	4 chains of PLL 5mers	130	0.7	30	20
1xPLL8	1 chain of PLL 8-mer	130	0.3	30	8
1xPLL12	1 chain of PLL 12mer	130	0.4	30	12
1xPLL16	1 chain of PLL 16mer	130	0.5	30	16
1xPLL20	1 chain of PLL 20mer	130	0.7	26	16
Grafted Oligolysine					
Poly2x4	Graft length 2, 4 repeat units	414	0.3	34	8
Poly3x4	Graft length 3, 4 repeat units	543	0.4	32	10
Poly4x4	Graft length 4, 4 repeat units	673	0.5	30	12
Poly5x4	Graft length 5, 4 repeat units	802	0.7	30	16
Poly2x8	Graft length 2, 8 repeat units	414	0.5	38	16
Poly3x8	Graft length 3, 8 repeat units	543	0.8	34	20
Poly4x8	Graft length 4, 8 repeat units	673	1.1	30	24
Poly5x8	Graft length 5, 8 repeat units	802	1.3	26	28

The N/P ratio and the charge neutrality of each system may not be immediately apparent for several reasons. First, the 16bp DNA in each simulation has a charge of –30 rather than –32 because the parameterization of the terminal base pairs excludes a negatively-charged phosphate group. Second, each linear lysine has a net positive charge equal to its length, e.g. +16 for PLL16. Third, each graft on PolyN has a net positive charge of *one less than the graft length*, e.g. +4 for each graft on Poly5 for a total charge of +16 for Poly5x4. The net charge of each oligolysine graft is one less than the graft length because of the negatively-charged C-terminus at the end of the graft. The grafts lack the positively-charged N-terminus due to the attachment to the poly(cyclooctene) backbone. In linear lysine, the N-terminus cancels the charge of the C-terminus, leading to net charge equal to the length of the polycation.

Regardless of the charge of the polycation, the N/P ratio is calculated as the ratio of the number of protonated nitrogens (i.e. the number of lysines) to the number of phosphates. For example, for Poly5x8 the number of lysines is 40, the number of phosphates is 30, and consequently the N/P ratio is 1.3 (rounded to one decimal place). The total charge of the Poly5x8 system is zero because the net charge of Poly5x8 is +32, the charge of the DNA is –30, and the net charge of the counterions is –2.

4.6.4 Effect of the smoothing parameter, κ, of the von Mises kernel density estimate on the calculated conformational entropy

Table 4.4 Results of conformational entropy calculations performed with several different values of the smoothing parameter, κ, which is analogous to σ^{-2} of the normal distribution, to determine the appropriate value to use. As κ increases, the kernel density estimate more closely resembles the actual distribution of a particular dihedral angle; the entropy calculation converges as κ increases. However, if κ is too large, the density estimate is no longer smooth, and the entropy calculations will not be accurate. We tested these values of κ using PLL20 as a model system and determined that κ = 225 is appropriate because it yielded the lowest ΔS value. We used κ = 225 for the remainder of the calculations, although values over the range 100 to 300 could also be acceptable.

κ	S, Complex	S, DNA	S, PLL-20	$\Delta S/k_B$
1.0	-20.7	-14.4	-9.1	-2.8
57.3	-580.8	-501.9	-93.1	-14.1
100.0	-751.9	-660.3	-106.1	-14.4
225.0	-1041.9	-929.7	-126.8	-14.5
306.3	-1161.9	-1041.4	-134.9	-14.4

4.6.5 Binding maps of PolyNx4 and PolyNx8 reveal qualitative graft length-dependence

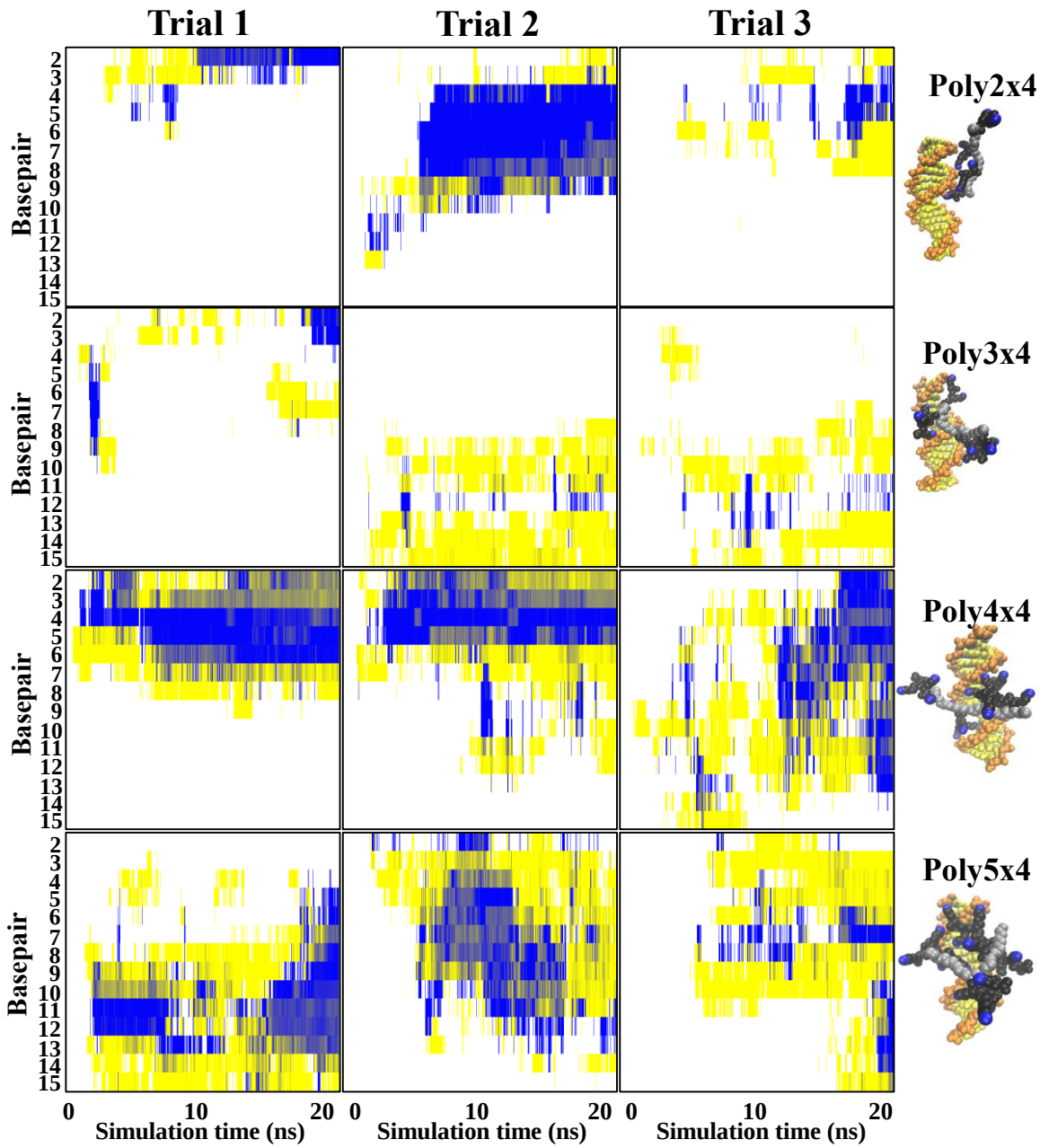


Figure 4.8 Binding heat maps of three trials of PolyNx4 (N = 2 to 5). The y-axis corresponds to individual basepairs; the terminal basepairs are omitted. The x-axis corresponds to simulation time. Representative snapshots of the 4 systems are shown at right. As the graft length increases, the grafted architecture increasingly interacts with the DNA grooves, as evidenced by an increase in the amount of blue visible going from Poly3x4 to Poly5x4. Trial 2 of Poly2x4 interacted substantially with the grooves because it was small enough to coil up entirely within the major groove; the other trials of Poly2x4, however, interacted mostly with the DNA backbone, albeit weakly.

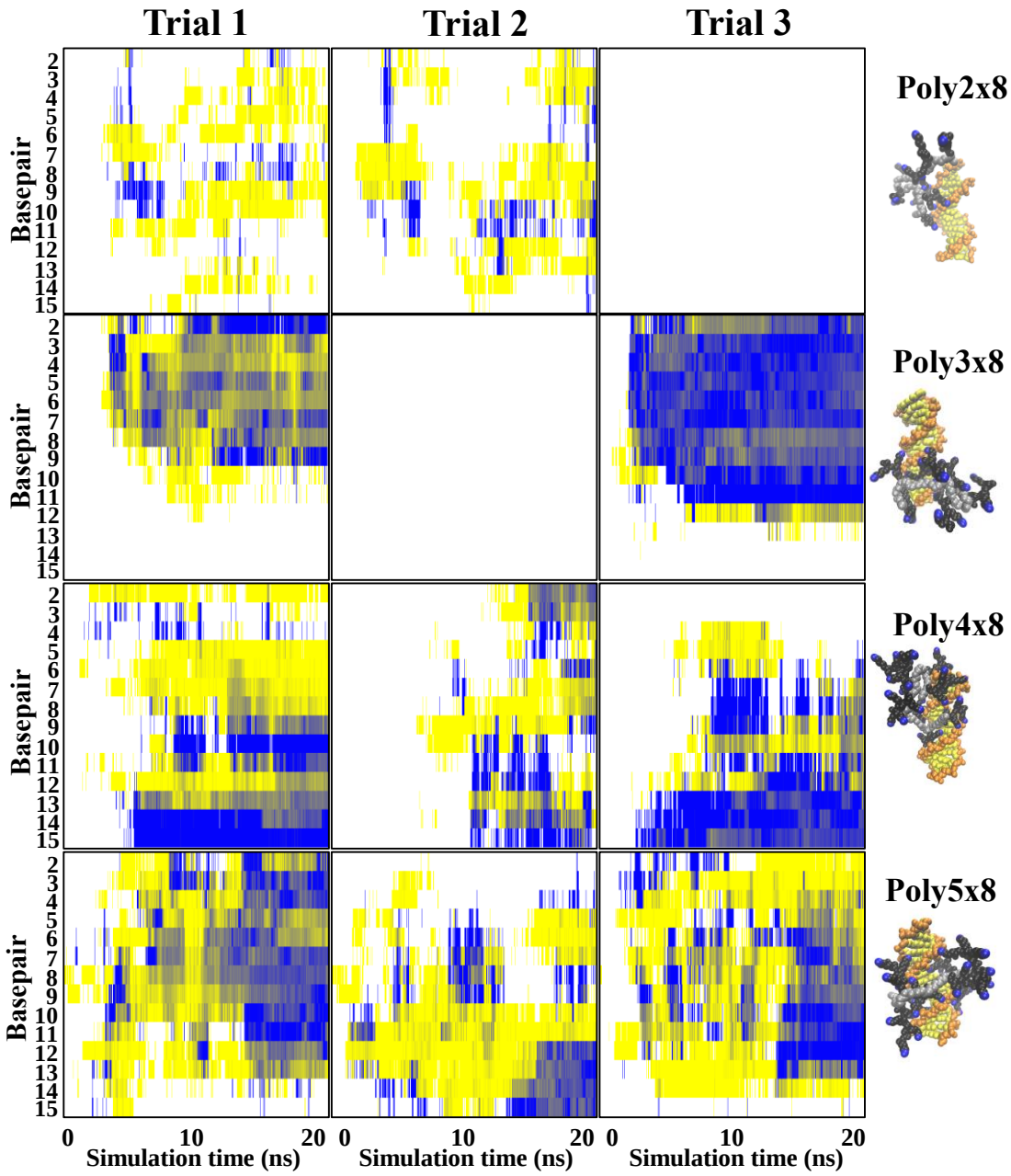


Figure 4.9 Binding heat maps of three trials of PolyNx8 (N = 2 to 5). The y-axis corresponds to individual basepairs; the terminal basepairs are omitted. The x-axis corresponds to simulation time. Representative snapshots of the 4 systems are shown at right. Trial 3 of Poly2x8 and Trial 2 of Poly3x8 did not bind to the DNA and have been excluded from analysis. We observe that Poly2x8 interacts almost entirely with the DNA backbones, albeit quite fleetingly: contacts are constantly forming and breaking, suggesting that the single charge of each Poly2 graft is insufficient to maintain a strong grasp on a phosphate group. In contrast, the Poly3-5x8 are able to maintain long-lasting contacts, although their increased bulk necessitates less-favorable interactions with the grooves in addition to the DNA backbone.

4.6.6 Effect of graft length on the individual contributions to binding energy

Table 4.5 List and description of all contributions to the polycation-DNA binding free energy.

BOND	Bond energy
ANGLE	Angle energy
DIHED	Dihedral energy
VDWAALS	vdW energy
EEL	Electrostatic energy
EPB	PB solvation free energy
1-4VDW	vdW energy of atoms with 3 bonds between them (1-4 interactions are calculated separately because they are scaled down relative to other interactions)
1-4EEL	Electrostatic energy of atoms with 3 bonds between them
RESTRAINT	Restraint energy (always 0 because no artificial restraints are applied)
ECAVITY	Free energy of forming a solvent cavity the size of the solute
EDISPER	Free energy of the solute-solvent interface
Total ΔE	Total energy; sum of all above terms (some contributions are actually free energies but we denote the total as simply energy for convenience)
TΔS: NMA Trans	Normal mode analysis (NMA) translational entropy
TΔS: NMA Rot	NMA rotational entropy
TΔS: NMA Vib	NMA vibrational entropy
TΔS: NMA Total	NMA total entropy (sum of previous 3 terms)
TΔS: Conform.	Conformational entropy
TΔS: Ion dissoc.	Ion dissociation entropy
Total TΔS	Total entropy (sum of previous 3 terms)
Total ΔG	Total free energy (= Total ΔE - Total TΔS)
Phos (P)	# of phosphates within 7 Å of a protonated N
Nitro (N)	# of protonated N with 7 Å of a phosphate
Total (T)	Sum of N + P
ΔG/P	ΔG normalized to # P
ΔG/N	ΔG normalized to # N
ΔG/T	ΔG normalized to # T

Table 4.6 Details of binding energy for PolyNx4 systems. All units are kcal/mol.

	Poly5x4		Poly4x4		Poly3x4		Poly2x4	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
BOND	-1.8	2.5	-0.8	2.6	-3.1	2.6	-0.4	2.3
ANGLE	-1.6	3.7	4.0	3.4	-1.5	3.4	-3.6	3.3
DIHED	11.1	1.9	-4.7	1.9	0.8	2.0	-3.0	2.0
VDWAALS	-30.6	2.2	-32.9	2.2	-34.7	2.3	-13.1	2.6
EEL	-7327.9	162.4	-5665.9	71.8	-3435.7	84.2	-1838.0	54.9
EPB	7223.0	160.3	5595.8	71.2	3389.7	83.1	1820.5	54.8
1-4VDW	-0.2	1.1	1.2	1.2	1.2	1.2	0.3	1.1
1-4EEL	8.7	3.6	4.3	3.8	7.7	3.0	-4.5	3.4
RESTRAINT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ECAVITY	-26.3	1.6	-26.9	1.3	-31.1	1.5	-10.9	1.7
EDISPER	58.8	3.0	55.4	1.9	55.6	2.7	23.8	3.2
Total ΔE	-86.7	4.9	-70.5	4.0	-51.2	4.3	-28.8	4.8
TΔS: NMA Trans	-14.7	0.0	-14.5	0.0	-14.4	0.0	-14.2	0.0
TΔS: NMA Rot	-14.1	0.0	-13.7	0.0	-13.6	0.1	-13.1	0.0
TΔS: NMA Vib	-11.1	3.5	-8.4	2.6	-8.7	4.0	-0.7	2.4
TΔS: NMA Total	-39.9	3.5	-36.6	2.7	-36.7	4.0	-28.0	2.4
TΔS: Conform.	-3.9	6.3	-6.8	9.0	-9.0	4.7	1.5	2.2
TΔS: Ion dissoc.	19.1	0.2	10.8	0.2	12.9	0.2	18.9	0.2
Total TΔS	-24.8	9.3	-32.6	11.2	-32.8	7.8	-7.6	4.5
Total ΔG	-61.9	10.5	-37.9	11.9	-18.4	8.9	-21.2	6.6
Phos	10.2	1.9	8.3	1.7	6.1	1.8	3.7	1.5
Nitro	9.5	1.9	7.6	1.6	5.1	1.6	3.0	1.4
Total	19.7	2.7	15.9	2.3	11.2	2.4	6.7	2.0
ΔG/P	-6.1	1.0	-4.6	1.4	-3.0	1.5	-5.7	1.8
ΔG/N	-6.5	1.1	-5.0	1.6	-3.6	1.7	-7.1	2.2
ΔG/T	-3.1	0.5	-2.4	0.7	-1.6	0.8	-3.2	1.0

Table 4.7 Details of binding energy for PolyNx8 systems. All units are kcal/mol.

	Poly5x8		Poly4x8		Poly3x8		Poly2x8	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
BOND	-0.9	2.9	2.5	2.6	6.3	3.0	3.9	2.8
ANGLE	5.1	4.3	2.9	4.1	6.5	3.7	-0.5	3.4
DIHED	-9.4	2.4	-6.1	2.4	0.2	2.2	-0.9	2.1
VDWAALS	-100.0	2.6	-82.3	3.1	-70.3	2.5	-14.9	2.8
EEL	-15015.6	76.4	-10537.1	36.2	-7631.1	32.4	-3657.6	37.7
EPB	14816.0	75.0	10386.5	35.5	7530.1	31.0	3616.9	36.5
1-4VDW	1.6	1.3	1.3	1.3	-1.2	1.3	0.3	1.4
1-4EEL	-5.7	3.9	6.9	4.0	-0.2	3.4	-7.0	3.7
RESTRAINT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ECAVITY	-93.8	2.0	-71.9	2.0	-59.2	1.7	-12.7	2.1
EDISPER	181.5	3.2	136.9	3.4	116.2	2.7	29.2	3.3
Total ΔE	-221.1	6.1	-160.3	5.4	-102.6	5.0	-43.1	4.8
TΔS: NMA Trans	-15.1	0.0	-15.0	0.0	-14.9	0.0	-14.7	0.0
TΔS: NMA Rot	-15.3	0.0	-14.9	0.0	-14.6	0.2	-14.2	0.1
TΔS: NMA Vib	-51.8	4.0	-32.7	3.8	-29.4	13.4	-14.0	3.2
TΔS: NMA Total	-82.2	4.0	-62.6	3.8	-58.9	13.6	-43.0	3.2
TΔS: Conform.	-16.6	2.8	-21.6	9.8	-26.3	0.9	-10.8	8.4
TΔS: Ion dissoc.	49.6	0.2	12.3	0.2	11.9	0.2	12.6	0.2
Total TΔS	-49.2	4.9	-72.0	10.5	-73.2	13.7	-41.2	9.0
Total ΔG	-171.9	7.8	-88.3	11.8	-29.4	14.6	-1.9	10.2
Phos	17.2	2.4	10.8	1.8	10.5	1.5	6.2	1.7
Nitro	17.6	2.8	11.0	1.8	10.2	1.2	5.1	1.7
Total	34.8	3.7	21.8	2.5	20.7	1.9	11.4	2.3
ΔG/P	-10.0	0.5	-8.2	1.1	-2.8	1.4	-0.2	1.0
ΔG/N	-9.8	0.4	-8.0	1.1	-2.9	1.4	-0.4	2.0
ΔG/T	-4.9	0.2	-4.1	0.5	-1.4	0.7	-0.2	0.9

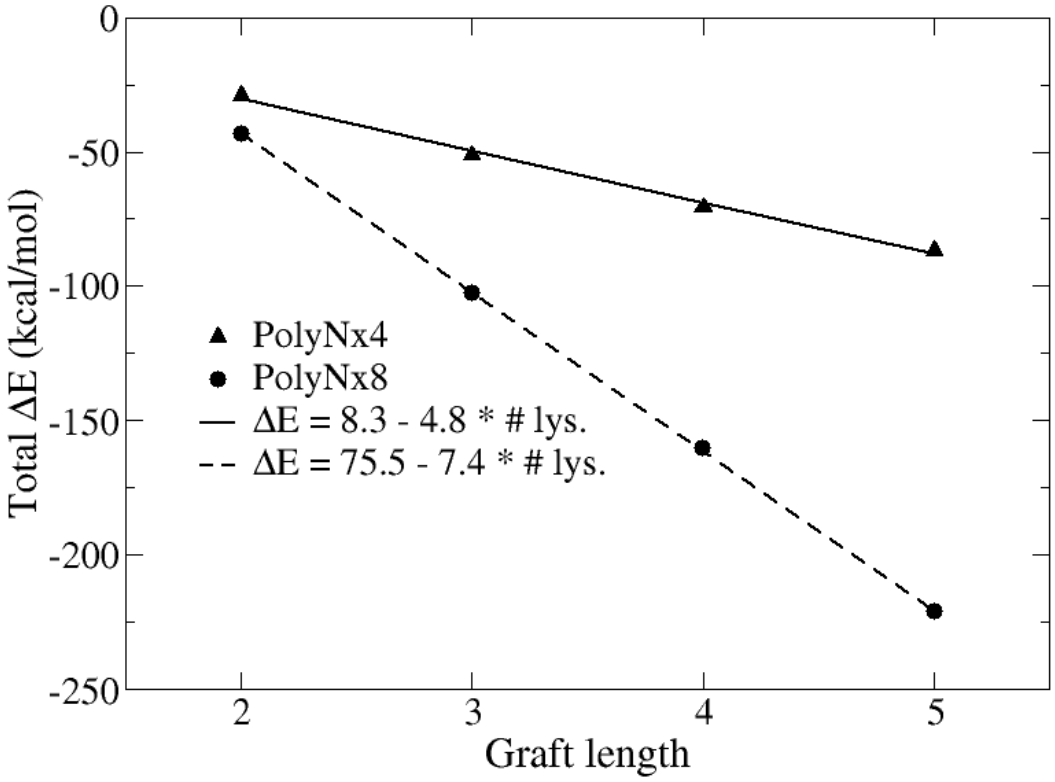


Figure 4.10 Total ΔE as a function of increasing graft length for PolyNx4 (triangles) and PolyNx8 (circles), with the solid and dashed lines showing linear fits for PolyNx4 and PolyNx8 respectively. Note that the slope of PolyNx8 appears much greater than the slope of PolyNx4 because the former has been shifted to have the same x-axis; the linear fit for the slope was, however, performed using the number of lysines in each system. As expected, total ΔE increases with graft length for both 4-mers and 8-mers because longer grafts possess greater total charge and therefore bind more strongly to the DNA. Interestingly, the slope (increase in ΔE with the number of lysines) is greater for 8-mers than for 4-mers: this demonstrates cooperative binding wherein the lysines in the 8-mer increase the binding energy of other lysines in the polycation.

4.6.7 Number and fraction of ions bound to both the polycation-DNA complex and the separate polyelectrolytes

Table 4.8 Number of ions bound to: DNA in the absence of polycation (only includes Na^+); polycation in the absence of DNA (only includes Cl^-); the polycation-DNA complex (includes both Na^+ and Cl^-); and the difference (delta) between the bound and unbound states. Although the DNA oligomers are all 16bp in length, they bind different numbers of counterions in different systems, due to the differing total number of counterions in each system. There is no discernible trend in the polycation-only systems, except that larger polycations bind more ions. In the bound state there appears to be no trend in the number of bound counterions. In all cases, there is a net loss of bound ions, resulting in an increase in the entropy, and the delta value very roughly correlates with $T\Delta S_{ion}$, with a greater number of displaced ions resulting in a greater entropic contribution (see Tables 4.5, 4.6, and 4.9).

System	DNA		Polycation		Complex		Delta	
	Mean	Std. err	Mean	Std. err	Mean	Std. err	Mean	Std. err
1xPLL8	9.71	0.06	0.72	0.05	5.92	0.06	-4.51	0.10
1xPLL12	9.02	0.07	2.04	0.06	8.43	0.08	-2.63	0.12
1xPLL16	7.47	0.07	2.81	0.06	7.36	0.07	-2.93	0.12
1xPLL20	7.23	0.04	3.87	0.04	6.37	0.06	-4.73	0.08
4xPLL2	7.47	0.07	0.43	0.04	5.56	0.06	-2.35	0.10
4xPLL3	7.86	0.06	1.06	0.06	5.31	0.05	-3.61	0.10
4xPLL4	7.51	0.08	1.72	0.08	3.61	0.04	-5.62	0.12
4xPLL5	6.72	0.05	2.69	0.05	4.55	0.05	-4.86	0.09
Poly2x4	9.24	0.05	0.38	0.07	6.37	0.06	-3.25	0.10
Poly3x4	8.60	0.05	0.65	0.05	7.21	0.07	-2.03	0.10
Poly4x4	8.00	0.05	1.78	0.08	8.22	0.08	-1.57	0.12
Poly5x4	7.71	0.04	2.40	0.06	6.53	0.07	-3.59	0.10
Poly2x8	8.85	0.06	1.14	0.07	7.52	0.08	-2.46	0.12
Poly3x8	8.02	0.06	2.27	0.05	7.98	0.08	-2.31	0.11
Poly4x8	6.58	0.06	4.67	0.05	8.63	0.09	-2.62	0.12
Poly5x8	6.18	0.07	7.48	0.07	4.47	0.04	-9.19	0.11

Table 4.9 Fraction of ions bound to: DNA in the absence of polycation (only includes Na^+); polycation in the absence of DNA (only includes Cl^-); the polycation-DNA complex (includes both Na^+ and Cl^-); and the difference (delta) between the bound and unbound states. The DNA-only systems, despite binding different numbers of ions, show less sensitivity in the fraction of bound ions. Larger polycations are able to bind a greater fraction of the counterions, as expected. As with the number of bound counterions, the fraction displays no obvious trends in the complex. In all cases, there is a net loss of bound ions, resulting in an increase in the entropy. The delta value very roughly correlates with $T\Delta S_{ion}$.

System	DNA		Polycation		Complex		Delta	
	Mean	Std. err	Mean	Std. err	Mean	Std. err	Mean	Std. err
1xPLL8	0.32	0.02	0.09	0.01	0.21	0.01	-0.20	0.01
1xPLL12	0.30	0.01	0.17	0.01	0.34	0.02	-0.13	0.01
1xPLL16	0.25	0.01	0.18	0.01	0.29	0.01	-0.14	0.01
1xPLL20	0.28	0.01	0.24	0.01	0.28	0.02	-0.24	0.01
4xPLL2	0.25	0.01	0.05	0.01	0.19	0.01	-0.11	0.01
4xPLL3	0.26	0.01	0.09	0.01	0.20	0.01	-0.15	0.01
4xPLL4	0.25	0.01	0.11	0.01	0.13	0.01	-0.23	0.01
4xPLL5	0.22	0.01	0.13	0.01	0.17	0.01	-0.19	0.01
Poly2x4	0.27	0.01	0.05	0.01	0.20	0.01	-0.12	0.01
Poly3x4	0.27	0.01	0.06	0.01	0.24	0.01	-0.09	0.01
Poly4x4	0.27	0.01	0.15	0.01	0.32	0.02	-0.09	0.01
Poly5x4	0.26	0.01	0.15	0.01	0.25	0.01	-0.15	0.01
Poly2x8	0.23	0.01	0.07	0.01	0.22	0.01	-0.09	0.01
Poly3x8	0.24	0.01	0.11	0.01	0.27	0.01	-0.08	0.01
Poly4x8	0.22	0.01	0.19	0.01	0.32	0.02	-0.10	0.01
Poly5x8	0.24	0.01	0.27	0.02	0.16	0.01	-0.34	0.02

4.6.8 Curvature of the DNA helical axis caused by polycation binding

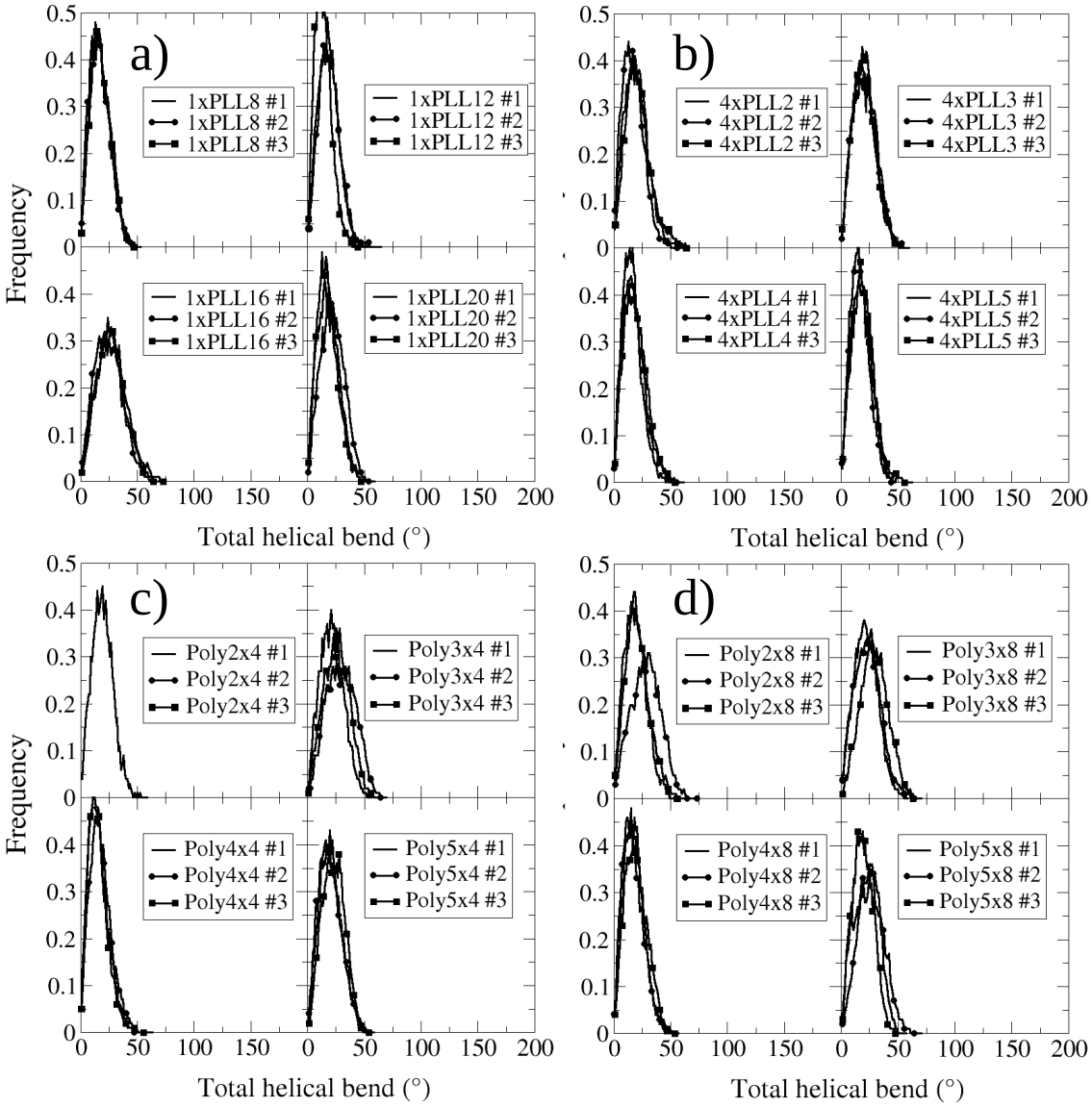


Figure 4.11 Curvature of the DNA helical axis caused by binding of various polycations: a) long linear PLL (length 8 to 20 lysines); b) short linear PLL (length 2 to 5 lysines, 4 oligomers); c) PolyNx4 (N = 2 to 5); and d) PolyNx8 (N = 2 to 5). Three trials of each system are shown to illustrate that bending of the DNA does not always occur but that several of the polycations are capable of inducing curvature. Specifically, Poly2x8 induces significant curvature in one trial; Poly5x8 induces significant curvature in two trials; Poly3x4 appears to induce some curvature in one or two trials; and PLL12 induces some curvature in one trial. Larger polycations can bend DNA more easily because they can interact with, say, both the top and the bottom of the DNA oligomers, thereby pulling the two ends together and bending the DNA. This does not occur for all trials because the polycation does not always bind in a manner (e.g. interacting with both ends of the DNA) that is conducive to bending.

4.6.9 Effect of architecture on the individual contributions to binding energy

Table 4.10 Details of the effects of architecture on binding energy. Each quadrant of the table shows the details long linear PLL, short linear PLL, and grafted PLL for systems with the same total number of lysines (from right to left and top to bottom: 20, 16, 12, and 8). All units are kcal/mol.

	1xPLL20		**4xPLL5**		**Poly5x4**		**1xPLL16**		**4xPLL4**		**Poly4x4**	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
BOND	2.4	2.7	3.3	2.4	-1.8	2.5	2.9	2.4	-3.1	2.3	-0.8	2.6
ANGLE	0.1	3.5	-1.1	3.2	-1.6	3.7	-0.6	3.4	-1.5	3.2	4	3.4
DIHED	3.8	1.9	3.8	2	11.1	1.9	-0.2	2	2	1.9	-4.7	1.9
VDWAALS	-60.1	2.6	-59.1	1.9	-30.6	2.2	-47	2.6	-52	1.7	-32.9	2.2
EEL	-9107.3	142.4	-9857.8	35.9	-7327.9	162.4	-7942	84.7	-8135.1	48.9	-5665.9	71.8
EPB	8989.7	140.3	9718	34.9	7223	160.3	7844.3	83.8	8030.7	48.3	5595.8	71.2
1-4VDW	2.3	1.1	-0.4	1.1	-0.2	1.1	1.2	1.1	0.1	1.1	1.2	1.2
1-4EEL	8.9	3.7	4.1	3.8	8.7	3.6	7	3.3	-7.3	3.4	4.3	3.8
RESTRAINT	0	0	0	0	0	0	0	0	0	0	0	0
ECAVITY	-48.7	1.5	-59.5	0.9	-26.3	1.6	-39.6	1.5	-48.8	1	-26.9	1.3
EDISPER	93.4	2.5	108.1	1.5	58.8	3	72.7	2.7	89.9	1.5	55.4	1.9
Total ΔE	-115.5	5.1	-140.6	4.3	-86.7	4.9	-101.4	4.7	-125	4.7	-70.5	4
TΔS: NMA Trans	-14.5	0	-14.5	0	-14.7	0	-14.4	0	-14.4	0	-14.5	0
TΔS: NMA Rot	-14.2	0.1	-16.1	0	-14.1	0	-13.8	0.1	-15.8	0.1	-13.7	0
TΔS: NMA Vib	-27.1	3.2	-64.8	3.6	-11.1	3.5	-17.9	2.8	-50.1	2.9	-8.4	2.6
TΔS: NMA Total	-55.8	3.2	-95.4	3.6	-39.9	3.5	-46.1	2.9	-80.2	2.9	-36.6	2.7
TΔS: Conform.	-7.7	6.6	-14.7	2.8	-3.9	6.3	-17.2	10.8	-6.1	0.4	-6.8	9
TΔS: Ion dissoc.	25.8	0.2	26.6	0.2	19.1	0.2	16.1	0.2	29.7	0.2	10.8	0.2
Total TΔS	-37.7	9.3	-83.5	5.4	-24.8	9.3	-47.2	13	-56.6	3.2	-32.6	11.2
Total ΔG	-77.7	10.6	-57	6.9	-61.9	10.5	-54.1	13.8	-68.4	5.7	-37.9	11.9
Phos	11.2	1.4	15.6	1.8	10.2	1.9	8.8	1.6	12.8	2.0	8.3	1.7
Nitro	11.5	1.4	14.7	1.9	9.5	1.9	8.8	1.7	12.5	2.0	7.6	1.6
Total	22.7	2.0	30.3	2.6	19.7	2.7	17.6	2.3	25.4	2.8	15.9	2.3
ΔG/P	-7	0.9	-3.7	0.4	-6.1	1	-6.1	1.6	-5.3	0.4	-4.6	1.4
ΔG/N	-6.7	0.9	-3.9	0.5	-6.5	1.1	-6.2	1.6	-5.5	0.5	-5	1.6
ΔG/T	-3.4	0.5	-1.9	0.2	-3.1	0.5	-3.1	0.8	-2.7	0.2	-2.4	0.7
	1xPLL12		**4xPLL3**		**Poly3x4**		**1xPLL8**		**4xPLL2**		**Poly2x4**	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
BOND	1.1	2.3	0.3	2.4	-3.1	2.6	3.2	2.1	-0.9	2.3	-0.4	2.3
ANGLE	1.1	3.5	4.7	3.2	-1.5	3.4	0.1	3.6	3.4	2.8	-3.6	3.3
DIHED	5.1	1.6	-3.0	2.0	0.8	2	1.8	1.8	1.7	1.6	-3	2
VDWAALS	-30.4	4.9	-49.6	2.0	-34.7	2.3	-19.1	2.0	-35.1	1.9	-13.1	2.6
EEL	-5716.5	124.2	-6446.2	24.2	-3435.7	84.2	-4308.8	25.2	-4274.1	25.6	-1838	54.9
EPB	5641.5	124.8	6356.1	23.9	3389.7	83.1	4253.6	24.7	4211.9	25.1	1820.5	54.8
1-4VDW	1.7	1.0	0.1	1.0	1.2	1.2	-0.8	0.9	1.2	1.0	0.3	1.1
1-4EEL	-5.2	3.8	4.9	3.0	7.7	3	-2.2	3.8	1.6	3.7	-4.5	3.4
RESTRAINT	0.0	0.0	0.0	0.0	0	0	0.0	0.0	0.0	0.0	0	0
ECAVITY	-28.2	3.3	-46.8	0.8	-31.1	1.5	-19.3	0.8	-34.0	0.8	-10.9	1.7
EDISPER	52.5	6.3	86.2	1.3	55.6	2.7	35.9	1.3	58.2	1.3	23.8	3.2
Total ΔE	-77.3	7.0	-93.4	4.4	-51.2	4.3	-55.6	4.0	-65.9	4.1	-28.8	4.8
TΔS: NMA Trans	-14.1	0.0	-14.2	0.0	-14.4	0	-13.8	0.0	-13.9	0.0	-14.2	0
TΔS: NMA Rot	-13.3	0.1	-15.5	0.1	-13.6	0.1	-12.7	0.0	-15.4	0.1	-13.1	0
TΔS: NMA Vib	-4.5	4.1	-39.8	4.0	-8.7	4	-4.2	1.5	-36.6	3.6	-0.7	2.4
TΔS: NMA Total	-32.0	4.1	-69.4	4.0	-36.7	4	-30.7	1.5	-65.8	3.6	-28	2.4
TΔS: Conform.	-1.5	6.4	-1.5	1.4	-9	4.7	5.8	2.4	0.9	0.3	1.5	2.2
TΔS: Ion dissoc.	12.9	0.2	19.0	0.2	12.9	0.2	22.1	0.2	12.2	0.2	18.9	0.2
Total TΔS	-20.6	7.6	-51.9	4.2	-32.8	7.8	-2.8	2.9	-52.7	3.6	-7.6	4.5
Total ΔG	-56.7	10.3	-41.5	6.1	-18.4	8.9	-52.8	4.9	-13.2	5.5	-21.2	6.6
Phos	8.6	1.6	10.8	2.0	6.1	1.8	6.7	1.3	8.6	1.7	3.7	1.5
Nitro	7.8	1.5	10.4	2.0	5.1	1.6	6.2	1.4	7.6	1.6	3.0	1.4
Total	16.4	2.2	21.2	2.8	11.2	2.4	12.9	1.8	16.1	2.4	6.7	2.0
ΔG/P	-6.6	1.2	-3.8	0.6	-3	1.5	-7.8	0.7	-1.5	0.6	-5.7	1.8
ΔG/N	-7.3	1.3	-4.0	0.6	-3.6	1.7	-8.6	0.8	-1.7	0.7	-7.1	2.2
ΔG/T	-3.5	0.6	-2.0	0.3	-1.6	0.8	-4.1	0.4	-0.8	0.3	-3.2	1

4.6.10 Effect of architecture on binding mode

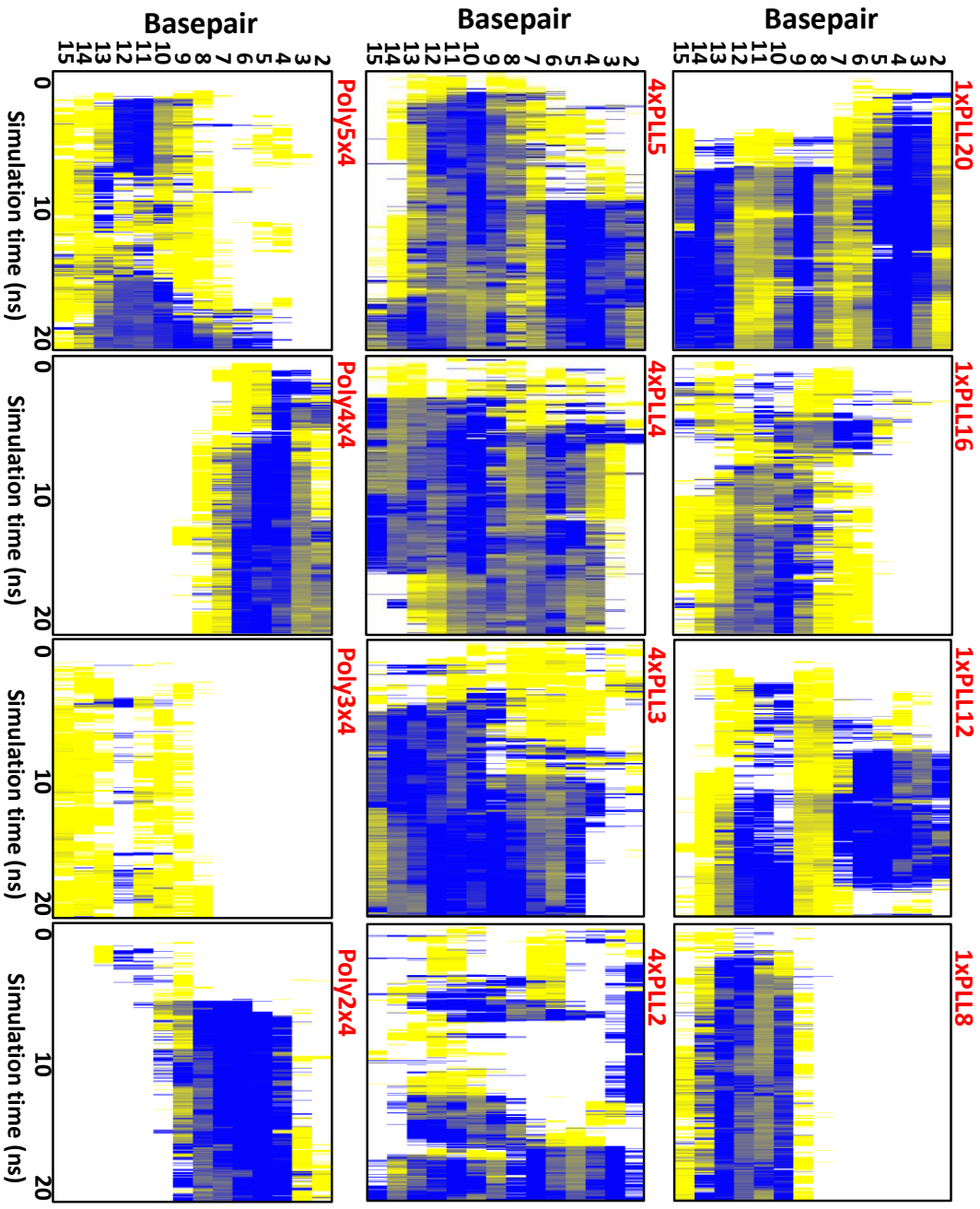


Figure 4.12 Binding maps showing the effect of architecture. Each column shows long linear PLL, short linear PLL, and grafted PLL (from top to bottom) with a given total number of lysines (20, 16, 12, 8 from left to right). The y-axis corresponds to individual basepairs; the terminal basepairs are omitted. The x-axis corresponds to simulation time. Long linear PLL tends to display an ordered pattern of binding. For instance, we can see that in these particular trials, PLL20 and PLL12 bound roughly parallel to the helical axis ("zipping") evidenced by the yellow stripes corresponding to the DNA backbones. Also in this set of trials, PLL16 and PLL8 happened to wrap around the DNA through one of the grooves ("winding"); this

is not as obvious from these plots, however. In contrast, short linear PLL displays no particular patterns, but is constantly moving, forming and reforming contacts with all portions of the DNA. PolyNx4 also displays somewhat ordered binding patterns, but no particular modes (e.g. zipping, winding) are apparent.

4.6.11 Effect of architecture on the binding mode: additional trials

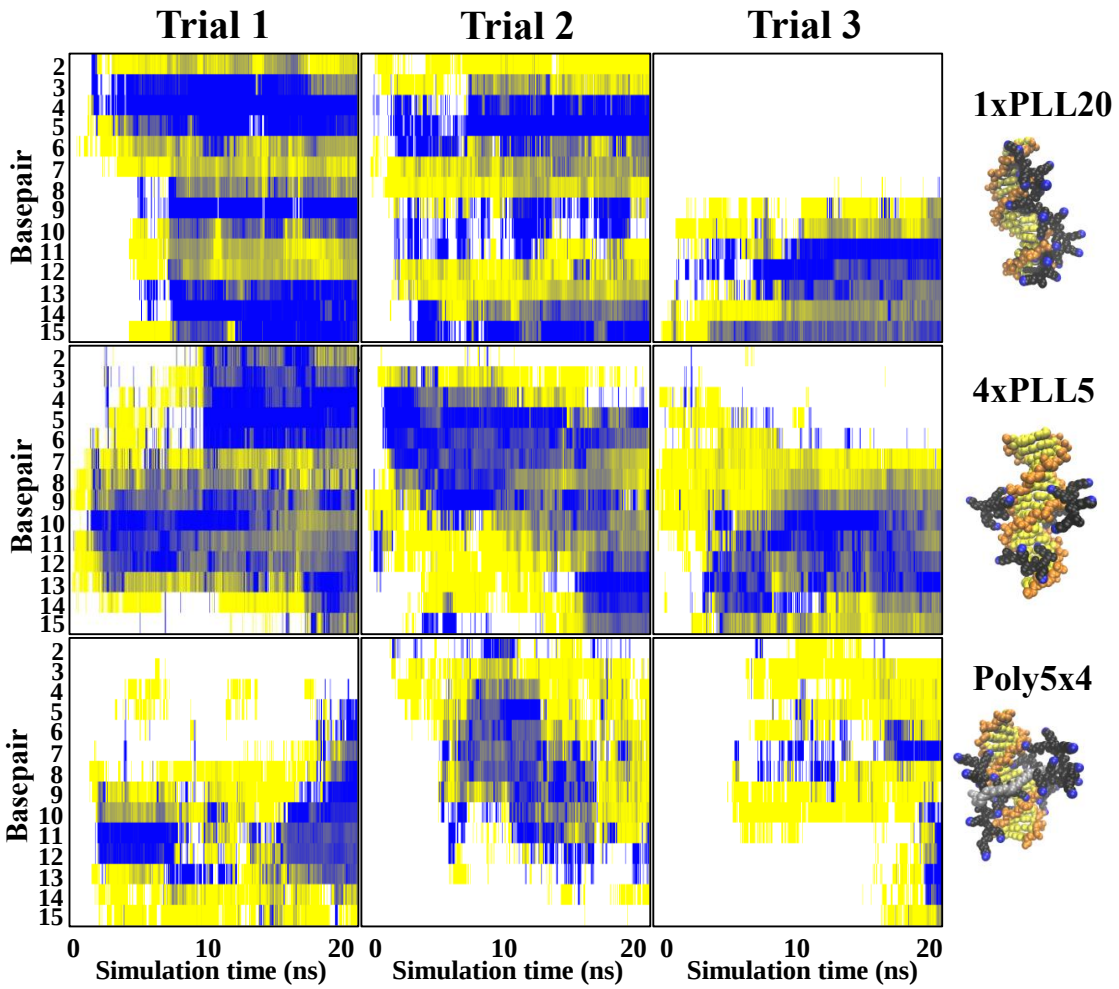


Figure 4.13 Binding maps showing the effect of architecture for three trials of systems with 20 total lysines. The rows, from top to bottom, are long linear PLL, short linear PLL, and grafted PLL.

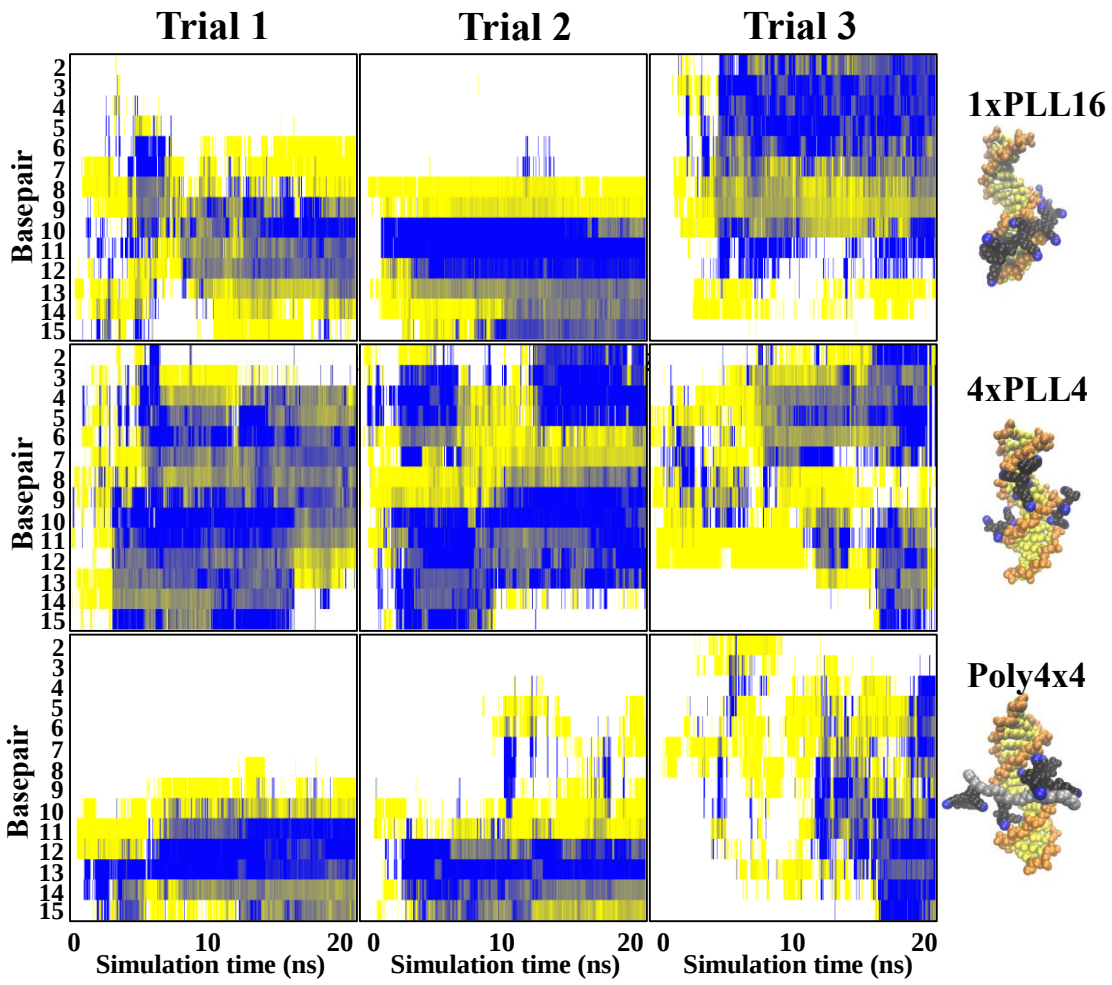


Figure 4.14 Binding maps showing the effect of architecture for three trials of systems with 16 total lysines. The rows, from top to bottom, are long linear PLL, short linear PLL, and grafted PLL.

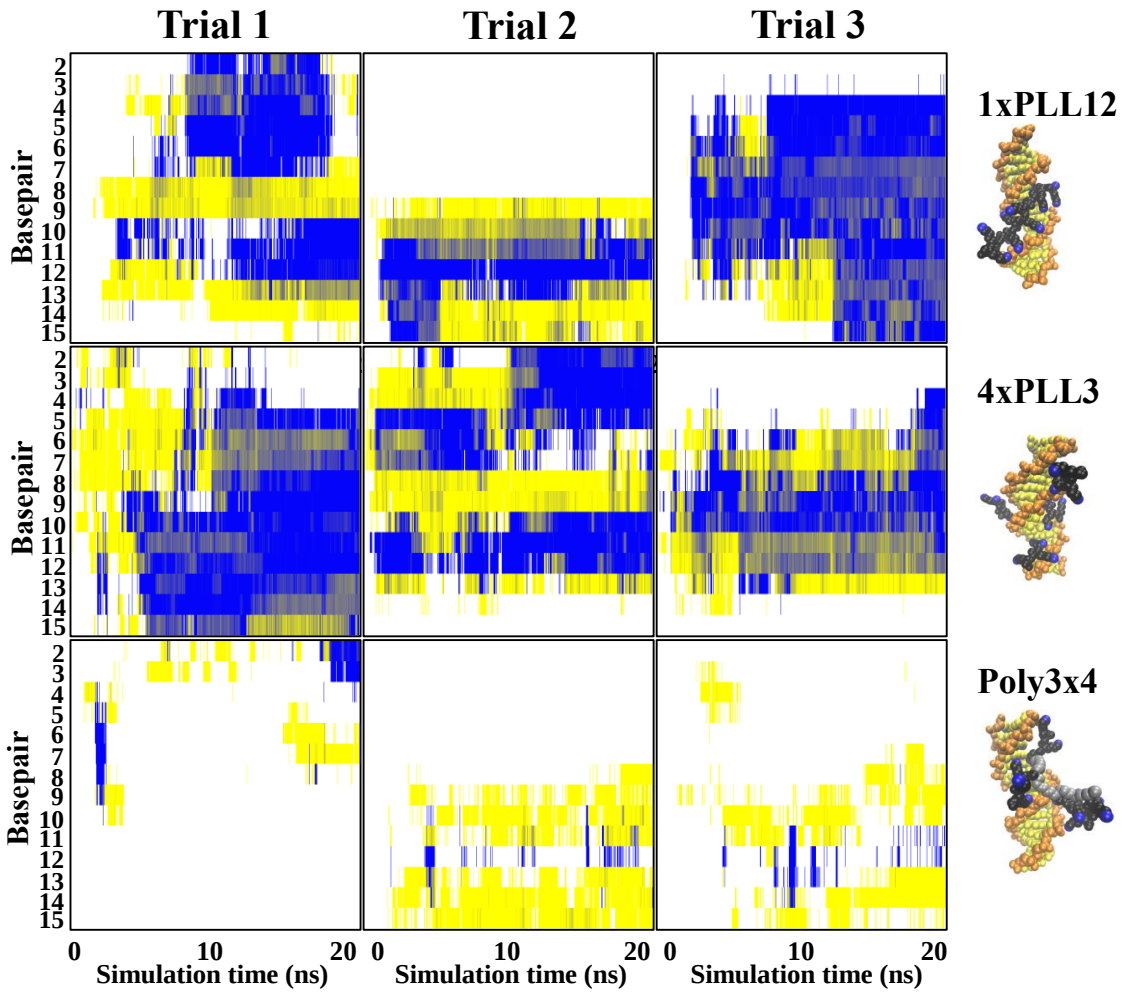


Figure 4.15 Binding maps showing the effect of architecture for three trials of systems with 12 total lysines. The rows, from top to bottom, are long linear PLL, short linear PLL, and grafted PLL.

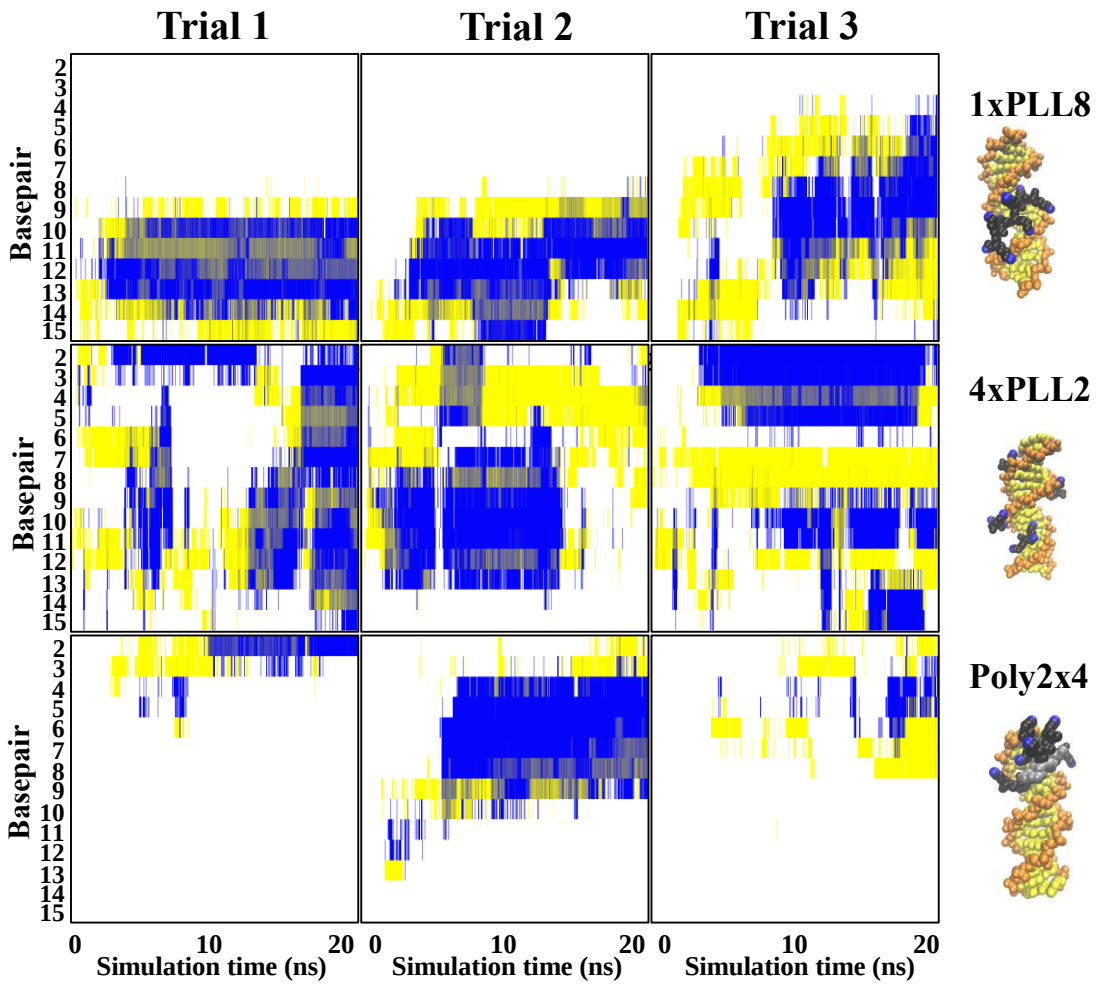


Figure 4.16 Binding maps showing the effect of architecture for three trials of systems with 8 total lysines. The rows, from top to bottom, are long linear PLL, short linear PLL, and grafted PLL. Interestingly, in Trial 3 of 4xPLL2 we observe a long-lasting contact (yellow patch at basepairs 7-8) in which one of the polycations remains stably bound between 2-3 phosphate groups for around 10 ns. A similar situation occurs in Trial 2, but the large yellow patch is caused by two polycations drifting around between the backbones near basepairs 3-6.

Chapter 5

Coarse-grained simulation studies of effects of polycation architecture on structure of the polycation and polycation-polyanion complexes

Adapted from: *Macromolecules* **45** 8083 (2012)

5.1 Introduction

Polyelectrolytes are chains of ionizable monomer units that impart periodic positive or negative charge along the chain in solution. The physical and chemical behavior of polyelectrolytes is quite complex due to interactions between these charges over multiple length scales. Understanding this complex behavior is of interest both from a fundamental and a practical standpoint. In particular, from a practical standpoint, polycations have been used as carriers, or vectors, for therapeutic gene delivery and are viewed as a preferred alternative to viral vectors, as polycations lack the immunogenic responses of the latter [1, 2]. By forming compact, positively-charged complexes (polyplexes) with polyanionic DNA or RNA, polycations allow therapeutic genetic information to cross the cell membrane to be integrated into the genome [1, 2]. A great deal of research effort has therefore gone towards synthesizing, and characterizing polycations that maintain biocompatibility while approaching the high level of transfection efficacy inherent to viral gene delivery vectors [3-7].

Previous experimental work has largely focused on designing polycations with novel chemistries and architectures to improve transfection efficiency [3, 5-8]. Many polycations have been synthesized using combinatorial and rational approaches by varying the chemistry (e.g., polylysine [4], polyamidoamine [9]) and architecture (e.g., linear [3], branched [10], comb-like [6]) of polycations, and by introducing biocompatiblizing copolymers or proteins into the formulations, such as polyethylene

glycol [11] or an oligopeptide designed to target the nucleus [6, 7]. One class of polycations recently synthesized and characterized by Emrick and coworkers shows striking architecture-dependent properties [5-7]. By reorganizing the structure of linear poly-L-lysine (PLL) into a comb-like architecture of oligolysine grafted to a polyolefin backbone, poly(cyclooctene-*g*-oligolysine), the transfection efficiency was increased 50-fold over PLL, reaching levels that are comparable to the best commercially-available transfection reagents [7]. Additionally, the stability of the polyplexes and transfection efficiency varied non-monotonically with the length of the oligolysine grafts: polycations with intermediate-length grafts produced polyplexes with substantially higher transfection efficiency than polycations with either long or short grafts [7]. In recent work, we sought to explain the molecular and thermodynamic reasons behind the improved transfection efficiency of the grafted architecture over the linear architecture using atomistic molecular dynamics simulations and calculations of binding free energy [12]. We found that the polyolefin backbone in the grafted architecture weakens the free energy of polycation-DNA binding relative to linear PLL, which by improving polyplex dissociation within the cell could partially account for the improved transfection efficiency of the grafted architecture over the linear architecture. We also matched the experimental trends in transfection efficiency as a function of architecture/graft length to trends in computationally obtained free energy of polycation-DNA binding as a function of architecture/graft length [12]. The non-monotonic variation is free energy of binding with increasing graft length stems from a non-monotonic trend in the entropy of binding, dominated by the counterion dissociation entropy [12].

While the atomistic simulations were effective for studying the fine details of polycation-DNA binding, the atomistic resolution comes at the expense of limited length and time scales, making it infeasible to study the structure of full-sized polyplexes. In the present study, we have leveraged our knowledge of the atomistic behavior of the lysine-based polycations into a coarse-grained (CG) model where less-important details of the system are subsumed into fewer degrees of freedom, or ‘beads’ that represent several atoms, allowing the study of greater length and time scales with reasonable

computational expense. Using these coarse-grained models, we study how the architecture and flexibility of polycations affect the structure of polyplexes, which is not feasible with atomistic simulations.

There are some previous theory and simulation studies of polyelectrolytes that have examined the effect of architecture and flexibility on the structure of isolated polyelectrolytes and polyelectrolyte complexes. Linear uncharged polymers and charged polyelectrolytes of varying flexibility and charge density have been widely-studied, both isolated in solution and forming complexes with oppositely-charged polyelectrolytes [13-16]. A number of studies of grafted neutral polymers and charged polyelectrolytes (also referred to as comb-like or bottle-brush polymers depending on grafting density) have been reported, showing the effects of rigidity of various components [17, 18], spacing and length of the grafts [19-21], the size of the monomers in the grafts [22], and the hydrophobicity of the grafted backbone [23, 24]. The effects of counterions on polyelectrolytes of varying architecture, both isolated in solution and complexed with an oppositely-charged polyelectrolyte, have been explored [13, 23, 25-28]. Most of these previous coarse-grained computational studies have focused on *model* polyelectrolytes with architectures and interactions that are not connected to a specific chemical structure. In this paper we have sought to generate a simplified CG model that maintains several key features of three specific polycations: 1) linear poly-L-lysine (PLL), 2) poly(cyclo*octene-g*-oligolysine) synthesized by Emrick and coworkers [5-7] which we denote as PCO-PolyX with X being the graft length, and 3) poly(cyclo*pentene-g*-oligolysine) denoted as PCP-PolyX with X being the graft length. Specifically, the polycation CG model accurately mimics the relative sizes of the CG beads, the structure of the lysine monomers, and the graft length and graft spacing. The intrinsic rigidity of the peptide backbone in PLL and the oligolysine grafts is also accounted for, relative to the intrinsic flexibility of the grafted polyolefin backbone. For this study, the polyanion is a linear negatively-charged chain of variable rigidity, modeled to mimic features of double stranded DNA with each monomer representing a base pair.

We find using Langevin dynamics simulations with the above CG models how the architecture of the polycation – linear versus grafted, varying graft length, and varying graft spacing – affects the size, shape and flexibility of the polycation and the polyplex formed with polyanions of varying flexibility.

First, we study the linear and grafted architectures described above both in a charged state and neutral state to isolate effects inherent to the architecture and effects due to electrostatic repulsion between adjacent grafts on the overall structure and flexibility of the polycation. Removing the charges from the linear semi-flexible polycation results in a compact and flexible polymer, in agreement with what has been demonstrated previously for model polycations [14, 29]. In the case of the grafted architecture, either increasing graft length or decreasing graft spacing increases the relative size and rigidity of the molecule. Removing the electrostatic effects, by eliminating charges from the grafted polycation, greatly reduces these architectural effects because of the greater range of electrostatic repulsion as compared to excluded volume repulsion. For both linear and grafted architectures, increasing the ionic strength or counterion valency can effectively neutralize the polycation charge and lead to behavior resembling that of the uncharged polymer, eliminating many architecture-dependent effects.

The size and shape of polyplexes formed by these lysine-based polycations with linear polyanions depend strongly on the polycation architecture, polyanion flexibility, and the positive-to-negative charge ratio (CR). In particular, polyplexes formed at charge neutrality (i.e., CR = 1, where the polycation charge exactly balances the polyanion charge) are small and globular in case of fully-flexible polyanion but are large and rod-like in case of semi-flexible, DNA-like, polyanion. For CR > 1, the polyplexes assume large, extended conformations for grafted architecture, but compact toroidal structures are observed for polycations with the longest grafts. The surface charge of the polyplexes is largely independent of the polyanion flexibility and depends mostly on polycation architecture and charge ratio: at CR = 1, all polyplexes possess neutral surface charge, but for CR > 1, the surface charge of the polyplex increases with increasing charge ratio, increasing graft length, or decreasing graft spacing. These results provide guidelines on how to tune the polycation architecture to modulate polyplex charge, size, and shape, which are important for processes like gene delivery involving polyplex transport into cells.

The remainder of this paper is organized as follows. In section 2, we describe the design and parameterization of our coarse-grained models, our simulation protocol, and the analysis methods we employ. In section 3.1, we characterize the behavior of the CG polycations in solution elucidating how

polycation architecture, ionic strength and counterion valency affect structure and flexibility of isolated polycations. In section 3.2, we characterize polycation-polyanion complexes (polyplexes) and relate polyplex size, shape and surface charge to polycation architecture and the charge ratio. We conclude with a summary of results and implications of this work on future applications.

5.2 Methods

5.2.1 Model definition

We build and parameterize a coarse-grained model of lysine based polymers in linear and grafted architecture by reproducing the key architectural features (the size, length, and spacing, and the flexibility) of various components in the polycations seen in our previous atomistic simulations [12] employing the Amber force field [30]. Additional atomistic simulations of poly(cyclopentene-*g*-oligolysine), beyond those included in our previous publication, also provide the basis for parameterizing our coarse-grained models. While the atomistic simulations used explicit TIP3P-model [31] water molecules, the coarse-grained model in this work takes into account the effects of water in an implicit manner and by doing so greatly reduces computational expense. We provide an overview of the coarse-grained model here, and additional details of the model development in supplementary section 5.6.1.

We model linear poly-L-lysine (PLL) using three coarse-grained beads per lysine monomer (see Figure 5.9 in supplementary section 5.6). The first bead represents the peptide link group (–C(=O)NH–). The two remaining beads represent the lysine α-carbon and R-group ($–CH(CH_2)_4NH_3^+$): the second bead represents the α-carbon, β-carbon, and γ-carbon atoms ($–CH(CH_2)_2–$), and the third bead represents the δ-carbon, ε-carbon, and charged amine group ($–(CH_2)_2NH_3^+$), as shown schematically in Figure 5.9 and Figure 5.10 (see supplementary section 5.6). All hydrogen atoms are included in the bead that includes their parent heavy atom. Mapping the atoms in this manner results in approximately equal mass and size for the resulting CG beads, as shown in Table 1 (supplementary section 5.6), and maintains the correct ratio of lysine monomer size to the spacing of the monomers, as shown in Figure 5.11 (supplementary section 5.6). To model the negatively-charged C-terminal ($–COO^-$) and the positively-charged N-terminal

($-NH_3^+$) of the peptide backbone, a negatively-charged bead is placed at one end of the peptide backbone and a positively-charged bead is place on the other end. The atoms comprising the C-terminal and N-terminal beads also have approximately the same mass and size as the other beads. To mimic the inherent rigidity of the peptide backbone, we calibrate a harmonic bending constant (k_{angle}, defined below) to impose an appropriate energetic bending penalty on the peptide backbone. We find that a value of k_{angle} = 30 energy/radian^2 best reproduces the atomistic results; the details of the procedure used to calibrate k_{angle} are available in Figure 5.12 (supplementary section 5.6).

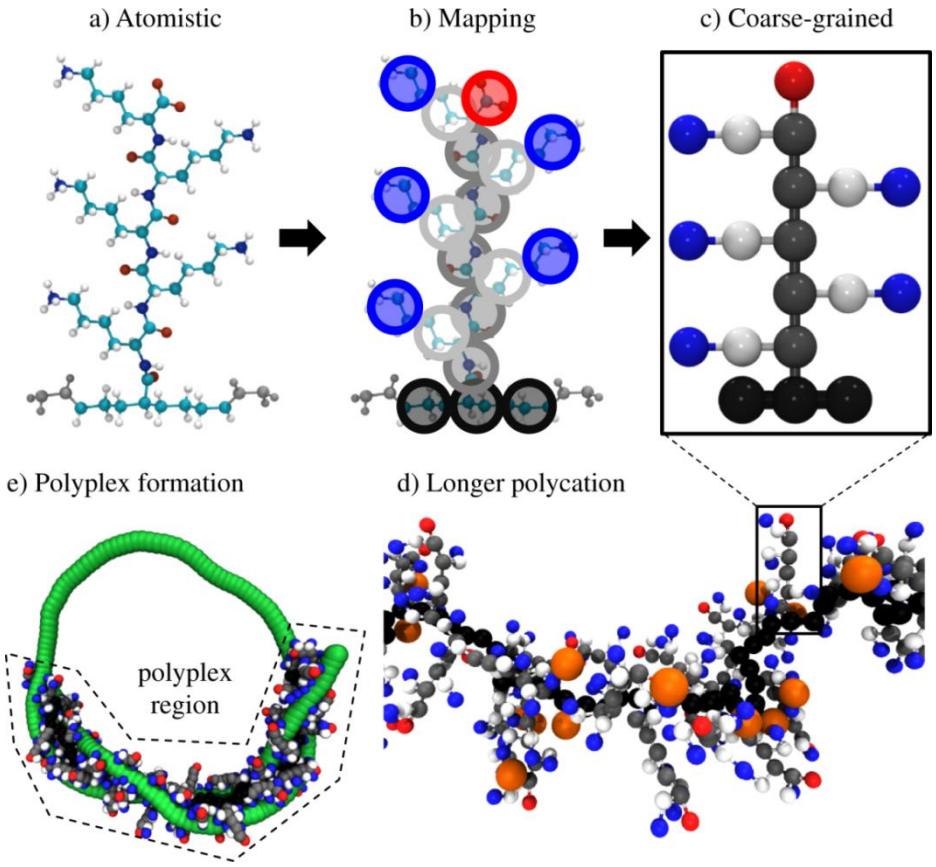


Figure 5.1 Schematic of the coarse-grained polycation model. The atomistic structure (a) pictured is poly(cyclooctene-g-pentalysine), PCO-Poly5. The overlay of transparent circles in (b) qualitatively demonstrates the correspondence of the atomistic structure to the final coarse-grained structure (c). A representative snapshot of a PCO-Poly5, including counterions (larger spheres), is shown in (d). An example snapshot of a polyplex and a diagram of the polyplex region are shown in (e). Figure 5.9 (in supplementary section 5.6) shows a similar diagram of our model of linear poly-L-lysine.

We also create coarse-grained models for two grafted polycations, poly(cyclopentene-*g*-oligolysine) and poly(cyclooctene-*g*-oligolysine) (Figure 5.1). Throughout this work, we abbreviate the

chemical name poly(cyclopentene-*g*-oligolysine) to PCP-PolyX, where X indicates the number of lysines in the oligolysine grafts; similarly, we abbreviate poly(cyclooctene-*g*-oligolysine) to PCO-PolyX. The only difference between the two grafted polycations is the number of carbon atoms in the backbone: PCP-PolyX has 5 carbon atoms, while PCO-PolyX has 8 carbon atoms. In the coarse-grained model of these grafted polycations, the three CG beads in the oligolysine grafts are the same as those in linear lysine described above, and the fourth bead type represents the polyolefin ($–(CH_2)_N–$) backbone. The atoms comprising each polyolefin backbone bead depend on the chemical structure of the backbone: for the pentene backbone in PCP-PolyX the five methylene groups ($–(CH_2)_5–$) are divided equally among two identical CG beads, while for the octene backbone in PCO-PolyX the eight methylene groups ($–(CH_2)_8–$) are divided equally among three identical CG beads. Regardless of the length of the polyolefin backbone, the properties of the polyolefin backbone beads are identical. The mass and size of the polyolefin backbone bead are similar to the beads defining the linear lysine model. Choosing three polyolefin backbone beads to represent octene and choosing two beads to represent pentene maintains the ratio of graft length to graft spacing seen in atomistic model, as shown in Table S2. The polyolefin backbone is completely flexible (k_{angle} = 0 energy/radian^2) and has the same van der Waals well depth as all other species. These choices are justified based on comparing the results of extensive coarse-grained simulations with atomistic simulation results, as detailed in supplementary section 5.6.3. The oligolysine grafts are bonded to the 2nd bead of every polyolefin backbone monomer, which each consist of 2 or 3 beads as mentioned above. The grafts possess a negatively-charged C-terminal CG bead, but grafting to the polyolefin backbone eliminates the positively-charged N-terminal.

We model a DNA-like polyanion using a simplified 1-bead coarse-grained model where each bead represents a single DNA base pair. Each CG bead has a valency of 2, corresponding to the two negatively-charged phosphate groups in each base pair. The mass of these beads is chosen to be 6.3 times the unit mass because the ratio of the average mass of a DNA base pair (263 g/mol) is 6.3 times our basis for the unit mass (41.8 g/mol). The distance between the beads is chosen to replicate the distance of 3.4 Å between consecutive base pairs, and the radius of the polyanion beads is 6.7 Å, a choice we discuss in

supplementary section 5.6.4. Finally, to parameterize the stiffness of the DNA backbone, we conducted a series of simulations of an *uncharged* CG 1-bead polymer with varying bending coefficient (k_{angle}, defined below), from which we determined that a value of k_{angle} = 460 energy/radian^2 best duplicates the DNA persistence length of approximately 50 base pairs in the high salt limit [32]. Specifics of this procedure are given in Figure 5.13 (supplementary section 5.6).

Mobile ions are modeled as charged spheres. In all cases, the positive and negative counterions have the same valency, either they are both monovalent or they are both divalent. Due to the level of coarse graining, it is not possible to assign a specific chemical identity to the counterions. Several possible biologically-relevant mobile ions (e.g., Na^+, Mg^{2+}, Ca^{2+},Cl^-, SO_4^{2-}, etc.) and the rationale for choosing the mass and van der Waals parameters of the counterions, are listed in supplementary section 5.6.1.

5.2.2 Unit definitions and model interactions

All units in our simulations are dimensionless and can be connected to physical quantities through the definitions of unit mass, unit length, and unit energy. We define unit mass as 41.8 g/mol, the average mass of the atoms comprising the polycation beads, unit length σ as 3 Å, the approximate average end-to-end distance of the atoms making up the polycation beads, and unit energy ε as 0.1 kcal/mol, the Lennard-Jones well depth of the atoms making up the polycation (excluding hydrogen), as defined in the Amber force field [30].

A harmonic potential is applied between bonded coarse-grained beads in both the polycation and polyanion,

$$U_{bond} = k_{bond}(r - r_0)^2 \quad (1)$$

where k_{bond} is the equilibrium constant and r_0 is the equilibrium bond length. k_{bond} is chosen as 500 ε/σ^2 such that the standard deviation around the equilibrium distance is approximately 10%, which prevents bonds from non-physically intersecting one another. The equilibrium bond lengths are all 1 σ, except for a bond length of 1.1 σ between DNA beads. Angular constraints are applied using a harmonic potential:

$$U_{angle} = k_{angle}(\theta - \theta_0)^2 \quad (2)$$

where k_{angle} is the equilibrium constant and θ_0 is the equilibrium angle. As noted earlier, for the peptide backbone k_{angle} = 30 ε/radian^2 and for the semi-flexible, DNA-like polyanion k_{angle} = 460 ε/radian^2, and θ_0 = π in both cases. k_{angle} is zero in all other cases. Van der Waals interactions between all CG beads are described using the Lennard-Jones potential:

$$U_{LJ} = 4\epsilon_{LJ}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad (3)$$

where ε is the well depth, σ is the location of the energy minimum, and *r* is the distance between two CG beads. The well depth of ε_{LJ} = 0.6 ε is chosen for all CG beads because it produces physically reasonable behavior (e.g., the linear polycation maintains the same conformation, as in the atomistic case), where ε is the unit energy defined above. The polyolefin backbone could be considered hydrophobic based on its chemical nature, and thus might require a stronger self-attraction and higher value of ε_{LJ} to mimic solvent effects. However, we find that hydrophobicity of the polyolefin backbone is not a major factor in dictating the structure and flexibility of the polyelectrolytes in our system, and therefore the value of ε_{LJ} between the polyolefin backbone beads is also 0.6 ε. The location of the energy minimum, σ, for a given pair of beads is chosen as the sum of their radii. Electrostatic interactions are calculated using the Coulomb equation:

$$U_{Coul} = \frac{q_i q_j}{\varepsilon_r r_{ij}} \quad (4)$$

where ε_r is the relative permittivity of the implicit solvent (defined as 80 here to mimic water at 300 K), q_i is the charge of the ith particle, and r_{ij} is the distance between the particles *i* and *j*. No dihedral or improper torsion terms are used in these models.

5.2.3 Simulation protocol

We conduct Langevin dynamics simulations of the models described above using the LAMMPS package [33]. Langevin dynamics are defined by the equation:

$$F = -\nabla U(\vec{r}) - \frac{m}{\gamma}\vec{v} + \sqrt{\frac{k_B T m}{\gamma\, dt}}R(t) \quad (5)$$

The first term originates from classical dynamics (the negative gradient of the potential), the second term represents friction between the particle and the implicit solvent, and the third term represents a random acceleration term due to collisions with the implicit solvent molecules. F is the force on a particle, U is the potential (defined by the positions *r* of all the particles), m is the particle mass, γ is the friction coupling coefficient, v is the particle velocity, k_B is the Boltzmann constant, T is the temperature, and R is a random number function with Gaussian distribution [34].

Our simulation protocol consists of two steps. First, the energy of the initial system configuration is minimized to a root-mean-square (RMS) gradient of 10^{-4} to remove any highly unfavorable contacts, which typically requires less than 1000 steps of conjugate gradient minimization. Then, after assigning initial temperature based on a Boltzmann distribution around the desired temperature, Langevin dynamics are carried out for 10^7 time steps, which we find is sufficient time for these systems to reach and sample their equilibrium state. The particle-particle particle-mesh (PPPM) method is used to calculate the electrostatic interactions, with a real space cutoff of 20 σ, a tolerance of 10^{-2}, and an interpolation order of 2 [35]. The parameters for the PPPM method, which is closely related to the particle mesh Ewald (PME) method, are chosen based on performance considerations and by showing that energy is conserved with these choices aside from fluctuations due to the Langevin thermostat. Van der Waals interactions are truncated at 10 σ and the CHARMM switching function is applied up to 12 σ. The Langevin friction coefficient, γ, is 5 (in dimensionless time units), and the dimensionless temperature, T*, is 6, which roughly corresponds to 300 K. A physically reasonable value for the damping coefficient is estimated by assuming a diffusion coefficient of 5×10^{-7} cm^2/s, roughly corresponding to a large DNA molecule in water, and using the relation $\gamma = \frac{Dm}{k_B T}$, where D is the diffusion coefficient, m is the particle mass, and k_BT is the thermal energy; with our bead mass of 41.8 g/mol, this calculation results in the drag coefficient of 5. The time step is 0.0075 dimensionless time units. The box size is $(100\ \sigma)^3$ for isolated polyelectrolytes and $(300\ \sigma)^3$ for systems modeling polycation-polyanion condensation, and the boxes are periodic in three dimensions.

Initial configurations of the polycations and polyanions are generated in helical configurations with parameters of the helices chosen to make the initial configurations relatively compact and to avoid undesirable steric clashes at the beginning of a simulation. Note that this helical configuration is merely for convenience and is unrelated to the structure of DNA. Counterions are placed randomly near the polyelectrolytes. In systems including both polycation and polyanion, the points for generating the helical polyelectrolyte initial structures are placed approximately 150 σ apart so that the polyelectrolytes are initially as far apart as possible.

5.2.4 Parameters varied

For the linear architecture, we simulated polycations with following number of monomers: 24, 32, 40, 48, 72, 96, 120, 144, 196, and 240. For the grafted architecture, we simulated polycations with all combinations of the following parameters: graft lengths of 2, 3, 4, or 5 lysines; graft spacings of 2 (PC**P**-PolyX) or 3 (PC**O**-PolyX); and polycations with 8, 24, 48, or 72 monomers. All of these polycations are simulated at multiple ionic strengths: neutralizing (i.e., with only the number of counterions required to neutralize the polycation charge, which we denote as 'neutral'), 50 mM, and 100 mM. Counterions are either monovalent or divalent. We also conducted simulations of all of these polycation architectures with all beads having no charge and in the absence of counterions.

Simulations of polycation-polyanion complexation are conducted with either fully-flexible or semi-flexible polyanion with 200 beads. Polycation-polyanion complexation simulations are conducted with all of the aforementioned polycation architectures: linear, and grafted with graft lengths of 2, 3, 4, or 5 and graft spacing of 2 or 3. The ratio of positive to negative polyelectrolyte charge (which we denote as charge ratio CR) is varied by changing the length of the polycation to achieve the following values of charge ratios: 0.5, 1.0, 1.5, and 2.0. In polycation-polyanion complexation simulations, neutralizing monovalent counterions for both polyelectrolytes are included: a constant 400 positively-charged counterions to balance the charge of the polyanion, and a variable number of negatively-charged counterions to balance the charge of the polycation.

5.2.5 Analysis methods

We have chosen a variety of methods to characterize the structure of the polycations. We calculate the end-to-end distance, R_{ee}, the average distance between the first and last monomers in a polymer, and the radius of gyration, R_g, the average distance of the monomers from the center of mass (R_{com}) of the polymer, mathematically defined by the equation:

$$\langle R_g^2 \rangle = \frac{1}{N} \langle \sum_{i=1}^{N} (R_i - R_{com})^2 \rangle \quad (6)$$

where N is the number of monomers and R_i is the position of monomer i. Angular brackets denote ensemble averages. Both R_{ee} and R_g vary with the length of a polymer in characteristic scaling relationships defined as $R \propto N^{\nu}$, where R may be R_{ee} or R_g, N is the number of monomers, and ν is termed the scaling factor. Scaling factors are determined from the slope of a log-log plot of R versus N. Notable examples of scaling factors are $\nu = 1/3$ for a sphere, $\nu = 0.588$ for a self-avoiding random walk, and $\nu = 1$ for a rigid rod. Therefore, flexible polymers in a globular shape have approximately $\nu = 0.6$, whereas more rigid, rod-like polymers approach $\nu = 1$ [36]. In this work, we report the radius of gyration of only the *backbone* of the polymer (either peptide in case of linear polycation or polyolefin in case of grafted polycation).

We also calculate the ratio of the above quantities, $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$, a measure of the shape of a polymer. While values of $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle \sim 12$ indicate a rod-like conformation, values near 6 indicate a globular conformation resembling a random walk. Finally, the ratio of R_{ee} to the contour length, R_{max}, gives an indication of how stretched or coiled the polymer is. R_{max} is the maximum distance between the first and last monomers of the chain if it is in a perfectly linear conformation with the bonds at equilibrium distances [36].

Another measure of the shape of a polymer is the relative shape anisotropy (RSA). To calculate the RSA, we first calculate the radius of gyration tensor, $T_{\alpha\beta}$, which is defined by:

$$T_{\alpha\beta} = \frac{1}{N} \sum_{i=1}^{N} (R_i - R_{com})_{\alpha} (R_i - R_{com})_{\beta} \quad (7)$$

where α and β denote the three Cartesian directions x, y, and z. The three eigenvalues of $T_{\alpha\beta}$, calculated by diagonalizing the gyration tensor, are denoted as λ_1^2, λ_2^2, and λ_3^2 and have the relation $\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2$ [37]. These eigenvalues, which sum to the radius of gyration squared, are the magnitudes of the eigenvectors pointing along the three principal axes of the structure in question. From these three eigenvalues, we compute the asphericity, A_S:

$$A_S = \lambda_1^2 - \frac{1}{2}(\lambda_2^2 + \lambda_3^2) \tag{8}$$

which is 0 for a perfect sphere. We also compute the acylindricity, A_C:

$$A_C = (\lambda_2^2 - \lambda_3^2) \tag{9}$$

which is 0 for a cylindrically symmetric structure [38]. Finally, the RSA is computed as:

$$RSA = \frac{A_S^2 + 0.75A_C^2}{R_g^4} \tag{10}$$

which is bounded between 0 (for a perfectly isotropic structure, such as a sphere) and 1 (for a perfectly anisotropic structure, such as a rod) [38].

A measure of the *local* flexibility, as opposed to the *global* values that arise from the ratio $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$ and the RSA, is the orientational correlation function. We have chosen two ways to examine the orientational correlation. The first is the scalar projection of each bond in the polymer onto the first bond, defined mathematically by:

$$\langle cos(\theta) \rangle^k = \langle \vec{b}_0 \cdot \vec{b}_k \rangle \tag{11}$$

where b_k is the k^{th} bond vector, the dot indicates the scalar product. This value is a function of the position along the polymer, *k*, and begins at a value of 1 at k=0 followed by a decay with increasing k. The manner of the decay can be indicative of a particular type of polymer. For example, the wormlike chain (WLC) model of semi-flexible polymers decays exponentially with a characteristic length known as the persistence length [36]. However, polyelectrolytes and flexible polymers are not well described by a simple exponential decay, and we do not focus on the standard WLC definition of the persistence length [39, 40]. Our second way of quantifying the orientational correlation is the scalar projection of the end-to-end vector onto each bond in the polymer. This is defined mathematically by:

$$L_P{}^k = \langle \hat{b}_k \cdot \hat{R}_{ee} \rangle \quad (12)$$

where circumflexes indicate unit vectors. Unlike the typical WLC exponential decay definition of the persistence length, the quantity $L_P{}^k$ does not assume a uniform persistence length and instead provides a local persistence length at each bond, which is more appropriate for flexible chains [21].

To examine the location and effect of counterions on the polymers, we calculate the number and fraction of counterions adsorbed to either the isolated polyelectrolyte or the polyplex, both the total number adsorbed as well as the number adsorbed to each monomer. Counterions are defined as 'adsorbed' when they are within the Bjerrum length of the corresponding oppositely charged bead on the polyelectrolytes. The Bjerrum length is the ratio of the attraction of two point charges to the thermal energy. The Bjerrum length in dimensionless units is defined as

$$\lambda_B \equiv \frac{e^2}{\epsilon_r k_B T}, \quad (13)$$

where e is the unit charge, ϵ_r is the dielectric of the medium, and $k_B T$ is the thermal energy. Using this equation and our simulation parameters yields a Bjerrum length of 2.3 σ. This is equal to the Bjerrum length in water at 300 K, approximately 7 Å, divided by our unit length of 3 Å/σ, providing a consistency check for our system of dimensionless units. In cases where a counterion is within the Bjerrum length of multiple monomers, we consider it to be bound to only the closest monomer, which eliminates double counting of counterions.

For all calculations on polyplexes, we consider only the region where the two polyelectrolytes are within the Bjerrum length of each other, which is termed the polyplex region in the discussion below. For instance, the radius of gyration of the polyplex includes only the polycation beads that are within the Bjerrum length of a polyanion bead and the polyanion beads that are within the Bjerrum length of a polycation bead, and these beads are considered the polyplex region, which is shown schematically in Figure 5.1e. Finally, we calculate the counterion-accessible surface area of the polyplex region using the VMD built-in measure command with a probe radius of 1 σ, which is the diameter of the counterions in our simulations [41].

5.3 Results

5.3.1 Polycation Structure

Effect of polycation architecture and charge. First, we consider the effect of architecture and charge on the size, shape, and local and global flexibility of the lysine-based polycations with only neutralizing counterions. Figure 5.2a shows the radius of gyration of the backbone of the linear and grafted polycations. R_g increases with the number of monomers, N, in the characteristic power-law behavior of polymers, which is evident from the linear nature of the relationship on a log-log plot. At constant N, the linear polycation is the smallest, PCP-PolyX is intermediate in size, and PCO-PolyX is the largest. This trend is expected because for a constant number of monomers N the number of backbone beads in each monomer increases from linear (1 bead per monomer) to PCP-PolyX (2 beads per monomer) to PCO-PolyX (3 beads per monomer), thus the overall sizes of the polymers also increase. As graft length (X) increases R_g increases but with decreasing returns, with the size difference between PCP-Poly2 and PCP-Poly3 being greater than the size difference between PCP-Poly4 and PCP-Poly5; the same trend is also observed for PC**O**-PolyX. The effect of increasing the graft length on polycation size decreases as the graft length increases because each added lysine on the graft has a larger free volume than the previous lysine in the graft, and therefore the lysine repels neighboring grafts less strongly.

Figure 5.2b shows the scaling factors, ν, of the polymers (i.e. the slopes of the lines in Figure 5.2a). We note that these values of the scaling factors apply to the polycations in the range of N we study (10-100 monomers); polyelectrolytes may show multiple scaling factors over a wider range of N. In the range of N we have studied, PCP-PolyX has a larger scaling factor than either PCO-PolyX or the linear architecture, suggesting that the grafted architecture with closer graft spacing has lower flexibility than the linear architecture or the grafted architecture with larger graft spacing. This trend is because closer graft spacing in PCP-PolyX leads to greater electrostatic repulsion between the charged oligolysine grafts. The scaling factors of uncharged polymers with the same architectures as the charged polycations show a universal decrease for all architectures (Figure 5.2b), clearly demonstrating that eliminating the charges

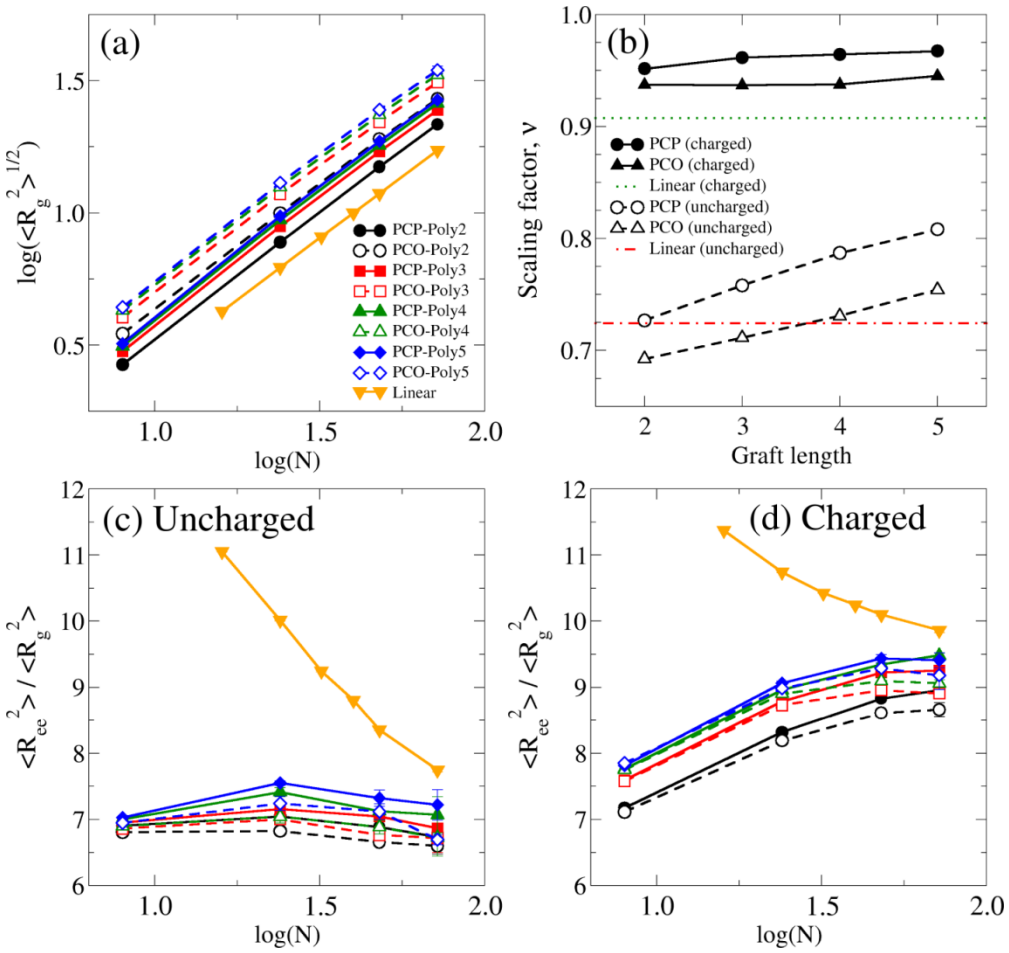


Figure 5.2 Effect of architecture and charge on polycation structure. (a) $\langle R_g^2 \rangle^{1/2}$ as a function of number of monomers, N. (b) The scaling factor, ν, from the relation $R_g \propto N^{\nu}$. (c) The ratio $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$ for uncharged polymers. (d) The ratio $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$ for charged polycations. Panels (a), (c), and (d) share the same legend.

leads to higher flexibility and smaller more compact conformations, which is expected because electrostatic repulsion is eliminated. Interestingly, the effects of both graft spacing and graft length on polycation size and flexibility are enhanced when charge is eliminated. Although uncharged PCP-PolyX has a smaller absolute radius of gyration than uncharged PCO-PolyX (Figure 5.14 in supplementary section 5.6), the difference in the scaling factor is greater than for the corresponding charged polycations, and the increase of the scaling factor with graft length is more pronounced. This is because in an uncharged state the polymers are smaller in size than in the charged state, and therefore excluded volume repulsion between grafts becomes more dominant within that compact state. When charges are present,

the size of the polycation is closer to the contour length (i.e., the maximum possible size), and any further expansion must compete with lengthening the bonds comprising the backbone (Figure 5.15 in supplementary section 5.6). The entropic unfavorability of maintaining a rod-like structure likely also plays a role. Therefore, the effects of the graft length and graft spacing on polycation size are lessened for the charged polycations compared to the uncharged polymers.

Polycation architecture has a substantial effect on the shape of the uncharged polymers and charged polycations, as quantified by the ratio of $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$ in Figure 5.2c and 5.2d, respectively. In the case of the uncharged linear polymer, with an intrinsically stiff peptide backbone, $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$ is near 12 for low N, indicating a rod-like shape, and decreases toward 6 for high N, indicating a globular polymer. The decrease occurs because at high N the length of the polymer backbone is greater than the persistence length of the peptide backbone. In contrast, the intrinsic flexibility of the polyolefin backbone of the uncharged grafted polymer leads to essentially globular behavior for all N studied here. Decreasing the graft spacing or increasing the graft length induces more rigidity, but the effect is minimal. The charged linear polycation behaves similar to the uncharged linear polymer, although it maintains more rod-like behavior at higher N due to electrostatic repulsion between charged monomers. The behavior of the charged *grafted* polymer is markedly different from their uncharged counterparts. First, for charged grafted polycation there is a shift from an essentially globular shape toward a more rod-like shape as N increases, in contrast with the globular shape at all N for uncharged grafted polymer. Second, increasing graft length increases the rod-like behavior, with the extent of increase decreasing with graft length. The graft spacing plays a weaker role in this overall shape factor because increased spacing increases both $\langle R_{ee}^2 \rangle$ and $\langle R_g^2 \rangle$, so the ratio is nearly the same regardless of graft spacing.

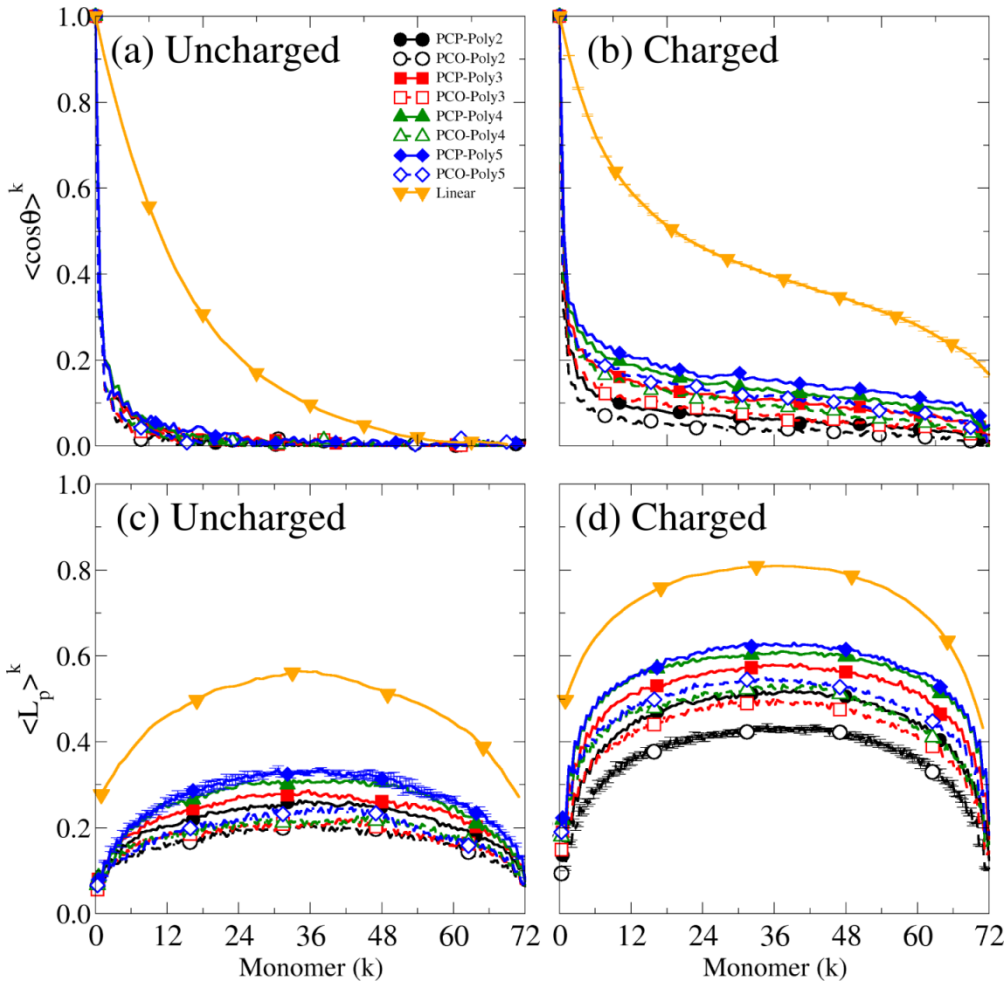


Figure 5.3 Effect of architecture and charge on polycation flexibility. The orientational correlation of uncharged polymers (a), and charged polycations (b), and the local orientational correlation of uncharged polymers (c), and charged polycations (d) as a function of monomer (k). All panels share the legend displayed in (a). Representative error bars are shown on only some lines for sake of clarity.

Next, we investigate in detail how polycation architecture affects *global* and *local* flexibility. The effect could be intrinsic (e.g. a rigid backbone) or architectural (e.g. induced by steric hindrance between grafts) or driven by electrostatics (e.g. due to like-charge repulsions) [21]. Using the orientational correlation function, $\langle cos(\theta) \rangle^k$, as a metric of *global* flexibility, we see that the value of $\langle cos(\theta) \rangle^k$ for the uncharged linear polymer (Figure 5.3a) decays smoothly over a comparatively long distance (or length of the polymer) due to the intrinsic rigidity of the backbone. In contrast, for the uncharged *grafted* polymer the correlation decays over a shorter distance due to the intrinsically flexible polyolefin backbone, and the correlation shows no dependence on either graft length or graft spacing. The

orientational correlation of the corresponding charged polymers (Figure 5.3b) decays over a longer distance than the uncharged polymers in all cases, indicating electrostatic repulsion-induced stiffness. One interesting behavior present in $\langle cos(\theta)\rangle^k$ for all of the polycations is the sharp decrease near the first and last monomers relative to the gradual decrease in the center. This inhomogeneity is attributable to end-effects: the monomers at the ends of the chain have greater free volume, thereby experiencing reduced repulsion from nearby monomers, and consequently they are more flexible [42]. The orientational correlation of the grafted polycations reveals slight increase in the stiffness with either increasing graft length or decreasing graft spacing, although the differences appear minor with this particular measure of flexibility.

The *local* orientational correlation function, ${L_P}^k$, is the correlation of each bond vector with the end-to-end vector of the polymer, yielding a metric of flexibility localized to each monomer. As with $\langle cos(\theta)\rangle^k$, the local orientational correlation typically takes on values between 0, indicating no correlation of that bond with the end-to-end distance and complete flexibility, and 1, indicating perfect alignment and complete rigidity. Figure 5.3c shows ${L_P}^k$ for the *uncharged* polymers, and the observed trends are different than with the orientational correlation, $\langle cos(\theta)\rangle^k$. As with $\langle cos(\theta)\rangle^k$, the linear polymer is clearly more rigid locally than the grafted polymers, but what is not evident with $\langle cos(\theta)\rangle^k$ is a slight increase in local rigidity with increasing graft length or decreasing graft spacing. For the uncharged polymer, the effect of graft spacing is greater than the effect of graft length, with PCP-PolyX being universally more rigid than PCO-PolyX, even at the lowest graft length. Additionally, the polymer chain ends are more flexible than the central monomers, which is not apparent from the orientational correlation function (Figure 5.3a). As stated earlier, increased flexibility near the chain ends is attributable to their greater free volume and reduced steric hindrance by adjacent grafts. Figure 5.3d shows ${L_P}^k$ for the corresponding *charged* polycations and several of the same trends as observed for the uncharged polymers are evident. As with $\langle cos(\theta)\rangle^k$, the same increase in rigidity with increased graft length or decreased graft spacing is observed, although the differences are clearer with ${L_P}^k$. Compared to the

uncharged polymers, two notable differences in the charged polycation ${L_P}^k$ may be observed: first, the significance of graft length and graft spacing in determining the local flexibility is increased; and second, the end-effects are more dramatic, with greater differences between the central and end monomers in the charged case. The reason for both of these observations is that electrostatic interactions extend over a longer range than van der Waals repulsions (i.e. excluded volume), so that graft-graft repulsions in charged polycations extend over a longer range, and therefore electrostatic repulsion rises more sharply with increasing chain length than excluded volume effects. The local orientational correlation increases with graft length with the same asymptotic behavior as is observed for the shape factor $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$ (i.e., the effect of increasing the graft length decreases with increasing graft length), which is sensible because the local rigidity defines the global shape.

To summarize the effects of architecture on the uncharged grafted polymers in the chain length range of 10-100 monomers, increasing graft length or decreasing graft spacing increases the size but has little effect on rigidity. Conversely, for the charged grafted polycations, increasing graft length increases the rigidity but has little effect on size, while increasing graft spacing decreases both the rigidity and the size scaling. These differences between the charged and uncharged molecules are due to the relative length scales and strength of electrostatic repulsion and excluded volume.

Effects of ionic strength and counterion valency. Having characterized the effect of polycation architecture and charge on structure with only neutralizing monovalent counterions, we now turn to the effects of both ionic strength and counterion valency. In addition to the simulations described in the previous section, which contained only the exact number of monovalent counterions necessary to neutralize the polycation charge, we have simulated the linear and grafted polycations at ionic strengths of 50 mM and 100 mM using either monovalent or divalent counterions, as well as systems with only neutralizing divalent counterions.

Figure 5.4 shows the effect of ionic strength and counterion valency on the R_g scaling factor, ν, of the different polycation architectures. We note again that these scaling factors are extracted from

simulations of grafted polycations in the chain length range of 10-100 monomers, and these polyelectrolytes may exhibit different scaling factors at larger length scales. As expected, increasing the ionic strength causes the polycations to become increasingly compact for all architectures, graft lengths, and graft spacings due to increased screening of the electrostatic repulsion between grafts [43, 44]. As the ionic strength increases, the Debye screening length (i.e. the length scale of interactions between charged groups) is reduced. Similarly, by comparing Figures 5.4a and 5.4b it is evident that increasing the counterion valency at a given ionic strength shrinks the polycations substantially: even with only neutralizing ionic strength, divalent counterions are more effective at compacting the polyelectrolyte, in agreement with some previous studies [13, 27, 45].

The degree of contraction occurring with increasing ionic strength or counterion valency depends on the architecture. With monovalent counterions, the linear architecture is the smallest due to the shorter length of the backbone per monomer in comparison with the grafted architecture (i.e., 1 bead for linear, 2 for PCP-PolyX, and 3 for PCP-PolyX). However, with divalent counterions the linear architecture is intermediate in size between Poly2 and Poly5. This change occurs because divalent counterions are able to condense Poly2, with a fully-flexible polyolefin backbone, more effectively than the linear polycation, which has a semi-flexible peptide backbone.

For the grafted architecture at a given ionic strength and graft length, regardless of counterion valency, the scaling factor of PCP-PolyX is always higher than that of PCO-PolyX, due to increased graft-graft repulsion when the grafts are spaced more closely. With monovalent counterions, as the ionic strength increases, polycations with long grafts maintain a higher scaling factor and the reduction in their scaling factor is lesser than that of polycations with short grafts (Figure 5.4a). With divalent counterions, a similar trend is observed, although the decrease in the scaling factor is less dramatic than with monovalent counterions (Figure 5.4b). The reason for both of these trends is that the charge of Poly2 is neutralized more effectively than the charge of Poly5. Figure 5.4c shows the effective charge per monomer of PCP-PolyX, which is calculated by subtracting the charge of any counterions bound to a monomer from the total charge per monomer. Clearly, longer grafts maintain a higher effective charge

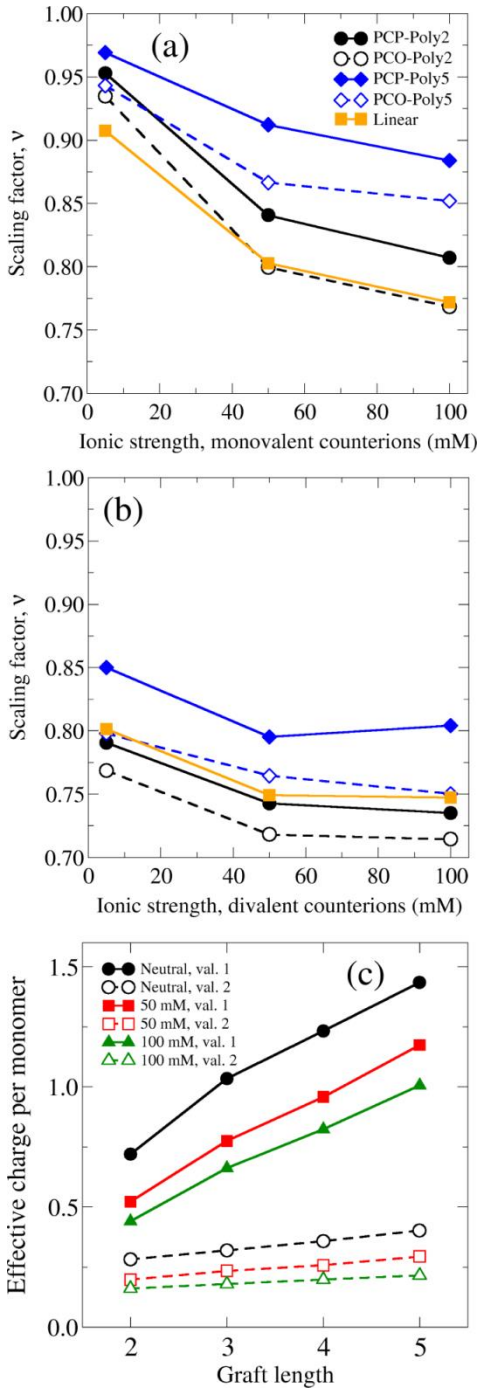


Figure 5.4 Effect of ionic strength and counterion valency on polycation size. Scaling factors for the relation $R_g \propto N^{\nu}$ at varying ionic strengths using (a) monovalent counterions and (b) divalent counterions. Only Poly2 and Poly5 are shown for the sake of clarity; the scaling factors of Poly3 and Poly4 are intermediate between the ones shown. Panels (a) and (b) share the same legend. Panel (c) shows the effective charge per monomer for PCP-PolyX, which is calculated by subtracting the charge of bound counterions from the charge per monomer. Short grafts are neutralized more effectively than long grafts as ionic strength or counterion valency are increased, leading to the smaller size of Poly2 compared to Poly5. PCO-PolyX is not shown because it is similar to PCP-PolyX.

than shorter grafts in all conditions, and this explains why Poly2 is condensed to a greater degree than Poly5 with increasing ionic strength. The higher effective charge per monomer results in higher effective graft-graft repulsion for longer grafts and therefore larger conformations. With divalent counterions, grafts of all lengths are mostly neutralized (effective charge near zero) even at the lowest ionic strengths, and the graft length has less of an effect on the effective charge than with monovalent counterions, which explains the reduced difference in size between Poly2 and Poly5 in the presence of divalent counterions.

To quantify the degree of counterion adsorption as a function of ionic strength and valency, we have calculated the number and fraction of counterions adsorbed to each individual monomer, as shown in Figure 5.5. Rather than examine the number of adsorbed counterions, which changes with the total number of counterions in the system and therefore in general does not allow for fair comparisons of different architectures, we present the fraction of counterions adsorbed to each monomer. The fraction of adsorbed counterions depends on the location of the monomer within the chain: monomers near the ends adsorb fewer counterions due to the lower electrostatic potential in those regions, as shown in Figure 5.16 in supplementary section 5.6. Away from the ends of the polycation, the fraction of adsorbed counterions reaches a plateau value, due to the consistent electrostatic potential. In the following discussion, we will consider only the plateau value because it most simply represents the behavior of the polycations.

Figure 5.5a shows the plateau values for the fraction of adsorbed counterions for PCP-PolyX versus the graft length; PCO-PolyX is not shown because it is essentially similar to PCP-PolyX. Increasing the ionic strength increases the actual number of adsorbed counterions (Figure 5.16 in supplementary section 5.6), but decreases the fraction of adsorbed counterions because of increased electrostatic screening. Also, since a given polycation can at most adsorb enough counterions to balance its charge, saturating the system with a large number of counterions will artificially lower the adsorbed fraction. Increasing the valency has the effect of dramatically increasing the fraction of adsorbed counterions for two reasons: first, there are a lower total number of divalent counterions required to reach a given ionic strength; and second, the adsorption of divalent counterions is more enthalpically favorable than the adsorption of monovalent counterions, while the entropic penalty for adsorption is the same.

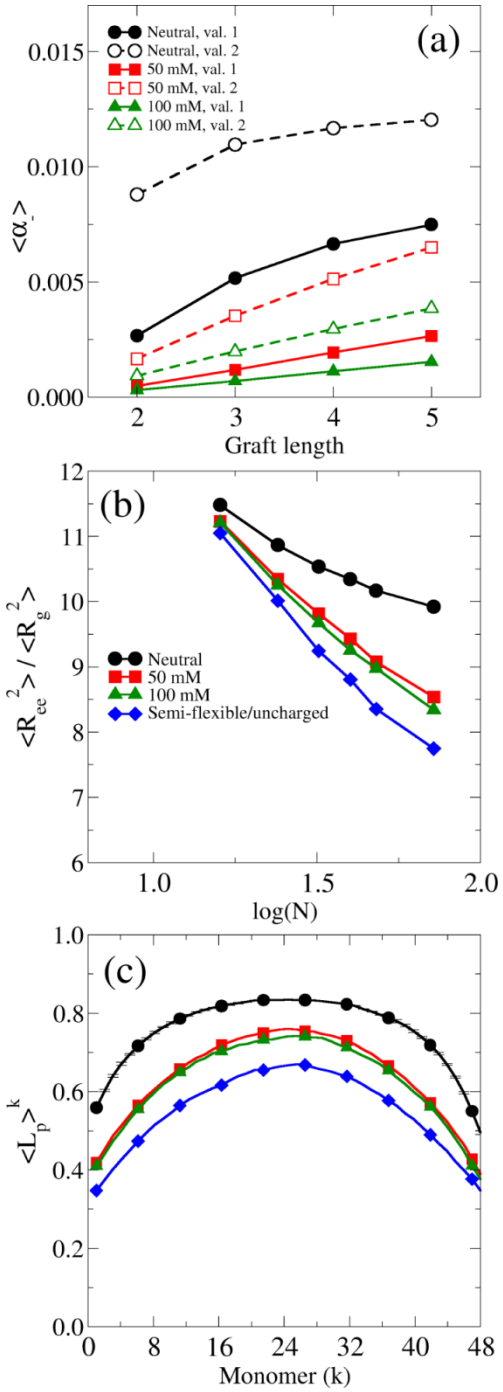


Figure 5.5 The effect of ionic strength and counterion valency on the fraction of adsorbed counterions and polycation structure. (a) The plateau values of the fraction of adsorbed counterions, $\langle\alpha_-\rangle$, as a function of graft length for PCP-PolyX. The trends for PCO-PolyX (not shown) are identical to those for PCP-PolyX. (b) Ratio $\langle R_{ee}^2\rangle/\langle R_g^2\rangle$ and (c) $\langle L_p\rangle^k$ for the linear polycation at various ionic strengths of monovalent counterions, and the linear uncharged polymer. The grafted polycation is not shown because the general trends are the same as with the linear polycation. The trends for divalent counterions are similar but more pronounced. Panels (b) and (c) share the same legend.

As the graft length increases, the fraction of adsorbed counterions monotonically increases. In the high ionic strength cases (50 mM and 100 mM) where the total number of counterions is identical for all graft lengths, the fraction adsorbed increases linearly with the length of the graft. However, in the neutral cases where the total number of counterions depends on the charge of the polycation (and hence on the graft length), the fraction adsorbed increase asymptotically because the total number of counterions also increases asymptotically: there are 2 times more counterions in the Poly3 systems than the Poly2 systems, only 1.5 times more counterions in the Poly4 systems than the Poly3 systems, and so on.

The structural properties of the polycations are also affected by ionic strength and counterion valency. As the ionic strength is increased, the polycations become more globular, as evidenced by a decrease in the ratio $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$ (Figure 5.5b). The polycations also become more flexible, as evidenced by a decrease in the local orientational correlation $\langle L_P \rangle^k$ (Figure 5.5c). All measures of polycation shape, size, and flexibility approach the values of the uncharged polymers as ionic strength or counterion valency are increased, because the polycation charge is largely neutralized. Figures 5.17 and 5.18 (supplementary section 5.6) contain additional data on the shape and flexibility of the linear polycation, including the relative shape anisotropy and $\langle cos(\theta) \rangle^k$. Here, we only present results for the linear polycation because it shows the general behavior of all the polycations, and we only show results for monovalent counterions for clarity. The effects of increasing ionic strength with divalent counterions are more pronounced (Figures 5.17 and 5.18, supplementary section 5.6), because the polycations are more effectively neutralized by divalent ions, as is evident in Figure 5.4c and Figure 5.5a. At high ionic strength with divalent counterions, the polycations behave like neutral polymers.

To summarize, the effect of increasing ionic strength or counterion valency is to increase screening between charges. Divalent counterions are substantially more effective at screening charges than monovalent counterions. Screening of charges causes the polycations to become increasingly compact for all architectures, graft lengths, and graft spacings, with the extent of compaction depending

on the architecture of the polycation and the valency of counterions. Lastly, as the ionic strength is increased, the grafted and linear polycations become more globular.

5.3.2 Complexation of Linear Polyanions and Polycations

Having extensively characterized the behavior of the polycations in isolation, we now describe their interactions and behavior during polycation-polyanion complexation. We have performed a series of simulations at increasing ratios of positive to negative charge, termed in this work the charge ratio (CR). The charge ratio can differ from the nitrogen:phosphate (N/P) ratio that is typically used in the field of gene delivery in reference to the common chemical nature of the species involved (i.e., amines and phosphates). For the grafted architecture studied here, the charge ratio and N/P ratio differ because each grafted monomer has a total charge that is 1 less than the number of positively-charged groups due to the presence of the negatively-charged C-terminus group. We found that the charge ratio, not the N/P ratio, is what actually determines the nature of the polyelectrolyte complex; in fact, the N/P ratio produces misleading results (see Figure 5.19, supplementary section 5.6, for a representative example). We varied the charge ratio by changing the length of the polycation while maintaining constant polyanion length. To explore the effects of polyanion flexibility on the binding process, we have conducted two sets of simulations: one with a fully-flexible polyanion, and another with a semi-flexible DNA-like polyanion. Initially, the polycation and polyanion are separated by some distance, and they condense into a polyplex during the course of the simulation. When characterizing the polyplexes, we consider only what we term the ‘polyplex region,’ (Figure 5.1e) which is the portions of the two polyelectrolytes that are within the Bjerrum length of one another. This distinction must be made because considering the entire structure of both polyelectrolytes could be misleading. An example is condensation at high charge ratios (CR > 1) when a large portion of the excess polycation may not interact with the polyanion at all, artificially increasing the apparent size and distorting the apparent shape of the polyplex if we did not consider only the ‘polyplex region’.

First, we characterize the size and shape of the polyplexes. The effects of graft length, graft spacing, and charge ratio on $\langle R_g^2 \rangle$ of the polyplexes are shown in Figure 5.6. Here, we only consider the grafted architecture for simplicity. Generally, the linear polycation produces similar results to the grafted architecture, although the polyplexes tend to be larger and more anisotropic; results pertaining to the linear architecture are available in supplementary section 5.6.8. For the fully-flexible polyanion (Figure 5.6a), the size of the polyplex is most sensitive to the charge ratio. At CR = 0.5, we observe that polycations with longer grafts are apparently able to condense the polyplex more effectively, regardless

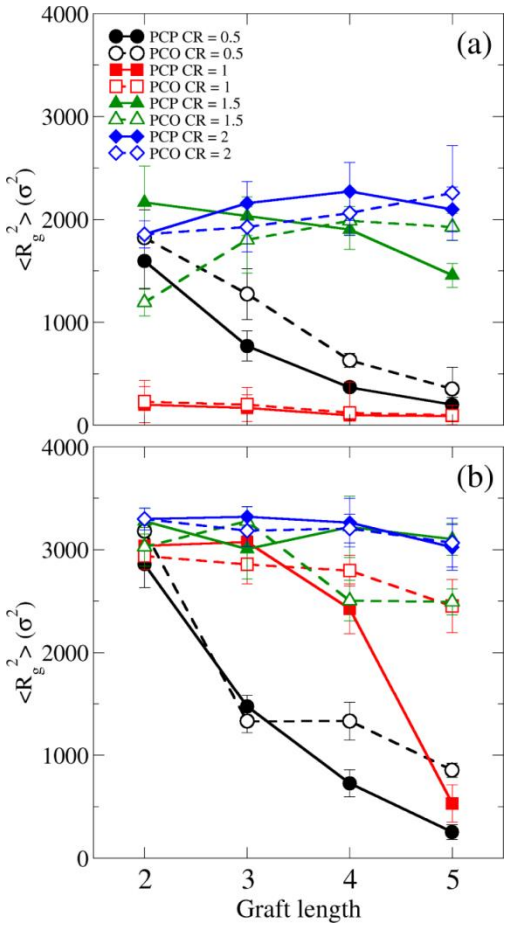


Figure 5.6 Effect of graft length, graft spacing, and charge ratio (CR) on polyplex size. $\langle R_g^2 \rangle$ for polyplexes formed by a polycation of the indicated graft spacing (PCP or PCO) and graft length (x-axis) with (a) a fully-flexible polyanion and (b) a semi-flexible, DNA-like polyanion. Both panels share the same legend.

of the flexibility of the polyanion. However, this trend in size at CR = 0.5 is an artifact of the nature of the calculation and the relative lengths of the polycations involved: at a given charge ratio, Poly2 is 4 times

longer than Poly5, and therefore the size of the region where the polycation and polyanion are in contact (the polyplex region) is larger for Poly2 than for Poly5. This artifact is not relevant at high charge ratios because then the length of the polyanion, which is constant, dictates the maximum size of the polyplex region. However, the results at CR = 0.5 is of little concern because the experimentally-relevant range of charge ratios for polyanion condensation is CR ≥ 1, where our discussion of the results now continues. At charge neutrality (CR = 1), the polyplex collapses due to the lack of electrostatic repulsion. For CR > 1, the polyplex expands due to the excess positive charge. For the semi-flexible polyanion (Figure 5.6b), the polyplex typically remains larger than any of the polyplexes formed with a fully-flexible polyanion due to the added intrinsic rigidity in the polyanion. Notably, Poly5 is able to substantially condense the semi-flexible polyanion at CR = 1, while the other architectures are incapable of this. This compaction occurs because Poly5 is substantially shorter than the polyanion at CR = 1, and it is therefore difficult to make favorable polycation-polyanion contacts if the polyanion is in an extended conformation. Poly5 exerts tension on the semi-flexible polyanion until it forms a more compact structure that allows for a greater number of enthalpically favorable contacts. By contrast, the other grafted polycations are closer in length to the polyanion at charge ratios near 1, so compaction of the polyanion is unnecessary to form favorable contacts.

The shape of the polyplexes, as quantified by the relative shape anisotropy (Figure 5.7), also depends on charge ratio, architecture, and polyanion flexibility. The fully-flexible polyanion (Figure 5.7a) is easily condensed into an isotropic globule at CR = 1, but the polyplex assumes an extended, anisotropic conformation for all other charge ratios. In the case of the semi-flexible polyanion (Figure 5.7b), the polyplexes are never isotropic. Poly5, however, is capable of forming toroidal structures at CR = 0.5, which is evident upon visualization of the simulation trajectories (Figure 5.7b) and which appears as a decrease in the relative shape anisotropy [37]. The formation of toroids occurs for the same reason that Poly5 reduces $\langle R_g^2 \rangle$ at CR = 0.5, as discussed above. For CR > 1, the polyplex favors a rod-like conformation for all of the grafted polycations (Figure 5.7b): because the flexible grafted polycation can

easily conform to the structure of the polyanion and the surface charge of the polyplex is inverted by the excess positive charge, the polyplex essentially behaves as a semi-flexible polycation.

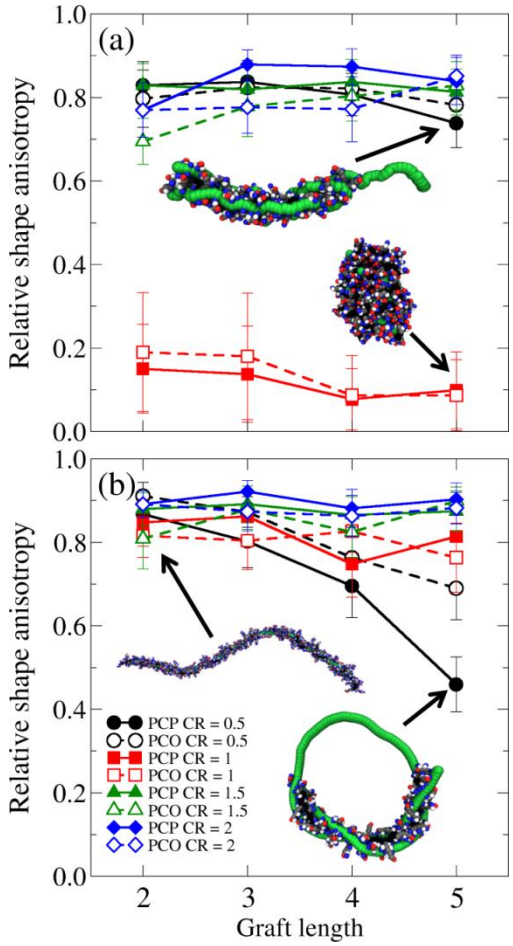


Figure 5.7 Effect of graft length, graft spacing, and charge ratio (CR) on polyplex shape. (a) Relative shape anisotropy of polyplexes formed between the indicated polycation architecture and a fully-flexible polyanion. (b) Relative shape anisotropy of polyplexes formed between the indicated polycation architecture and a semi-flexible polyanion, which could be considered a model for DNA. Representative snapshots of polyplex conformations are shown, with arrows indicating the corresponding value of the relative shape anisotropy. Both panels share the same legend.

So far we have quantified the size and shape of the polyplexes as a function of polycation architecture and charge ratio. Next, we quantify the surface charge of the polyplex as a function of the same parameters. We quantify the surface charge by calculating the average total charge of counterions adsorbed to the polyplex region and dividing this quantity by the average surface area of the polyplex

region. Because this quantity is defined in terms of counterions bound to the polyplex, it has opposite sign to the charge of the polyplex. We define a counterion as adsorbed to the polyplex if it is within the Bjerrum length of the polyplex region (i.e., those regions of the polyelectrolytes that are within the Bjerrum length of one another). Normalizing the charge of bound counterions to the surface area is necessary because increasing the surface area increases the available space for counterions to bind, and the surface area can depend on the particular conformation of the polyplex. An example of a polyplex conformation with lower surface area is a torus: since toroids typically involve multiple overlapping layers of the polyelectrolytes, a greater portion of the polyelectrolytes are buried within a dense polyplex region than would be in a rod-like conformation, thereby decreasing the surface area. The average total charge of adsorbed counterions and the polyplex surface area are available in Figures 5.23 and 5.24 in supplementary section 5.6, respectively.

Figure 5.8 shows the adsorbed charge per surface area for the grafted polycation. In all cases, the surface charge decreases with charge ratio, being positive at CR < 1, neutral at CR = 1, and negative at CR > 1, which occurs because the increase in polycation length with increased charge ratio inverts the net charge of the polyplex. The surface charge displays interesting polycation-dependent effects, which are most clearly evident for polyplexes formed with the fully-flexible polyanion (Figure 5.8a): increasing the graft length or decreasing the graft spacing increases the magnitude of the polyplex surface charge. For the semi-flexible polyanion (Figure 5.8b), the same trend is evident for CR ≥ 1, but an apparently non-monotonic trend with graft length is evident for CR = 0.5. However, this trend is due to particular conformations taken on by polyplexes with the semi-flexible polyanion. The semi-flexible polyanion can form kinetically-trapped structures, particularly with highly charged polycations such as PCP-Poly5, and the innate rigidity prevents rearrangement and redistribution of the polyplex charge. In contrast, the fully-flexible polyanion can easily deform to accommodate the local geometry of the polyplex, ensuring consistent neutralization of the surface charge by evenly distributing the polyanion charge throughout the polyplex region. Regardless, the issue of kinetic trapping at CR = 0.5 is of little practical significance because experimentally relevant charge ratios are above charge neutrality (CR ≥ 1).

From a fundamental standpoint, the trends for the fully-flexible polyanion are of interest (Figure 5.8a). The relatively short fully-flexible polyanion simulated here could be considered a model for a semi-flexible polyanion with a length substantially greater than the persistence length, like a long strand of

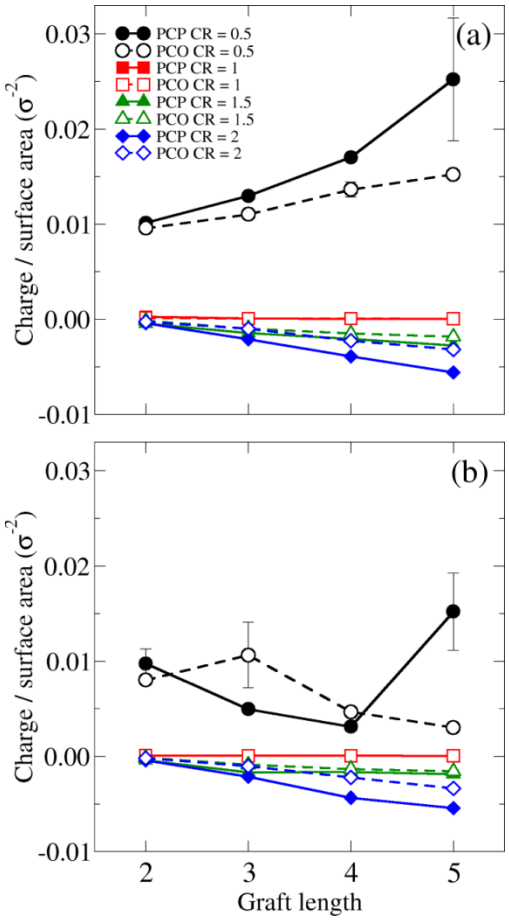


Figure 5.8 Effect of graft length, graft spacing, and charge ratio (CR) on the total charge of counterions adsorbed to the polyplex normalized to the surface area of the polyplex region for (a) the fully-flexible polyanion and (b) the semi-flexible polyanion. The total charge is calculated by summing the charge of all counterions that are within the Bjerrum length of the polyplex region, i.e. those portions of the polyelectrolytes that are within the Bjerrum length of each other. This quantity is a proxy for the polyplex surface charge, although of opposite sign because we are considering the counterions adsorbed to the polyplex surface. Both panels share the same legend.

DNA, since such a polyanion would be flexible over the length scale of the resulting polyplex. It may then be reasonable to extend the architecture-dependent results observed for these relatively small systems to larger scales. Therefore, the surface charge of polyplexes may be modulated by changing the linear charge density of the polycation, e.g. by changing the graft length or graft spacing for this particular

architecture. Considering the importance of polyplex net charge and surface charge on DNA transfection efficiency, in particular, the ability of a polycation to impart a positive surface charge to the polyplex to allow favorable interactions with the cell membrane, it is interesting to find such architecture-dependent behavior.

5.4 Conclusions

Using Langevin dynamics simulations and coarse-grained models, we have studied linear poly-L-lysine (PLL), poly(cyclopentene-*g*-oligolysine) (PCP-PolyX), and grafted poly(cyclooctene-*g*-oligolysine) (PCO-PolyX) to understand the combined effects of architecture and charge on the structure and flexibility of these lysine-based polycations. Comparing the linear and grafted polycations, we find that the linear polycation, although universally smaller than the grafted polycations with the same number of monomers, is universally more rod-like and less flexible than the grafted polycations, owing to the intrinsically rigid linear peptide backbone compared to the intrinsically flexible grafted polyolefin backbone. The same trend is observed for neutral polymers with the same architectures for the same reason. Considering only the grafted architecture, we find that increasing graft length or decreasing graft spacing increases the size and rigidity of the polycations, due to increased electrostatic repulsion between longer or more closely spaced grafts. The effect of increasing the graft length decreases as the graft length increases, because each additional lysine added to a graft has a greater free volume and therefore repels adjacent grafts less effectively.

Interestingly, the size of the uncharged grafted polymers is more dependent on architecture (i.e., graft length and graft spacing) than the charged grafted polycations, while the rigidity of the charged grafted polycations is more dependent on architecture than the corresponding uncharged polymers. This reversal occurs because of the relative strength and length scales of excluded volume and electrostatic repulsion: excluded volume repulsion (i.e., Lennard-Jones repulsion) is weaker and acts over a shorter range than electrostatic repulsion. The uncharged polymers are smaller than the charged polycations so the effect of excluded volume is more prominent because the grafts are generally closer together.

Increasing the graft length or decreasing the graft spacing increases the excluded volume repulsion between adjacent grafts, which in turn increases the size of the polymer. On the other hand, the size of the charged polycations already approaches the contour length (the maximum possible size) due to the electrostatic repulsion, and it is difficult to further expand the polycations because any further increase in size would compete with unfavorable lengthening of the bonds in the backbone. Therefore, increasing graft length or decreasing graft spacing produce weaker effects on the size of the charged polycations. The rigidity of the uncharged grafted polymers is lower than the charged grafted polycations due to the relative weakness of excluded volume repulsion, and the effects of graft length and graft spacing are lessened because of the shorter length scale of excluded volume: the polymer can bend substantially before distant monomers interact. The opposite is true when electrostatic repulsion dominates, leading to the greater rigidity and greater dependence of rigidity on graft length and graft spacing for the grafted polycations. These trends in rigidity and the molecular reasons for the varying rigidity with varying architecture can guide future chemical synthesis of linear/grafted polycations and in use of these polycations in applications other than gene delivery, such as in directing assembly of particles into nanostructures, as probes in biosensors, or drug delivery, where the flexibility of charged polymers/oligomers may play an important role.

We have also quantified the effect of ionic strength and counterion valency and find that increasing either value sufficiently will effectively neutralize the polycation and lead to behavior resembling the uncharged polymer, which results in a decrease in size and rigidity. Increasing the graft length or decreasing the graft spacing reduces the sensitivity to ionic strength and counterion valency because longer and more closely spaced grafts resist neutralization and maintain higher electrostatic repulsion between adjacent monomers, despite the fact that longer grafts adsorb a greater fraction of the total number of counterions. While past studies have shown that ionic strength and counterion valency can strongly affect the behavior of polyelectrolytes [13, 43] our work differs from previously published studies in that our polycation architectures are directly related to specific chemistries. By including details

specific to the chemistries of the polycations we are modeling, we have been able to discern architecture-dependent properties beyond the general trends described in previous studies.

Polyplexes formed by complexation of these polycations with fully-flexible and semi-flexible linear polyanions exhibit sizes, shapes and surface charges that depend strongly on the architecture and the charge ratio (CR) of the two polyelectrolytes. Polyplexes formed with a fully-flexible polyanion at charge neutrality (CR = 1) are smaller and more globular compared to these polyplexes at CR > 1, which are larger and anisotropic. Polyplexes formed with the semi-flexible polyanion maintain larger, more rod-like conformations than polyplexes formed with the fully-flexible polyanion, except for the polycations possessing the longest grafts, which are capable of forming smaller toroidal polyplexes. The surface charge of the polyplexes is independent of the nature of the polyanion but depends strongly on the charge ratio and the architecture: at CR = 1, all polyplexes are neutralized, regardless of the graft length and graft spacing, whereas at CR > 1, either increasing the linear charge density of the polycation, whether by increasing the graft length or decreasing the graft spacing, increases the surface positive charge of the polyplex. These results of the effect of linear charge density on surface charge density are particularly relevant for gene delivery and other biomedical applications because polyplexes must possess net-positive charge to interact favorably with the negatively charged cell membrane. Although it is not known how the magnitude of the polyplex charge affects transfection, the ability to tune surface charge may offer a route to improve transfection efficiency.

We have employed a simplified DNA-like polyanion model to study the fundamental behavior of these polycations complexing with polyanion. However, a more detailed DNA model may be required to accurately replicate polycation-DNA complexation. To this end, work to incorporate a more realistic model of DNA, with a level of detail comparable to our CG polycation models, is underway. Also, despite using coarse-grained models, the length scales in this study are smaller than the molecules studied experimentally. In experiments, plasmid-DNA with a length greater than 1000 base pairs is typically used, but the length of the DNA-like polyanion in this study is only 200 base pairs. Although the length of the polycations we have studied is comparable to the degree of polymerization used in experiments [5-7],

in experiments multiple such polycations are used to condense several strands of DNA into a polyplex, while only one such polycation is used here. Despite these limitations, this study serves as a good first step towards providing valuable guidelines as to how to modulate the polyplex charge, size and shape by employing the appropriate graft length, graft spacing, and charge ratio: longer, more closely spaced grafts yield a higher surface charge than the converse, and higher charge ratios also lead to greater surface charge. Since a positive surface charge is requisite for transfection, one implication of these findings is that increased graft length may require lower polycation concentrations for efficient transfection, possibly reducing toxicity concerns in some applications. Most importantly this work provides a fundamental understanding of how polycation architecture affects polycation structure and polyplex size and shape at various ionic strengths, which is useful not only for designing gene delivery agents but also for engineering ligands for bioseparations, flocculants, biocompatible coatings, and nanoparticle assembly.

5.5 References

[1] D. J. Glover, H. J. Lipps, and D. A. Jans, Nature Reviews Genetics 6 (2005) 299.
[2] D. W. Pack, A. S. Hoffman, S. Pun, and P. S. Stayton, Nature Reviews Drug Discovery 4 (2005) 581.
[3] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, and J. P. Behr, Proceedings of the National Academy of Sciences 92 (1995) 7297.
[4] E. Wagner, M. Ogris, and W. Zauner, Advanced drug delivery reviews 30 (1998) 97.
[5] R. B. Breitenkamp, Z. Ou, K. Breitenkamp, M. Muthukumar, and T. Emrick, Macromolecules 40 (2007) 7617.
[6] R. B. Breitenkamp and T. Emrick, Biomacromolecules 9 (2008) 2495.
[7] S. S. Parelkar, D. Chan-Seng, and T. Emrick, Biomaterials 32 (2011) 2432.
[8] E. Wagner, Proceedings of the National Academy of Sciences 88 (1991) 4255.
[9] J. Dennig and E. Duncan, Reviews in Molecular Biotechnology 90 (2002) 339.
[10] Y. Wang, M. Zheng, F. Meng, J. Zhang, R. Peng, and Z. Zhong, Biomacromolecules 12 (2011) 1032.
[11] S. Mishra, P. Webster, and M. E. Davis, European Journal of Cell Biology 83 (2004) 97.
[12] R. M. Elder, T. Emrick, and A. Jayaraman, Biomacromolecules 12 (2011) 3870.
[13] Z. Ou and M. Muthukumar, Journal of Chemical Physics 123 (2005) 074905.
[14] R. S. Dias, P. Linse, and A. A. C. C. Pais, Journal of Computational Chemistry 32 (2011) 2697.
[15] R. S. Dias and A. Pais, Advances in colloid and interface science 158 (2010) 48.
[16] D. Angelescu, M. Vasilescu, and G. Staikos, Colloid and Polymer Science 289 (2011) 871.
[17] L. Feuz, F. A. M. Leermakers, M. Textor, and O. Borisov, Macromolecules 38 (2005) 8891.
[18] S. Bolisetty, C. Airaud, Y. Xu, A. H. E. Müller, L. Harnau, S. Rosenfeldt, P. Lindner, and M. Ballauff, Physical Review E 75 (2007) 040803.
[19] M. Saariaho, O. Ikkala, I. Szleifer, I. Erukhimovich, and G. ten Brinke, Journal of Chemical Physics 107 (1997) 3267.

[20] E. Y. Kramarenko, O. S. Pevnaya, and A. R. Khokhlov, Journal of Chemical Physics 122 (2005) 084902.
[21] R. Connolly, G. Bellesia, E. G. Timoshenko, Y. A. Kuznetsov, S. Elli, and F. Ganazzoli, Macromolecules 38 (2005) 5288.
[22] M. Saariaho, I. Szleifer, O. Ikkala, and G. ten Brinke, Macromolecular Theory and Simulations 7 (1998) 211.
[23] P. Košovan, Z. Limpouchová, and K. Procházka, Journal of Physical Chemistry B 111 (2007) 8605.
[24] O. V. Borisov and E. B. Zhulina, Macromolecules 38 (2005) 2506.
[25] Q. Cao, C. Zuo, L. Li, and G. Yan, Biomicrofluidics 5 (2011) 044119.
[26] Q. Cao, C. Zuo, L. Li, and N. Zhang, Macromolecular Theory and Simulations 19 (2010) 298.
[27] Q. Cao, C. Zuo, Y. Ma, L. Li, S. Chen, and Z. Hu, Journal of Polymer Science Part B: Polymer Physics 49 (2011) 882.
[28] Z. Ou and M. Muthukumar, Journal of Chemical Physics 124 (2006) 154902.
[29] C. F. Narambuena, E. P. M. Leiva, M. Chávez-Páez, and E. Pérez, Polymer 51 (2010) 3293.
[30] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, Journal of the American Chemical Society 117 (1995) 5179.
[31] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, Journal of Chemical Physics 79 (1983) 926.
[32] M. Ullner, in DNA Interactions with Polymers and Surfactants, John Wiley & Sons, Inc., 2007, p. 1.
[33] S. Plimpton, Journal of Computational Physics 117 (1995) 1.
[34] T. Schlick, Molecular modeling and simulation: an interdisciplinary guide, Springer-Verlag, 2002.
[35] R. W. Hockney and J. W. Eastwood, Computer Simulation Using Particles, Taylor & Francis Group, New York, 1988.
[36] M. Rubinstein and R. H. Colby, Polymer Physics, Oxford University Press, 2008.
[37] H. Noguchi and K. Yoshikawa, Journal of Chemical Physics 109 (1998) 5070.
[38] W. L. Mattice and U. W. Suter, Conformational theory of large molecules: The rotational isomeric state model in macromolecular systems, John Wiley & Sons, Ltd., 1994.
[39] A. V. Dobrynin, Macromolecules 38 (2005) 9304.
[40] M. Ullner and C. E. Woodward, Macromolecules 35 (2002) 1437.
[41] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics and Modelling 14 (1996) 33.
[42] H. Limbach, J. Chem. Phys. 114 (2001) 9674.
[43] A. Kundagrami and M. Muthukumar, Macromolecules 43 (2010) 2574.
[44] J. L. Barrat and J. F. Joanny, Europhysics Letters 24 (1993) 333.
[45] N. Korolev, A. P. Lyubartsev, and L. Nordenskiöld, Advances in Colloid and Interface Science 158 (2010) 32.

5.6 Supplementary information

5.6.1 Additional model details: Polycation model design

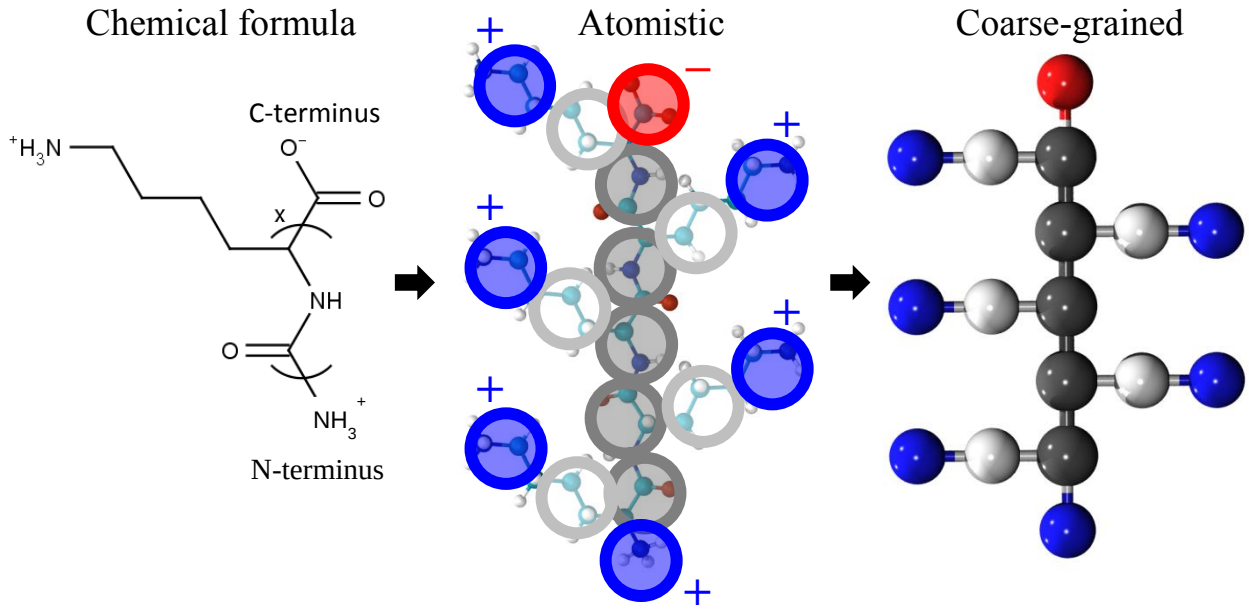


Figure 5.9 Diagram of the coarse-grained mapping of linear poly-L-lysine (PLL). The mapping is largely similar to that used for the grafted polycations. The two differences are the lack of the polyolefin backbone and the presence of a positively charged bead that represents the amine-terminus (i.e., N-terminus) of the polypeptide. Figure 5.10 gives additional details of the mapping.

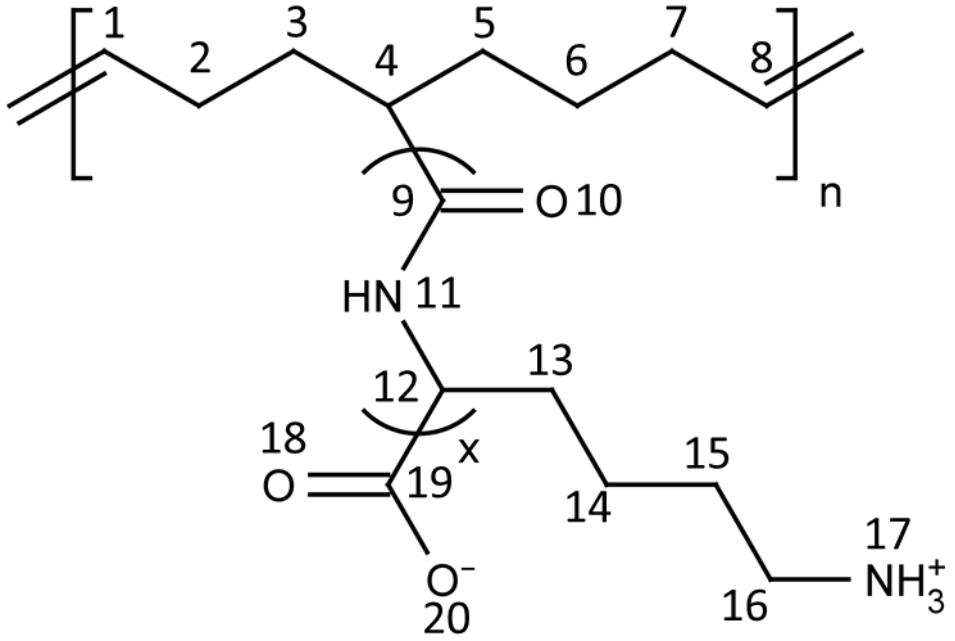


Figure 5.10 Mapping of atomistic structure to coarse-grained (CG) structure. Shown is PCO-PolyX.

The CG beads contain the following atoms:

- Positively-charged group (bead type P): atoms 15, 16, 17
- Negatively-charged group (bead type N): atoms 18, 19, 20
- Uncharged lysine spacer (bead type US): atoms 12, 13, 14
- Uncharged peptide backbone (bead type UP): atoms 9, 10, 11
- Uncharged polyolefin backbone (bead type UB):
    - atoms 1-8 for PCO-PolyX
    - atoms 1-5 for PCP-PolyX

All hydrogen atoms are included in the bead that includes their parent heavy atom. For PCP-PolyX, the oligolysine is grafted at atom 3 rather than atom 4. The N-terminus of linear lysine (see Figure 5.9) consists of an $–NH_3^+$ group, which is represented by the CG positively-charged group (bead type P) for simplicity.

Table 5.1 Atomistic sizes and masses of the atoms comprising the CG beads. The size of the groups is approximated by the end-to-end distance of heavy atoms in each polycation bead. The atomistic sizes of counterions are the Lennard-Jones radii of Na^+ and Cl^- from the Amber force field [30]. The sizes are similar enough that we find it reasonable to assign each of these groups the unit size, 1 σ. We define σ ≡ 3 Å, which is also similar to the height of a DNA base pair. Similarly, the masses of the CG groups are similar enough that it is reasonable to assign them unit mass, which we define as 41.8 g/mol because this is the average mass of the beads in the grafted architecture when averaging over all graft lengths and graft spacings. We note that the monovalent counterions may be considered wither Na^+ or K^+ ions, and the coions may be considered Cl^- ions. The divalent counterions may be considered Mg^{2+} or Ca^{2+}, and the divalent coions may be considered sulfate ions (SO_4^{2-}). Due to the level of coarse graining, it is not possible to assign a specific chemical identity to the counterions. Besides the valency, divalent counterions are identical to monovalent ions.

Bead type (name)	Atomistic Size (Å)	Mass (g/mol)
P (positive group)	2.2	45
N (negative group)	2.3	44
US (uncharged lysine spacer)	2.4	41
UP (peptide backbone)	2.1	43
UB (polyolefin backbone)	3.0	33 (PCP); 36 (PCO)
IP (positive ion)	2.5 (Na^+)	23 (Na^+)
IN (negative ion)	1.9 (Cl^-)	36 Cl^-)

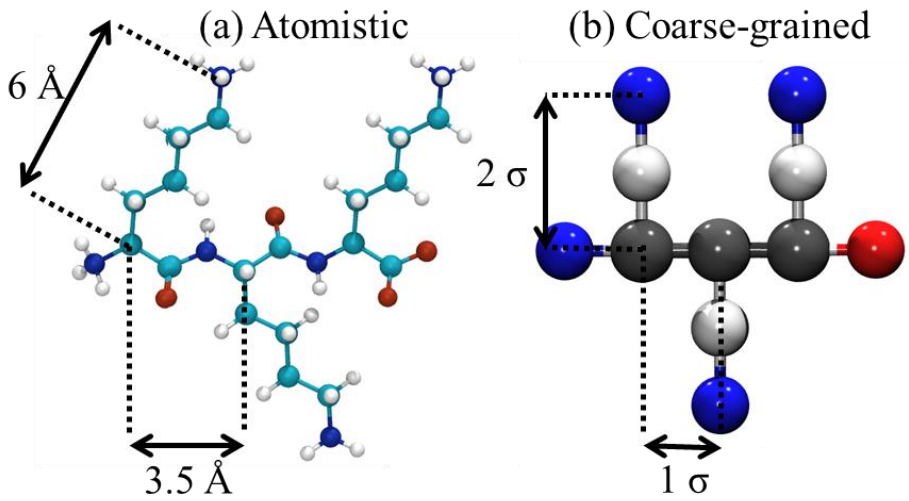


Figure 5.11 Spacing of linear architecture. For the atomistic structure, approximately 3.5 Å separates each R-group of approximately 6 Å length, leading to a length:spacing (L/S) ratio of 6/3.5 = 1.7. For the coarse-grained structure, the side chains are separated by 1 σ and are 2 σ in length, yielding an L/S ratio of 2.0. Therefore, the ratio is approximately commensurate between the atomistic and coarse-grained models.

Table 5.2 Spacing of grafted architecture. The spacing and length of grafts in both the atomistic (AA) and coarse-grained (CG) structures are shown in the top portion of the table. Length:spacing (L/S) ratios are shown in the bottom portion of the table. Finally, the ratio of the L/S ratios for CG and AA systems are shown. Because this final ratio is close to one, the ratio of graft length to graft spacing is commensurate between the atomistic and coarse-grained structures.

Lengths		
Spacing (S)	Atomistic (AA)	Coarse-grained (CG)
PCO	9.1	3
PCP	5.6	2
Length (L)	Atomistic (AA)	Coarse-grained (CG)
2	9	3
3	12.5	4
4	15.9	5
5	19.2	6

Ratios			
Graft length, PCO	L/S (AA)	L/S (CG)	**L/S (CG/AA)**
2	1.0	1.0	**1.0**
3	1.4	1.3	**1.0**
4	1.7	1.7	**1.0**
5	2.1	2.0	**0.9**
Graft length, PCP	L/S (AA)	L/S (CG)	**L/S (CG/AA)**
2	1.6	1.5	**0.9**
3	2.2	2.0	**0.9**
4	2.8	2.5	**0.9**
5	3.4	3.0	**0.9**

5.6.2 Model calibration: Linear polycation peptide backbone flexibility

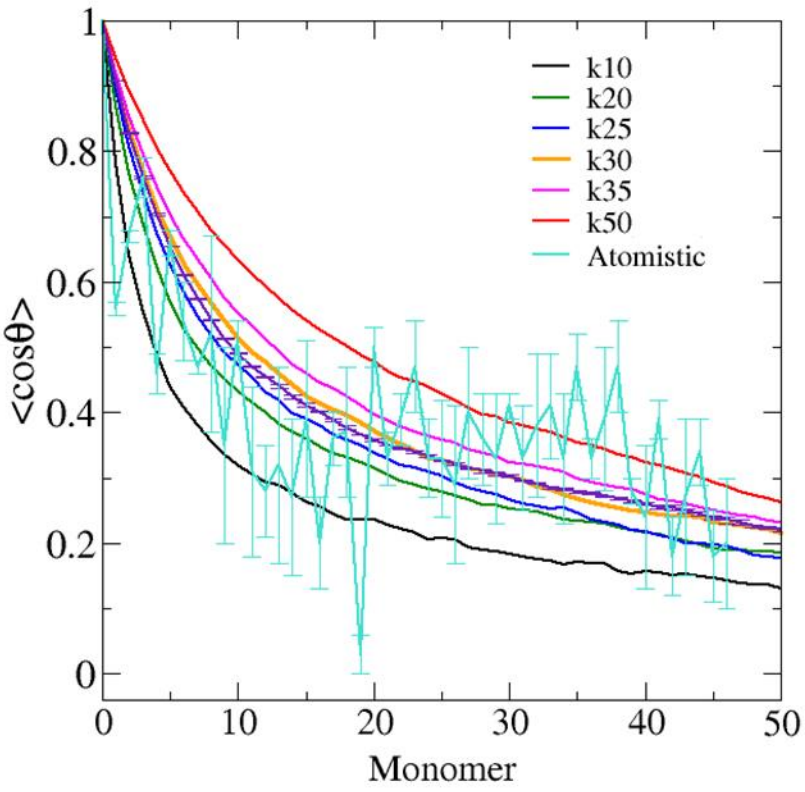


Figure 5.12 Calibration of linear backbone bending constant. The orientational correlation function of the CG linear model with varying k_{angle} (with the value indicated in the legend) versus the orientational correlation of atomistic poly-L-lysine. Although the error bars for the atomistic data are rather large, a value of k_{angle} = 30 for the CG model provides a value that neither consistently overestimates nor underestimates the flexibility of the atomistic molecule.

5.6.3 Model calibration: Grafted architecture polyolefin backbone hydrophobicity and flexibility

From our previous atomistic simulations, we have data on the size of the grafted architecture with all graft lengths and graft spacings at several lengths, N. To reproduce this data with our CG model, we had to alter both the backbone flexibility and the backbone hydrophobicity, because the relative importance of these two properties was unclear. These two properties combine to dictate the behavior of the grafted polycation. In general, the grafted architecture explores conformations that are of intermediate size, being neither globular nor rod-like. Two general situations may lead to this intermediate size: 1) the combination of weak hydrophobicity (weak motivation to coil up) and intrinsic flexibility (weak motivation to stretch out), or 2) by the competing combination of high hydrophobicity (strong motivation to coil up) and intrinsic rigidity (strong motivation to stretch out). Therefore, we conducted a series of simulations of all the grafted architectures with varying polyolefin backbone flexibility and hydrophobicity. Specifically, all combinations of k_{angle} = [0, 0.1, 0.5, 5.0] and ε_{LJ} = [0.6, 1.2, 1.8]; this combinatorial set of parameters ranges from fully-flexible (k_{angle} = 0) to semi-flexible (k_{angle} = 5) and from athermal interactions (ε_{LJ} = 0.6, which is the same as all other Lennard-Jones interactions) to hydrophobic (ε_{LJ} = 1.8). We found that the combination of ε_{LJ} = 0.6 and k_{angle} = 0 produced the greatest number of results that fell within 10% of the atomistic results over the entire range of graft lengths, graft spacings, and polycation lengths, when considering the end-to-end distance of the grafted polycations. It should be noted that all combinations of k_{angle} = [0, 0.1, 0.5] and ε_{LJ} = [0.6, 1.2] performed nearly as well as ε_{LJ} = 0.6 and k_{angle} = 0, indicating that the polycation size is relatively insensitive to these parameters over these ranges. However, using either a high bending constant or a high Lennard-Jones well depth consistently produced poor results, even when these two conditions are combined. We chose the values ε_{LJ} = 0.6 and k_{angle} = 0, but we are confident that our results would be insensitive to small changes in these values.

5.6.4 Additional model details and calibration: linear polyanion

To most accurately represent DNA within the confines of the 1-bead model, the diameter of the DNA beads should be approximately 20 Å to match the width of DNA. However, a width of 20 Å and a bond distance of 3.4 Å leads to substantial van der Waals overlaps between consecutive beads, which artificially increases both the actual bond distance between adjacent beads and the innate rigidity of the polyanion by imposing an effective bending potential. This is undesirable because we wish to have control over both the linear charge density, which is decreased by increasing the average bonded distance, and the intrinsic rigidity of the DNA, which is increased by the van der Waals repulsions. To overcome this, we reduced the diameter of the DNA beads to 2.2 σ, the largest value that eliminates overlaps with the same bonded distance of 1.1 σ. This compromise allows for the greatest correspondence between the atomistic structure of DNA and our CG model within the confines of the simplified 1-bead model. Once we had designed the model, we proceeded to parameterize the rigidity of the polyanion to best represent DNA.

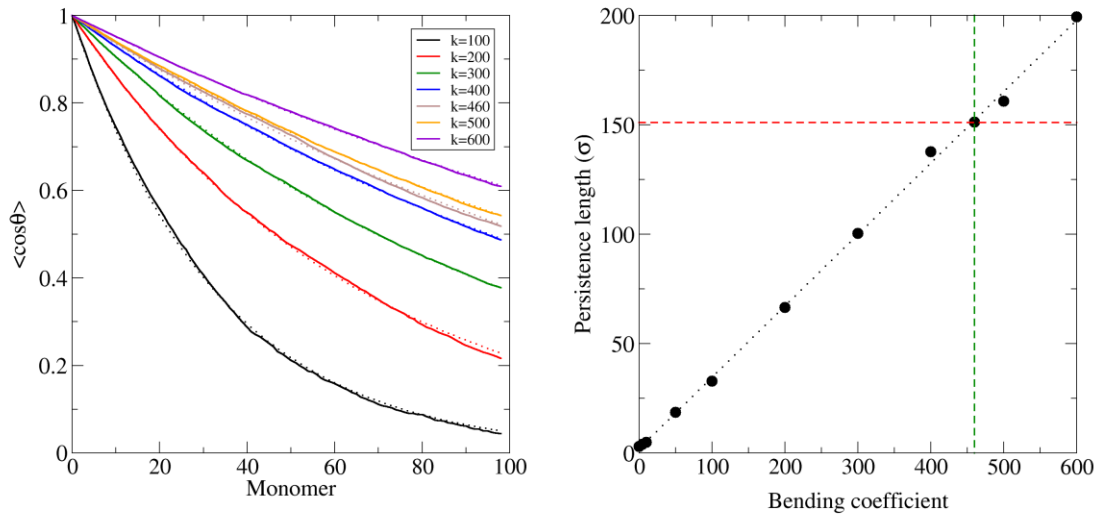


Figure 5.13 Orientational correlation of uncharged linear polymers with varying k_{angle} (a) and the persistence length of these polymers (b). The persistence length (L_P) was calculated by fitting $\langle\cos\theta\rangle$ with an exponential decay of the form $\langle\cos\theta\rangle = \exp(-s/L_P)$, where s is the monomer number. An uncharged polymer was used to obtain the persistence length in the high salt limit, which is approximately 450-500 Å depending on the definition and measurement method [32]. A persistence length of 150 σ is equivalent to a persistence length of 450 Å since we have defined σ = 3 Å.

5.6.5 Effect of architecture and charge on polycation structure

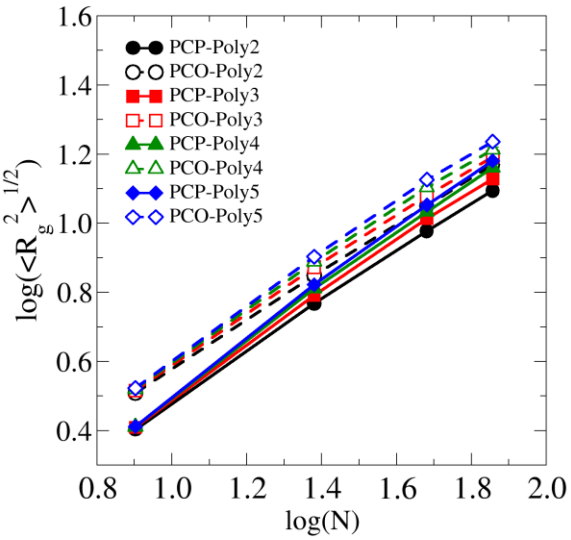


Figure 5.14 Radius of gyration (R_g) versus the length of uncharged grafted polymers of varying graft length and spacing. PCP-PolyX, possessing a shorter polyolefin backbone per monomer, has a smaller absolute size than PCO-PolyX, although the scaling factor ν (in the relation $R_g \sim N^{\nu}$) of PCP-PolyX is universally larger than that of PCO-PolyX.

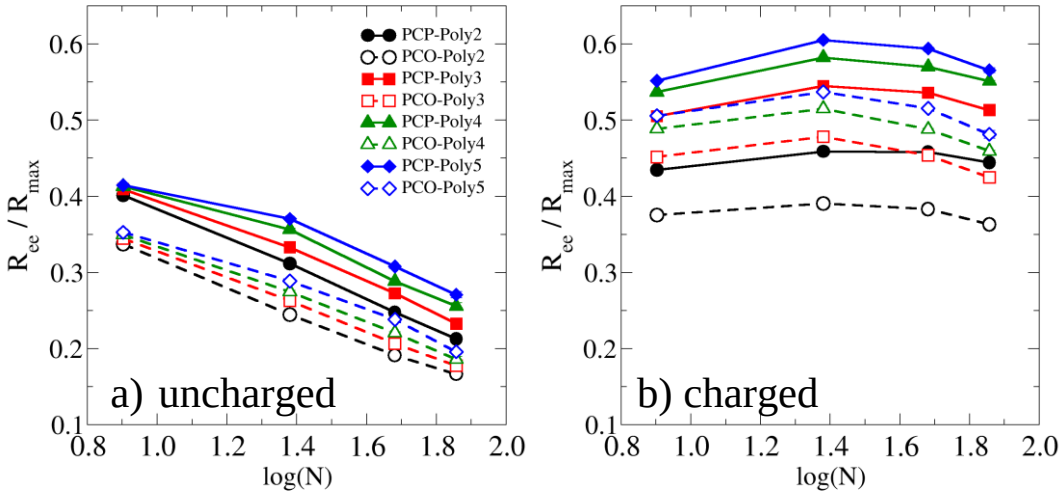


Figure 5.15 The end-to-end distance (R_{ee}) normalized to the contour length (R_{max}), which is the maximum possible extension of each molecule. For both charged and uncharged molecules, decreasing the spacing increases the relative size of the molecule. Increasing the graft length has a small effect on the uncharged polymer but a significant effect on the charged polycations. Finally, panel (b) demonstrates how the charged polycations are substantially closer to full extension ($R_{ee}/R_{max} = 1$), which likely makes any further expansion unfavorable for both enthalpic (e.g. lengthening of bonds) and entropic (e.g. rod-like structures have less conformational entropy than globular structures) reasons.

5.6.6 Effect of ionic strength and counterion valency

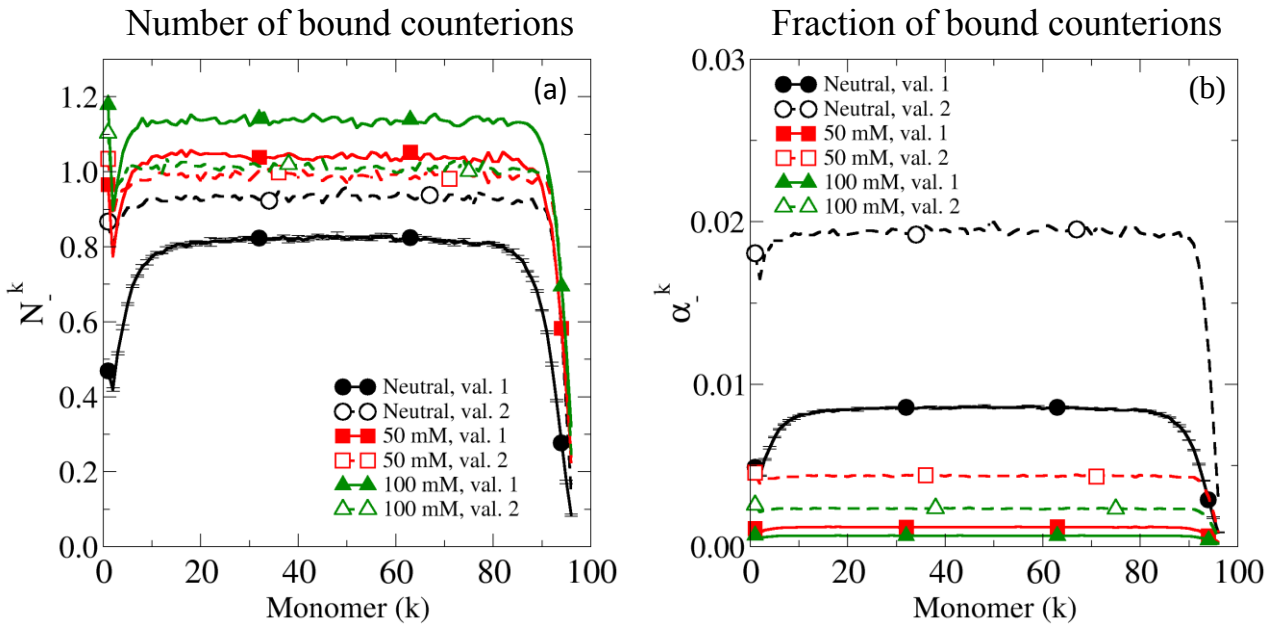


Figure 5.16 Effect of ionic strength and counterion valency on (a) the number and (b) the fraction of counterions adsorbed to the linear polycation. First, we observe three distinct regions: the region near the 1st monomer (N-terminal), the plateau region in the center, and the region near the Nth monomer (C-terminal). The plateau region occurs because the electrostatic potential near the center of a sufficiently long polycation becomes constant, and so all monomers in this region adsorb the same number of counterions. In the two tail regions the electrostatic potential is weaker, so fewer counterions are adsorbed. We also note that the specific model we are using for the linear polycation possesses a natural asymmetry. Since the N-terminal is modeled as an additional positive group on the 1st monomer, that monomer adsorbs more counterions than it otherwise would; and conversely, since the C-terminal is modeled as a negative group on the Nth monomer, that monomer adsorbs fewer counterions than it otherwise would. If neither of these terminal groups were present, the number of adsorbed counterions would be symmetric. Considering the plateau values of the data in (b), we find that by increasing the ionic strength, the fraction adsorbed decreases, because of increased electrostatic screening. Also, since a given polycation can at most adsorb enough counterions to balance its charge, saturating the system with a large number of counterions will artificially lower the adsorbed fraction. Increasing the valency has the effect of dramatically increasing the fraction of adsorbed counterions for two reasons: first, there are a lower total number of divalent counterions required to reach a given ionic strength; and second, the adsorption of divalent counterions is more enthalpically favorable than the adsorption of monovalent counterions while the entropic penalty for adsorption remains the same.

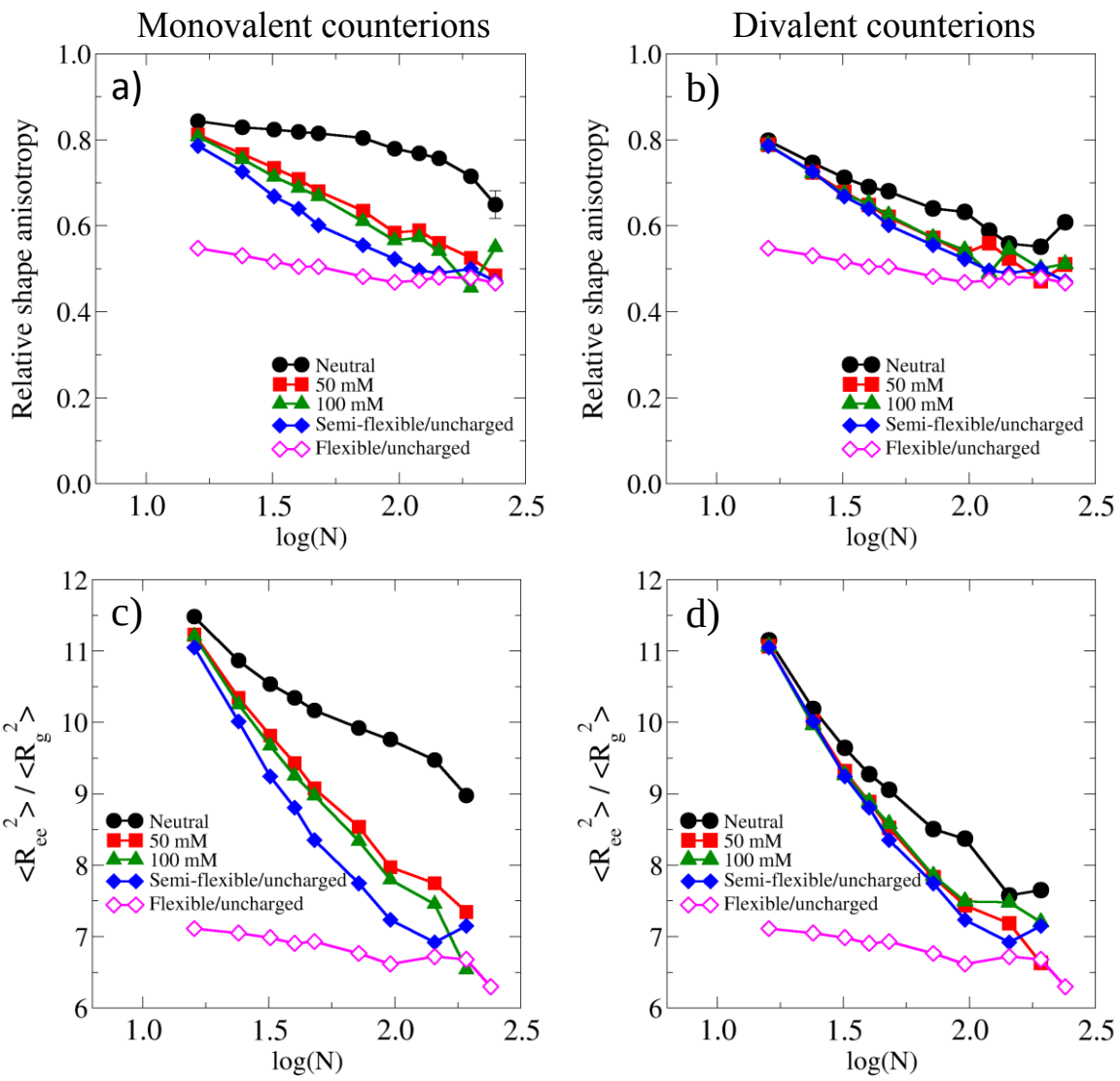


Figure 5.17 Measures of linear polycation shape with varying ionic strength and counterion valency. The 'semi-flexible/uncharged' polymer has the same peptide backbone rigidity as the linear polycation but lacks charges. The 'flexible/uncharged' polymer lacks charges and a backbone bending penalty. Panels (a) and (b) show the relative shape anisotropy, while panels (c) and (d) show the ratio $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$. Both measures of shape demonstrate that either increasing the ionic strength or the counterion valency neutralizes the polycation charge and results in shapes more closely resembling the semi-flexible/uncharged polymer. At high enough N, all of the molecules approach the same behavior (that of the flexible/uncharged polymer) because their length has exceeded the length scales of the peptide backbone rigidity and the induced electrostatic rigidity and they are able to form more globular shapes.

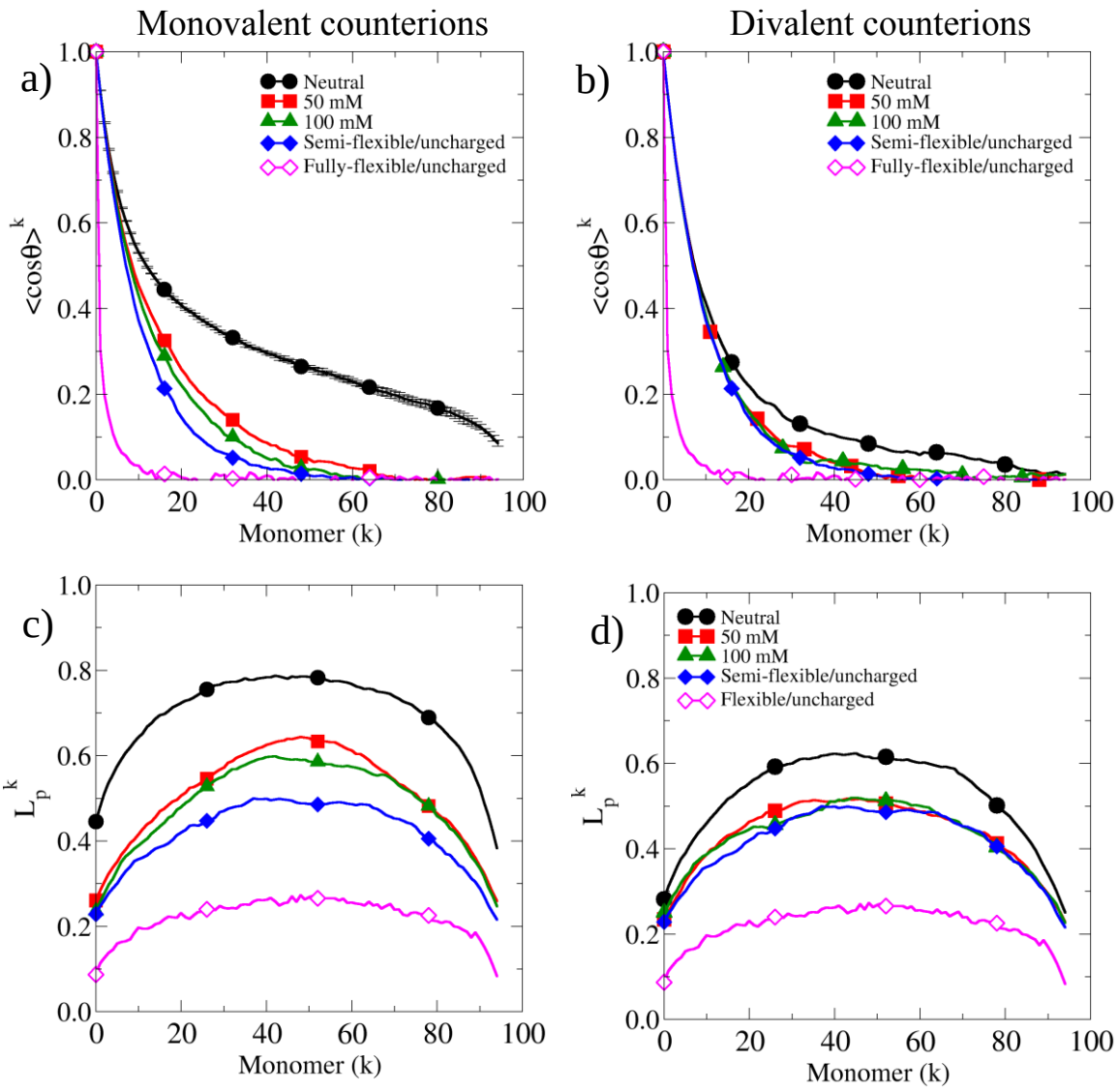


Figure 5.18 Measures of flexibility for the linear polycation at varying ionic strength and counterion valency. The 'semi-flexible/uncharged' polymer has the same peptide backbone rigidity as the linear polycation but lacks charges. The 'flexible/uncharged' polymer lacks charges and a backbone bending penalty. Panels (a) and (b) show the orientational correlation function, $\langle \cos\theta \rangle^k$, while panels (c) and (d) show the local orientational correlation, L_p^k. Both measures of flexibility show that increasing ionic strength or counterion valency reduce the rigidity of the molecules by reducing electrostatic repulsion. Note that the semi-flexible molecules (i.e. all of them except for the flexible/uncharged polymer) retain their intrinsic rigidity even at high ionic strengths.

5.6.7 N/P ratio produces misleading results compared to the charge ratio (CR)

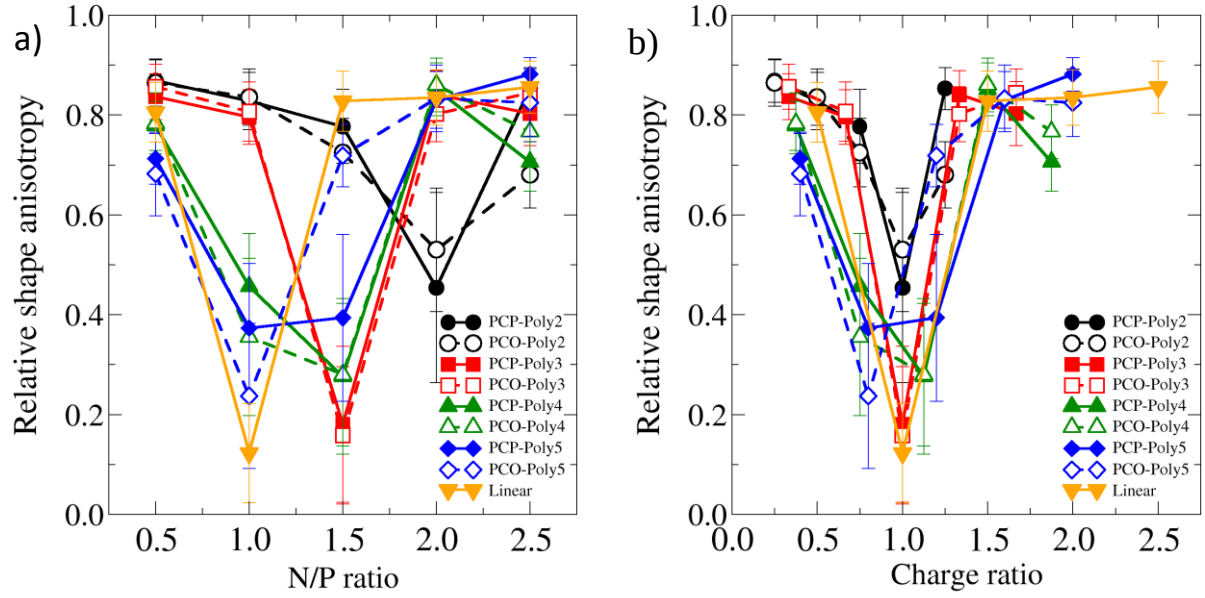


Figure 5.19 Relative shape anisotropy of polyplexes formed with a fully-flexible polyanion at varying N/P and charge ratios (CR). In panel (a), we present data for all polycation architectures (linear and grafted with varying graft length and graft spacing) over the range of N/P ratios from 0.5 to 2.5. For each architecture, the N/P ratio where the relative shape anisotropy indicates the point at which that polycation is able to substantially condense the polyanion into a compact polyplex, which always occurs at charge neutrality (i.e. when the charge of the polycation and polyanion are equal). Because the N/P ratio corresponding to charge neutrality depends on the architecture, the resulting plot yields a confusing trend at best. However, when the same data are plotted versus the corresponding charge ratio, the decrease in relative shape anisotropy occurs at CR = 1.0 (charge neutrality), as expected. It is possible that using N/P ratio to report data may be inappropriate in both experimental and simulation studies, depending on the particular molecules involved.

5.6.8 Linear architecture polyanion condensation

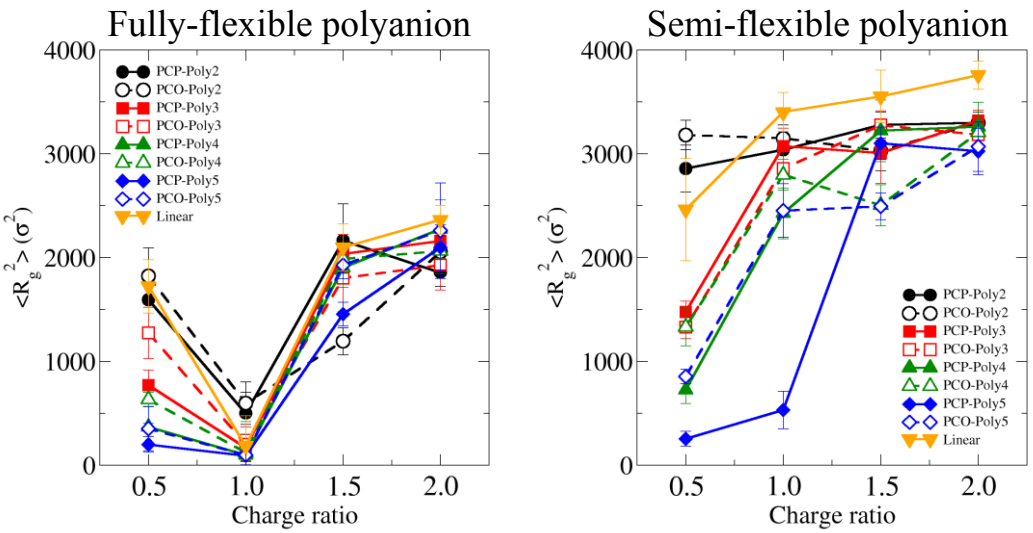


Figure 5.20 Radius of gyration (R_g^2) of polyplexes formed with fully-flexible and semi-flexible polyanions by polycations of varying architecture at varying charge ratio (CR). Regardless of polyanion flexibility, the linear architecture follows the same general trends as the grafted architecture. However, the linear architecture tends to produce larger polyplexes due to the rigidity of the peptide backbone, which is especially noticeable for the semi-flexible polyanion.

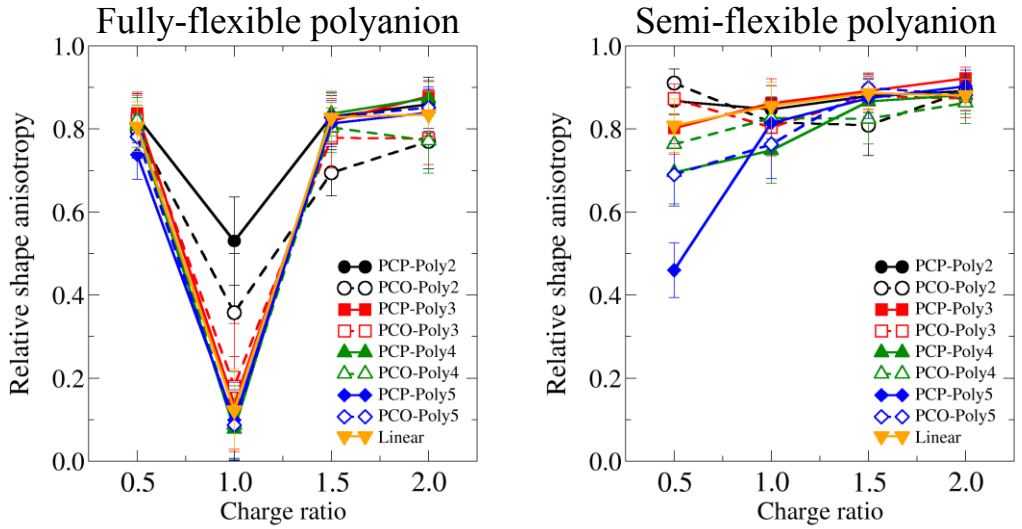


Figure 5.21 Relative shape anisotropy of polyplexes formed with fully-flexible and semi-flexible polyanions by polycations of varying architecture at varying charge ratio (CR). The shapes of polyplexes formed by the linear architecture with the fully-flexible polyanion are similar to those formed by the grafted architecture in most cases. The linear polycation does not condense the semi-flexible polyanion into a more globular form at any N/P ratio, likely because the length of the linear polycation is comparable to or greater than the polyanion, which also limits the ability of the grafted architecture to form globular polyplexes in most cases.

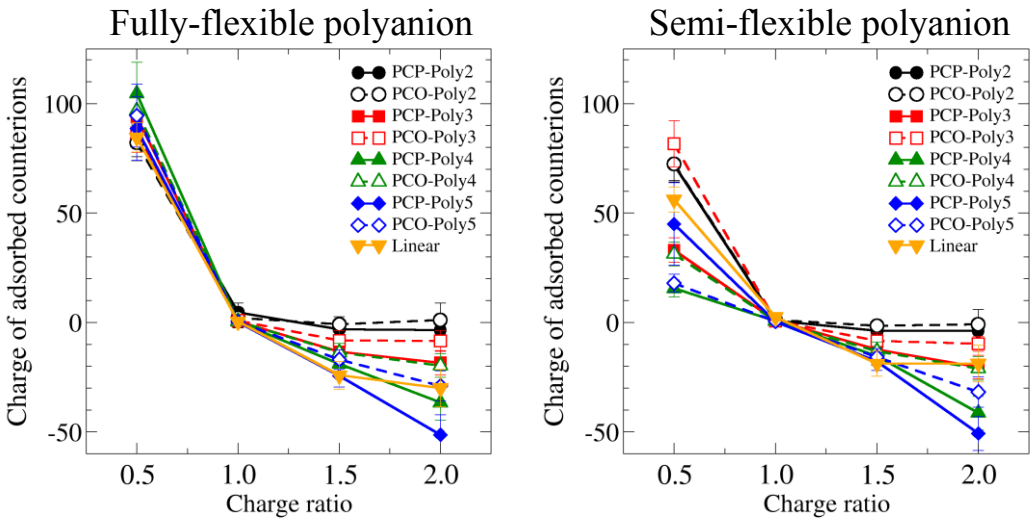


Figure 5.22 Charge of counterions adsorbed to polyplexes formed with fully-flexible and semi-flexible polyanions by polycations of varying architecture at varying charge ratio (CR). For CR > 1, the linear architecture produces polyplexes with intermediate charge compared to the grafted polycations. This is due to the intermediate charge density of the linear architecture with respect to the grafted architectures.

5.6.9 Additional details on total charge of adsorbed counterions and polyplex surface area

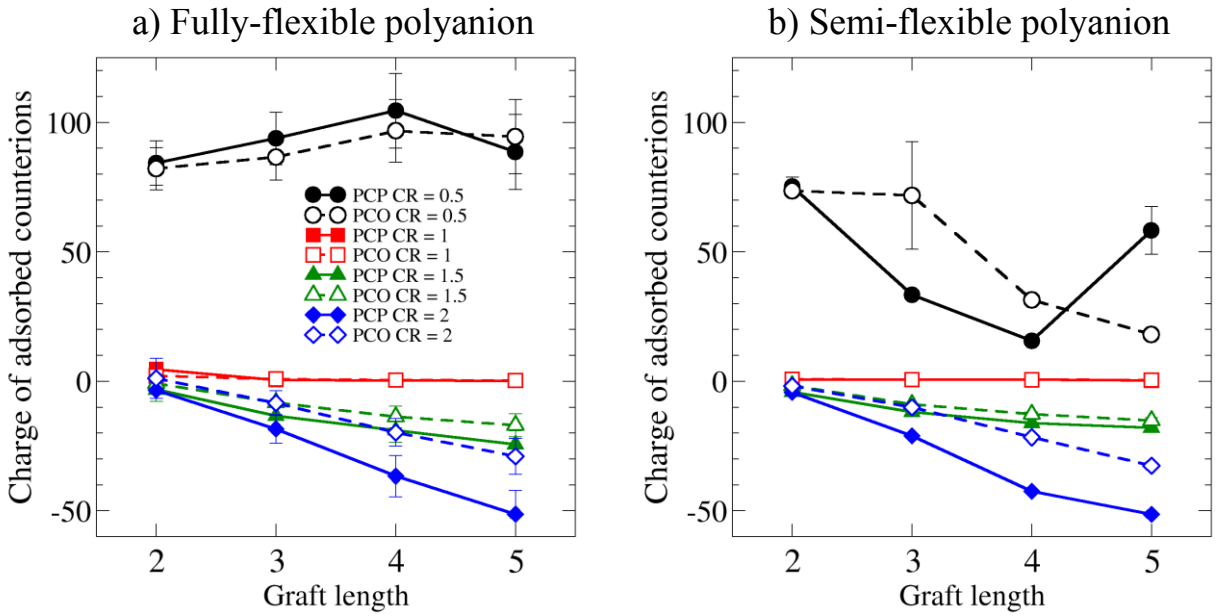


Figure 5.23 Total charge of counterions adsorbed to the polyplex region. The charge of adsorbed counterions is a proxy for the charge of the polyplex. Comparing with Figure 8 in the main text, which shows this quantity normalized to the polyplex surface area, many of the same trends are evident. First, as charge ratio (CR) increases, the charge of adsorbed counterions decreases due to the increasing positive charge of the polyplex. Second, at CR = 1, all polyplexes are neutral. Third, for CR > 1, the charge of adsorbed counterions decreases with increasing graft length or decreasing graft spacing. For the fully-flexible polyanion (a) at CR = 0.5, the total adsorbed charge is approximately the same for all graft lengths and graft spacings, whereas the charge normalized to the polyplex surface area increases with graft length and graft spacing (Figure 8a). For the semi-flexible polyanion at CR = 0.5, the apparent non-monotonic trends in adsorbed charge are more pronounced than when normalized to the surface area (Figure 8b). Both panels have the same legend.

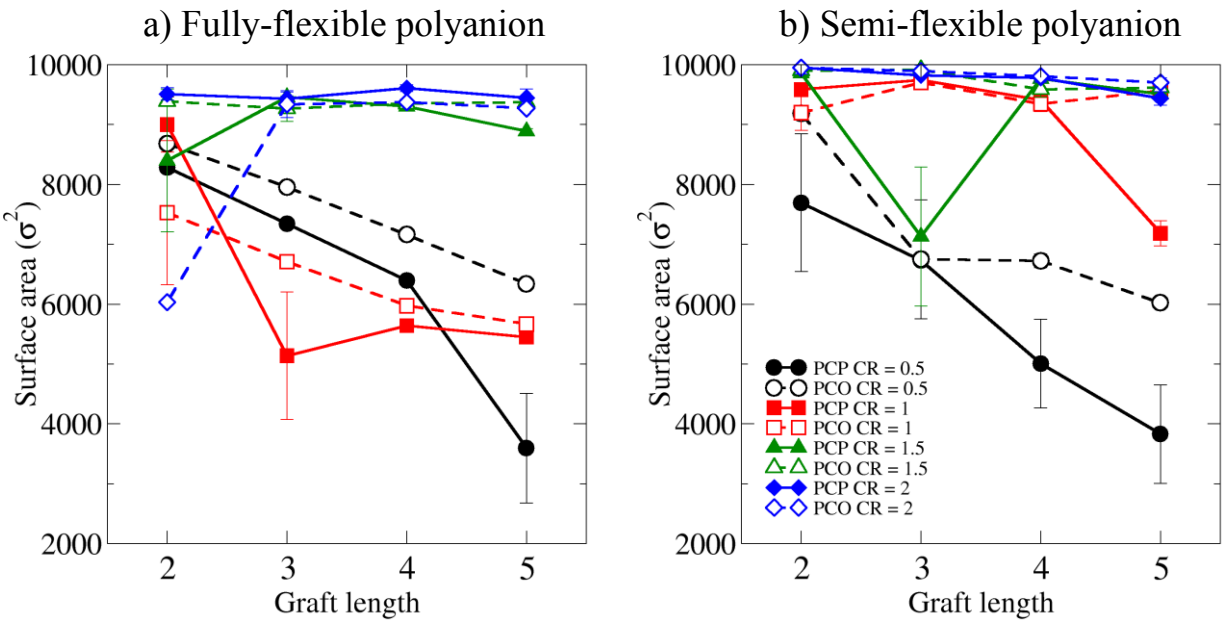


Figure 5.24 Surface area of the polyplex region. For CR > 1, the surface area is generally the largest due to charge-induced expansion of the net positively charged polyelectrolyte complex. In some cases the polyplex assumes a conformation with lower surface area, highlighting the importance of normalizing the adsorbed charge to the surface area. For the fully-flexible polyanion at CR = 1, the surface area decreases with graft length because polycations with longer grafts form smaller, more globular complexes. For the semi-flexible polyanion, which mostly forms rod-like polyplexes at CR = 1, the surface area remains high except for PCP-Poly5, which induces a toroidal polyplex shape. At CR = 0.5, the surface area of the polyplex decreases with graft length because polycations with longer grafts are necessarily shorter at a given charge ratio, as discussed in the main text, and the size of the polyplex region is dictated by the polycation length for CR < 1. Both panels have the same legend.

Chapter 6

Molecular Simulations of Polycation-DNA Binding Exploring the Effect of Peptide Chemistry and Sequence in Nuclear Localization Sequence Based Polycations

Adapted from: *J. Phys. Chem. B* **117** 11988 (2013)

6.1 Introduction

Gene therapy is an emerging technology that aims to cure disease by introducing therapeutic nucleic acids into human cells [1, 2]. Cationic polymers, a promising class of gene delivery agents, bind to polyanionic DNA or RNA molecules to form complexes (termed 'polyplexes'), which are often capable of gene delivery [1, 2]. In recent work, it was found that rearranging the architecture of a lysine-based polycation from a linear to a grafted architecture yielded an impressive improvement in gene delivery efficiency [3, 4]. It was also found that changing the graft length in this grafted architecture changed the efficiency [5], which we subsequently investigated using simulation methods to examine the structure of polyplexes and the strength of polycation-DNA binding [6, 7]. We found that, although each of the polycations can competently bind to DNA, the graft length affects the strength of binding to DNA, which in turn affects the release of the DNA during gene delivery. It is important to note that even if a polycation can competently form a polyplex with DNA, a number of barriers may prevent successful gene delivery. For example, the polyplex must undergo endocytosis to enter the cell, then escape the endosome into the cytosol, and finally, while avoiding destruction by intracellular nucleases, transport its genetic payload into the nucleus [8]. To overcome the last barrier – transport into the nucleus – polycations have been synthesized to include nuclear localization sequences (NLS) [1, 2, 5], which are oligopeptides that facilitate nuclear entry [9]. Nuclear localization sequences are typically cationic, making them an ideal

addition to existing synthetic polycations used for gene delivery [1, 2, 9]. In particular, Parelkar *et al.* introduced the SV40 large T antigen NLS into the previously studied oligolysine-grafted polycation [4-7]. The inclusion of NLS yields a polycation more effective than even the best commercially available products [5]. Surprisingly, the excellent transfection efficiency of these NLS-based polycations is apparently *not* driven by increased nuclear localization but rather by the DNA-binding behavior of NLS [5]. The present work extends our previous study on the effects of polycation *architecture* on DNA binding [6] by examining the effects of polycation *chemistry*, specifically the peptide sequence within the NLS-based polycations, on DNA binding.

There are several features of the chemistry of the NLS-based polycations studied by Parelkar *et al.* that likely affect DNA-binding behavior. In addition to lysine (single letter code K) in the original oligolysine-grafted polycations, the polycation containing the SV40 NLS (peptide sequence PKKKRKV) also contains the positively charged residue arginine (R) and the hydrophobic residues proline (P) and valine (V) (Figure 6.1). Although both arginine and lysine possess a single positive charge, the chemical nature of that charge differs: the guanidinium group of arginine has a delocalized charge that yields a higher hydrogen bonding capability than the primary amine of lysine [10, 11]. There are many consequences of this chemical difference between arginine and lysine. For example, because of its greater ability to form hydrogen bonds, arginine binds to the phosphate groups in DNA more strongly than lysine [11]. In contrast to lysine and arginine, the proline and valine residues have much weaker affinity for DNA and are also substantially more hydrophobic, suggesting that these residues may interact with the hydrophobic poly(cyclooctene) backbone in these polycations or the hydrophobic nucleobases in DNA (Figure 6.1). Altogether, the chemical differences between the P, V, K, and R residues are expected to significantly affect the DNA-binding behavior of NLS-based polycations.

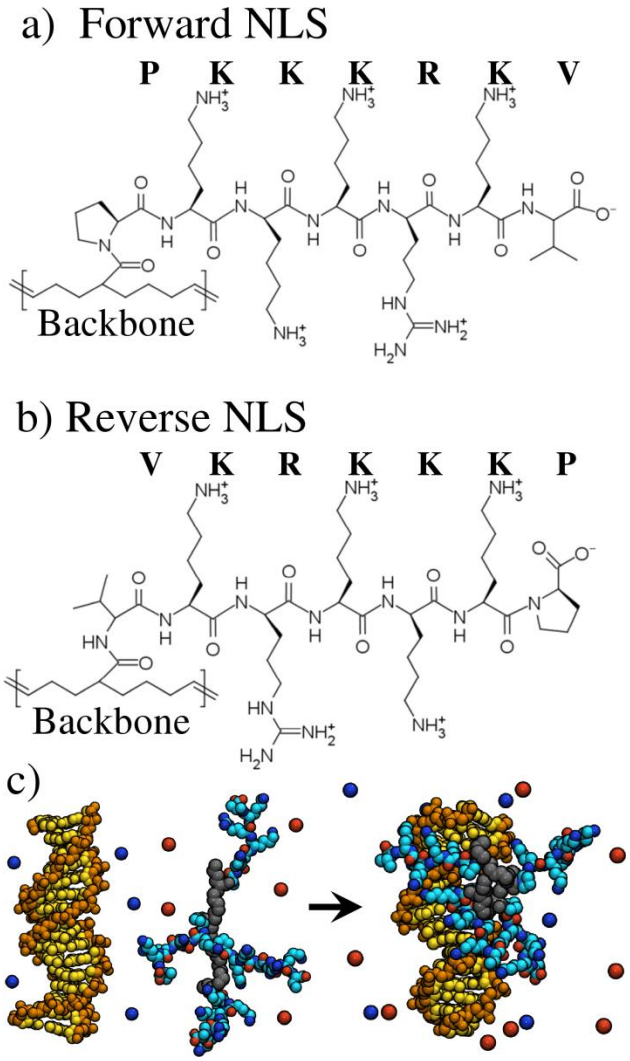


Figure 6.1 a) 'Forward' NLS-based grafted polycation with chemical name poly(cyclooctene-*g*-PKKKRKV) with grafts of the SV40 NLS. The hydrophobic poly(cyclooctene) backbone is labeled. Because the arginine (R) is the 4th charged residue from the hydrophobic backbone, we refer to this polycation as the 'forward' 'R4' case. b) 'Reverse' NLS-based grafted polycation with chemical name poly(cyclooctene-*g*-VKRKKKP). We refer to this polycation as the 'reverse' 'R2' case because arginine (R) is the 2nd charged residue from the hydrophobic backbone. c) A schematic of 16 bp double-stranded DNA, an NLS-based polycation, counterions, and water before binding (left) and after binding (right). DNA backbones are colored orange and DNA bases are colored yellow for clarity. The hydrophobic backbone of the polycation is colored gray and the NLS-based grafts are colored according to atomic element (carbon, cyan; nitrogen, blue; oxygen, red). Positive and negative counterions are shown as blue and red spheres, respectively. All water molecules and hydrogen atoms are omitted for clarity.

In addition to the differences in peptide *chemistry*, the effect of the peptide *sequence* on DNA binding is not well understood. It is possible that other peptide sequences based on the SV40 NLS will possess more favorable DNA-binding behavior and will in turn yield higher DNA transfection efficiency than the SV40 NLS. However, exploring the large combinatorial space of peptide sequence and peptide

chemistry using experimental tools is a daunting prospect. For this reason, we use computational tools, specifically atomistic molecular dynamics simulations, to explore how changing the peptide sequence and chemistry in the grafts affects changes in the enthalpy (examined through hydrophobic and electrostatic contacts) and entropy (specifically, conformational entropy measured through the distribution of dihedral angles) upon DNA binding.

We simulate the DNA-binding behavior of grafted polycations with the chemical name poly(cyclooctene-*g*-NLS), where the graft (NLS) attached to the poly(cyclooctene) backbone is a heptapeptide similar to the SV40 NLS (Figure 6.1). Based on preliminary experimental results suggesting that the directionality of the NLS peptide sequence with respect to the hydrophobic poly(cyclooctene) backbone affects polycation-DNA binding (T. Emrick, private communication), we first examine DNA-binding behavior of grafted polycations with 'forward' and 'reverse' versions of the SV40 NLS as grafts (Figure 6.1). In the 'forward' cases, the proline (P) residues are near the hydrophobic backbone and the valine (V) residues are at the end of the grafts; in the 'reverse' cases, the placement of these residues is switched. Within the 'forward' and 'reverse' cases, we also explore the role of arginine (R) on DNA binding by placing it at different positions in the graft peptide sequence and also by removing it entirely via mutation to lysine (K). Finally, we further examine the effects of the placement and chemistry of the proline and valine residues on DNA binding by mutating both hydrophobic residues to either proline or valine.

6.2 Simulation and Analysis Methods

6.2.1 Systems studied and simulation protocol

We use atomistic molecular dynamics to study the DNA-binding behavior of tetramers of grafted polycations with a hydrophobic backbone and grafts similar to that shown in Figure 6.1. In the first part of this study, we examine the effect of arginine position and NLS orientation by simulating polycations with grafts with the following peptide sequences: PKKKKKV/VKKKKKP ('No R' forward/reverse); PRKKKKV/VRKKKKP (R1 forward/reverse); PKRKKKV/VKRKKKP (R2 forward/reverse);

PKKRKKV/VKKRKKP (R3 forward/reverse); PKKKRKV/VKKKRKP (R4 forward/reverse); and PKKKKRV/VKKKKRP (R5 forward/reverse). In the second part of the study, we examine the effect of point mutations to the hydrophobic P and V residues with the following graft peptide sequences: PKKKKKP/PKKKKKV/VKKKKKP/VKKKKKV ('No R' P-P/P-V/V-P/V-V); PKRKKKP/PKRKKKV/VKRKKKP/VKRKKKV (R2 P-P/P-V/V-P/V-V); and PKKRKKP/PKKRKKV/VKKRKKP/VKKRKKV (R3 P-P/P-V/V-P/V-V). Tetramers of these polycations are simulated with a 16 base pair double-stranded DNA oligomer with the sequence 5'-d(CGCGCGAATTCGCGCG)-3' in a solution of explicit TIP3P water molecules and explicit neutralizing Na^+ and Cl^- counterions. We use 30 Na^+ ions to neutralize the DNA and 16 Cl^- ions to neutralize the polycation (i.e., an ionic strength of zero).

The polycation tetramers are constructed in the same manner as in our previous publication [6]. Briefly, the hydrophobic poly(cyclooctene) backbone is parameterized with the general Amber force field (*gaff*) and the AM1-BCC charge method by using the *antechamber* module of the Amber suite, and the remainder of the system (peptide grafts, DNA, TIP3P water molecules, counterions) is parameterized using the Amber *ff10* combination of force fields [12-18]. The limitations of this parameterization are discussed in our previous publication [6]. All simulations are conducted with NAMD, and all visualization and analysis is conducted with VMD [19, 20]. The simulation protocol (i.e., equilibration steps and parameters for production simulation) is identical to that in our previous publication [6]. Briefly, production simulations are performed in the constant pressure and temperature ensemble (NPT) at 300 K and 1 atm by using a Langevin thermostat with a temperature damping coefficient of 1 ps^{-1}, piston oscillation period of 100 fs, and piston decay coefficient 50 fs. We use the SHAKE algorithm to constrain all bonds with hydrogen atoms to enable 2 fs time steps, and snapshots are recorded every 2 ps [21]. Short-range non-bonded interactions are cut off at 9 Å, and the particle mesh Ewald (PME) method (with a tolerance of 10^{-6}, interpolation order of 4, and a maximum grid spacing of 1 Å) is used for long-range electrostatic interactions [22]. Initially in each trial, the polycation and DNA are separated by approximately 30 Å in a solvent box with side lengths of 80-90 Å. Polycation-DNA binding occurs over

approximately 10 ns, and the polycation-DNA systems are simulated for a total of 20 ns in each independent trial. We conduct 10 independent trials for each polycation-DNA system with the same initial coordinates but different initial velocities. Only the last 10 ns of each trial are used for analysis, so there are a total of 100 ns of sampling for each polycation-DNA complex.

We use the 'multi-trajectory' approach, wherein separate simulations are conducted of the DNA in solution, the polycation in solution, and the polycation-DNA complex in solution. This approach allows us to more effectively sample the unbound state (i.e., the polycation and DNA prior to complexation) than the 'single-trajectory' approach that is more commonly used [6, 23]. We conduct only a single 20 ns trial of the DNA in solution and each of the polycations in solution, in contrast to the combined polycation-DNA systems where we conduct 10 trials.

6.2.2 Analysis methods

We quantify the change in conformational entropy of the polycation and DNA upon polycation-DNA binding, which is one contribution to the free energy of polycation-DNA binding. Conformational entropy losses arise from the polycation and DNA having greater conformational freedom in the unbound state than in the polycation-DNA bound state. One would expect a great loss of conformational entropy in a tightly bound polycation-DNA complex where the polycation and DNA have greatly reduced conformational freedom.

To approximate the conformational entropy loss upon polycation-DNA binding, we use a method that estimates the conformational entropy from the breadth of the distributions of dihedral angles, or, equivalently, the flexibility of dihedral angles [24]. The entropy of the k^{th} dihedral angle, S_k, is given by:

$$\frac{-S_k}{k_B} = \int_0^{2\pi} P_k(q) \ln\big(P_k(q)\big)\, dq \qquad (1)$$

where $P_k(q)$ is the von Mises kernel density estimate of the probability distribution (i.e., histogram) of the k^{th} dihedral angle and k_B is the Boltzmann constant. The von Mises distribution function has the following form:

$$f(q,\mu,\kappa) = \frac{e^{\kappa\cos(q-\mu)}}{2\pi I_0(\kappa)} \quad (2)$$

where I_0 is the modified Bessel function of the first kind of order 0. The von Mises distribution is the periodic equivalent of the Gaussian distribution and is described by a mean value, μ, and a width parameter, κ, which is analogous to σ^{-2} of the Gaussian distribution. We use a value of κ = 225 as in our previous publication [6]. The von Mises kernel density estimate, $P_k(q)$, combines multiple von Mises distributions to describe the actual distribution of each dihedral angle *k*:

$$P_k(q) = \frac{1}{N}\sum_{j=1}^{N} f\left(q, q_{k,j}, \kappa\right) \quad (3)$$

where *N* is the number of simulation snapshots and $q_{k,j}$ is the value of the dihedral angle *k* in snapshot *j*. Finally, the sum over all *K* dihedral angles yields the total absolute conformational entropy, *S*:

$$S = \sum_{k=1}^{K} (S_k - S_0) \quad (4)$$

where S_k is the conformational entropy of dihedral *k*, and the quantity S_0 is obtained when $P_k(q)$ is the von Mises kernel function, rather than the density estimate, with any value for μ (e.g. π, which we use here) and the same value for κ as in the calculation of S_k. This method has been used to successfully quantify the conformational entropy of proteins such as dialanine and the Villin Head Piece and to relate protein sidechain entropy to NMR order parameters [24, 25]. We calculate the absolute conformational entropy in the bound and unbound states using all dihedrals not involving hydrogen atoms in both the polycation and the DNA. The difference between the absolute conformational entropy in the bound and unbound states yields the change in conformational entropy (ΔS_{conf}).

We also quantify other contributions to the free energy of binding, including the vibrational entropy, ion dissociation entropy, and the molecular mechanics energy. The methodology is identical to that in our previous publication [6], and we only describe it briefly here.

The molecular mechanics energy is calculated using the MMPBSA functionality in the Amber suite. We do not present the solvation free energy (i.e., the PBSA component) in this work in order to

examine only the electrostatic attraction between the polycation and DNA as a function of the polycation chemistry [26]. We note that the molecular mechanics energy is dominated by the electrostatic attraction between the polycation and DNA, with other contributions (e.g., bond stretching energy, van der Waals energy) making up less than 1% of the total molecular mechanics energy.

The vibrational entropy is estimated using a harmonic-approximation normal mode analysis using the *nmode* program in the Amber suite [26]. The counterion dissociation entropy is calculated by using a mean-field approximation of the translational entropy of all counterions *not* bound to either the DNA or the polycation, and by making the assumption that counterions bound to either polyelectrolyte have negligible translational entropy [6, 27]. Counterions are defined as being bound to a polyelectrolyte when they are within the Bjerrum length of the appropriate oppositely charged group on the polyelectrolyte. The Bjerrum length is the distance at which the electrostatic attraction between two point charges is equal to the thermal energy (k_BT), which is approximately 7 Å in water at 300 K.

Motivated by our observations of the thermodynamics of polycation-DNA binding, we examine structural aspects of polycation-DNA binding by quantifying the aggregation of hydrophobic chemical groups within the polycation and DNA and the formation of contacts between oppositely charged chemical groups in the polycation and DNA. Pair-wise hydrophobic contacts and electrostatic contacts constitute favorable interactions. At the same time, these contacts are expected to restrict the conformational freedom of the polycation and DNA and in turn reduce the conformational entropy.

We quantify hydrophobic aggregation by calculating the number of hydrophobic contacts, which we define as the number of pairs of hydrophobic atoms within 5 Å of each other. The choice of a 5 Å cut off distance for hydrophobic contacts is arbitrary, but this choice provides a balance between observing very few contacts (e.g., with a shorter cut off of 3.5 Å, approximately corresponding to the van der Waals contact distance between two carbon atoms) and observing many spurious contacts between atoms that are near each other merely because of proximity in the chemical structure (e.g., with a longer cut off of 7 Å, as used for bound counterions and electrostatic contacts). We find that our results regarding hydrophobic contacts are essentially insensitive to small changes in the cut off distance (i.e., between 4.5

Å and 6.0 Å), as shown in Figure 6.6 (supplementary section 6.6). We define hydrophobic atoms as the carbon atoms in the poly(cyclooctene) backbone of the polycation (general Amber force field atom types c2 and c3), the three consecutive carbon atoms in the arginine and lysine side chains (names CB, CG, CD), the carbon atoms in the valine side chain (names CB, CG1, CG2) and proline side chain (names CB, CG, CD), and non-polar carbon atoms in DNA (i.e., those with a partial charge less than +0.25 and greater than -0.25 as defined in the current parameterization of the Amber force field). Defining the hydrophobic atoms in DNA in this manner is arbitrary but provides a reasonable method of quantifying hydrophobic (i.e., non-polar) contacts within and between the polycation and DNA. We note that the nucleobases, which are the most hydrophobic portion of DNA, do not strongly interact with the polycation in our simulations because the DNA essentially maintains a canonical B-DNA structure with the nucleobases sequestered at the center of the double helix. We also calculate the duration of hydrophobic contacts, which we define as the length of simulation time that a particular pair of hydrophobic atoms remains continuously within the 5 Å cut off distance of each other.

We define electrostatic contacts (also termed *charged contacts*) as occurring when the distance between the center of mass of oppositely charged groups (e.g., phosphates, amines, guanidinium groups) is less than 7 Å, which is the Bjerrum length in water at 300 K, as discussed above. Arginine has a diffuse charge distribution and an abundance of hydrogen bond donors, leading to more favorable arginine-phosphate interactions compared to lysine-phosphate interactions [10, 11, 28]. Analogously, the geometry of arginine is excellent for forming charged contacts with other negatively charged groups, such as the peptide carboxyl-terminus. Lysine, by contrast, generally forms weaker contacts with negatively charged moieties, including carboxyl groups [29]. To further examine this difference between lysine and arginine in our simulations, we calculate the number of hydrogen bonds formed by the side chains of lysine or arginine with the DNA, and we calculate the electrostatic attraction between the lysine and arginine side chains and DNA. A hydrogen bond is considered to be formed when the distance between the donor and acceptor atoms is less than 3.5 Å, and the angle formed by the three involved atoms (donor, hydrogen, and acceptor atoms) is greater than 120°. We calculate the electrostatic attraction between the primary

amine of lysine or the guanidinium group of arginine and the entirety of the DNA molecule using the pair interaction feature of NAMD [20].

Statistical significance of trends (e.g., number of electrostatic contacts as a function of position of arginine) is assessed by performing one-way analysis of variance to calculate the F-statistic and corresponding p-value. Error bars are either the standard error of the mean calculated using the means of 10 trials (polycation-DNA simulations), or the standard error of the mean calculated by block averaging over 10 blocks within a single trial (polycation-only and DNA-only simulations).

6.3 Results and Discussion

6.3.1 Role of Arginine, Proline and Valine in Polycation-DNA Binding

Conformational entropy loss upon polycation-DNA binding. First, we describe how the position and presence of arginine (R), and the 'forward' and 'reverse' orientation of the grafts (*f* and *r*, respectively) in the grafted polycations affect the conformational entropy loss upon polycation-DNA binding in Figure 6.2.

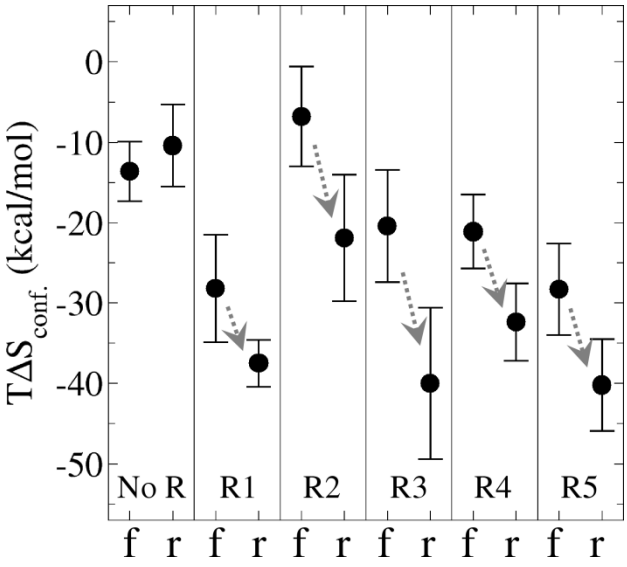


Figure 6.2 Conformational entropy change upon polycation-DNA binding for several NLS-based polycations. The peptide sequences represented in the x-axis are PKKKKKV/VKKKKKP ('No R' forward/reverse); PRKKKKV/VRKKKKP ('R1' forward/reverse); PKRKKKV/VKRKKKP ('R2' forward/reverse); PKKRKKV/VKKRKKP ('R3' forward/reverse); PKKKRKV/VKKKRKP ('R4' forward/reverse); and PKKKKRV/VKKKKRP ('R5' forward/reverse).

With respect to the position and presence of arginine, for the sequences containing arginine it is reasonable to expect that placing R farther from the hydrophobic backbone (i.e., going from R1 to R5), irrespective of the 'forward' or 'reverse' orientation of the sequence, should increase the exposure of arginine, which in turn should increase arginine-DNA interactions. Since arginine's diffuse charge distribution allows it to bind DNA more tightly than lysine [28], increased arginine-DNA interactions would lead to tighter polycation-DNA binding, thereby reducing the polycation's conformational freedom. This would be indicated by a downward trend in $T\Delta S_{conf}$ as arginine's position along the graft moves farther from the backbone (i.e., R1 to R5). Similarly, it is a reasonable expectation that sequences *without* arginine (i.e., the 'No R' case) would show the smallest loss of conformational entropy due to the complete elimination of arginine-DNA interactions. We observe some of the above expected behavior in our simulations, as shown in Figure 6.2. We see that for both the 'forward' and 'reverse' orientations, the 'No R' case shows a significantly smaller loss of conformational entropy than the sequences containing arginine, as expected. Surprisingly, as arginine moves away from the backbone (i.e., R1 to R5), the conformational entropy loss does not significantly increase, i.e $T\Delta S_{conf}$ is similar going from R1 to R5, for 'forward' and 'reverse' orientations. Likewise, the electrostatic interaction energy between the polycation and the DNA (the most significant contribution to the molecular mechanics energy) does not consistently increase or decrease in magnitude as arginine moves away from the backbone (Figure 6.7 in supplementary section 6.6). The absence of a significant relationship between the position of arginine and the conformational entropy loss or the molecular mechanics energy is contrary to our expectation that moving arginine away from the backbone would reduce hindrance by other peptides in the graft, would make arginine more available to interact with the DNA, and would in turn strengthen/tighten polycation-DNA binding. Reasons for the absence of these trends are discussed in the following sections addressing the formation of intra- and intermolecular contacts. Lastly, we find that the ion dissociation entropy and vibrational entropy are not significantly affected by moving arginine along the graft sequence (Figure 6.7 in supplementary section 6.6).

Interestingly, we find that the position of the hydrophobic residues, P and V, within the peptide sequence (i.e., the 'forward' or 'reverse' orientation of the graft sequence) significantly affects $T\Delta S_{conf}$ (Figure 6.2). Although the 'forward' and 'reverse' sequences without arginine ('No R') show a similar value of $T\Delta S_{conf}$, for sequences including R the 'reverse' sequences consistently lose more conformational entropy than the 'forward' sequences with the same placement of R, shown using the gray arrows in Figure 6.2. This finding is interesting for two reasons: first, polyelectrolyte condensation is typically dominated by electrostatic attraction and counterion dissociation entropy rather than by hydrophobic interactions [6, 27]; and second, the 'forward' and 'reverse' sequences are compositionally identical, and yet they show differences in the conformational entropy loss. Thus, to probe reasons why the 'reverse' sequences lose more conformational entropy than the corresponding 'forward' sequences, we now examine changes in the aggregation of hydrophobic groups upon binding.

Hydrophobic aggregation connected with conformational entropy losses. We show the change upon binding in the number of hydrophobic contacts made by proline (P) and valine (V) residues in Figure 6.3a. This metric shows the change in the aggregation of hydrophobic groups upon binding, which we hypothesize to be related to changes in conformational entropy upon binding. Relative to the 'forward' sequences, the 'reverse' sequences show a greater increase in hydrophobic aggregation upon binding, as indicated by the increase in the difference in contacts between the bound and unbound states (see gray arrows in Figure 6.3a). Because increased aggregation may imply decreased conformational freedom, the greater increase in hydrophobic aggregation for the 'reverse' sequences is consistent with the greater loss of conformational entropy observed for the 'reverse' sequences relative to the 'forward' sequences. Additionally, consistent with the lack of difference in conformational entropy for the 'forward' and 'reverse' sequences lacking arginine, there is no difference in the change in hydrophobic contacts for the 'No R' cases. We note that both the absolute value and the change upon binding in the *total* number of hydrophobic contacts (Figure 6.8 in supplementary section 6.6) follows similar trends to those seen for the P/V hydrophobic contacts shown in Figure 6.3a.

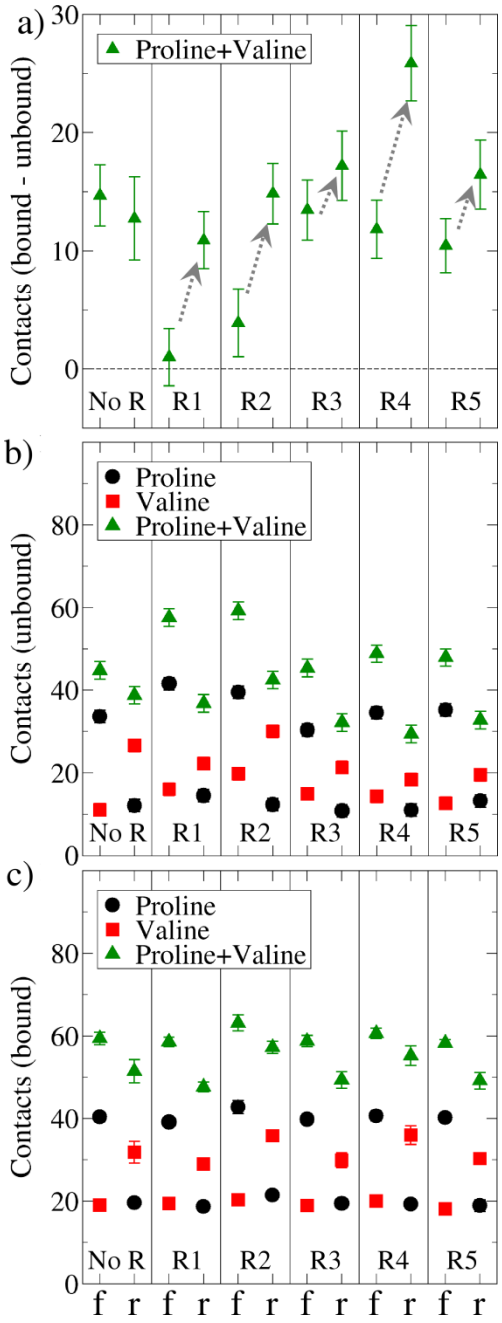


Figure 6.3 a) Change in the number of hydrophobic contacts made by proline and valine upon polycation-DNA binding. b) Number of hydrophobic contacts made by proline, valine, and the combination of proline and valine prior to binding (i.e., the unbound state). c) Number of hydrophobic contacts made by proline, valine, and the combination of proline and valine after polycation binds to DNA (i.e., the bound state). The peptide sequences represented in the x-axis are PKKKKKV/VKKKKKP ('No R' forward/reverse); PRKKKKV/VRKKKKP ('R1' forward/reverse); PKRKKKV/VKRKKKP ('R2' forward/reverse); PKKRKKV/VKKRKKP ('R3' forward/reverse); PKKKRKV/VKKKRKP ('R4' forward/reverse); and PKKKKRV/VKKKKRP ('R5' forward/reverse).

To explain these differences in the change in hydrophobic aggregation upon binding, we now consider the number of hydrophobic contacts made by proline and valine before and after DNA binding (Figure 6.3b). Before binding to the DNA (i.e., the unbound state), the 'forward' sequences form a greater number of P/V hydrophobic contacts than the 'reverse' sequences (green triangles in Figure 6.3b). After binding to DNA (i.e., the bound state), the difference between the P and V hydrophobic contacts for 'forward' and 'reverse' sequences is smaller (green triangles in Figure 6.3c) than in the unbound state, because, although both sequences gain contacts upon binding, they do not gain an equal number of contacts. Having more contacts initially, the 'forward' sequences tend to gain fewer contacts upon DNA binding; conversely, the 'reverse' sequences, having fewer contacts initially, tend to gain more contacts upon DNA binding. The difference in the number of contacts in the unbound state appears to be why the 'reverse' sequences gain more hydrophobic contacts upon binding (Figure 6.3a).

We now examine the cause of the greater number of hydrophobic contacts in the unbound state observed for the 'forward' sequences relative to the 'reverse' sequences (Figure 6.3b). First, the residue closest to the hydrophobic backbone forms a greater number of hydrophobic contacts than the residue at the end of the grafts. In the 'forward' sequences, P forms a greater number of contacts than V (black circle vs. red square in *f* columns in Figure 6.3b), and in the 'reverse' sequences V forms a greater number of contacts than P (red square vs. black circle in *r* columns in Figure 6.3b). Second, when P is near the hydrophobic backbone ('forward' sequences), it forms a greater number of hydrophobic contacts than V near the hydrophobic backbone (black circle in *f* columns vs. red square in *r* columns in Figure 6.3b). Third, in many sequences where V is placed at the end of the graft ('forward' sequences) V forms a significantly higher number of hydrophobic contacts than P at the end of the graft (red square in *f* columns vs. black circle in *r* columns in Figure 6.3b).

The first observation that the residue closest to the hydrophobic backbone makes more hydrophobic contacts is reasonable because the hydrophobic backbone contains nearly half of the total hydrophobic atoms of the polycation in a small region, facilitating the formation of a large number of contacts with nearby residues. The second observation that P near the backbone forms more contacts than

V near the backbone is somewhat surprising because V is considered more hydrophobic than P, and therefore V can be expected to form a greater number of hydrophobic contacts. We suggest that this observation may be explained by the greater conformational rigidity of P compared to V. Because proline's hydrophobic atoms are directly bonded to each other and to the peptide backbone, they are less flexible and less mobile than valine's hydrophobic sidechain. A related explanation is that the hydrogen atoms on the methyl groups of valine's side chain can rotate relatively freely, which may disrupt hydrophobic contacts, but the hydrogen atoms on proline's side chain have little conformational freedom. Due to the lower conformational freedom in proline, other hydrophobic groups (e.g., the hydrophobic backbone) might form longer-lasting contacts with proline than with valine. To support this conjecture, we quantify the average duration of hydrophobic contacts with P and with V, and we find that, in both the unbound and bound states, proline near the hydrophobic backbone does in fact form longer-lasting contacts than valine placed near the hydrophobic backbone (Figure 6.9 in supplementary section 6.6). Our observations on this trend highlight the known fact that the apparent hydrophobicity of peptides depend on the molecular-level details of their structure and local chemical environment: although valine is more hydrophobic than proline, in the local context of these polycations the rigidity of proline induces greater hydrophobic aggregation than occurs with valine [30, 31]. The third observation that V at the graft end forms more contacts than P at the graft end may be explained by the greater hydrophobicity of valine compared to proline: when located at the solvent-exposed end of the graft, hydrophobic contacts form to reduce interactions between valine and water. This third trend is weaker than the other two trends, which may be because the grafts are relatively inflexible, limiting the ability of the hydrophobic residue at the end of the graft to form contacts with the hydrophobic backbone. We speculate that this third trend may be of greater importance in polycations longer than the tetramers studied here: if the polycation is long enough to assume a coiled conformation where chemically distant monomers are spatially adjacent, the hydrophobic residue at the end of a graft on one coil of the polycation could aggregate with the hydrophobic backbone in another coil. The net effect of these three trends is that the sum of the proline

and valine hydrophobic contacts, and the total number of hydrophobic contacts, is larger for the 'forward' sequences than the 'reverse' sequences prior to binding.

After polycation-DNA binding (i.e., the bound state), two of the three trends in hydrophobic contacts remain: the residue closest to the hydrophobic backbone forms more contacts, and P forms more contacts when close to the hydrophobic backbone than V in the same position (Figure 6.3c). The continued presence of these trends in the bound state indicates that proximity to the backbone and the higher conformational rigidity of P, respectively, still exercise control over the behavior of the polycation in the bound state. Interestingly, in the bound state the two hydrophobic residues form the same number of contacts when placed at the end of the graft (red square in *f* columns vs. black circle in *r* columns in Figure 6.3c), so the weaker third trend that is present in the unbound state completely disappears in the bound state. The disappearance of this third trend suggests that the tendency of V to form hydrophobic contacts to reduce solvent exposure is overwhelmed by the formation of charged contacts between the polycation and the DNA. It also appears that polycation-DNA binding largely eliminates the dependence on the position of R of the number of hydrophobic contacts, whereas in the unbound state the presence and placement of R do affect the number of hydrophobic contacts. The latter observation suggests that R and/or K may be involved in the formation of hydrophobic contacts in the unbound state, and the former observation suggests that the formation of charged contacts in the bound state largely eliminates this postulated role of R and K in hydrophobic contacts. Overall, polycation-DNA binding changes the relative abundance of hydrophobic contacts, leading to greater hydrophobic aggregation for the 'reverse' sequences relative to the 'forward' sequences (Figure 6.3a), which in turn leads to the greater loss of conformational entropy for the 'reverse' sequences relative to the 'forward' sequences (Figure 6.2).

Charged contacts connected with conformational entropy loss. Because electrostatic contacts between oppositely charged groups (termed *charged contacts*) clearly play an important role in polycation-DNA binding, and because we postulate interactions between charged and hydrophobic residues, we now examine the effect of sequence on the formation of charged contacts. These contacts form between the positively charged peptides on the polycation (lysine and arginine) and the negatively charged groups in

the system (DNA phosphates and the peptide carboxyl-terminus, COO^-). Figure 6.4a shows a diagram of charged contacts with the peptide carboxyl-terminus, demonstrating the greater hydrogen bonding ability of arginine compared to lysine [10]. The formation of charged contacts, either within the polycation or between the polycation and DNA, is expected to tighten the polycation-DNA binding and thus reduce conformational entropy of the polycation and DNA.

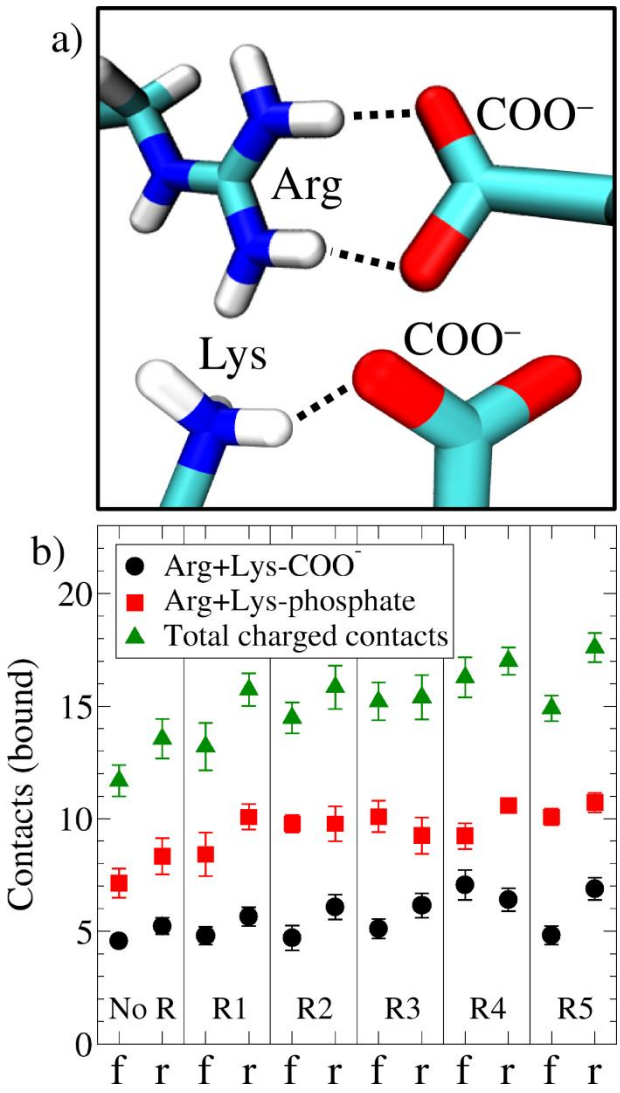


Figure 6.4 a) Diagram of intrapolycation charged contacts between the positively charged peptides arginine (Arg) and lysine (Lys) and the negatively charged peptide carboxyl-terminus (COO^-). b) Number of charged contacts in the bound state between Arg/Lys and the carboxyl-terminus (COO^-), between Arg/Lys and DNA phosphates, and the total of these two types of contacts. The peptide sequences represented in the x-axis are PKKKKKV/VKKKKKP ('No R' forward/reverse); PRKKKKV/VRKKKKP ('R1' forward/reverse); PKRKKKV/VKRKKKP ('R2' forward/reverse); PKKRKKV/VKKRKKP ('R3' forward/reverse); PKKKRKV/VKKKRKP ('R4' forward/reverse); and PKKKKRV/VKKKKRP ('R5' forward/reverse).

As the position of arginine shifts away from the hydrophobic backbone (i.e., R1 to R5), the number of charged contacts formed with the carboxyl-terminus (COO^-) increases slightly, because arginine interacts more favorably with negatively charged groups than does lysine and because arginine moves closer to the carboxyl-terminus (Figure 6.10 in supplementary section 6.6). Contrary to our expectation, as the position of arginine moves away from the hydrophobic backbone, the number of charged contacts with the phosphate groups of the DNA does not significantly increase (Figure 6.4b). Similarly, there is not a significant relationship between the position of arginine and the number of hydrogen bonds formed with DNA or the electrostatic attraction between arginine and DNA (Figure 6.11 in supplementary section 6.6). We note that the absence of a significant relationship between the position of arginine and arginine-DNA contacts is consistent with the absence of a significant relationship between the position of arginine and both the conformational entropy loss and the polycation-DNA electrostatic attraction, as discussed before. However, in accord with expectations, in the 'No R' case there are substantially fewer polycation-DNA charged contacts, suggesting that lysine cannot completely replace the contacts lost due to the removal of arginine. The stronger DNA-binding ability of arginine compared to lysine is further supported by the observation that, on a per-residue basis, arginine forms a greater number of charged contacts with phosphates, forms a greater number of hydrogen bonds with DNA, and has a greater electrostatic attraction to DNA (Figure 6.11 in supplementary section 6.6). These three observations are, as stated before, due to the diffuse positive charge in arginine's guanidinium group, which allows for stronger interaction with negatively charged groups than the localized positive charge of lysine's primary amine group.

In addition, in several cases ('No R', R1, and R5) the 'reverse' sequences form a significantly greater number of total charged contacts than the corresponding 'forward' sequences (Figure 6.4). To examine the origin of this difference, and considering our above suggestion that the formation of charged contacts disrupts the formation of hydrophobic contacts, we now investigate whether the formation of hydrophobic contacts can likewise disrupt the formation of charged contacts. Specifically, we quantify the

number of hydrophobic contacts between P/V and R/K with the hypothesis that a larger number of these hydrophobic contacts in the unbound state could limit the formation of charged contacts in the bound state. In one case (R1) where the 'reverse' sequence forms more charged contacts than the 'forward' sequence in the bound state (Figure 6.4b), we do find that the 'forward' sequence forms more hydrophobic contacts between P/V and R/K than the 'reverse' sequence in the unbound state (Figure 6.12 in supplementary section 6.6). However, in the other two cases where the 'reverse' sequence forms more charged contacts than the 'forward' sequence in the bound state, the 'reverse' sequence forms a greater number of hydrophobic contacts in the unbound state ('No R'), or the 'forward' and 'reverse' sequences form the same number of hydrophobic contacts (R5). Therefore, we cannot clearly explain why some 'reverse' sequences form a greater number of charged contacts than the corresponding 'forward' sequences. Nonetheless, these findings suggest that the formation of hydrophobic contacts in the unbound state does not significantly impede the formation of charged contacts in the bound state. As a caveat to this suggestion, we note that hydrophobic peptides containing aromatic groups (i.e. phenylalanine, tyrosine, and tryptophan) can participate in cation-π interactions with arginine and lysine, which may lead to stronger interactions of the charged peptides with those aromatic hydrophobic peptides than with the non-aromatic hydrophobic peptides in this work (i.e. proline and valine) [32].

In summary, both the greater number of charged contacts formed by some 'reverse' sequences and the differences in hydrophobic contacts described above appear to contribute to the greater loss of conformational entropy for the 'reverse' sequences compared to the 'forward' sequences (Figure 6.2). It is interesting that the difference between the 'forward' and 'reverse' sequences is caused by differences in both charged *and* hydrophobic contacts, since polyelectrolyte condensation is typically driven by electrostatics [6, 27]. Finally, it appears that direct interactions between the charged and hydrophobic residues appear to have only a weak effect on polycation-DNA binding.

6.3.2 Effects of Point Mutations to Proline and Valine on DNA Binding

To attempt to elucidate the effect of the hydrophobic groups on polycation-DNA binding, we explore the effects of point mutations to the hydrophobic residues, proline and valine. Specifically, we mutate both

hydrophobic residues to either P or V. To simultaneously examine the effect of the presence and position of arginine, we perform these mutations with three of the NLS-based polycations: the 'No R' case, where arginine is absent; the R2 case, which is asymmetric with respect to arginine placement within the graft; and the R3 case, which is symmetric with respect to arginine placement within the graft. The 'No R' case tests the behavior in the absence of R. The R2 and R3 cases test the effect of R when it is placed closer to one of the hydrophobic residues than the other and when it is placed equidistantly from both hydrophobic residues, respectively. We have also simulated the R1, R4, and R5 sequences with these point mutations to P and V, but these additional sequences do not contribute information beyond what we find with the 'No R', R2, and R3 sequences, so we exclude R1, R4, and R5 from this discussion for the sake of brevity. We present the behavior of the 'No R', R2, and R3 mutated polycations in Figure 6.5, where the polycations with the indicated placement of arginine are labeled P-P, P-V ('forward' sequences in previous sections), V-P ('reverse' sequences in previous sections), and V-V. The hyphen in each label represents the sequence of the central charged residues, which is KKKKK for the 'No R' cases, KRKKK for the R2 cases, and KKRKK for the R3 cases. For example, P-P 'No R' denotes PKKKKKP sequence along the graft, and V-P 'R2' denotes VKRKKKP sequence along the graft.

Conformational entropy loss upon polycation-DNA binding. We present the conformational entropy loss upon polycation-DNA binding, $T\Delta S_{conf}$, in Figure 6.5a. We note that the central sequences (P-V/V-P) for each placement of R in Figure 6.5a are the 'forward' and 'reverse' sequences from the previous sections. As stated previously, for the 'No R' cases there is no difference in $T\Delta S_{conf}$ between 'forward' (P-V) and 'reverse' (V-P) sequences, but, when R is present (R2/R3), we observe that the V-P sequences lose a greater amount of entropy than the P-V sequences. Now considering the mutated P-P/V-V sequences, regardless of the presence or placement of R within the graft, all of the P-P and V-V sequences show an equal loss of conformational entropy within the uncertainty. For the mutated P-P and V-V sequences, the presence and placement of R and the identity of the hydrophobic residue do not affect $T\Delta S_{conf}$. This observation is in contrast with the P-V and V-P cases, where the placement of the hydrophobic residues does affect the conformational entropy loss in the presence of arginine.

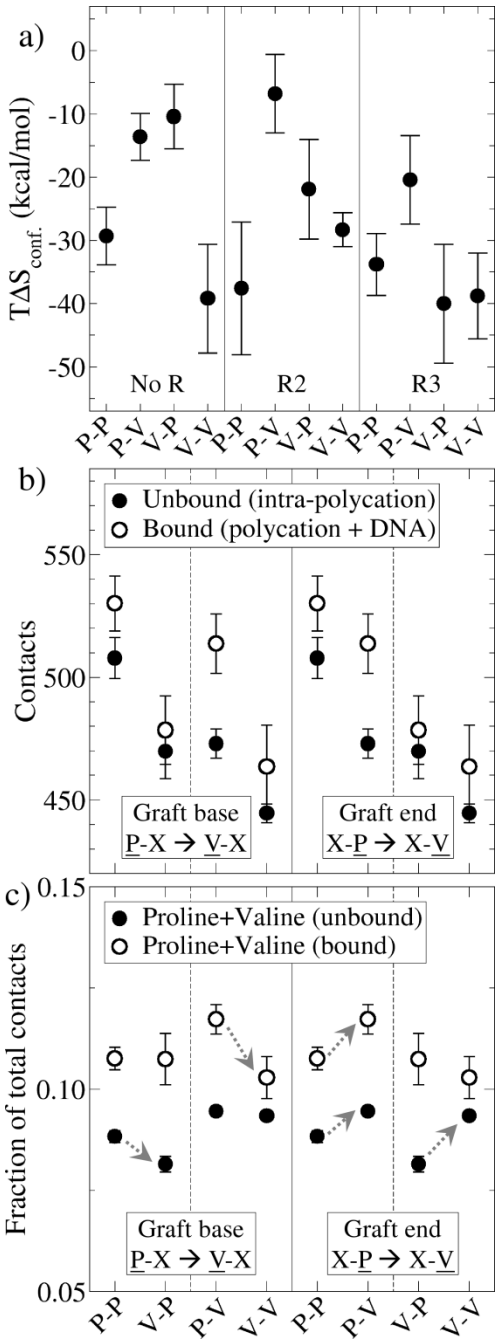


Figure 6.5 Behavior of NLS-based polycations with altered hydrophobicity. a) Conformational entropy change upon polycation-DNA binding. b) Total number of hydrophobic contacts before and after binding for the 'No R' case, arranged to facilitate comparisons of the effects of point mutations to P and V. c) Fraction of the total number of hydrophobic contacts that are formed by P and V, arranged to facilitate comparisons of the effects of point mutations. The peptide sequences represented in the x-axis of b) and c) are PKKKKKP/PKKKKKV/VKKKKKP /VKKKKKV ('No R'); PKRKKKP/PKRKKKV/VKRKKKP /VKRKKKV (R2); and PKKRKKP/PKKRKKV/VKKRKKP /VKKRKKV (R3).

We now discuss possible interpretations of the observed trends in $T\Delta S_{conf}$. Our expectation for the behavior of the symmetric P-P/V-V sequences was that, rather than showing similar behavior when compared to each other, they would behave as extreme cases compared to the asymmetric P-V/V-P sequences. That is, the P-V/V-P cases would show intermediate entropy loss, while one of the P-P/V-V cases would show much larger entropy loss than the P-V/V-P cases, and the other P-P/V-V case would show much smaller entropy loss than P-V/V-P. As a thought experiment to expand upon this idea, the P-P case would exhibit more 'proline-like' behavior and the V-V case would exhibit more 'valine-like' behavior compared to the P-V/V-P cases. The 'proline-like' and 'valine-like' behaviors in this thought experiment could respectively correspond to 'less hydrophobic' and 'more hydrophobic' behaviors, for example. Since our above observations for the P-V/V-P cases suggest that greater hydrophobic aggregation before binding is associated with a smaller loss of conformational entropy upon binding, our expectation would then be a greater loss of conformational entropy for the less hydrophobic P-P case and a smaller loss for the more hydrophobic V-V case. However, we do not observe that the symmetric P-P/V-V cases behave as extremes of the asymmetric P-V/V-P cases. Rather, we observe that the symmetric P-P/V-V sequences behave similarly to each other but differently compared to the asymmetric P-V/V-P sequences. Therefore, we cannot clearly explain the observed trends in the conformational entropy loss. Leaving aside the effect of the placement of P and V on $T\Delta S_{conf}$ in the context of these mutated polycations, we now examine the formation of inter- and intramolecular contacts to further characterize the behavior of these polycations.

Formation of charged contacts upon polycation-DNA binding. We first briefly comment on the formation of charged contacts. In the previous section, we observe that the 'No R' P-V/V-P cases form fewer charged contacts than the sequences containing arginine. Similarly, we find that the 'No R' P-P/V-V cases form significantly fewer charged contacts than the R2 or R3 P-P/V-V cases (Figure 6.13 in supplementary section 6.6). In the previous section, we also observe that several of the 'reverse' (V-P) sequences form more charged contacts than the corresponding 'forward' (P-V) sequences. However, the mutated polycations that are our focus in this section do not show any coherent trends in the number of

charged contacts with the placement or chemistry of the hydrophobic residues. In fact, almost all of the sequences with a given placement of R form the same number of charged contacts within the uncertainty (Figure 6.13 in supplementary section 6.6). Despite the lack of differences in the number of charged contacts between P-P and V-V polycations, we examine the number of hydrophobic contacts between the hydrophobic (P/V) and charged (R/K) residues with the hypothesis that hydrophobic contacts with the charged residues may limit the ability of the charged residues to form charged contacts. However, we find no correlation between the formation of these hydrophobic contacts in the unbound state (Figure 6.14 in supplementary section 6.6) and the formation of charged contacts in the bound state (Figure 6.13 in supplementary section 6.6). In summary, the placement and chemistry of the hydrophobic residues do not appear to influence the behavior of the charged residues.

Effects of hydrophobic residue sequence and chemistry on hydrophobic aggregation. Finally, we discuss the effects of peptide sequence on hydrophobic aggregation and on polycation-DNA binding for these mutated polycations. In the previous sections, we observed that the V-P ('reverse') sequences gain a greater number of hydrophobic contacts upon binding than the P-V ('forward') sequences, which is connected with greater conformational entropy losses for the V-P ('reverse') sequences. Considering the current collection of sequences, however, we do not find coherent trends in the *change* in hydrophobic contacts upon binding, nor do we observe a correlation between hydrophobic contacts and the conformational entropy within the uncertainty of these simulations (Figure 6.15 in supplementary section 6.6). Nonetheless, we do find relationships between the *number* of hydrophobic contacts and the peptide sequence that are relevant for interpreting the behavior of these polycations.

In our previous discussion of P-V/V-P, we find that P near the hydrophobic backbone forms *more* contacts than V near the hydrophobic backbone, while P at the end of the graft forms *fewer* contacts than V at the end of the graft. This is rationalized on the basis of the relatively lower conformational flexibility of P and the relatively greater hydrophobicity of V, respectively. By including the mutated P-P and V-V sequences, we can now compare two broader types of point mutation to generalize these observations. These two point mutations are: mutating P near the backbone to V, irrespective of the hydrophobic

residue at the end of the graft (graft base, P-X → V-X); and mutating P at the end of the graft to V, irrespective of the hydrophobic residue near the hydrophobic backbone (graft end, X-P → X-V). The underlined residue undergoes the point mutation, and X denotes the other hydrophobic residue, which is unchanged. We show this comparison for the 'No R' case in Figure 6.5b, and for the R2 and R3 cases in (Figure 6.16 in supplementary section 6.6).

First, we consider the mutation of P at the hydrophobic backbone to V (P-X → V-X). We find that mutating P near the hydrophobic backbone to V significantly decreases the total number of hydrophobic contacts, regardless of R position and in both the bound and unbound states. This observation is consistent with our previous observation that V near the backbone forms fewer contacts than P at that location and can be similarly rationalized on the basis of the lesser conformational flexibility of P leading to longer-lasting hydrophobic contacts (Figure 6.9 in supplementary section 6.6). The greater hydrophobic aggregation with P than with V near the hydrophobic backbone demonstrates that the apparent hydrophobicity of a peptide can depend on the molecular-level details of both its structure and local environment [30, 31].

Second, we consider the mutation of P at the graft end to V (X-P → X-V). We find that mutating P near the graft end to V does not significantly change the total number of hydrophobic contacts in the unbound state and bound state (in most cases). This observation is not consistent with our previous observation that V at the end of the graft forms more hydrophobic contacts than P; if it were consistent, the X-P → X-V mutation would increase the number of contacts. However, if we examine the number of hydrophobic contacts made specifically by P and V (i.e., P/V-specific contacts), instead of the total number of hydrophobic contacts, we find that the X-P → X-V point mutation often, but not always, increases the number of P/V-specific hydrophobic contacts (Figure 6.17 in supplementary section 6.6) and, more importantly, significantly increases the *fraction of the total* number of hydrophobic contacts *that are P/V-specific* (Figure 6.5c). The fact that the fraction of P/V-specific contacts increases shows that the increase in P/V-specific contacts is not caused by changes in the total number of contacts but is actually caused by the X-P → X-V point mutation. So, although mutating P at the graft end to the more

hydrophobic V generally does not change the *total* number of hydrophobic contacts (Figure 6.5b), the mutation does increase the relative importance of hydrophobic contacts made by the two hydrophobic residues (Figure 6.5c). This finding further supports the idea that valine, a highly hydrophobic residue, seeks out hydrophobic contacts to limit solvent exposure when placed at the end of the grafts. The fact that the *total* number of hydrophobic contacts is apparently insensitive to the X-P → X-V point mutation may be due to the small number of P/V-specific contacts relative to the total, which limits our ability to discern the relatively small sequence-dependent differences in the total. Additionally, the weak effect of the X-P → X-V point mutation may be due to the rigidity of the peptide backbone of the grafts, which may impede the formation of hydrophobic contacts between the hydrophobic residue at the end of the graft and the hydrophobic backbone. As in the discussion of Figure 6.3b, we speculate that the effect of the X-P → X-V point mutation may be of greater importance in polycations longer than the tetramers studied here: in a longer polycation, hydrophobic contacts may be able to form between chemically distant but spatially adjacent portions of the polycation, circumventing the restriction imposed by the rigid peptide backbone. Finally, there is a noteworthy nuance to our observations about the X-P → X-V point mutation: valine at the graft end forms more P/V-specific hydrophobic contacts only in the unbound state, and this trend (mostly) disappears in the bound state (Figure 6.17 in supplementary section 6.6). As noted in our discussion of Figure 6.3c, this suggests that the formation of charged contacts upon polycation-DNA binding can disrupt the formation of hydrophobic contacts.

6.4 Conclusions

We have used atomistic molecular dynamics simulations of grafted polycations containing peptides based on a nuclear localization sequence (NLS) to examine the effect of peptide sequence and chemistry on polycation-DNA binding behavior. Starting with a naturally occurring NLS peptide sequence (PKKKRKV), we examine the effect of the position of arginine (R) along the graft. We also examine the effect of the position of the hydrophobic residues with respect to the backbone of the polycation by examining 'forward' and 'reverse' versions of the NLS by exchanging proline (P) and valine (V). Then,

we further probe the effect the hydrophobic residues by replacing both of the hydrophobic residues (P and V) with either proline or valine. We describe the DNA-binding behavior of these polycations by calculating the conformational entropy loss upon binding, the extent of aggregation between the hydrophobic species, and the formation of electrostatic contacts within and between polycation and DNA.

We find that the presence of arginine affects the conformational entropy loss. When arginine is absent, the polycation binds less tightly to the DNA and consequently loses less conformational entropy upon binding to DNA as compared to when arginine is present. This difference is caused by the diffuse charge distribution of arginine's guanidinium group, which we find to increase the electrostatic attraction between the polycation and DNA through increased hydrogen bonds and electrostatic contacts with DNA as compared to the more localized charge of lysine's primary amine group, in agreement with previous findings [10, 11]. Surprisingly, the placement of arginine within the grafts does not significantly affect the loss of conformational entropy or the gain of electrostatic interactions upon polycation-DNA binding. Our expectation was that moving arginine farther from the hydrophobic backbone would make arginine more available to interact with the DNA, which would increase the polycation-DNA electrostatic attraction and, due to tighter DNA-polycation binding, would increase the conformational entropy loss. However, moving arginine farther from the hydrophobic backbone does not increase arginine-DNA electrostatic interactions or affect the conformational entropy loss.

We find that the positions of the hydrophobic proline and valine residues affect the conformational entropy loss, which is interesting because polycation-DNA binding is typically dominated by electrostatic interactions [6, 27]. Placing the more rigid proline residue near the hydrophobic backbone ('forward' sequences) allows longer-lasting hydrophobic contacts to form with proline, which causes the polycation in the absence of DNA (i.e., the *unbound* state) to form a greater number of hydrophobic contacts with itself relative to 'reverse' sequences, in which valine is near the hydrophobic backbone. Conversely, placing the more hydrophobic valine residue at the end of the graft ('forward' sequences) where it is more exposed to solution significantly increases the number of hydrophobic contacts the polycation forms with itself in the unbound state relative to 'reverse' sequences, in which proline is at the

end of the graft. These two trends in the hydrophobic contacts are further supported by our observations on the effects of point mutations to proline and valine on hydrophobic contacts. Relative to 'reverse' sequences, these two trends cause 'forward' sequences to form more hydrophobic contacts in the unbound state, which causes 'forward' sequences to gain fewer hydrophobic contacts upon binding, which is related to the smaller conformational entropy losses observed for 'forward' sequences.

Based on these findings, we suggest that altering the sequence or chemistry of the grafts to increase hydrophobic aggregation in the unbound state will decrease the conformational entropy loss that occurs upon polycation-DNA binding. Since we also find that hydrophobic contacts with the charged residues in the unbound state do not impede the ability of the charged residues to form electrostatic contacts in the bound state, we suggest that chemical changes to hydrophobic residues may influence the conformational entropy loss by changing hydrophobic aggregation while having minimal impact on the electrostatic interactions that dominate polycation-DNA binding strength. We note that the inclusion of aromatic peptides, which are capable of participating in cation-π interactions with Arg/Lys, may obscure the effects of hydrophobic aggregation [32]. This complication notwithstanding, we suggest that judicious changes to polycation hydrophobicity may provide a method to tailor the strength of polycation-DNA binding.

Based on this suggestion, we propose two ways in which the conformational entropy of these NLS-based polycations in the unbound state might be tailored by mutating the hydrophobic residues to change hydrophobic aggregation. First, since the greater conformational rigidity of proline increases hydrophobic aggregation when proline is placed near the hydrophobic backbone, other rigid hydrophobic residues with greater rigidity or greater hydrophobicity may increase aggregation in the unbound state, thereby decreasing the conformational entropy loss upon binding. Since the structure and rigidity of proline are unique among the naturally occurring amino acids, a synthetically modified proline residue may be required. The side chain of this synthetic proline residue could be modified by the addition of alkyl substituents to increase hydrophobicity or by the addition of an aromatic ring to increase both hydrophobicity and rigidity. However, as noted above, the design of a novel polycation using this

suggestion may be complicated by the introduction of cation-π interactions between Arg/Lys and aromatic peptides, which can be energetically comparable to hydrogen bonds [32]. Second, limiting the conformational entropy of the polycation in the absence of DNA could be accomplished by including two hydrophobic groups that are distant from each other in the chemical structure but that are free to associate; in the 'forward' NLS-based polycations studied here, these two hydrophobic groups are the poly(cyclooctene) backbone and the valine residues at the end of the grafts. Mutating valine to a more hydrophobic residue, such as leucine or isoleucine, may increase hydrophobic aggregation in the unbound state and decrease conformational entropy loss. Conversely, mutating valine to a less hydrophobic residue, such as alanine, may decrease the hydrophobic aggregation in the unbound state and increase conformational entropy loss. These modifications to graft chemistry might be used to tailor the strength of polycation-DNA binding.

While we have focused purely on the effect of peptide sequence and chemistry on DNA binding, a larger set of polycation design parameters may affect the DNA-binding behavior of these grafted polycations, including graft length, graft spacing, and the chemical nature of the polycation backbone. The graft length could be changed by adding or removing lysine residues, which may reveal a non-monotonic trend in polycation-DNA binding strength with graft length similar to findings in previous studies of oligolysine-based grafted polycations [5, 6]. The hydrophobic poly(cyclooctene) backbone could be replaced with a hydrophilic poly(ethylene glycol) backbone, which would likely decrease the importance of hydrophobic aggregation in polycation-DNA binding and, similar to what has been observed with poly(ethylene imine)-*g*-poly(ethylene glycol) polycations [33], may decrease the strength of polycation-DNA binding. The graft spacing could be decreased by using a shorter poly(cyclopentene) backbone, which, by increasing the charge density of the polycation, may increase the stability of polycation-DNA complexes. Ongoing work focuses on exploring these design parameters using computational tools to help guide experimental synthesis and transfection efforts in the most promising directions.

6.5 References

[1] M. Thomas and A. M. Klibanov, Applied Microbiology and Biotechnology 62 (2003) 27.
[2] D. W. Pack, A. S. Hoffman, S. Pun, and P. S. Stayton, Nature Reviews Drug Discovery 4 (2005) 581.
[3] R. B. Breitenkamp, Z. Ou, K. Breitenkamp, M. Muthukumar, and T. Emrick, Macromolecules 40 (2007) 7617.
[4] R. B. Breitenkamp and T. Emrick, Biomacromolecules 9 (2008) 2495.
[5] S. S. Parelkar, D. Chan-Seng, and T. Emrick, Biomaterials 32 (2011) 2432.
[6] R. M. Elder, T. Emrick, and A. Jayaraman, Biomacromolecules 12 (2011) 3870.
[7] R. M. Elder and A. Jayaraman, Macromolecules 45 (2012) 8083.
[8] M. Nishikawa and L. Huang, Human Gene Therapy 12 (2001) 861.
[9] A. Lange, R. E. Mills, C. J. Lange, M. Stewart, S. E. Devine, and A. H. Corbett, Journal of Biological Chemistry 282 (2007) 5101.
[10] M. J. Betts and R. B. Russell, in Bioinformatics for Geneticists (M. R. Barnes and I. C. Gray, eds.), John Wiley & Sons, Ltd, Chichester, UK, 2003, p. 289.
[11] K. G. Wagner and H.-A. Arfmann, European Journal of Biochemistry 46 (1974) 27.
[12] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, Journal of Molecular Graphics and Modelling 25 (2006) 247.
[13] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Journal of Computational Chemistry 25 (2004) 1157.
[14] A. Jakalian, D. B. Jack, and C. I. Bayly, Journal of Computational Chemistry 23 (2002) 1623.
[15] I. S. Joung and T. E. Cheatham, Journal of Physical Chemistry B 112 (2008) 9020.
[16] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco, Biophysical journal 92 (2007) 3817.
[17] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, Journal of the American Chemical Society 117 (1995) 5179.
[18] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins 65 (2006) 712.
[19] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics and Modelling 14 (1996) 33.
[20] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of Computational Chemistry 26 (2005) 1781.
[21] C. Andersen, Journal of Computational Physics 52 (1983) 24.
[22] T. Darden, D. York, and L. D. A. J. Pedersen, Journal of Chemical Physics 98 (1993) 10089.
[23] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case, Journal of the American Chemical Society 120 (1998) 9401.
[24] D. W. Li and R. Bruschweiler, Physical Review Letters 102 (2009) 118108.
[25] D. W. Li and R. Bruschweiler, Journal of the American Chemical Society 131 (2009) 7226.
[26] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, Accounts of Chemical Research 33 (2000) 889.
[27] Z. Y. Ou and M. Muthukumar, Journal of Chemical Physics 124 (2006) 154902.
[28] J. DeRouchey, B. Hoover, and D. C. Rau, Biochemistry 52 (2013) 3000.
[29] J. R. Fromm, R. E. Hileman, E. E. O. Caldwell, J. M. Weiler, and R. J. Linhardt, Archives of Biochemistry and Biophysics 323 (1995) 279.
[30] M. C. J. Wilce, M.-I. Aguilar, and M. T. W. Hearn, Analytical Chemistry 67 (1995) 1210.
[31] K. A. Dill, Journal of Biological Chemistry 272 (1997) 701.
[32] A. S. Mahadevi and G. N. Sastry, Chemical Reviews 113 (2012) 2100.

[33] S. Mao, M. Neu, O. Germershaus, O. Merkel, J. Sitterberg, U. Bakowsky, and T. Kissel, Bioconjugate Chemistry 17 (2006) 1209.

6.6 Supplementary information

6.6.1 Insensitivity of number of hydrophobic contacts to cut off distance

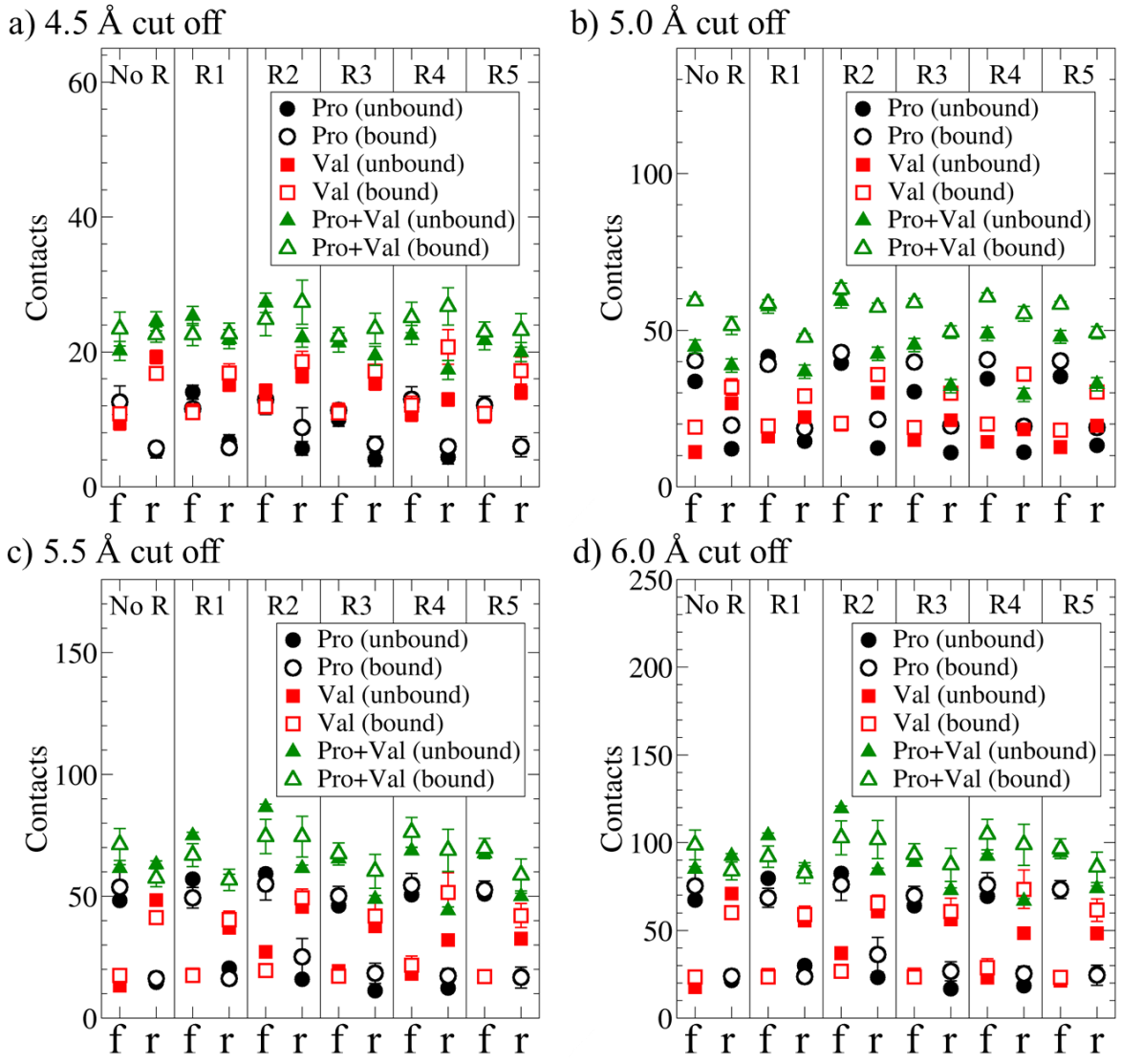


Figure 6.6 Comparison of the number of hydrophobic contacts formed with Pro and Val with differing values of the cut off distance used to define a hydrophobic contact. Panels a) through d) correspond to cut off distances of 4.5, 5.0, 5.5, and 6.0 Å, respectively. A cut off distance of 5.0 Å is used in all figures in the other sections of this chapter. We find that the differences in hydrophobic contacts caused by differences in peptide sequence that are evident with a cut off distance of 5.0 Å are, by and large, also evident with cut off distances of 4.5, 5.5, or 6.0 Å. These differences include: Pro forms a greater number of hydrophobic contacts in 'forward' sequences, while Val forms a greater number of hydrophobic

contacts in 'reverse' sequences (both unbound and bound states); Pro placed near the backbone ('forward' sequences) forms a greater number of hydrophobic contacts compared to Val near the backbone ('reverse' sequences); and forward sequences form a greater number of combined Pro+Val hydrophobic contacts in the unbound state. See also Figure 6.9, which discusses the insensitivity of the duration of hydrophobic contacts with Pro/Val to the cut off distance.

6.6.2 Energetic and entropic contributions to polycation-DNA binding strength

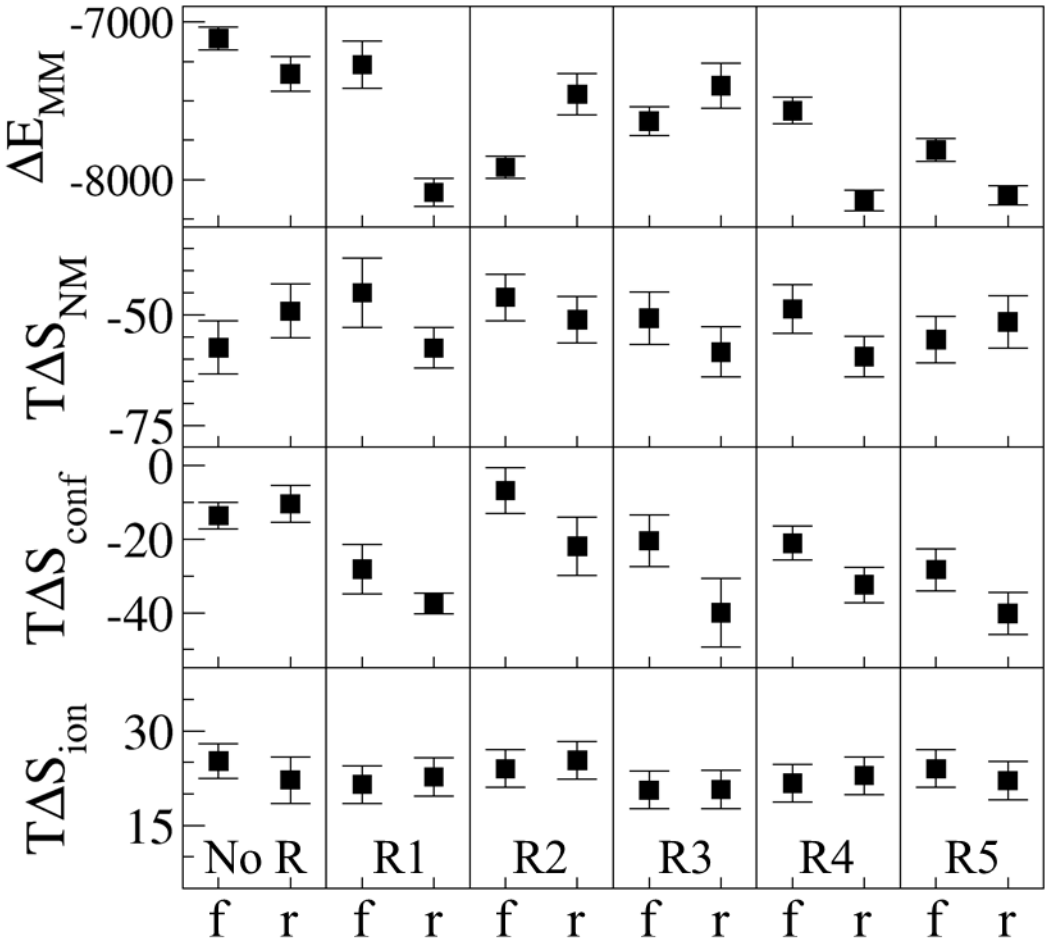


Figure 6.7 The change in molecular mechanics energy upon binding (ΔE_{MM}) is calculated in the absence of water and counterions and is therefore dominated by the electrostatic attraction between the polycation and the DNA. The magnitude of ΔE_{MM} is smallest (i.e., least favorable) for the 'No R' case where arginine is absent and is significantly larger ($p \approx 0.07$) when R is present (i.e., R1 to R5). As discussed in the main manuscript, the reason for this trend is the diffuse charge distribution of arginine's positively charged guanidinium group, which binds to the DNA more effectively than lysine's positively charged primary amine. When R is present, the magnitude of ΔE_{MM} does not change significantly as R is moved out along the graft: including R1, R2, R3, R4, and R5, $p = 0.34$; excluding R1, $p = 0.23$. Apparently, the position of R within the grafts does not significantly affect the DNA-binding ability of these polycations. This finding is further corroborated by the lack of a relationship between the position of arginine and various types of interactions between arginine and DNA (i.e., electrostatic contacts, hydrogen bonds,

energy), as shown in Figure 6.11. We note that E_{MM} does not include interactions with counterions or with water molecules; it only includes inter- and intramolecular interactions of the polycation and DNA. The ion dissociation entropy ($T\Delta S_{ion}$) and vibrational entropy ($T\Delta S_{NM}$), with the subscript NM signifying the normal mode analysis used to calculate vibrational entropy, are not significantly affected by the graft sequence. The conformational entropy loss ($T\Delta S_{conf}$) is shown for comparison with the other terms and is discussed in detail in the main manuscript. Considering sequences containing arginine, there is not a significant effect of moving R out along the graft (i.e., R1 → R5). However, there is a significant increase in the conformational entropy loss between the 'No R' case and the sequences containing R ($p \approx 0.08$).

6.6.3 Effect of arginine, proline, and valine position on change in total number of hydrophobic contacts upon polycation-DNA binding

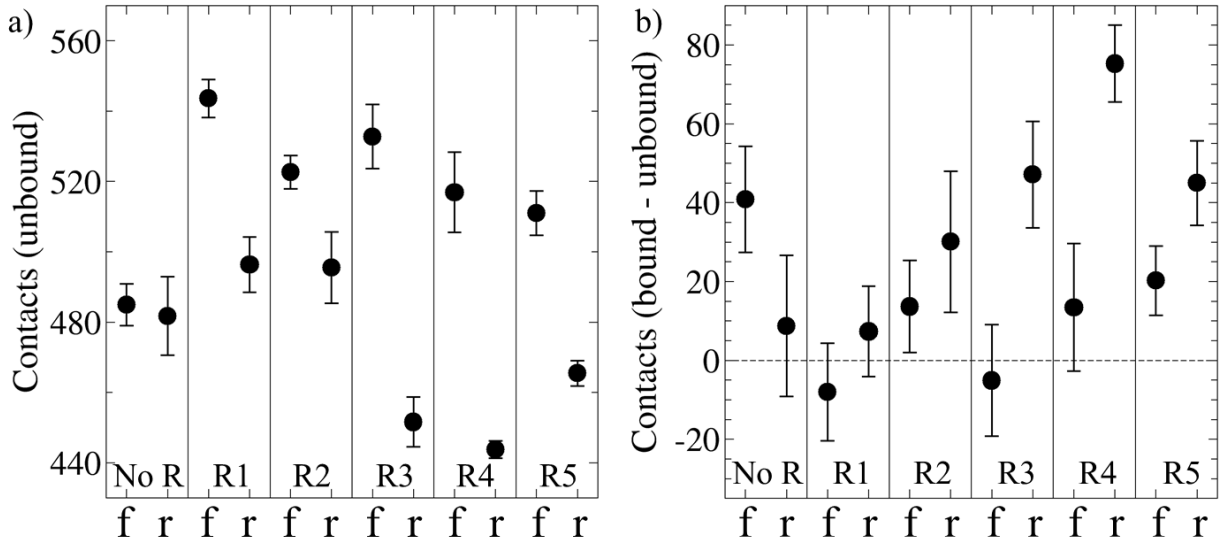


Figure 6.8 a) The total number of hydrophobic contacts in the unbound state. b) The change upon binding in the total number of hydrophobic contacts. Both quantities show similar trends with peptide sequence as those exhibited by the number of hydrophobic contacts formed specifically by proline and valine, which is shown in Figure 3 of the main manuscript. 'Reverse' sequences have fewer hydrophobic contacts prior to binding than 'forward' sequences, and consequently 'reverse' sequences tend to gain more hydrophobic contacts upon binding than 'forward' sequences.

6.6.4 Duration of hydrophobic contacts with proline and valine

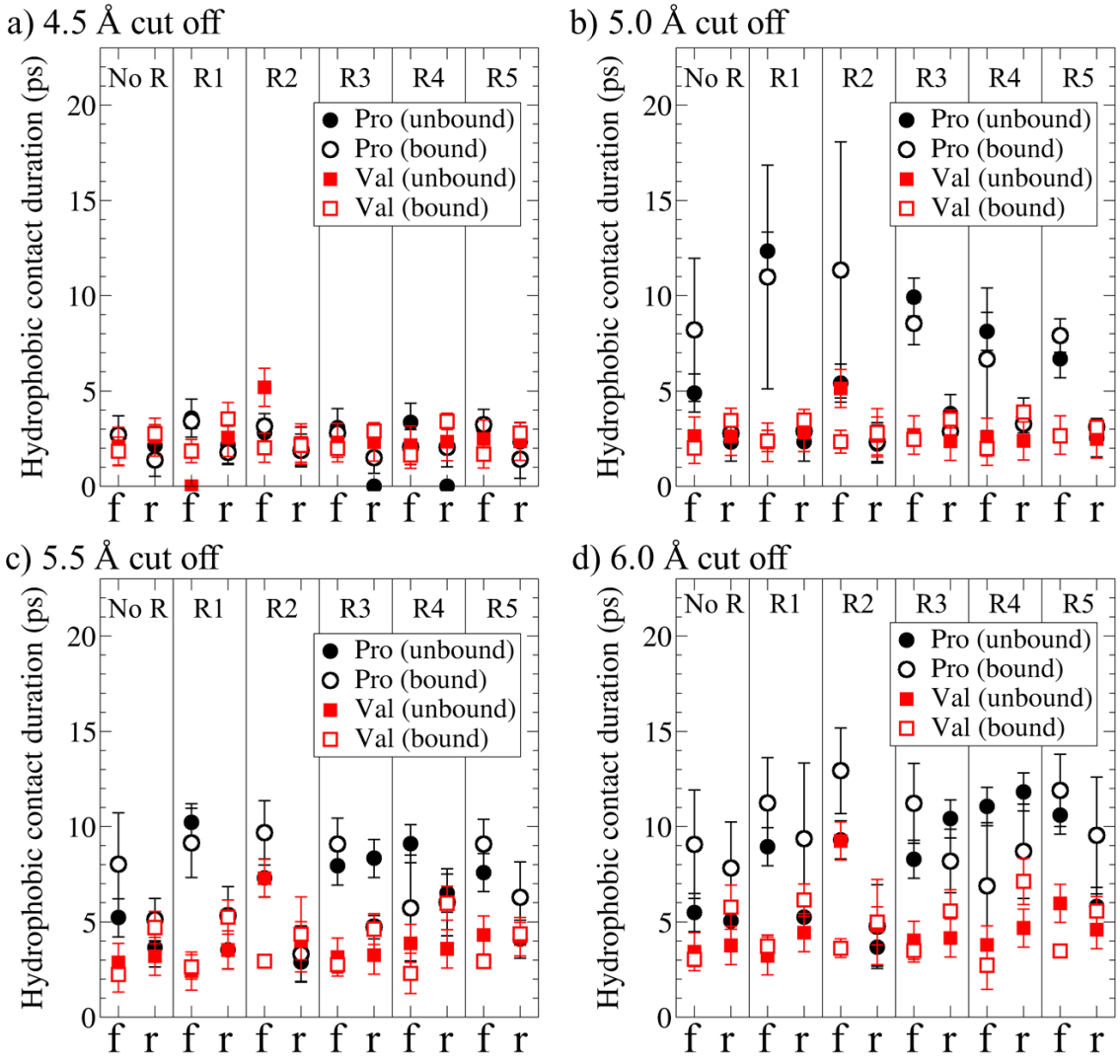


Figure 6.9 Average duration of hydrophobic contacts with proline (Pro) and valine (Val) at various cut off distances. Panels a) through d) correspond to cut off distances of 4.5, 5.0, 5.5, and 6.0 Å, respectively. A cut off distance of 5.0 Å is used in all figures in the other sections of this chapter. a) With the short cut off distance of 4.5 Å, all hydrophobic contacts have approximately the same short duration, which may be related to the relatively small number of contacts observed using this cut off (Figure 6.6). b) With a slightly larger cut off of 5.0 Å, when Pro is placed close to the hydrophobic backbone (black circles in 'forward', or *f*, sequences), the duration of its hydrophobic contacts is greater than hydrophobic contacts with Val when it is placed close to the hydrophobic backbone (red squares in 'reverse', or *r*, sequences). This finding suggests that the lower conformational rigidity of Pro compared to Val allows other hydrophobic groups (e.g., the hydrophobic poly(cyclooctene) backbone) to form longer-lasting contacts with Pro compared to Val, leading to the greater number of contacts forming with Pro compared to Val when placed near the hydrophobic backbone (main manuscript Figure 3). Still considering panel b), we also note that both Pro and Val at the end of the graft form equally short-lived contacts, which we

speculate is because the environment at the end of the grafts is more dynamic than the environment near the hydrophobic backbone. For example, perhaps water molecules and counterions frequently interact with the negatively charged carboxyl-terminus at the end of the graft, disrupting hydrophobic contacts. Panels c) and d), with the larger cut off distances of 5.5 and 6.0 Å, respectively, show essentially the same behavior as discussed for panel b). Compared to panel b), the main difference in panels c) and d) is that Pro at the end of the graft forms longer-lived contacts than Val at the end of the graft in many cases, which is further evidence that longer-lived contacts can form with Pro than with Val.

6.6.5 Number of intrapolycation charged contacts formed by arginine and lysine with peptide carboxyl groups

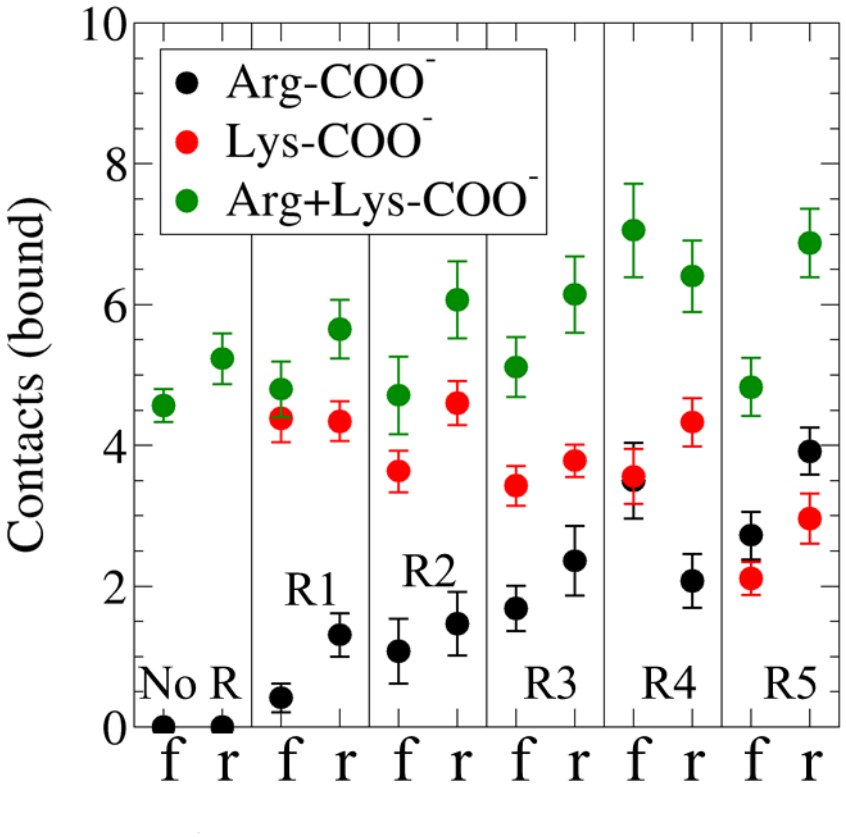


Figure 6.10 Charged contacts between arginine (Arg), lysine (Lys), and the peptide carboxyl-terminus (COO^-). The total number of these contacts (Arg+Lys-COO^-) contacts increases as Arg moves closer to the graft end, where the carboxyl-terminus is located. This upward trend is because the increase in Arg-COO^- contacts as Arg moves closer to the end of the grafts is greater than the decrease in Lys-COO^- contacts as Arg moves closer to the end of the graft, which is due to the stronger interactions of arginine with negatively charged groups.

6.6.6 Arginine interacts more favorably with DNA than lysine

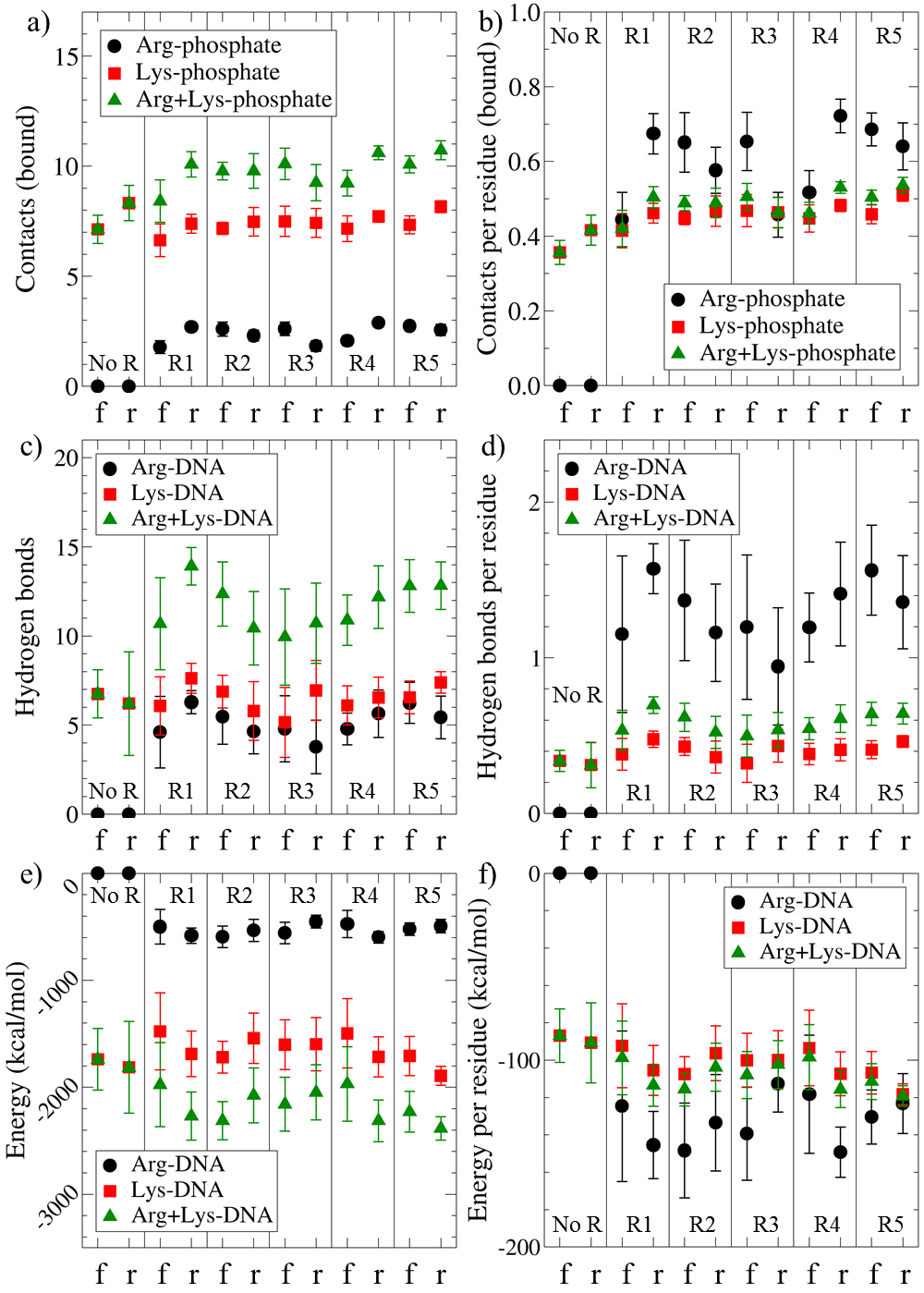


Figure 6.11 a) Number of charged contacts forming between arginine (Arg), lysine (Lys), and DNA phosphate groups. The elimination of Arg-phosphate contacts in the 'No R' case lowers the total number of polycation-DNA contacts. For sequences with R, the position of R along the graft does not affect the number of Arg-phosphate contacts, as discussed in the main manuscript. b) Number of Arg-phosphate and Lys-phosphate charged contacts divided by the number of those residues in the polycation. As expected,

Arg generally forms a greater number of contacts per residue than Lys, owing to the more diffuse charge distribution on the positively charged Arg side chain, as discussed in the main manuscript. c) Number of hydrogen bonds between Arg/Lys and DNA. We examined the number of hydrogen bonds with DNA to supplement the electrostatic contacts: even though a charged residue may not be engaged in an electrostatic contact with a phosphate group, it may still be forming electrostatically favorable hydrogen bonds with other portions of the DNA. The 'No R' case forms a significantly smaller number of hydrogen bonds with DNA than the cases containing R. For the sequences containing arginine, the arginine residues form approximately the same number of hydrogen bonds as the lysine residues, which is quite striking considering that there are four times as many lysine residues as there are arginine residues. d) Number of hydrogen bonds between Arg/Lys and DNA normalized to the number of Arg/Lys residues in the polycation. Each arginine residue forms a greater number of hydrogen bonds than each lysine residue, confirming the greater hydrogen bonding ability of arginine. e) Electrostatic energy between the side chains of Arg/Lys and the entire DNA molecule (i.e., not just phosphates). We calculate the electrostatic energy to supplement our observations on electrostatic contacts and hydrogen bonds: even though a positively charged residue may not meet the geometric criteria for a charged contact or a hydrogen bond, it may still be interacting favorably with the DNA. The total Lys-DNA electrostatic attraction is substantially greater than the Arg-DNA attraction, which is reasonable because there are more lysine residues than arginine residues. f) Electrostatic energy between the side chains of Arg/Lys and the entire DNA molecule normalized to the number of Arg/Lys residues in the polycation. Arginine residues generally show a greater electrostatic attraction to the DNA than lysine residues, further confirming the greater DNA-binding ability of arginine residues.

6.6.7 Effect of arginine, proline, and valine position on hydrophobic contacts formed between hydrophobic residues (P, V) and charged residues (R, K)

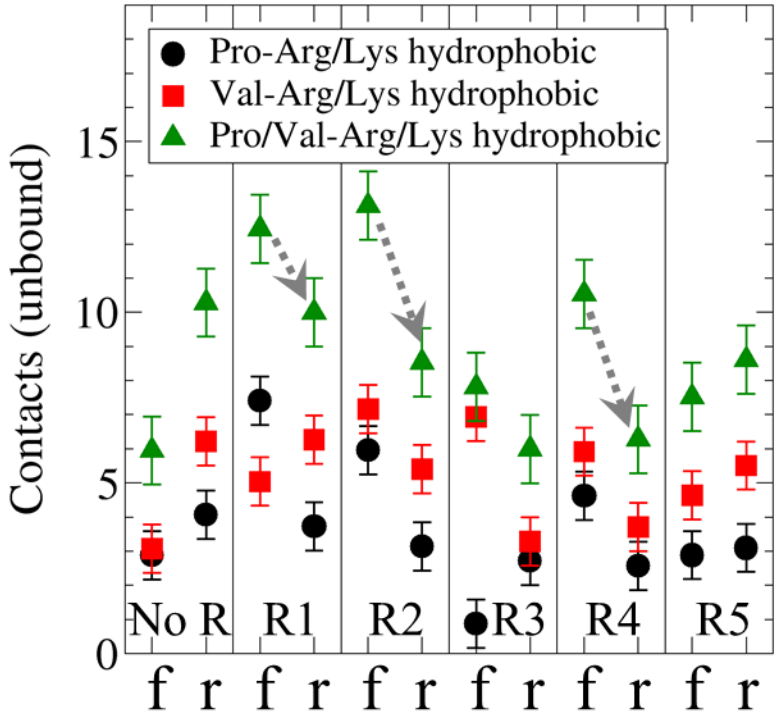


Figure 6.12 Number of hydrophobic contacts between the non-polar carbon atoms of Pro/Val and Arg/Lys in the unbound state. Our hypothesis is that a greater number of hydrophobic contacts with the charged residues R/K in the unbound state could limit their ability to form charged contacts with the DNA in the bound state. We focus on the hydrophobic contacts originating from P/V since changing the position of those residues appears to cause changes in charged contacts: the 'No R', R1, and R5 'forward' sequences form fewer charged contacts than the corresponding 'reverse' sequences (main manuscript Figure 4b). In the unbound state (shown here), we see that three of the 'reverse' sequences (R1, R2, R4) form *fewer* hydrophobic contacts between P/V and R/K than the 'forward' sequences, as indicated by the gray arrows. The smaller number of hydrophobic contacts for the R1 'reverse' sequences in the unbound state could explain why that sequence forms a greater number of charged contacts with the DNA in the bound state. However, in the other two cases where 'reverse' sequences form more charged contacts than the 'forward' sequences in the bound state, the 'reverse' sequence forms a *greater* number of hydrophobic contacts in the unbound state ('No R'), or the 'forward' and 'reverse' sequences form an *equal* number of hydrophobic contacts (R5). These findings suggest that the formation of hydrophobic contacts in the unbound state does not significantly impede the formation of charged contacts in the bound state.

6.6.8 Effect of point mutations to hydrophobic residues (P, V) on formation of charged contacts

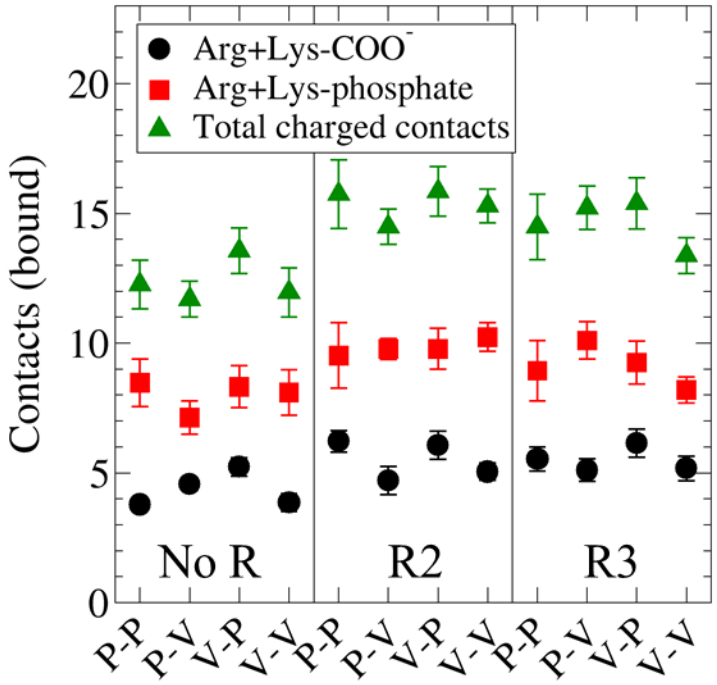


Figure 6.13 The number of charged contacts forming between arginine (Arg), lysine (Lys), the peptide carboxyl-terminus (COO^-), and DNA phosphate groups in the bound state. As noted in the main manuscript, the 'No R' cases form a significantly smaller number of charged contacts due to the absence of Arg. Comparing the R2 and R3 cases the position of R has little effect. The relative position and abundance of P and V residues similarly has little effect, suggesting that the hydrophobic residues have little effect on the charged residues in these particular polycations.

6.6.9 Effect of point mutations to hydrophobic residues (P, V) on hydrophobic contacts formed between hydrophobic residues (P, V) and charged residues (R, K)

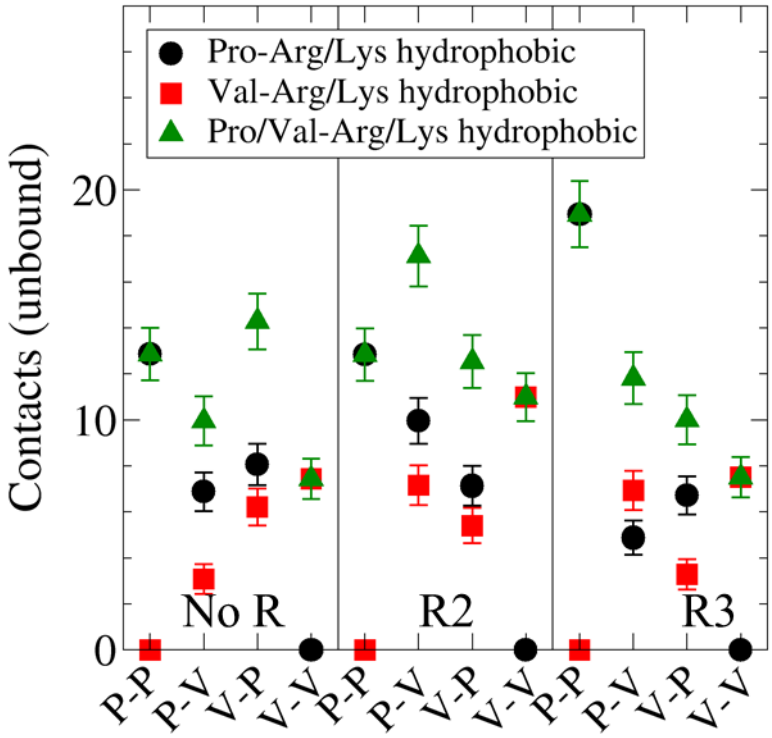


Figure 6.14 The number of hydrophobic contacts formed between proline and valine (Pro/Val) and arginine and lysine (Arg/Lys) in the unbound state. As in Figure 6.12, our intent is to examine whether the formation of these hydrophobic contacts in the unbound state hinders the formation of charged contacts by Arg/Lys in the bound state, which would be indicated by a large number of hydrophobic contacts for a particular sequence in this plot corresponding to a small number of charged contacts for the same sequence in Figure 6.13. We find no such correspondence, and conclude that the influence of the placement and chemistry of the hydrophobic residues on the charged residues appears to be small.

6.6.10 Effect of point mutations to hydrophobic residues (P, V) on change in total number of hydrophobic contacts upon polycation-DNA binding

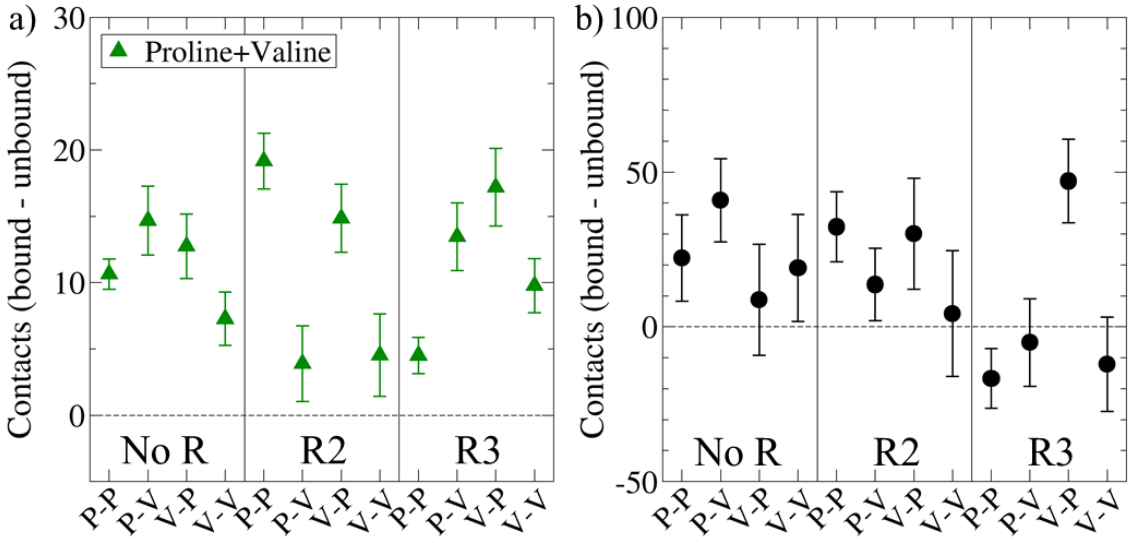


Figure 6.15 a) The change upon binding in the number of hydrophobic contacts made by proline and valine. b) The change upon binding in the *total* number of hydrophobic contacts. As stated in the main manuscript, we do not observe any coherent trends in the *change* in hydrophobic contacts. Nor do we observe a correlation between the change in the number of hydrophobic contacts and the conformational entropy loss (Figure 6.5).

6.6.11 Effect of point mutations to hydrophobic residues (P, V) on total number of hydrophobic contacts for R2 and R3 sequences

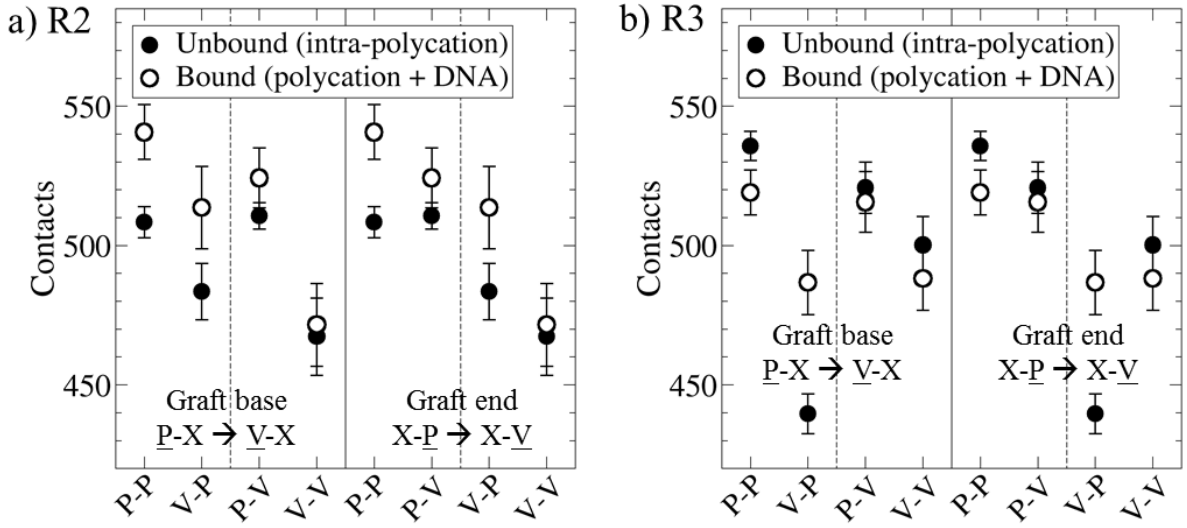


Figure 6.16 a) Total number of hydrophobic contacts before and after binding for the **R2** case, with the x-axis arranged to facilitate comparisons of the effects of point mutations to P and V. b) Total number of hydrophobic contacts before and after binding for the **R3** case, with the x-axis arranged to facilitate comparisons of the effects of point mutations to P and V. Compare this figure to main manuscript Figure 5b, which shows the same comparisons for the 'No R' sequences. We find that mutating P near the hydrophobic backbone to V (P-X → V-X) significantly decreases the total number of hydrophobic contacts, regardless of R position and in both the bound and unbound states. This trend can be rationalized on the basis of the lesser conformational flexibility of P compared to V, as discussed in the main manuscript. We find that mutating P near the graft end to V (X-P → X-V) does not significantly change the *total* number of hydrophobic contacts.

6.6.12 Effect of point mutations to hydrophobic residues (P, V) on hydrophobic contacts formed by proline and valine

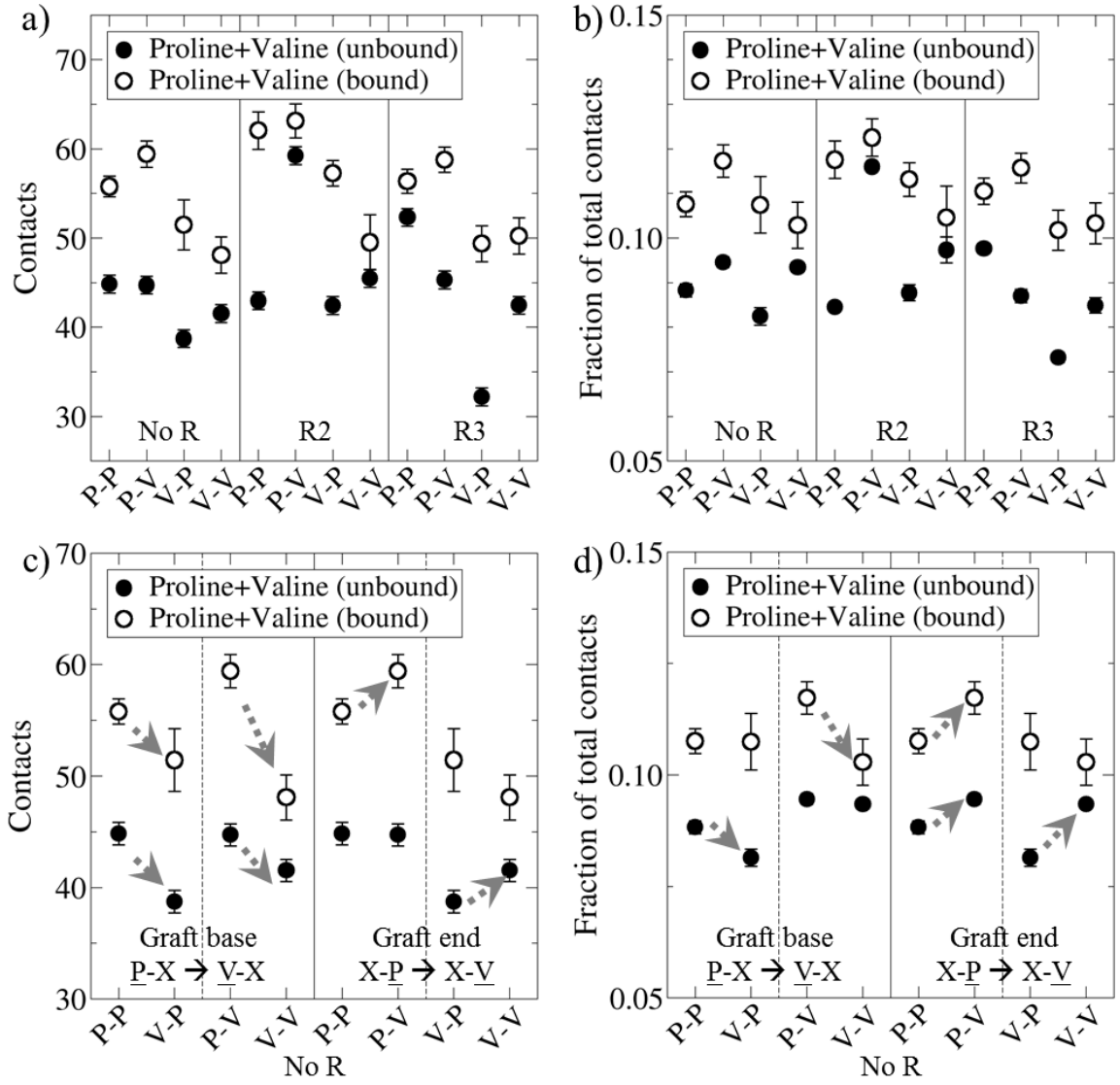


Figure 6.17 a) Number of hydrophobic contacts formed by proline and valine in the unbound and bound states. b) Number of hydrophobic contacts formed by proline and valine in the unbound and bound states *normalized to the total number of hydrophobic contacts*. c) Number of hydrophobic contacts formed by proline and valine in the unbound and bound states for the 'No R' case, arranged to facilitate comparisons of the indicated point mutations. d) Number of hydrophobic contacts formed by proline and valine in the unbound and bound states *normalized to the total number of hydrophobic contacts* for the 'No R' case, arranged to facilitate comparisons of the indicated point mutations. (Figure 6.17d is identical to main manuscript Figure 5c and is shown here for the purpose of comparison.) We find that changing P near hydrophobic backbone to V (P-X → V-X) significantly decreases the number of P/V hydrophobic contacts, regardless of R position (c). This trend appears in both the unbound and bound states. Similarly,

this point mutation decreases the fraction of the total hydrophobic contacts that are P/V-specific (d). We find that the X-P → X-V point mutation often, but not always, increases the number of P/V-specific hydrophobic contacts (c) and significantly increases ($p \approx 0.1$) the fraction of the total number of hydrophobic contacts that are P/V-specific (d). This mutation increases the relative importance of hydrophobic contacts made by the proline and valine, supporting the idea that valine, a highly hydrophobic residue, seeks out hydrophobic contacts to limit solvent exposure when placed at the end of the grafts. Interestingly, the effect of this point mutation only appears in the unbound state; the 'No R' case shown here is the only sequence displaying an effect of this point mutation in the bound state. By showing that the differences in hydrophobic contacts are eliminated upon binding, this observation supports the idea that the formation of charged contacts disrupts the formation of hydrophobic contacts.

Chapter 7

Structure and Thermodynamics of ssDNA Oligomers Adsorbing to Hydrophobic and Hydrophilic Surfaces

Adapted from: *Soft Matter* **9** 11521 (2013)

7.1 Introduction

Understanding the behavior of biomolecules near surfaces of varying chemistry has relevance for many applications, such as preventing protein adsorption in pharmaceutical production [1], reducing biofouling on marine vessels [2], controlling cell adhesion on implantable medical devices [3], assembly of DNA into two- and three-dimensional shapes (termed "DNA origami") [4, 5], and self-assembly of DNA-grafted nanoparticles into desired arrangements [6-10]. In many of these applications involving nucleic acids, it is desirable to use solid surfaces as templates to guide self-assembly through hybridization of single-stranded DNA (ssDNA) into double-stranded DNA (dsDNA) or to maintain the stability of DNA-based nanostructures near surfaces [11]. The behavior of nucleic acids near surfaces has been studied less extensively than their behavior in solution [12], and a fundamental understanding of the interactions of ssDNA with the substrate is needed to guide many of the above applications.

Past studies aimed at investigating the interactions of biomolecules with surfaces have shown that the complexity of the problem arises from the chemical and physical heterogeneity of biomolecules, which can contain positively and negatively charged groups, hydrogen bonding groups, and hydrophobic groups of varying size and hydrophobicity. The nature of the surface is also of critical importance: for example, many proteins readily adsorb to most hydrophobic surfaces but are largely repelled by most hydrophilic surfaces, with many exceptions and complicating factors [13-15]. In the case of DNA, some studies have shown that hydrophilic surfaces resist nonspecific adsorption and hydrophobic surfaces

promote adsorption [16, 17], while others have shown that DNA can adsorb to hydrophilic surfaces [18]. Clearly, DNA adsorption shows many of the same complexities as protein adsorption. However, in contrast to most proteins, DNA has a well-defined, regular chemical structure that allows for a more controlled exploration of the nuanced effects of surface hydrophobicity on this amphiphilic biomolecule.

Although generally more homogeneous in composition than amino acids, nucleic acids are fundamentally amphiphilic in nature. The nucleobases themselves are flat, relatively hydrophobic surfaces but also possess considerable hydrophilic character because of the presence of hydrogen bonding partners (e.g., the canonical Watson-Crick hydrogen bonds between base pairs). DNA as a whole is amphiphilic because it comprises the relatively hydrophobic bases and the relatively hydrophilic phosphate backbone (Figure 7.1a). These opposing chemical functionalities – hydrophilic and

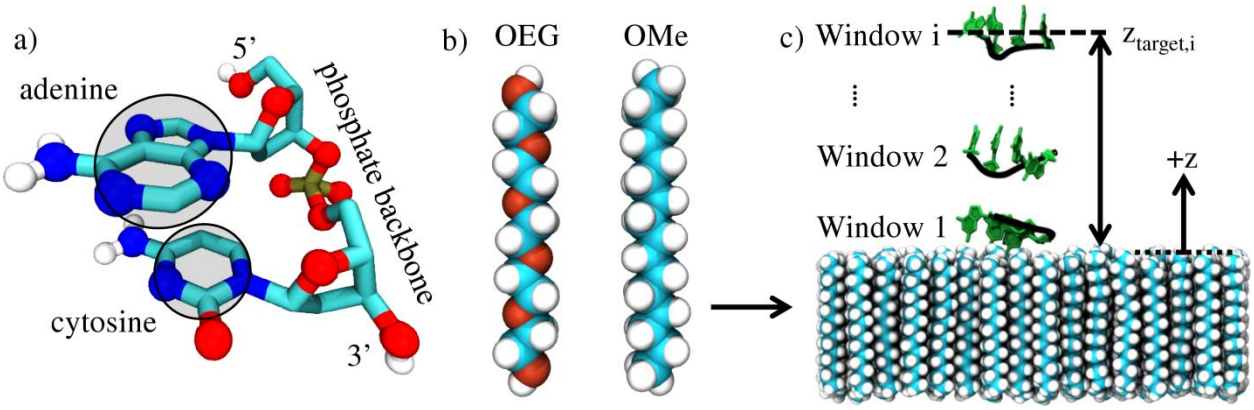


Figure 7.1 a) Diagram of the hydrophobic and hydrophilic portions of adenine and cytosine. The nucleobases (adenine and cytosine), which are the most hydrophobic chemical functionality of DNA, are indicated by transparent gray circles, the hydrogen bonding donors and acceptors are shown as red (oxygen) and blue (nitrogen) spheres, and the hydrogen atoms involved in hydrogen bonding are shown as white spheres. Other hydrogen atoms are hidden for clarity. b) Functionalized surfaces are constructed of 256 strands of oligoethyleneglycol (OEG, chemical formula $H(OCH_2CH_2)_5OH$) or oligomethylene (OMe, chemical formula $CH_3(CH_2)_{14}CH_3$) in a 6.5x7 nm^2 area to yield model hydrophilic and hydrophobic self-assembled monolayers (SAMs), respectively. The bottommost heavy atom of each oligomer is constrained to the xy-plane to mimic attachment to a solid surface. The topmost heavy atom of each oligomer is similarly constrained to maintain similar surface roughness and stiffness. c) DNA-surface separation distance is defined as the z-component of the distance between the center of mass of the ssDNA oligomer and the center of mass of the topmost surface heavy atoms, i.e. the top of the SAM. The free energy method we use, known as umbrella sampling, involves applying a series of energetic constraints at increasing values (z_{target}) of DNA-surface separation distance and yields the free energy as a function of DNA-surface separation distance. We also calculate a variety of other quantities as a function of DNA-surface distance. We study adenine and cytosine dimers (A_2, C_2) and tetramers (A_4, C_4) with explicit solvent and counterions (hidden for clarity) on the two surfaces (OEG, OMe) using this methodology.

hydrophobic – cannot be decoupled and result in complex interactions with surfaces that must be harnessed to direct the behavior of DNA. Studying the subtle interplay between hydrophobic and hydrophilic interactions at the nanoscale will afford a better fundamental understanding of the behavior of DNA and allow greater control of DNA behavior near surfaces.

To elucidate the molecular-level interactions between DNA and surfaces of varying chemistry, we adopt an atomistic molecular dynamics (MD) simulation approach, which allows us to discern the motions and interactions of individual molecules at time and length scales that are often inaccessible with experimental techniques (i.e., sub-nanosecond and sub-nanometer). Previous simulation studies have used similar approaches to examine DNA near surfaces but have largely focused on interactions with carbon nanotubes [19] or solid surfaces [20], or on hybridization with surface-tethered ssDNA in the context of DNA microarrays [21, 22]. Only a few previous studies have examined DNA interactions with functionalized solid surfaces or self-assembled monolayers (SAMs) [23, 24]. In this work, we focus on the fundamental hydrophobic and hydrophilic interactions underlying ssDNA adsorption, which, to the best of our knowledge, has received comparatively little attention with computational methods. We approach this problem by using model hydrogen-bonding hydrophilic (hydroxyl-terminated oligoethyleneglycol, OEG) and non-hydrogen-bonding hydrophobic (oligomethylene, OMe) SAMs (Figure 7.1b). Our approach attempts to isolate the hydrophobic and hydrophilic interactions from other specific properties of the surface, such as oligomer chemistry, the chemical identity of the solid substrate, and surface coverage. We conduct free energy calculations (Figure 7.1c) to determine the effects of nucleobase identity (i.e., adenine vs. cytosine, see Figure 7.1a) and surface hydrophobicity on the strength of ssDNA adsorption and simultaneously observe the non-specific interactions that govern adsorption, such as DNA-surface hydrogen bonds and the behavior of water molecules.

Our free energy calculations show that ssDNA, which is amphiphilic in nature, readily adsorbs to both hydrophobic (OMe SAMs) and hydrophilic (OEG SAMs) surfaces with similar strength. On OEG SAMs, the ssDNA forms attractive DNA-surface hydrogen bonds but also experiences repulsion due to

the high water density near OEG, similar to what has been shown previously for proteins [25, 26]. By contrast, adsorption to OMe SAMs is primarily driven by weak hydrophobic attraction to the surface and water plays a smaller role; consequently, the strength of ssDNA adsorption to hydrophobic surfaces increases with increasing nucleobase hydrophobicity (e.g., adenine is more hydrophobic, and adsorbs more strongly, than cytosine). On both surfaces, the behavior of water molecules that mediate adsorption is affected by the approach of ssDNA, showing a decrease in diffusion of the water molecules and a surprising increase in water ordering, changes which appear to be caused by the replacement of water-water hydrogen bonds with longer-lived DNA-water hydrogen bonds. Unexpectedly, we find that a region of *low* water density near OMe exerts a subtle repulsive force on approaching ssDNA – an occurrence typically associated with the *high* water density near hydrophilic surfaces.

7.2 Simulation and Analysis Methods

7.2.1 Systems studied

We construct model hydrophilic (oligoethyleneglycol, OEG) and hydrophobic (oligomethylene, OMe) self-assembled monolayers (SAMs) from oligomers with chemical formulas $H(OCH_2CH_2)_5OH$ (hydroxyl-terminated OEG) or $CH_3(CH_2)_{14}CH_3$ (methyl-terminated OMe) (Figure 7.1b). We study the adsorption of single-stranded DNA (ssDNA) oligonucleotides (dimers and tetramers) of adenine (A_2, A_4) and cytosine (C_2, C_4). Our philosophy in choosing these systems is as follows. Adenine is a large, more-hydrophobic base with a smaller number of hydrogen bonding partners than other bases, while cytosine is a smaller, less-hydrophobic base with a greater number of hydrogen bonding partners than other bases. Therefore, adenine and cytosine may be said to represent opposite ends of the spectrum of hydrophobicity of the nucleobases. Our SAMs were likewise chosen to provide model hydrophobic and hydrophilic environments. The SAMs have a depth greater than 1 nm, ensuring that ideal hydrophobic and hydrophilic environments exist at the top surface of the SAMs: if the SAMs were thinner, the water and DNA molecules at the top of the SAM would be affected by the water molecules at the bottom of the SAM, altering the apparent hydrophobicity of the environment at the top surface of the SAM. Similarly,

having such a thick SAM makes it unnecessary to consider ssDNA interactions with a specific solid substrate (e.g., gold, silica), simplifying the simulation design and interpretation of results. We chose a high packing density of oligomers (~5 oligomers / nm^2) to minimize the effects of intercalation into the surface, although further work may be aimed at studying these effects because they may play an important role in adsorption [18]. Additionally, our surfaces do not have fixed coordinates as in some studies [26], allowing the oligomers to rearrange to maximize the formation of favorable contacts with adsorbing molecules.

7.2.2 Construction and equilibration of surfaces

The oligomers comprising the surface are initially arranged in a hexagonal grid consisting of 256 oligomers in a cross-sectional area of approximately 6.5 nm by 7 nm, yielding a grafting density of approximately 5 chains / nm^2, which is comparable to a functionalized silica or gold surface [27]. The oligomers are constructed using VegaZZ, and the *antechamber* program of the Amber suite is used to assign general Amber force field (*gaff*) atom types and atomic partial charges using the AM1-BCC charge method [28-30]. The SAMs are constructed in the xy-plane such that they are perpendicular to the z-axis. After constructing the SAMs from the oligomers using VMD, we conducted 1000 steps of conjugate gradient minimization using NAMD to eliminate any unfavorable atomic contacts [31, 32]. We further relaxed the SAMs with the protocol provided in supplementary section 7.6.9.

7.2.3 Construction and equilibration of DNA-surface systems

To prepare systems consisting of DNA, water, counterions, and the SAM, we place the DNA approximately 15 Å from the surface, add neutralizing counterions but no additional salt, and solvate with a sufficiently thick layer of TIP3P water molecules (placed only in the z-direction above and below the SAM) to yield a simulation box height of approximately 80 Å, of which the surface occupies 15-20 Å. This large water box ensures that the DNA can move large distances in the z-direction before interacting with the periodic image of the surface. After constructing these systems, we perform a minimization and equilibration procedure that is described in supplementary section 7.6.10 prior to conducting production

umbrella sampling simulations. During all production simulations, both the topmost and bottommost heavy atoms of the oligomers comprising the surface are constrained to constant z-values with harmonic restraints, using the *collective variables* module of NAMD and a force coefficient of 2.5 kcal/mol-Å [31]. The distance between the topmost and bottommost heavy atoms was 14 Å for OEG and 19 Å for OMe, which we determined to be the preferred brush heights by simulating the surfaces with the bottommost heavy atoms restrained to a constant z-value but the topmost heavy atoms unrestrained. In these simulations, the OEG oligomers assume a helical conformation with a height of approximately 14 Å while the OMe oligomers assume an extended all-*trans* conformation with a height of approximately 19 Å, which are reasonable conformations and brush heights for these oligomers [27, 33]. The reasons for constraining the top and bottom of the oligomers comprising the surface are as follows: the bottom of each oligomer is constrained to mimic attachment to a solid surface, while the top of each oligomer is constrained to maintain a similar stiffness and surface roughness between OEG and OMe and to prevent dissolution of OEG into the bulk water.

7.2.4 Umbrella sampling simulation protocol

Because of the long timescales involved in ssDNA adsorption and desorption, we found it necessary to employ a biased simulation method to enhance sampling of energetically unfavorable states, and we use the umbrella sampling methodology for reasons we discuss below [34]. Since our goal is to study the adsorption process, our reaction coordinate is the DNA-surface separation distance, which we define as the distance in the z-direction between the center of mass of the ssDNA and the center of mass of the topmost heavy atoms of the surface (Figure 7.1c). The umbrella sampling simulations are conducted with DNA-surface separation distances between 5 Å and 25 Å with windows spaced in 1 Å increments and a force constant of 2.5 kcal/mol/Å. In each window, we use two different initial conformations for the ssDNA, and conduct three independent simulations with different initial velocities and each of 20 ns length. The first 5 ns of each independent simulation are excluded from analysis, and therefore each umbrella sampling window, with two initial conformations of the ssDNA and three independent trials, comprises 90 ns of simulation time. One of the initial ssDNA conformations has the bases in a ‘stacked’

state, where the ssDNA has the same conformation as one of the DNA strands in B-form double-stranded DNA (dsDNA) and there is complete base-base stacking. The second initial ssDNA conformation has the bases in an 'unstacked' state, where the ssDNA is in a 'reverse helix' conformation with no base-base stacking. For the lengths of ssDNA studied here, both of the initial conformations yield similar results, and we therefore combine the results of all independent simulations. All energetic constraints are applied with the *collective variables* module of NAMD [31]. The weighted histogram analysis method (WHAM) [35, 36], using a freely available implementation [37], is used to calculate the free energy of adsorption from the results of the umbrella sampling simulations. We estimate the uncertainty of our free energy profiles using Monte Carlo bootstrap error analysis, which is implemented in the freely available WHAM code [37]. The bootstrap procedure consists of the following steps. First, the free energy profile is generated with WHAM by the standard method that uses all data points in the time series of the value of the reaction coordinate in each window; this profile is the 'average' free energy profile. Then, multiple free energy profiles are calculated with WHAM using multiple randomly sampled subsets of the data points in each window. The number of data points sampled in each subset of each window is equal to the total number of data points in that window divided by the correlation time of the time series in that window, which accounts for time correlations in the data and thereby ensures, on average, statistical independence of the sampled points. We estimate the correlation time in each window by fitting the autocorrelation function of the data in each window with a single exponential decay function and using double the decay constant as the correlation time. The correlation times range from 0.2 ns near the surface to 50 ps far from the surface, which results in much lower uncertainty in the free energy at distances far from the surface than at distances close to the surface. Finally, the standard deviation of the free energy in each window of the randomly sampled free energy profiles is used as the estimate for the uncertainty (i.e. the error bars in our plots of free energy).

Our choice of umbrella sampling, although it may not be the most computationally efficient approach to calculating free energies, was motivated by a distinct advantage it possesses compared to non-equilibrium methods, such as metadynamics: in each umbrella sampling window, one can calculate

equilibrium properties of the system, whereas dynamic approaches to calculating free energies (e.g., metadynamics) also dynamically alter the properties of the system [38]. Umbrella sampling allows us to simultaneously determine the free energy landscape of DNA adsorption and various system properties, such as DNA lateral diffusion or water-water hydrogen bonds, as a function of DNA-surface separation distance.

7.2.5 Analysis details

Additional details and references for our methods of characterizing the structure and forces of these systems are available in supplementary section 7.6.12 [31, 39-44]. A discussion on the limitations of our approach is available in supplementary section 7.6.13.

7.3 Results and Discussion

7.3.1 Behavior of water on the two surface chemistries in the absence of ssDNA

We first characterize the properties of water near the two surfaces of interest, and we find that our model hydrophobic and hydrophilic surfaces produce the expected behavior of water, including higher water density near the hydrophilic surface [40, 45], higher density fluctuations near the hydrophobic surface [46], reduced diffusion near the hydrophilic surface [27, 40], and a decrease in the tetrahedral ordering of water near both surfaces [42]. Details of the properties of water near the surfaces are presented in the Figure 7.8 (supplementary section 7.6). Having established that our model surfaces yield the expected properties of hydrophilic and hydrophobic surfaces, we now discuss the behavior of ssDNA on these two surfaces.

7.3.2 Behavior of ssDNA on the two surface chemistries

Figure 7.2a shows that ssDNA exhibits favorable binding free energy to both the hydrophilic (OEG) and the hydrophobic (OMe) surfaces, as evidenced by the decrease in free energy as the DNA approaches both OEG and OMe, which is in accord with results from recent experimental studies of ssDNA on surfaces of varying hydrophobicity [17, 18, 47, 48]. We find that on the hydrophilic OEG surface, there is no energetic barrier to adsorption, and there is a broad attractive well with a minimum near 5 Å. On the

hydrophobic OMe surface, the attractive well is much narrower than on OEG and the minimum is located between 9-10 Å. Interestingly, there is a small energetic barrier located at 12-13 Å above OMe. Although the height of this barrier is low (~0.5 kcal/mol), it appears consistently as a qualitative feature of the free energy landscape of ssDNA adsorption to OMe, as will be further discussed. The two DNA tetramers (A_4 and C_4) adsorb with equal strength on OEG, with a free energy well depth of approximately 2 kcal/mol relative to the free energy of the ssDNA in bulk water, which we define as zero kcal/mol. Interestingly, A_4 adsorbs more strongly than C_4 on OMe, with free energy well depths of 1.8 and 0.5 kcal/mol, respectively. This difference of approximately 1.3 kcal/mol is more than double the thermal energy (k_BT) at room temperature, and is therefore considered significant. We will discuss a possible cause for the difference in adsorption strength between A_4 and C_4 below.

The broadness of the attractive well on OEG and narrowness of the attractive well on OMe can be explained by the chemical and physical features of the two surfaces. Our OEG SAMs, with only a single hydrogen atom on the hydroxyl group that faces the solvent, have a less tightly packed upper surface than OMe SAMs, where each oligomer has three hydrogen atoms on the methyl group that faces the solvent. Consequently, there is more open space between the hydroxyl headgroups of the OEG SAM surface than between the methyl headgroups of the OMe SAM surface. Therefore, the DNA can more easily insert itself into the OEG surface than the OMe surface, despite the fact that we designed the surfaces to have the same grafting density, stiffness, and roughness, and hence the OMe surface becomes more repulsive at a shorter distance than the OEG surface. Although this behavior very close to the surfaces (less than about 7 Å) is the actual behavior of the systems we have simulated, it is not of great interest because behavior in this region is influenced so strongly by the *physical structure* of the surfaces, which is due to our choice of surface chemistry rather than the fundamental hydrophobic and hydrophilic interactions that are our focus. This observation highlights the importance of carefully considering the properties of a surface (e.g., grafting density, stiffness, physical structure, chemistry) when studying adsorption to that particular surface.

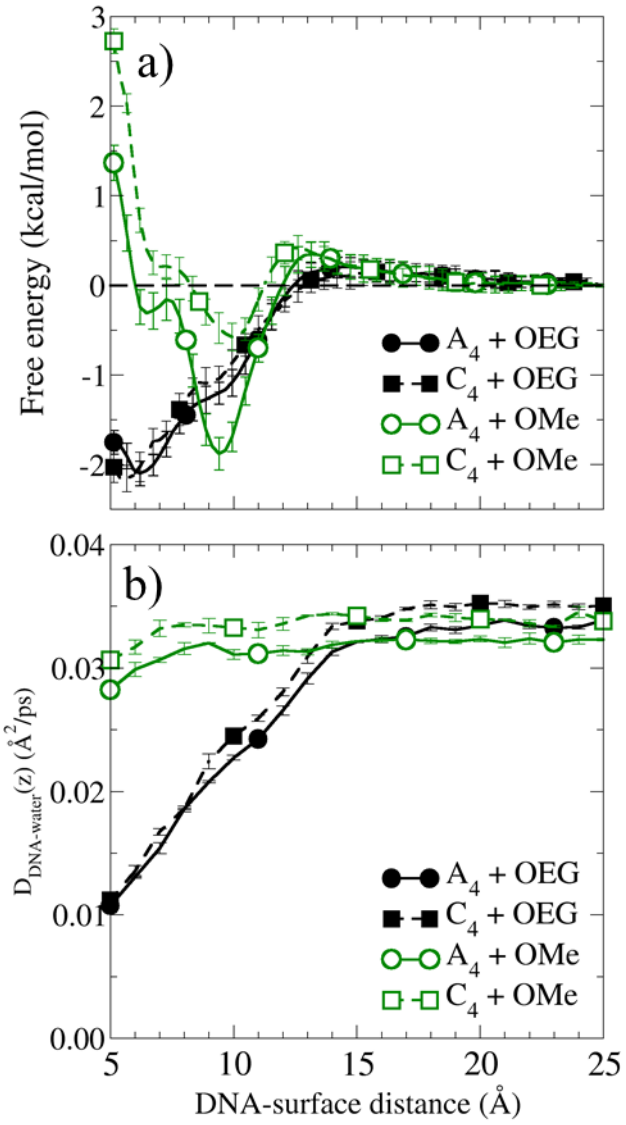


Figure 7.2 a) Free energy and b) lateral (2D) diffusion coefficient of ssDNA tetramers, A_4 and C_4, on OEG and OMe SAMs as a function of DNA-surface separation distance. The free energy is defined to be zero at a DNA-surface separation distance of 25 Å (i.e., in bulk water). Error bars in a) are the standard deviation obtained from the bootstrap analysis described in the Methods section; error bars in b) are the standard error of the mean of three independent trials.

Figure 7.2b shows the lateral (2D in the xy-plane) diffusivity of ssDNA as a function of distance from the surface. As expected, the diffusivity of C_4 in bulk water (DNA-surface distances > 15 Å) is slightly higher than A_4 because cytosine (a pyrimidine) has a lower mass than adenine (a purine). Of greater interest is the behavior near the two surfaces, where we find striking differences: the lateral

diffusion of ssDNA is slowed greatly near hydrophilic OEG but is almost unaffected when approaching hydrophobic OMe, where the diffusion remains almost the same as in bulk solution.

We can compare our results in Figure 7.2 with recent experiments conducted by Monserud *et al.*, where total internal reflectance microscopy was used to study the dynamics of individual ssDNA molecules on hydrophobic and hydrophilic surfaces [18]. Our lateral diffusivity results are in accord with the experimental observation of Monserud *et* al., where it was found that a cytosine pentamer at ~300 K diffuses more quickly on a hydrophobic surface (octadecyltriethoxysilane, OTES) than on a hydrophilic surface (oligoethyleneglycol, OEG) [18]. However, we observe slightly stronger adsorption of C_4 to our hydrophilic surface than to our hydrophobic surface, while Monserud *et al.* observe longer residence time (i.e., stronger adsorption) for C_5 on their hydrophobic OTES surface than their hydrophilic OEG surface [18]. While their functionalized surfaces have similar chemical functionality to ours, their OEG surface is methoxy-terminated (non-hydrogen-bonding) compared to our hydroxy-terminated OEG surface (hydrogen-bonding), possibly increasing the strength of hydrophilic interaction with our OEG SAM. Similarly, our OMe surface is more densely grafted than their OTES surface, possibly reducing the strength of hydrophobic interaction with our hydrophobic surface (e.g., by not allowing intercalation of the relatively hydrophobic bases into the hydrophobic surface). Our surface grafting density is approximately 5.2 oligomers per nm^2, whereas functionalized silica surfaces such as those used by Monserud *et al.* have a maximum grafting density of 4.5 oligomers per nm^2 at full surface coverage [18, 27]. These differences between the simulated and experimental surfaces, and the covalent attachment of bulky fluorescent probes to the experimentally studied ssDNA, may explain the differences in adsorption strength. However, rather than comparing strengths of adsorption, our focus here is to elucidate the underlying molecular-level interactions that mediate the adsorption of ssDNA to hydrophobic and hydrophilic surfaces, which should be broadly applicable to understanding ssDNA adsorption.

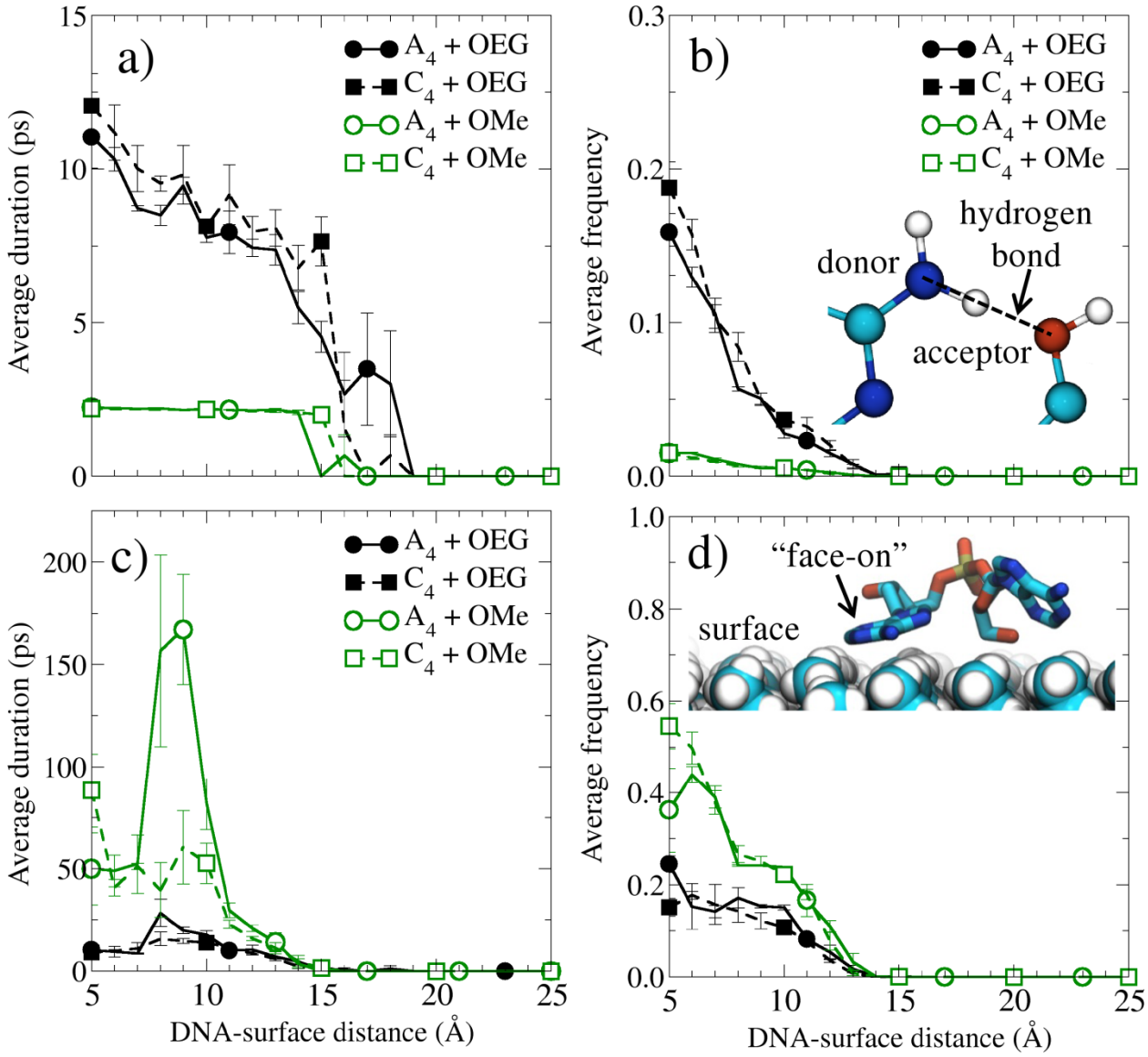


Figure 7.3 Quantification of surface-DNA interactions: hydrogen bonding a) duration and b) frequency, and 'face-on' conformation c) duration and d) frequency. Inset of (b) is a diagram of a hydrogen bond, where the distance between the donor and acceptor atoms is less than 3.5 Å, and the angle formed by the three involved atoms is greater than 120° (an angle of 180° is shown). Inset of (d) shows an diagram of a face-on conformation occurring with an ssDNA dimer, where a DNA base is close to the surface (< 4.5 Å) and approximately parallel to the surface (< 45° between nucleobase and surface normal vectors). Error bars are the standard error of the mean of three independent trials.

We now examine the DNA-surface interactions that could be responsible for the near-surface behaviors of ssDNA that we have described so far. We quantify two major types of non-specific interactions between the DNA and the surfaces: a) hydrogen bonding (H-bonding) interactions and b) base-surface interactions facilitated through "face-on" conformations near the surface. A face-on

conformation occurs when a DNA base is nearly parallel to and is in close proximity to the surface (see Figure 7.3d inset). H-bonding states and face-on states are identified using the geometric criteria described in supplementary section 7.6.12. We show the duration and frequency of these two types of interactions as a function of DNA-surface separation distance in Figure 7.3. Hydrogen bonding interactions, which are typically associated with hydrophilic molecules, dominate on the OEG SAM, while face-on conformations associated with hydrophobic interactions dominate on the OMe SAM. Both the duration and frequency of states that facilitate H-bonds are much higher on OEG than on OMe (Figure 7.3a and 3b), which is reasonable because the atoms comprising the OMe oligomers have negligible partial charges and are therefore incapable of hydrogen bonding. Likewise, the duration and frequency of face-on states are much higher on OMe than on OEG (Figure 7.3c and 7.3d). This is because a) hydrogen bonding interactions disfavor the formation of face-on interactions with OEG and b) the hydrophobic surface of OMe innately favors face-on states, which also conveniently displace water away from the hydrophobic surface. The fact that one type of interaction dominates on each surface suggests a possible explanation for the observed differences in the lateral diffusion of DNA, which is fast on hydrophobic OMe and slow on hydrophilic OEG (Figure 7.2b). Hydrogen bonds, being dependent on the angle and distance between three specific atoms (the donor, hydrogen, and acceptor atoms), must break and reform to allow lateral motion, which slows diffusion, as has been shown elsewhere for peptides [49, 50]. Conversely, the interactions between the surface and a base in a face-on state depend on the distance and angle between the plane of the base and the surface, not between specific atoms, and consequently those interactions are easily maintained during lateral motion, allowing rapid diffusion [49]. Therefore, the diffusion of DNA is slower on the surface on which hydrogen bonds dominate (OEG), and faster on the surface where face-on conformations dominate (OMe). Furthermore, the formation of a substantial number of hydrogen bonds with OEG (Figure 7.3b) corresponds to the distance at which the diffusion of ssDNA begins to decrease (both occur at approximately 12 Å), which is beyond the region where OEG slows the diffusion of water (up to 10 Å, see Figure 7.8 in supplementary section 7.6). Therefore, it is

likely that DNA-surface interactions, rather than solely DNA-water interactions, are responsible for the reduction in ssDNA diffusion speed near OEG.

We also observe that at the distance of onset of surface-DNA interactions (~12 Å), the ssDNA tetramers assume an upright conformation where the ssDNA is more perpendicular to the surface than it is in bulk solution (see Figure 7.9 in supplementary section 7.6). This conformation occurs on both surfaces and for both adenine and cytosine tetramers. This upright conformation is possible because an ssDNA tetramer is innately anisotropic, having a slightly rod-like shape due to charge-charge repulsion and base-base stacking. In bulk solution, the ssDNA has no preferred orientation because the local environment is isotropic. However, when the ssDNA reaches a close enough distance to interact with the surface (~12 Å), the ssDNA orients itself in such a manner so as to maintain the required center-of-mass constraint from the umbrella sampling procedure while also forming favorable contacts with the surface, leading to the upright conformation. When the DNA more closely approaches the surface (< 9 Å), this upright conformation is no longer necessary to form surface-DNA contacts, and the DNA shifts into a conformation parallel to the surface to maximize the number of contacts and to minimize steric repulsion from the surface (Figure 7.9 in supplementary section 7.6). At these small DNA-surface separation distances, the ssDNA preferentially adsorbs to OMe in a "bases-down" conformation where the bases are closer to the surface than the phosphate backbone, but no such preference is observed on OEG (Figures 7.10 and 7.11 in supplementary section 7.6). The implications of these conformations can be appreciated more clearly in the context of applications involving DNA hybridization near surfaces. Adsorption in a "bases-down" conformation may inhibit hybridization on OMe by making the bases of adsorbed ssDNA unavailable for hybridization with other DNA single strands in bulk solution. Conversely, by inducing a conformation where the DNA bases are stacked and unobstructed by the surface, the upright conformation may favor hybridization with complementary DNA single strands in solution.

Now, we relate the relative strength of adsorption of adenine and cytosine tetramers to their non-specific interactions with the surfaces. First, there is no difference in the hydrogen bonding behavior of A_4 and C_4 with OEG (Figure 7.3a and 7.3b), despite the fact that cytosine forms three Watson-Crick

hydrogen bonds with guanine in double-stranded DNA (dsDNA) but adenine forms only two Watson-Crick hydrogen bonds with thymine in dsDNA. This is because when all possible hydrogen bond donors and acceptors (e.g., phosphate groups) are included, the number of hydrogen bonding partners is very similar between adenine and cytosine. This lack of difference in hydrogen bonding frequency or duration suggests an explanation for the equal adsorption strength of A_4 and C_4 on OEG. In contrast to the hydrogen bonding behavior of the two ssDNA tetramers, there is a large difference in the duration and frequency of hydrophobic interactions mediated by the face-on states of A_4 and C_4 on hydrophobic OMe (Figure 7.3c and 7.3d). Near the location of the free energy minimum (~9 Å), A_4 forms longer-lasting face-on interactions than C_4 (150 ps vs. 50 ps, respectively). This observation provides a plausible explanation for the stronger adsorption of A_4 on OMe, and also leads us to consider the relative strength of the surface-DNA interactions and the other interactions, such as water-DNA interactions, that are involved in adsorption.

7.3.3 Forces between ssDNA, water, and surfaces

To quantify the strength of interactions between DNA and other components of the system, we calculate the non-bonded forces between DNA and the following: water molecules, the oligomers comprising the SAM, and the Na^+ counterions (Figure 7.4). We only consider the z-component of the force because the x- and y-components average to zero as a result of the planar symmetry of the SAMs. Therefore, in Figure 7.4, a positive force indicates repulsion away from the surface while a negative force indicates attraction toward the surface.

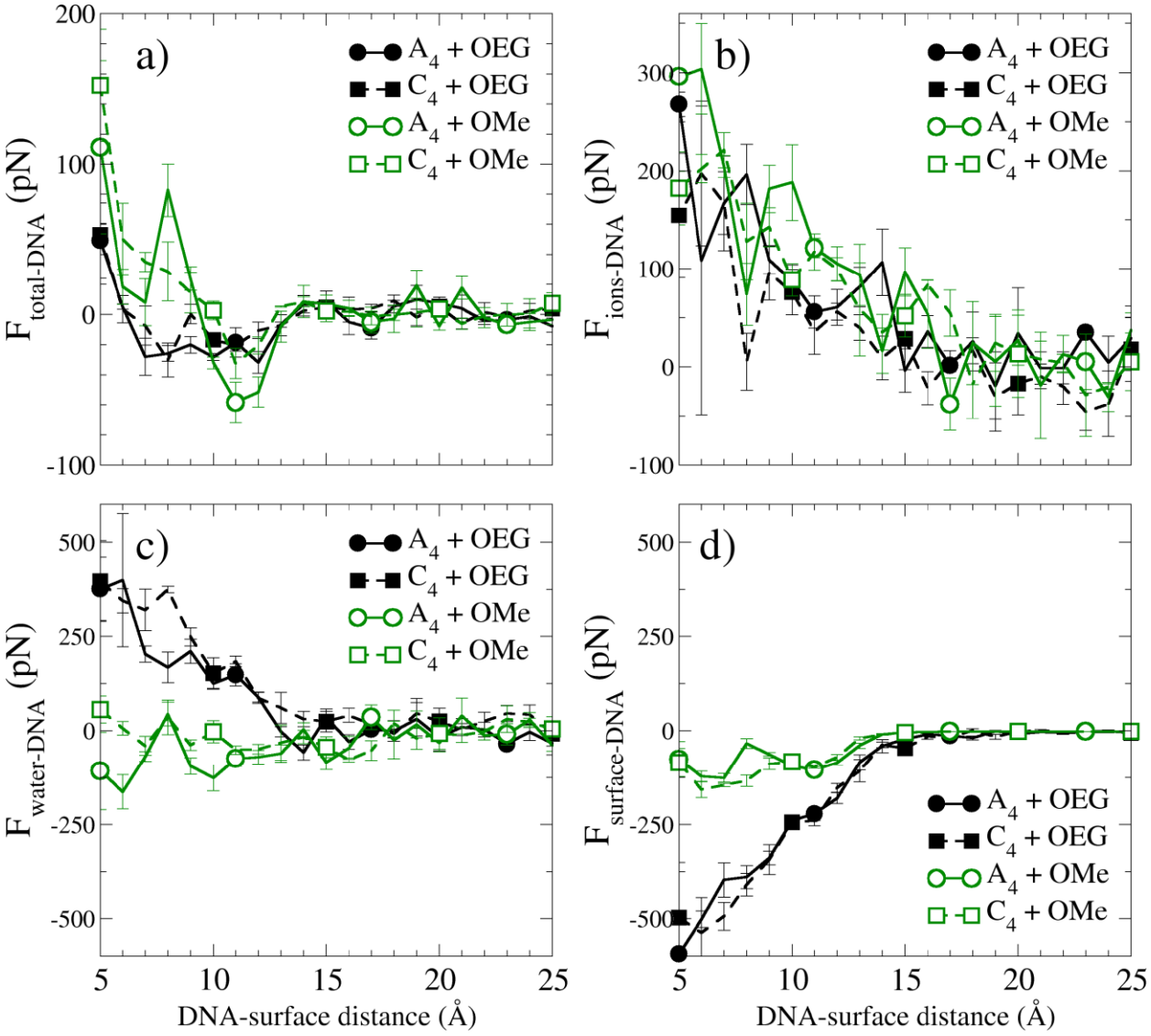


Figure 7.4 Non-bonded forces (van der Waals and electrostatics) of a) all components (the total force), b) the counterions, c) the water, and d) the surface on the ssDNA. Only the z-component of the force is shown: therefore, positive values indicate a repulsive force and negative values indicate an attractive force relative to the surface. Error bars are the standard error of the mean of three independent trials.

In Figure 7.4a we present the total force on the DNA, which is the sum of the three individual contributions (water, surface, and counterions), as a function of DNA-surface distance. On both OEG and OMe, the total force (Figure 7.4a) is zero at DNA-surface separation distances greater than approximately 15 Å, where the DNA is incapable of interacting with the surface. At separation distances less than 15 Å, the total force becomes negative (attractive toward the surface) in the same region as the ssDNA experiences a favorable free energy of adsorption (Figure 7.2), and the total force becomes zero at

approximately the location of the free energy minima, as expected. These findings are not surprising because the negative gradient of the free energy of adsorption should equal the force experienced by the DNA, and these observations are consistent with that fact. While the total force on the DNA strand shows the expected behavior on both surfaces, the individual contributions to the total force show interesting behavior, as we discuss next.

First, we examine the forces originating from the counterions. As stated in the methods section, we only add sufficient Na^+ counterions to the simulation box to neutralize the negative charge of the ssDNA. As is clear in Figure 7.4b, the force exerted by the Na^+ ions on the ssDNA becomes increasingly repulsive (positive) on both OEG and OMe as the DNA approaches the surface. The cause of this repulsive force is that Na^+ ions are incapable of entering these tightly packed self-assembled monolayers. When the DNA is in bulk solution (> 15 Å), the counterions are distributed equally above and below the ssDNA, and the counterbalancing positive and negative forces of the ions above and below the DNA, respectively, yield a net zero force. However, as the ssDNA approaches the surface more closely, there are fewer counterions below the ssDNA because they are excluded from the SAM. Without the negative force of ions below the ssDNA to counterbalance the positive force of the ions above the ssDNA, the net force becomes positive (repulsive). Surprisingly, the magnitude of the counterion force is of the same order of magnitude as the total force, suggesting an important role for counterions in adsorption.

Next we consider the force of water and the force of the surface on the DNA. On OEG, the water exerts a strong repulsive force on the ssDNA (Figure 7.4c) due to the excess density of water near the hydrophilic surface as well as favorable water-surface interactions, a phenomenon that has been documented in previous studies [25, 26]. It has been shown that the high-density water near a hydrophilic surface can shield a surface from protein adsorption, creating desirable anti-fouling properties [13]. In this case, the favorable DNA-OEG interactions, largely consisting of hydrogen bonds (Figure 7.3), counteract the repulsive force of water and lead to a net attraction. If, like many proteins, this short ssDNA could assume a robust, globular conformation that prevented the DNA from conforming to the local topography of the surface, or if the DNA or surface possessed a weaker ability to hydrogen bond, the ssDNA might

interact less favorably with the surface and experience a net repulsion, as has been observed studies of proteins [25, 26]. However, the large number of hydrogen bonding partners on both the ssDNA and our OEG surface leads to the strong net attraction to the surface (Figure 7.4d).

Near the hydrophobic OMe surface, the forces of water and the surface on ssDNA are markedly different from those found for the hydrophilic OEG surface (Figure 7.4c and 7.4d). As with OEG, attractive surface forces are present with OMe, but they are much weaker than those found on the hydrophilic OEG surface. This likely indicates that hydrophobic interactions are energetically weaker than hydrogen bonding interactions, at least for this particular combination of ssDNA and surface: another surface chemistry (e.g., styrene-terminated oligomers) might have more favorable or more numerous hydrophobic interactions with ssDNA. The greater relative strength of hydrogen bonds compared to hydrophobic interactions for this system is corroborated by comparing the DNA-surface interaction energy divided by the area of contact between the DNA and the surface, which shows that OEG exerts approximately double the attractive energy per contact area compared to OMe (Figure 7.12 in supplementary section 7.6). Rather than counteracting the attractive surface forces as on OEG, the force of water on OMe is much weaker and may be negligible, repulsive, or even slightly attractive, depending on the nature of the DNA base: adenine experiences generally attractive water-DNA forces, whereas cytosine experiences generally negligible forces, which is sensible since adenine (a purine) should experience stronger hydrophobic attraction than cytosine (a pyrimidine).

7.3.4 Interactions between water and ssDNA near surfaces

Since water clearly plays an important role in mediating adsorption to both hydrophilic and hydrophobic surfaces, we now shift our focus to the interactions between the ssDNA and water molecules. First, we quantify the behavior of the water molecules that mediate adsorption, i.e., water molecules sandwiched between the DNA and the surface during adsorption. To accomplish this, we define "mediating water molecules" as those water molecules within a cylindrical region of space between the DNA and the surface (Figure 7.5a). The radius of the cylinder is given by the two-dimensional radius of gyration in the xy-plane ($R_{g,2D}$) of the DNA, and the cylinder extends vertically (perpendicular to the surface in the

positive z-direction) from the center of mass of the topmost heavy atoms of the surface upward to the center of mass of the DNA. Furthermore, we limit our analysis to only those water molecules that are close to the surface (< 5 Å in the positive z-direction from the center of mass of the topmost heavy atoms

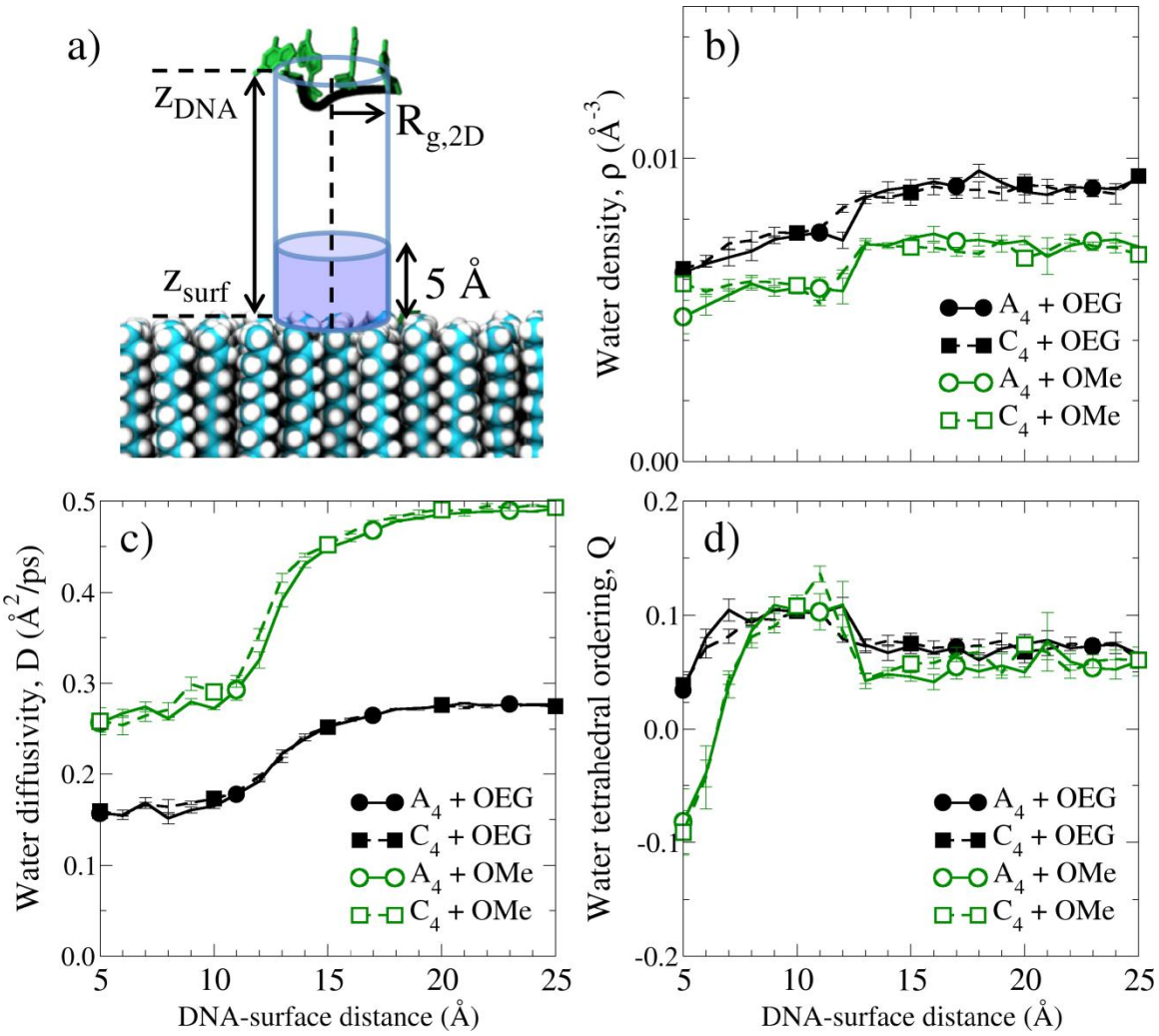


Figure 7.5 a) Diagram of spatial region used to define water molecules that mediate adsorption of DNA, i.e. water molecules within 5 Å of the surface and limited to the region between the DNA and surface. b) Number density, ρ, c) diffusivity, D, and d) tetrahedral ordering, Q, of the water molecules in the shaded region of the cylinder as a function of DNA-surface separation distance. Error bars are the standard error of the mean of three independent trials.

of the surface). We characterize the behavior of the water molecules mediating DNA adsorption in a variety of ways, and focus our discussion on their density, diffusivity, and tetrahedral ordering. In Figure 7.5, we show these measures as a function of DNA-surface separation distance to observe how the

behavior of the mediating water changes as the ssDNA approaches the surface. Note that the behavior of these water molecules is unaffected by the ssDNA until approximately 15 Å, the minimum distance at which an ssDNA tetramer is able to interact with the surface and the water molecules near the surface.

Figure 7.5b shows the density of the adsorption-mediating water molecules as a function of DNA-surface separation distance. Independent of the position of the ssDNA, the density of water near OEG is higher than the density of water near OMe. The high water density near OEG is partly responsible for the strong repulsive force of water on the ssDNA (Figure 7.4c), as described in the preceding section. As the ssDNA approaches the surface more closely, the density of water on both surfaces decreases as the ssDNA displaces water from the surface, unsurprisingly. Figure 7.5c shows the diffusivity of the adsorption-mediating water molecules. As expected, water diffusion near the hydrophobic OMe surface is substantially faster than near the hydrophilic OEG surface, but the diffusivity near both surfaces shows a non-intuitive decrease as the ssDNA approaches. Similarly, the tetrahedral ordering of the water near both surfaces unexpectedly *increases* as the ssDNA approaches the surface (Figure 7.5d). This increased ordering persists until the DNA reaches very small surface separation distances (< 7 Å), at which point the ordering decreases. The reason for this decrease is the nature of the tetrahedral ordering calculation and the small number of water molecules that remain near the surface. When there are not many water molecules, the nearest neighbors of a particular water molecule may be separated by large distances and be arranged in particularly non-tetrahedral geometries due to interference by the DNA and the surface. Therefore, while the reduction in tetrahedral ordering when the DNA is very close to the surface can be explained, the increase in water tetrahedral ordering as the DNA first begins to interact with the surface at ~15 Å is unexpected.

To better understand how the ssDNA influences the behaviors of the near-surface water during adsorption, such as the decrease in diffusion speed and the surprising *increase* in water ordering, we now quantify the number of hydrogen bonds between the mediating water molecules (i.e., those water molecules between the DNA and surface and within 5 Å of the surface, as described in the preceding section) and the following: water molecules, the SAM, and the DNA (Figure 7.6). To eliminate the effect

of water density and thereby allow fair comparison of the two surfaces, we normalize the number of hydrogen bonds to the number of water molecules. First, we comment on the number of hydrogen bonds when the DNA is far from the surface (> 15 Å). There are more water-water hydrogen bonds near OMe than near OEG (Figure 7.6a), but the water molecules near OEG are compensated by water-surface H-bonds (Figure 7.6b). In fact, the OEG-bound water molecules are *overcompensated* by water-surface H-bonds and have a greater total number of hydrogen bonds than those near OMe (3.3 for OEG vs. 3.1 for OMe). Comparing the total number of water hydrogen bonds per water near the surfaces to the number of water-water hydrogen bonds in bulk water (approximately 3.7 with our geometric criteria; see Figure 7.13 in supplementary section 7.6) reveals that the water near OEG is more bulk-like than the water near OMe. The greater total number of water hydrogen bonds on OEG is related to the higher density of water near OEG and is also partly responsible (along with the higher water density) for the resistance of the mediating water to the adsorption of the DNA, as has been shown for proteins [25, 26].

Next, we discuss the changes in water hydrogen bonds as the DNA approaches the surface. As the DNA reaches a close enough distance to interact with the surface and the mediating water molecules (~15 Å), the hydrogen bonding network of water is altered as DNA-water hydrogen bonds form. The number of water-water hydrogen bonds decreases (Figure 7.6a) while the number of water-surface hydrogen bonds increases slightly (Figure 7.6b), although the latter remains negligible on OMe, which is essentially incapable of forming hydrogen bonds. In total, water molecules near either OEG or OMe lose approximately 0.2 water-water hydrogen bonds but gain approximately 0.4 water-DNA hydrogen bonds (Figure 7.6c): surprisingly, the overall effect of DNA adsorption is to *increase* the total number of hydrogen bonds per water. This increase may provide a small favorable enthalpic motivation for adsorption on both surfaces. The increase in the total number of hydrogen bonds involving the mediating water molecules, as well as the replacement of water-water hydrogen bonds with water-DNA hydrogen bonds, may also be responsible for the changes in water behavior – reduced diffusion and increased ordering – shown in Figure 7.5. Water-water hydrogen bonds break and reform quickly whereas water-DNA hydrogen bonds are much longer-lived (Figure 7.14 in supplementary section 7.6). The reduction in

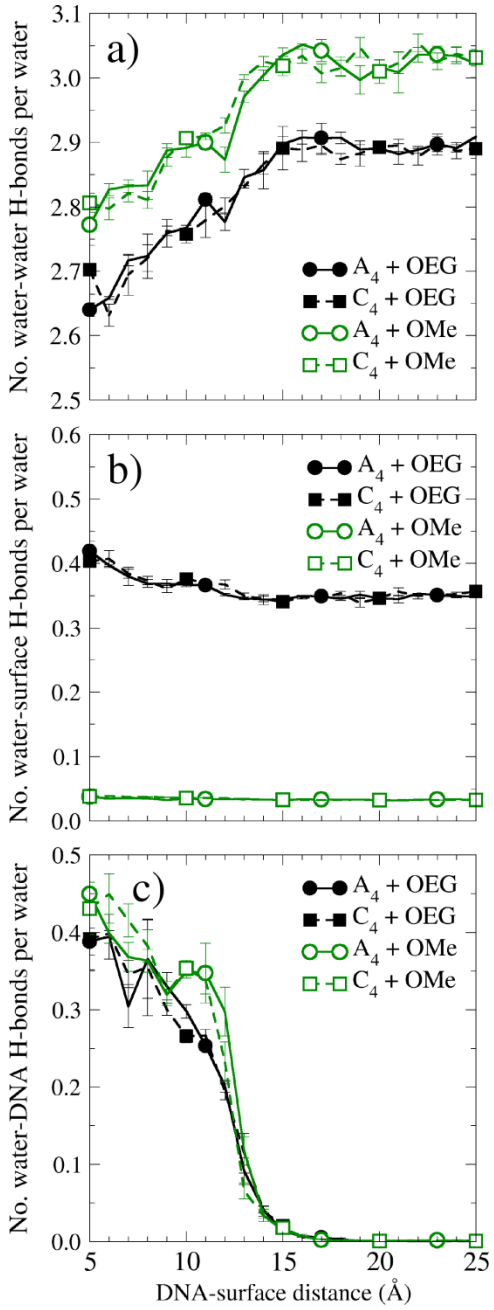


Figure 7.6 Number of hydrogen bonds between water molecules (normalized to the number of water molecules) with a) other water molecules, b) the surface, and c) the ssDNA. Only water molecules within 5 Å of the surface and in the spatial region between the DNA and the surface are considered (see Figure 7.5 for diagram). In part a), all water-water hydrogen bonds to/from the water molecules in the limited spatial region are included, not just water-water hydrogen bonds between the water molecules within the limited spatial region. Error bars are the standard error of the mean of three independent trials.

diffusion of water near both surfaces as the ssDNA approaches the surface (Figure 7.5c) may be explained by the formation of these longer-lived water-DNA hydrogen bonds. The increased tetrahedral ordering of water near both surfaces (Figure 7.5d) may be explained in a similar manner. Because the ssDNA both diffuses and changes orientation more slowly than water molecules, hydrogen bonds with DNA will slow the reorientation of DNA-bound water molecules relative to water molecules that are not hydrogen bonded to the DNA. Consequently, we hypothesize that the faster-moving water molecules may rearrange around the slower-moving, DNA-bound water molecules, enhancing the preferred tetrahedral arrangement of water.

7.3.5 Comparison of ssDNA tetramers to ssDNA dimers

Having thoroughly examined the interactions of ssDNA tetramers with the surfaces and the water molecules that mediate adsorption, we now verify that our qualitative observations are consistent for all short ssDNA oligomers (i.e., oligomers that are so short that they cannot adopt compact globular or hairpin-like conformations). To this end, we present a comparison of ssDNA tetramers with ssDNA dimers (Figure 7.7). We find that essentially all of the interactions and changes in behavior that occur for tetramers (e.g., qualitative shape of free energy landscape, formation of DNA-surface interactions, reduction in water density and diffusion) also occur for dimers, but at a shorter DNA-surface distance, as expected (see Figure 7.7a and 7.7b for adenine, and Figure 7.15 in supplementary section 7.6 for cytosine). The difference in the distance at which ssDNA-surface interactions begin (~12 Å for dimers and ~15 Å for tetramers, leading to a ~3 Å difference) corresponds to the difference in the size of the ssDNA (i.e. the width of the density distributions). Examining the density distribution of DNA held at a surface separation distance of 15 Å (Figure 7.7c for A_2 and Figure 7.7e for A_4), the distribution of the dimers extends approximately 6 Å in the negative z-direction from the peak density at 15 Å, while the density distribution of the tetramers extends approximately 9 Å in the negative z-direction. The difference between 6 Å and 9 Å yields the observed difference of ~3 Å. Aside from these expected differences due to molecular size, our qualitative observations about short ssDNA oligomer adsorption on the two surface chemistries are mostly independent of the oligomer length. We find two other differences in behavior

between the dimer and tetramer that are easily explained. First, the "upright" conformation observed for tetramers does not occur because ssDNA dimers are essentially isotropic (Figure 7.16 in supplementary

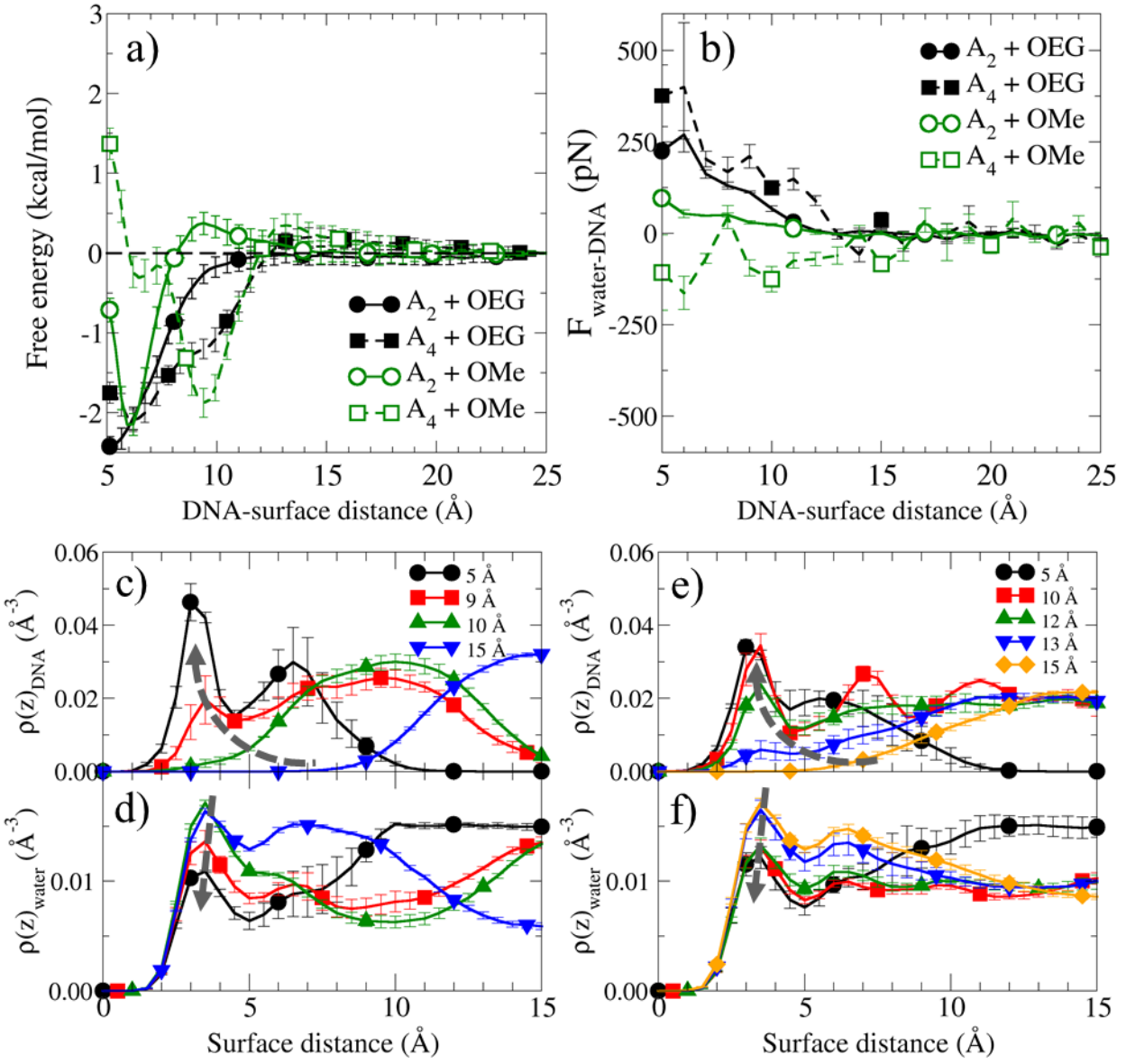


Figure 7.7 Comparison of (a) free energy of adsorption and (b) force of water on ssDNA for the adenine dimer (circles) or tetramer (squares) adsorbing to OEG or OMe. The density profiles of water (c, e) and DNA (d, f) when the adenine dimer (c, d) or tetramer (e, f) is held at various distances from the OMe SAM. The dashed gray arrows in (c, e) indicate the increasing density of DNA at the surface as the DNA approaches the surface, and the dashed arrows in (d, f) indicate the decreasing water density as the DNA approaches. Error bars in a) are the standard deviation obtained from the bootstrap analysis described in the Methods section; error bars in the other panels are the standard error of the mean of three independent trials.

section 7.6). Second, the diffusivity of the dimers is approximately twice that of the tetramers due to the lower overall mass of the dimers (Figure 7.17 in supplementary section 7.6).

However, there is a difference between dimers and tetramers that is *not* easily explained. The force of water on the ssDNA tetramers adsorbing to OMe depends on the nucleobase identity and may be zero, repulsive, or attractive (Figure 7.4c), whereas for ssDNA dimers the force of water is consistently repulsive regardless of the DNA base identity (Figure 7.7b for adenine and Figure 7.15 in supplementary section 7.6 for cytosine). This finding is unexpected because the water near OMe neither interacts favorably with the surface nor forms a dense, repulsive layer, whereas water near OEG repels ssDNA due to its high density and numerous hydrogen bonds with the surface. Furthermore, there is also a *similarity* between dimers and tetramers that warrants closer inspection: the presence of the small free energy barrier to adsorption on OMe. Although this feature of the free energy landscape is quantitatively small (~0.5 kcal/mol in all cases), the fact that it is qualitatively present in all the free energy landscapes involving OMe, regardless of DNA length or base identity, suggests an important underlying phenomenon.

To explain the above observations about ssDNA adsorption to OMe – purely repulsive water forces on dimers and the free energy barrier to adsorption – in Figure 7.7c-f we present the ssDNA density and water density distribution functions when the ssDNA is held at various surface separation distances above the OMe SAM. These density distributions are calculated within a cylindrical spatial region similar to that shown in Figure 7.5a, except that we calculate the density of atoms in thin slices of the cylinder with increasing distance from the surface rather than the density within the region 5 Å above the surface shown in Figure 7.5a. First, we note that the water density distribution near OMe follows a pattern of low density very close to the surface (<3 Å), followed by a peak of increased density near 3 Å, and then a second region of reduced density near 5 Å before reaching the bulk density at approximately 7 Å (see Figure 7.7d with the ssDNA held at 15 Å, or Figure 7.8 in supplementary section 7.6). This hydration pattern is a general feature of hydrophobic surfaces, although the magnitude of the peaks and troughs depends on the surface chemistry [46]. Next, we consider the two extremes of DNA-surface

distance shown in Figure 7.7c-f: when the ssDNA (dimer or tetramer) is far from the surface (15 Å), and when it is close to the surface (5 Å). When the ssDNA (dimer or tetramer) is far from the surface (15 Å) the water density is decreased in the bulk solution because of the presence of the DNA, as expected. Additionally, at 15 Å the ssDNA is not close enough to alter the density of the water near the surface, and the water distribution close to the surface is identical to the distribution in the absence of DNA. When the ssDNA (dimer or tetramer) is very close to the surface (5 Å), there are two peaks in DNA density: one peak is near the surface at approximately 3 Å, and the other peak is farther from the surface at approximately 6 Å. There are two critical observations about the density distributions when the ssDNA is located at 5 Å. First, the peak in DNA density at 3 Å corresponds to the peak in water density close to OMe. Second, there is a region of lower DNA density between the two DNA density peaks, and this region corresponds to the second region of low water density that exists at approximately 5 Å above OMe. The ssDNA resides more frequently in the regions of higher water density than in the region of lower water density.

We now connect these observations about water and ssDNA density to our previous observations of the repulsive force of water on ssDNA dimers and the presence of a free energy barrier on OMe by examining the DNA density distributions when the ssDNA is held at surface separation distances between 5 Å and 15 Å from OMe. More specifically, the DNA-surface distances of 10 Å and 9 Å are the approximate locations where the ssDNA *dimers* are just reaching the peak of the free energy barrier and just surpassing the peak of the free energy barrier on OMe (Figure 7.7a), respectively. The DNA-surface distances of 13 Å and 12 Å bear the same significance for ssDNA *tetramers*. When the ssDNA is held at the farther of these distances (10 Å for dimers, 13 Å for tetramers), the leading edge of DNA density has begun to contact the region of low water density, but the ssDNA has not quite penetrated the region of low water density near OMe. At the closer of the two aforementioned distances (9 Å for dimers, 12 Å for tetramers), the DNA penetrates the low water density region and enters the high water density region at the OMe surface, drastically reducing the water density at the surface. This increase in DNA density near the surface, and the concomitant decrease in water density at the surface, with decreasing surface-DNA

distance are indicated schematically with dashed arrows in Figure 7.7c-f. For the ssDNA to pass over the free energy barrier (10 to 9 Å for dimers, 13 to 12 Å for tetramers), it must cross through the unfavorable region of low water density. The low-density region of water appears to be the driving force for the free energy barrier to ssDNA adsorption to or desorption from OMe.

Additionally, this repulsive, low-density region of water could explain the purely repulsive force of water on ssDNA *dimers*, as opposed to the variable force of water experienced by ssDNA *tetramers*. First, ssDNA dimers are smaller than tetramers, so a greater proportion of the dimer must reside in the unfavorable region of low-density water: tetramers might experience stabilizing forces from the bulk water outside of the dehydrated layer that minimize the importance of repulsion from the dehydrated layer. Second, ssDNA dimers have a lower frequency of base-base stacking interactions than tetramers (Figure 7.18 in supplementary section 7.6), probably because base stacking is stabilized by stacking of adjacent bases. The higher base-base stacking frequency may shield the tetramer bases from the low-density layer of water, possibly by providing a more hydrophobic environment (base/base stacking interaction) instead of a more hydrophilic environment (base/low-density-water interaction). Whatever the reason for this difference in water repulsion between ssDNA dimers and tetramers, it is nonetheless surprising to discover a repulsive role for water on a hydrophobic surface, since hydrophilic surfaces are typically associated with hydration repulsion.

Repulsive forces caused by the hydration pattern on hydrophobic surfaces are not common in the literature, although this may be partly due to the difficulty in decoupling the force of water from the total force with experimental measurements. Some simulation studies report repulsive forces directly caused by water in graphene-based systems (e.g., carbon nanotubes, fullerenes), but these repulsions appear because of the strong dispersion attraction between the densely-arranged carbon atoms and water molecules, which is not typical of hydrophobic surfaces [51, 52]. To the best of our knowledge, this repulsive force due to a dehydrated region of water above a hydrophobic surface has not been explicitly described. It is not clear why this low-density region of water necessarily generates a repulsive force. It is possible that, on average, DNA prefers more hydrated/hydrophilic environments, or it is possible that the

location of the dehydrated region – between two high-density regions of water – is not an ideal location for the relatively hydrophobic bases. The ssDNA may prefer to reside within the region of high water density above OMe because that distance (~3 Å) is approximately the closest distance of approach if one considers the van der Waals radii of the atoms involved. Additionally, adopting a face-on conformation at that distance (~3 Å) would release the greatest number of water molecules from the surface, since the water density is highest there prior to adsorption, while still maintaining sufficient hydration of the hydrophilic portions of the nucleobases. A related explanation may lie in the amphiphilic nature of DNA. Since hydrophobic surfaces experience long-range *repulsive* forces from *hydrophilic* surfaces but experience mid-range *attractive* forces from other *hydrophobic* surfaces [53], an amphiphilic molecule like DNA may experience a combination of these forces as it approaches a hydrophobic surface, resulting in distance-dependent attractions and repulsions.

7.4 Conclusions

We have used atomistic molecular dynamics (MD) simulations combined with free energy calculations to understand adsorption of single-stranded DNA (ssDNA) oligomers of adenine and cytosine to a model hydrogen-bonding hydrophilic surface (hydroxyl-terminated oligoethyleneglycol, OEG) and a model non-hydrogen-bonding hydrophobic surface (methyl-terminated oligomethylene, OMe). Our approach is designed to minimize some complications of real self-assembled monolayers (e.g., surface coverage) and to focus on the molecular-level interactions governing the adsorption of amphiphilic ssDNA to hydrophilic and hydrophobic surfaces, such as DNA-surface hydrogen bonds, hydrophobic interactions, and the behavior of water molecules. We find that ssDNA behaves as an amphiphile, adsorbing with similar strength to both OEG and OMe SAMs. On OEG, there is little difference in the adsorption strength of adenine (a purine) and cytosine (a pyrimidine), likely because of the approximately equal number of hydrogen bonding partners on each nucleotide. On OMe, however, there are substantial differences between cytosine and adenine, with adenine adsorbing more strongly because of its greater

hydrophobicity relative to cytosine. Adsorption to OEG and OMe is mediated by the formation of hydrogen bonds (hydrophilic attraction) and face-on conformations (hydrophobic attraction), respectively.

During ssDNA adsorption, the behavior of the water molecules mediating adsorption (i.e., water molecules sandwiched between the surface and the DNA) is altered: water density and diffusivity are reduced, and water ordering is surprisingly increased. The reduction in density occurs simply because the DNA displaces water molecules from the near-surface environment, but the other changes appear to be related to the formation of longer-lived DNA-water hydrogen bonds and the concomitant reduction in shorter-lived water-water hydrogen bonds. While the behavior of water is influenced by ssDNA adsorption, water simultaneously plays a critical role in mediating the strength of adsorption. On the hydrophilic surfaces the highly favorable hydrogen bonding interactions between ssDNA and OEG is opposed by repulsion by the dense water at the OEG surface that resists ssDNA adsorption, as has been found in simulation studies of protein adsorption [25, 26]. In contrast, the weaker hydrophobic interactions with OMe are not counterbalanced by a strong contribution from water molecules. The balance of forces on OEG (strong hydrogen bonding and strong repulsion by water) and OMe (weak face-on interactions and a smaller role for water) leads to similar adsorption strength on both surfaces. Despite the weaker role of water near the OMe SAM, we unexpectedly find that water near the hydrophobic surface can present a repulsive force to adsorbing ssDNA, and we find that this repulsion appears for the opposite reason that hydrophilic surfaces typically resist adsorption (i.e., strong hydration of the surface). That is, the mechanism underlying this repulsive force on our hydrophobic surface appears to be a partially *dehydrated* region near the OMe SAM. Similar dehydrated regions exist near other hydrophobic surfaces and even on some hydrophilic surfaces [46], suggesting that this force might be present at many liquid-solid interfaces.

This study shows that adsorption of short ssDNA to hydrophobic and hydrophilic surfaces results from a combination of weak hydration repulsion on hydrophobic surfaces, strong hydration repulsion on hydrophilic surfaces, hydrogen bonding interactions with hydrophilic surfaces, and face-on conformations at hydrophobic surfaces. The understanding obtained from this work can be used in a variety of contexts,

such as patterning a surface with hydrophobic and hydrophilic regions to guide DNA self-assembly. For example, since adenine bases adsorb more strongly to hydrophobic surfaces than do cytosine bases, a hydrophobic patch on a surface may favor the attachment of adenine-rich sequences, possibly serving as an anchor point to guide further bottom-up assembly or to direct the placement of existing nanostructures. To expand on this work, we will next explore the behavior of longer ssDNA oligomers (i.e., oligomers long enough to adopt compact globular or hairpin-like structures) to understand the effect of surface hydrophobicity on near-surface DNA conformation and hybridization. For instance, it has been hypothesized that long DNA single strands (≥ 10 bases) can assume globular conformations that sequester the relatively hydrophobic nucleobases at the core of a micelle-like structure, thereby reducing interactions with hydrophobic surfaces and increasing interactions with hydrophilic surfaces [18]. Similarly, the hydrophobicity of a surface can affect DNA hybridization, with hydrophobic surfaces favoring hairpin formation compared to hydrophilic surfaces [17]. Additionally, the non-monotonic dependence of the hydrophobic effect on temperature [54] can non-monotonically affect the adsorption strength and diffusion speed of DNA near surfaces [18]. Using simulations to understand such behaviors at the molecular level will aid the rational design of surfaces that promote not only the desired adsorption behavior but also the desired hybridization behavior.

7.5 References

[1] C. Pinholt, R. A. Hartvig, N. J. Medlicott, and L. Jorgensen, Expert Opinion on Drug Delivery 8 (2011) 949.
[2] C. S. Gudipati, J. A. Finlay, J. A. Callow, M. E. Callow, and K. L. Wooley, Langmuir 21 (2005) 3044.
[3] C. J. Wilson, R. E. Clegg, D. I. Leavesley, and M. J. Pearcy, Tissue Engineering 11 (2005) 1.
[4] P. W. K. Rothemund, Nature 440 (2006) 297.
[5] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. M. Shih, Nature 459 (2009) 414.
[6] S. Y. Park, J. S. Lee, D. Georganopoulou, C. A. Mirkin, and G. C. Schatz, Journal of Physical Chemistry B 110 (2006) 12673.
[7] S. Y. Park, K. R. A. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin, Nature 451 (2008) 553.
[8] H. Noh, A. M. Hung, and J. N. Cha, Small 7 (2011) 3021.
[9] A. Striolo, Small 3 (2007) 628.
[10] A. Seifpour, S. R. Dahl, B. Lin, and A. Jayaraman, Molecular Simulation (2013) 1.
[11] X. Sun, S. Hyeon Ko, C. Zhang, A. E. Ribbe, and C. Mao, Journal of the American Chemical Society 131 (2009) 13248.

[12] P. Gong and R. Levicky, Proceedings of the National Academy of Sciences of the United States of America 105 (2008) 5301.
[13] E. Ostuni, R. G. Chapman, R. E. Holmlin, S. Takayama, and G. M. Whitesides, Langmuir 17 (2001) 5605.
[14] R. J. Marsh, R. A. L. Jones, and M. Sferrazza, Colloids and Surfaces B 23 (2002) 31.
[15] E. Ostuni, R. G. Chapman, M. N. Liang, G. Meluleni, G. Pier, D. E. Ingber, and G. M. Whitesides, Langmuir 17 (2001) 6336.
[16] T. W. Cha, V. Boiadjiev, J. Lozano, H. Yang, and X. Y. Zhu, Analytical Biochemistry 311 (2002) 27.
[17] M. Kastantin and D. K. Schwartz, Small 9 (2012) 933.
[18] J. H. Monserud and D. K. Schwartz, Biomacromolecules 13 (2012) 4002.
[19] X. Zhao and J. K. Johnson, Journal of the American Chemical Society 129 (2007) 10438.
[20] S. Monti and T. R. Walsh, Journal of Physical Chemistry C 115 (2011) 24238.
[21] S. Monti, I. Cacelli, A. Ferretti, G. Prampolini, and V. Barone, Journal of Physical Chemistry B 114 (2010) 8341.
[22] K.-Y. Wong and B. Montgomery Pettitt, Theoretical Chemistry Accounts 106 (2001) 233.
[23] H. Liang, Z. Li, and J. Yang, Physical Chemistry Chemical Physics 12 (2010) 4431.
[24] K.-Y. Wong, A. Vainrub, T. Powdrill, M. Hogan, and B. M. Pettitt, Molecular Simulation 30 (2004) 121.
[25] J. C. Hower, Y. He, and S. Jiang, Journal of Chemical Physics 129 (2008) 215101.
[26] N. Schwierz, D. Horinek, S. Liese, T. Pirzer, B. N. Balzer, T. Hugel, and R. R. Netz, Journal of the American Chemical Society 134 (2012) 19628.
[27] M. J. Stevens and G. S. Grest, Biointerphases 3 (2008) FC13.
[28] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, Journal of Molecular Graphics and Modelling 25 (2006) 247.
[29] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Journal of Computational Chemistry 25 (2004) 1157.
[30] A. Jakalian, D. B. Jack, and C. I. Bayly, Journal of Computational Chemistry 23 (2002) 1623.
[31] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of Computational Chemistry 26 (2005) 1781.
[32] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics and Modelling 14 (1996) 33.
[33] A. Ulman, Chemical Reviews 96 (1996) 1533.
[34] G. M. Torrie and J. P. Valleau, Journal of Computational Physics 23 (1977) 187.
[35] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, Journal of Computational Chemistry 13 (1992) 1011.
[36] B. Roux, Computer Physics Communications 91 (1995) 275.
[37] A. Grossfield, "WHAM: the weighted histogram analysis method," version 2.0.4, membrane.urmc.rochester.edu/content/wham.
[38] C. Micheletti, A. Laio, and M. Parrinello, Physical Review Letters 92 (2004) 170601.
[39] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, Nucleic Acids Research 37 (2009) 5917.
[40] J. Feng, K.-Y. Wong, K. Dyer, and B. M. Pettitt, Journal of Chemical Physics 131 (2009) 125102.
[41] M. F. Hagan, A. R. Dinner, D. Chandler, and A. K. Chakraborty, Proceedings of the National Academy of Sciences 100 (2003) 13922.
[42] D. Bandyopadhyay and N. Choudhury, Journal of Chemical Physics 136 (2012) 224505.
[43] A. Luzar and D. Chandler, Nature 379 (1996) 55.
[44] S. Balasubramanian, S. Pal, and B. Bagchi, Physical Review Letters 89 (2002) 115505.
[45] A. Poynor, L. Hong, I. K. Robinson, S. Granick, Z. Zhang, and P. A. Fenter, Physical Review Letters 97 (2006) 266101.

[46] S. N. Jamadagni, R. Godawat, and S. Garde, in Annual Review of Chemical and Biomolecular Engineering, Vol 2, Vol. 2 (J. M. Prausnitz, ed.), Annual Reviews, Palo Alto, 2011, p. 147.
[47] M. L. Carot, R. M. Torresi, C. D. Garcia, M. J. Esplandiu, and C. E. Giacomelli, Journal of Physical Chemistry C 114 (2010) 4459.
[48] V. Chan, S. E. McKenzie, S. Surrey, P. Fortina, and D. J. Graves, Journal of Colloid and Interface Science 203 (1998) 197.
[49] A. Serr, D. Horinek, and R. R. Netz, Journal of the American Chemical Society 130 (2008) 12408.
[50] A. Hung, M. Mager, M. Hembury, F. Stellacci, M. M. Stevens, and I. Yarovsky, Chemical Science 4 (2013) 928.
[51] L. Li, D. Bedrov, and G. D. Smith, Journal of Physical Chemistry B 110 (2006) 10509.
[52] S. Ou, S. Patel, and B. A. Bauer, Journal of Physical Chemistry B 116 (2012) 8154.
[53] G. B. Kaggwa, P. C. Nalam, J. I. Kilpatrick, N. D. Spencer, and S. P. Jarvis, Langmuir 28 (2012) 6589.
[54] D. Chandler, Nature 437 (2005) 640.
[55] P. Mark and L. Nilsson, Journal of Physical Chemistry A 105 (2001) 9954.
[56] D. Frenkel and B. Smit, Understanding Molecular Simulations: From Algorithms to Applications, Academic Press, 2001.
[57] Z. M. Raspopovic, S. Sakadzic, S. A. Bzenic, and Z. L. Petrovic, IEEE Transactions on Plasma Science 27 (1999) 1241.

7.6 Supplementary information

7.6.1 Water shows expected behaviors near a hydrophilic and hydrophobic surface in the absence of DNA

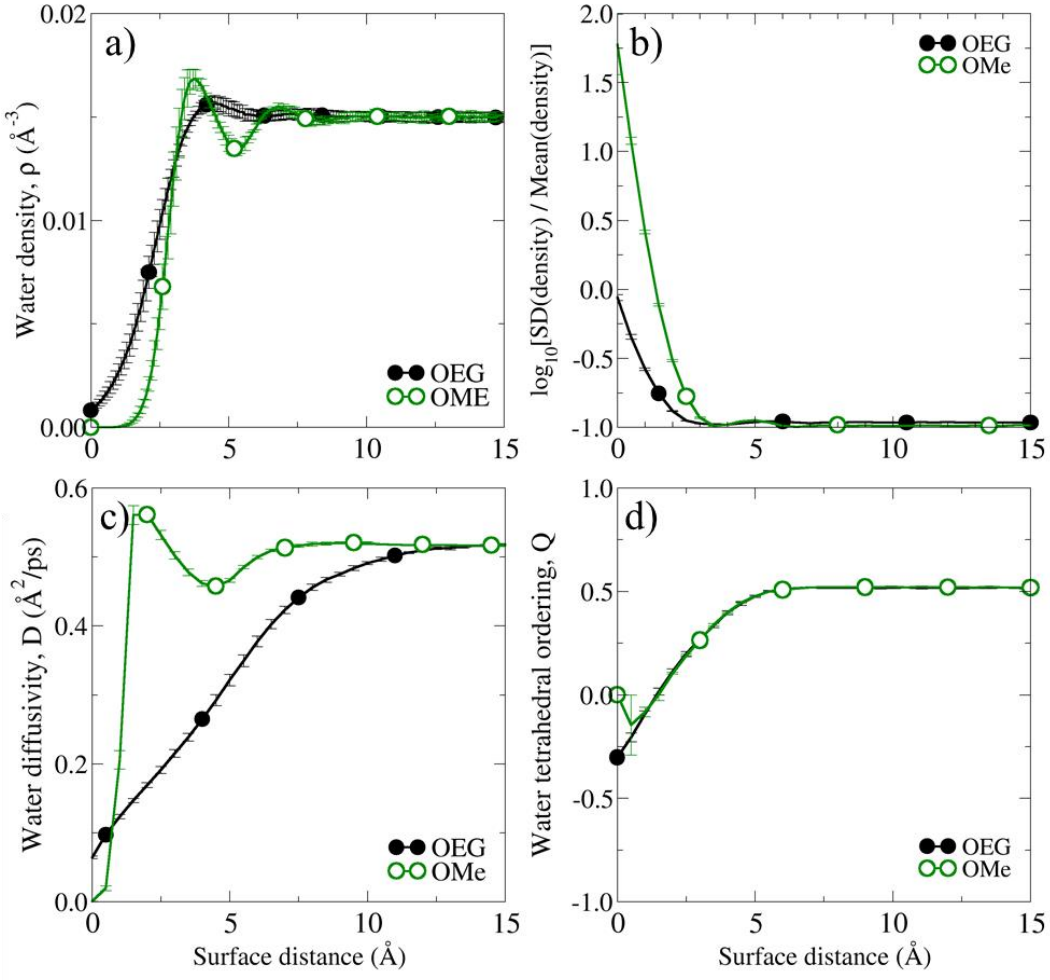


Figure 7.8 Behavior of water near model hydrophilic (OEG) and hydrophobic (OMe) surfaces. Shown are a) number density, ρ, b) density fluctuations, c) diffusion coefficient, D, and d) tetrahedral ordering, Q, of water as a function of distance from the surface. Density fluctuations are defined as the base-10 logarithm of the standard deviation of the density divided by the mean of the density. With this measure, it is clear that the density fluctuations are much higher near the hydrophobic surface.

Compared to the hydrophilic surface (OEG), water density is lower and density fluctuations are higher near the hydrophobic surface (OMe), as expected (Figure 7.8a and 7.8b) [40, 45, 46]. The effects of the surface on density only extend 5 Å above the surface. Diffusion is slowed substantially from the bulk

value near OEG, an effect which extends farther into the bulk water (~10 Å) than the effects on density (Figure 7.8c). Conversely, water diffusion is unaffected by the hydrophobic surface (OMe) except at shorter ranges, and the effect on diffusion appears to be caused by the changes in water density near the hydrophobic surface: at 3-4 Å, water density increases and diffusion decreases; below 2.5 Å, where the water density is reduced, diffusion increases. These observations are in accord with previous studies of water diffusion near surfaces of varying hydrophobicity [27, 40]. The water self-diffusion coefficient in bulk (~15 Å from the surface) is in quantitative agreement with previous studies of the TIP3P water model in bulk [55]. As expected, the tetrahedral ordering of water is reduced by the presence of a surface (Figure 7.8d) [42].

7.6.2 Single-stranded DNA (ssDNA) tetramers assume an upright conformation (perpendicular to the surface) at a specific distance from the surface

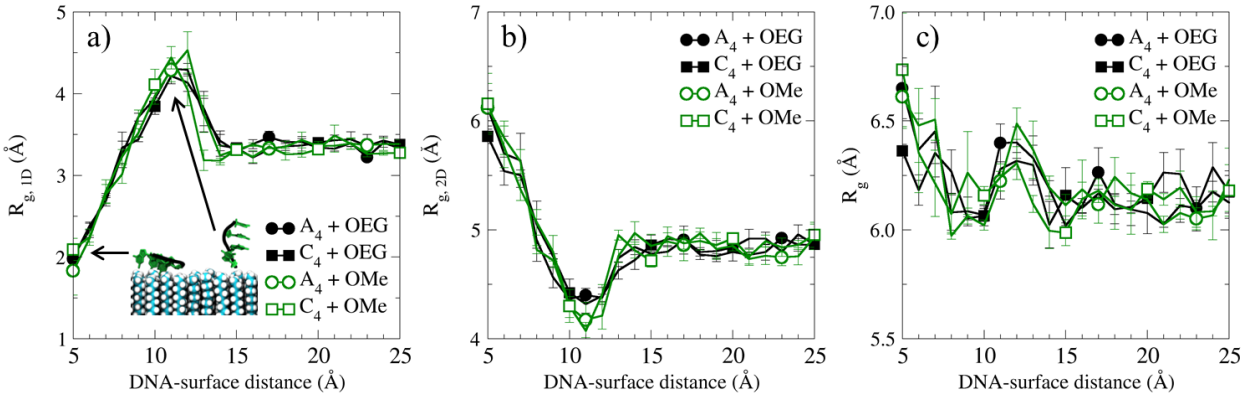


Figure 7.9 The radius of gyration of the ssDNA in a) the z-direction, $R_{g,1D}$, b) the xy-plane, $R_{g,2D}$, and c) in all three dimensions, R_g, as a function of DNA-surface separation distance. Diagram in panel (a) shows parallel and perpendicular (or upright) conformations of ssDNA on the surface at the corresponding values of $R_{g,1D}$ and DNA-surface distance.

The simultaneous increase of $R_{g,1D}$ and decrease of $R_{g,2D}$ between 8-15 Å, relative to the bulk values at distances > 15 Å, indicate that the DNA assumes a conformation perpendicular to the surface. The increase in $R_{g,2D}$ for distances < 8 Å, and the corresponding decrease in $R_{g,1D}$, indicate that the DNA must orient itself parallel to the surface as it approaches very closely, probably due to steric constraints imposed by the surface. The overall R_g does not change significantly until distances < 8 Å are reached,

below which the value of R_g increases by roughly 10%, probably due to steric constraints imposed by the surface.

7.6.3 ssDNA adsorbs to a hydrophobic surface in a 'bases-down' orientation, but there is no preferential orientation of the ssDNA on a hydrophilic surface

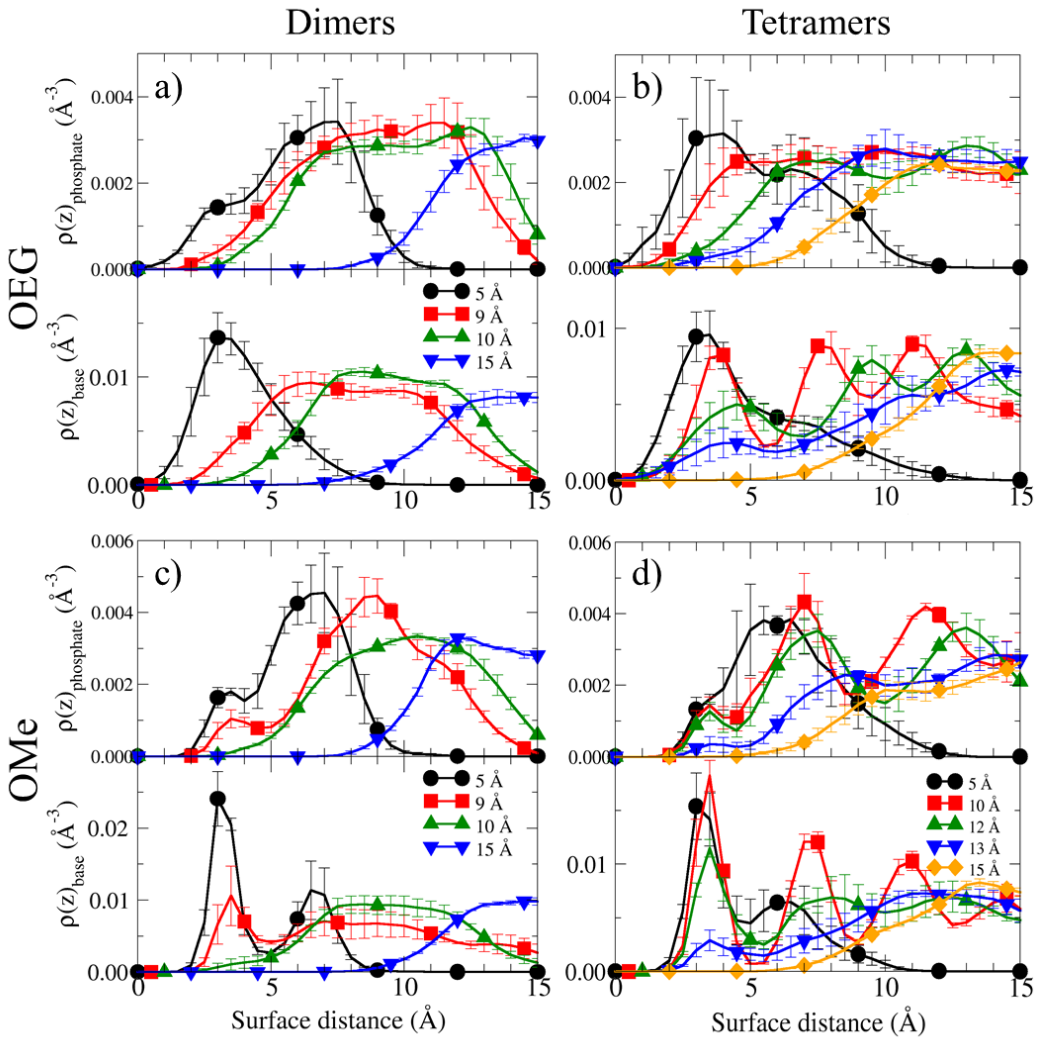


Figure 7.10 Density distributions of phosphate atoms (top half of each panel) and nucleobase atoms (bottom half of each panel) for **adenine**-based dimers and tetramers when the ssDNA is held at various distances from the indicated surfaces. The legends indicate the distance at which the ssDNA is held from the surface. Phosphate atoms include the phosphorous atom and the four oxygen atoms bonded to the phosphorous atom. Nucleobase atoms include only the atoms that comprise the nucleobases and exclude all sugar and phosphate backbone atoms.

Examining the density distributions on OEG when the ssDNA is held at 5 Å from the surface (black lines in panels (a) and (b)), neither the bases nor the phosphates show a preference for the surface or the bulk; i.e., the base and phosphate density distributions occupy roughly the same spatial region. The same is true on OEG at larger DNA-surface separation distances (e.g., the red lines in panels (a) and (b), where the ssDNA is held at 10 Å from the surface). We interpret this to mean that there is no preferred orientation for adsorption to OEG.

Examining the distributions on OMe, however, shows that the bases preferentially locate at the surface and the phosphates remain farther from the surface. For both ssDNA lengths held at 5 Å from the OMe surface (black lines in panels (c) and (d)), there is a sharp peak in the base density around 3 Å, while the peak in the phosphate density is farther from the surface at 6 Å. For the tetramers on OMe, this pattern persists even at larger DNA-surface separation distances: there is a peak in the base density at 3 Å, and the phosphate density is primarily located at farther distances, at DNA-surface separation distances of 10, 12, and 13 Å. We interpret this to mean that the ssDNA adsorbs preferentially in a "bases-down" conformation where the bases are close to the surface and the phosphates are farther from the surface. This conformation might leave the nucleobases unavailable for base-pairing with other ssDNA strands in bulk solution.

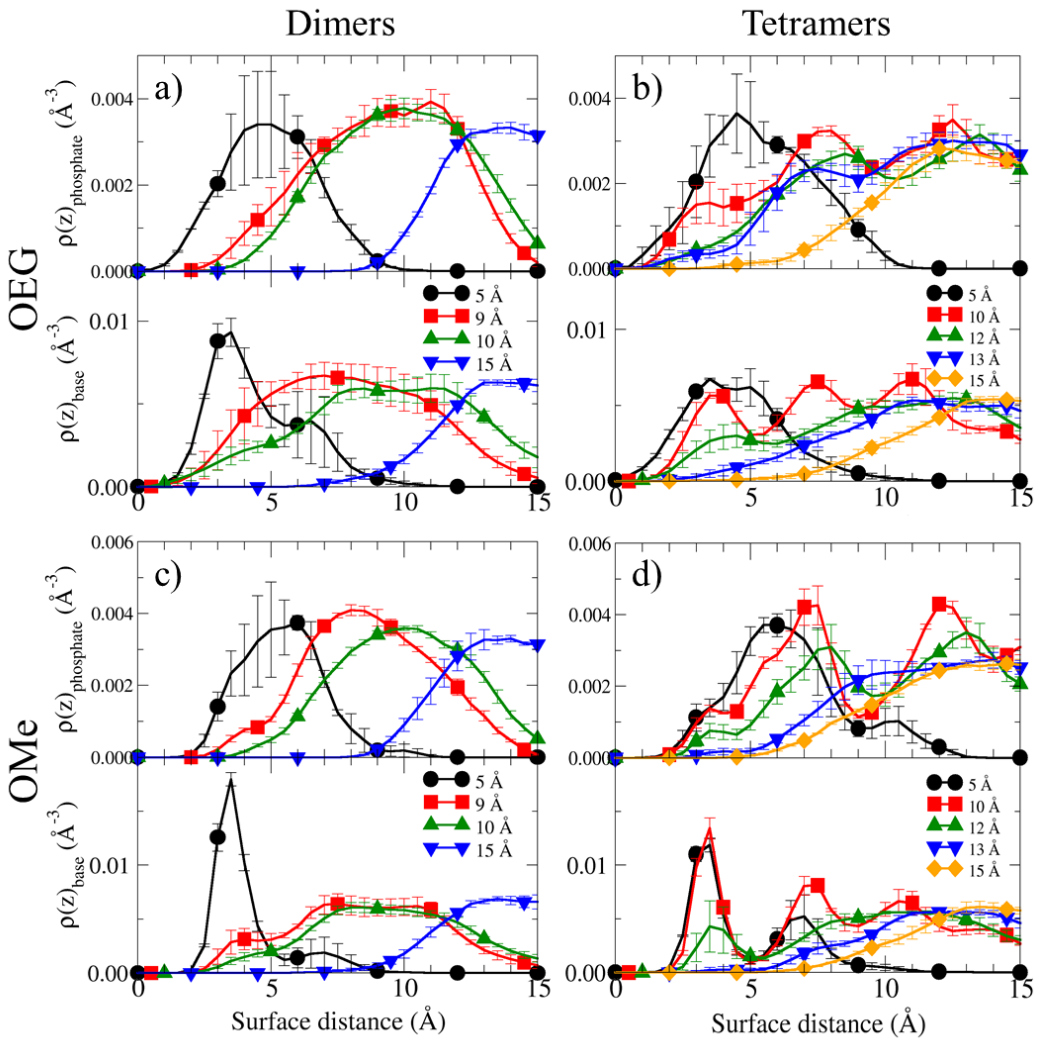


Figure 7.11 Density distributions of phosphate atoms (top half of each panel) and nucleobase atoms (bottom half of each panel) for **cytosine**-based dimers and tetramers when the ssDNA is held at various distances from the indicated surfaces. The legends indicate the distance at which the ssDNA is held from the surface. Phosphate atoms include the phosphorous atom and the four oxygen atoms bonded to the phosphorous atom. Nucleobase atoms include only the atoms that comprise the nucleobases and exclude all sugar and phosphate backbone atoms.

The same commentary as for Figure 7.10 applies here: on OMe, the bases locate preferentially to the surface ("base-down") and the phosphates are farther from the surface; while on OEG, there is no preferred orientation of the bases or phosphates.

7.6.4 More favorable energy per ssDNA-surface contact area on OEG than on OMe indicates greater strength of hydrophilic interactions compared to hydrophobic interactions

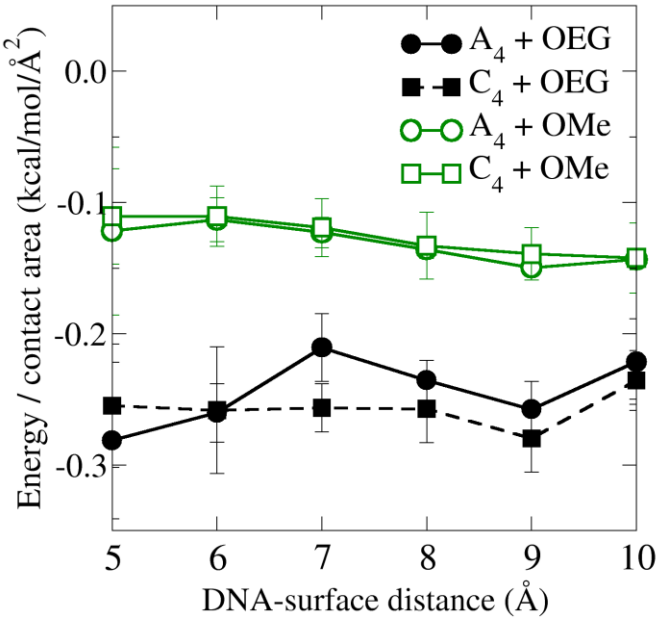


Figure 7.12 Total non-bonded energy (van der Waals and electrostatic) between ssDNA and the surface divided by the contact area between ssDNA and the surface. Only DNA-surface distances less than 10 Å are shown to omit regions with very low DNA-surface contact area, which lead to physically unreasonable values of this quantity (i.e., long-range interactions do not require direct contact, leading to non-finite values when the contact area is zero). Energy values are negative because the energy between the ssDNA and the surfaces is favorable. The average value on OMe is approximately -0.125 kcal/mol/Å^2, while the average value on OEG is approximately -0.25 kcal/mol/Å^2, showing that the energy per contact area on OEG is approximately double that on OMe. We interpret the greater energy per contact area with OEG as evidence that hydrophilic interactions have a stronger enthalpic component than hydrophobic interactions for these systems.

7.6.5 Number of water-water hydrogen bonds per water as a function of surface distance in the absence of DNA

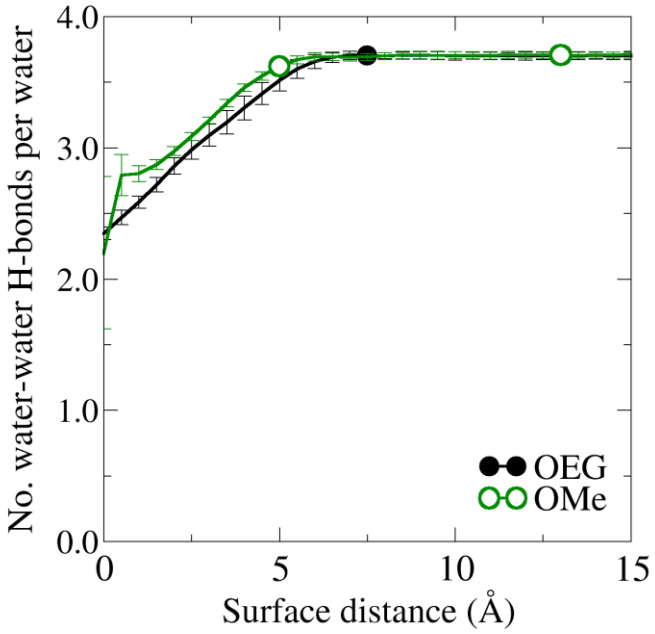


Figure 7.13 Number of water-water hydrogen bonds per water molecule in the absence of ssDNA as a function of surface distance. The average bulk value (surface distances > 5 Å) with our geometric criteria is approximately 3.7 water-water hydrogen bonds per water molecule. Compare this to Figure 7.6, which shows the number of water hydrogen bonds with various molecules, reveals that the water near OEG is more bulk-like than the water near OMe: in the absence of DNA, water molecules near OEG have 3.3 total hydrogen bonds per water molecule, while water molecules near OMe have only 3.1 total hydrogen bonds per water molecule.

7.6.6 Water-DNA hydrogen bonds have a higher duration than water-water hydrogen bonds

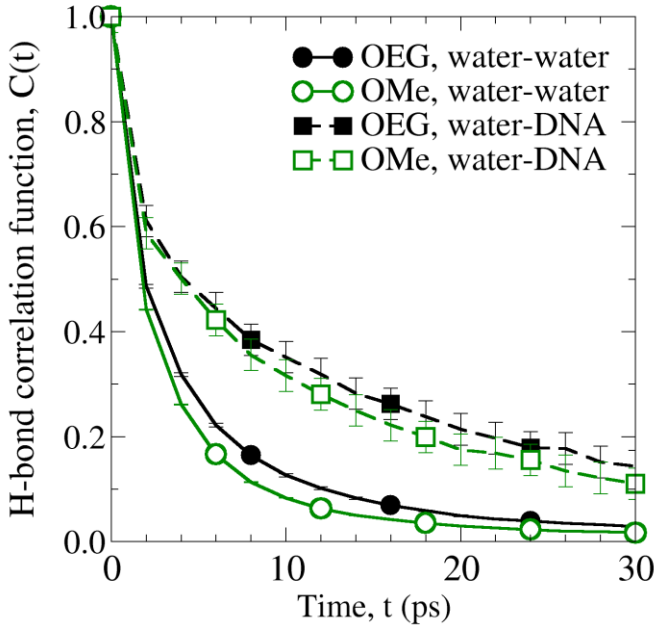


Figure 7.14 Hydrogen bond time correlation function, C(t), of water-water and water-DNA hydrogen bonds. The time correlation function shows the probability that a hydrogen bond still exists at time *t* given that it existed at time 0 [43]. In calculating the correlation function, we have allowed for hydrogen bonds to break and subsequently reform rather than requiring that they remain continuously intact. This manner of calculation emphasizes the fact that water-DNA hydrogen bonds tend to last for longer periods of time, and re-form after temporarily breaking, by showing a long-lasting tail in the correlation function [44]. Since our interest lies here in the behavior of water molecules near the surface, we only consider water molecules within 5 Å of the topmost heavy atoms of the surfaces. Here, we use simulations of an adenine dimer at a DNA-surface separation distances of 5 Å to calculate the water-DNA hydrogen bond correlation functions; results for cytosine and for tetramers are similar. Clearly, water-water hydrogen bonds last for a shorter period of time than water-DNA hydrogen bonds. Over 50% of water-water hydrogen bonds break within 2 ps, and only 10% remain after 10 ps. By contrast, approximately two-thirds of the water-DNA hydrogen bonds still exist after 2 ps, and more than 10% remain even after 30 ps.

7.6.7 Cytosine oligomers show similar behavior to adenine oligomers

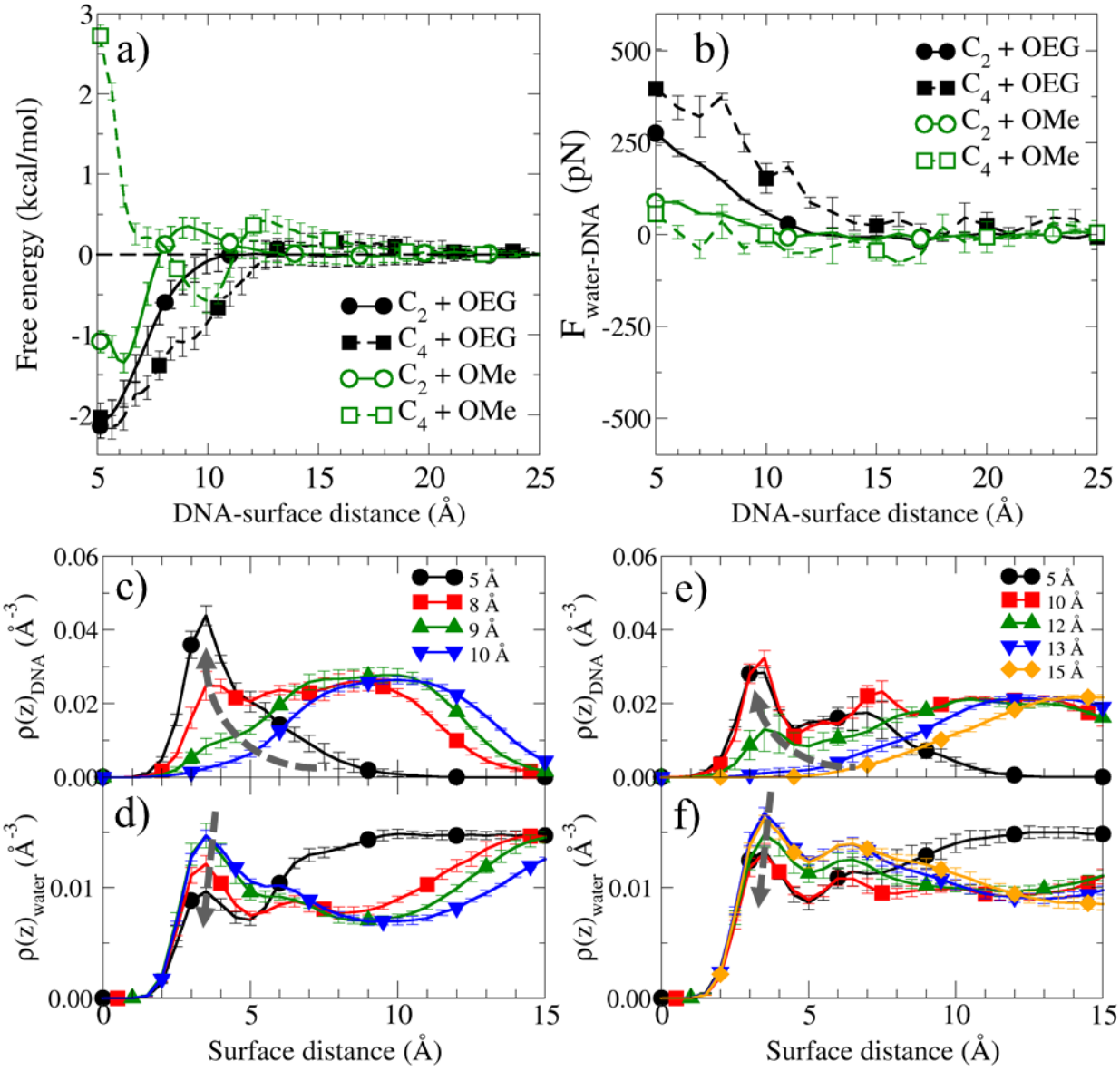


Figure 7.15 Comparison of (a) free energy of adsorption and (b) force of water on ssDNA for the ***cytosine*** dimer (circles) or tetramer (squares) adsorbing to OEG or OMe. The density profiles of water (c, e) and DNA (d, f) when the adenine dimer (c, d) or tetramer (e, f) is held at various distances from the OMe SAM. The dashed gray arrows in (c, e) indicate the increasing density of DNA at the surface as the DNA approaches the surface, and the dashed arrows in (d, f) indicate the decreasing water density as the DNA approaches. This figure is intended to demonstrate that the behaviors shown in Figure 7.7, which pertains to ***adenine***-based ssDNA, are not limited to adenine bases. See discussion of the behaviors observed for adenine bases shown in Figure 7.7, which also apply to cytosine.

7.6.8 Additional comparisons of ssDNA dimers and tetramers showing expected and unexpected differences

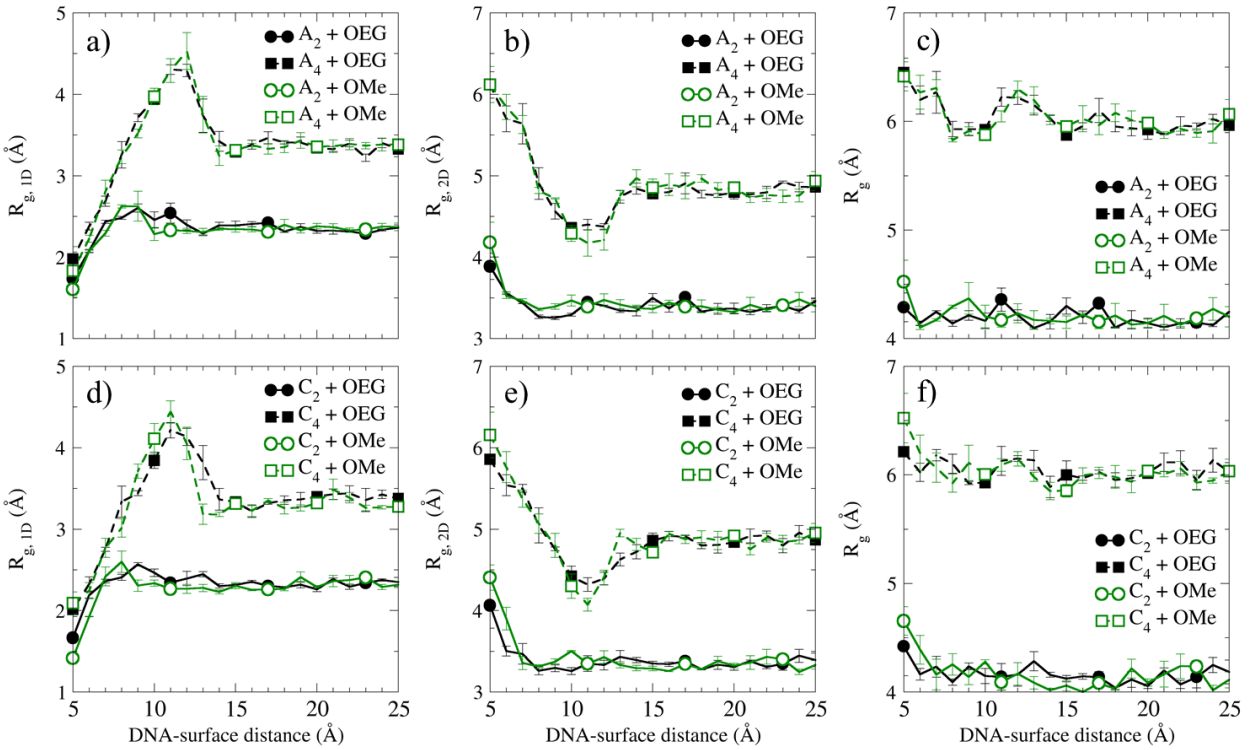


Figure 7.16 The radius of gyration of the ssDNA in (a, d) the z-direction, $R_{g,1D}$, (b, e) the xy-plane, $R_{g,2D}$, and (c, f) in all three dimensions, R_g, as a function of DNA-surface separation distance for (a, b, c) adenine and (d, e, f) cytosine oligomers.

While the tetramers (A_4, C_4) show significant changes in these three measures of size as the DNA-surface distance decreases below 15 Å (discussed in Figure 7.9 and in section 7.3), the ssDNA dimers (A_2, C_2) only show changes in size at very small surface separation distances (< 6 Å). The changes in ssDNA dimer size are most likely due to steric constraints imposed by the surface.

Both the dimers and tetramers reach approximately the same value of $R_{g,1D}$ at a DNA-surface distance of 5 Å, while the value of $R_{g,2D}$ for tetramers is nearly double that of dimers at a distance of 5 Å. These observations suggest that both lengths of oligomer lie parallel to the surface at a DNA-surface distance of 5 Å. In a parallel conformation, the long axis of each oligomer (roughly oriented along the DNA backbone) should be parallel to the surface, and consequently the shorter axes of the oligomers should be oriented perpendicular to the surface. The "width" or "depth" of the oligomers (i.e., the shorter

axes) should be the same for dimers and tetramers (i.e., approximately the "width" or "depth" of a nucleobase), while the "length" of the oligomers (i.e., the long axis) should be different (i.e., approximately the length of the oligomer backbone). Having $R_{g,1D}$ similar between the two oligomer lengths and $R_{g,2D}$ approximately double for the tetramers is therefore consistent with a parallel orientation for both ssDNA lengths at very close surface separation distances.

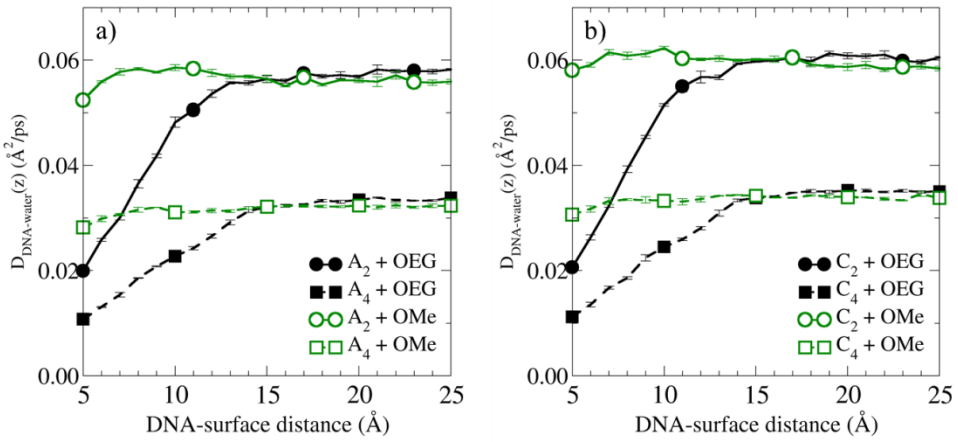


Figure 7.17 Diffusion coefficients of ssDNA dimers and tetramers of a) adenine and b) cytosine as a function of DNA-surface distance. The behavior of dimers and tetramers is similar, but the diffusion coefficients of tetramers are approximately half that of dimers because the ssDNA tetramers have approximately double the mass of dimers.

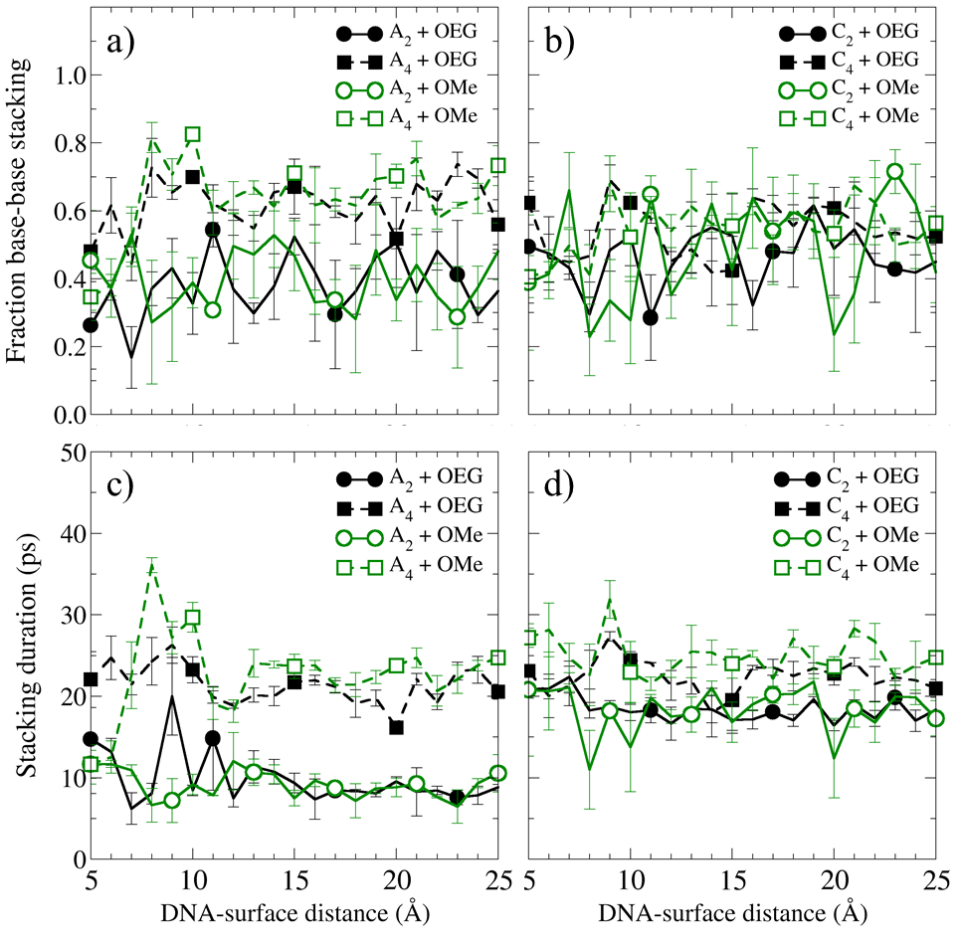


Figure 7.18 Frequency (a, b) and duration (c, d) of base-base stacking of adenine (a, c) and cytosine (b, d) dimers and tetramers as a function of DNA-surface distance. Please note that the frequency of stacking is normalized to the number of base-base stacking interactions that are possible. Tetramers exhibit both greater frequency and duration of base-base stacking than dimers, corroborating our supposition that increased base-base stacking protects tetramer ssDNA from the repulsive force of the low-density region of water, while ssDNA dimers experience the full effect of this repulsive force because they are less capable of base-base stacking (see Figure 7.7).

7.6.9 Details of surface construction and relaxation

After constructing the surfaces and minimizing their energies to eliminate unfavorable atomic contacts, we use the following protocol to further relax the surfaces into their preferred arrangement. The coordinates of the energy minimized surfaces were solvated with TIP3P water molecules to create a 20 Å thick layer on both the +z and –z sides of the surface. It is convenient to place water molecules on both sides of the surface when using the LEaP program of the Amber suite, and the periodic boundary condition in the z-direction ensures that this is equivalent to having water molecules on only one side (top or bottom) of the surface. Then, to further relax the surfaces, we conducted a series of MD simulations. The initial size of the periodic box for these simulations was carefully chosen to ensure that the periodic images of the surfaces were in close contact, thereby eliminating any gaps in the surfaces and creating an infinite surface when combined with periodic boundary conditions in the x- and y-directions, yielding a cross-sectional area of approximately 6.5 nm by 7 nm. During these simulations, the bottommost heavy atom of each oligomer was constrained to a constant z-value to prevent dissolution of the monolayer while still allowing lateral diffusion and rearrangement in the xy-plane. This constraint on the oligomers was applied with the *collective variables* module of NAMD and a force constant of 2.5 kcal/mol-Å was used. First, we conducted constant temperature (300 K), volume, and number of particles (NVT) MD for 100 ps to allow the water molecules to equilibrate. Next, we conducted constant temperature, constant pressure (1 atm), and constant number of particles (NPT) MD for 100 ps with constant xy-area to allow the pressure to equilibrate while maintaining constant grafting density. Finally, we used the coordinates of these relaxed surfaces – with a predetermined grafting density and now equilibrated in the presence of liquid water – for the remainder of our simulations.

7.6.10 Details of production simulation protocol

After constructing fully-solvated systems including ssDNA, neutralizing Na^+ counterions, the self-assembled monolayer (SAM), we used the following procedure to carefully equilibrate the systems prior to use in umbrella sampling simulations. First, in constant volume and temperature conditions (NVT ensemble), the water and ions were minimized with the conjugate gradient algorithm for 2000 steps, with the DNA and SAM subjected to a 500 kcal/mol-A^2 harmonic restraining potential. Second, in the NVT ensemble, with the DNA/SAM restraining potential reduced to 100 kcal/mol-A^2, the system was heated from 0 to 300 K over 20 ps. Third, in constant pressure and temperature conditions (NPT ensemble) and with the restraint reduced to 50 kcal/mol-A^2 the system was relaxed for another 20 ps. Fourth, in the NVT ensemble, the restraining potential was reduced in three steps (50, 10, and 5 kcal/mol-A^2) and minimized for 2000 steps each. Fifth, in the NVT ensemble with the same 5 kcal/mol-A^2 restraint the system was heated from 10 to 300 K over 20 ps. Sixth, in the NPT ensemble, the restraint was reduced from 5 to 1 to 0.1 to 0 kcal/mol-A^2 over 20 ps. Seventh, in the NVT ensemble with no restraints, the system was heated from 10 to 300 K over 20 ps. Finally, the system was heated from 100 to 300 K over 20 ps in the NPT ensemble at the beginning of each production run. A Langevin thermostat and barostat were used to control the temperature (damping coefficient of 1 ps^{-1}) and pressure (piston period of 100 fs and piston decay constant of 50 fs) at 300 K and 1 atm. The SHAKE [56] algorithm was used to constrain all bonds involving hydrogen, and a time step of 2 fs was used. Electrostatic interactions were treated with the particle-mesh Ewald (PME) summation method [56], with a tolerance of 1e-6 and interpolation order of 4. The non-bonded cutoff was 9.0 Å and the non-bonded list was updated every 10 steps. Snapshots were recorded every 2 ps.

7.6.11 Diagram of atoms used to define nucleobase plane

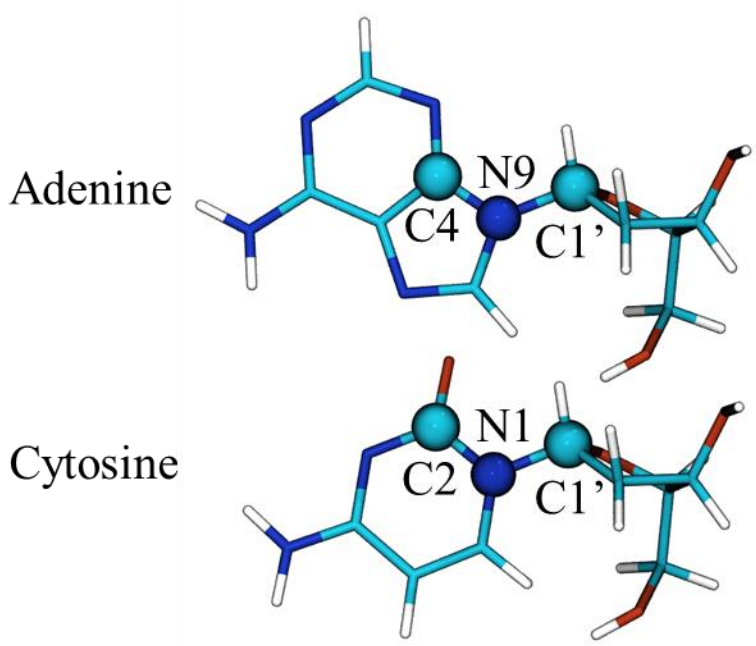


Figure 7.19 Diagram of atoms (shown as spheres) used to define the nucleobase plane of adenine and cytosine. The direction of the base plane normal is given by the cross product (N1–C1') × (N1–C2) for cytosine and (N9–C1') × (N9–C4) for adenine, where (X–Y) is the vector pointing from atom Y to atom X [39].

7.6.12 Details of calculations used for analysis

We assess hydrogen bonding with a geometric criterion involving the three atoms that participate in the hydrogen bond (donor, acceptor, and hydrogen): if the distance between the donor and acceptor is less than 3.5 Å and the angle formed by the three atoms is greater than 120° (i.e., the angle is within 60° of 180°), they are characterized as forming a hydrogen bond (see Figure 7.3b for diagram). We quantify the hydrogen bonding duration and frequency, the latter of which is defined as the percentage of simulation time that a particular type of hydrogen bond is formed. When we calculate hydrogen bonds between the DNA and other molecules, we consider all possible hydrogen bonding partners (donors and acceptors) on the ssDNA rather than only the Watson-Crick hydrogen bonding partners. We also normalize the frequency to the total number of hydrogen bonding partners on the ssDNA to allow comparisons between different DNA bases with different total numbers of hydrogen bonding partners.

Base-base stacking is quantified by applying a geometric criterion involving the distance between two bases and the normal vectors of the planes of the two bases involved. In keeping with standard practice [39], we define the base plane by the plane formed between three atoms in each nucleobase: N9, C4, and C1' for purines and N1, C2, and C1' for pyrimidines (see Figure 7.19 for diagram). If the angle between the normal of two base planes is less than 45° and the distance between the centers of mass of the two nucleobases is less than 4.5 Å, we consider the bases stacked. These geometric cut-offs are chosen because they yield approximately 90% stacking in B-form double-stranded DNA. We introduce another form of analysis, related to base-base stacking, to quantify hydrophobic interactions between the DNA and the surface. These "face-on" conformations, where a DNA base is lying parallel against the surface, facilitate hydrophobic interactions between the DNA and the surface. If the center of mass of a base is within 4.5 Å in the z-direction of the center of mass of the topmost heavy atoms in the surface and the base normal is within 45° of the surface normal, which we define as the z-axis for simplicity, we consider the base to be interacting "face-on" with the surface (see Figure 7.3d for diagram). The base normal for a face-on conformation is defined in the same way as for base-base stacking. As with hydrogen bonds, we quantify the duration and frequency of base-base stacking and face-on conformations, and we normalize the frequency (i.e., the percentage of the entire simulation time that an interaction occurs) to the total possible number of base-base stacking interactions or face-on conformations.

We calculate the radius of gyration (R_g) of the ssDNA as: $R_g = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(r_k - r_{COM})^2}$, where N is the number of atoms comprising the ssDNA, r_k is the coordinates of atom k, and r_{COM} is the center-of-mass of the ssDNA. We choose r_k and r_{COM} to be three-dimensional (x-, y-, and z-coordinates) for the standard measure of R_g, two-dimensional (x- and y-coordinates) to calculate the radius of gyration in the xy-plane ($R_{g,2D}$), or one-dimensional (z-coordinate only) to calculate the radius of gyration in the z-direction ($R_{g,1D}$).

For some calculations, we wish to limit the analysis to only those water molecules that are between the DNA and the surface. To accomplish this, we only use water molecules that are within a cylindrical region of space between the DNA and the surface (see Figure 7.5a for a diagram). The radius of the cylinder is given by the two-dimensional radius of gyration in the xy-plane ($R_{g,2D}$) of the ssDNA, and the cylinder extends vertically (perpendicular to the surface in the z-direction) from the center of mass of the topmost heavy atoms of the surface upward to the center of mass of the DNA.

We calculate the axial density distribution function, ρ(z), of all water and ssDNA atoms as a function of distance from the surface in the +z-direction: $\rho(z) = N(z)/V_{slab}$, where *N(z)* is the number of atoms within 0.5 Å of the z-value within, and V_{slab} is the volume of the rectangular slab at each z-distance (0.5 Å slab height multiplied by the cross-sectional area of the simulation box). For some calculations, we calculate the density distribution function in the same cylindrical region defined in the previous paragraph; for these calculations, we use cylindrical slices of the cylindrical region instead of rectangular slabs of the entire simulation box. We report the density as the number density of the atoms of interest (i.e., water or DNA atoms).

The diffusivity, *D*, of a molecule is calculated using a finite-difference approximation: $D = (\langle|r(t_2) - r(t_0)|^2\rangle - \langle|r(t_1) - r(t_0)|^2\rangle)/(2d(t_2 - t_1))$, where *r* is the coordinate vector of the molecule in question, which may be 1-, 2-, or 3-dimensional, *t* is the time, for which the subscripts (0, 1, 2) indicate three consecutive simulation snapshots, and *d* is the dimensionality of *r* (1, 2, or 3) [40]. Angular brackets denote averaging over all times. Note that the quantity (t_2-t_1) is the length of a single simulation snapshot. To calculate the diffusivity of ssDNA, we use the center of mass of the DNA as the coordinate *r*, while for water molecules we use the coordinate of the oxygen atom. To ensure that the lateral diffusion of the DNA was not affected by the umbrella sampling bias (i.e., no coupling of diffusion in the z-direction to diffusion in the x- and/or y-directions, or, equivalently, a diagonal diffusion tensor) we calculated the off-diagonal components of the diffusion tensor (D_{xy}, D_{xz}, D_{yz}). We used the formalism shown in eq. 6 of Raspopovic *et al.* [57]. In this formalism, the off-diagonal components are calculated as $D_{ij} =$

$\frac{1}{2}\frac{d(\langle ij\rangle-\langle i\rangle\langle j\rangle)}{dt} = \frac{1}{2}\left(\langle iv_j\rangle + \langle jv_i\rangle - \langle i\rangle\langle v_j\rangle - \langle j\rangle\langle v_i\rangle\right)$, where i and j are the x, y, and z coordinates, v_i is the velocity in the i direction, and angular brackets indicate ensemble averaging. This calculation assumes that the diffusion tensor is symmetric. We estimated the velocity at each time step t using a central difference approximation with time steps t+1 and t–1. We found that all three off-diagonal components are essentially zero (data not shown), indicating that there is no coupling between the umbrella sampling bias and lateral diffusion.

We also calculate the tetrahedral ordering of water. The tetrahedral ordering parameter of water oxygen atom k, Q_k, is defined by: $Q_k = 1 - \frac{3}{8}\sum_i^3 \sum_{j=i+1}^4 \left[cos\phi_{ikj} + \frac{1}{3}\right]^2$, where i and j iterate over the 4 nearest water oxygen atoms to water oxygen atom k, and φ_{ikj} is the angle between atoms i, k, and j [41]. For this calculation to function properly, the two vectors defining φ_{ikj} must point from the central atom k to the atoms i and j (i.e., $r_{ik} = r_i - r_k$, where r_k is the Cartesian coordinate of the atom k). A value of $Q_k = 1$ indicates perfect tetrahedral ordering, a value of $Q_k = 0$ indicates random ordering (e.g. an ideal gas), and values of $Q_k < 0$ are possible. We average the value of Q_k over all water oxygen atoms of interest and over all simulation times, and we denote this average value simply as Q. Note that we do not include non-water atoms (e.g., atoms in the surface) in the calculation of Q as is sometimes done [42].

The non-bonded forces and energies between the ssDNA and other components of the system are calculated using the pair interaction feature of NAMD [31]. We only report the forces in the z-direction because the x- and y-components average to zero.

7.6.13 Discussion of limitations to our approach

We have touched on the limitations to our approach in the preceding sections, but we explicitly state and discuss some of these limitations here. First, there are the limitations inherent in any simulation approach, such as the approximate nature of the force field defining the intra- and intermolecular interactions, including the estimation of atomic partial charges. However, our primary interest is in creating idealized hydrophilic and hydrophobic surfaces rather than studying self-assembled monolayers with specific chemistries, and we have shown that our idealized surfaces produce the intended environments by

verifying that they produce the expected behaviors of water in hydrophilic and hydrophobic environments. Second, care must be taken in making comparisons between our idealized surfaces and experimentally studied self-assembled monolayers (SAMs). For instance, most OEG SAMs have somewhat lower coverage than the surface we have constructed, and the most effective surface coverage for preventing protein adsorption has been shown to be approximately 2/3 [27]. Therefore, real surfaces with lower coverage may exhibit weaker ssDNA adsorption than we have found. Similarly, surfaces with different chemistry, such as a methoxy-terminated OEG instead of our hydroxyl-terminated OEG, will likely have different properties [18]. Bearing these limitations in mind, we have found interesting behaviors that could be used to understand more complicated systems and to optimize the adsorption of ssDNA.

Chapter 8

Simulation Study of the Effects of Surface Chemistry and Temperature on the Conformations of ssDNA Oligomers near Hydrophilic and Hydrophobic Surfaces

8.1 Introduction

Many technologies rely on the unique properties of single-stranded DNA (ssDNA). The spontaneous self-assembly of ssDNA can be used to construct predefined two- and three-dimensional nanostructures [1, 2], to direct DNA-grafted nanoparticles into specific spatial arrangements [3-7], and to screen solutions for numerous target nucleic acid sequences simultaneously using DNA microarrays [8, 9]. In each of these DNA-based technologies, the ssDNA is in close proximity to a surface, such as the surface of the microarray or nanoparticle, and the interactions of the ssDNA with these surfaces will impact the success of the application. Similarly, elevated temperatures or controlled temperature annealing procedures might be used control the behavior of DNA to improve the performance of these technologies [10, 11]. However, the behavior of DNA near surfaces, either at room temperature or at elevated temperatures, has received comparatively less attention than the behavior of DNA in bulk solution [12]. Therefore, the present work aims to elucidate some of the effects of surface chemistry and temperature on the interfacial behaviors of DNA.

Previous experimental work has shown that ssDNA exhibits complex interfacial behaviors. Because ssDNA is an amphiphilic polyelectrolyte, it can adsorb to both hydrophilic and hydrophobic surfaces [13-15], and both the conformations and dynamic behaviors of adsorbed ssDNA depend on the hydrophobicity of the substrate, the sequence of DNA bases, and the temperature [14, 15]. To elucidate the molecular-level basis for the complex behaviors observed using experimental methods, simulation

methods have been employed to study ssDNA adsorbed to carbon nanotubes [16-18] and solid surfaces [19], and to study the hybridization behavior of surface-grafted ssDNA strands with ssDNA strands in solution [20-23]. However, few studies have examined the molecular-level interactions of ssDNA with functionalized surfaces or self-assembled monolayers [24, 25], and the temperature-dependence of DNA-surface interactions has also received relatively little attention using simulation methods [18, 26]. We have previously studied the adsorption of short single-stranded DNA oligomers (dimers and tetramers) to hydrophilic and hydrophobic self-assembled monolayers [27], and the current work is an extension to longer ssDNA oligomers (hexadecamers) and also to the effects of temperature.

Using atomistic molecular dynamics simulations combined with umbrella sampling and temperature-replica exchange, we study how the presence and chemistry of a surface affect the conformations and interactions of ssDNA (e.g., size, exposure of nucleobases, base-base stacking, hydrogen bonding). We study an ssDNA oligomer composed of sixteen cytosine nucleotides (C_{16}) near a model hydrophilic oligo(ethylene glycol) surface and a model hydrophobic oligo(methylene) surface (Figure 8.1). Using umbrella sampling, we calculate the free energy as a function of DNA size ($R_{g,DNA}$) in bulk solution and near both of these surfaces, while simultaneously observing changes in DNA structure and interactions as a function of DNA size and surface chemistry. Temperature-replica exchange, in which a set of simulations (replicas) with increasing temperatures are run in parallel and are periodically allowed to exchange temperatures, allows a biomolecular system to overcome energy barriers in high-temperature replicas, which enhances conformational sampling in the low-temperature replicas [28]. Furthermore, using temperature-replica exchange allows us to examine the temperature-dependence of the DNA size, structure, and interactions in bulk and on the two model surfaces.

Overall, we find that the presence and chemistry of the surfaces do not strongly affect the preferred size of the ssDNA ($R_{g,DNA}$) or the structure of the ssDNA (e.g., exposure of the nucleobases, base-base stacking). Neither hydrogen bonds with the hydrophilic surface nor hydrophobic "face-on" interactions with the hydrophobic surface, which are the expected DNA-surface interactions based on our previous work [27], are greatly affected by the size of the ssDNA oligomer. However, the hydrophilic

surface does favor slightly smaller ssDNA sizes than the hydrophobic surface, which appears to be connected to stronger energetic interactions between the more-concentrated ssDNA charge and the hydrophilic surface at smaller ssDNA sizes. In general, increasing temperature has subtle effects that do not depend on the surface presence or chemistry, although several reasonable changes in ssDNA behavior with increasing temperature are observed, such as a reduction in base-base stacking and decreased hydrogen bonding with water molecules at higher temperatures. Lastly, while increasing temperature flattens the entire free energy landscape of ssDNA size, making both compact and extended conformations less unfavorable at higher temperature, the more compact ssDNA conformations show a greater decrease in free energy than the more extended ssDNA conformations. The differential change in free energy between compact and extended ssDNA conformations appears to be related to the entropically favorable release of DNA-bound water molecules by the compact ssDNA molecules.

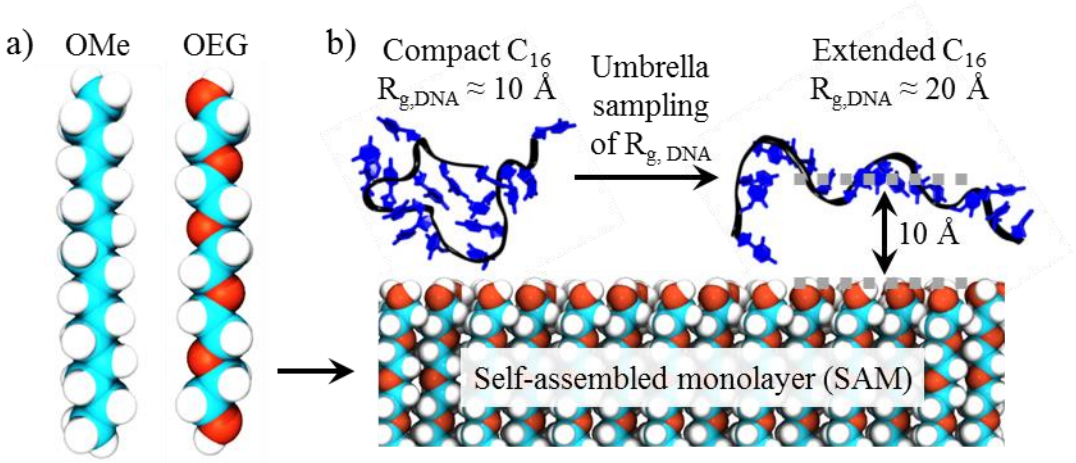


Figure 8.1 a) The hydrophobic (OMe) and hydrophilic (OEG) oligomers used to construct the model hydrophobic and hydrophilic self-assembled monolayers (SAMs). b) Representative simulation snapshots of the cytosine hexadecamer (C_{16}) ssDNA oligomer at two different values of the reaction coordinate used for umbrella sampling of ssDNA size, $R_{g,DNA}$. The reaction coordinate is the radius of gyration of the O5' atoms in the DNA. A schematic of the OEG SAM is also shown (not to scale with ssDNA). In each umbrella sampling window, the value of $R_{g,DNA}$ is biased with a harmonic constraint between 9 Å and 23 Å in 2 Å increments. For the simulations of C_{16} near the OEG or OMe surfaces, the center of mass of the ssDNA oligomer is also constrained to be 10 Å from the top of the SAM (i.e. the center of mass of the topmost heavy atoms in the SAM). The surface-distance constraint is applied for all values of $R_{g,DNA}$ although it is only shown for the extended conformation in this diagram.

8.2 Methods

8.2.1 Surfaces and systems

We constructed hydrophilic (oligoethyleneglycol, OEG) and hydrophobic (oligomethylene, OMe) self-assembled monolayers (SAMs) from oligomers of hydroxyl-terminated OEG (chemical formula $H(OCH_2CH_2)_5OH$) and methyl-terminated OMe (chemical formula $CH_3(CH_2)_{14}CH_3$), as shown in Figure 8.1. These SAMs, which are identical to those in our previous publication [27], were designed to provide model hydrophobic and hydrophilic interfacial environments. The SAMs have a depth greater than 1 nm, ensuring that ideal hydrophobic and hydrophilic environments exist at the top surface of the SAMs: if the SAMs were thinner, the water and DNA molecules at the top of the SAM would be affected by the water molecules at the bottom of the SAM, affecting the apparent hydrophobicity at the top surface of the SAM. Similarly, having such a thick SAM makes it unnecessary to consider ssDNA interactions with a specific solid substrate (e.g., gold, silica), simplifying the simulation design and interpretation of results. We chose a high packing density of oligomers (~5 oligomers / nm^2) to minimize the effects of intercalation into the surface, although further work could be aimed at studying these effects because they may play an important role in adsorption [15]. Additionally, our surfaces do not have fixed coordinates as in some studies [29], allowing the oligomers to rearrange to maximize the formation of favorable contacts with adsorbing molecules. Additional details of surface construction and equilibration are available in supplementary section 8.6.1.

We have studied the conformations (e.g., size, surface area composition) and interactions (e.g., hydrogen bonds) of a single-stranded DNA (ssDNA) oligonucleotide of 16 cytosine bases (C_{16}) near both of the model SAMs and in bulk solution (i.e. without a surface). For the bulk solution simulations, we added 15 Na^+ counterions to neutralize the ssDNA charge and then solvated the system with approximately 15,000 TIP3P water molecules in a $(77.83\ Å)^3$ box. For the simulations containing a surface, we placed the ssDNA oligomer 10 Å above the OEG or OMe surface, added 15 Na^+ counterions, and solvated the system with approximately 7,500 TIP3P water molecules in (65.1 x 68.7 x 71.3) $Å^3$ box.

The distance between the ssDNA oligomer and the SAM is defined by the distance between the center of mass of the ssDNA oligomer and the center of mass of the topmost heavy atoms (i.e. non-hydrogen atoms) in the SAM. The DNA, TIP3P water molecules, and counterions are parameterized using the Amber *ff10* combination of force fields [30-32]. Additional details of system construction are available in supplementary section 8.6.2.

8.2.2 Simulation protocol

After constructing the above systems, they were subjected to the minimization and equilibration procedure described in supplementary section 8.6.2 prior to use in production simulations where we combined umbrella sampling with temperature-replica exchange molecular dynamics.

Umbrella sampling and temperature-replica exchange protocols. To accelerate the sampling of ssDNA conformations, we used the umbrella sampling method [33]. The reaction coordinate defining ssDNA conformation was the radius of gyration of the O5' atoms in the DNA ($R_{g,DNA}$). Only the O5' atoms were used to increase computational efficiency relative to using all of the DNA atoms in the definition of $R_{g,DNA}$. Umbrella sampling windows with values of $R_{g,DNA}$ from 9 Å to 23 Å in 2 Å increments were used, and a harmonic spring with a force constant of 2.5 kcal/mol-Å^2 was used to bias the conformation of the ssDNA near each value of $R_{g,DNA}$. Representative conformations of the ssDNA oligomer at two different values of $R_{g,DNA}$ are shown schematically in Figure 8.1b. Initial configurations for each umbrella sampling window were generated in a stepwise fashion as follows. First, the C_{16} ssDNA oligomer was simulated in bulk solution at 300 K without any biasing potential for 20 ns; the initial configuration for this simulation was a single strand from a dsDNA helix in canonical B-DNA conformation generated using *nucleic acid builder* in the Amber software suite. The final configuration from this unbiased simulation had roughly $R_{g,DNA}$ = 15 Å and was used as the initial configuration for the umbrella sampling window with a biasing potential at $R_{g,DNA}$ = 15 Å. Then, the ssDNA configuration after 1 ns of equilibration in the $R_{g,DNA}$ = 15 Å window was used as the initial configuration for the adjacent umbrella sampling windows (13 Å and 17 Å). In those adjacent windows, the $R_{g,DNA}$ constraint was gradually changed from 15 Å to the value of the current window (13 Å or 17 Å) over 200 ps before

beginning umbrella sampling in that window. The initial configurations for the next set of adjacent windows (13 Å to 11 Å, and 17 Å to 19 Å) were generated in the same fashion, and so on. Temperature-replica exchange, described below, was used during this process to accelerate equilibration, and the initial structure for each temperature-replica came from the corresponding temperature-replica from the preceding umbrella sampling window (i.e. all of the temperature-replicas for a given umbrella sampling window had different starting conformations).

As an alternative to $R_{g,DNA}$, we could have used the end-to-end distance of the ssDNA oligomer to define ssDNA conformation (e.g. the distance between the oxygen atoms at the 5' and 3' terminals). However, the end-to-end distance has a larger range of values than the radius of gyration: the end-to-end distance ranges from approximately 3 Å (i.e. the van der Waals contact distance of two oxygen atoms) up to approximately 54 Å (i.e. 16 bases by 4.3 Å per base, which is the distance between bases in ssDNA [34]), while the radius of gyration only spans the range of 9 Å to 23 Å. The smaller range of relevant values of $R_{g,DNA}$ requires a smaller number of umbrella sampling windows and hence lesser computational expense than the end-to-end distance. Additionally, the radius of gyration defines the size of the ssDNA more uniquely than the end-to-end distance would: for example, in a globular conformation, where $R_{g,DNA}$ is unequivocally low, the end-to-end distance can take on a wide range of values from approximately 3 Å up to values larger than $R_{g,DNA}$). Hence, although the end-to-end distance of the ssDNA oligomer could certainly be useful in other situations, we chose to use $R_{g,DNA}$ to define ssDNA conformations in this case.

The weighted histogram analysis method (WHAM) [35] was used to calculate the free energy as a function of $R_{g,DNA}$, using a freely available implementation [36]. In the WHAM procedure, we used 111 bins to generate the histogram. We estimate the uncertainty of our free energy profiles using Monte Carlo bootstrap error analysis, which is implemented in the freely available WHAM code [36]. The Monte Carlo bootstrap procedure requires a correlation time, which we estimated by observing where the autocorrelation function of $R_{g,DNA}$ reaches zero (approximately 200 ps) and using twice this value (400 ps) as the correlation time (Figure 8.10 in supplementary section 8.6). The error bars on the free energy as a function of $R_{g,DNA}$ are the standard deviation produced by the bootstrapping method. In each window, 7 ns

of simulation were conducted, and the first 1 ns was discarded as equilibration time, yielding 6 ns of sampling per window. Although this is a relatively short simulation time, the sampling was accelerated by the using replica exchange, as described next.

In each umbrella sampling window, we also used the replica exchange method to accelerate sampling of the internal degrees of freedom of the ssDNA and to obtain information about the temperature dependence of DNA conformations [28]. The replicas had temperatures ranging from 273 K to 473 K. The highest temperature, 473 K, was chosen by observing the frequency of base-base stacking over a range of temperatures, which reaches a minimum value at around 473 K. Temperatures higher than 473 K did not substantially decrease the prevalence of base-base stacking compared to 473 K, so we reasoned that using higher temperatures would unduly increase the computational cost without a corresponding increase the speed of conformational sampling. The number of replicas was chosen to yield exchange acceptance ratio of approximately 20%, which is the most computationally efficient acceptance ratio [37, 38]. The temperatures of the replica are distributed exponentially between the lowest and highest temperatures to maintain constant acceptance ratio across the replicas [28]. In bulk solution simulations, 36 replicas were used, while in the near surface simulations, which have a greater number of atoms and therefore require a greater number of replicas to achieve a given exchange acceptance ratio, 72 replicas were used. Adjacent replicas were allowed to exchange temperatures every 1 ps if they met the Metropolis criterion [28].

We conducted several analyses to assess whether these umbrella sampling temperature-replica exchange simulations yielded converged free energy profiles and sampled many different ssDNA conformations. First, we verified that the free energy profiles as a function of $R_{g,DNA}$ have converged within the given sampling time of 7 ns per window by monitoring the change in the free energy profiles with increasing amounts of sampling time (Figure 8.10 in supplementary section 8.6). Second, the autocorrelation of $R_{g,DNA}$ is approximately 200 ps, suggesting that we are sampling a wide variety of independent conformations over the 7 ns simulations (Figure 8.10 in supplementary section 8.6). Third, we observed the movement of each replica through the range of temperatures, and found that each replica

traverses the entire temperature range several times over the course of the 7 ns simulations (not shown). Finally, we conducted a cluster analysis of the DNA structure in the $R_{g,DNA}$ = 15 Å umbrella sampling window at 300 K, using the functionality in VMD with a root-mean-square-deviation (RMSD) cut off of 5.0 Å of all non-hydrogen atoms in the ssDNA. We found that several clusters are generated during 7 ns of sampling (10 clusters containing 5% or more of the structures per cluster). The most populous cluster, containing approximately 40% of the ssDNA structures, does not contain the starting structure, which suggests that the ssDNA has relaxed from its initial configuration and explored many different conformations.

Additional simulation details. After the equilibration protocol described in supplementary section 8.6.2, production simulations were conducted in the constant temperature and constant volume ensemble (NVT). We discuss the choice and limitations of constant volume simulations rather than constant pressure simulations in the conclusion. A Langevin thermostat was used to control the temperature (damping coefficient of 1 ps^{-1}) at the temperature of each replica. The SHAKE [39] algorithm was used to constrain all bonds involving hydrogen, and a time step of 2 fs was used. Electrostatic interactions were treated with the particle mesh Ewald (PME) summation method [39], with a tolerance of 1e-6 and interpolation order of 4. The non-bonded cutoff was 9.0 Å, and the non-bonded list, which had a cutoff of 10 Å, was updated every 10 steps. Snapshots were recorded every 2 ps. In the near surface simulations, the center of mass of the ssDNA was held at approximately 10 Å from the topmost heavy atoms in the SAM using a harmonic constraint with a 10 kcal/mol-$Å^2$ force constant. In the bulk solution simulations, no such center-of-mass constraint was applied to the ssDNA.

8.2.3 Analysis methods

We assess hydrogen bonding with geometric criteria involving the three atoms that participate in the hydrogen bond (donor, acceptor, and hydrogen): if the distance between the donor and acceptor is less than 3.5 Å and the angle formed by the three atoms is greater than 120° (i.e., the angle is within 60° of 180°), they are characterized as forming a hydrogen bond. We quantify the hydrogen bonding frequency, which is defined as the percentage of simulation time that a particular type of hydrogen bond is formed.

When we calculate hydrogen bonds between the DNA and other molecules, we consider all possible hydrogen bonding partners (donors and acceptors) on the ssDNA rather than only the Watson-Crick hydrogen bonding partners. We normalize the frequency to the total number of hydrogen bonding partners on the ssDNA. Hydrogen bonding partners are defined as all non-hydrogen atoms with a partial charge greater than 0.25 or less than –0.25 (i.e. the polar atoms).

Base-base stacking is quantified by applying a geometric criterion involving the distance between two bases and the normal vectors of the planes of the two bases involved. We define the base plane by the plane formed between three atoms in the cytosine nucleobases: N1, C2, and C1' [40]. If the angle between the normal of two base planes is less than 45° and the distance between the centers of mass of the two nucleobases is less than 4.5 Å, we consider the bases stacked. These geometric cutoffs are chosen because they yield approximately 90% stacking in B-form double-stranded DNA. We introduce another form of analysis, related to base-base stacking, to quantify hydrophobic interactions between the DNA and the surface. These "face-on" conformations, where a nucleobase is lying parallel against the surface, facilitate hydrophobic interactions between the nucleobase and the surface. If the center of mass of a base is within 4.5 Å in the z-direction of the center of mass of the topmost heavy atoms in the surface and the base normal is within 45° of the surface normal, which we define as the z-axis for simplicity, we consider the base to be interacting "face-on" with the surface. The base normal for a face-on conformation is defined in the same way as for base-base stacking. As with hydrogen bonds, we quantify the frequency of base-base stacking and face-on conformations, and we normalize the frequency (i.e., the percentage of the entire simulation time that an interaction occurs) to the total possible number of base-base stacking interactions or face-on conformations.

We calculate the surface area of the DNA, and of various components of the DNA, using the solvent-accessible surface area (SASA) calculation method in VMD [41], with the default solvent probe radius of 1.4 Å. We calculate the total surface area of the DNA, the surface area of the nucleobases (i.e. all the atoms in the nucleobase including hydrogen atoms, but no atoms in the sugar-phosphate backbone), and the surface area of the phosphate backbone (i.e. the atoms in the sugar-phosphate

backbone, as defined in VMD, but also including hydrogen atoms, which are not defined as 'backbone' atoms by VMD). We note that the C1', C2', and O4' atoms of the deoxyribose sugar, and the hydrogen atoms attached to those heavy atoms, are not included in the definition of either the nucleobases or the sugar-phosphate backbone, so the sum of the nucleobase and phosphate backbone surface areas is slightly less than the total ssDNA surface area. For the near surface simulations, we note that these surface areas not the solvent-accessible surface areas because we ignore the SAMs in these calculations, so portions of the DNA that are in contact with the SAMs, and hence are not accessible by the solvent, contribute to the surface area. Although we have also calculated the strictly solvent-accessible surface area, we do not find any noteworthy results, so we exclude them from our discussion for the sake of brevity.

The non-bonded forces and energies between the ssDNA and other components of the system are calculated using the pair interaction feature of NAMD [42]. We only report the forces in the z-direction because the x- and y-components average to zero due to the xy-symmetry of the system.

The uncertainty of the above quantities is calculated as the standard deviation by block averaging with 10 blocks. That is, the data were divided into 10 blocks of equal size, the average of each block was calculated, and the standard deviation of the 10 averages is shown as the error bars on all plots except for the free energy, where the Monte Carlo bootstrapping analysis described above is used.

8.3 Results

8.3.1 Effect of surface chemistry on ssDNA structure and size at 300 K

First, we examine how the presence and chemistry of a hydrophilic or hydrophobic surface affect the conformation of the C_{16} ssDNA oligomer at 300 K.

In Figure 8.2, we show the free energy profile as a function of $R_{g,DNA}$ (i.e. the size of the DNA oligomer) for the C_{16} oligomer in bulk water and near (10 Å from the top of the surfaces) the OEG and OMe SAMs. Qualitatively, the shape of the free energy profile is qualitatively the same for the absence and presence of a surface of hydrophilic and hydrophobic chemistry. For all three conditions, the free energy minimum is between 13 Å and 17 Å, and compacting the ssDNA to smaller sizes ($R_{g,DNA}$ < 13 Å)

or extending the ssDNA to larger sizes ($R_{g,DNA}$ > 17 Å) leads to an unfavorable increase in the free energy of several kcal/mol. The preferred values of $R_{g,DNA}$ (13 to 17 Å) are a reasonable range for a semi-flexible polymer (i.e. the worm-like chain model [43], for which the radius of gyration is given by equation 17 of ref. [44]) with persistence length of 40 Å [34] and a contour length of 69 Å (i.e. 4.3 Å per base [34]). There are small *quantitative* differences in the free energy profile between OEG and OMe: the entire free energy profile for OEG, although qualitatively the same shape as that on OMe, is shifted to slightly smaller values of $R_{g,DNA}$ relative to the free energy profile for OMe. On the hydrophilic OEG surface, the free energy minimum is located at $R_{g,DNA}$ = 13 Å while on the hydrophobic OMe surface the free energy minimum is located at $R_{g,DNA}$ = 17 Å. At the highest value of $R_{g,DNA}$ (~23 Å), the value of the free energy on OEG is approximately 0.5 kcal/mol greater than the corresponding value on OMe, indicating that larger ssDNA sizes are more unfavorable on OEG than on OMe. At the lowest value of $R_{g,DNA}$ (~9 Å), the free energy on OEG is approximately 2 kcal/mol lower than the corresponding value on OMe, indicating that smaller ssDNA sizes are less unfavorable on OEG than on OMe. Overall, these observations indicate that slightly smaller ssDNA sizes are favored on OEG relative to OMe. We now describe other aspects of the ssDNA conformation to explain the above observations on the free energy.

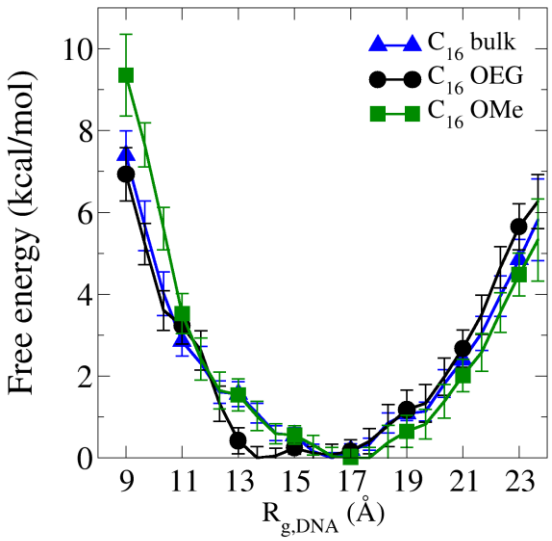


Figure 8.2 Free energy as a function ssDNA size, $R_{g,DNA}$, for the C16 oligomer in bulk solution, 10 Å from the hydrophilic OEG surface, and 10 Å from the hydrophobic OMe surface. Error bars are calculating using the Monte Carlo bootstrapping method described in the methods section.

The next aspect of the conformations of the C_{16} ssDNA oligomer that we investigate is the surface area composition. The composition of the surface area is the fraction of the total surface area that is associated with either the relatively hydrophobic nucleobases or the hydrophilic phosphate backbone. We note that the total surface area of the C_{16} ssDNA oligomer decreases with decreasing $R_{g,DNA}$, as expected, and the total DNA surface area is not significantly affected by the presence or the chemistry of the surfaces (Figure 8.11 in supplementary section 8.6). Figure 8.3 shows the fraction of the total surface area that is composed of phosphate backbone (Figure 8.3a) or nucleobases (Figure 8.3b). We note that these two surface area fractions do not sum to 1 because we have excluded several atoms in the deoxyribose sugar groups from this analysis as stated in the methods section, and these atoms account for approximately 5% of the total ssDNA surface area. In general, the surface area composition is only weakly dependent on $R_{g,DNA}$ or on the presence and chemistry of the surfaces.

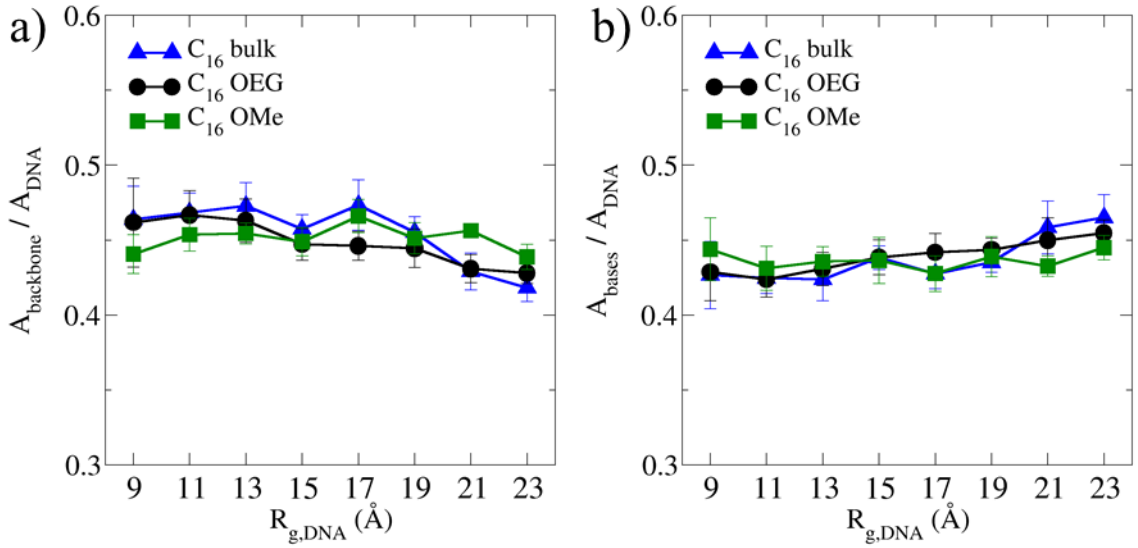


Figure 8.3 Surface area composition of the C_{16} ssDNA oligomer as a function of $R_{g,DNA}$ in bulk solution, 10 Å from the hydrophilic OEG surface, and 10 Å from the hydrophobic OMe surface. The surface area composition is the fraction of the total ssDNA surface area that is composed of a) the hydrophilic phosphate backbone and b) the relatively hydrophobic nucleobases. Error bars are the standard deviation calculated by block averaging over 10 blocks.

In bulk solution, we had expected that the ssDNA would form a micelle-like conformation, with the relatively hydrophobic nucleobases sequestered at the center of a globule and the hydrophilic phosphate backbone at the outside of the globule, as hypothesized based on recent experimental findings

on the behavior of ssDNA oligomers near surfaces of varying hydrophobicity [15]. We find some evidence of such a micellar conformation in these simulations, which we now describe. The conformations with $R_{g,DNA}$ < 19 Å show slightly reduced hydrophobic base surface area (Figure 8.3a) compared to the most extended conformations ($R_{g,DNA}$ > 19 Å). Because the free energy minimum as a function of ssDNA size in bulk is located at approximately $R_{g,DNA}$ = 16 Å (Figure 8.2), which is less than $R_{g,DNA}$ = 19 Å, the most favorable ssDNA size corresponds to a conformation where the nucleobases are somewhat sequestered (i.e. somewhat micelle-like) relative to larger sizes. However, we note that the difference in the exposure of the nucleobases between the smaller, less exposed ($R_{g,DNA}$ < 19 Å) and larger, more exposed ($R_{g,DNA}$ > 19 Å) ssDNA sizes is quite small (< 5%). The fact that we do not observe stronger evidence of a micellar conformation under these conditions may be explained by considering that the molecular modeling conducted by Monserud *et al.* was, as they acknowledge, an energy minimization procedure that neglected entropic effects and therefore yielded limiting cases for the molecular conformation [15]. In molecular dynamics simulations, the effects of entropy are included, and the ssDNA conformations do not reach the much lower nucleobase exposure achieved in those limiting cases. As hypothesized by Monserud *et al.* [15], it seems likely that a micellar conformation will be more evident for longer ssDNA oligomers (> 16 bases), for which the greater number of relatively hydrophobic nucleobases will lead to a greater driving force for hydrophobic aggregation, but testing this hypothesis will require further investigation with simulation methods in the future.

The surface area composition is not significantly affected by the presence or chemistry of the surfaces. One reasonable expectation is that the hydrophilic OEG surface would cause the hydrophobic bases to aggregate to avoid the more hydrophilic environment, which would decrease the exposure of the bases, or that the hydrophilic surface would cause the phosphate backbone to become more exposed to interact more strongly with the polar surface. However, we do not observe any significant differences between OEG and bulk. Near the hydrophobic OMe surface, a reasonable expectation is that the hydrophobic surface would interact strongly with the relatively hydrophobic nucleobases, thereby increasing the exposed hydrophobic DNA surface area with respect to bulk. However, as with OEG, we

do not observe any significant differences in the surface composition between OMe and bulk. The absence of any major differences in surface composition of the DNA in bulk and near surfaces is consistent with the similar free energy profiles as a function of $R_{g,DNA}$ (Figure 8.2), but does not suggest an explanation for why slightly smaller conformations are favored on OEG compared to OMe. To further explore the origin of smaller conformations on OEG, we now consider the molecular-level interactions of the ssDNA with itself and with other components in the system (i.e., the surfaces, water, and ions).

To see if more favorable interactions within the ssDNA molecule could explain the smaller conformations on OEG, we examine the number of intra-DNA hydrogen bonds and base-base stacking. Although the frequency of intra-DNA hydrogen bonds does slightly increase with decreasing $R_{g,DNA}$, which is reasonable because the hydrogen bonding donors and acceptors are in closer proximity when the DNA is smaller, the presence and chemistry of the surfaces does not strongly affect the number of these interactions (Figure 8.4a). Therefore, intra-DNA hydrogen bonds do not appear to explain why smaller DNA conformations are favored on OEG compared to OMe.

Next, we consider base-base stacking, which is somewhat affected by the presence and chemistry of the surfaces for certain values of $R_{g,DNA}$ (Figure 8.4b). At the largest ($R_{g,DNA} \geq 21$ Å) and smallest ($R_{g,DNA} \leq 11$ Å) DNA sizes, the frequency of base-base stacking is unaffected by the surfaces. However, at intermediate DNA sizes, the presence and chemistry of the surfaces affect base-base stacking. At the intermediate DNA sizes (13 Å $\leq R_{g,DNA} \leq$ 17 Å), the ssDNA shows the greatest amount of base-base stacking in bulk solution, which is consistent with the hypothesis of a micelle-like conformation that limits the exposure of the relatively hydrophobic nucleobases. Near OMe the frequency of base-base stacking is not significantly different from bulk solution, which is somewhat surprising because it might be expected that the "face-on" interactions with a hydrophobic surface would compete with base-base stacking. "Face-on" interactions are a measure of hydrophobic interactions that we previously observed for short ssDNA oligomers near the OMe surface [27], and we discuss the prevalence of these interactions for this longer ssDNA oligomer below. Near OEG the frequency of base-base stacking is significantly reduced compared to bulk solution, possibly due to DNA-surface interactions, which we discuss next.

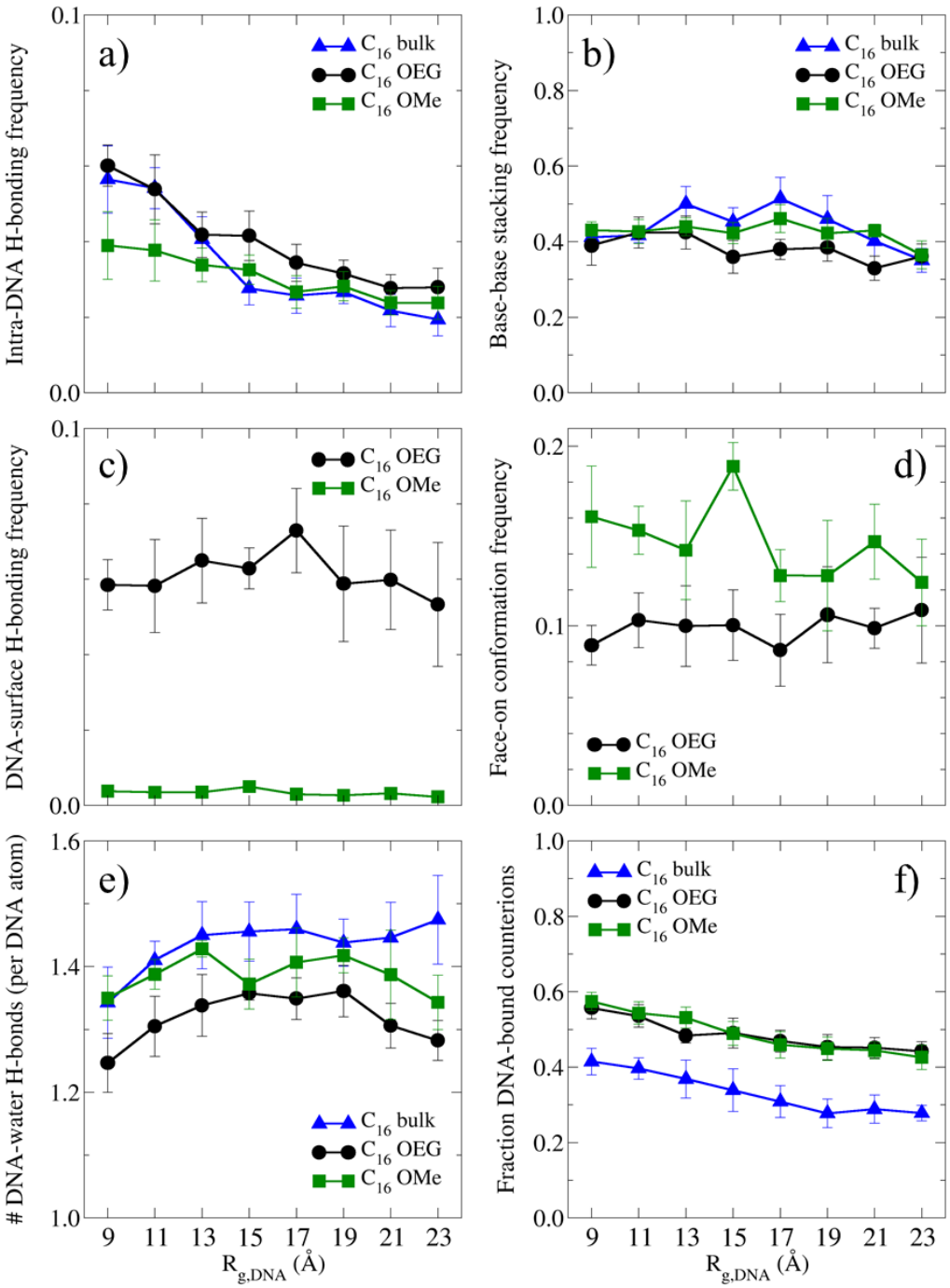


Figure 8.4 Molecular-level interactions of the C_{16} ssDNA oligomer as a function of $R_{g,DNA}$ in bulk solution, 10 Å from the hydrophilic OEG surface, and 10 Å from the hydrophobic OMe surface. a) the frequency of intra-DNA hydrogen bonds b) the frequency of base-base stacking c) the frequency of DNA-surface hydrogen bonds d) the frequency of face-on conformations, which are a type of hydrophobic interaction e) the number of DNA-water hydrogen bonds, normalized to the number of hydrogen bonding partners on the ssDNA f) the fraction of the total number of Na^+ counterions that are bound to the DNA. A counterion is defined as being bound to the DNA if it is within the Bjerrum length (7 Å in water at 300 K) of a phosphate group on the DNA. Error bars are the standard deviation calculated by block averaging over 10 blocks.

Regardless of ssDNA size, the ssDNA oligomer forms a substantial number of hydrogen bonds with the OEG surface but a negligible number of hydrogen bonds with the OMe surface (Figure 8.4c), consistent with our previous study of short ssDNA oligomers [27]. The polar OEG oligomers are capable of forming hydrogen bonds with the DNA, whereas the OMe oligomers are nonpolar and therefore incapable of hydrogen bonding. We hypothesize that DNA-OEG hydrogen bonds destabilize base-base stacking on the OEG surface relative to bulk solution (Figure 8.4b). To quantify hydrophobic DNA-surface interactions, we calculate the frequency of "face-on" states, wherein a nucleobase is close to and parallel to the surface. A nucleobase in a face-on conformation displaces a large number of water molecules from the surface, which makes face-on states a reasonable measure of hydrophobic interaction. Regardless of ssDNA size, we find that the DNA more frequently forms face-on interactions with the hydrophobic OMe surface than with the hydrophilic OEG surface (Figure 8.4d), which is also consistent with previous study of short ssDNA oligomers [27]. We note that the ssDNA forms a significant number of face-on interactions with the hydrophilic OEG surface because of the relatively large tolerances we employ in our geometric definition of a face-on state (4.5 Å surface distance and 45° between base and surface normal vectors). In contrast with DNA-surface hydrogen bonds, face-on interactions do not appear to destabilize base-base stacking relative to bulk (Figure 8.4a). It is interesting to note that $R_{g,DNA}$ does not significantly affect these DNA-surface interactions. One reasonable expectation is that the ssDNA would form more face-on interactions with the OMe surface when in extended conformations, which we do not observe. The absence of this finding is consistent with our observation that base-base stacking, which might compete with face-on interactions, is not affected by $R_{g,DNA}$ near the OMe surface and is also consistent with our observation that the exposed hydrophobic base surface area does not significantly increase with increasing $R_{g,DNA}$. Although these DNA-surface interactions are of fundamental interest, they do not clearly explain why slightly smaller DNA conformations are favored on the OEG surface.

We now consider interactions between the DNA and water. DNA-water interactions, measured through the number of DNA-water hydrogen bonds per polar DNA atom, do not show an obvious trend

with $R_{g,DNA}$ (Figure 8.4e). The average number of DNA-water hydrogen bonds is largest in bulk because the presence of a surface necessarily decreases the number of water molecules near the DNA molecule compared to bulk solution (Figure 8.12 in supplementary section 8.6). The average number of DNA-water hydrogen bonds is lowest near the OEG surface because some DNA-water hydrogen bonds are replaced by DNA-surface hydrogen bonds (Figure 8.4c), and because the OEG surface, by interacting strongly with the DNA through hydrogen bonding, displaces some water molecules from near the DNA (Figure 8.12 in supplementary section 8.6). The average number of DNA-water hydrogen bonds when the DNA is near the OMe surface is intermediate between bulk and OEG because the OMe surface does not form hydrogen bonds with the DNA but, simply by its presence, reduces the number of water molecules near the DNA. Overall, water-DNA interactions do not suggest an obvious explanation for the slightly smaller conformations of DNA on OEG compared to OMe.

Next, we examine DNA-ion interactions, which are quantified by the number of DNA-bound counterions. An Na^+ counterion is defined as being "bound" to the DNA if it is within the Bjerrum length (7 Å in water at 300 K) of a negatively charged phosphate. Both in bulk and on the two surfaces, we observe an increase in the fraction of DNA-bound counterions as $R_{g,DNA}$ decreases, indicating that more compact DNA conformations have increased interactions with counterions. This finding can be rationalized by considering that at lower $R_{g,DNA}$ the negatively charged phosphate groups are closer together, and the greater density of negative charge (per volume) attracts a greater number of counterions (i.e. cooperative binding of the counterions by the phosphates). We note that the fraction of DNA-bound counterions is essentially the same near both OEG and OMe at all values of $R_{g,DNA}$ but is considerably lower for the DNA in bulk. The reason for this is the greater concentration of counterions in the surface systems, which is due to the smaller total amount of water in the surface-containing systems relative to the bulk systems.

To better understand the significance of increasing DNA-counterion interactions with decreasing $R_{g,DNA}$, and of the other molecular-level interactions of DNA discussed above, we now examine the energetics of the interactions of DNA with the other components in the systems. In Figure 8.5 we present

the energy of interactions between the C_{16} ssDNA oligomer and other components in the systems (i.e. water, counterions, and, the hydrophilic OEG or hydrophobic OMe surfaces) as a function of DNA size.

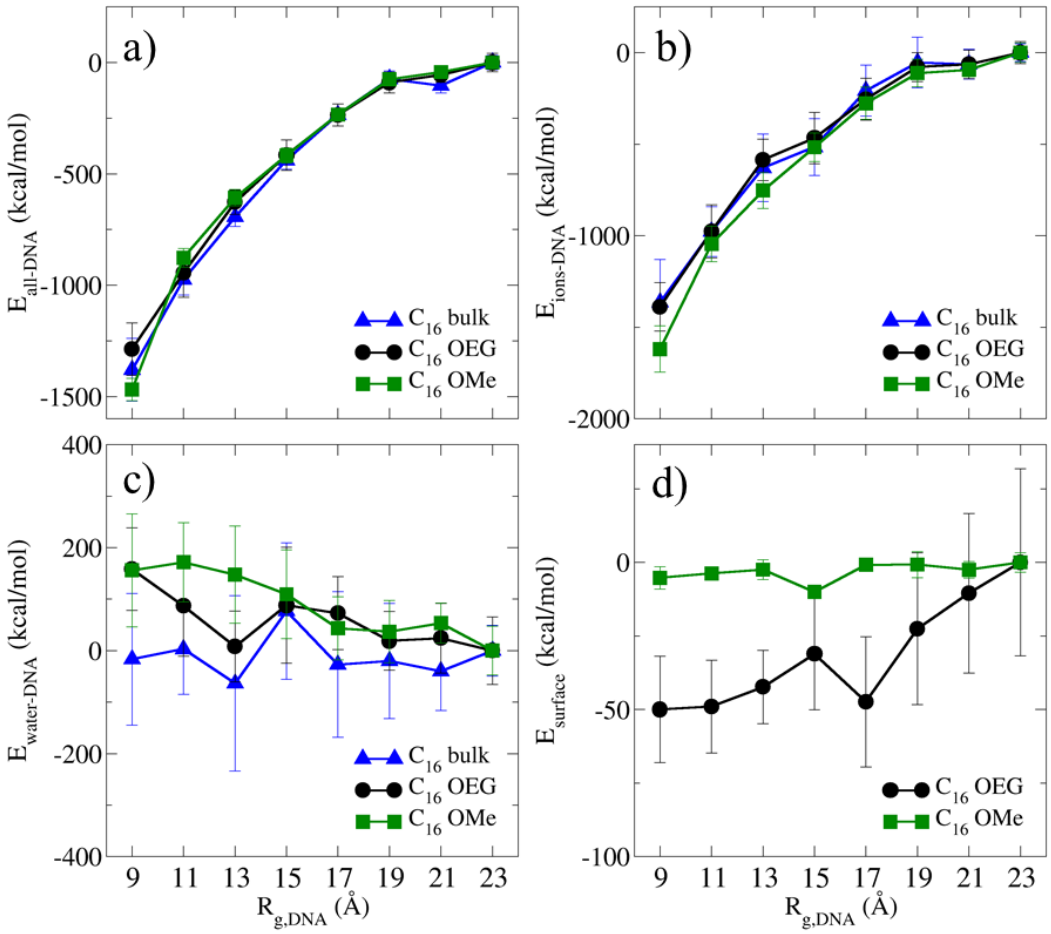


Figure 8.5 Energetic interactions between the C_{16} ssDNA oligomer and a) everything in the system (water, counterions, and surface), b) Na^+ counterions, c) water molecules, and d) the OEG or OMe surfaces as a function of $R_{g,DNA}$ in bulk solution, 10 Å from the hydrophilic OEG surface, and 10 Å from the hydrophobic OMe surface. Error bars are the standard deviation calculated by block averaging over 10 blocks. Energies are shifted vertically such that the energy is zero at $R_{g,DNA}$ = 23 Å.

The energies have been shifted vertically so that the energy for largest $R_{g,DNA}$ is zero kcal/mol, which allows us to visualize the effect of decreasing DNA size on the interaction energies. First, in Figure 8.5a, we present the total interactions *of DNA with everything else* in the system (water, counterions, surface). We emphasize that this is not the potential energy of the entire system. We do not include a discussion of the total potential energy of these systems because, as is common in biological simulations, the value and fluctuations of the total potential energy are largely dictated by the relatively large volume of water. As

such, it is difficult to discern differences in the total potential energy (Figure 8.13 in supplementary section 8.6). In bulk and near both surfaces, the total DNA interaction energy decreases substantially (i.e. becomes more energetically favorable) with decreasing $R_{g,DNA}$ (Figure 8.5a). The large decrease in the total interaction energy with decreasing $R_{g,DNA}$ is primarily due to the decrease in the DNA-ion interaction energy that occurs in bulk and on both surfaces (Figure 8.5b). The large decrease in the DNA-ion interaction energy is reasonable given that the fraction of bound counterions increases as $R_{g,DNA}$ decreases (Figure 8.4f): the closer proximity of the ions to the DNA leads to stronger DNA-ion energetic interaction. We note that the *absolute* values of DNA-ion interaction energy are less favorable in bulk solution, where the fraction of bound counterions is lower, and more favorable for the ssDNA near the surfaces, where the fraction of bound counterions is higher, but the *change* in the DNA-ion interaction energy, and the change in the fraction of bound counterions, relative to the largest ssDNA conformations (Figure 8.4f) is the same for bulk and both surfaces.

Another contribution to the total interaction energy is the DNA-water interaction energy (Figure 8.5c). The change in the DNA-water interaction energy with $R_{g,DNA}$ is much smaller than the change in the DNA-ion interaction energy, and the error bars for bulk and both surfaces overlap, making it difficult to distinguish any differences. However, the remaining contribution to the total interaction energy, the DNA-surface interaction energy, shows clear surface-dependent effects (Figure 8.5d). As the size of the DNA decreases, the interactions with OMe do not change significantly, but the interactions with the hydrophilic OEG surface become much more favorable. The increased favorability of DNA-OEG interaction energy with decreasing $R_{g,DNA}$ suggests an energetic motivation for the slightly smaller ssDNA sizes that are favored on OEG compared to OMe (Figure 8.2). We propose some possible explanations for the increase in DNA-OEG interaction with decreasing DNA size. We hypothesize that the higher charge density of the DNA at lower $R_{g,DNA}$ increases DNA-OEG interactions in the same way that the more compact DNA molecule binds more counterions. Additionally, it may be the higher charge density of the DNA-counterion complex, rather than just the DNA itself, that increases DNA-surface interactions with the polar OEG surface.

Another factor contributing to the smaller DNA size on OEG may be the entropic penalty of confining the counterions near the DNA. DNA-bound counterions have reduced translational entropy, and because the fraction of DNA-bound counterions is the same for both OEG and OMe, the entropic penalty of confining the ions is equal for both surfaces. We suggest that the loss of counterion translational entropy is partially offset by favorable energetic interactions between the DNA-bound counterions and the polar OEG surface, but *not* offset by the essentially negligible interactions between DNA-bound counterions and the non-polar OMe surface. Overall, the increased energetic interactions between the ssDNA and the OEG surface with decreasing $R_{g,DNA}$ provide a plausible explanation for the slightly smaller conformations of the ssDNA on the hydrophilic OEG surface relative to the hydrophobic OMe surface.

8.3.2 Effect of temperature on ssDNA structure, energetics, and size

We now discuss the temperature-dependence of the structure, size, and interactions of the C_{16} ssDNA oligomer, which we obtained using temperature-replica exchange molecular dynamics. In general, we note that the qualitative temperature-dependent behavior of the ssDNA is independent of the presence or chemistry of the surfaces, but we highlight cases where differences between bulk and the two surfaces exist.

We now present discuss aspects of the molecular-level structure and interactions as a function of temperature and $R_{g,DNA}$. The ssDNA surface area composition (i.e. the fraction of the total DNA surface area that is composed of nucleobases or phosphate backbone) shows temperature-dependence in bulk and near both surfaces: as temperature increases, the fraction of backbone surface area decreases (Figures 8.14a for bulk, 8.15a for OEG, and 8.16a for OMe in supplementary section 8.6.7) and the fraction of nucleobase surface area slightly increases (Figures 8.14b, 8.15b, and 8.16b in supplementary section 8.6.7). We note that the trend of decreasing nucleobase surface area with decreasing $R_{g,DNA}$ for the ssDNA in bulk at 300 K (Figure 8.3a) is also observed at most of the other temperatures, which indicates that some micelle-like behavior (i.e. sequestering of the nucleobases) may also be present at these temperatures. The cause of the reduction in the phosphate backbone exposure with increasing temperature

is complemented by the corresponding increase in the nucleobase exposure. The cause of the slight increase in the nucleobase surface area appears to be a decrease in base-base stacking with increasing temperature (Figure 8.6a for bulk, and Figures 8.15c and 8.16c in supplementary section 8.6.7), which was an expected with increasing temperature.

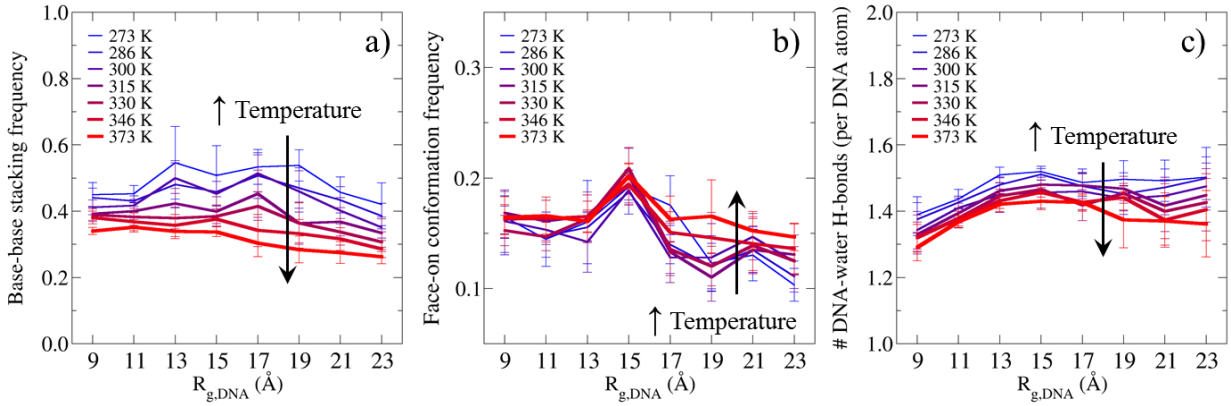


Figure 8.6 Effect of temperature on several structural characteristics of the C_{16} ssDNA oligomer as a function of the ssDNA size ($R_{g,DNA}$). The effect of increasing temperature is indicated by the color of the lines changing from blue to red, by an increase in line thickness, and by the direction of the arrows. a) Frequency of base-base stacking in bulk solution, which is representative of the behavior near OEG and OMe. b) Frequency of face-on conformations with the hydrophobic OMe surface, which increases slightly with temperature for the extended ssDNA conformations ($R_{g,DNA} \geq 19$ Å). c) Number of DNA-water hydrogen bonds, normalized to the number of hydrogen bonding partners on the ssDNA in bulk solution, which is representative of the behavior near OEG and OMe. Error bars are the standard deviation calculated by block averaging over 10 blocks.

Along with the decrease in base-base stacking with increasing temperature (Figure 8.6a), the prevalence of face-on interactions with the hydrophobic OMe surface increases slightly with temperature for the most extended ssDNA conformations (Figure 8.6b). This finding is interesting because it could indicate stronger hydrophobic interactions of the ssDNA with a hydrophobic surface at increased temperature, but we must note that, because of the constant-volume simulations we have used and because the hydrophobic effect is temperature- and pressure-dependent [45, 46], it is difficult to draw conclusions about hydrophobic interactions in constant-pressure conditions. The DNA-water hydrogen bonds decrease with increasing temperature (Figure 8.6c), which is reasonable given that increasing temperature increases the translational and rotational motion of water molecules, making it easier to break hydrogen bonds with water.

Next, we consider the interaction energy of the ssDNA with other components in the system. Figure 8.7 shows the energetic interactions of the DNA near OEG, which are representative of the temperature-dependence of these interactions in bulk and near OMe, which are shown in supplementary section 8.6.8, with the exception of the DNA-surface interaction energy as discussed below. We note that Figure 8.7 shows the absolute values of the energy, unlike Figure 8.5 which shows the energy relative to

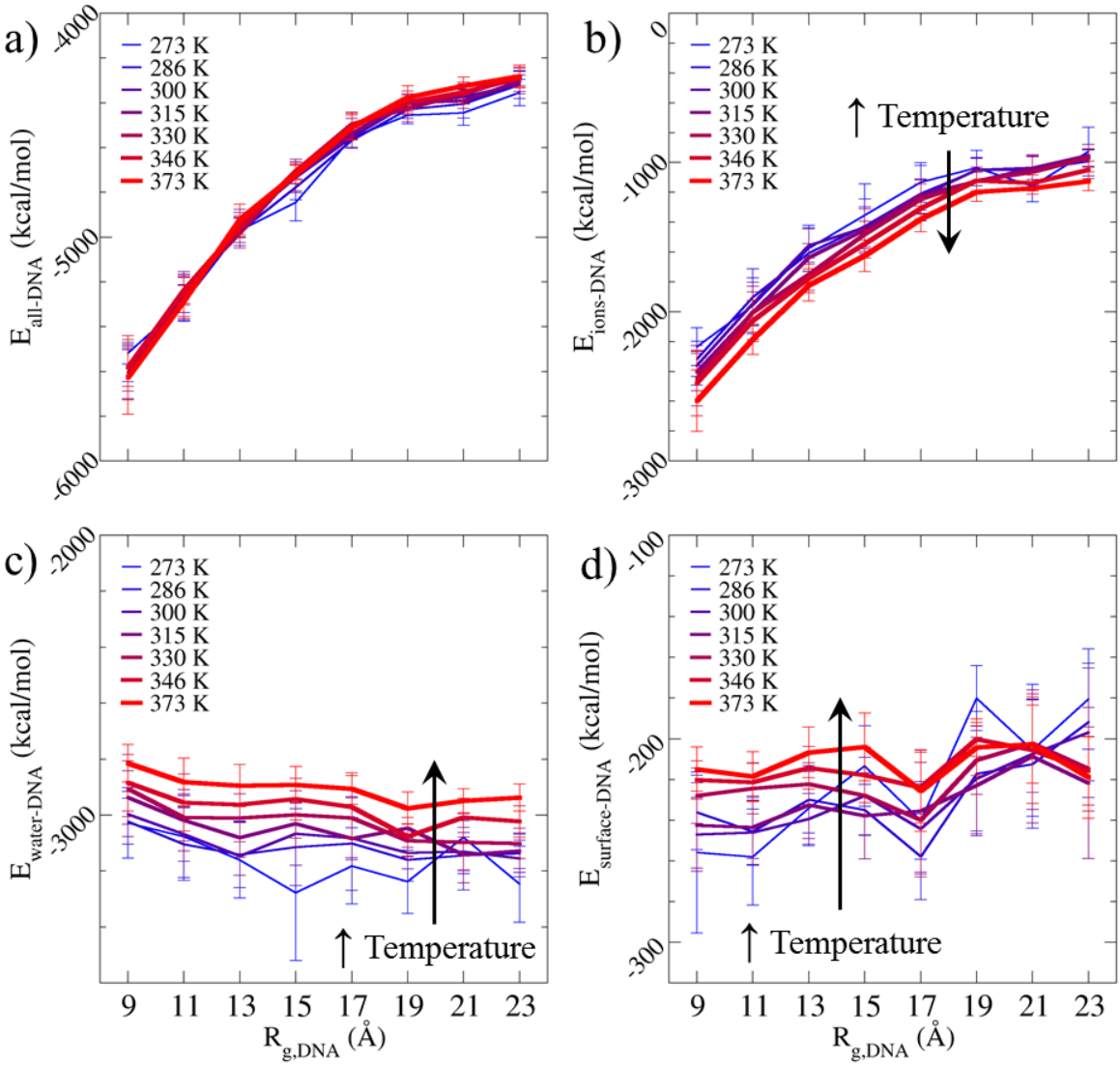


Figure 8.7 Effect of temperature on energetic interactions of the C_{16} ssDNA oligomer 10 Å from the hydrophilic OEG surface, which is representative of the behavior in bulk and near the hydrophobic OMe surface, as a function of the ssDNA size ($R_{g,DNA}$). The effect of increasing temperature is indicated by the color of the lines changing from blue to red, by increasing thickness of the lines, and by the direction of the arrows. The energetic interactions are between the C_{16} ssDNA oligomer and a) everything in the system (water, counterions, and surface), b) Na^+ counterions, c) water molecules, and d) the OEG or OMe surfaces. Error bars are the standard deviation calculated by block averaging over 10 blocks.

the largest value of $R_{g,DNA}$. The total interaction energy does not change significantly with increasing temperature, although there is a very slight upward trend with increasing temperature (Figure 8.7a), which is expected because the increasing importance of entropy with increasing temperature should generally allow the system to sample states with less favorable energetic interactions (i.e. increased total interaction energy).

The DNA-ion interaction energy becomes slightly *more* favorable with increasing temperature (Figure 8.7b), and DNA-water interaction energy becomes *less* favorable with increasing temperature (Figure 8.7c). The latter is reasonable because of the expected increase in water entropy with increasing temperature and because the number of DNA-water hydrogen bonds decreases with increasing temperature (Figure 8.6c). We conjecture that the increase DNA-water interaction energy with increasing temperature is related to the decrease in DNA-ion interaction energy. First, as the water interacts less strongly with the DNA at higher temperatures, the counterions are able to interact more strongly with the DNA at higher temperatures. Second, the counterions, each bearing a positive charge, have a much stronger electrostatic drive to interact with the negatively charged phosphate backbone than the polar-but-uncharged water molecules do. Consequently, the increase in the importance of entropy with increasing temperature disfavors water-DNA binding more strongly than it disfavors ion-DNA binding.

Near OEG (shown in Figure 8.7d), the DNA-surface interaction energy becomes less favorable with increasing temperature, likely due to the same reasons that the water-DNA interaction energy becomes less favorable with increasing temperature, which is that dominant entropic contributions steer the system away from states with low entropy but favorable energy. The reduction in DNA-surface interactions with increasing temperature is not evident in the number of DNA-surface hydrogen bonds (Figure 8.19 in supplementary section 8.6), possibly because of the imprecise nature of the geometric definition hydrogen bonds. Near OMe (Figure 8.18 in supplementary section 8.6), the change in DNA-surface interaction energy with temperature and $R_{g,DNA}$ is negligible.

Having described the effect of temperature on the molecular-level structure and interactions, we now shift our focus to the effect of temperature on the free energy as a function of $R_{g,\,DNA}$ at varying

temperatures in bulk solution and near the OEG and OMe surfaces in Figure 8.8. Qualitatively, since the free energy profile is essentially the same shape irrespective of the temperature, increasing the temperature does not drastically change free energy landscape (e.g., by the introduction of a metastable state). Clearly, and not unexpectedly, increasing the temperature generally flattens the free energy profile (i.e. decreases the free energy for all values of $R_{g,DNA}$). Interestingly, the amount by which the free energy decreases is different for bulk, OEG, and OMe, and within each of the three cases the amount of the decrease is different for the compact conformations ($R_{g,DNA}$ < 16 Å) and the extended conformations ($R_{g,DNA}$ > 16 Å). Qualitatively, it appears that the free energy profile decreases more for the compact conformations than for the extended conformations, a difference that we now quantify and investigate in greater detail.

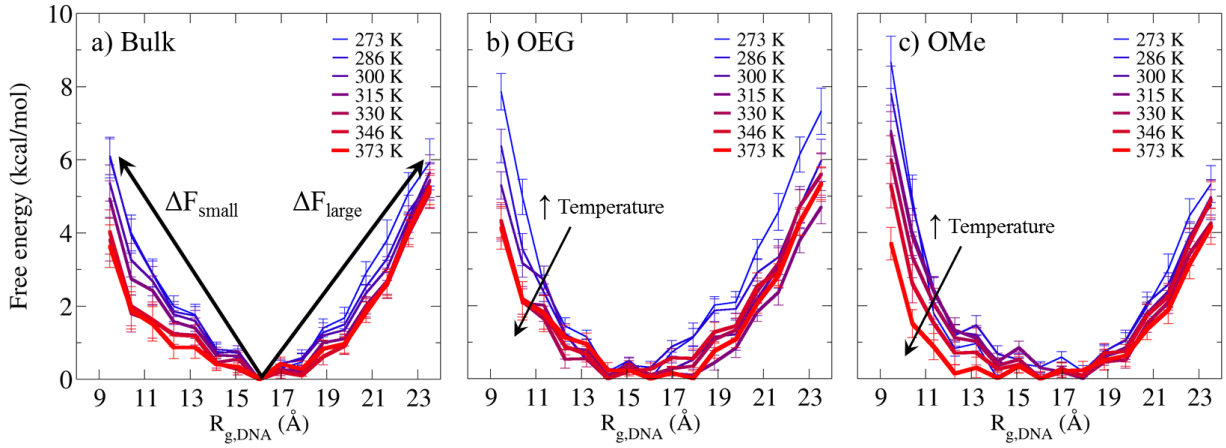


Figure 8.8 Effect of temperature on the free energy as a function of DNA size ($R_{g,DNA}$) for the C_{16} ssDNA oligomer a) in bulk solution, b) 10 Å from the hydrophilic OEG surface, and c) 10 Å from the hydrophobic OMe surface. The effect of increasing temperature is indicated by the color of the lines changing from blue to red and by increasing thickness of the lines, and by the direction of the arrows in b) and c). The arrows in a) indicate values of free energy for which the free energy difference is plotted in Figure 8.9. ΔF_{small} is the free energy change from intermediate (16 Å) to small (9 Å) values of $R_{g,DNA}$, and ΔF_{large} is the free energy change from intermediate to large (23 Å) values of $R_{g,DNA}$. Error bars are calculating using the Monte Carlo bootstrapping analysis method described in the methods section.

To compare the behavior of the compact and extended conformations for ssDNA oligomer in bulk, in Figure 8.9 we present the free energy difference between the free energy at 16 Å (roughly the minimum free energy in all three cases: bulk, OEG and OMe) and (i) the most extended DNA conformation (~23 Å, ΔF_{large}, Figure 8.9a) and (ii) the most compact DNA conformation (~9 Å, ΔF_{small},

Figure 8.9b). The points on the free energy landscape between which we calculate this difference are shown schematically in Figure 8.8a. The free energy difference between the intermediate (16 Å) and the extended (23 Å) conformations (ΔF_{large}) decreases only slightly with temperature, and remains close to a value of 6 kcal/mol in bulk and on both surfaces (Figure 8.9a). However, the free energy difference between the intermediate (16 Å) and the compact (9 Å) conformations (ΔF_{small}) decreases significantly with increasing temperature (Figure 8.9b). In bulk, this free energy difference decreases from approximately 6 kcal/mol to approximately 4 kcal/mol, while on both surfaces this free energy difference decreases from approximately 8 kcal/mol to approximately 4 kcal/mol. Although there are quantitative differences between bulk and the surfaces, the qualitative downward trend in ΔF_{small} with increasing temperature is similar for bulk and the surfaces (Figure 8.9b).

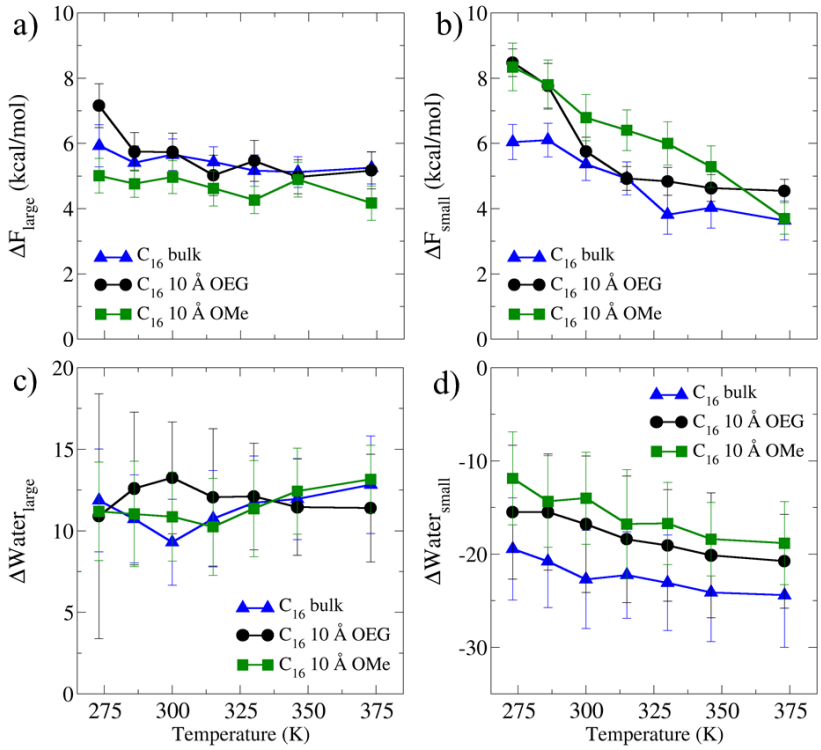


Figure 8.9 Effect of temperature on changes in free energy (a, b) and number of water molecules within 3 Å of the ssDNA (c, d). The changes are between intermediate (16 Å) and large (23 Å) values of $R_{g,DNA}$ (a, c) and between intermediate (16 Å) and small (9 Å) values of $R_{g,DNA}$ (b, d). ΔF indicates a free energy difference, while ΔWater indicates a difference in the number of DNA-bound water molecules. The subscript "large" indicates that the difference is between intermediate (16 Å) and large (23 Å) values of $R_{g,DNA}$, while the subscript "small" indicates that the difference is between intermediate (16 Å) and small (9 Å) values of $R_{g,DNA}$. The uncertainty is propagated as the root sum of the squares of the constituent uncertainties.

The reason for the difference in the temperature dependence of the compact and extended conformations may be the difference in the number of water molecules that are adsorbed to the DNA in different conformations. As temperature increases, the translational and rotational entropy of water molecules becomes increasingly important, but water molecules that are adsorbed to the DNA (e.g. through hydrogen bonds) have substantially reduced entropy. As such, it is reasonable to suggest that DNA conformations that adsorb a greater number of water molecules will be more entropically unfavorable. To relate this idea to the difference in the free energy change between the intermediate and extreme (most compact, most extended) conformations, we quantify the change in the number of DNA-bound water molecules from the intermediate to the extreme conformations as a function of temperature (Figure 8.9c and 8.9d). DNA-bound water molecules are defined as those water molecules within 3 Å of the DNA. The 3 Å cutoff distance is arbitrary but we note that the trends in the change in the number of adsorbed water molecules with temperature is qualitatively similar with a 7 Å cutoff distance, suggesting that this trend is independent of the cutoff distance (Figure 8.20 in supplementary section 8.6). This quantity is the number of water molecules that are adsorbed to or displaced from the DNA during a change in DNA conformation. For the change from intermediate (16 Å) to extended (23 Å) conformations ($\Delta Water_{large}$), additional water molecules adsorb to the ssDNA, as expected, and the number of additional adsorbed water molecules (approximately 12) is essentially independent of temperature (Figure 8.9c). So, irrespective of temperature, the entropic cost of confining additional water molecules near the DNA in an extended conformation is the same. This is consistent with the temperature independence of the free energy change between intermediate and extended conformations (Figure 8.9a).

For the change from intermediate (16 Å) to compact (9 Å) conformations ($\Delta Water_{small}$), water molecules are displaced from the ssDNA oligomer, as expected, but the number of displaced water molecules increases with increasing temperature (Figure 8.9d). At the highest temperature shown here (373 K), approximately 5 additional water molecules are displaced compared to the lowest temperature (273 K). Therefore, the entropically favorable release of water molecules increases with increasing temperature, which is consistent with the decreasing free energy cost of forming a compact ssDNA

conformation (ΔF_{small}, Figure 8.9b). Other contributions to the relative changes in ΔF_{large} and ΔF_{small} with increasing temperature may include the increasing entropic unfavorability of highly extended conformations, which could cause the decrease in ΔF_{large} with increasing temperature to be smaller, and the reduction in base-base stacking with increasing temperature, which could make it easier for the ssDNA oligomer to assume a globular conformation and could thereby allow ΔF_{small} to decrease by a greater extent.

8.4 Conclusion

We used atomistic molecular dynamics simulations, combined with umbrella sampling and temperature-replica exchange to enhance conformational sampling, to study the conformations and interactions of a ssDNA oligomer consisting of 16 cytosine nucleotides (C_{16}) in bulk solution, near a model hydrophilic oligo(ethylene glycol) (OEG) surface, and near a model hydrophobic oligo(methylene) (OMe) surface. We calculated the free energy as a function of ssDNA size, $R_{g,DNA}$. Then, as a function of $R_{g,DNA}$, surface presence and chemistry, and temperature, we quantified the molecular-level structure (e.g., surface area composition), interactions (e.g. DNA-surface hydrogen bonds), and energetics of the ssDNA oligomer.

Although the free energy as a function of $R_{g,DNA}$ is only weakly affected by surface presence and chemistry, slightly smaller ssDNA conformations are favored on the hydrophilic OEG surface relative to the hydrophobic OMe surface. More compact ssDNA conformations concentrate the negative charge of the DNA in a smaller spatial region, which increases the favorable electrostatic interactions between the ssDNA and the polar OEG surface. This increase in ssDNA-OEG interactions appears to be the energetic driving force for the slightly smaller conformations favored by the OEG surface.

The ssDNA surface composition (i.e. the relative exposure of hydrophobic nucleobase and hydrophilic phosphate backbone), which shows some evidence for a micellar conformation in bulk solution, is not significantly affected by the presence of either the hydrophilic or the hydrophobic surface. DNA self-interactions (i.e., intra-DNA hydrogen bonds and base-base stacking) are only weakly affected

by the two surfaces, although it appears that hydrogen bonds between the ssDNA and the hydrophilic OEG surface may disrupt base-base stacking.

Increasing the temperature has qualitatively similar effects in bulk solution and near both the hydrophilic OEG surface and the hydrophobic OMe surface. Increasing temperature decreases base-base stacking and slightly increases the amount of exposed nucleobase surface area. The increase in the exposure of the relatively hydrophobic nucleobases could lead to increased hydrophobic interactions at higher temperatures, although we note limitations in this interpretation below. DNA-water interactions decrease with increasing temperature, likely due to entropic effects. Interestingly, DNA-counterion interactions actually increase with increasing temperature, possibly because of the reduction of DNA-water interactions with increasing temperature and because of the greater electrostatic interaction of counterions with DNA relative to water with DNA. Finally, while increasing temperature flattens the entire free energy landscape as a function of $R_{g,DNA}$, more compact ssDNA conformations see a greater reduction in the free energy than do more extended ssDNA conformations. The reason for this is the entropically favorable release of DNA-bound water molecules, which we find increases with increasing temperature for the more compact ssDNA conformations.

Lastly, we note the limitations to the simulation approach we have used and future directions for this work. One, we have only looked at the behavior of the ssDNA when it is held at 10 Å from the two surfaces. The reason for this was the greatly increased computational expense of simulating additional surface separation distances. Constructing a two-dimensional free energy landscape of DNA conformation and surface distance would be a valuable future direction that could reveal how DNA adsorption and conformation are related. However, obtaining sufficient sampling for such a 2D free energy landscape would be computationally costly, and would likely require the use of methods that improve the efficiency of replica exchange molecular dynamics [47, 48]. Two, the use of temperature-replica exchange subjects the biomolecular force fields to conditions for which they have not been rigorously tested. For example, the TIP3P water model does not reproduce phase behavior of water [49], and the Amber force fields are parameterized at ambient conditions (i.e. 300 K and atmospheric pressure)

[31]. As such, the force fields may not accurately describe the behavior of ssDNA at elevated temperatures. However, it seems reasonable that the force field for ssDNA will still serve as a qualitative model of an amphiphilic polyelectrolyte even at high temperatures, and, in any case, the purpose of the high-temperature replicas is primarily to accelerate conformational sampling in the low-temperature replicas where the force fields are more reliable. Three, the temperature-replica exchange simulations were conducted at constant volume, so the pressure increases with temperature. We used constant volume conditions rather than constant pressure conditions because the high-temperature replicas would require a greatly increased volume to maintain constant atmospheric pressure (i.e. water would vaporize at the highest temperatures employed in the temperature-replica exchange protocol at atmospheric pressure). The large differences in system volume and phase could make it difficult to successfully exchange between high-temperature replicas. However, because the replicas at very high temperatures (> 373 K) are intended only as a means to increase conformation sampling in the lower-temperature replicas, rather than as a means to sample the actual behavior of the ssDNA at the highest temperatures, the use of constant volume conditions seems reasonable. Nonetheless, because the hydrophobic effect is both temperature- and pressure-dependent [45, 46] care must be taken in comparing our results at elevated temperatures and correspondingly elevated pressures with experimental results obtained at atmospheric pressure. It would be worth examining the effects of temperature on ssDNA conformations under constant pressure conditions in future work. Finally, this study is an extension of our previous work on short ssDNA oligomers [27], and to connect to various applications involving double-stranded DNA near surfaces our future work will involve studies of the thermal stability of dsDNA near surfaces of various chemistries.

8.5 References

[1] P. W. K. Rothemund, Nature 440 (2006) 297.
[2] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. M. Shih, Nature 459 (2009) 414.
[3] S. Y. Park, J. S. Lee, D. Georganopoulou, C. A. Mirkin, and G. C. Schatz, Journal of Physical Chemistry B 110 (2006) 12673.
[4] S. Y. Park, K. R. A. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin, Nature 451 (2008) 553.

[5] H. Noh, A. M. Hung, and J. N. Cha, Small 7 (2011) 3021.
[6] A. Striolo, Small 3 (2007) 628.
[7] A. Seifpour, S. R. Dahl, B. Lin, and A. Jayaraman, Molecular Simulation (2013) 1.
[8] D. Gresham, M. J. Dunham, and D. Botstein, Nat Rev Genet 9 (2008) 291.
[9] R. B. Stoughton, Annual Review of Biochemistry 74 (2005) 53.
[10] L. M. Dillenback, G. P. Goodrich, and C. D. Keating, Nano Letters 6 (2005) 16.
[11] C. Chen, W. Wang, J. Ge, and X. S. Zhao, Nucleic Acids Research 37 (2009) 3756.
[12] P. Gong and R. Levicky, Proceedings of the National Academy of Sciences of the United States of America 105 (2008) 5301.
[13] T. W. Cha, V. Boiadjiev, J. Lozano, H. Yang, and X. Y. Zhu, Analytical Biochemistry 311 (2002) 27.
[14] M. Kastantin and D. K. Schwartz, Small 9 (2012) 933.
[15] J. H. Monserud and D. K. Schwartz, Biomacromolecules 13 (2012) 4002.
[16] X. Zhao and J. K. Johnson, Journal of the American Chemical Society 129 (2007) 10438.
[17] R. R. Johnson, A. T. C. Johnson, and M. L. Klein, Nano Letters 8 (2007) 69.
[18] R. R. Johnson, A. Kohlmeyer, A. T. C. Johnson, and M. L. Klein, Nano Letters 9 (2009) 537.
[19] S. Monti and T. R. Walsh, Journal of Physical Chemistry C 115 (2011) 24238.
[20] S. Monti, I. Cacelli, A. Ferretti, G. Prampolini, and V. Barone, Journal of Physical Chemistry B 114 (2010) 8341.
[21] K.-Y. Wong and B. Montgomery Pettitt, Theoretical Chemistry Accounts 106 (2001) 233.
[22] A. Jayaraman, C. K. Hall, and J. Genzer, Biophysical Journal 91 (2006) 2227.
[23] A. Jayaraman, C. K. Hall, and J. Genzer, Journal of Chemical Physics 127 (2007) 144912.
[24] H. Liang, Z. Li, and J. Yang, Physical Chemistry Chemical Physics 12 (2010) 4431.
[25] K.-Y. Wong, A. Vainrub, T. Powdrill, M. Hogan, and B. M. Pettitt, Molecular Simulation 30 (2004) 121.
[26] W. Martin, W. Zhu, and G. Krilov, The Journal of Physical Chemistry B 112 (2008) 16076.
[27] R. M. Elder and A. Jayaraman, Soft Matter 9 (2013) 11521.
[28] Y. Sugita and Y. Okamoto, Chemical Physics Letters 314 (1999) 141.
[29] N. Schwierz, D. Horinek, S. Liese, T. Pirzer, B. N. Balzer, T. Hugel, and R. R. Netz, Journal of the American Chemical Society 134 (2012) 19628.
[30] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco, Biophysical journal 92 (2007) 3817.
[31] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, Journal of the American Chemical Society 117 (1995) 5179.
[32] I. S. Joung and T. E. Cheatham, Journal of Physical Chemistry B 112 (2008) 9020.
[33] G. M. Torrie and J. P. Valleau, Journal of Computational Physics 23 (1977) 187.
[34] B. Tinland, A. Pluen, J. Sturm, and G. Weill, Macromolecules 30 (1997) 5763.
[35] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, Journal of Computational Chemistry 13 (1992) 1011.
[36] A. Grossfield, "WHAM: the weighted histogram analysis method," version 2.0.4, membrane.urmc.rochester.edu/content/wham.
[37] A. Kone and D. A. Kofke, The Journal of Chemical Physics 122 (2005)
[38] N. Rathore, M. Chopra, and J. J. de Pablo, The Journal of Chemical Physics 122 (2005)
[39] D. Frenkel and B. Smit, Understanding Molecular Simulations: From Algorithms to Applications, Academic Press, 2001.
[40] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, Nucleic Acids Research 37 (2009) 5917.
[41] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics and Modelling 14 (1996) 33.

[42] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of Computational Chemistry 26 (2005) 1781.
[43] M. Rubinstein and R. H. Colby, Polymer Physics, Oxford University Press, 2008.
[44] D. Wang, G. Zhu, X. Huang, and S. J. Lippard, Proceedings of the National Academy of Sciences 107 (2010) 9584.
[45] H. S. Chan and K. A. Dill, Proteins: Structure, Function, and Bioinformatics 30 (1998) 2.
[46] G. Hummer, S. Garde, A. E. García, M. E. Paulaitis, and L. R. Pratt, Proceedings of the National Academy of Sciences 95 (1998) 1552.
[47] M. Deighan and J. Pfaendtner, Langmuir 29 (2013) 7999.
[48] A. J. Lee and S. W. Rick, The Journal of Chemical Physics 131 (2009)
[49] C. Vega and J. L. F. Abascal, Physical Chemistry Chemical Physics 13 (2011) 19663.
[50] M. J. Stevens and G. S. Grest, Biointerphases 3 (2008) FC13.
[51] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, Journal of Molecular Graphics and Modelling 25 (2006) 247.
[52] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Journal of Computational Chemistry 25 (2004) 1157.
[53] A. Jakalian, D. B. Jack, and C. I. Bayly, Journal of Computational Chemistry 23 (2002) 1623.
[54] A. Ulman, Chemical Reviews 96 (1996) 1533.

8.6 Supplementary Information

8.6.1 Additional details of surface construction and equilibration

The oligomers comprising the SAMs (either $H(OCH_2CH_2)_5OH$ (hydroxyl-terminated OEG) or $CH_3(CH_2)_{14}CH_3$ (methyl-terminated OMe)) are initially arranged in a hexagonal grid consisting of 256 oligomers in a cross-sectional area of approximately 6.5 nm by 7 nm, yielding a grafting density of approximately 5 chains / nm^2, which is comparable to a functionalized silica or gold surface [50]. The oligomers are constructed using VegaZZ, and the *antechamber* program of the Amber suite is used to assign general Amber force field (*gaff*) atom types and atomic partial charges using the AM1-BCC charge method [51-53]. The SAMs are constructed in the xy-plane such that they are perpendicular to the z-axis. After constructing the SAMs from the oligomers using VMD, we conducted 1000 steps of conjugate gradient minimization using NAMD to eliminate any unfavorable atomic contacts [41, 42]. Then, we used the following protocol to further relax the surfaces into their preferred arrangement. The coordinates of the energy minimized surfaces were solvated with TIP3P water molecules to create a 20 Å thick layer on both the +z and –z sides of the surface. It is convenient to place water molecules on both sides of the surface when using the LEaP program of the Amber suite, and the periodic boundary condition in the z-direction ensures that this is equivalent to having water molecules on only one side (top or bottom) of the surface. Then, to further relax the surfaces, we conducted a series of MD simulations. The initial size of the periodic box for these simulations was carefully chosen to ensure that the periodic images of the surfaces were in close contact, thereby eliminating any gaps in the surfaces and creating an infinite surface when combined with periodic boundary conditions in the x- and y-directions, yielding a cross-sectional area of approximately 6.5 nm by 7 nm. During these simulations, the bottommost heavy atom of each oligomer was constrained to a constant z-value to prevent dissolution of the monolayer while still allowing lateral diffusion and rearrangement in the xy-plane. This constraint on the oligomers was applied with the *collective variables* module of NAMD and a force constant of 2.5 kcal/mol-Å was used. First, we conducted constant temperature (300 K), volume, and number of particles (NVT) MD for 100 ps to allow

the water molecules to equilibrate. Next, we conducted constant temperature, constant pressure (1 atm), and constant number of particles (NPT) MD for 100 ps with constant xy-area to allow the pressure to equilibrate while maintaining constant grafting density. Finally, we used the coordinates of these relaxed surfaces – with a predetermined grafting density and now equilibrated in the presence of liquid water – for the remainder of our simulations.

8.6.2 Additional details of system set-up and equilibration

To prepare systems consisting of DNA, water, counterions, and the SAM, we place the ssDNA oligomer, which consists of 16 repeat units of cytosine (C_{16}), 10 Å from the surface, add neutralizing counterions (15 Na^+ ions) but no additional salt, and solvate with a sufficiently thick layer of TIP3P water molecules (placed only in the z-direction above and below the SAM) to yield a simulation box height of approximately 80 Å, of which the surface occupies approximately 15-20 Å. This large water box ensures that the DNA can move large distances in the z-direction before interacting with the periodic image of the surface. After constructing these systems, we perform a minimization and equilibration procedure that is described below prior to conducting production umbrella sampling simulations. During all production simulations, both the topmost and bottommost heavy atoms of the oligomers comprising the surface are constrained to constant z-values with harmonic restraints, using the *collective variables* module of NAMD and a force coefficient of 2.5 kcal/mol-$Å^2$ [42]. The distance between the topmost and bottommost heavy atoms was 14 Å for OEG and 19 Å for OMe, which we determined to be the preferred brush heights by simulating the surfaces with the bottommost heavy atoms restrained to a constant z-value but the topmost heavy atoms unrestrained. In these simulations, the OEG oligomers assume a helical conformation with a height of approximately 14 Å while the OMe oligomers assume an extended all-*trans* conformation with a height of approximately 19 Å, which are reasonable conformations and brush heights for these oligomers [50, 54]. The reasons for constraining the top and bottom of the oligomers comprising the surface are as follows: the bottom of each oligomer is constrained to mimic attachment to a solid surface, while the top of each oligomer is constrained to maintain a similar stiffness and surface roughness between OEG and OMe and to prevent dissolution of OEG into the bulk water.

After constructing fully-solvated systems including ssDNA, neutralizing Na^+ counterions, the self-assembled monolayer (SAM), we used the following procedure to equilibrate the systems prior to use in production simulations. First, in constant volume and temperature conditions (NVT ensemble), the water and ions were minimized with the conjugate gradient algorithm for 2000 steps, with the DNA and SAM subjected to a 500 kcal/mol-A^2 harmonic restraining potential. Second, in the NVT ensemble, with the DNA/SAM restraining potential reduced to 100 kcal/mol-A^2, the system was heated from 0 to 300 K over 20 ps. Third, in constant pressure and temperature conditions (NPT ensemble) and with the restraint reduced to 50 kcal/mol-A^2 the system was relaxed for another 20 ps. Fourth, in the NVT ensemble, the restraining potential was reduced in three steps (50, 10, and 5 kcal/mol-A^2) and minimized for 2000 steps each. Fifth, in the NVT ensemble with the same 5 kcal/mol-A^2 restraint the system was heated from 10 to 300 K over 20 ps. Sixth, in the NPT ensemble, the restraint was reduced from 5 to 1 to 0.1 to 0 kcal/mol-A^2 over 20 ps. Seventh, in the NVT ensemble with no restraints, the system was heated from 10 to 300 K over 20 ps.

8.6.3 Convergence and sampling of umbrella sampling temperature-replica exchange simulations

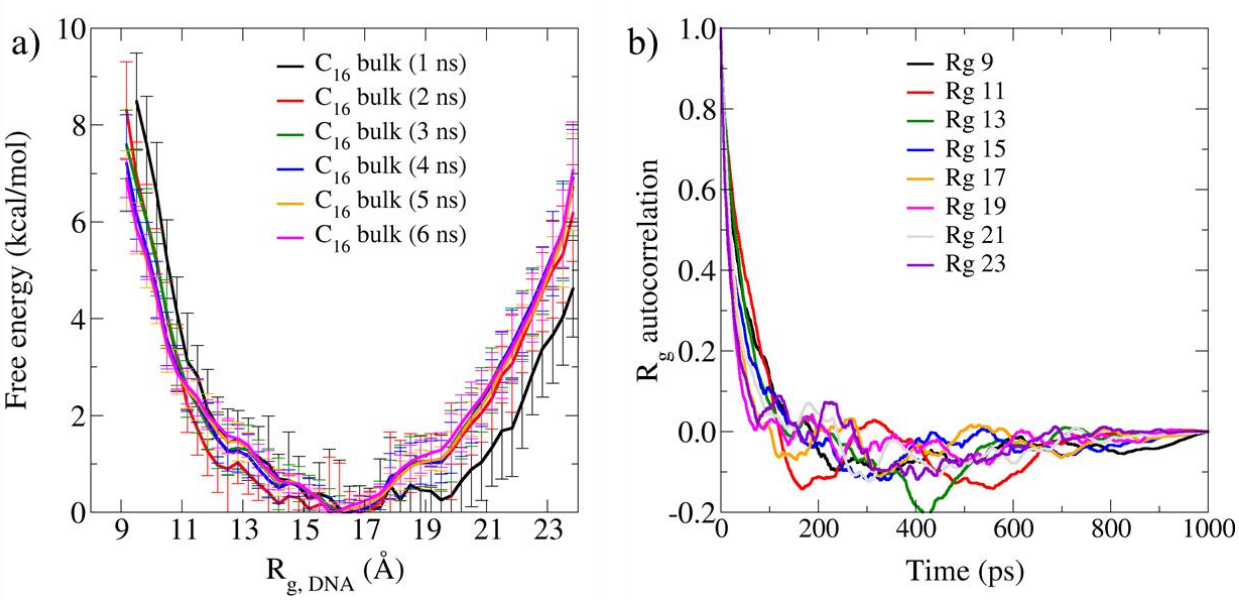


Figure 8.10 a) Convergence of the free energy as a function of $R_{g,DNA}$ with increasing length of simulation time in bulk solution. Although the majority of the free energy profile does not significantly change after 3 ns of sampling, the value of free energy at the lowest values of $R_{g,DNA}$ continues to decrease up to 5 ns of sampling, but appears stable by 6 ns of sampling (i.e. the free energy appears to have converged with 6 ns of sampling). Note that the sampling times indicated exclude a 1 ns equilibration period; there are 7 ns of simulation in each umbrella sampling window. b) Autocorrelation function of $R_{g,DNA}$ in one replica in all of the umbrella sampling windows. Note that this plot shows results for a particular *replica*, not a particular *temperature*, so the autocorrelation decays due to ssDNA conformational changes, not due to coordinate swaps between replicas. If this analysis is performed for a given temperature, the autocorrelation function decays much more quickly because of coordinate swaps. In all cases, the autocorrelation function reaches zero and begins to oscillate around zero at approximately 200 ps.

8.6.4 Total ssDNA surface area as a function of ssDNA size

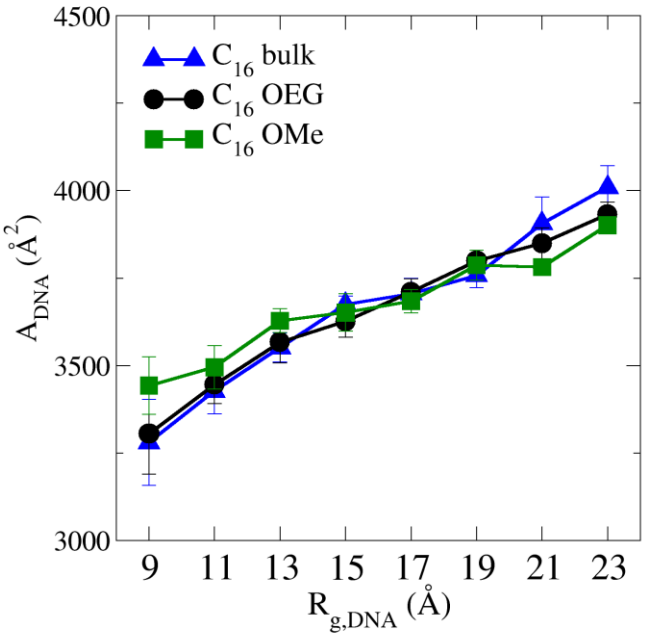


Figure 8.11 The total surface area of the C_{16} ssDNA oligomer decreases with decreasing $R_{g,DNA}$. The total DNA surface area is not significantly affected by the presence or the chemistry of the surfaces.

8.6.5 Number of water molecules near the ssDNA oligomer

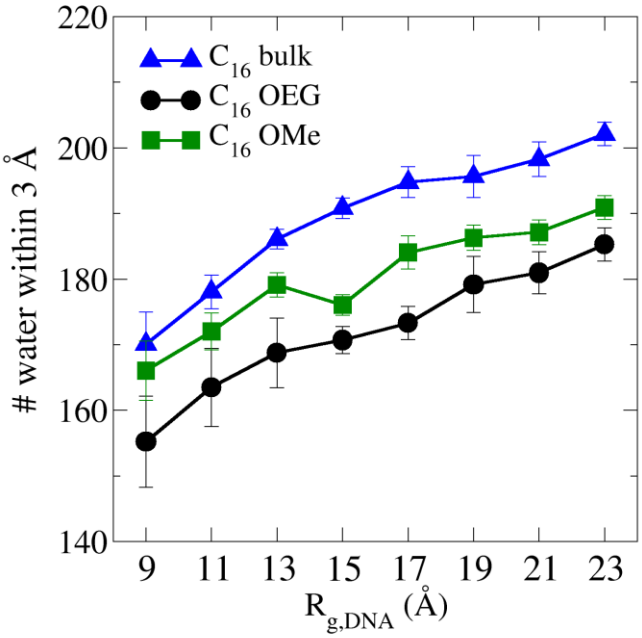


Figure 8.12 The number of water molecules within 3 Å of any DNA atom as a function of $R_{g,DNA}$ in bulk solution, 10 Å from the OEG surface, and 10 Å from the OMe surface. The number of water molecules decreases with decreasing $R_{g,DNA}$ because there is less DNA surface area available for the water molecules

to interact with when DNA is in a globular conformation. The number of water molecules near the DNA when the DNA is close to either surface is lower than when the DNA is in bulk because the surfaces displace water molecules from the vicinity of the ssDNA molecule. The number of DNA-bound water molecules is lowest when the DNA is near OEG because the hydrophilic OEG surface interacts strongly with the ssDNA (e.g. through hydrogen bonds), which further decreases the available surface area for DNA-water interactions.

8.6.6 Total potential energy, free energy, and entropy

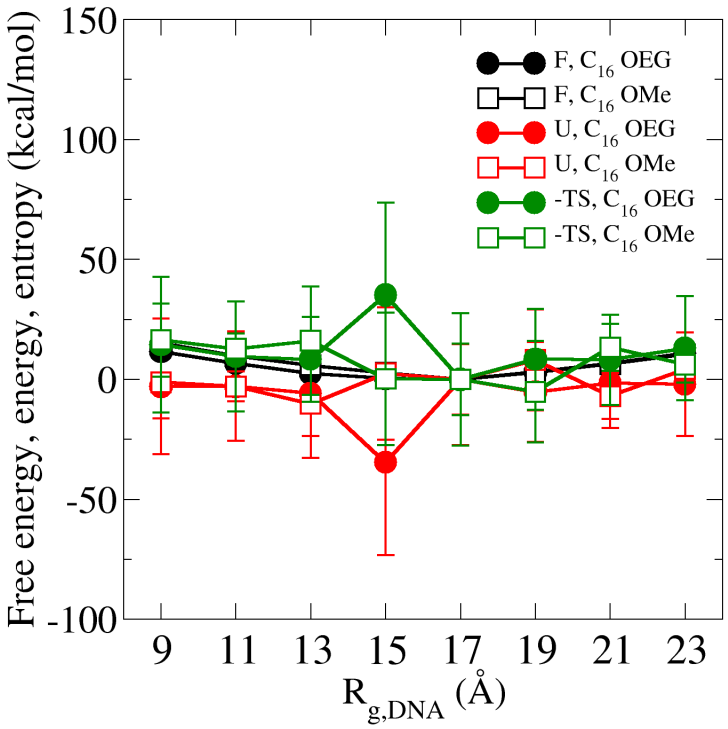


Figure 8.13 Free energy calculated using umbrella sampling, total potential energy of the system, and entropy. The entropy is calculated as the difference between the free energy and the total potential energy. The uncertainty in the total potential energy is too large to draw any conclusions about the effect of ssDNA conformation on the total potential energy or, consequently, about the total entropy of the system.

8.6.7 Additional details of the effect of temperature on ssDNA structure

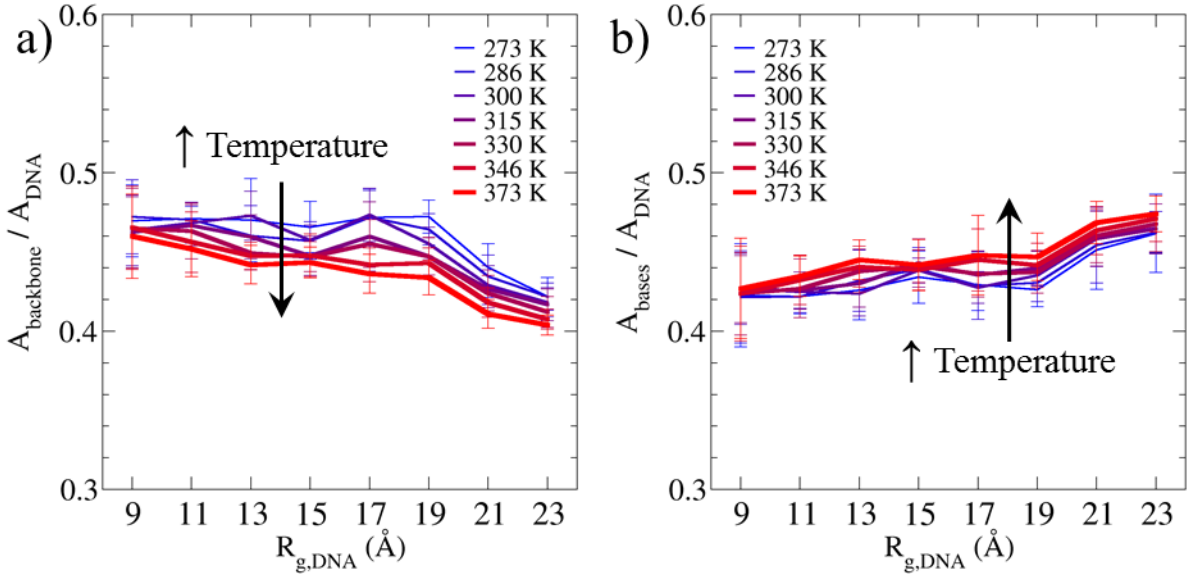


Figure 8.14 The surface area composition of a) the backbone and b) the nucleobases of the ssDNA *in bulk solution*. The fraction of the surface area that is composed of backbone atoms decreases with increasing temperature while the fraction of the surface area that is composed of nucleobase atoms increases slightly with increasing temperature.

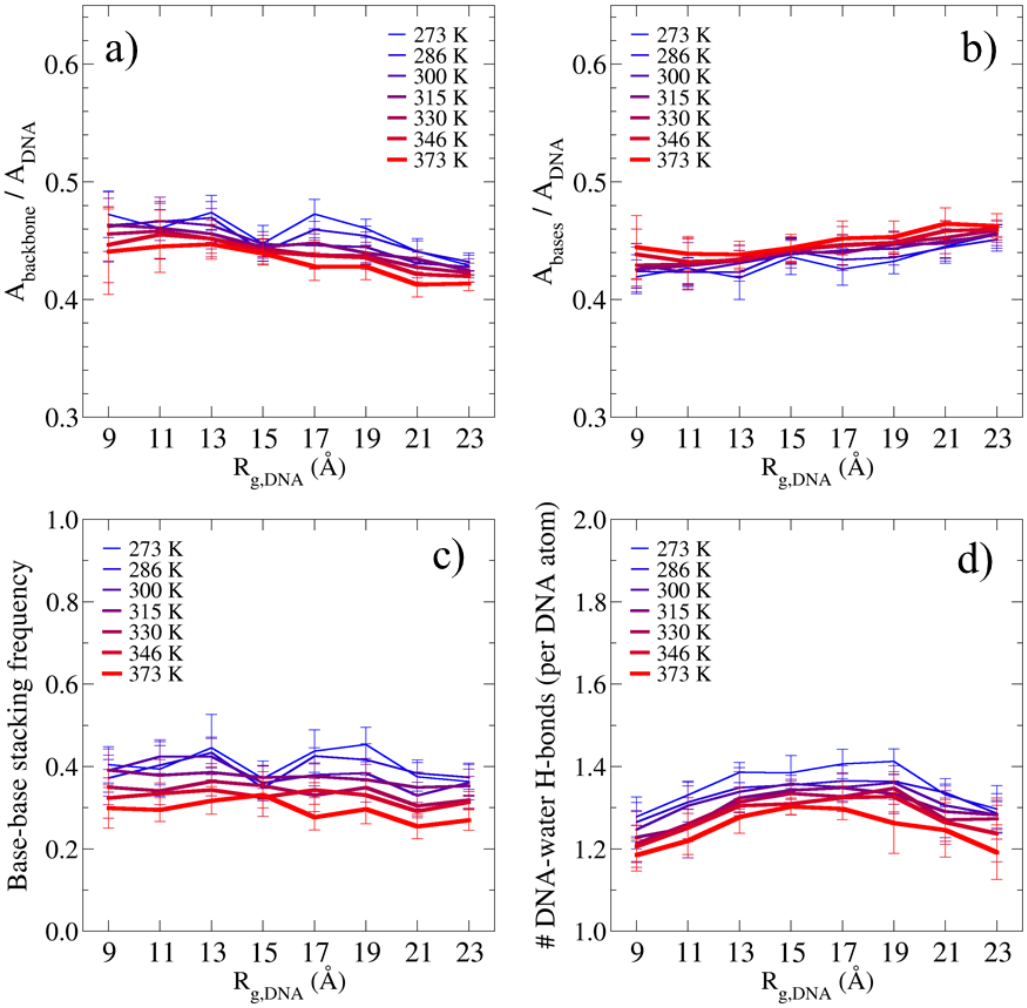


Figure 8.15 Several aspects of the structure of the C16 ssDNA oligomer *10 Å from the hydrophilic OEG surface*. Similar to what occurs in bulk solution, with increasing temperature the DNA shows a) decreased backbone surface area, b) slightly increased nucleobase surface area, c) decreased base-base stacking, and d) decreased DNA-water hydrogen bonds.

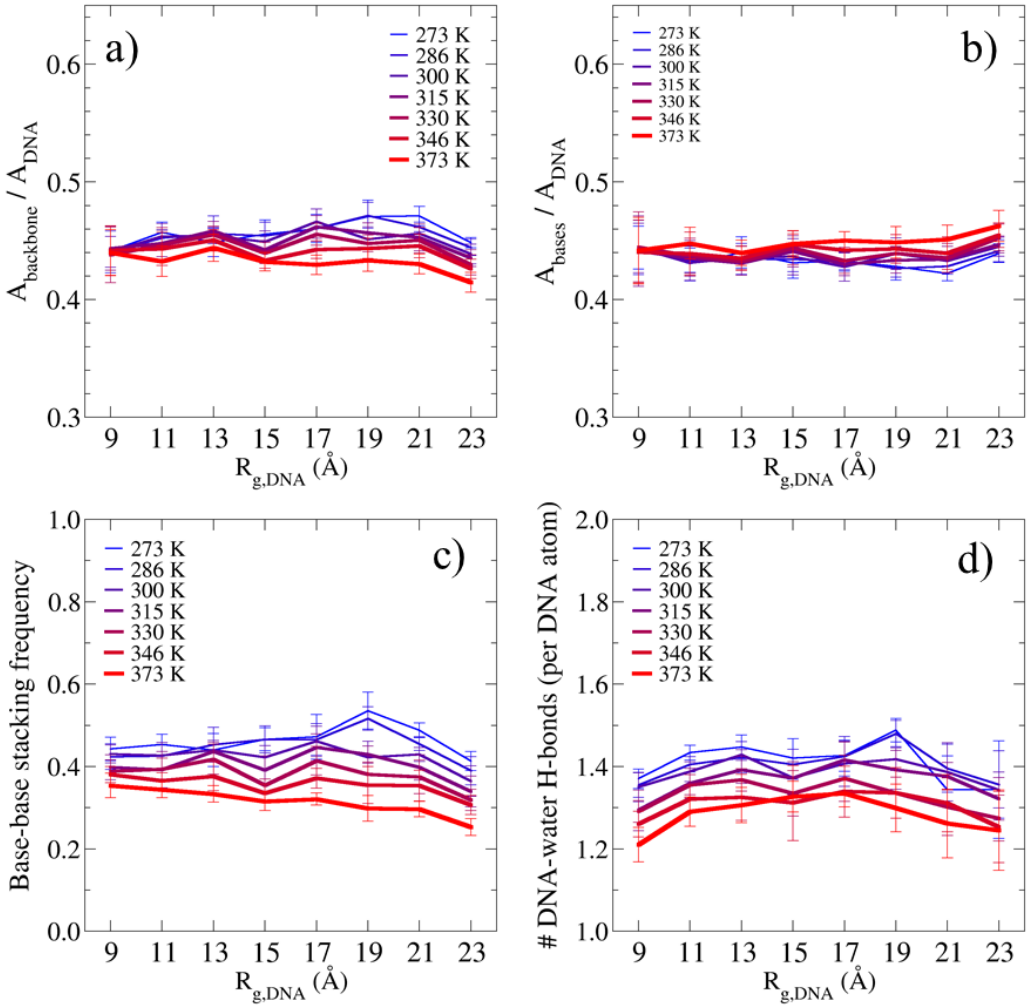


Figure 8.16 Several aspects of the structure of the C16 ssDNA oligomer *10 Å from the hydrophobic OMe surface*. Similar to what occurs in bulk solution, with increasing temperature the DNA shows a) decreased backbone surface area, b) slightly increased nucleobase surface area, c) decreased base-base stacking, and d) decreased DNA-water hydrogen bonds.

8.6.8 Effect of temperature on ssDNA energetic interactions in bulk and on OMe surface

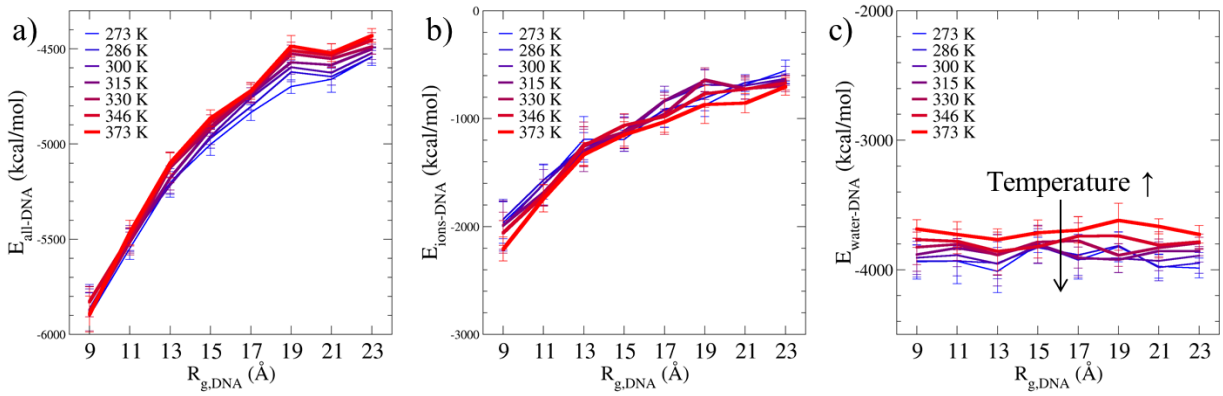


Figure 8.17 Analogous to Figure 8.7, which shows the energetic interactions of the C_{16} ssDNA oligomer near the OEG surface, this figure shows the energetic interactions of the ssDNA *in bulk solution*. Similar to the trends observed on the OEG surface, with increasing temperature a) the total energetic interaction of the ssDNA with everything else in the system is mostly unchanged, although it decreases slightly, b) the energetic interaction with counterions becomes slightly more favorable, and c) the energetic interaction with water becomes less favorable.

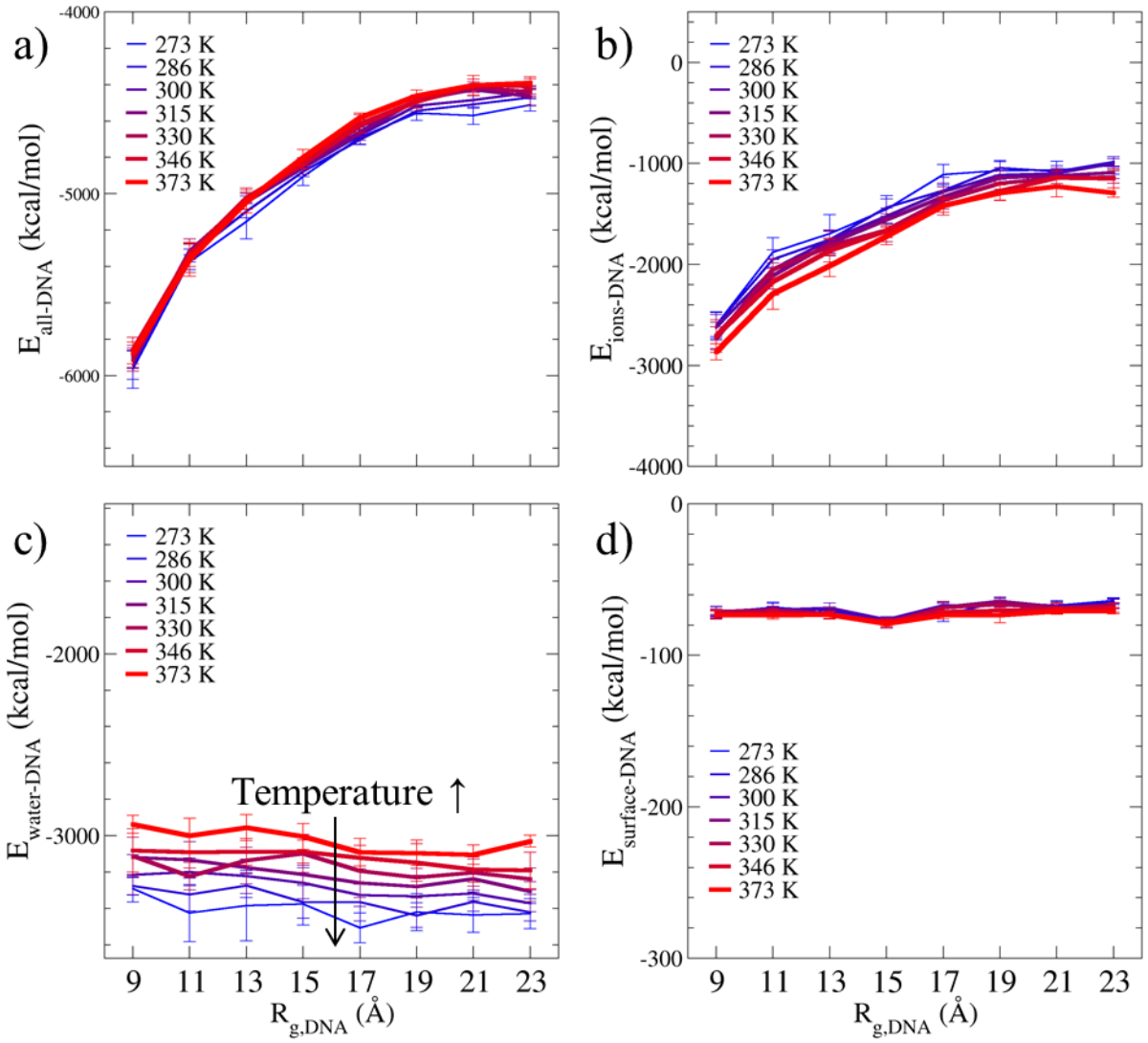


Figure 8.18 Analogous to Figure 8.7, which shows the energetic interactions of the C_{16} ssDNA oligomer near the OEG surface, this figure shows the energetic interactions of the ssDNA *10 Å from the hydrophobic OMe surface*. Similar to the trends observed on the OEG surface, with increasing temperature a) the total energetic interaction of the ssDNA with everything else in the system is mostly unchanged, although it decreases slightly, b) the energetic interaction with counterions becomes slightly more favorable, and c) the energetic interaction with water becomes less favorable. As discussed in the main text, the energetic interaction of the ssDNA with the OMe surface is unaffected by temperature.

8.6.9 No effect of temperature on DNA-OEG hydrogen bonds

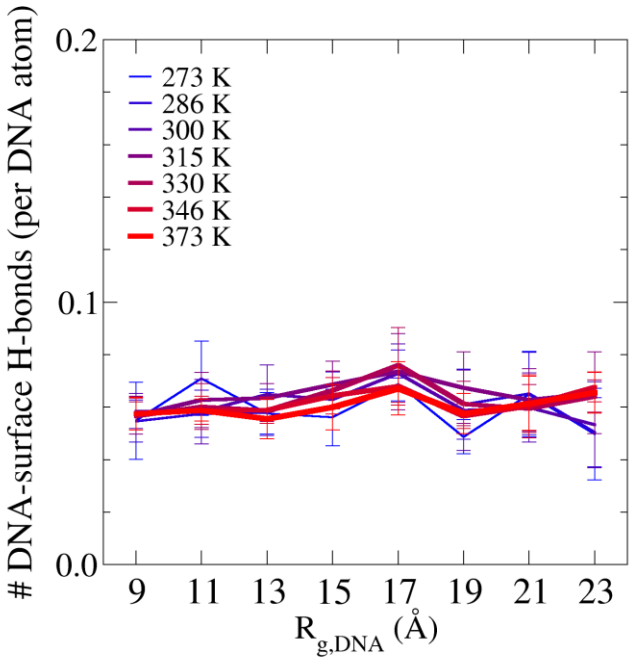


Figure 8.19 The number of DNA-surface hydrogen bonds near the OEG surface is unaffected by temperature or by $R_{g,DNA}$. The fact that the number of DNA-OEG hydrogen bonds does not reflect the increase in DNA-surface interaction energy with increasing temperature (i.e. higher temperature makes the DNA-surface interaction energy less favorable) may be due to the imprecise nature of the geometric criteria used to define a hydrogen bond.

8.6.10 Number of water molecules within 7 Å of DNA as a function of temperature

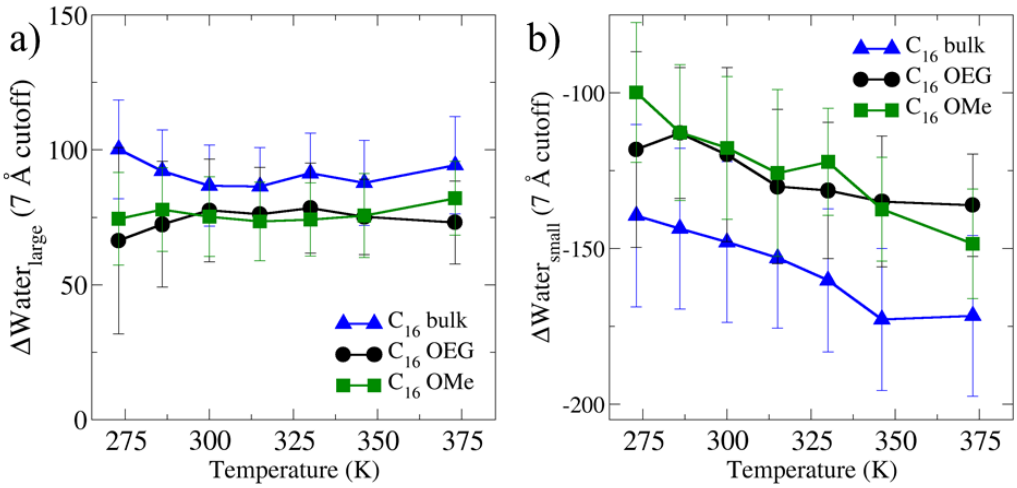


Figure 8.20 The change in the number of water molecules within 7 Å of any DNA atom between intermediate $R_{g,DNA}$ (16 Å) and either small $R_{g,DNA}$ (9 Å, $\Delta Water_{small}$) or large $R_{g,DNA}$ (23 Å, $\Delta Water_{large}$). As with the cutoff distance of 3 Å (Figure 8.9), $\Delta Water_{large}$ does not change with temperature but $\Delta Water_{small}$ becomes increasingly negative with temperature, demonstrating that these results are independent of the arbitrary cutoff distance.

Chapter 9

Conclusions and Future Directions

9.1 Protein-DNA interactions

Summary of work in this thesis. In the first section of this thesis (chapters 2 and 3), we studied the ability of the DNA damage recognition protein HMGB1a to recognize anticancer drug-DNA damage. Our goal was to explain experimentally observed differences in the drug-DNA binding affinity of the protein HMGB1a for two anticancer drugs, cisplatin and oxaliplatin, covalently attached to DNA in several DNA sequence contexts. In the first study (chapter 2), using atomistic molecular dynamics we elucidated the structure and conformational dynamics of the drug-DNA molecules in the presence and absence of HMGB1a and in multiple sequence contexts. In the presence of HMGB1a, the drug-DNA structure was independent of the drug identity and the sequence context, leading us to suggest that differences in the drug-DNA molecule in the absence of the protein are the cause of differences in the binding affinity of HMGB1a. In the absence of HMGB1a, we found drug- and sequence-dependent differences in drug-DNA structure that allowed us to rationalize some of the observed differences in binding affinity. In particular, oxaliplatin in the TGGA sequence context, for which HMGB1a has negligible binding affinity [1], displayed structural characteristics consistent with extremely poor recognition by the protein HMGB1a. Based on this first study, we hypothesized that the deformability of the drug-DNA molecule may be an additional factor affecting HMGB1a binding affinity. We addressed this hypothesis in the second study (chapter 3) by calculating the free energy required to deform the drug-DNA molecule from the structure in the absence of HMGB1a to the structure in the presence of the protein, which may be thought of as the work the protein must exert to deform the drug-DNA molecule during binding. The results of this study suggest that the greater free energy penalty for deforming oxaliplatin in the TGGA context (compared to cisplatin-DNA molecules and oxaliplatin in other sequence contexts) is partly responsible for the

extremely low binding affinity of HMGB1a for oxaliplatin in the TGGA context. The mechanistic reason for the sequence-dependent deformation free energy of oxaliplatin-DNA adducts appears to be steric hindrance by nearby thymine methyl groups, the locations of which depend on the sequence context. Recent efforts focused on steric hindrance between the drug and DNA [2-5] and our work in these studies support the idea of designing novel drugs that exploit steric hindrance to obstruct the binding of damage recognition proteins.

Limitations. The primary limitation in these studies of drug-DNA adducts is the force field for the platinum-based drugs, cisplatin and oxaliplatin. In particular, a platinum atom has a complex electronic structure that is poorly represented by a static point charge and a van der Waals sphere. To partially overcome this limitation, in the first study we made modifications to the published force fields for these drugs to better match experimental structures that had become available since the initial publication of the force fields. While the modified force fields are undoubtedly still quantitatively inaccurate, as with all such approximate models, our qualitative observations on the role of steric hindrance are supported by recent experimental studies [2-5], which suggests that these force fields are reasonable models. Several other limitations are discussed in detail in chapter 3.

Future directions. These studies were focused on the recognition of drug-DNA damage caused by the first-generation platinum-based anticancer drugs, cisplatin and oxaliplatin, by a single DNA damage recognition protein, HMGB1a. As next-generation anticancer drugs are currently in development, it would be valuable to study how these novel drugs affect DNA structure and flexibility to understand their improved efficacy. The monofunctional platinum-based drug phenanthriplatin [5, 6], which was designed to induce less deformation of the DNA structure and increase steric hindrance between the drug and DNA compared to the first-generation drugs, is an excellent starting point for additional studies. Subsequently, the tricyclic phenanthridine moiety of phenanthriplatin could be systematically modified by rearrangement of the existing cyclic groups or by including additional bulky substituent groups, such as methyl or cyclic groups, to determine the effect of these modified drugs on DNA deformability. One could also examine the interactions of other relevant proteins with the first- and next-generation

anticancer drugs. One of the primary targets for the next-generation drugs is RNA polymerase II, a major component of the intracellular transcription machinery that is partially impeded by the presence of drug-DNA damage. An X-ray crystal structure of this polymerase bound at the site of next-generation drug-DNA damage is available [7], which could serve as an initial configuration for molecular simulation study aimed at understanding how existing and putative next-generation drugs affect polymerase activity [6].

9.2 Polycation-DNA interactions

Summary of work in this thesis. In the second section of this thesis (chapters 4 through 6), we described our studies of polycations with varying architecture (i.e., linear and grafted), chemistry (i.e., lysine-based and oligopeptide-based), and sequence (i.e., ordering of peptides within the grafts) binding to DNA. Our overarching goal was to explain experimental findings showing that architecture, chemistry, and sequence affect polycation-DNA binding and DNA transfection efficiency, and through our computational understanding provide guidelines for future synthesis of new and improved polycations with optimal DNA binding and transfection. In chapter 4, we used atomistic molecular dynamics simulations combined with free energy calculations to examine the differences in polycation-DNA binding caused by rearranging a linear polylysine-based architecture into a grafted oligolysine-based architecture. Experimental results show that the linear architecture binds to DNA more strongly than the grafted architecture [8], a result we were able to qualitatively reproduce. We rationalized this finding on the basis that the linear architecture binds in a concerted fashion while the oligolysine grafts in the grafted architecture bind independently, leading to weaker binding and, in turn, improved DNA release within the cell and improved transfection efficiency for the grafted architecture. Additionally, we found that the experimentally observed non-monotonic trend in the binding strength with increasing graft length in the grafted architecture is caused by non-monotonic contributions to the entropy of binding. In chapter 5, we developed a coarse-grained (CG) model that allowed us to simulate greater length and time scales than the atomistic models used in chapter 4. We used this CG model to study the effect of polycation architecture

on the properties of the polycation, such as flexibility, and on the structure of complexes formed between the CG polycations and a DNA-like linear polyanion. We found that increasing the linear charge density of the grafted polycation (i.e., by increasing graft length or decreasing graft spacing) decreased the flexibility of the polycation, which could decrease the strength of polycation-DNA binding since a less-flexible polycation may be less able to conform to the local structure of DNA. However, increasing the linear charge density of the polycation should also increase the strength of polycation-DNA binding due to increased electrostatic attraction. We therefore suggest that these opposing trends with increasing graft length – decreasing polycation flexibility and increasing electrostatic attraction – may be the cause of the non-monotonic trend in the polycation-DNA binding strength observed both in our simulations (chapter 4) and in experiments [8]. Regarding the structure of polycation-DNA complexes (polyplexes), these CG studies suggested that increasing the linear charge density of the polycation increases the surface charge density of the polyplex, which may provide a way to tune the electrostatic interactions of polyplexes with the cell membrane. In the final study in this section (chapter 6), we transition from studying the effect of polycation *architecture* to studying the effect of polycation *chemistry* and *sequence* on polycation-DNA binding. Rather than a pure oligolysine-based grafted polycation, we began with a nuclear localization sequence (NLS, peptide sequence PKKKRKV) oligopeptide, systematically varied the chemistry and sequence of the oligopeptide grafts, and used atomistic molecular dynamics simulations to examine polycation-DNA binding. We found that replacing arginine (R) with lysine (K) reduces the strength of polycation-DNA binding, but that the position of arginine within the grafts had little effect on the strength of polycation-DNA binding. Changing the positions of the hydrophobic residues proline (P) and valine (V) systematically changed hydrophobic aggregation within the polycation, which we rationalized by considering molecular-level properties of proline and valine (i.e., hydrophobicity and rigidity). Because differences in hydrophobic aggregation led to differences in the loss of conformational entropy that occurs upon polycation-DNA binding, we hypothesized that the strategic inclusion of hydrophobic groups in polycations may be used to tailor the strength of polycation-DNA binding.

Limitations. One major limitation of our simulation approach is that we can only model the extracellular environment, not the complex, protein-rich intracellular environment, and therefore we can only draw conclusions about polycation-DNA binding outside the cell rather than overall transfection efficiency. However, because polycation-DNA binding strength is a major contributor to overall transfection efficiency, we can infer how our findings on polycation-DNA binding might affect transfection [9, 10]. As addressed in chapter 4, other limitations of our atomistic approach include the small size of the polycation and DNA that can feasibly be simulated with atomistic detail, and difficulties in sampling the multitude of polycation-DNA binding modes due to the essentially irreversible binding that occurs over a simulation of reasonable duration. These two limitations are partially overcome, respectively, by employing the coarse-grained model in chapter 5 and in ongoing studies, and by conducting multiple independent atomistic simulations of polycation-DNA binding to improve the sampling of different binding modes.

Future directions. In these studies we have examined how several features of polycation architecture, chemistry, and sequence affect polycation-DNA binding, but many other features of polycations remain to be investigated. Given the success of the grafted architecture as a transfection agent [8, 11], the most profitable route of inquiry will likely be further examination of the role of peptide chemistry and sequence within the grafted architecture. First, the hypotheses regarding hydrophobic aggregation set forth in chapter 6 could be tested using atomistic simulations by mutating proline and valine to other hydrophobic peptides (e.g. alanine, leucine, isoleucine) to see if these mutations have the expected effect on hydrophobic aggregation. Synthetic peptides with even greater hydrophobicity (e.g. polycyclic aromatic groups) could also be investigated to gain a deeper understanding of the effects of hydrophobicity on polycation-DNA binding. Second, the structure of polyplexes formed with NLS-based polycations could be investigated using a coarse-grained model that includes arginine, proline, and valine. Furthermore, coarse-grained simulations of the grafted polycations, both lysine- and NLS-based, with multiple short oligoanions, rather than a single long polyanion, could be used to study the structure of polyplexes formed with small interfering RNA (siRNA), which are relatively short dsRNA molecules (approximately

20 base pairs) that show great promise for gene therapy [9]. The accuracy of the coarse-grained simulations could be improved by implementing a more realistic model of DNA, such as the "three-site-per-nucleotide" (3SPN) model [12, 13]. Coarse-grained simulations of multiple polyplexes, formed with polycations of varying architecture and chemistry, could be used to examine the aggregation of polyplexes, which is undesirable in a therapeutic context. Finally, the effect of the chemical nature of the grafted backbone could be considered at both the atomistic and coarse-grained levels by using a hydrophilic poly(ethylene glycol) (PEG) backbone instead of the hydrophobic poly(cyclooctene) backbone used in these studies. A PEG backbone is of particular interest because of experimental efforts to increase the stability and biocompatibility of polyplexes by incorporating PEG chains [14, 15].

9.3 Surface-DNA interactions

Summary of work in this thesis. In the third section of this thesis (chapters 7 and 8), we studied how surface chemistry affects the adsorption and stability DNA. First, we constructed hydrophilic and hydrophobic self-assembled monolayers to serve as model surfaces with differing chemistry. Then, in chapter 7, we used atomistic molecular dynamics simulations combined with an enhanced sampling technique to calculate the free energy of adsorption of short single-stranded DNA (ssDNA) oligomers of adenine and cytosine to the model surfaces. This simulation approach allowed us to simultaneously observe the behaviors and molecular-level interactions of the DNA, the surface, and water molecules. The amphiphilic ssDNA oligomers adsorb to both hydrophobic and hydrophilic surfaces with roughly similar strength. Both adenine (a purine) and cytosine (a pyrimidine) adsorb to the hydrophilic surface with equal strength because these two nucleobases form roughly the same number of hydrogen bonds with the surface. However, because adenine is more hydrophobic than cytosine, adenine forms longer-lasting hydrophobic interactions ("face-on" conformations) with the hydrophobic surface than cytosine does, and consequently adenine adsorbs to the hydrophobic surface with greater strength than cytosine. Differences in the molecular-level DNA-surface interactions lead to surface-dependent differences in the behavior of

adsorbed ssDNA. DNA-surface hydrogen bonds with the hydrophilic surface, which are short-range directional interactions, cause the ssDNA to diffuse more slowly when adsorbed to the hydrophilic surface. Conversely, the ssDNA can easily maintain face-on conformations with the hydrophobic surface during lateral motion, allowing rapid diffusion when adsorbed to the hydrophobic surface. Differences in the hydration pattern of the hydrophilic and hydrophobic surfaces lead to different DNA-water interactions during adsorption. The water near the hydrophilic surface, being dense and strongly adsorbed to the surface, exerts a strong repulsive force on the adsorbing DNA. In contrast, the water near the hydrophobic surface plays a smaller role in adsorption, but this water can exert a weakly attractive force (due to the hydrophobic effect) and, surprisingly, a weakly repulsive force (due to a region of low-density water above the hydrophobic surface) depending on the location and length of the ssDNA oligomer. Finally, we observed that the short ssDNA oligomers assume a conformation perpendicular to the surface during adsorption, which suggests that adsorption events, where the ssDNA bases are aligned perpendicular to the surface, may be associated with hybridization events. Many of these findings for the adsorption of ssDNA oligomers to hydrophilic and hydrophobic surfaces may also apply more generally to other amphiphilic polymers adsorbing to surfaces of varying hydrophobicity. In chapter 8, we present our initial studies on the effects of surface chemistry on the conformations of longer ssDNA oligomers. We used atomistic molecular dynamics, combined with umbrella sampling and replica exchange to enhance conformational sampling, to calculate the free energy profile of DNA size in bulk solution and near the same model surfaces as in chapter 7. Although the surfaces have relatively little effect on the size of the DNA, the hydrophilic surface slightly favors more compact DNA conformations compared to the hydrophobic surface, which appears to be caused by stronger DNA-surface interactions with the hydrophilic surface when the DNA is in a compact conformation. Using the replica exchange methodology also allowed us to extract information about the temperature dependence of DNA conformations. Increased temperature favors more compact ssDNA conformations relative to more extended conformations, which appears to be related to the entropically favorable release of DNA-bound water molecules from the more compact DNA molecules. These studies represent promising first steps in

understanding the molecular-level interactions of DNA molecules with surfaces, and in understanding how these interactions affect the adsorption and hybridization of DNA molecules near surfaces.

Limitations. In these studies, we have simulated self-assembled monolayers with higher grafting density than is typically obtained on surfaces used in experiments. While our intention with this choice was to create ideal hydrophobic and hydrophilic interfacial environments, such high grafting density eliminates the possibility of DNA intercalation into the SAM, which may be important for reproducing experimentally observed phenomena. Studies investigating the prevalence and effects of DNA intercalation into less-dense SAMs are warranted. Additionally, the hydrophilic and hydrophobic SAMs we have studied have subtly different mechanical properties due to their chemical structures. The hydrophilic oligo(ethylene glycol) molecules have fewer hydrogen atoms than the hydrophobic oligo(methylene) molecules, creating a more crowded environment within the hydrophobic SAM and hence causing the hydrophobic SAM to be more rigid than the hydrophilic SAM. It is worth studying whether this difference in mechanical properties significantly affects DNA adsorption, which might be accomplished by using a less-densely grafted hydrophobic SAM to match the mechanical properties of the hydrophilic SAM, or by using oligomers with the same chemical structure except for the head group functionalization (e.g. instead of oligo(ethylene glycol), use hydroxyl-terminated oligoethylene). While we have employed several computational techniques to enhance conformational sampling, recently developed methods hold great promise for improving both the quality and computational efficiency of sampling, such as parallel tempering metadynamics in the well-tempered ensemble [16]. Finally, in our replica exchange simulations we have used constant volume conditions rather than constant pressure conditions, which lead to elevated pressures at elevated temperatures. Because the hydrophobic effect is both temperature- and pressure-dependent, caution must be used when comparing our results with experimental measurements taken at constant pressure. It would be worth studying the effects of temperature on DNA conformation at constant pressure, both out of fundamental interest and to allow a fairer comparison between simulations and experiments.

Future directions. In these studies, we have considered only two surface chemistries: hydrogen-bonding hydrophilic oligo(ethylene glycol) and non-hydrogen bonding hydrophobic oligo(methylene). A natural extension of the current work is to study other surface chemistries with differing hydrophobicity and hydrogen-bonding ability, such as non-hydrogen-bonding hydrophilic methoxy-terminated oligomers, more-hydro*philic* amide-terminated oligomers, or more-hydro*phobic* trifluoromethyl-terminated oligomers. Studying these additional surface chemistries might help generalize our findings regarding ssDNA adsorption, as has been done to study the behavior of water near surfaces of varying hydrophobicity [17]. Additionally, based on experimental observations that the balance of hydrogen bond donors versus acceptors affects protein adsorption [18], modifying the hydrophilic surface from hydroxyl-terminated (hydrogen bond donating) to aldehyde-terminated (hydrogen bond accepting), or even methoxy-terminated (non-hydrogen bonding), could be worthwhile. Another modification of the SAMs that does not involve chemistry is to decrease the grafting density, which may allow DNA intercalation into the surface and lead to substantial changes in DNA behavior. For example, intercalation of the relatively hydrophobic nucleobases into a hydrophobic SAM may strengthen hydrophobic interactions and increase the strength of adsorption. In addition to the above permutations of surface properties, many questions remain about the behavior of ssDNA and dsDNA on the two model surfaces we have studied. First, the hypothesis in chapter 7 that adsorption events are associated with hybridization events could be studied by simulating the simultaneous adsorption of two complementary ssDNA oligomers, and observing the effect of adsorption on the strength or rate of hybridization or, alternatively, on the stability of a premade dsDNA oligomer. Second, a recent experimental study shows that DNA hairpins are more stable on a hydrophobic surface than a hydrophilic surface [19], and the molecular-level basis of this finding could be studied using computational methods similar to those employed in chapter 8 or other recently developed methods [16].

9.4 References

[1] M. Wei, S. M. Cohen, A. P. Silverman, and S. J. Lippard, Journal of Biological Chemistry 276 (2001) 38774.

[2] N. Margiotta, G. Natile, F. Capitelli, F. P. Fanizzi, A. Boccarelli, P. De Rinaldis, D. Giordano, and M. Coluccia, Journal of Inorganic Biochemistry 100 (2006) 1849.

[3] J. Kasparkova, T. Suchankova, A. Halamikova, L. Zerzankova, O. Vrana, N. Margiotta, G. Natile, and V. Brabec, Biochemical Pharmacology 79 (2010) 552.

[4] K. S. Lovejoy and S. J. Lippard, Dalton Transactions (2009) 10651.

[5] G. Y. Park, J. J. Wilson, Y. Song, and S. J. Lippard, Proceedings of the National Academy of Sciences 109 (2012) 11987.

[6] M. W. Kellinger, G. Y. Park, J. Chong, S. J. Lippard, and D. Wang, Journal of the American Chemical Society 135 (2013) 13054.

[7] D. Wang, G. Zhu, X. Huang, and S. J. Lippard, Proceedings of the National Academy of Sciences 107 (2010) 9584.

[8] S. S. Parelkar, D. Chan-Seng, and T. Emrick, Biomaterials 32 (2011) 2432.

[9] D. W. Pack, A. S. Hoffman, S. Pun, and P. S. Stayton, Nature Reviews Drug Discovery 4 (2005) 581.

[10] D. Lechardeur, A. S. Verkman, and G. L. Lukacs, Advanced Drug Delivery Reviews 57 (2005) 755.

[11] R. B. Breitenkamp and T. Emrick, Biomacromolecules 9 (2008) 2495.

[12] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo, The Journal of Chemical Physics 139 (2013)

[13] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, Biophysical Journal 96 (2009) 1675.

[14] S. Mao, M. Neu, O. Germershaus, O. Merkel, J. Sitterberg, U. Bakowsky, and T. Kissel, Bioconjugate Chemistry 17 (2006) 1209.

[15] T. Merdan, K. Kunath, H. Petersen, U. Bakowsky, K. H. Voigt, J. Kopecek, and T. Kissel, Bioconjugate chemistry 16 (2005) 785.

[16] M. Deighan and J. Pfaendtner, Langmuir 29 (2013) 7999.

[17] S. N. Jamadagni, R. Godawat, and S. Garde, in Annual Review of Chemical and Biomolecular Engineering, Vol 2, Vol. 2 (J. M. Prausnitz, ed.), Annual Reviews, Palo Alto, 2011, p. 147.

[18] E. Ostuni, R. G. Chapman, R. E. Holmlin, S. Takayama, and G. M. Whitesides, Langmuir 17 (2001) 5605.

[19] M. Kastantin and D. K. Schwartz, Small 9 (2012) 933.

Bibliography

1. Aaqvist, J. (1990) Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.*, **94**, 8021-8024.

2. Andersen, C. (1983) RATTLE: A "Velocity" Version of the SHAKE Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.*, **52**, 24-34.

3. Angelescu, D., Vasilescu, M. and Staikos, G. (2011) Characterization of the water-soluble comb–linear interpolyelectrolyte nanoaggregates by Monte Carlo simulations and fluorescence probe techniques. *Colloid. Polym. Sci.*, **289**, 871-879.

4. Arioka, H., Nishio, K., Ishida, T., Fukumoto, H., Fukuoka, K., Nomoto, T., Kurokawa, H., Yokote, H., Abe, S. and Saijo, N. (1999) Enhancement of cisplatin sensitivity in high mobility group 2 cDNA-transfected human lung cancer cells. *Jpn. J. Cancer Res.*, **90**, 108-115.

5. Balasubramanian, S., Pal, S. and Bagchi, B. (2002) Hydrogen-Bond Dynamics near a Micellar Surface: Origin of the Universal Slow Relaxation at Complex Aqueous Interfaces. *Phys. Rev. Lett.*, **89**, 115505.

6. Bandyopadhyay, D. and Choudhury, N. (2012) Characterizing hydrophobicity at the nanoscale: A molecular dynamics simulation study. *J. Chem. Phys.*, **136**, 224505-224510.

7. Barrat, J.L. and Joanny, J.F. (1993) Persistence Length of Polyelectrolyte Chains. *Europhys. Lett.*, **24**, 333.

8. Bassett, E., King, N.M., Bryant, M.F., Hector, S., Pendyala, L., Chaney, S.G. and Cordeiro-Stone, M. (2004) The role of DNA polymerase eta in translesion synthesis past platinum-DNA adducts in human fibroblasts. *Cancer Res.*, **64**, 6469-6475.

9. Betts, M.J. and Russell, R.B. (2003) In Barnes, M. R. and Gray, I. C. (eds.), *Bioinformatics for Geneticists*. John Wiley & Sons, Ltd, Chichester, UK, pp. 289-316.

10. Bhattacharyya, D., Ramachandran, S., Sharma, S., Pathmasiri, W., King, C.L., Baskerville-Abraham, I., Boysen, G., Swenberg, J.A., Campbell, S.L., Dokholyan, N.V. *et al.* (2011) Flanking Bases Influence the Nature of DNA Distortion by Platinum 1,2-Intrastrand (GG) Cross-Links. *PLoS ONE*, **6**, e23582.

11. Bolisetty, S., Airaud, C., Xu, Y., Müller, A.H.E., Harnau, L., Rosenfeldt, S., Lindner, P. and Ballauff, M. (2007) Softening of the stiffness of bottle-brush polymers by mutual interaction. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, **75**, 040803.

12. Bondi, A. (1964) VAN DER WAALS VOLUMES + RADII. *J. Phys. Chem.*, **68**, 441-&.

13. Bondi, A. (1964) van der Waals Volumes and Radii. *J. Phys. Chem.*, **68**, 441-451.

14. Borisov, O.V. and Zhulina, E.B. (2005) Amphiphilic Graft Copolymer in a Selective Solvent: Intramolecular Structures and Conformational Transitions. *Macromolecules*, **38**, 2506-2514.

15. Boussif, O., Lezoualc'h, F., Zanta, M.A., Mergny, M.D., Scherman, D., Demeneix, B. and Behr, J.P. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 7297-7297.

16. Breitenkamp, R.B. and Emrick, T. (2008) Pentalysine-grafted ROMP Polymers for DNA Complexation and Delivery. *Biomacromolecules*, **9**, 2495-2500.

17. Breitenkamp, R.B., Ou, Z., Breitenkamp, K., Muthukumar, M. and Emrick, T. (2007) Synthesis and Characterization of Polyolefin-graft-oligopeptide Polyelectrolytes. *Macromolecules*, **40**, 7617-7624.

18. Brukner, I., Dlakic, M., Savic, A., Susic, S., Pongor, S. and Suck, D. (1993) Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. *Nucleic Acids Res.*, **21**, 1025-1029.

19. Cao, Q., Zuo, C., Li, L. and Yan, G. (2011) Effects of chain stiffness and salt concentration on responses of polyelectrolyte brushes under external electric field. *Biomicrofluidics*, **5**, 044119-044112.

20. Cao, Q., Zuo, C., Li, L. and Zhang, N. (2010) Conformational Behavior of Bottle-Brush Polyelectrolytes with Charged and Neutral Side Chains. *Macromol. Theory Simul.*, **19**, 298-308.

21. Cao, Q., Zuo, C., Ma, Y., Li, L., Chen, S. and Hu, Z. (2011) Molecular dynamics simulations of end-grafted polymers with charged side chains. *J. Polym. Sci., Part B: Polym. Phys.*, **49**, 882-889.

22. Carot, M.L., Torresi, R.M., Garcia, C.D., Esplandiu, M.J. and Giacomelli, C.E. (2010) Electrostatic and hydrophobic interactions involved in CNT biofunctionalization with short ss-DNA. *J. Phys. Chem. C*, **114**, 4459-4465.

23. Cha, T.W., Boiadjiev, V., Lozano, J., Yang, H. and Zhu, X.Y. (2002) Immobilization of oligonucleotides on poly(ethylene glycol) brush-coated Si surfaces. *Anal. Biochem.*, **311**, 27-32.

24. Chan, H.S. and Dill, K.A. (1998) Protein folding in the landscape perspective: Chevron plots and non-arrhenius kinetics. *Proteins: Structure, Function, and Bioinformatics*, **30**, 2-33.

25. Chan, V., McKenzie, S.E., Surrey, S., Fortina, P. and Graves, D.J. (1998) Effect of Hydrophobicity and Electrostatics on Adsorption and Surface Diffusion of DNA Oligonucleotides at Liquid/Solid Interfaces. *J. Colloid Interface Sci.*, **203**, 197-207.

26. Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature*, **437**, 640-647.

27. Chaney, S.G., Campbell, S.L., Bassett, E. and Wu, Y.B. (2005) Recognition and processing of cisplatin- and oxaliplatin-DNA adducts. *Crit. Rev. Oncol. Hematol.*, **53**, 3-11.

28. Cheatham, T.E., Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845-862.

29. Chen, C., Wang, W., Ge, J. and Zhao, X.S. (2009) Kinetics and thermodynamics of DNA hybridization on gold nanoparticles. *Nucleic Acids Res.*, **37**, 3756-3765.

30. Chijiwa, S., Masutani, C., Hanaoka, F., Iwai, S. and Kuraoka, I. (2010) Polymerization by DNA polymerase eta is blocked by cis-diamminedichloroplatinum(II) 1,3-d(GpTpG) cross-link: implications for cytotoxic effects in nucleotide excision repair-negative tumor cells. *Carcinogenesis*, **31**, 388-393.

31. Cohen, S.M., Mikata, Y., He, Q. and Lippard, S.J. (2000) HMG-Domain protein recognition of cisplatin 1,2-intrastrand d(GpG) cross-links in purine-rich sequence contexts. *Biochemistry*, **39**, 11771-11776.

32. Connolly, R., Bellesia, G., Timoshenko, E.G., Kuznetsov, Y.A., Elli, S. and Ganazzoli, F. (2005) "Intrinsic" and "Topological" Stiffness in Branched Polymers. *Macromolecules*, **38**, 5288-5299.

33. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2ND GENERATION FORCE-FIELD FOR THE SIMULATION OF PROTEINS, NUCLEIC-ACIDS, AND ORGANIC-MOLECULES. *J. Am. Chem. Soc.*, **117**, 5179-5197.

34. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, **117**, 5179-5197.

35. Curuksu, J., Zacharias, M., Lavery, R. and Zakrzewska, K. (2009) Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res.*, **37**, 3766-3773.

36. Curuksu, J., Zakrzewska, K. and Zacharias, M. (2008) Magnitude and direction of DNA bending induced by screw-axis orientation: influence of sequence, mismatches and abasic sites. *Nucleic Acids Res.*, **36**, 2268-2283.

37. Czapla, L., Peters, J.P., Rueter, E.M., Olson, W.K. and Maher Iii, L.J. (2011) Understanding Apparent DNA Flexibility Enhancement by HU and HMGB Architectural Proteins. *J. Mol. Biol.*, **409**, 278-289.

38. Darden, T., York, D. and Pedersen, L.D.A.J. (1993) Particle Mesh Ewald: An N-Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.*, **98**, 10089-10092.

39. Deighan, M. and Pfaendtner, J. (2013) Exhaustively Sampling Peptide Adsorption with Metadynamics. *Langmuir*, **29**, 7999-8009.

40. Dennig, J. and Duncan, E. (2002) Gene transfer into eukaryotic cells using activated polyamidoamine dendrimers. *Rev. Mol. Biotechnol.*, **90**, 339-347.

41. DeRouchey, J., Hoover, B. and Rau, D.C. (2013) A Comparison of DNA Compaction by Arginine and Lysine Peptides: A Physical Basis for Arginine Rich Protamines. *Biochemistry*, **52**, 3000-3009.

42. Dias, R.S., Linse, P. and Pais, A.A.C.C. (2011) Stepwise disproportionation in polyelectrolyte complexes. *J. Comput. Chem.*, **32**, 2697-2707.

43. Dias, R.S. and Pais, A. (2010) Polyelectrolyte condensation in bulk, at surfaces, and under confinement. *Adv. Colloid Interface Sci.*, **158**, 48-62.

44. Dill, K.A. (1997) Additivity Principles in Biochemistry. *J. Biol. Chem.*, **272**, 701-704.

45. Dillenback, L.M., Goodrich, G.P. and Keating, C.D. (2005) Temperature-Programmed Assembly of DNA:Au Nanoparticle Bioconjugates. *Nano Lett.*, **6**, 16-23.

46. Dobrynin, A.V. (2005) Electrostatic Persistence Length of Semiflexible and Flexible Polyelectrolytes. *Macromolecules*, **38**, 9304-9314.

47. Douglas, S.M., Dietz, H., Liedl, T., Hogberg, B., Graf, F. and Shih, W.M. (2009) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, **459**, 414-418.

48. Dunham, S.U. and Lippard, S.J. (1997) DNA sequence context and protein composition modulate HMG-domain protein recognition of ciplatin-modified DNA. *Biochemistry*, **36**, 11428-11436.

49. E, W., Ren, W. and Vanden-Eijnden, E. (2002) String method for the study of rare events. *Phys. Rev. B.*, **66**, 052301.

50. E, W., Ren, W. and Vanden-Eijnden, E. (2007) Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, **126**, 164103-164108.

51. Elder, R.M., Emrick, T. and Jayaraman, A. (2011) Understanding the Effect of Polylysine Architecture on DNA Binding Using Molecular Dynamics Simulations. *Biomacromolecules*, **12**, 3870-3879.

52. Elder, R.M. and Jayaraman, A. (2012) Role of structure and dynamics of DNA with cisplatin and oxaliplatin adducts in various sequence contexts on binding of HMGB1a. *Mol. Sim.*, **38**, 793-808.

53. Elder, R.M. and Jayaraman, A. (2012) Coarse-Grained Simulation Studies of Effects of Polycation Architecture on Structure of the Polycation and Polycation-Polyanion Complexes. *Macromolecules*, **45**, 8083-8096.

54. Elder, R.M. and Jayaraman, A. (2013) Structure and thermodynamics of ssDNA oligomers near hydrophobic and hydrophilic surfaces. *Soft Matter*, **9**, 11521-11533.

55. Elizondo-Riojas, M.A. and Kozelka, J. (2001) Unrestrained 5 ns molecular dynamics simulation of a cisplatin-DNA 1,2-GG adduct provides a rationale for the NMR features and reveals increased conformational flexibility at the platinum binding site. *J. Mol. Biol.*, **314**, 1227-1243.

56. Feng, J., Wong, K.-Y., Dyer, K. and Pettitt, B.M. (2009) Transport properties of water at functionalized molecular interfaces. *J. Chem. Phys.*, **131**, 125102-125108.

57. Feuz, L., Leermakers, F.A.M., Textor, M. and Borisov, O. (2005) Bending Rigidity and Induced Persistence Length of Molecular Bottle Brushes: A Self-Consistent-Field Theory. *Macromolecules*, **38**, 8891-8901.

58. Fichtingerschepman, A.M.J., Vanderveer, J.L., Denhartog, J.H.J., Lohman, P.H.M. and Reedijk, J. (1985) Adducts of the antitumor drug cis-diamminedichloroplatinum(II) with DNA - Formation, identification, and quantitation. *Biochemistry*, **24**, 707-713.

59. Frederick, K.K., Marlow, M.S., Valentine, K.G. and Wand, A.J. (2007) Conformational entropy in molecular recognition by proteins. *Nature*, **448**, 325-329.

60. Frenkel, D. and Smit, B. (2001) *Understanding Molecular Simulations: From Algorithms to Applications*. Academic Press.

61. Fromm, J.R., Hileman, R.E., Caldwell, E.E.O., Weiler, J.M. and Linhardt, R.J. (1995) Differences in the Interaction of Heparin with Arginine and Lysine and the Importance of these Basic Amino Acids in the Binding of Heparin to Acidic Fibroblast Growth Factor. *Arch. Biochem. Biophys.*, **323**, 279-287.

62. Gelasco, A. and Lippard, S.J. (1998) NMR solution structure of a DNA dodecamer duplex containing a cis-diammineplatinum(II) d(GpG) intrastrand cross-link, the major adduct of the anticancer drug cisplatin. *Biochemistry*, **37**, 9230-9239.

63. Glover, D.J., Lipps, H.J. and Jans, D.A. (2005) Towards safe, non-viral therapeutic gene expression in humans. *Nat. Rev. Genet.*, **6**, 299-U229.

64. Gohlke, H. and Case, D.A. (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.*, **25**, 238-250.

65. Gong, P. and Levicky, R. (2008) DNA surface hybridization regimes. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 5301-5306.

66. Gopal, S.M., Mukherjee, S., Cheng, Y.-M. and Feig, M. (2010) PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins*, **78**, 1266-1281.

67. Gresham, D., Dunham, M.J. and Botstein, D. (2008) Comparing whole genomes using DNA microarrays. *Nat Rev Genet*, **9**, 291-302.

68. Grossfield, A. WHAM: the weighted histogram analysis method. *"WHAM: the weighted histogram analysis method," version 2.0.4, membrane.urmc.rochester.edu/content/wham.*

69. Gudipati, C.S., Finlay, J.A., Callow, J.A., Callow, M.E. and Wooley, K.L. (2005) The Antifouling and Fouling-Release Perfomance of Hyperbranched Fluoropolymer (HBFP)−Poly(ethylene glycol) (PEG) Composite Coatings Evaluated by Adsorption of Biomacromolecules and the Green Fouling Alga Ulva. *Langmuir*, **21**, 3044-3053.

70. Haensler, J. and Szoka Jr, F.C. (1993) Polyamidoamine cascade polymers mediate efficient transfection of cells in culture. *Bioconjugate Chem.*, **4**, 372-379.

71. Hagan, M.F., Dinner, A.R., Chandler, D. and Chakraborty, A.K. (2003) Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 13922-13927.

72. Hambley, T.W. (1998) van der Waals radii of Pt(II) and Pd(II) in molecular mechanics models and an analysis of their relevance to the description of axial M center dot center dot center dot H(-C), M center dot center dot center dot H(-N), M center dot center dot center dot S and M center dot center dot center dot M (M = Pd(II) or Pt(II)) interactions. *Inorg. Chem.*, **37**, 3767-3774.

73. Hambley, T.W. (1998) van der Waals Radii of Pt(II) and Pd(II) in Molecular Mechanics Models and an Analysis of Their Relevance to the Description of Axial M.H(-C), M.H(-N), M.S, and M.M (M = Pd(II) or Pt(II)) Interactions. *Inorg. Chem.*, **37**, 3767-3774.

74. He, Q., Ohndorf, U.M. and Lippard, S.J. (2000) Intercalating residues determine the mode of HMG1 domains A and B binding to cisplatin-modified DNA. *Biochemistry*, **39**, 14426-14435.

75. Hinckley, D.M., Freeman, G.S., Whitmer, J.K. and de Pablo, J.J. (2013) An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *The Journal of Chemical Physics*, **139**, -.

76. Hockney, R.W. and Eastwood, J.W. (1988) *Computer Simulation Using Particles*. Taylor & Francis Group, New York.

77. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins*, **65**, 712-725.

78. Hower, J.C., He, Y. and Jiang, S. (2008) A molecular simulation study of methylated and hydroxyl sugar-based self-assembled monolayers: Surface hydration and resistance to protein adsorption. *J. Chem. Phys.*, **129**, 215101-215107.

79. Huang, J.C., Zamble, D.B., Reardon, J.T., Lippard, S.J. and Sancar, A. (1994) HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 10394-10398.

80. Hummer, G., Garde, S., García, A.E., Paulaitis, M.E. and Pratt, L.R. (1998) The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 1552-1555.

81. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graphics Modell.*, **14**, 33-38.

82. Hung, A., Mager, M., Hembury, M., Stellacci, F., Stevens, M.M. and Yarovsky, I. (2013) Amphiphilic amino acids: a key to adsorbing proteins to nanopatterned surfaces? *Chem. Sci.*, **4**, 928-937.

83. Jakalian, A., Jack, D.B. and Bayly, C.I. (2002) Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.*, **23**, 1623-1641.

84. Jamadagni, S.N., Godawat, R. and Garde, S. (2011) In Prausnitz, J. M. (ed.), *Annual Review of Chemical and Biomolecular Engineering, Vol 2*. Annual Reviews, Palo Alto, Vol. 2, pp. 147-171.

85. Jamieson, E.R. and Lippard, S.J. (1999) Structure, recognition, and processing of cisplatin-DNA adducts. *Chem. Rev.*, **99**, 2467-2498.

86. Jayaraman, A., Hall, C.K. and Genzer, J. (2006) Computer simulation study of molecular recognition in model DNA microarrays. *Biophys. J.*, **91**, 2227-2236.

87. Jayaraman, A., Hall, C.K. and Genzer, J. (2007) Computer simulation study of probe-target hybridization in model DNA microarrays: Effect of probe surface density and target concentration. *J. Chem. Phys.*, **127**, 144912.

88. Johnson, R.R., Johnson, A.T.C. and Klein, M.L. (2007) Probing the Structure of DNA−Carbon Nanotube Hybrids with Molecular Dynamics. *Nano Lett.*, **8**, 69-75.

89. Johnson, R.R., Kohlmeyer, A., Johnson, A.T.C. and Klein, M.L. (2009) Free Energy Landscape of a DNA−Carbon Nanotube Hybrid Using Replica Exchange Molecular Dynamics. *Nano Lett.*, **9**, 537-541.

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

91. Joung, I.S. and Cheatham, T.E. (2008) Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B*, **112**, 9020-9041.

92. Jung, Y.W. and Lippard, S.J. (2007) Direct cellular responses to platinum-induced DNA damage. *Chem. Rev.*, **107**, 1387-1407.

93. Juo, Z.S., Chiu, T.K., Leiberman, P.M., Baikalov, I., Berk, a.J. and Dickerson, R.E. (1996) How proteins recognize the TATA box. *J. Mol. Biol.*, **261**, 239-254.

94. Kaggwa, G.B., Nalam, P.C., Kilpatrick, J.I., Spencer, N.D. and Jarvis, S.P. (2012) Impact of Hydrophilic/Hydrophobic Surface Chemistry on Hydration Forces in the Absence of Confinement. *Langmuir*, **28**, 6589-6594.

95. Kasparkova, J., Delalande, O., Stros, M., Elizondo-Riojas, M.A., Vojtiskova, M., Kozelka, J. and Brabec, V. (2003) Recognition of DNA interstrand cross-link of antitumor cisplatin by HMGB1 protein. *Biochemistry*, **42**, 1234-1244.

96. Kasparkova, J., Suchankova, T., Halamikova, A., Zerzankova, L., Vrana, O., Margiotta, N., Natile, G. and Brabec, V. (2010) Cytotoxicity, cellular uptake, glutathione and DNA interactions of an antitumor large-ring Pt-II chelate complex incorporating the cis-1,4-diaminocyclohexane carrier ligand. *Biochem. Pharmacol.*, **79**, 552-564.

97. Kastantin, M. and Schwartz, D.K. (2012) DNA Hairpin Stabilization on a Hydrophobic Surface. *Small*, **9**, 933-941.

98. Kellinger, M.W., Park, G.Y., Chong, J., Lippard, S.J. and Wang, D. (2013) Effect of a Monofunctional Phenanthriplatin-DNA Adduct on RNA Polymerase II Transcriptional Fidelity and Translesion Synthesis. *J. Am. Chem. Soc.*, **135**, 13054-13061.

99. Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S.H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W. *et al.* (2000) Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.*, **33**, 889-897.

100. Kone, A. and Kofke, D.A. (2005) Selection of temperature intervals for parallel-tempering simulations. *The Journal of Chemical Physics*, **122**, -.

101. Korolev, N., Lyubartsev, A.P. and Nordenskiöld, L. (2010) Cation-induced polyelectrolyte–polyelectrolyte attraction in solutions of DNA and nucleosome core particles. *Adv. Colloid Interface Sci.*, **158**, 32-47.

102. Košovan, P., Limpouchová, Z. and Procházka, K. (2007) Conformational Behavior of Comb-like Polyelectrolytes in Selective Solvent: Computer Simulation Study†. *J. Phys. Chem. B*, **111**, 8605-8611.

103. Kozelka, J., Savinelli, R., Berthier, G., Flament, J.P. and Lavery, R. (1993) Force field for platinum binding to adenine. *J. Comput. Chem.*, **14**, 45-53.

104. Kramarenko, E.Y., Pevnaya, O.S. and Khokhlov, A.R. (2005) Stoichiometric polyelectrolyte complexes as comb copolymers. *J. Chem. Phys.*, **122**, 084902-084910.

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

106. Kundagrami, A. and Muthukumar, M. (2010) Effective Charge and Coil−Globule Transition of a Polyelectrolyte Chain. *Macromolecules*, **43**, 2574-2581.

107. Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E. and Corbett, A.H. (2007) Classical Nuclear Localization Signals: Definition, Function, and Interaction with Importin α. *J. Biol. Chem.*, **282**, 5101-5105.

108. Lange, S.S. and Vasquez, K.M. (2009) HMGB1: The Jack-of-All-Trades Protein Is a Master DNA Repair Mechanic. *Mol. Carcinog.*, **48**, 571-580.

109. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917-5929.

110. Lechardeur, D., Verkman, A.S. and Lukacs, G.L. (2005) Intracellular routing of plasmid DNA during non-viral gene transfer. *Adv. Drug Delivery Rev.*, **57**, 755-767.

111. Lee, A.J. and Rick, S.W. (2009) Improving replica exchange using driven scaling. *The Journal of Chemical Physics*, **131**, -.

112. Li, D.W. and Bruschweiler, R. (2009) A Dictionary for Protein Side-Chain Entropies from NMR Order Parameters. *J. Am. Chem. Soc.*, **131**, 7226-7227.

113. Li, D.W. and Bruschweiler, R. (2009) In silico Relationship between Configurational Entropy and Soft Degrees of Freedom in Proteins and Peptides. *Phys. Rev. Lett.*, **102**, 118108.

114. Li, L., Bedrov, D. and Smith, G.D. (2006) Water-Induced Interactions between Carbon Nanoparticles. *J. Phys. Chem. B*, **110**, 10509-10513.

115. Liang, H., Li, Z. and Yang, J. (2010) Single-stranded DNA adsorption on chiral molecule coated Au surface: a molecular dynamics study. *Phys. Chem. Chem. Phys.*, **12**, 4431-4434.

116. Lilley, D.M.J. (1996) Cisplatin adducts in DNA: Distortion and recognition. *J. Biol. Inorg. Chem.*, **1**, 189-191.

117. Limbach, H. (2001) End effects of strongly charged polyelectrolytes: A molecular dynamics study. *J. Chem. Phys.*, **114**, 9674.

118. Liu, Z., Ensing, B. and Moore, P.B. (2010) Quantitative Assessment of Force Fields on Both Low-Energy Conformational Basins and Transition-State Regions of the (ϕ−ψ) Space. *J. Chem. Theory Comput.*, **7**, 402-419.

119. Lovejoy, K.S. and Lippard, S.J. (2009) Non-traditional platinum compounds for improved accumulation, oral bioavailability, and tumor targeting. *Dalton Transactions*, 10651-10659.

120. Luzar, A. and Chandler, D. (1996) Hydrogen-bond kinetics in liquid water. *Nature*, **379**, 55-57.

121. Macke, T.J. and Case, D.A. (1997), *Molecular Modeling of Nucleic Acids*. American Chemical Society, Vol. 682, pp. 379-393.

122. Magdalena, M., Pawel, S., Ewa, N.M., Marcin, M., Piotr, M., Wojciech, R., Helena, K., Maciej, Z. and Marek, S. (2008) Analysis of hMLH1 and hMSH2 expression in cisplatin-treated ovarian cancer patients. *Ginekologia Polska*, **79**, 826-834.

123. Mahadevi, A.S. and Sastry, G.N. (2012) Cation−π Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chem. Rev.*, **113**, 2100-2138.

124. Malina, J., Novakova, O., Vojtiskova, M., Natile, G. and Brabec, V. (2007) Conformation of DNA GG intrastrand cross-link of antitumor oxaliplatin and its enantiomeric analog. *Biophys. J.*, **93**, 3950-3962.

125. Malina, J., Voitiskova, M., Brabec, V., Diakos, C.I. and Hambley, T.W. (2005) DNA adducts of the enantiomers of the Pt(II) complexes of the ahaz ligand (ahaz=3-aminohexahydroazepine) and recognition of these adducts by HMG domain proteins. *Biochem. Biophys. Res. Commun.*, **332**, 1034-1041.

126. Mao, S., Neu, M., Germershaus, O., Merkel, O., Sitterberg, J., Bakowsky, U. and Kissel, T. (2006) Influence of Polyethylene Glycol Chain Length on the Physicochemical and Biological Properties of Poly(ethylene imine)-graft-Poly(ethylene glycol) Block Copolymer/siRNA Polyplexes. *Bioconjugate Chem.*, **17**, 1209-1218.

127. Margiotta, N., Natile, G., Capitelli, F., Fanizzi, F.P., Boccarelli, A., De Rinaldis, P., Giordano, D. and Coluccia, M. (2006) Sterically hindered complexes of platinum(II) with planar heterocyclic nitrogen donors. A novel complex with 1-methyl-cytosine has a spectrum of activity different from cisplatin and is able of overcoming acquired cisplatin resistance. *J. Inorg. Biochem.*, **100**, 1849-1857.

128. Mark, P. and Nilsson, L. (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A*, **105**, 9954-9960.

129. Marlow, M.S., Dogan, J., Frederick, K.K., Valentine, K.G. and Wand, A.J. (2010) The role of conformational entropy in molecular recognition by calmodulin. *Nature Chemical Biology*, **6**, 352-358.

130. Marples, B., Adomat, H., Billings, P.C., Farrell, N.P., Koch, C.J. and Skov, K.A. (1994) Recognition of platinum-induced DNA-damage by nuclear proteins - Screening for mechanisms. *Anti-Cancer Drug Des.*, **9**, 389-399.

131. Marsh, R.J., Jones, R.A.L. and Sferrazza, M. (2002) Adsorption and displacement of a globular protein on hydrophilic and hydrophobic surfaces. *Colloids Surf., B*, **23**, 31-42.

132. Martin, W., Zhu, W. and Krilov, G. (2008) Simulation Study of Noncovalent Hybridization of Carbon Nanotubes by Single-Stranded DNA in Water†. *The Journal of Physical Chemistry B*, **112**, 16076-16089.

133. Mascotti, D.P. and Lohman, T.M. (1990) Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 3142-3146.

134. Mattice, W.L. and Suter, U.W. (1994) *Conformational theory of large molecules: The rotational isomeric state model in macromolecular systems*. John Wiley & Sons, Ltd.

135. Mazur, A.K. (2006) Evaluation of elastic properties of atomistic DNA models. *Biophys. J.*, **91**, 4507-4518.

136. Mello, J.A., Acharya, S., Fishel, R. and Essigmann, J.M. (1996) The mismatch-repair protein hMSH2 binds selectively to DNA adducts of the anticancer drug cisplatin. *Chem. Biol. (Cambridge, MA, United States)*, **3**, 579-589.

137. Merdan, T., Kunath, K., Petersen, H., Bakowsky, U., Voigt, K.H., Kopecek, J. and Kissel, T. (2005) PEGylation of poly(ethylene imine) affects stability of complexes with plasmid DNA under in vivo conditions in a dose-dependent manner after intravenous injection into mice. *Bioconjugate Chem.*, **16**, 785-792.

138. Micheletti, C., Laio, A. and Parrinello, M. (2004) Reconstructing the Density of States by History-Dependent Metadynamics. *Phys. Rev. Lett.*, **92**, 170601.

139. Mikata, Y., He, Q. and Lippard, S.J. (2001) Laser-induced photo-cross-linking of cisplatin-modified DNA to HMG-domain proteins. *Biochemistry*, **40**, 7533-7541.

140. Mishra, S., Webster, P. and Davis, M.E. (2004) PEGylation significantly affects cellular uptake and intracellular trafficking of non-viral gene delivery particles. *Eur. J. Cell Biol.*, **83**, 97-111.

141. Monserud, J.H. and Schwartz, D.K. (2012) Effects of Molecular Size and Surface Hydrophobicity on Oligonucleotide Interfacial Dynamics. *Biomacromolecules*, **13**, 4002-4011.

142. Monti, S., Cacelli, I., Ferretti, A., Prampolini, G. and Barone, V. (2010) Simulating DNA Hybridization on an Amine-Functionalized Silicon Substrate. *J. Phys. Chem. B*, **114**, 8341-8349.

143. Monti, S. and Walsh, T.R. (2011) Molecular Dynamics Simulations of the Adsorption and Dynamical Behavior of Single DNA Components on TiO2. *J. Phys. Chem. C*, **115**, 24238-24246.

144. Narambuena, C.F., Leiva, E.P.M., Chávez-Páez, M. and Pérez, E. (2010) Effect of chain stiffness on the morphology of polyelectrolyte complexes. A Monte Carlo simulation study. *Polymer*, **51**, 3293-3302.

145. Nishikawa, M. and Huang, L. (2001) Nonviral Vectors in the New Millennium: Delivery Barriers in Gene Transfer. *Hum. Gene Ther.*, **12**, 861-870.

146. Noguchi, H. and Yoshikawa, K. (1998) Morphological variation in a collapsed single homopolymer chain. *J. Chem. Phys.*, **109**, 5070-5077.

147. Noh, H., Hung, A.M. and Cha, J.N. (2011) Surface-Driven DNA Assembly of Binary Cubic 3D Nanocrystal Superlattices. *Small*, **7**, 3021-3025.

148. Noid, W.G., Chu, J.W., Ayton, G.S., Krishna, V., Izvekov, S., Voth, G.A., Das, A. and Andersen, H.C. (2008) The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.*, **128**.

149. Ohndorf, U.M., Rould, M.A., He, Q., Pabo, C.O. and Lippard, S.J. (1999) Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature*, **399**, 708-712.

150. Ostuni, E., Chapman, R.G., Holmlin, R.E., Takayama, S. and Whitesides, G.M. (2001) A Survey of Structure−Property Relationships of Surfaces that Resist the Adsorption of Protein. *Langmuir*, **17**, 5605-5620.

151. Ostuni, E., Chapman, R.G., Liang, M.N., Meluleni, G., Pier, G., Ingber, D.E. and Whitesides, G.M. (2001) Self-Assembled Monolayers That Resist the Adsorption of Proteins and the Adhesion of Bacterial and Mammalian Cells. *Langmuir*, **17**, 6336-6343.

152. Ou, S., Patel, S. and Bauer, B.A. (2012) Free Energetics of Carbon Nanotube Association in Pure and Aqueous Ionic Solutions. *J. Phys. Chem. B*, **116**, 8154-8168.

153. Ou, Z. and Muthukumar, M. (2005) Langevin dynamics of semiflexible polyelectrolytes: rod-toroid-globule-coil structures and counterion distribution. *J. Chem. Phys.*, **123**, 074905-074905.

154. Ou, Z. and Muthukumar, M. (2006) Entropy and enthalpy of polyelectrolyte complexation: Langevin dynamics simulations. *J. Chem. Phys.*, **124**, 154902-154902.

155. Ou, Z.Y. and Muthukumar, M. (2006) Entropy and Enthalpy of Polyelectrolyte Complexation: Langevin Dynamics Simulations. *J. Chem. Phys.*, **124**, 154902.

156. Pack, D.W., Hoffman, A.S., Pun, S. and Stayton, P.S. (2005) Design and Development of Polymers for Gene Delivery. *Nat. Rev. Drug Discovery*, **4**, 581-593.

157. Parelkar, S.S., Chan-Seng, D. and Emrick, T. (2011) Reconfiguring Polylysine Architectures for Controlling Polyplex Binding and Non-viral Transfection. *Biomaterials*, **32**, 2432-2444.

158. Park, G.Y., Wilson, J.J., Song, Y. and Lippard, S.J. (2012) Phenanthriplatin, a monofunctional DNA-binding platinum anticancer drug candidate with unusual potency and cellular activity profile. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 11987-11992.

159. Park, S. and Lippard, S.J. (2011) Redox State-Dependent Interaction of HMGB1 and Cisplatin-Modified DNA. *Biochemistry*, **50**, 2567-2574.

160. Park, S.Y., Lee, J.S., Georganopoulou, D., Mirkin, C.A. and Schatz, G.C. (2006) Structures of DNA-linked nanoparticle aggregates. *J. Phys. Chem. B*, **110**, 12673-12681.

161. Park, S.Y., Lytton-Jean, K.R.A., Lee, B., Weigand, S., Schatz, G.C. and Mirkin, C.A. (2008) DNA programmable nanoparticle crystallization. *Nature*, **451**, 553-556.

162. Parkinson, J.A., Chen, Y., del Socorro Murdoch, P., Guo, Z., Berners-Price, S.J., Brown, T. and Sadler, P.J. (2000) Sequence-Dependent Bending of DNA Induced by cisplatin: NMR structures of an AT-Rich 14-mer duplex. *Chem. Eur. J.*, **6**, 3636-3644.

163. Pasheva, E.A., Ugrinova, I., Spassovska, N.C. and Pashev, I.G. (2002) The binding affinity of HMG1 protein to DNA modified by cis-platin and its analogs correlates with their antitumor activity. *Int. J. Biochem. Cell Biol.*, **34**, 87-92.

164. Pavan, G.M., Danani, A., Pricl, S. and Smith, D.K. (2009) Modeling the multivalent recognition between dendritic molecules and DNA: understanding how ligand "sacrifice" and screening can enhance binding. *J. Am. Chem. Soc.*, **131**, 9686-9694.

165. Pérez, A., Luque, F.J. and Orozco, M. (2011) Frontiers in Molecular Dynamics Simulations of DNA. *Acc. Chem. Res.*, **45**, 196-205.

166. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817-3829.

167. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.*, **92**, 3817-3829.

168. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K. (2005) Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781-1802.

169. Pilch, D.S., Dunham, S.U., Jamieson, E.R., Lippard, S.J. and Breslauer, K.J. (2000) DNA sequence context modulates the impact of a cisplatin 1,2-d(GpG) intrastrand cross-link an the conformational and thermodynamic properties of duplex DNA. *J. Mol. Biol.*, **296**, 803-812.

170. Pinholt, C., Hartvig, R.A., Medlicott, N.J. and Jorgensen, L. (2011) The importance of interfaces in protein drug delivery - why is protein adsorption of interest in pharmaceutical formulations? *Expert Opin. Drug Deliv.*, **8**, 949-964.

171. Plimpton, S. (1995) Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.*, **117**, 1-19.

172. Poynor, A., Hong, L., Robinson, I.K., Granick, S., Zhang, Z. and Fenter, P.A. (2006) How Water Meets a Hydrophobic Surface. *Phys. Rev. Lett.*, **97**, 266101.

173. Ramachandran, S., Temple, B.R., Chaney, S.G. and Dokholyan, N.V. (2009) Structural basis for the sequence-dependent effects of platinum-DNA adducts. *Nucleic Acids Res.*, **37**, 2434-2448.

174. Raspopovic, Z.M., Sakadzic, S., Bzenic, S.A. and Petrovic, Z.L. (1999) Benchmark calculations for Monte Carlo simulations of electron transport. *IEEE Transactions on Plasma Science*, **27**, 1241-1248.

175. Rathore, N., Chopra, M. and de Pablo, J.J. (2005) Optimal allocation of replicas in parallel tempering simulations. *The Journal of Chemical Physics*, **122**, -.

176. Rodriguez-Gomez, D., Darve, E. and Pohorille, A. (2004) Assessing the efficiency of free energy calculation methods. *J. Chem. Phys.*, **120**, 3563-3578.

177. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B. and Mann, R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem*, **79**, 233-269.

178. Rothemund, P.W.K. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature*, **440**, 297-302.

179. Roux, B. (1995) The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.*, **91**, 275-282.

180. Rubinstein, M. and Colby, R.H. (2008) *Polymer Physics*. Oxford University Press.

181. Saariaho, M., Ikkala, O., Szleifer, I., Erukhimovich, I. and ten Brinke, G. (1997) On lyotropic behavior of molecular bottle-brushes: A Monte Carlo computer simulation study. *J. Chem. Phys.*, **107**, 3267-3276.

182. Saariaho, M., Szleifer, I., Ikkala, O. and ten Brinke, G. (1998) Extended conformations of isolated molecular bottle-brushes: Influence of side-chain topology. *Macromol. Theory Simul.*, **7**, 211-216.

183. Salsbury, F.R., Clodfelter, J.E., Gentry, M.B., Hollis, T. and Scarpinato, K.D. (2006) The molecular mechanism of DNA damage recognition by MutS homologs and its consequences for cell death response. *Nucleic Acids Res.*, **34**, 2173-2185.

184. Sambriski, E.J., Schwartz, D.C. and de Pablo, J.J. (2009) A Mesoscale Model of DNA and Its Renaturation. *Biophys. J.*, **96**, 1675-1690.

185. Scheeff, E.D., Briggs, J.M. and Howell, S.B. (1999) Molecular modeling of the intrastrand guanine-guanine DNA adducts produced by cisplatin and oxaliplatin. *Mol. Pharmacol.*, **56**, 633-643.

186. Schlick, T. (2002) *Molecular modeling and simulation: an interdisciplinary guide*. Springer-Verlag.

187. Schwierz, N., Horinek, D., Liese, S., Pirzer, T., Balzer, B.N., Hugel, T. and Netz, R.R. (2012) On the Relationship between Peptide Adsorption Resistance and Surface Contact Angle: A Combined Experimental and Simulation Single-Molecule Study. *J. Am. Chem. Soc.*, **134**, 19628-19638.

188. Serr, A., Horinek, D. and Netz, R.R. (2008) Polypeptide Friction and Adhesion on Hydrophobic and Hydrophilic Surfaces: A Molecular Dynamics Case Study. *J. Am. Chem. Soc.*, **130**, 12408-12413.

189. Sharma, S., Gong, P., Temple, B., Bhattacharyya, D., Dokholyan, N.V. and Chaney, S.G. (2007) Molecular dynamic simulations of cisplatin- and oxaliplatin-d(GG) intrastand cross-links reveal differences in their conformational dynamics. *J. Mol. Biol.*, **373**, 1123-1140.

190. Spiegel, K., Magistrato, A., Carloni, P., Reedijk, J. and Klein, M.L. (2007) Azole-bridged diplatinum anticancer compounds. Modulating DNA flexibility to escape repair mechanism and avoid cross resistance. *J. Phys. Chem. B*, **111**, 11873-11876.

191. Spiegel, K., Rothlisberger, U. and Carloni, P. (2004) Cisplatin binding to DNA oligomers from hybrid Car-Parrinello/molecular dynamics simulations. *J. Phys. Chem. B*, **108**, 2699-2707.

192. Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A. and Case, D.A. (1998) Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *J. Am. Chem. Soc.*, **120**, 9401-9409.

193. Stefl, R., Wu, H., Ravindranathan, S., Sklenár, V. and Feigon, J. (2004) DNA A-tract bending in three dimensions: Solving the dA4T4 vs. dT4A4 conundrum. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 1177-1182.

194. Stehlikova, K., Kostrhunova, H., Kasparkova, J. and Brabec, V. (2002) DNA bending and unwinding due to the major 1,2-GG intrastrand cross-link formed by antitumor cis-diamminedichloroplatinum(II) are flanking-base independent. *Nucleic Acids Res.*, **30**, 2894-2898.

195. Stevens, M.J. and Grest, G.S. (2008) Simulations of water at the interface with hydrophilic self-assembled monolayers. *Biointerphases*, **3**, FC13-FC22.

196. Stoughton, R.B. (2005) Applications of DNA Microarrays in Biology. *Annu. Rev. Biochem*, **74**, 53-82.

197. Striolo, A. (2007) Controlled assembly of spherical nanoparticles: Nanowires and spherulites. *Small*, **3**, 628-635.

198. Striolo, A., Jayaraman, A., Genzer, J. and Hall, C.K. (2005) Adsorption of comb copolymers on weakly attractive solid surfaces. *J. Chem. Phys.*, **123**.

199. Stros, M. (2010) HMGB proteins: Interactions with DNA and chromatin. *Biochim. Biophys. Acta, Gene Regul. Mech.*, **1799**, 101-113.

200. Stumpff-Kane, A.W., Maksimiak, K., Lee, M.S. and Feig, M. (2008) Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. *Proteins*, **70**, 1345-1356.

201. Sugita, Y. and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **314**, 141-151.

202. Sun, X., Hyeon Ko, S., Zhang, C., Ribbe, A.E. and Mao, C. (2009) Surface-Mediated DNA Self-Assembly. *J. Am. Chem. Soc.*, **131**, 13248-13249.

203. Takahara, P.M., Rosenzweig, A.C., Frederick, C.A. and Lippard, S.J. (1995) Crystal-structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. *Nature*, **377**, 649-652.

204. Tan, C., Tan, Y.-H. and Luo, R. (2007) Implicit nonpolar solvent models. *J. Phys. Chem. B*, **111**, 12263-12274.

205. Tang, M.X. and Szoka, F.C. (1997) The influence of polymer structure on the interactions of cationic polymers with DNA and morphology of the resulting complexes. *Gene Therapy*, **4**, 823-832.

206. Teletchea, S., Skauge, T., Sletten, E. and Kozelka, J. (2009) Cisplatin Adducts on a GGG Sequence within a DNA Duplex Studied by NMR Spectroscopy and Molecular Dynamics Simulations. *Chem. Eur. J.*, **15**, 12320-12337.

207. Thomas, M. and Klibanov, A.M. (2003) Non-viral Gene Therapy: Polycation-mediated DNA Delivery. *Appl. Microbiol. Biotechnol.*, **62**, 27-34.

208. Tinland, B., Pluen, A., Sturm, J. and Weill, G. (1997) Persistence Length of Single-Stranded DNA. *Macromolecules*, **30**, 5763-5765.

209. Todd, R.C. and Lippard, S.J. (2010) Structure of duplex DNA containing the cisplatin 1,2-{Pt(NH3)(2)}(2+)-d(GpG) cross-link at 1.77 angstrom resolution. *J. Inorg. Biochem.*, **104**, 902-908.

210. Torrie, G.M. and Valleau, J.P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, **23**, 187-199.

211. Tyagi, P., Gahlot, P. and Kakkar, R. (2008) Structural aspects of the anti-cancer drug oxaliplatin: A combined theoretical and experimental study. *Polyhedron*, **27**, 3567-3574.

212. Ullner, M. (2007), *DNA Interactions with Polymers and Surfactants*. John Wiley & Sons, Inc., pp. 1-39.

213. Ullner, M. and Woodward, C.E. (2002) Orientational Correlation Function and Persistence Lengths of Flexible Polyelectrolytes. *Macromolecules*, **35**, 1437-1445.

214. Ulman, A. (1996) Formation and Structure of Self-Assembled Monolayers. *Chem. Rev.*, **96**, 1533-1554.

215. Vaisman, A., Lim, S.E., Patrick, S.M., Copeland, W.C., Hinkle, D.C., Turchi, J.J. and Chaney, S.G. (1999) Effect of DNA polymerases and high mobility group protein 1 on the carrier ligand specificity for translesion synthesis past platinum-DNA adducts. *Biochemistry*, **38**, 11026-11039.

216. Vaisman, A., Masutani, C., Hanaoka, F. and Chaney, S.G. (2000) Efficient translesion replication past oxaliplatin and cisplatin GpG adducts by human DNA polymerase eta. *Biochemistry*, **39**, 4575-4580.

217. Vanden-Eijnden, E. and Venturoli, M. (2009) Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.*, **130**, 194103-194117.

218. Vega, C. and Abascal, J.L.F. (2011) Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.*, **13**, 19663-19688.

219. Wagner, E. (1991) Transferrin-Polycation-DNA Complexes: The Effect of Polycations on the Structure of the Complex and DNA Delivery to Cells. *Proc. Natl. Acad. Sci. U. S. A.*, **88**, 4255-4259.

220. Wagner, E., Ogris, M. and Zauner, W. (1998) Polylysine-based transfection systems utilizing receptor-mediated delivery. *Adv. Drug Delivery Rev.*, **30**, 97-113.

221. Wagner, K., Harries, D., May, S., Kahl, V., Rädler, J.O. and Ben-Shaul, a. (2000) Direct Evidence for Counterion Release upon Cationic Lipid−DNA Condensation. *Langmuir*, **16**, 303-306.

222. Wagner, K.G. and Arfmann, H.-A. (1974) Properties of Basic Amino-Acid Residues. *Eur. J. Biochem.*, **46**, 27-34.

223. Wang, D. and Lippard, S.J. (2005) Cellular processing of platinum anticancer drugs. *Nat. Rev. Drug Discovery*, **4**, 307-320.

224. Wang, D., Zhu, G., Huang, X. and Lippard, S.J. (2010) X-ray structure and mechanism of RNA polymerase II stalled at an antineoplastic monofunctional platinum-DNA adduct. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 9584-9589.

225. Wang, J., Wang, W., Kollman, P.A. and Case, D.A. (2006) Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.*, **25**, 247-260.

226. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and Testing of a General Amber Force Field. *J. Comput. Chem.*, **25**, 1157-1174.

227. Wang, Y., Zheng, M., Meng, F., Zhang, J., Peng, R. and Zhong, Z. (2011) Branched polyethylenimine derivatives with reductively cleavable periphery for safe and efficient in vitro gene transfer. *Biomacromolecules*, **12**, 1032-1040.

228. Wei, M., Cohen, S.M., Silverman, A.P. and Lippard, S.J. (2001) Effects of spectator ligands on the specific recognition of intrastrand platinum-DNA cross-links by high mobility group box and TATA-binding proteins. *J. Biol. Chem.*, **276**, 38774-38780.

229. Wiggins, P.A., van der Heijden, T., Moreno-Herrero, F., Spakowitz, A., Phillips, R., Widom, J., Dekker, C. and Nelson, P.C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature nanotechnology*, **1**, 137-141.

230. Wilce, M.C.J., Aguilar, M.-I. and Hearn, M.T.W. (1995) Physicochemical Basis of Amino Acid Hydrophobicity Scales: Evaluation of Four New Scales of Amino Acid Hydrophobicity Coefficients Derived from RP-HPLC of Peptides. *Anal. Chem.*, **67**, 1210-1219.

231. Wilson, C.J., Clegg, R.E., Leavesley, D.I. and Pearcy, M.J. (2005) Mediation of biomaterial-cell interactions by adsorbed proteins: A review. *Tissue Eng.*, **11**, 1-18.

232. Wong, K.-Y. and Montgomery Pettitt, B. (2001) A study of DNA tethered to a surface by an all-atom molecular dynamics simulation. *Theor. Chem. Acc.*, **106**, 233-235.

233. Wong, K.-Y., Vainrub, A., Powdrill, T., Hogan, M. and Pettitt, B.M. (2004) A Non-Watson–Crick Motif of Base-pairing on Surfaces for Untethered Oligonucleotides. *Mol. Sim.*, **30**, 121-129.

234. Woynarowski, J.M., Faivre, S., Herzig, M.C.S., Arnett, B., Chapman, W.G., Trevino, A.V., Raymond, E., Chaney, S.G., Vaisman, A., Varchenko, M. *et al.* (2000) Oxaliplatin-induced damage of cellular DNA. *Mol. Pharmacol.*, **58**, 920-927.

235. Wozniak, K. and Blasiak, J. (2002) Recognition and repair of DNA-cisplatin adducts. *Acta Biochim. Pol.*, **49**, 583-596.

236. Wu, Y.B., Bhattacharyya, D., King, C.L., Baskerville-Abraham, I., Huh, S.H., Boysen, G., Swenberg, J.A., Temple, B., Campbell, S.L. and Chaney, S.G. (2007) Solution structures of a DNA dodecamer duplex with and without a cisplatin 1,2-d(GG) intrastrand cross-link: Comparison with the same DNA duplex containing an oxaliplatin 1,2-d(GG) intrastrand cross-link. *Biochemistry*, **46**, 6477-6487.

237. Wu, Y.B., Pradhan, P., Havener, J., Boysen, G., Swenberg, J.A., Campbell, S.L. and Chaney, S.G. (2004) NMR solution structure of an oxaliplatin 1,2-d(GG) intrastrand cross-link in a DNA dodecamer duplex. *J. Mol. Biol.*, **341**, 1251-1269.

238. Yang, D., van Boom, S.S.G.E., Reedijk, J., van Boom, J.H. and Wang, A.H.J. (1995) Structure and Isomerization of an Intrastrand Cisplatin-Cross-Linked Octamer DNA Duplex by NMR Analysis. *Biochemistry*, **34**, 12912-12920.

239. Yao, S. and Plastaras, J.P. (1994) A molecular mechanics AMBER-type force field for modeling platinum complexes of guanine derivatives. *Inorg. Chem.*, **33**, 6061-6077.

240. Zacharias, M. (2006) Minor groove deformability of DNA: A molecular dynamics free energy simulation study. *Biophys. J.*, **91**, 882-891.

241. Zhao, X. and Johnson, J.K. (2007) Simulation of Adsorption of DNA on Carbon Nanotubes. *J. Am. Chem. Soc.*, **129**, 10438-10445.

242. Zhou, J., Thorpe, I.F., Izvekov, S. and Voth, G.A. (2007) Coarse-grained peptide modeling using a systematic multiscale approach. *Biophys. J.*, **92**, 4289-4303.

243. Ziebarth, J. and Wang, Y.M. (2009) Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.*, **97**, 1971-1983.