

**MOLECULAR DYNAMICS SIMULATIONS OF
DNA: FORCE FIELD EVALUATION AND
BACKBONE SUBSTATE DYNAMICS
IN FREE AND PROTEIN-BOUND
DNA**

by

James C. Robertson

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Medicinal Chemistry

The University of Utah

May 2016

Copyright © James C. Robertson 2016

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of _____ **James C. Robertson** _____
has been approved by the following supervisory committee members:

_____ **Thomas E. Cheatham** _____, Chair _____ **3/1/2016** _____
Date Approved

_____ **Darrell R. Davis** _____, Member _____ **3/1/2016** _____
Date Approved

_____ **Amy M. Barrios** _____, Member _____ **3/1/2016** _____
Date Approved

_____ **Eric W. Schmidt** _____, Member _____ **3/1/2016** _____
Date Approved

_____ **Julio Cesar Facelli** _____, Member _____ **3/1/2016** _____
Date Approved

and by _____ **Darrell R. Davis** _____, Chair/Dean of
the Department/College/School of _____ **Medicinal Chemistry** _____

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Computer simulations of biomolecules can provide insight into biological structure and dynamics at the atomic level. For simulations to be accurate and reliable, the underlying force field that describes that system has to be tested and assessed against experimental values. Evaluating the latest AMBER nucleic acids force fields at microsecond time scales is important for determining how updated force field parameter sets compare to earlier models and updated, competing models. The latest two improvements to the AMBER nucleic acid force field were compared to each other and previous, widely used versions. Both of the latest versions, bsc1 and OL15, showed improvement over earlier versions and reproduced many structural properties in agreement with nuclear magnetic resonance (NMR) and X-ray crystallography experiments. Having a force field that accurately models duplex DNA is important, but it is also necessary to validate protein-nucleic acid simulations. To examine this, the E2-DNA system was chosen for simulations to see if DNA backbone substates observed in the X-ray structure were reproducible in simulations. In particular, the BI/BII substates were scrutinized. The BI substate is dominant in duplex DNA but ϵ and ζ dihedrals populate both trans/gauche- (t/g-) and gauche-/trans (g-/t) conformations, the latter being the BII state. It was determined that dinucleotide steps crystallized in the BII were correspondingly highly-populated in simulations. Additionally, ϵ and ζ dihedrals were manipulated to decoy states and found to converge to native distributions on the microsecond time scale in E2-bound DNA, and in less time for free DNA. During the investigation of BII substates in E2-DNA, the surprising observation was made that BI/BII transitions are dominant modes of motion in E2-bound DNA. This inspired research into whether these modes are sequence-dependent or completely induced by the E2 protein, and if the dynamics are conserved across E2 types. Simulations of E2-DNA found that the BII state is determined by a combination of DNA sequence and E2 binding partner.

For my parents who worked hard so that I could have this opportunity.

To my wife for support along the way.

“The beginning of knowledge is the discovery of something we do not understand.”

– Frank Herbert

CONTENTS

ABSTRACT	iii
LIST OF FIGURES	viii
LIST OF TABLES	x
ACKNOWLEDGMENTS	xii
CHAPTERS	
1. INTRODUCTION	1
1.1 Overview and Background	1
1.2 DNA Backbone Substates in Free and Protein-Bound DNA	3
1.2.1 BI/BII Substates	3
1.2.2 E2-DNA	4
1.3 Summary of Research Aims and Major Findings	4
2. ASSESSING THE CURRENT STATE OF AMBER FORCE FIELD MODIFICATIONS FOR DNA	8
2.1 Chapter Notes	8
2.2 Abstract	8
2.3 Introduction	9
2.4 Methods	11
2.5 Results and Discussion	13
2.5.1 DDD	13
2.5.2 1FZX	18
2.5.3 1SK5	18
2.5.4 3GGI	19
2.5.5 1I0T	19
2.6 Conclusion	20
2.7 Acknowledgments	20
3. DNA BACKBONE BI/BII DISTRIBUTION AND DYNAMICS IN E2 PROTEIN-BOUND ENVIRONMENT DETERMINED BY MOLECULAR DYNAMICS SIMULATIONS	59
3.1 Chapter Notes	59
3.2 Abstract	59
3.3 Introduction	60
3.4 Methods	62
3.4.1 Native State Molecular Dynamics Simulations	62
3.4.2 Decoy State Molecular Dynamics Simulations	64

3.4.3	BI/BII Restraints at C ₉ pG ₁₀ pT ₁₁ Steps	64
3.4.4	Analysis of Trajectories and DNA Structures	65
3.5	Results and Discussion	65
3.5.1	Convergence of Percent BII for Native and Decoy Simulations	65
3.5.2	Percent BII Distribution	66
3.5.3	Frequency and Lifetime of BII States	69
3.5.4	DNA Dynamics in Free DNA vs. E2-Bound DNA	70
3.6	Conclusion	71
4.	THE ROLE OF BI/BII TRANSITIONS IN E2-DNA RECOGNITION	89
4.1	Abstract	89
4.2	Introduction	89
4.3	Methods	92
4.3.1	Preparing Systems for Simulations	92
4.3.2	Analysis and Visualization	93
4.4	Results and Discussion	96
4.4.1	Overview of Dynamics and E2-DNA Interactions	96
4.4.2	DNA Bending and Minor Groove Widths	98
4.4.3	BI/BII Backbone Substate Dynamics	100
4.5	Conclusion	103
5.	CONCLUSION	121
5.1	Significance	121
5.2	Future Directions	122
	REFERENCES	123
	CURRICULUM VITAE	132

LIST OF FIGURES

1.1 DNA dihedral angles	6
1.2 E2-DNA and internal 12 base pair sequences with consensus half-sites and ACGT/AATT spacer DNA	7
2.1 Recent history of force field development in AMBER.	30
2.2 DDD RMSDs, average WC hydrogen bonding, and RMSFs.	31
2.3 Comparison of average structures extracted from the full aggregated trajectory of DDD.	32
2.4 DDD twist and propeller values.	33
2.5 DDD base pair parameters.	34
2.6 DDD base pair step parameters.	35
2.7 DDD helical parameters.	36
2.8 DDD groove widths.	37
2.9 DDD dihedral histograms.	38
2.10 DDD BII populations per base pair step.	39
2.11 DDD PCA histograms	40
2.12 RMSD histograms for 1FZX, 1I0T, 1SK5, and 3GGI.	41
2.13 Twist and propeller values for 1FZX, 1I0T, 1SK5, and 3GGI.	42
2.14 1FZX base pair parameters.	43
2.15 1FZX base pair step parameters.	44
2.16 1FZX helical parameters.	45
2.17 1FZX groove widths.	46
2.18 1SK5 base pair parameters.	47
2.19 1SK5 base pair step parameters.	48
2.20 1SK5 helical parameters.	49
2.21 1SK5 groove widths.	50
2.22 3GGI base pair parameters.	51
2.23 3GGI base pair step parameters.	52
2.24 3GGI helical parameters.	53
2.25 3GGI groove widths.	54

2.26	1I0T base pair parameters.	55
2.27	1I0T base pair step parameters.	56
2.28	1I0T helical parameters.	57
2.29	1I0T groove widths.	58
3.1	Convergence of percent BII over time.	78
3.2	Convergence of percent BII per dinucleotide step over time for protein-DNA complex.	79
3.3	Helical parameters for free DNA and complex DNA: native and decoys.	80
3.4	Groove widths for free DNA and complex DNA: native and decoys.	81
3.5	Convergence of percent BII for over time for 4 steps.	82
3.6	Root mean square deviations (RMSD) measurements.	83
3.7	Percent BII distributions in free DNA and E2-bound DNA.	84
3.8	Percent BII in facing and consecutive steps.	85
3.9	Correlation between percent BII and BII frequency.	86
3.10	Principal component projections calculated from DNA heavy atoms.	87
3.11	Force fields and BI/BII states.	88
4.1	X-ray structures of E2-DNA complexes and DNA sequences.	111
4.2	RMSD over time for molecular dynamics trajectories.	112
4.3	Free E2 RMSD, RMSF, and PCA from molecular dynamics trajectories.	113
4.4	RMSF of free DNA, E2-bound DNA, and DNA-bound E2.	114
4.5	BPV E2 loop regions hug DNA minor grooves.	115
4.6	RMSD of $\alpha 1$ helix between free E2 and DNA-bound E2.	116
4.7	DNA major and minor groove widths.	117
4.8	Convergence of fraction BII in free and E2-bound DNA.	118
4.9	Fraction BII in free and E2-bound DNA.	119
4.10	BI/BII dynamics in E2-bound DNA are sequence-dependent.	120

LIST OF TABLES

2.1 List of MD simulations and simulation conditions. All systems were performed with net-neutral ion concentration using the Joung-Cheatham ion model and with TIP3P and OPC water models.	22
2.2 Average structural parameters for the DDD system with $\epsilon_{\zeta_{OL1}\chi_{OL4}}$ parameters. The two terminal base pairs on each side of the DDD were excluded for the average value calculation. NMR and X-ray values obtained from Reference [78,79]. RMSD values calculated using an average structure over 10 μ s of simulation time.	23
2.3 Average and standard deviation (SD) of structural parameters for the DDD system. The two terminal base pairs on each side of the DDD were excluded for the average value calculation. NMR and X-ray values obtained from Reference [78,79]. RMSD values calculated using an average structure over 10 μ s of simulation time.	24
2.4 Comparison of average BII fraction from simulations. Because 1I0T is Z-DNA, the 1I0T values are for ZI rather than BII.	25
2.5 Average and standard deviation (SD) of structural parameters for the 1FZX system. The terminal base pairs on each side of the 1FZX were excluded for the average value calculation. NMR values were obtained from Reference [80]. RMSD values calculated using an average structure over 10 μ s of simulation time.	26
2.6 Average and standard deviation (SD) of structural parameters for the 1SK5 system. The terminal base pairs on each side of the 1SK5 were excluded for the average value calculation. RMSD values calculated using an average structure over 10 μ s of simulation time.	27
2.7 Average structural parameters for the 3GGI system. The terminal base pairs on each side of the 3GGI were excluded for the average value calculation. RMSD values calculated using an average structure over 10 μ s of simulation time.	28
2.8 Average structural parameters for the 1I0T system. The terminal base pairs on each side of the 1I0T were excluded for the average value calculation. X-ray values were obtained from Reference [81]. RMSD values calculated using an average structure over 10 μ s of simulation time.	29
3.1 Epsilon and zeta from average structures: free DNA. Epsilon and zeta dihedrals for native, decoy1, and decoy2. Each is from the zeta-v-epsilon data in Figure 2.1, i.e. the epsilon and zeta were measured from an average structure. The epsilon for free DNA shows nearly complete convergence but the complex values differ. Data shown for W strand only.	73

3.2 Epsilon and zeta from average structures: E2-bound DNA. Epsilon and zeta dihedrals for native, decoy1, and decoy2. Each is from the zeta-vs.-epsilon data in Figure 2.1, i.e. the epsilon and zeta were measured from an average structure. The epsilon for free DNA shows nearly complete convergence but the complex values differ. Data shown for W strand only.	74
3.3 Hydrogen Bonding Interactions from E2-DNA Simulations and X-ray Reference Structure. Hydrogen bonds between two conserved residues in the recognition helix and DNA are shown. The H-bond fraction is based on the presence of hydrogen bond within angle cutoff of 135 ° and heavy atom distance of 3 Å.	75
3.4 Overall bend (in degrees) and percent BII from the internal 12 bp of unrestrained E2-bound DNA and free DNA, as well as free DNA with restraints at various combinations of steps 9 and 10. For example, CpGpT-BI means that both CpG and GpT steps on both strands were restrained to BI state, CpGpT_W-BI means that those steps were restrained on only a single strand etc.	76
3.5 Average lifetime of BII states per dinucleotide step, the maximum BII lifetime per step, and the frequency of BII states per step for free DNA and E2-bound DNA. Values are averages of three simulation copies with standard deviations.	77
4.1 Descriptions of each system simulated. All molecular dynamics were run in net-neutralizing ion conditions with Joung-Cheatham ions [60] and fully solvated with TIP3P water model [54]. Three copies of each were run. Spacer sequences with an asterisk are the sequences that were co-crystallized with the E2.	105
4.2 RMSD of average structures from simulations with X-ray structures as reference (Å). The average structure was calculated from an aggregated trajectory of three independent copies of each system, totaling 9000 ns of sampling per average structure calculation. RMSD was calculated over nucleic acid and protein backbone atoms.	106
4.3 Center of mass distance between the disordered regions between $\beta 2$ and $\beta 3$ sheets (Å).	107
4.4 Total helical bending in free and E2-bound DNA. Bending was calculated for the internal 12 base pairs of each system using Curves+ [70].	108
4.5 Average minor groove widths for the spacer region of free and E2-bound DNA.	109
4.6 BI/BII transitions in the spacer region of free and E2-bound DNA from principal component analysis pseudotrajectories. Principal component analysis was calculated on all heavy atoms of DNA for 20 modes. n/d means that BI/BII transitions were not observed in any mode.	110

ACKNOWLEDGMENTS

This research was enabled by computational resources from the Blue Waters sustained-petascale computing project (NSF OCI 07-25070 and PRAC OCI-1036208), the NSF Extreme Science and Engineering Discovery Environment (XSEDE, OCI-10535575), and allocation MCA01S027P from the Center for High Performance Computing at the University of Utah.

CHAPTER 1

INTRODUCTION

1.1 Overview and Background

Understanding the structure of biological molecules has been of primary interest to scientists throughout contemporary history. The discovery of the DNA double helix was one of the major scientific breakthroughs of the last century and intensive efforts continue to elucidate molecular structures through X-ray crystallography, NMR, and cryo-EM [1–3]. Molecular dynamics (MD) simulations can complement these experimental approaches by simulating biomolecular dynamics on the submillisecond time scale [4,5] and provide insight to structure, dynamics, and energetics of biomolecules. Simulations can be routinely run into the microsecond time scale, for example a 25 kDa protein like dihydrofolate reductase can achieve sampling of over 100 ns/day with explicit solvent and ions, ~25,000 atoms, on graphical processing units (GPUs) running AMBER software [6,7]. This is remarkable considered that 20 years earlier, simulation times for even smaller systems was 1000 times slower [8]. These advances in hardware and simulation methods have led to renewed interest and optimism that suggest computer simulations will continue play an important role in researching biological structure, dynamics, and function.

Molecular dynamics are based on Newtonian mechanics to simulate the motions of particles and provide information about how the system evolves over time [9–11]. A potential energy function comprised of terms for bonds, angles, dihedrals, and nonbonded van der Waals and electrostatics, along with an accompanying parameter set, is known as a force field. The pairwise additive force field used in AMBER [12] is shown below

$$V(r) = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} V_n(1 + \cos[n\phi - \gamma]) + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (1.1)$$

with a harmonic expression describing the bonds and angles, a Fourier expression for dihedrals, a Lennard-Jones term for van der Waals, and a Coloumbic term for electrostatics.

The bonding term contains the force constant k_b (kcal/mol) and equilibrium bond length r_0 (Å); parameters for the angle terms are the force constant k_θ (kcal/mol/radian²) and the equilibrium angle θ_0 ; torsions are described by the barrier height V_n (kcal/mol), the phase γ (degrees), and the periodicity n . The nonbonded terms describe the van der Waals and electrostatic interactions. The A term is defined as ϵR_{min}^{12} , where ϵ (kcal/mol) is the depth of the potential well that separates atoms i and j , and R_{min} is the distance (Å) separating atoms i and j at ϵ . Similarly, B is $2\epsilon R_{min}^6$, where ϵ and R_{min} have the same values as in A. The electrostatic terms contain partial charges q_i and q_j . A detailed discussion of AMBER parameters can be found in the AMBER15 manual [12].

Newton’s equations of motion use the potential energy to calculate forces, which are integrated to update the coordinates for the next step, and the process repeats, to propagate the system over time. Time steps are limited by the highest frequency motions in the system, but advances like hydrogen mass repartitioning allow a 4 fs time step [13]. Force field development for biomolecules is an active and progressing area of research. As simulations are expanded to longer time scales, weaknesses in the force field are often revealed and reparameterized. Increasing computational power also allows for higher level quantum mechanical calculations used in parameter development. Limitations in molecular dynamics are generally due to insufficient sampling or inaccuracies in the force field. Progress continues to be made in overcoming both, leading to improved simulation results.

Current methods in molecular dynamics simulations provide insight to systems of up to millions of atoms and on time scales reaching to the millisecond. Molecular processes such as predicting folded protein structures [14,15], converging DNA dynamics [5,16], and observing drug molecules find their target binding site [17] are notable examples of recent applications and demonstrate the role of simulations in providing biological insight.

Molecular dynamics studies of nucleic acids, specifically DNA, demonstrated stability at the nanosecond time scale in 1995 [18], and the microsecond time scale in 2007 [19]. With extended time scales came new understanding of weaknesses in the underlying models. The inclusion of particle mesh Ewald (PME) methods for the treatment of long-range electrostatics was an important fix to a problem obscured by low sampling times, and led to reliable simulations > 1 ns. More recently, the challenge has been accurately representing dihedrals in the backbone [20–24], simulating the crystal environment [25], understanding sequence specific behavior [26–29], and understanding terminal base pair fraying [30]. Despite these challenges, state of the art B-form duplex DNA simulations show strong agreement with experiment, are stable over long time scales, and inform new

views of biological structure and function.

1.2 DNA Backbone Substates in Free and Protein-Bound DNA

1.2.1 BI/BII Substates

The DNA double helix is as familiar as any image in the biological sciences. The structure appears simple and elegant, and the genetic information encoded within it is awe inspiring. The dynamics of DNA, however, are complex and polymorphic [31, 32]. Duplex DNA bends, in plane and out of plane, the strands twist like an extended sponge, and DNA breathes, the base pairs open, internally and at the termini.

The DNA backbone, consisting of α , β , γ , δ , ϵ , and ζ dihedrals (Figure 1.1), contributes to these dynamics. The α and γ dihedrals are dynamically coupled, and posited to be involved in protein-DNA complex formation [33]. The α and γ dihedral parameters were also targeted for force field improvement in the bsc0 parameter set because the previous force field overpopulated α in $g+$ and γ in t conformations [20]. The alternative Z-DNA conformation also undergoes backbone transitions, largely through the β dihedral [24]. The δ dihedral is involved in north and south pucker conformations. Finally, the ϵ and ζ dihedrals populate BI ($\epsilon=t$ and $\zeta=g-$) and BII ($\epsilon=g-$ and $\zeta=t$) states [34].

A structural survey determined that the BII state exists in both free and protein-bound DNA, unlike the unusual α and γ states that only appear in protein-bound structures [35]. This strengthened the idea that BII states were not due to protein-induced conformations or crystal packing, and that BII could be observed in solution, though this was difficult because NMR measurements could not directly confirm the BII state. More recent work has confirmed BII in free DNA in solution, due to better characterization and quantification of BII states [36–38].

The dynamics have been implicated in specific recognition of protein-binding partners [39]. Typically, specific protein-DNA interactions are due to the interaction of conserved protein residues with the chemical signature presented by the faces of DNA base pairs, this is referred to as direct readout and plays a major role in protein-DNA interactions [40]. Shape and dynamics of DNA, in bases not directly contacted by the protein, have been found to play a role as well, this is referred to as indirect readout [41]. This mechanism is less well understood than direct readout, but BI/BII dynamics were shown to be important for the recognition of NF- κ B to DNA binding [39]. A picture has emerged of DNA as a conformationally diverse and active molecule, that bends, twists, breaths, and rearranges its backbone conformation upon protein binding.

1.2.2 E2-DNA

To explore BI/BII dynamics in protein-bound DNA we needed a system that had experimental data to compare to and was well-studied so we could see how our results compared to previous reports. The E2 family of proteins are master regulators in the papillomaviruses, and most importantly for this work, they bind DNA in a sequence specific manner [42–44]. E2 proteins recognize ACCGN₄CGGT, where N are any 4 bases called the spacer region, and on either side are two consensus half-sites, ACCG and CGGT, see Figure 1.2. X-ray crystal structures exist for bovine papillomavirus type 1 (BPV) E2 bound to DNA with spacer sequence ACGT [43], human papillomavirus (HPV) type 6 E2 bound to DNA with spacer AATT [45], and HPV 18 E2 bound to DNA with spacer AATT [46]. DNA binding studies determined that the HPV E2 proteins preferentially bind DNA with the AATT spacer over the ACGT spacer, and that BPV E2 showed no preference for AATT over ACGT [45–48].

E2-bound DNA smoothly bends around the protein and makes direct contact with E2 at the two consensus half-sites. E2 is a homodimer, preformed prior to DNA binding, with a β barrel fold decorated with α 1 and α 2 helices. The α 1 helices insert into consecutive major grooves of DNA, where they make base-specific, and direct, contact with DNA. The DNA spacer sequence, between the consensus half-sites, does not make direct contact with DNA. The E2-DNA system was chosen for the studies in this work because it was well characterized, the X-ray crystal structure of BPV E2-DNA was high-resolution (1.7 Å), the crystal structure contained dinucleotides in the BII substate, and the differential binding to different DNA spacer sequences was poorly understood.

1.3 Summary of Research Aims and Major Findings

The work presented in this dissertation builds on the improvements in simulation methods and force fields that enable routine and reproducible simulations of nucleic acids and extends them to investigate protein-DNA interactions. A crucial component of any simulation is the underlying model, and Chapter 2 is an assessment of the two latest AMBER DNA force fields. The force fields are shown to be an improvement over earlier models and demonstrate close agreement with NMR reference structures. Evaluating and validating the model in duplex DNA is important, but determining how well the model performs in more complex systems adds interest. The DNA backbone BI/BII substates were investigated in a protein-bound environment to determine whether dinucleotide steps crystallized in the BII state could be reproduced in high populations by the simulations, and to determine

BII convergence times in a protein-bound system. High BII populations were observed in dinucleotide steps that corresponded to steps crystallized in the BII state, and BII substates converged at $\sim 1 \mu\text{s}$ time scale. BI/BII backbone dynamics were observed to be dominant modes of motion in the protein-DNA complex (a papillomavirus E2 transcription factor-DNA complex), so simulations were run with two DNA sequences to determine whether the motions were E2 protein-induced or DNA sequence-dependent. The BI/BII transitions were found to depend on both E2 binding and on sequence. The BII populations differed between free and E2-bound DNA, and in E2-bound DNA the BII populations differed between the two sequences. In addition, the BI/BII transitions were found to dominate in one sequence, with an ACGT spacer, and be almost entirely absent in another, with an AATT spacer. The biological implications of sequence-dependent backbone substate dynamics in E2-bound DNA are discussed.

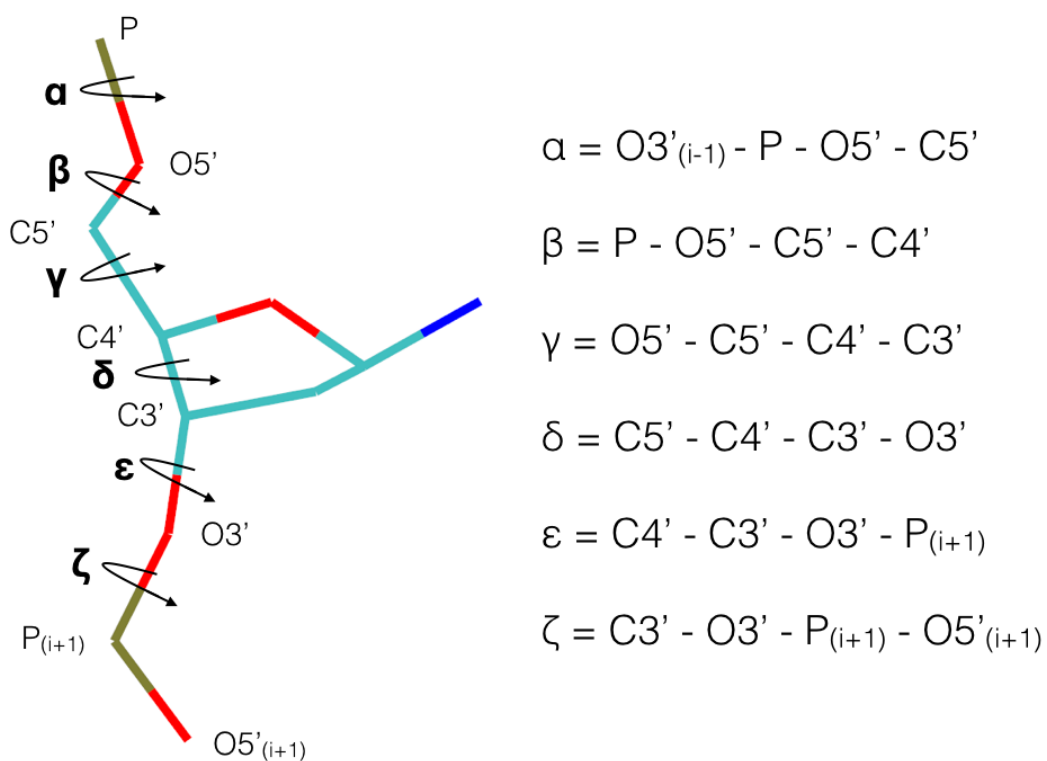
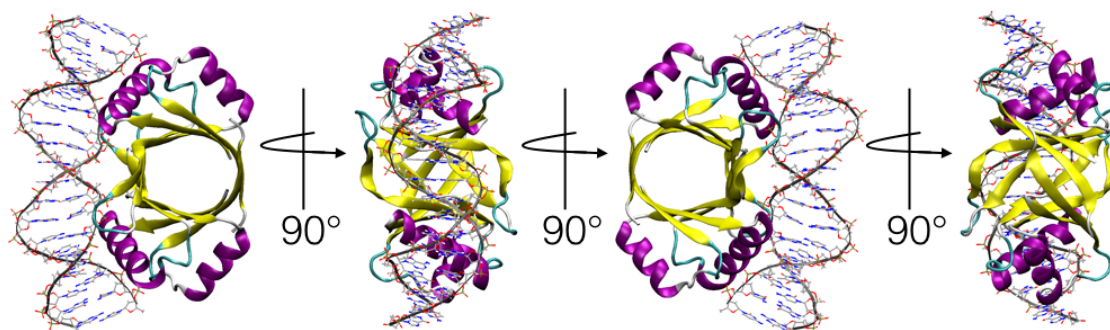


Figure 1.1: DNA dihedral angles. The DNA sugar phosphate backbone is shown with backbone atoms and dihedrals labeled. Chain direction is 5' to 3', top to bottom. A nucleotide unit is from P to O3'. Phosphorus are colored tan, oxygen red, carbon cyan, and nitrogen blue.



ACGT Spacer 5'- A C C G A C G T C G G T -3'

AATT Spacer 5'- A C C G A A T T C G G T -3'

Figure 1.2: E2-DNA and internal 12 base pair sequences with consensus half-sites and ACGT/AATT spacer DNA. The E2-DNA complex is shown from four perspectives. The molecular representations show the E2 homodimer and 18 base pairs of duplex DNA. The sequences below show the DNA recognition sequences with two different spacers, ACGT and AATT, which are between consensus half-sites. α 1 helices insert into consecutive major grooves of DNA, where the consensus half-sites are located. The spacer region of DNA lies between the consensus half-sites, in the region where the minor groove faces E2. The model shown is from the BPV X-ray crystal structure (PDB ID 2BOP). The protein is in newcartoon representation and colored by secondary structure; α helices are purple, β sheets yellow, turns cyan, and loops white. DNA backbone atoms are represented as a gray ribbon, and all atoms represented as lines. Phosphorus are tan, oxygen red, carbon silver, and nitrogen blue.

CHAPTER 2

ASSESSING THE CURRENT STATE OF AMBER FORCE FIELD MODIFICATIONS FOR DNA

2.1 Chapter Notes

This chapter has been adapted and submitted as:

R. Galindo-Murillo, J.C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, T.E. Cheatham III. Assessing the Current State of AMBER Force Field Modifications for DNA. *J. Chem. Theory Comput.*, Under Review

R. Galindo-Murillo, J.C. Robertson, and T.E. Cheatham III designed the experiments. R. Galindo-Murillo ran the simulations. R. Galindo-Murillo and J.C. Robertson performed the analysis and wrote the manuscript. M. Zgarbová, J. Šponer, M. Otyepka, and P. Jurečka revised the manuscript.

2.2 Abstract

Molecular dynamics (MD) simulations have made enormous advances due to the availability of optimized MD software and access to significant computational power including GPU multicore computing engines. This has led researchers to routinely extend the sampling time to the microsecond level and beyond. The extended sampling time has allowed the community to discover deficiencies and overcome problems with the force fields. Accuracy of the force fields is a key component, along with sampling, towards being able to generate accurate and stable structures of biopolymers. The AMBER force field for nucleic acids has been used extensively since the 1990s and multiple artifacts have been discovered, corrected, and reassessed by different research groups. We present a direct comparison of two of the most recent and state of the art AMBER force field modifications, bsc1 and OL15, that focus on accurate modeling of double-stranded DNA. After extensive MD simulations with five test cases and two different water models, we conclude that both modifications are a

remarkable improvement over the previous bsc0 force field. Both force field modifications show better agreement when compared to experimental structures. To assure convergence, one of the test systems was simulated using 100 independent MD simulations, each extended to at least 11 μ s, with the independent MD simulations concatenated a single one-millisecond long trajectory file. The bsc1 and OL15 force fields render average structures that are less than 1 Å in deviation from the average experimental structures, with only small and point differences between the two. Any future research involving double-stranded DNA simulations using the AMBER force field should switch to the bsc1 or OL15 modifications.

2.3 Introduction

The ability to simulate nucleic acids, especially DNA at atomic resolution over biologically relevant time scales, has led to new insight to the richness and complexity of dynamics on the submillisecond time scale. The ability to achieve advances in DNA simulations has been driven by improvements to hardware, simulation methods, and continual development and optimization of the underlying physical model that describes the system, the force field. As improvements to force fields are made it is important to understand the strengths and weaknesses of the models, and evaluate them with simulations that reach well into the microsecond time scale.

Convergence and reproducibility of the structure and dynamics of the internal portions of a DNA helix on the 1-5 μ s time scale were recently demonstrated [5, 16] and microsecond-length simulations of fully solvated, atomistic duplex DNA are now routine [6, 7], so the previous limitations of not being able to achieve enough sampling have been eliminated (for modestly sized systems). This provides an opportunity to focus on developing as accurate a force field as possible. The efforts of multiple research groups have led to two force fields that deserve close scrutiny to determine how well each reproduce experimental observables so that users in the community can determine the best force field to use to suit their needs. If the past experience serves, it is expected that longer simulations and usage by a larger community will inevitably uncover further deficiencies in the current force fields: nevertheless, it is prudent to learn as much about the current models as possible. Even on the microsecond time scale, simulations reveal limitations of the force fields that can provide insight into where to focus efforts for even further improvement.

Since the original parm94 force field [49, 50], was updated to parm98 [51] and parm99 [52], the development of new parameters has progressed along two primary paths (Figure 2.1). One fork follows the efforts led by the group at the Barcelona Supercomputing Center (BSC).

The bsc0 [20] modifications published in 2007 improved upon parm99 by updating the α and γ dihedrals. This corrected α overpopulated in g+ and γ overpopulated in t conformations. The most recent bsc1 [21] force field developed by this group, released in 2015, includes the bsc0 modifications and additional modifications to the sugar pucker, the χ glycosidic torsion, and the ϵ and ζ dihedrals. The other fork follows the collective work from research groups in Czech Republic, and include "OL" in the name based on the territory of Olomouc in the Czech Republic, where these parameters were generated. The number represents the version for that particular parameter, for example, χ OL4 represents the 4th version for the χ dihedral. Development along this path progressed in an incremental fashion as improvements from parm99 were made to the χ glycosidic torsion with χ OL4 [53]. The next OL improvements came when updates were made to the ϵ and ζ dihedrals [23]. At this point, AMBER15 was released and the recommended DNA force field was the combination bsc0 + χ OL4 + ϵ/ζ OL1. These parameters resulted in improvements over bsc0 by increasing populations of BII, increasing twist, and reducing major groove width. Since then, the β dihedral has been reparameterized to improve ZI and ZII substates in Z-DNA and the new ff-nucleic-OL15 parameter set, a combination of parm99 + bsc0 + χ OL4 + ϵ/ζ OL1 + β OL1, has been released [24]. The OL15 version thus represents, compared to the original 1995 Cornell et al. force field [49], a complete one-dimensional (uncoupled) reparameterization of all the DNA dihedral backbone potentials and may reach the accuracy limits obtainable by tuning the dihedral parameters of the Cornell et al. parameters.

We present an exhaustive comparison using relatively long time scale MD simulations of recent AMBER force field modifications developed to improve the accuracy of molecular dynamics simulations of double-stranded DNA. The systems tested include two with solution NMR reference structures, which allows comparison to experiment. The two latest, and most advanced versions of the force field, bsc1 and ff-nucleic-OL15, inspired these in-depth comparisons. The different approaches to developing the force fields, and the resulting parameter sets, demonstrate that an accurate force field is a difficult problem to solve. The updated parameters do show improvement over previous versions and, in general, demonstrate increased agreement with experiment. Nucleic acids and protein simulations are routinely performed in explicit solvent, mainly the TIP3P [54], SPC/E [55] and TIP4P/Ewald [56] water models. The Optimal Point Charge "OPC" water model developed in 2014 [57] has been shown to optimize simulations for certain systems [58]. This model was parameterized to capture the charge asymmetry of a water molecule, and indeed, it shows improvement in simulating bulk water properties [59]. We included the force field

comparison with the commonly used TIP3P water model and the OPC water model and detected a slight overall improvement in all of the tested systems when using OPC, although, with a considerable hit in the simulation speed ($\sim 30\%$ slower). Our extended simulations (around 7 ms of combined sampling time on 5 tested systems) confirm improvements of simulated DNA with both bsc1 and OL15 force fields when compared to high-resolution experimental X-ray structures and NMR spectroscopy structures. Both bsc1 and OL15 perform in a remarkable similar way and only very detailed, point differences were detected.

2.4 Methods

The systems considered to evaluate the different force field modifications are presented in Table 2.1. All systems were modeled using three different force fields: the current parm99 with the bsc0 correction (referred as **bsc0**), the recent **bsc1**, and the **OL15** (combination of ff99bsc0 with modifications ϵ/ζ OL1, χ OL4 and β OL1). Additionally, we include information on simulations of the test systems using the ϵ/ζ OL1 χ OL4 (without the β dihedral adjustment Table 2.2) modifications, which has been in use since 2013 by a large community of users who may want to better understand the implications for their unpublished work of the force field comparisons.

Crystallographic water molecules and counter ions were removed in each case. This starting structure was then solvated with either the TIP3P [54] or the OPC [57] water model in a truncated octahedral box using a 10 Å buffer distance between the solute and the edges of the box. Sodium counter ions were added to neutralize the charge using the Joung-Cheatham [60, 61] model and an excess of NaCl was added to achieve a final concentration of ~ 200 mM. Ten individual copies were created for systems 1SK5, 3GGI, and 1I0T, each copy with a total simulation time of 11 μ s. For systems 1BNA and 1FZX, 100 individual copies were built, each copy with a total simulation time of 11 μ s for 1BNA and 6 μ s for 1FZX.. In all copies of each system, the ions were randomized using CPPTRAJ: a random water molecule was swapped for an ion at least 4.0 Å from each other and no closer than 6 Å from solute atoms. Initial equilibration for each copy was achieved using incremental minimization steps where the solute was remained fixed with a harmonic restraint of 5 kcal/mol-Å² for 1 ns. This restraint was decreased in subsequent 1 ns long simulations until the full system was unrestrained. Equilibration was performed using a time step value of 0.001 fs. Production simulations were run in the NTP ensemble at 300K using Langevin dynamics for temperature control [62, 63]. The SHAKE methodology was used to restrain hydrogen atoms (tolerance of 0.0000001). Hydrogen mass repartitioning was

used in all simulations allowing a time step of 0.004 ps [13]. Periodic boundary conditions were used and long-range electrostatics were treated using the Ewald methodology with a cutoff value of 10 Å and default parameters [64, 65]. Aggregated trajectories used to perform the analysis were created deleting the first 1000 ns of sampling time for each copy and concatenating the remaining frames into a single trajectory file. Example of the CPPTRAJ analysis script for the DDD system is available in the supporting information. In addition to the scripts in the Supporting Information, the topologies, the raw (solvent and ion stripped and aggregated) trajectories, and all of the analysis files are available for download at <http://www.amber.utah.edu/FF-DNA-bsc1-OL15/>.

All simulations were performed using the CPU and GPU version of pmemd as available in Amber14 [6, 7, 66–68]. Analysis was performed using CPPTRAJ v16 [69] and Curves+ [70]. Molecular graphics were rendered in VMD [71] and principal component pseudo-trajectories visualized with the help of the Normal Mode Wizard plugin [72]. An example of the CPPTRAJ analysis script for the DDD system is shown below.

```
#
# Example for the d(CGCGAATTCGCG) DNA structure (12 base pairs)
# Input the simulation topology and simulation trajectory file (NETcdf format)
# Input reference PDB (downloaded from the PDB for each system)
# Read RMSd using the reference for each frame,
# including all residues, no hydrogens
# Read RMSd using the reference for each frame,
# only internal base pairs, no hydrogens
# Calculate atomic fluctuations
#
parm vacuo.prmtop
trajin aggregate.nc 1 last 10
check @N1,N2,P,OP1,OP2,C1',C2',C3',C4',O3',O4' skipbadframes silent
reference ddd-nmr-average.pdb
rms reference :1-24&!@H= out data/rms_ref.dat time 200
rms reference :3-10,15-22&!@H= out data/rms_ref-internal.dat time 20
atomicfluct out data/fluct.byatom_bfactor.dat :1-24 bfactor
atomicfluct out data/fluct.byres_bfactor.dat :1-24 byres bfactor
atomicfluct out data/fluct.byatom.dat :1-24
```

```

atomicfluct out data/fluct.byres.dat :1-24 byres
#
# Calculate interhelical and intrahelical parameters,
# internal base pairs only, using the
# NAstruct command, output to files inside data folder.
#
nastruct NA1 reference resrange 3-10,15-22
run
writedata data/major.dat NA1[major]
writedata data/minor.dat NA1[minor]
writedata data/helical-twist.dat NA1[htwist]
writedata data/helical-inclination.dat NA1[incl]
writedata data/hb.dat NA1[hb]
writedata data/shear.dat NA1[shear]
writedata data/buckle.dat NA1[buckle]
writedata data/stretch.dat NA1[stretch]
writedata data/stagger.dat NA1[stagger]
writedata data/prop.dat NA1[prop]
writedata data/opening.dat NA1[open]
writedata data/shift.dat NA1[shift]
writedata data/tilt.dat NA1[tilt]
writedata data/slide.dat NA1[slide]
writedata data/rise.dat NA1[rise]
writedata data/roll.dat NA1[roll]
writedata data/twist.dat NA1[twist]

```

2.5 Results and Discussion

2.5.1 DDD

The Drew-Dickerson dodecamer (DDD) is the defacto benchmark system regarding B-form DNA duplex structure and dynamics due to the availability of very high-resolution NMR data in the solution phase, and therefore DDD was commonly studied as new force field parameters were developed and optimized [20, 21, 23, 24, 53].

The root-mean-square-deviation of the eight internal base pairs of the DDD sequence (neglecting the two terminal base pairs on each end that tend to fray on the microsecond

time scale), using the NMR structure as a reference and compared to converged average structures from the MD simulations, shows that in general both bsc1 and OL15 have root-mean-square-deviation (RMSD) values that are closer to the reference structure than bsc0. This was true for both water models (Figure 2.2 and Table 2.3). The sub 1 Å deviation in all cases is remarkable, especially considering that the instantaneous deviations on the nanosecond time scale are considerably larger. Visual inspection of the trajectory files show stable duplex structures throughout the entire simulation time, and consistent with previous MD simulation reports, the observation of transient fraying events of the first and second base pairs on either side of the DNA chain. These fraying events are present in the short-lived bumps in the RMSD plots [5] and have been well characterized [30]. As noted previously, terminal base pair fraying can cause long-lived noncanonical structure conformation that affects the end results (See, e.g., Figure 6 of [23] and references [30] and [73]). The amount of fraying (separation of strands) for each force field modification was calculated based on the average number of Watson-Crick (WC) bonds between base pairs (Figure 2.2B). From the figure we can see that the internal section of AATT residues show a steady count of 2 WC bonds and that G₄C₂₁ and C₉G₁₆ base pairs average three WC hydrogen bonds. This is consistent for the three force fields and expected based on low RMSD values and visual inspection showing stable duplexes. Maintenance of the internal base pairs is expected since internal base pair opening times are on the order milliseconds. The terminal base pairs, however, show decreased average WC hydrogen bonds, instead of a canonical value of three. This is caused by multiple fraying events that break canonical hydrogen bonds and lead to distorted conformations such as trans WC/sugar edge, stacked, and so forth. [30]. Although fraying is more frequent with bsc0 and least frequent in bsc1, it is difficult to ascribe less frequent terminal base pair opening to "better" behavior since the characteristics and frequency of terminal base pair openings are not well characterized on these rapid time scales. As fraying tends to distort the canonical structure, for MD simulations on time scales that are currently routine (ns to μ s), better representation of the expected structure can be expected from simulations that fray less. The fraying frequency trend with the various force fields is observed regardless of the water model applied. However, for the TIP3P water model, more variation is detected even up into the third base pair from each end of the DNA. As this affects the overall structure (in the absence of more complete sampling) of the backbone and the base pairing, perhaps simulations with OPC may be recommended in shorter simulations, however, this is a tradeoff due to the increased costs. The best method would be to run the MD simulations to sufficient length to converge the structure

and dynamics, which if not possible choose the force field and water model appropriate to the resources or consider inhibiting terminal base pair fraying through application of weak WC base pair restraints on the termini.

The root-mean-square-fluctuations (RMSF) representing the fluctuations above the average structure for each tested force field are presented in Figure 2.2C, and show nearly complete agreement for the inner base steps. A small increase in atomic fluctuation is detected for the bsc0 force field at the terminal base pairs, which is in agreement with the measurement of WC bonds for each base pair (as previously discussed). The more frequent fraying events present when using the TIP3P water model for both bsc1 and OL15 could partially explain why the simulations using OPC present an overall better agreement when compared to experimental structures. Overall, the structures look extremely similar when compared with the NMR experimental reference. The RMSD difference using only the heavy atoms and the eight internal base pairs are 0.74 Å, 0.52 Å, and 0.42 Å (TIP3P) and 0.69 Å, 0.50 Å, and 0.37 Å (OPC) for bsc0, bsc1, and OL15, respectively. These sub 1 Å deviations are remarkable considering the absence of experimental restraints. An overlay of these average structures is presented in Figure 2.3. A detail of dA6 with the four structures (NMR reference, bsc0, bsc1, and OL15) shows strong similarity to the reference. We notice that the bsc1 and OL15 structures are very similar and are overall closer to the experimental structures. Average values for structural, inter, and intra helical properties from our simulations using bsc0 and the optimized bsc1, and OL15 force fields using two water models for the DDD are presented alongside NMR and X-ray reference values in Table 2.3. Overall, and as already suggested by the RMSD plot, both bsc1 and OL15 force field modifications generate structures that are in a better agreement with experimental structural properties than the earlier bsc0 force field. The most notable differences from the experimental values are observed in the twist, propeller, and groove widths.

Underestimation of the twist value has been a concern since the Cornell et al. conception of the AMBER DNA force field and its posterior bsc0 modification [74]. The helical twist values predicted by both bsc1 and OL15 are on average closer to the observed experimental values (NMR and X-ray average value of 34.9°). Helical twist of the eight internal base pairs shows better agreement for OL15 in TIP3P with a value of 35.2° and 35.6° for the bsc1 modification. Regardless of the water model, and despite these small differences, both force field modifications represent improvement over bsc0 of the helical twist structural parameter of DNA. Base step detail of the twist value provides further information (Figure 2.4). The data shows that bsc0 is consistently off in almost every step compared to both

bsc1 and OL15. Symmetrical behavior of the per-base step twist value around the central step is seen, characteristic of a palindromic sequence. This symmetrical feature is obtained only in a fully converged ensemble of simulations as presented on this work. CpG steps are underestimated by bsc1, and overestimated by OL15 regardless of water model used, but closer to the reference value in both cases compared to bsc0. GpA base steps stand in very good agreement in both force fields, especially in the OPC water model where the values are less than 0.3° of the reference value. The same behavior is seen in ApA base step where the values for bsc1 and OL15 for OPC are off by 0.6° and 0.9° , respectively. For the ApT base step the bsc1 modification has a value of 33.1° and OL15 of 33.0° with TIP3P (OPC is 32.8° and 32.7° , respectively), which makes them essentially identical, yet below the NMR experimental value of 34.6° . Propeller twist in the DDD system is consistently off by $\sim 5^\circ$ for OL15 and $\sim 7^\circ$ for bsc1 (assuming the NMR reference value, Table 2.3). A per base pair comparison shows a general overestimation (Figure 2.4), especially for base pairs near the end of the DNA chain. The high propeller twist near the termini is likely due to fraying effects influencing the structure and modifying the base pairing. Improvement is also obtained in both grooves with differences of ~ 0.1 - 0.5\AA for bsc1 and OL15, regardless of water model used. This is expected after the improvements on the majority of helical parameters. Plots showing all of the DDD structural parameters are in Figures 2.5, 2.6, 2.7, and 2.8.

Backbone dihedrals distributions for DDD are displayed in Figure 2.9. Close agreement with the experimental value, and between the compared force fields, is observed for α and γ dihedrals. The OL15 modifications that are centered on ϵ/ζ , χ , and β dihedrals appear to have reduced completely a low trans population in the γ dihedral, evident in bsc0 and bsc1 (in a very low population, however). The bsc1 modification for the β dihedral is in good agreement with the NMR value of 173° and shows an increased population close to 60° in comparison with bsc0. The β dihedral representation for OL15 is also in good agreement with the experimental value and displays the low population at $\sim 130^\circ$ which was observed in previous work¹³ and aids in the BI/BII substate balance [75]. The δ dihedral, which is part of the furanose ring, presents improvement over the earlier bsc0 version in both bsc1 and OL15 although slightly over estimated by 5° and 8° , respectively. Good agreement with experimental values is also detected in the ϵ/ζ dihedrals for bsc1 and OL15, and the later shows increased populations in the g- region for ϵ and t for ζ , which corresponds to an increased percentage of the DNA BII state. The fraction of the ensemble in the BII substate was calculated by taking the $\epsilon - \zeta$ difference, frames that had $\epsilon - \zeta > 0$ were considered to

be in the BII state. The fraction of BII populations for the DDD systems is shown in Table 2.4 and displays an almost identical increment of BII population for both bsc1 and OL15 in both water models used. Per base step analysis (Figure 2.10) shows similar BII estimates for OL15 and bsc1. Improvements over bsc0 at the CpG and GpC steps at both ends of the DDD are evident although under representation of the central AT, TT, and TC is still present in both bsc1 and OL15.

To directly compare the modes of motion present in the simulations, principal component analysis (PCA) over all heavy atoms of each trajectory was calculated and is shown in Figure 2.11. The combined PC analysis consisted of calculating the covariance matrix from both trajectories of the two force fields being compared (an example of a CPPTRAJ analysis script can be found in reference [76]). This technique provides insight into the collective dynamics that are sampled during a simulation and is able to rank the contribution of each mode of motion [16]. Each of the histograms represent the projection of an individual mode of motion for each force field tested. For clarity, only the first three projections of the principal components are shown. These three projections contribute with $\sim 80\%$ of the eigenvalue, and hence $\sim 80\%$ of the overall motion of the system. Trajectories that explore equivalent dynamical processes will have overlapping PC projections. The best overlap for TIP3P systems is observed with bsc1 and OL15, which is expected since bsc1 and OL15 consist of newer and updated parameters, and based on the agreement with average properties shown in Table 2.3. However, the poor overlap between bsc0/bsc1 and bsc0/OL15 projections, led us to explore the principal modes in the individual force field trajectories and compare the top modes through visualizing the pseudo trajectories. We found that the top mode for each force field was twisting. Bsc0 and bsc1 showed agreement with the second mode, bending toward the major groove (asymmetric bend), while this motion was the third ranked mode for OL15. Again, bsc0 and bsc1 exhibited agreement with the third mode, another bending mode, but more symmetric than the previous bending. OL15 demonstrated this mode as well, but it was the second-ranked mode rather than the third. The eigenvalue fractions for the twisting mode were 0.39, 0.35, and 0.40; for the asymmetric bend the fractions were 0.21, 0.21, and 0.16; the symmetric bending mode eigenvalue fractions were 0.19, 0.19, and 0.22 for bsc0, bsc1, and OL15, respectively. The top three modes showed near equivalent dynamics in the three force fields but the ranking of the bending modes differed between OL15 and the other parameter sets.

Additional systems considered to assess the performance between bsc0, bsc1, and OL15 were chosen mainly to evaluate the capacity of these force field modifications in their ability

to reproduce experimental data, hence, we used DNA structures with the highest resolution available from the Protein Data Bank. It is important to mention that known artifacts occur in X-ray crystallography data mainly due to crystal packing, causing small but noticeable variations in the conformation; with that in mind, we also included the 1FZX model which was obtained by NMR spectroscopy. In the same way as the DDD NMR structure already mentioned, we calculated an average structure from the 10 submitted conformers available in the PDB and used this average structure as a reference. RMSD histograms of each of the tested systems is shown in Figure 2.12. Overall, and as previously detected with the DDD analysis, both bsc1 and OL15 have lower RMSD values, which translate to a better agreement on the reference experimental structure.

2.5.2 1FZX

For the case of the 1FZX system (NMR structure), both bsc1 and OL15 are within 1.3 (TIP3P) and 1.2 (OPC) Å from the reference structure and 0.5 Å between the two force field modifications. Although the agreement with experiment appears not as good as with the high-resolution DDD structure, this could be due to the fact that the 1NAJ DDD PDB structure was solved with considerably more NMR RDC and 31^P chemical shift anisotropy restraints and is arguably a higher resolution structure. With the 1FZX system, the OPC water model increases the performance of the bsc0, which is now close to bsc1 (Figure 2.12). This appears to be caused by increased twist values in the central base pairs, effectively reducing the population of structures over an RMSD value of ~ 2 Å. Propeller twist does not appear to be as influenced by the OPC water model as the helical twist (Figure 2.13). As previously discussed in the DDD case, it has been observed an underestimation of the population of BII state with bsc0. This DNA substate is increased considerable in a similar fashion for bsc1 and OL15 (Table 2.5). Structural properties for 1FZX can be found in Figures 2.14, 2.15, 2.16, and 2.17.

2.5.3 1SK5

The 1SK5 system, bsc1, and OL15 are within ~ 1.4 - 1.6 Å from the reference regardless of water model, although the OPC simulations reduced considerably the population of structures with an RMSD of ~ 3.5 Å for bsc1, effectively reducing the difference between bsc1 and OL15 to less than 0.2 Å (Figure 2.12). Twist value shows improvement over bsc0 for TT and AA steps and an over estimation of TA base steps of $\sim 4^\circ$. Propeller twist is not improved by bsc1 (Figure 2.13) and shows underestimation of TA and AT base pairs (in both water models). Bsc0 and OL15 render values very close to each other and an also

a general underestimation. Structural properties for 1SK5 can be found in Figures 2.18, 2.19, 2.20, and 2.21 as well as Table 2.6.

2.5.4 3GGI

The 3GGI system is also benefited by bsc1 and OL15 although not as much as the previously mentioned systems. The RMSD values for bsc1 and OL15 with respect to the experimental structure shows values of ~ 3.8 - 4.1 Å for TIP3P and ~ 3.8 - 4.1 Å for OPC (Figure 2.12). Bsc0 shows maximum population on 4.4 and 4.3, respectively, with a broader histogram profile, which translates in a higher appearance of structures with RMSD values higher than 4.3 Å. Twist is underpopulated in bsc0 as previously discussed with both bsc1 and OL15 in a better overall agreement with experiment. GG and GC steps are off by $\sim 5^\circ$ in both cases and with both water models (Figure 2.13). Propeller, on the other hand, is slightly over represented by bsc1 and under represented by OL15, still both force field modification are within $\sim 2^\circ$ of the observed experimental value in each base pair. BII population is almost without change with OL15 showing a slight increased fraction (Table 2.7). Structural properties for 3GGI can be found in Figures 2.22, 2.23, 2.24, and 2.25.

2.5.5 1I0T

The last system we included in the analysis is the peculiar Z-DNA structure 1I0T, which is notoriously difficult for simulations. In this particular case, the OL15 modification performs in a more adequate manner than bsc1. RMSD values on the performance of the modified force fields on this structure show modest improvement (Figure 2.12). OL15 shows increased population close to the 3.2 Å value followed by bsc0 in the same region, although with a $\sim 1\%$ less population. A maximum of 3.3 is present for the bsc1 force field. Between water models, OPC reduced a very small population of structures within 3.8 Å for the bsc1 force field but little to none influence besides that. The twist values are in good agreement with the experimental reference, showing a difference of $\sim 2^\circ$ for the central CG/GC step (Figure 2.13), which is not the case with propeller twist. The internal CG and GC values are off by ~ 5 - 6° for the three tested force fields, which helps explain the high RMSD values observed before. One of the main goals of the last torsion adjustment for the OL15 (β OL1) was to influence the ZI/ZII equilibrium [24]. Values for these different substates are presented in Table 2.8. The OL15 generates a significant increase in ZI population compared to bsc0, while bsc1 performs poorly with the Z-DNA system. This indicates that the β refinement that is entirely absent in bsc1 may significantly improve description of some noncanonical DNAs and may be potentially important for deformed structures.

Structural properties for 1I0T can be found in Figures 2.26, 2.27, 2.28, and 2.29

2.6 Conclusion

We assessed the performance of two recent modifications to the AMBER force field, designed to enhance the representation of double-stranded DNA. The individual modifications (bsc1 and OL15) reviewed here increase the overall performance of DNA simulations, and are to be considered a general force field for most of the DNA systems. It is then important to have a direct comparison of these two force fields aimed to enhance the ability to simulate DNA and be confident that the rendered simulations are adequate for research. No force field is perfect and, as experience over the years has suggested, we doubt that there will be a universal force field capable of representing the enormous structural diversity of biopolymers. Inclusion of polarization terms that allow the representation of the dynamic redistribution of electronic charge over time could lead to a more universal force field but still lack the computational efficiency of the nonpolarizable potential [77]. This work concludes that both OL15 and bsc1 force field modifications increase the accuracy of representing averaged structures of DNA compared to the earlier models (or, as we would like to comment; they are both less wrong). It is recommended that users switch to either bsc1 or OL15 force field for molecular dynamics simulations of DNA with the AMBER force field. Moreover, we would like to stress that bsc1 and OL15 render such similar structures that the difference for the majority of parameters measured are judged within tenths of an angstrom or tenths of a degree. This makes structural comparisons complicated since measuring nucleic acids properties depends heavily on the used reference system (global vs. helical axis) and the software used (Curves vs. 3DNA, to name the most common). These individual differences are to be taken into consideration when a particular research problem is planned to take advantage of the strength and weakness of either bsc1 or OL15 to better suit the simulations.

2.7 Acknowledgments

This research was enabled by NIH R-01 GM081411, NIH R-01 GM098102, and computational resources from the Blue Waters sustained-petascale computing project (National Science Foundation (NSF) OCI 07-25070 and PRAC OCI-1036208), the NSF Extreme Science and Engineering Discovery Environment (XSEDE, OCI-1053575) and allocation MCA01S027P, the Center for High Performance Computing at the University of Utah. M. Z. acknowledges support by the Grant Agency of the Czech Republic (grant no. 14-29874P).

Further funding was provided by project LO1305 of the Ministry of Education, Youth and Sports of the Czech Republic (M. O., M. Z., J.Š., P. J.)

Table 2.1: List of MD simulations and simulation conditions. All systems were performed with net-neutral ion concentration using the Joung-Cheatham ion model and with TIP3P and OPC water models.

PDB ID	Sequence	Simulation Time (aggregated)	Experimental Reference	X-Ray Res. (Å)	Type	Reference
1BNA/1NAJ	d(CGCGAATTGCGG)	1 ms	X-ray/NMR	1.9	B-DNA	[78, 79]
1FZX	d(GGC AAAAACGG)	500 μ s	NMR	N/A	B-DNA	[80]
1SK5	d(CTTTTAAAAG)	100 μ s	X-ray	0.89	B-DNA	[81]
3GGI	d(CCAGGCCTGG)	100 μ s	X-ray	0.6	B-DNA	[82]
1I0T	d(CGCGCG)	100 μ s	X-ray	1.74	Z-DNA	[83]

Table 2.2: Average structural parameters for the DDD system with $\epsilon\zeta_{OL1\chi OL4}$ parameters. The two terminal base pairs on each side of the DDD were excluded for the average value calculation. NMR and X-ray values obtained from Reference [78,79]. RMSD values calculated using an average structure over 10 μ s of simulation time.

Property of Interest	NMR		X-ray		TIP3P		OPC	
	Avg	Std Dev	Avg	Std Dev	bsc0 $\epsilon\zeta_{OL1\chi OL4}$			
	Avg	Std Dev	Avg	Std Dev	Avg	Std Dev	Avg	Std Dev
Shear (Å)	0.0	0.2	0.1	0.2	0.0	0.1	0.0	0.1
Stretch (Å)	-0.3	0.1	-0.2	0.1	0.0	0.0	0.0	0.0
Stagger (Å)	-0.1	0.1	0.1	0.2	0.0	0.1	0.0	0.1
Buckle (°)	0.0	2.9	0.2	5.9	0.0	3.7	0.0	4.0
Propeller (°)	-17.6	3.9	13.3	5.6	12.4	6.4	12.1	6.0
Opening (°)	-1.1	1.2	1.3	3.4	0.2	0.5	0.2	0.6
Shift (Å)	0.0	0.3	0.0	0.3	0.0	0.3	0.0	0.2
Tilt (°)	0.0	1.1	0.3	2.5	0.0	2.0	0.0	1.8
Slide (Å)	-0.2	0.2	0.1	0.5	-0.1	0.4	-0.2	0.3
Rise (Å)	3.2	0.2	3.3	0.1	3.3	0.0	3.3	0.0
Roll (°)	3.0	3.5	2.0	3.2	3.3	2.5	2.4	2.9
Twist (°)	35.7	1.9	34.2	5.2	35.3	1.3	35.7	1.4
X-disp (Å)	-0.8	0.3	-0.2	0.5	-0.7	0.3	-0.7	0.2
Y-disp (Å)	0.0	0.4	0.1	0.9	0.0	0.4	0.0	0.3
Hel. Rise (Å)	3.2	0.1	3.3	0.1	3.3	0.0	3.3	0.0
Hel. Incl. (°)	5.0	5.9	4.0	6.7	5.5	4.2	4.1	4.8
Tip (°)	0.0	1.9	-0.7	4.7	0.0	3.2	0.0	2.8
Hel. Twist (°)	36.0	1.7	34.6	4.8	36.3	1.4	36.6	1.5
Maj. Width (Å)	19.6	0.2	19.1	0.5	19.4	0.2	19.5	0.2
Min. Width (Å)	12.2	0.3	12.2	0.3	12.3	0.1	12.3	0.1
Pucker (°)	137.1	14.2	129.5	26.7	148.1	11.2	147.9	11.0
α (°)	298.8	6.2	299.2	6.6	290.1	3.3	291.1	3.5
β (°)	172.4	4.9	175.7	10.4	170.6	4.2	171.8	4.2
γ (°)	50.3	4.9	56.5	7.3	54.4	1.8	54.1	1.7
δ (°)	126.7	9.0	122.8	19.6	131.0	5.8	130.7	6.0
ϵ (°)	188.5	5.2	190.4	27.4	194.5	12.7	191.3	12.0
ζ (°)	257.1	12.0	251.3	37.6	249.4	15.9	251.4	14.4
χ (°)	249.2	7.9	243.9	14.7	250.4	5.0	249.9	4.7
BII %	N/A	N/A	N/A	N/A	12.7		16.2	
RMSD all (Å)	N/A	N/A	N/A	N/A	1.9	0.4	1.9	0.3
RMSD int. (Å)	N/A	N/A	N/A	N/A	1.2	0.2	1.1	0.2

Table 2.3: Average and standard deviation (SD) of structural parameters for the DDD system. The two terminal base pairs on each side of the DDD were excluded for the average value calculation. NMR and X-ray values obtained from Reference [78,79]. RMSD values calculated using an average structure over 10 μ s of simulation time.

Property of Interest	NMR		X-ray		TIP3P						OPC						
	Avg	SD	Avg	SD	bsc0	oll5	bsc1	bsc0	oll5	bsc1	bsc0	oll5	bsc1	bsc0	oll5	bsc1	
Shear (Å)	0.0	0.2	0.1	0.2	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.1
Stretch (Å)	-0.3	0.1	-0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Stagger (Å)	-0.1	0.1	0.1	0.2	0.1	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.1
Buckle (°)	0.0	2.9	0.2	5.9	0.0	4.5	0.0	3.9	0.0	4.2	0.0	4.0	0.0	4.0	0.0	4.1	4.1
Propeller (°)	-17.6	3.9	-13.3	5.6	-12.7	6.4	-12.3	6.3	-10.4	5.9	-12.6	5.7	-11.9	6.0	-9.9	5.6	5.6
Opening (°)	-1.1	1.2	1.3	3.4	0.3	0.5	0.1	0.5	-0.4	0.4	0.2	0.6	0.1	0.6	-0.5	0.4	0.4
Shift (Å)	0.0	0.3	0.0	0.3	0.0	0.2	0.0	0.3	0.0	0.4	0.0	0.4	0.0	0.2	0.0	0.3	0.3
Tilt (°)	0.0	1.1	0.3	2.5	0.0	1.5	0.0	2.1	0.0	1.8	0.0	1.5	0.0	1.8	0.0	1.8	1.8
Slide (Å)	-0.2	0.2	0.1	0.5	-0.4	0.3	0.0	0.4	-0.3	0.4	-0.5	0.3	-0.1	0.3	-0.3	0.3	0.3
Rise (Å)	3.2	0.2	3.3	0.1	3.3	0.1	3.3	0.0	3.3	0.1	3.3	0.1	3.3	0.0	3.3	0.1	0.1
Roll (°)	3.0	3.5	2.0	3.2	3.4	2.9	2.8	2.2	2.3	2.2	2.9	3.5	2.1	2.6	2.1	2.4	2.4
Twist (°)	35.7	1.9	34.2	5.2	32.9	2.9	35.2	1.5	34.6	2.2	33.4	2.2	35.6	1.5	34.7	2.1	2.1
X-disp (Å)	-0.8	0.3	-0.2	0.5	-1.5	0.4	-0.6	0.3	-0.9	0.3	-1.5	0.4	-0.6	0.2	-1.0	0.2	0.2
Y-disp (Å)	0.0	0.4	0.1	0.9	0.0	0.3	0.0	0.4	0.0	0.4	0.0	0.2	0.0	0.3	0.0	0.3	0.3
Hel. Rise (Å)	3.2	0.1	3.3	0.1	3.2	0.2	3.3	0.0	3.3	0.1	3.2	0.2	3.3	0.0	3.2	0.1	0.1
Hel. Incl. (°)	5.0	5.9	4.0	6.7	6.7	6.3	4.9	3.8	4.3	4.2	5.8	6.9	3.7	4.5	4.0	4.6	4.6
Tip (°)	0.0	1.9	-0.7	4.7	0.0	2.6	0.0	3.3	0.0	3.0	0.0	2.6	0.0	2.9	0.0	2.8	2.8
Hel. Twist (°)	36.0	1.7	34.6	4.8	34.1	2.3	36.2	1.4	35.6	1.8	34.5	1.7	36.4	1.5	35.6	1.8	1.8
Maj. Width (Å)	19.6	1.2	19.1	0.5	19.9	0.2	19.3	0.2	19.5	0.2	20.0	0.2	19.4	0.2	19.6	0.2	0.2
Min. Width (Å)	12.2	0.3	12.2	0.3	12.5	0.1	12.3	0.2	12.4	0.2	12.5	0.1	12.3	0.1	12.4	0.2	0.2
Pucker (°)	137.1	14.2	129.5	26.7	130.5	12.0	148.3	9.7	149.7	7.7	130.7	11.4	148.4	10.0	149.7	7.5	7.5
α (°)	298.8	6.2	299.2	6.6	288.6	4.5	289.5	3.4	285.6	3.9	289.3	4.8	290.5	3.6	286.1	4.5	4.5
β (°)	172.4	4.9	175.7	10.4	168.1	5.1	167.3	7.2	164.4	7.8	168.4	5.2	169.6	7.0	165.5	7.3	7.3
γ (°)	50.3	4.9	56.5	7.3	57.3	2.7	54.0	1.8	60.1	5.1	57.4	3.0	53.8	1.9	59.7	5.0	5.0
δ (°)	126.7	9.0	122.8	19.6	121.0	7.1	131.7	5.4	134.6	4.9	120.9	6.8	131.4	5.8	134.5	5.1	5.1
ϵ (°)	188.5	5.2	190.4	27.4	197.6	13.2	199.0	15.1	201.3	13.0	194.8	10.7	194.8	14.2	199.8	12.2	12.2
ζ (°)	257.1	12.0	251.3	37.6	254.5	18.2	244.6	17.7	245.5	19.0	256.8	13.7	247.5	15.8	245.9	16.6	16.6
χ (°)	249.2	7.9	243.9	14.7	241.7	7.5	251.3	5.1	247.5	6.9	240.9	6.8	250.9	4.7	247.4	6.3	6.3
BII %	N/A	N/A	N/A	N/A	12.7	N/A	23.9	N/A	22.3	N/A	10.8	N/A	21.6	N/A	21.9	N/A	N/A
RMSD all (Å)	N/A	N/A	N/A	N/A	2.5	0.5	1.8	0.3	1.7	0.4	2.3	0.4	1.8	0.3	1.7	0.4	0.4
RMSD int. (Å)	N/A	N/A	N/A	N/A	1.4	0.3	1.1	0.2	1.2	0.2	1.3	0.3	1.1	0.2	1.2	0.2	0.2

Table 2.4: Comparison of average BII fraction from simulations. Because 1I0T is Z-DNA, the 1I0T values are for ZI rather than BII.

	TIP3P			OPC		
	bsc0	bsc1	ol15	bsc0	bsc1	ol15
DDD	0.13	0.22	0.24	0.11	0.22	0.21
1FZX	0.09	0.17	0.18	0.08	0.16	0.15
1SK5	0.15	0.17	0.20	0.11	0.15	0.18
3GGI	0.20	0.21	0.28	0.15	0.22	0.27
ZI						
1I0T	0.29	0.08	0.79	0.25	0.06	0.80

Table 2.5: Average and standard deviation (SD) of structural parameters for the IFZX system. The terminal base pairs on each side of the IFZX were excluded for the average value calculation. NMR values were obtained from Reference [80]. RMSD values calculated using an average structure over 10 μ s of simulation time.

Property of Interest	NMR		TIP3P						OPC					
	Avg	SD	bsc0		oll5		bsc1		bsc0		oll5		bsc1	
Shear (Å)	0.2	0.2	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0
Stretch (Å)	-0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Stagger (Å)	0.3	0.3	0.0	0.1	-0.1	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1
Buckle (Å)	0.5	8.3	3.9	2.9	6.4	3.6	4.2	3.0	2.2	4.3	3.7	4.1	3.1	3.2
Propeller (°)	-13.9	5.5	-14.3	3.6	-13.6	4.8	-11.0	4.0	-14.6	3.9	-13.6	4.8	-11.2	4.1
Opening (°)	0.2	3.9	0.9	0.6	1.2	0.8	0.4	0.6	0.6	0.6	0.8	0.7	0.1	0.6
Shift (Å)	-0.1	0.3	-0.1	0.1	-0.2	0.1	-0.2	0.2	-0.1	0.1	-0.2	0.1	-0.1	0.2
Tilt (°)	-1.8	1.8	-0.6	1.3	-0.9	1.2	-1.1	1.2	-0.5	1.3	-0.9	1.2	-1.0	1.2
Slide (Å)	-0.9	0.2	-0.6	0.1	-0.2	0.2	-0.5	0.2	-0.6	0.2	-0.3	0.2	-0.6	0.2
Rise (Å)	3.1	0.1	3.3	0.0	3.3	0.0	3.3	0.1	3.3	0.0	3.3	0.0	3.3	0.0
Roll (°)	0.4	3.6	2.5	3.3	2.9	2.2	1.9	2.5	1.3	3.4	1.5	2.7	1.2	2.8
Twist (°)	36.2	2.0	32.8	2.0	34.2	1.9	33.9	2.1	33.6	1.9	34.5	2.0	34.1	2.1
X-disp (Å)	-1.5	0.5	-1.5	0.6	-1.0	0.3	-1.3	0.4	-1.3	0.6	-0.9	0.4	-1.3	0.5
Y-disp (Å)	-0.1	0.3	0.1	0.3	0.2	0.3	0.1	0.4	0.1	0.3	0.2	0.3	0.1	0.4
Hel. Rise (Å)	3.0	0.2	3.2	0.2	3.2	0.1	3.2	0.1	3.2	0.2	3.2	0.1	3.2	0.1
Hel. Incl. (°)	0.8	6.1	5.1	6.7	5.4	4.6	3.8	5.1	2.9	6.9	3.1	5.5	2.6	5.6
Tip (°)	2.7	2.8	0.8	2.4	1.4	2.1	1.7	2.2	0.7	2.5	1.2	2.1	1.5	2.1
Hel.Twist (°)	36.5	1.8	33.9	1.2	35.1	1.5	34.8	1.6	34.5	1.3	35.4	1.5	35.0	1.6
Maj. Width (Å)	19.4	0.1	20.0	0.2	19.6	0.1	19.7	0.1	20.1	0.2	19.6	0.2	19.8	0.2
Min. Width (Å)	12.4	0.2	12.6	0.1	12.3	0.1	12.4	0.1	12.6	0.1	12.3	0.1	12.4	0.1
Pucker (°)	130.5	21.3	127.2	11.2	144.0	11.0	145.6	8.6	129.1	11.1	144.3	11.0	146.4	8.2
α (°)	290.1	5.8	288.0	7.2	290.2	3.7	286.4	2.8	288.6	9.6	291.3	3.9	287.2	3.8
β (°)	183.1	9.3	169.1	2.1	170.1	5.2	166.7	3.5	169.4	2.8	172.4	7.0	167.6	4.6
γ (°)	58.2	4.4	59.0	5.0	54.9	2.2	60.2	3.0	59.2	7.2	54.5	2.2	59.9	3.0
δ (°)	132.5	12.2	118.8	6.4	128.7	6.4	131.8	5.5	119.7	6.4	128.3	6.3	132.0	5.4
ϵ (°)	173.0	6.2	193.9	8.6	194.5	13.1	197.4	9.6	191.6	9.2	189.4	13.9	194.9	10.3
ζ (°)	261.5	9.1	259.6	12.7	250.1	15.7	250.2	14.9	260.3	12.1	253.9	16.1	251.5	14.9
χ (°)	238.2	9.2	239.5	6.2	249.3	5.9	245.1	5.1	239.7	6.3	248.7	5.7	245.1	5.1
BII %	N/A	N/A	8.7	18.0	18.0	18.0	17.4	17.4	7.8	14.8	14.8	14.8	15.8	15.8
RMSD all (Å)	N/A	N/A	2.8	0.6	2.1	0.4	1.8	0.3	2.7	0.5	1.9	0.4	1.7	0.3
RMSD int. (Å)	N/A	N/A	1.4	0.3	1.1	0.2	1.2	0.2	1.3	0.3	1.1	0.2	1.2	0.2

Table 2.6: Average and standard deviation (SD) of structural parameters for the 1SK5 system. The terminal base pairs on each side of the 1SK5 were excluded for the average value calculation. RMSD values calculated using an average structure over 10 μ s of simulation time.

Property of Interest	X-Ray		TIP3P						OPC					
	Avg	SD	bsc0		oll5		bsc1		bsc0		oll5		bsc1	
Shear (Å)	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1
Stretch (Å)	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Stagger (Å)	0.0	0.1	0.0	0.1	0.0	0.2	0.1	0.1	0.0	0.1	0.0	0.2	0.0	0.1
Buckle (°)	-2.5	4.9	0.4	6.1	-0.5	8.6	0.0	5.5	0.0	5.0	-0.1	7.9	-0.2	5.5
Propeller (°)	-15.9	3.8	-13.5	2.0	-14.4	3.9	-11.5	3.4	-13.3	2.3	-15.6	4.1	-12.8	4.0
Opening (°)	2.2	1.1	2.5	0.7	2.0	0.5	1.9	0.5	1.9	0.9	1.9	0.7	1.0	0.5
Shift (Å)	-0.1	0.2	0.0	0.2	0.0	0.3	0.0	0.3	0.1	0.2	0.0	0.3	0.0	0.3
Tilt (°)	0.2	1.7	0.4	2.4	-0.6	3.1	0.0	1.9	-0.8	1.9	-0.1	1.9	-0.1	1.7
Slide (Å)	0.2	0.5	-0.2	0.3	0.1	0.4	-0.3	0.3	-0.3	0.3	0.0	0.4	-0.4	0.2
Rise (Å)	3.2	0.1	3.2	0.1	3.3	0.1	3.3	0.1	3.3	0.1	3.3	0.1	3.3	0.1
Roll (°)	-1.0	4.2	2.1	3.7	3.0	1.4	1.5	1.6	1.8	4.1	1.7	1.9	0.9	2.0
Twist (°)	37.3	4.3	31.5	1.7	34.6	1.6	34.8	1.2	32.2	2.2	35.1	1.8	34.8	1.4
X-disp (Å)	0.4	0.6	-1.0	0.2	-0.4	0.3	-0.8	0.2	-1.1	0.4	-0.4	0.3	-0.8	0.1
Y-disp (Å)	0.1	0.3	-0.1	0.1	0.0	0.3	-0.1	0.3	-0.1	0.2	0.0	0.3	0.0	0.3
Hel. Rise (Å)	3.2	0.1	3.0	0.2	3.2	0.1	3.2	0.1	3.1	0.1	3.2	0.1	3.2	0.1
Hel. Incl. (°)	-0.8	7.2	5.0	6.5	5.0	2.8	2.8	2.9	4.5	7.4	3.4	3.7	1.6	4.1
Tip (°)	-0.2	2.7	0.6	2.8	0.3	3.5	0.1	2.9	0.5	2.2	0.1	3.0	0.0	2.7
Hel.Twist (°)	37.6	4.1	33.3	2.0	36.0	1.8	35.8	1.1	34.2	1.7	36.0	1.6	35.7	1.1
Maj. Width (Å)	19.1	0.4	17.9	1.0	19.3	0.1	18.1	0.3	18.0	0.7	19.4	0.1	19.5	0.1
Min. Width (Å)	12.2	0.3	11.2	0.6	12.1	0.1	11.3	0.2	11.2	0.4	12.1	0.2	12.2	0.1
Pucker (°)	133.9	34.6	117.4	14.5	143.3	13.8	135.5	8.7	117.9	12.2	143.6	13.5	145.7	8.8
α (°)	274.2	66.9	284.7	5.9	290.1	3.1	286.4	1.9	287.9	4.6	290.6	3.5	286.8	3.0
β (°)	176.4	19.9	169.0	2.4	168.8	4.7	166.3	4.7	169.1	2.4	170.4	5.8	167.2	5.8
γ (°)	70.4	45.0	60.5	4.1	55.9	2.3	60.8	3.1	58.8	2.6	55.6	2.4	60.7	2.5
δ (°)	125.8	16.0	120.5	8.6	128.5	7.8	132.5	5.6	120.5	7.9	128.5	7.6	132.2	5.7
ϵ (°)	188.4	23.0	200.3	9.6	196.1	12.1	196.8	9.2	195.1	10.2	192.7	13.9	194.9	10.4
ζ (°)	257.4	32.2	248.6	17.5	248.1	15.1	251.3	13.2	255.9	15.7	250.5	16.9	252.2	14.3
χ (°)	247.8	16.2	241.4	8.4	250.1	8.2	245.9	5.3	242.2	8.0	250.3	8.0	245.4	5.1
BII %	N/A	N/A	15.0		20.1		16.7		10.8		18.1		15.4	
RMSD all (Å)	N/A	N/A	3.0	1.0	1.9	0.3	2.1	0.8	2.8	0.9	1.7	0.4	1.9	0.6
RMSD int. (Å)	N/A	N/A	1.4	0.3	1.1	0.2	1.2	0.2	1.3	0.3	1.1	0.2	1.2	0.2

Table 2.7: Average structural parameters for the 3GGI system. The terminal base pairs on each side of the 3GGI were excluded for the average value calculation. RMSD values calculated using an average structure over 10 μ s of simulation time.

Property of Interest	X-Ray		TIP3P						OPC					
	Avg	SD	bsc0		oll5		bsc1		bsc0		oll5		bsc1	
Shear (Å)	0.0	0.2	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1
Stretch (Å)	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0
Stagger (Å)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Buckle (°)	0.0	7.4	-0.2	4.8	0.0	6.2	0.0	6.5	0.9	5.7	0.2	6.9	0.0	6.8
Propeller (°)	-8.7	2.5	-5.8	5.0	-9.0	3.0	-6.4	2.6	-6.1	5.4	-8.7	3.7	-6.6	2.8
Opening (°)	-1.6	1.6	0.7	0.8	0.0	0.3	-0.1	0.2	0.5	1.3	-0.1	0.5	-0.2	0.3
Shift (Å)	0.0	0.3	-0.1	0.4	0.0	0.2	0.0	0.2	0.1	0.3	0.0	0.2	0.0	0.2
Tilt (°)	0.0	3.9	0.9	2.0	0.1	1.5	0.1	1.5	-0.4	1.9	-0.1	1.6	-0.1	1.6
Slide (Å)	1.3	1.0	-0.3	0.7	-0.1	0.3	-0.4	0.3	-0.3	0.6	-0.1	0.4	-0.4	0.3
Rise (Å)	3.3	0.2	3.3	0.3	3.3	0.1	3.3	0.2	3.3	0.2	3.3	0.2	3.3	0.2
Roll (°)	1.4	6.3	2.0	1.7	2.6	1.5	2.3	1.6	3.7	2.1	2.6	1.3	2.5	1.5
Twist (°)	37.3	12.0	29.3	4.2	33.9	2.4	33.4	2.7	29.1	5.1	34.2	2.5	34.1	2.7
X-disp (Å)	1.0	2.1	-2.0	0.5	-0.8	0.3	-1.3	0.3	-2.3	0.6	-0.7	0.4	-1.2	0.3
Y-disp (Å)	0.0	1.5	0.1	0.4	0.0	0.3	0.0	0.3	0.0	0.4	0.0	0.2	0.0	0.3
Hel. Rise (Å)	3.2	0.3	2.9	0.6	3.2	0.2	3.2	0.2	3.0	0.4	3.2	0.2	3.2	0.2
Hel. Incl. (°)	5.6	11.9	3.1	3.8	5.0	2.8	4.3	3.2	4.6	3.4	4.9	2.3	4.7	2.9
Tip (°)	0.0	6.9	-1.1	2.7	-0.1	2.6	-0.1	2.7	0.3	3.2	0.2	2.9	0.0	2.8
Hel. Twist (°)	38.2	11.5	31.7	2.7	34.9	2.2	34.5	2.5	30.8	4.2	35.1	2.4	35.2	2.6
Maj. Width (Å)	18.2	0.7	18.1	2.6	19.4	0.1	19.4	0.3	18.6	2.3	19.3	0.2	19.4	0.3
Min. Width (Å)	11.7	0.3	11.4	1.6	12.2	0.2	12.2	0.3	11.8	1.4	12.2	0.2	12.2	0.3
Pucker (°)	151.8	17.7	125.1	11.8	147.8	8.9	147.5	8.1	131.5	13.6	148.9	9.7	148.6	8.0
α (°)	289.9	5.3	277.5	12.0	287.9	2.3	283.5	3.3	278.7	13.4	288.1	2.1	283.5	3.7
β (°)	157.9	18.5	170.3	3.0	167.2	6.7	166.9	7.7	170.5	3.0	167.8	7.7	167.2	7.7
γ (°)	47.7	5.2	63.1	7.9	54.1	1.4	60.2	4.1	61.6	6.3	54.2	1.8	59.6	4.4
δ (°)	135.1	9.5	122.5	6.9	132.2	5.3	134.5	4.8	123.2	6.9	132.8	5.5	135.3	4.6
ϵ (°)	224.5	40.9	206.4	14.9	202.1	11.8	200.1	7.7	201.5	11.8	200.1	12.8	199.8	8.3
ζ (°)	218.5	56.9	239.2	31.0	240.8	15.2	247.7	10.5	249.1	20.7	241.8	15.4	246.7	11.4
χ (°)	266.5	12.8	240.1	10.9	251.3	6.1	245.8	5.4	241.9	9.0	252.3	6.1	246.7	5.3
BII %	N/A	N/A	20.3		27.7		20.9		15.3		26.9		21.9	
RMSD all (Å)	N/A	N/A	5.1	0.5	4.1	0.3	4.4	0.3	5.0	0.5	4.1	0.3	4.4	0.3
RMSD int. (Å)	N/A	N/A	1.4	0.3	1.4	0.2	1.2	0.2	1.3	0.3	1.1	0.2	1.2	0.2

Table 2.8: Average structural parameters for the 1I0T system. The terminal base pairs on each side of the 1I0T were excluded for the average value calculation. X-ray values were obtained from Reference [81]. RMSD values calculated using an average structure over 10 μ s of simulation time.

Property of Interest	X-Ray		TIP3P						OPC					
	Avg	SD	bsc0		oll15		bsc1		bsc0		oll15		bsc1	
Shear (Å)	0.0	0.3	0.0	0.2	0.0	0.3	0.0	0.1	0.0	0.2	0.0	0.3	0.0	0.2
Stretch (Å)	-0.2	0.0	-0.1	0.0	-0.1	0.0	-0.1	0.1	-0.1	0.0	-0.1	0.0	-0.1	0.0
Stagger (Å)	0.0	0.1	-0.1	0.0	0.0	0.0	-0.2	0.0	-0.1	0.0	0.0	0.0	-0.2	0.0
Buckle (°)	-0.1	4.2	0.0	1.6	0.0	3.6	0.0	1.9	0.0	2.0	0.0	3.9	0.0	1.8
Propeller (°)	1.4	5.5	0.8	0.8	0.2	0.1	-1.0	1.9	-0.2	1.3	-0.8	0.2	-1.9	1.5
Opening (°)	-2.0	0.6	-0.5	0.1	-0.3	0.1	-0.9	0.5	-0.7	0.2	-0.5	0.1	-0.4	0.1
Shift (Å)	0.2	0.6	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.2
Tilt (°)	-0.1	1.2	0.0	0.1	0.1	0.1	0.2	0.3	0.0	0.2	0.0	0.0	-0.1	0.2
Slide (Å)	1.4	3.0	0.8	3.1	1.1	3.0	1.5	2.4	0.8	3.2	1.0	3.1	1.4	2.3
Rise (Å)	-3.4	0.2	-3.5	0.1	-3.4	0.2	-3.4	0.2	-3.4	0.1	-3.4	0.2	-3.4	0.1
Roll (°)	-2.1	3.3	-2.3	2.5	-3.9	1.7	0.4	1.2	-2.9	1.9	-4.7	1.4	-1.3	1.6
Twist (°)	37.8	21.7	34.1	15.8	31.8	13.4	32.1	12.9	33.9	16.5	32.1	13.8	32.1	12.5
X-disp (Å)	14.9	22.5	6.2	11.9	4.9	9.8	5.7	8.8	6.4	12.2	4.3	9.3	5.4	8.6
Y-disp (Å)	-2.9	4.1	0.0	0.1	0.0	0.0	-0.1	0.3	0.0	0.2	0.0	0.0	0.0	0.3
Hel. Rise (Å)	-3.0	0.3	-3.1	0.2	-3.4	0.4	-3.1	0.2	-3.2	0.0	-3.4	0.6	-3.2	0.1
Hel. Incl. (°)	0.1	6.5	-2.5	4.2	-7.4	0.1	0.4	3.4	-4.3	1.7	-9.4	1.9	-1.5	3.0
Tip (°)	-3.4	6.3	0.1	0.2	0.0	0.2	-0.1	0.2	0.0	0.2	-0.1	0.1	0.2	0.3
Hel.Twist (°)	38.0	21.7	35.1	15.3	32.9	13.0	33.5	12.7	34.9	15.9	33.2	13.3	32.9	12.0
Maj. Width (Å)	14.8	0.2	15.1	0.1	15.1	0.1	15.8	0.0	15.0	0.1	15.0	0.1	15.8	0.0
Min. Width (Å)	11.8	0.1	12.0	0.0	12.1	0.0	12.8	0.1	11.9	0.0	12.0	0.0	12.8	0.1
Pucker (°)	93.3	61.3	109.6	31.2	102.5	31.6	152.0	6.5	109.4	32.5	101.4	33.3	153.6	7.9
α (°)	130.8	64.5	153.5	57.1	158.6	26.3	202.1	13.6	147.4	63.8	150.6	37.9	201.8	13.9
β (°)	194.1	24.2	185.7	2.7	201.4	20.8	181.6	5.4	185.0	3.2	202.5	19.0	180.8	4.2
γ (°)	116.1	63.0	106.3	49.9	94.0	30.9	62.5	4.7	111.4	55.1	101.9	38.2	61.1	3.8
δ (°)	118.9	23.5	116.2	17.4	111.6	17.0	138.5	3.7	116.3	17.9	111.9	17.4	139.8	4.4
ϵ (°)	247.6	24.6	256.9	14.8	243.1	28.2	225.4	30.2	256.5	14.6	246.4	26.2	225.6	31.1
ζ (°)	157.6	114.7	170.8	93.4	177.1	97.2	231.6	90.4	163.4	87.7	174.4	100.9	234.4	90.6
χ (°)	131.1	72.1	131.9	70.0	132.2	64.4	159.9	46.4	131.0	71.1	131.4	65.1	160.4	45.5
RMSD all (Å)	N/A	N/A	3.4	0.5	3.3	0.2	4.4	0.3	3.6	0.6	3.2	0.0	4.3	0.2
RMSD int. (Å)	N/A	N/A	1.4	0.3	1.1	0.2	1.2	0.2	1.3	0.3	1.1	0.2	1.2	0.2

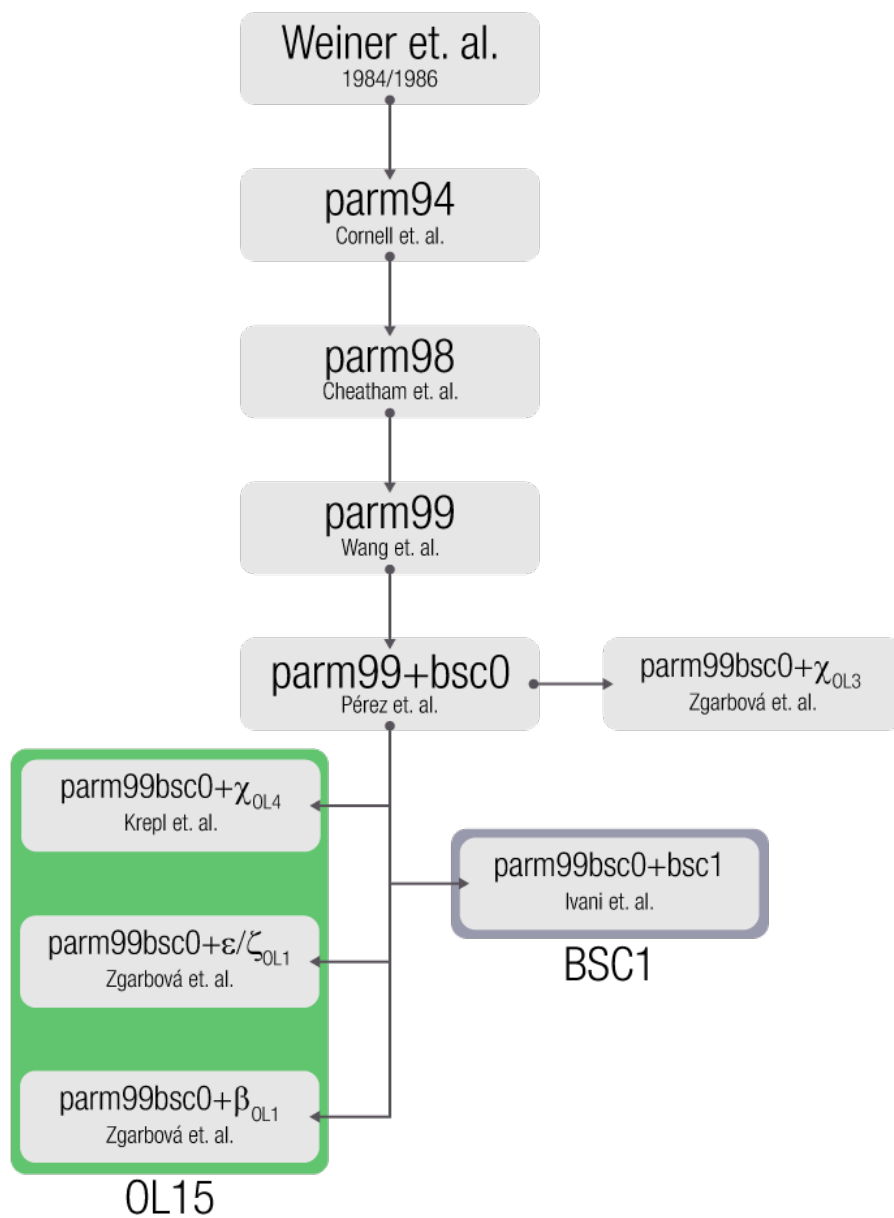


Figure 2.1: Recent history of force field development in AMBER. Historic flowchart of the two main forks of the AMBER force field development for DNA. The combination of parameter modifications that now comprise OL15 is highlighted in green. The parm99bsc0 χ OL3 is also shown because it was an important improvement to the RNA force field.

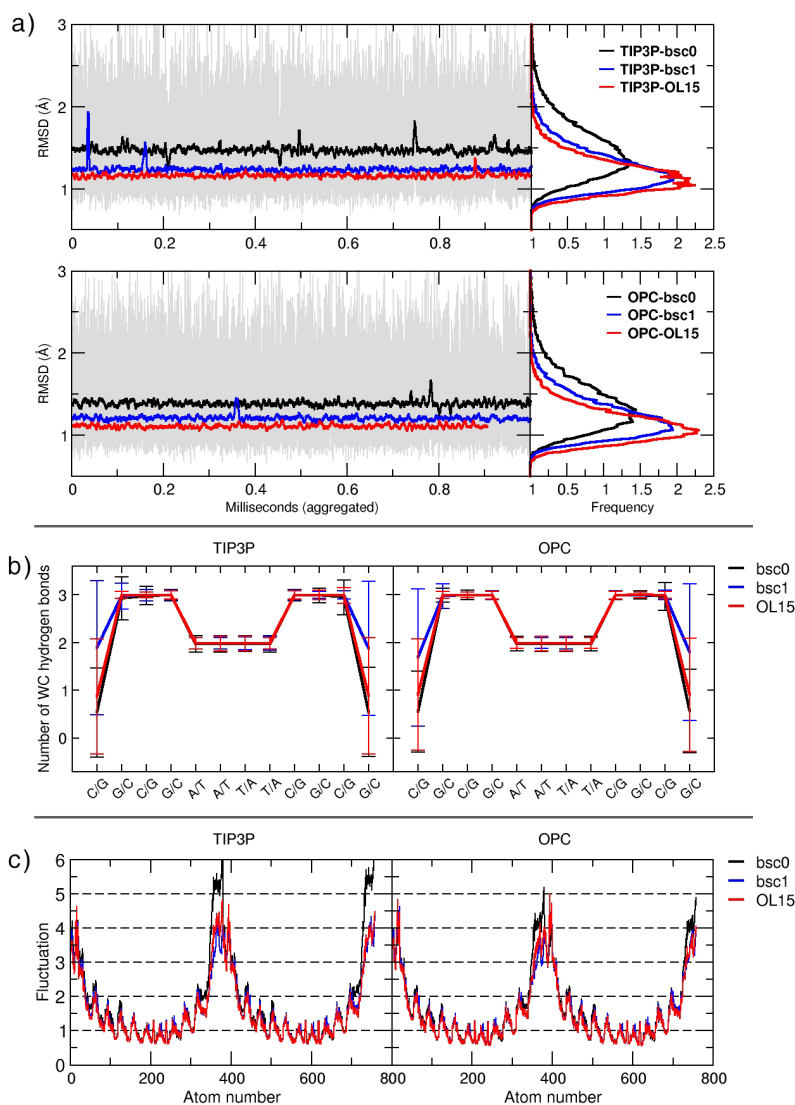


Figure 2.2: DDD RMSDs, average WC hydrogen bonding, and RMSFs. a) Running average of the root-mean-square-deviation and histogram plots for the DDD system. The first and second base pairs at each end of the DNA sequence are not included in the calculation. Top plot is using the TIP3P water model, bottom is with OPC water model. RMSD measurements were made using the average structure from the 1NAJ NMR as a reference. The raw RMSD data is shown in gray in light gray, running average using 500 frames. b) The average amount of Watson-Crick hydrogen bonds for each base pair in the DDD system using the full aggregated trajectory. Error bars represent standard deviations. c) Atomic fluctuations for each force field modification using the aggregated trajectories.

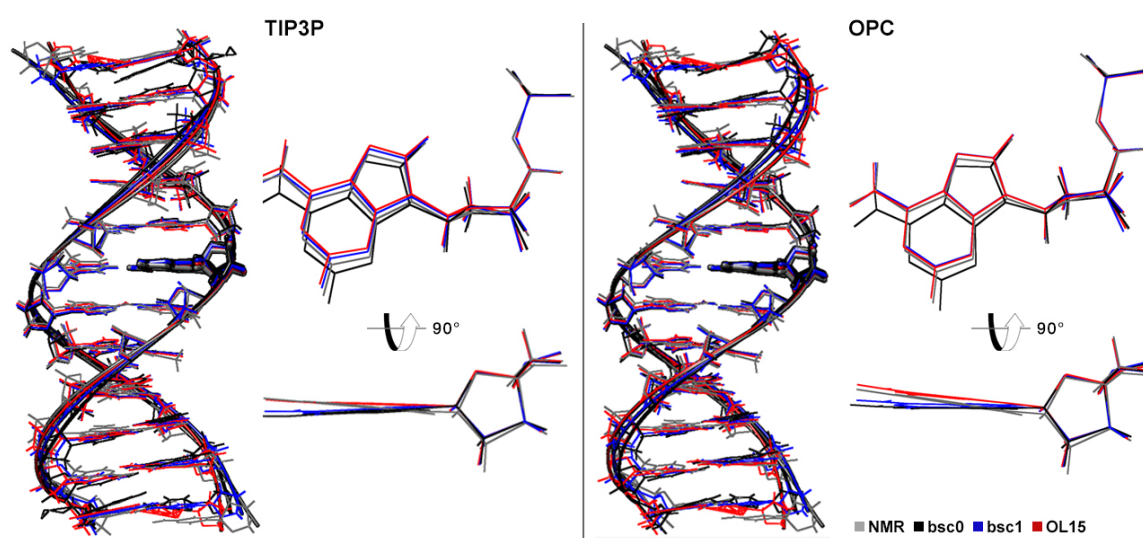


Figure 2.3: Comparison of average structures extracted from the full aggregated trajectory of DDD. The detail represents D6 (licorice in the full helix) in wire representation.

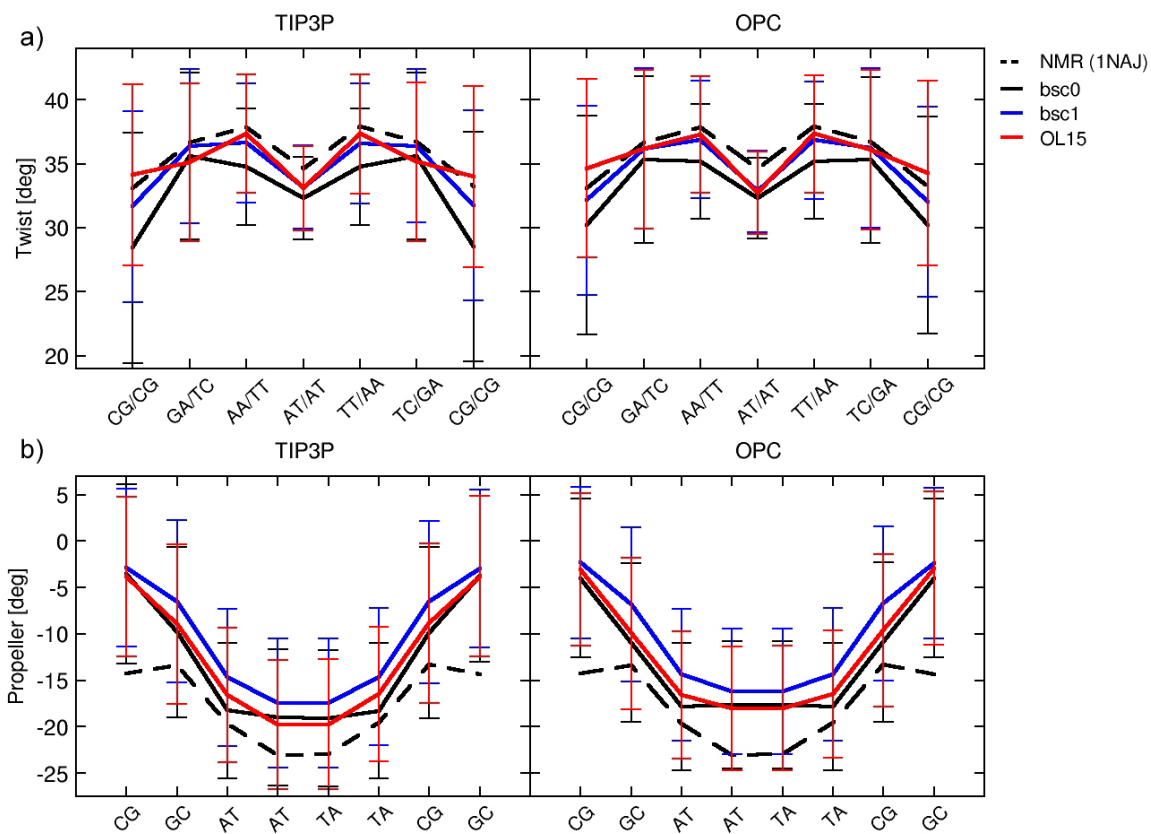


Figure 2.4: DDD twist and propeller values. a) Twist values for the seven internal base pair steps. b) Propeller values at each base pair for the internal 8 bp of DDD. MD simulations using the different force field modifications are compared with the experimental NMR structure (dotted-line). The entire aggregated trajectory (1 ms) was used. The two terminal base pairs at each end were excluded from the analysis.

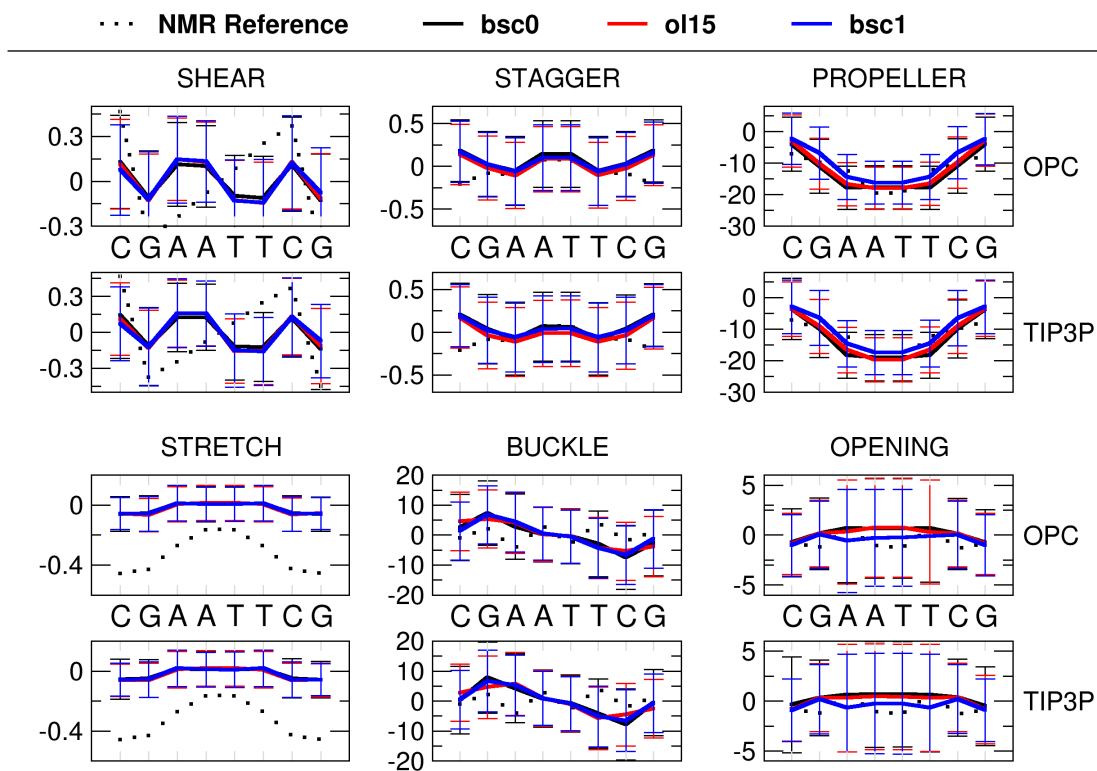


Figure 2.5: DDD base pair parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

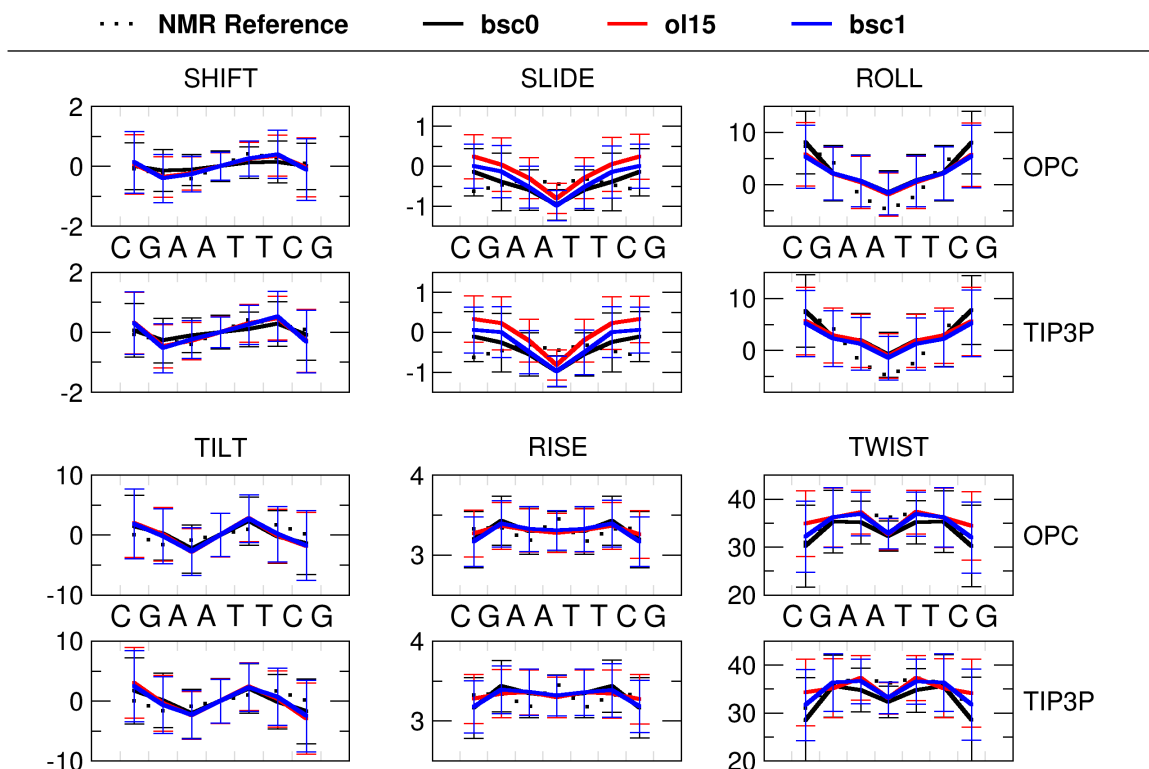


Figure 2.6: DDD base pair step parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

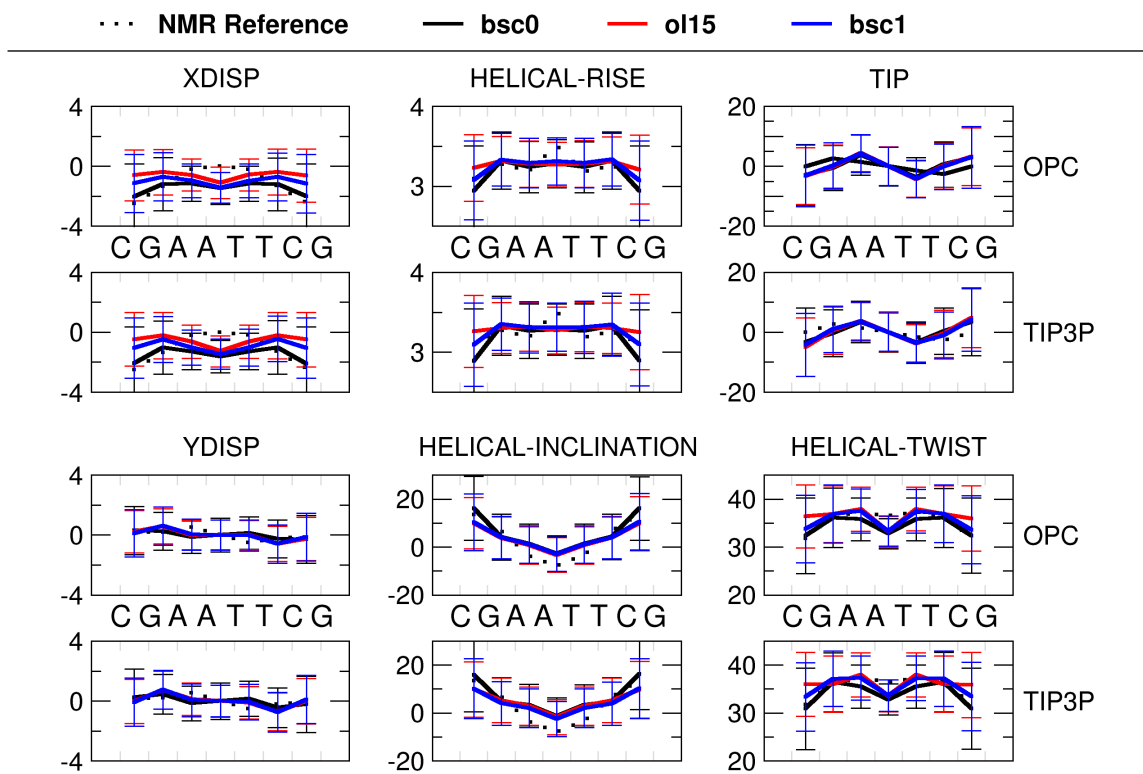


Figure 2.7: DDD helical parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

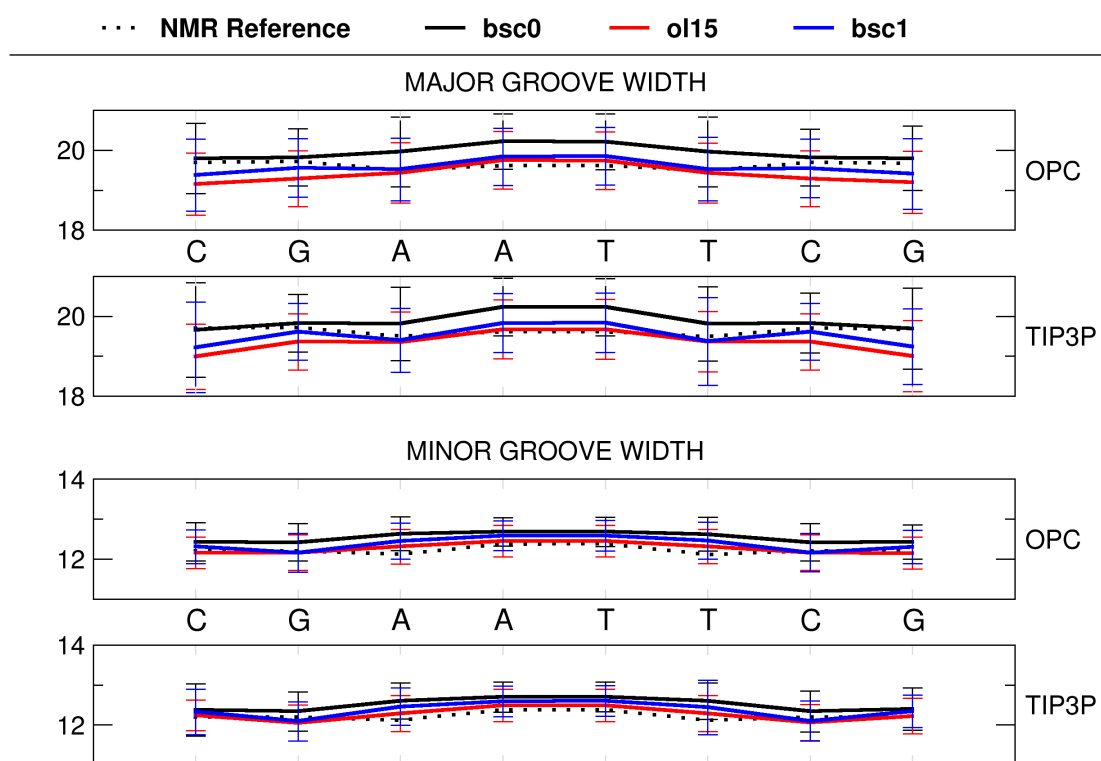


Figure 2.8: DDD groove widths. Each property was calculated using the CPPTRAJ implementation of 3DNA.

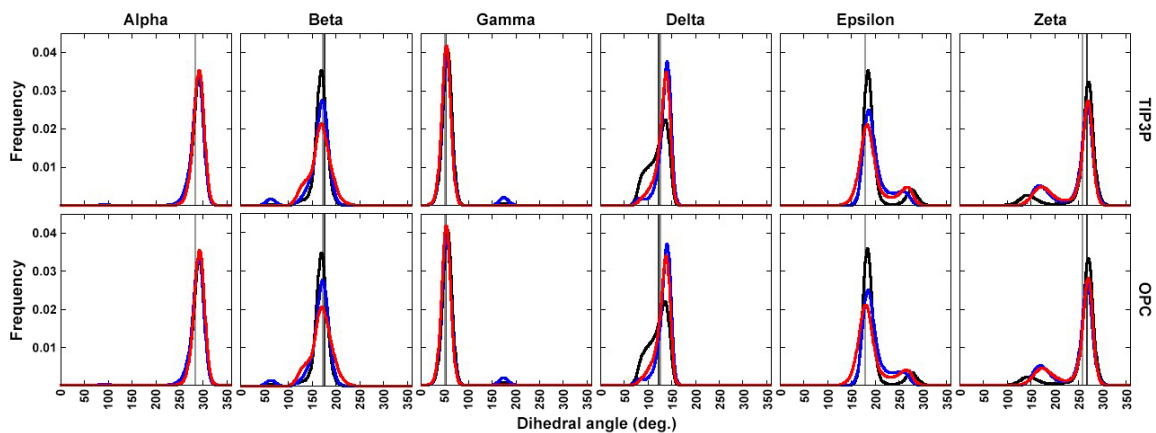


Figure 2.9: DDD dihedral histograms. Populations of the six dihedrals of DDD using the full aggregated trajectory. Only the inner 8 base pairs were included for analysis. The black line corresponds to bsc0, bsc1 is blue, and ol15 is red. The X-ray and NMR reference values are shown as vertical lines (gray and black, respectively).

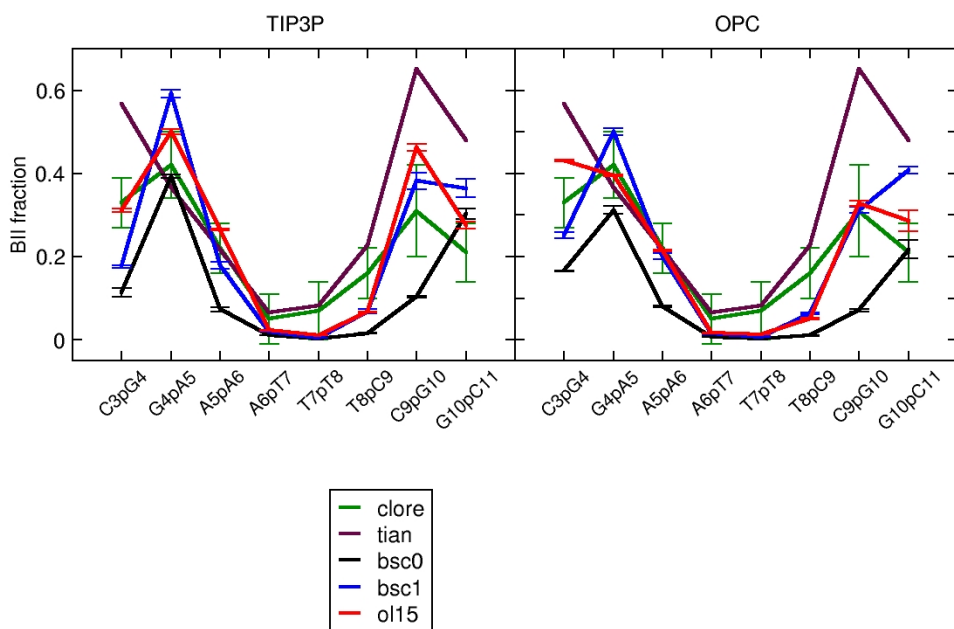


Figure 2.10: DDD BII populations per base pair step. Comparison of the BII fraction for each of the internal 8 base pair steps for DDD. Bsc0 (black), bsc1 (blue), and ol15 (red) is shown with NMR data (green is Tian et. al. [37], maroon is from Schwieters and Clore [36])

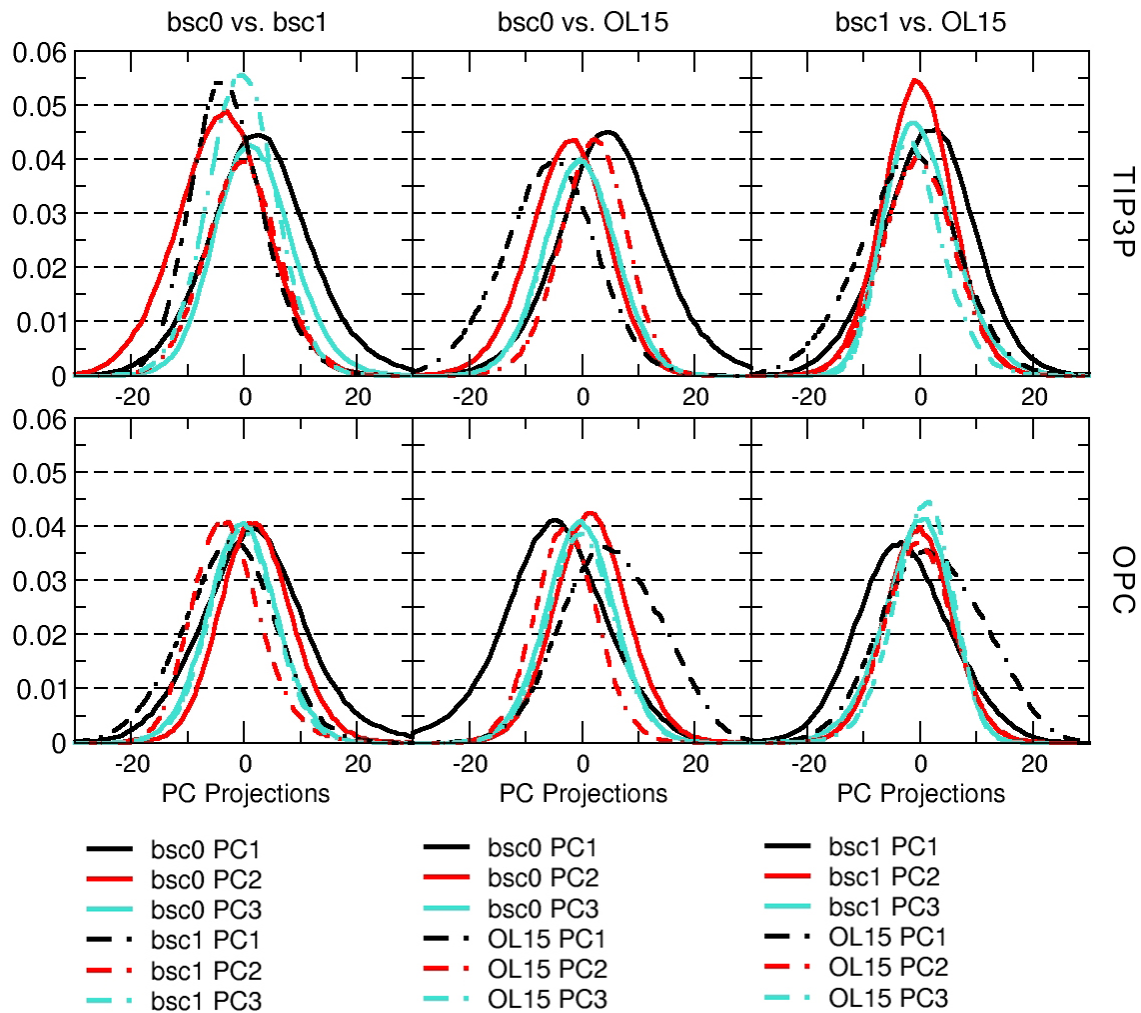


Figure 2.11: DDD PCA histograms. Principal component analysis projections for the eight internal base pairs of the DDD simulations.

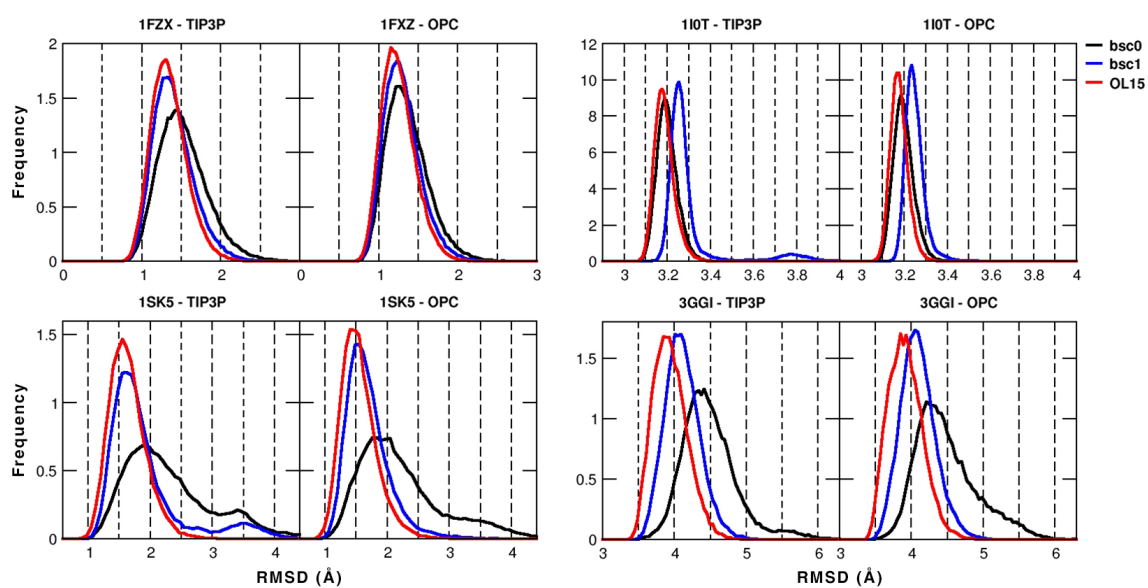


Figure 2.12: RMSD histograms for 1FZX, 1I0T, 1SK5, and 3GGI. Values were calculated using the full aggregated trajectories. The RMSDs were calculated with the experimental X-ray or NMR structures as reference.

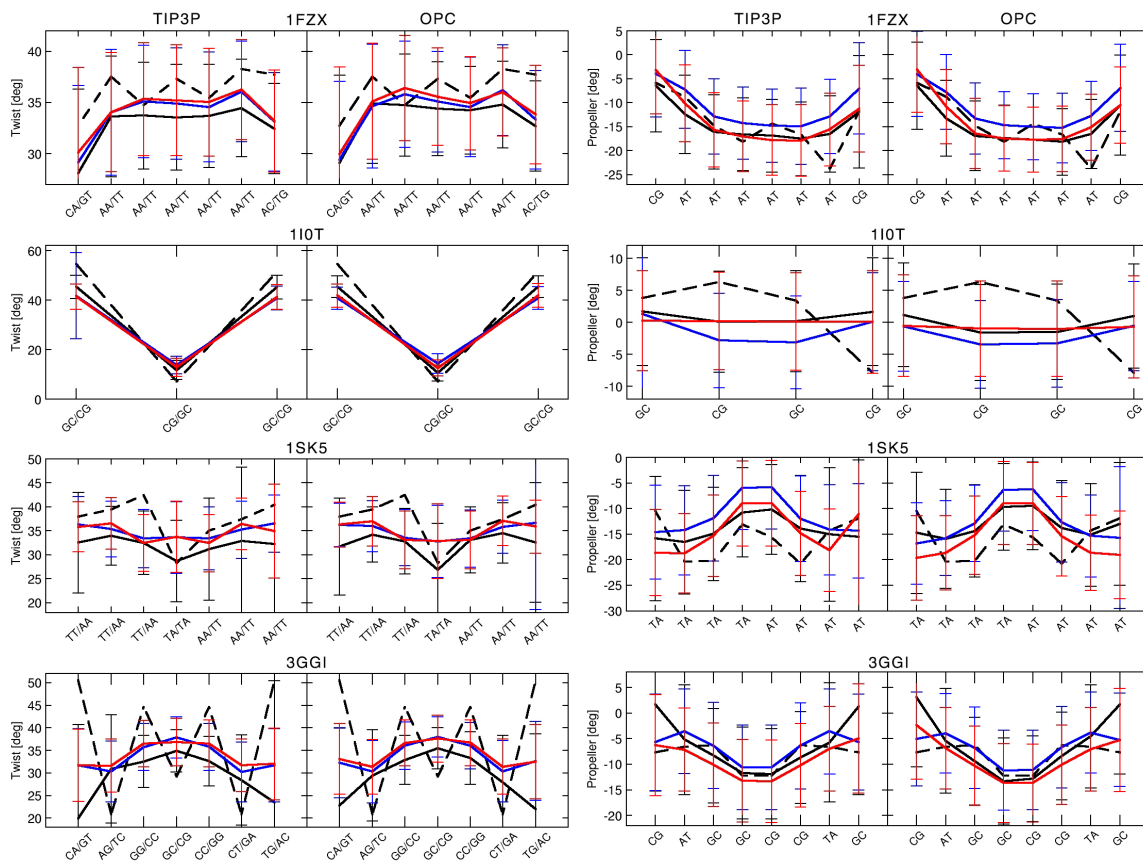


Figure 2.13: Twist and propeller values for 1FZX, 1I0T, 1SK5, and 3GGI. TIP3P and OPC are left and right column, respectively. The black dashed lines are the X-ray (and NMR for 1FZX) reference values. Black is bsc0, blue is bsc1, and red is ol15.

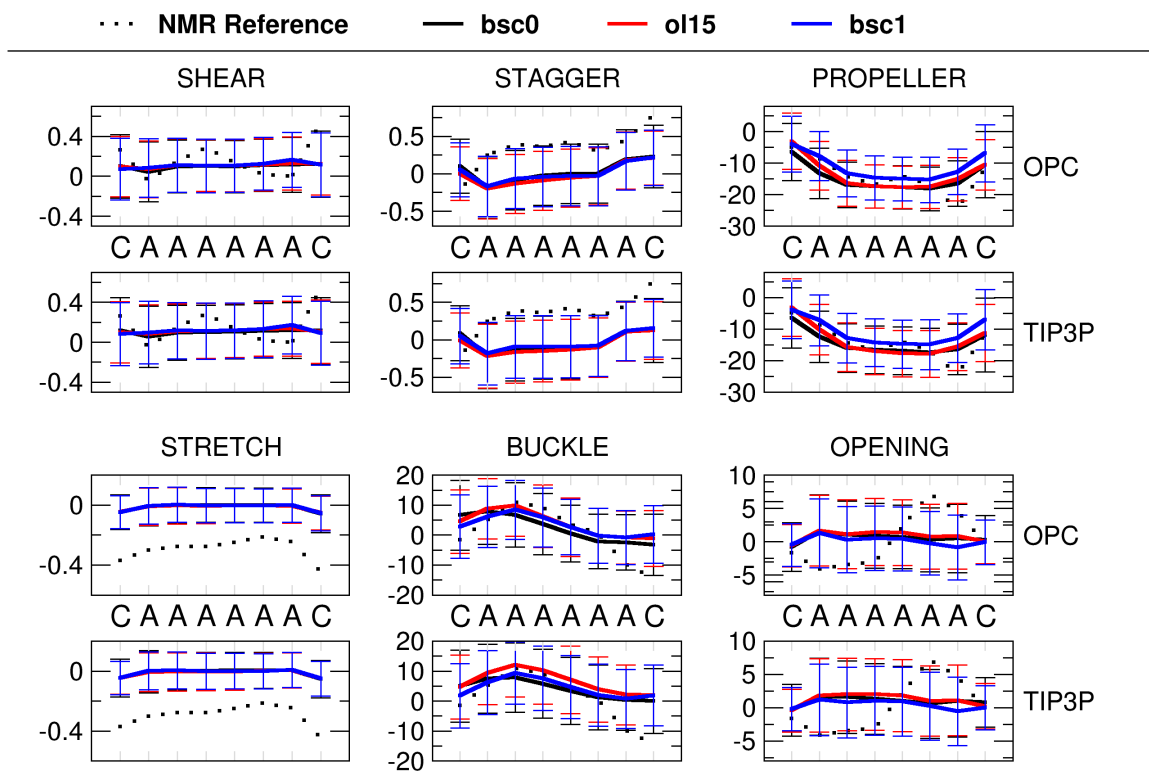


Figure 2.14: 1FZX base pair parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

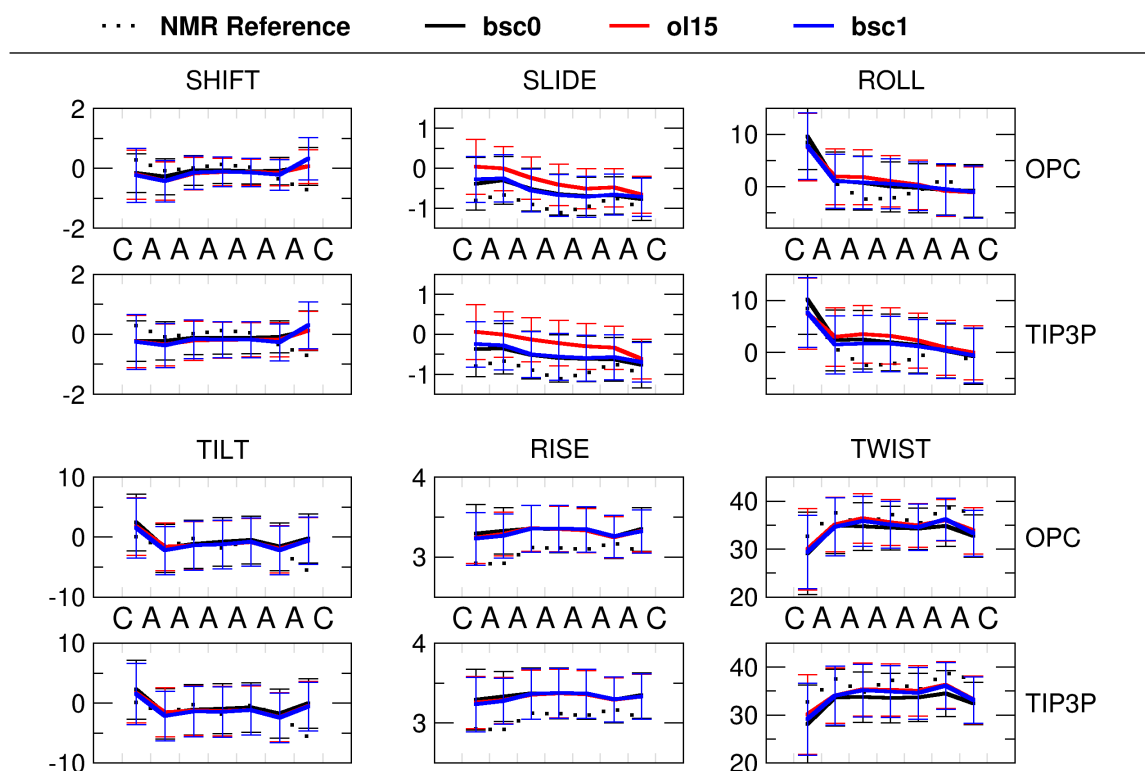


Figure 2.15: 1FZX base pair step parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

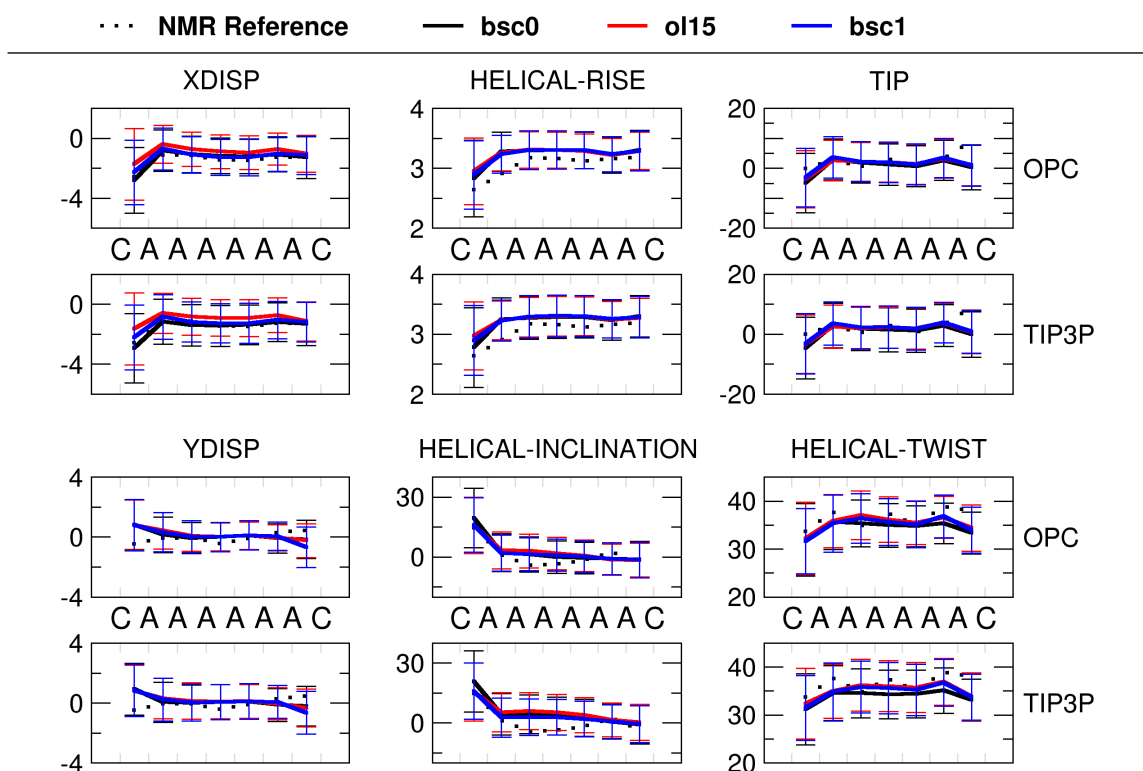


Figure 2.16: 1FZX helical parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

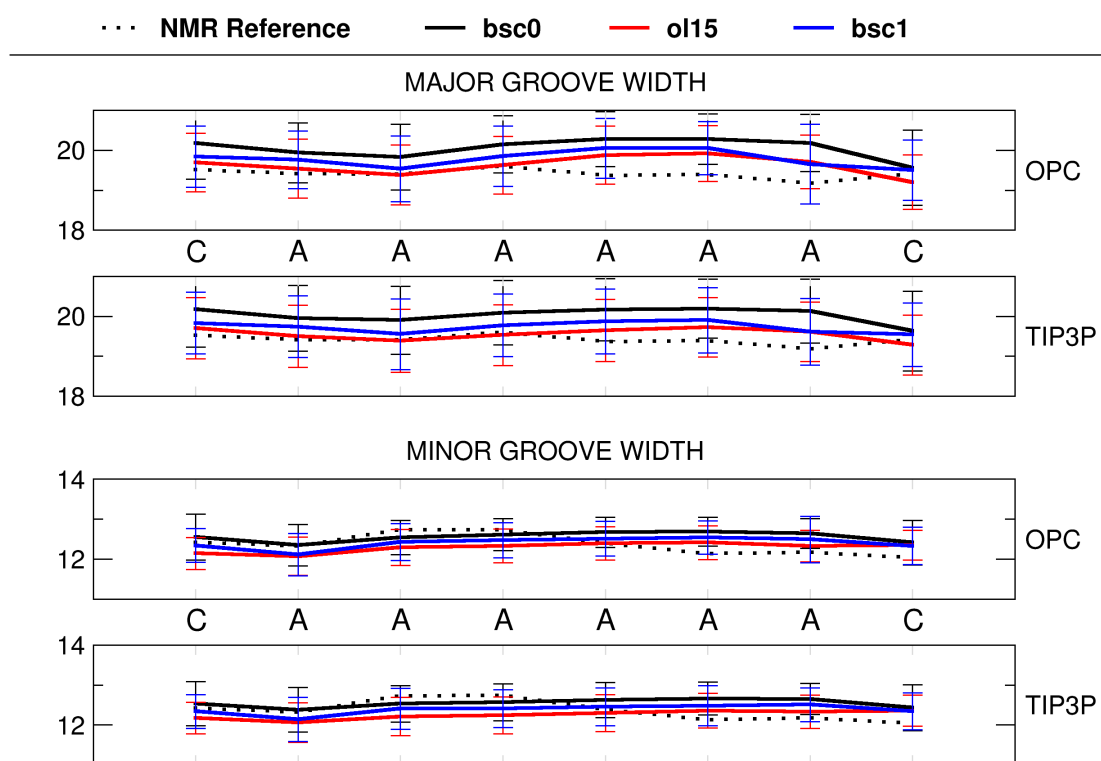


Figure 2.17: 1FZX groove widths. Each property was calculated using the CPPTRAJ implementation of 3DNA.

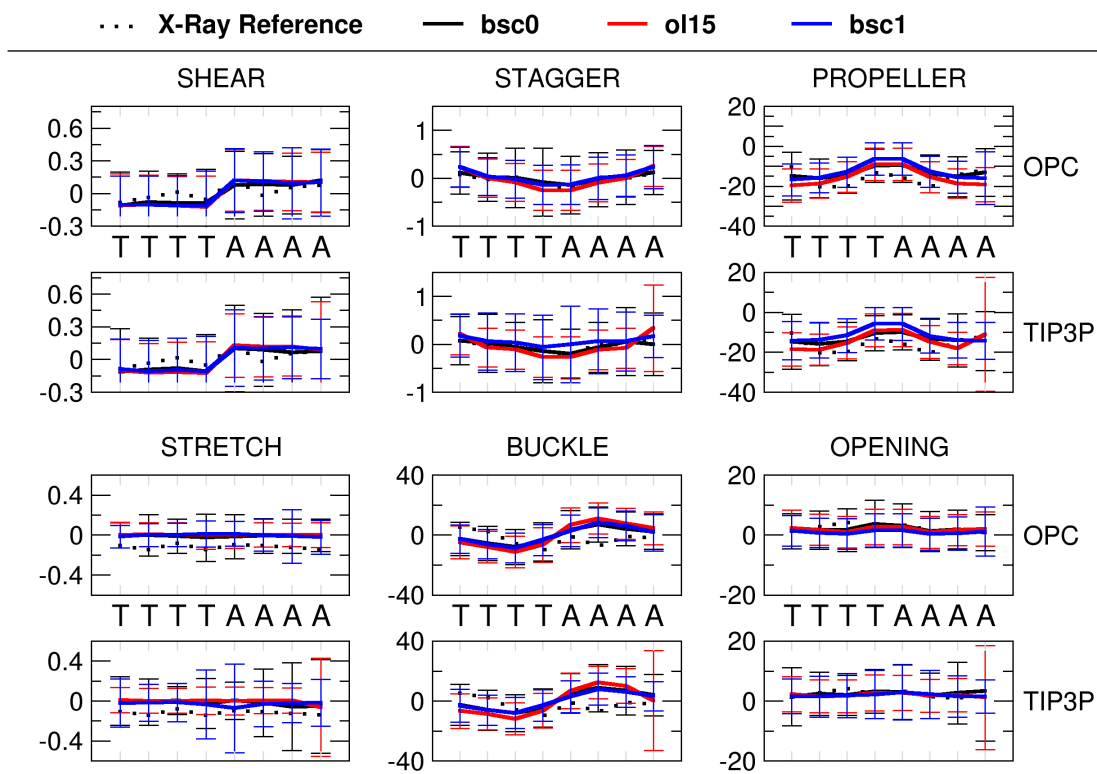


Figure 2.18: 1SK5 base pair parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

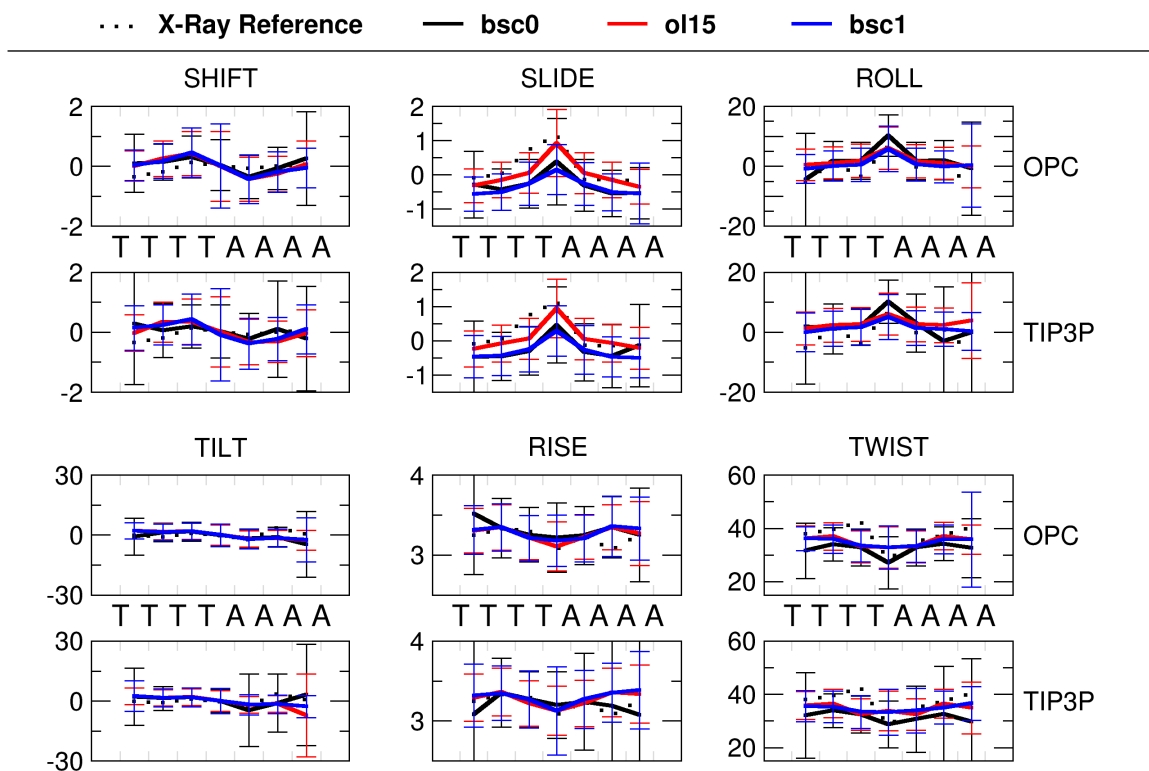


Figure 2.19: 1SK5 base pair step parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

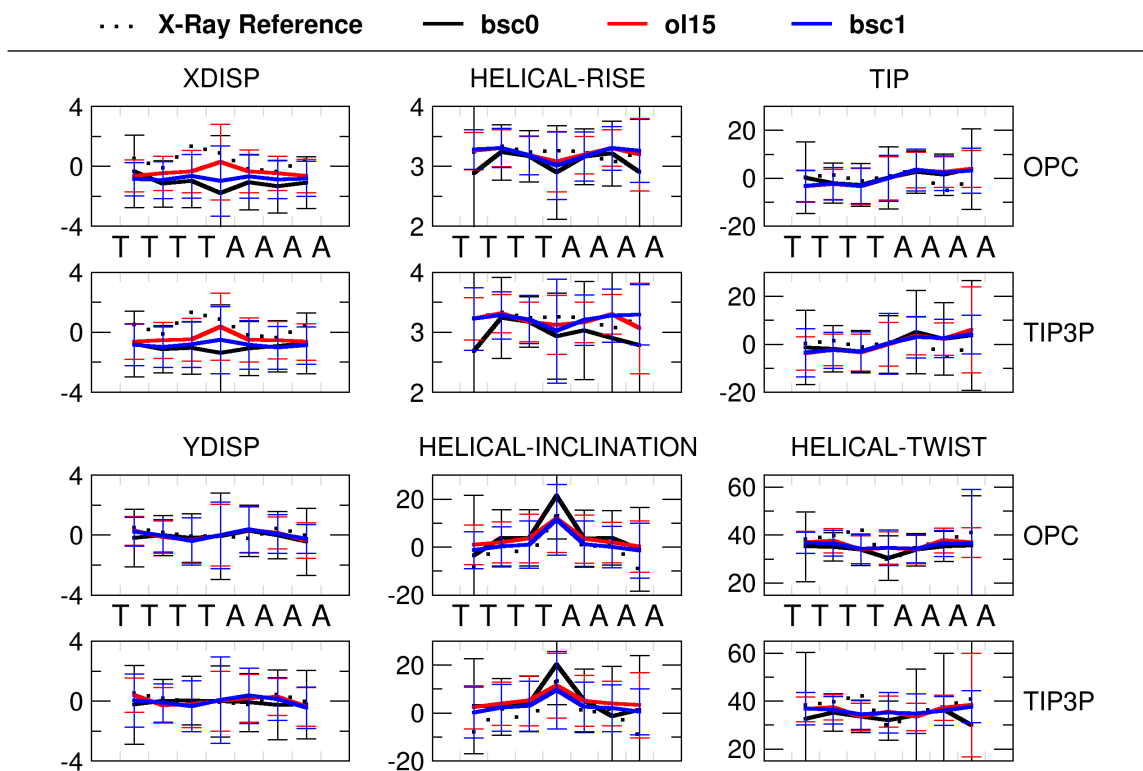


Figure 2.20: 1SK5 helical parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

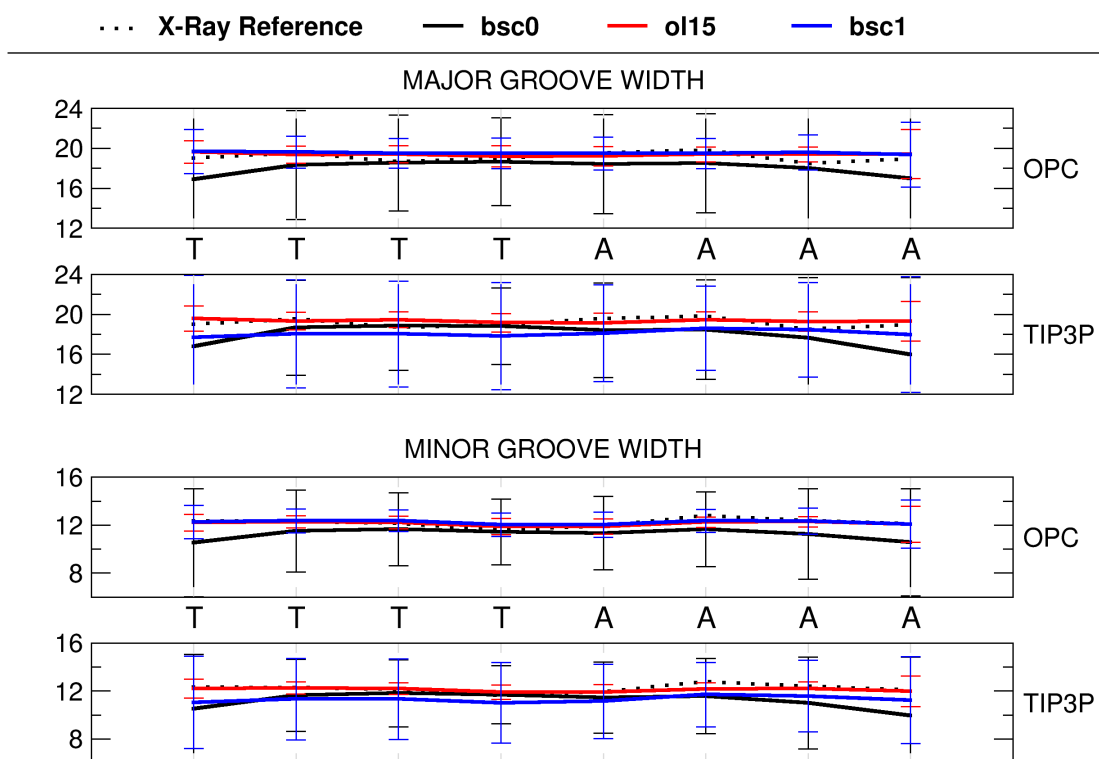


Figure 2.21: 1SK5 groove widths. Each property was calculated using the CPPTRAJ implementation of 3DNA.

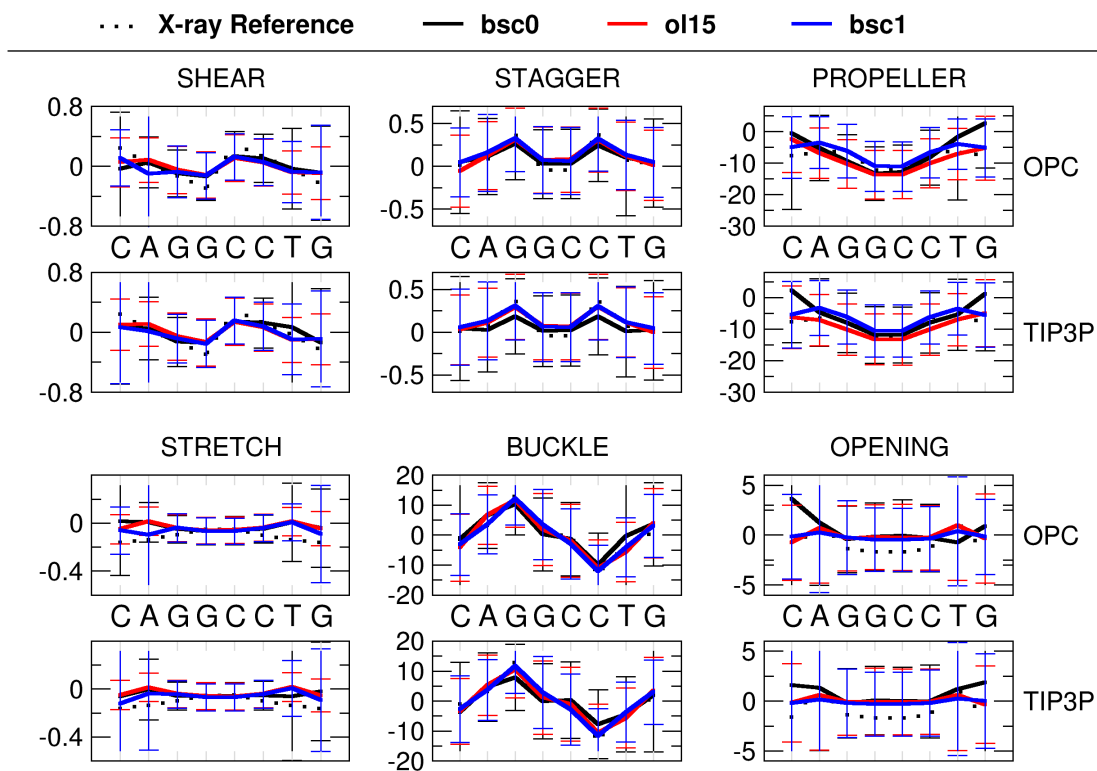


Figure 2.22: 3GGI base pair parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

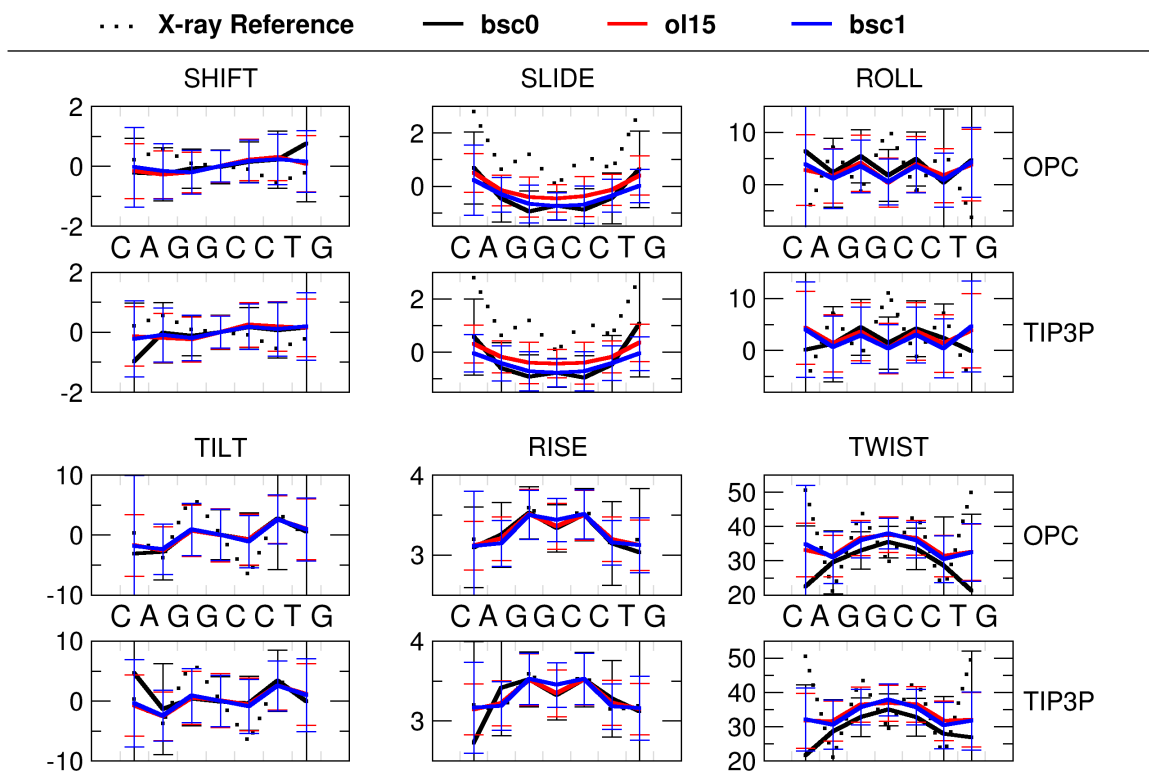


Figure 2.23: 3GGI base pair step parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

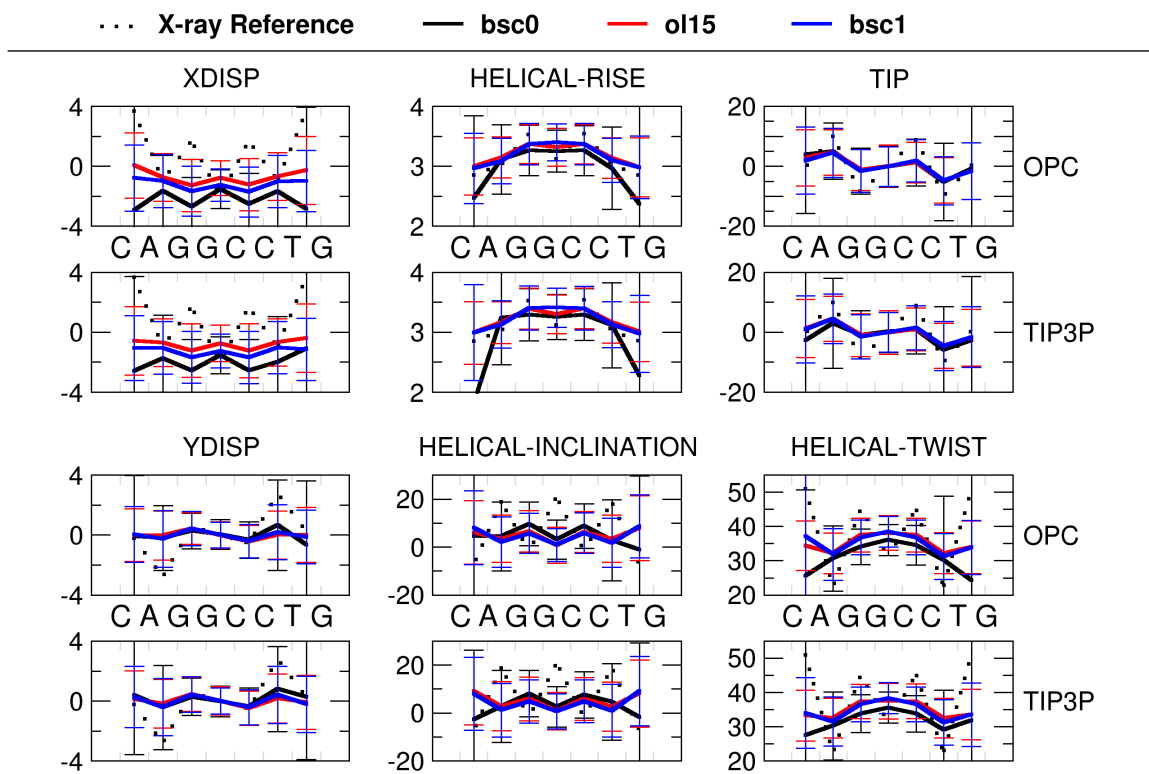


Figure 2.24: 3GGI helical parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

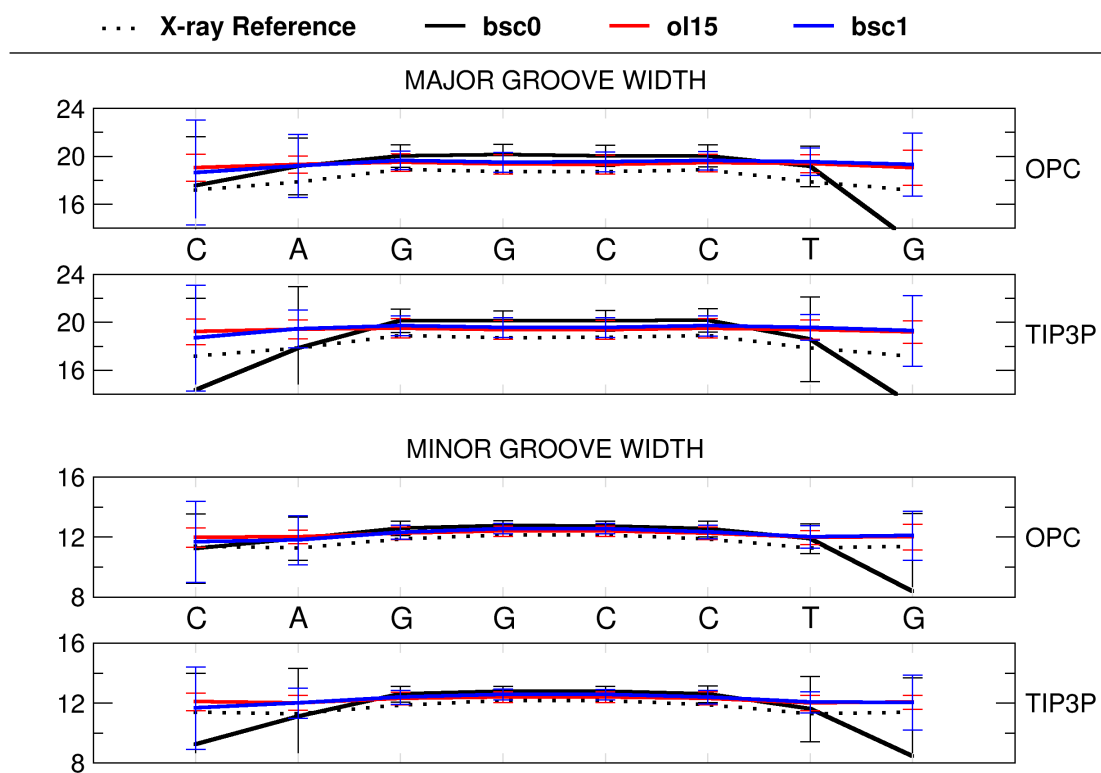


Figure 2.25: 3GGI groove widths. Each property was calculated using the CPPTRAJ implementation of 3DNA.

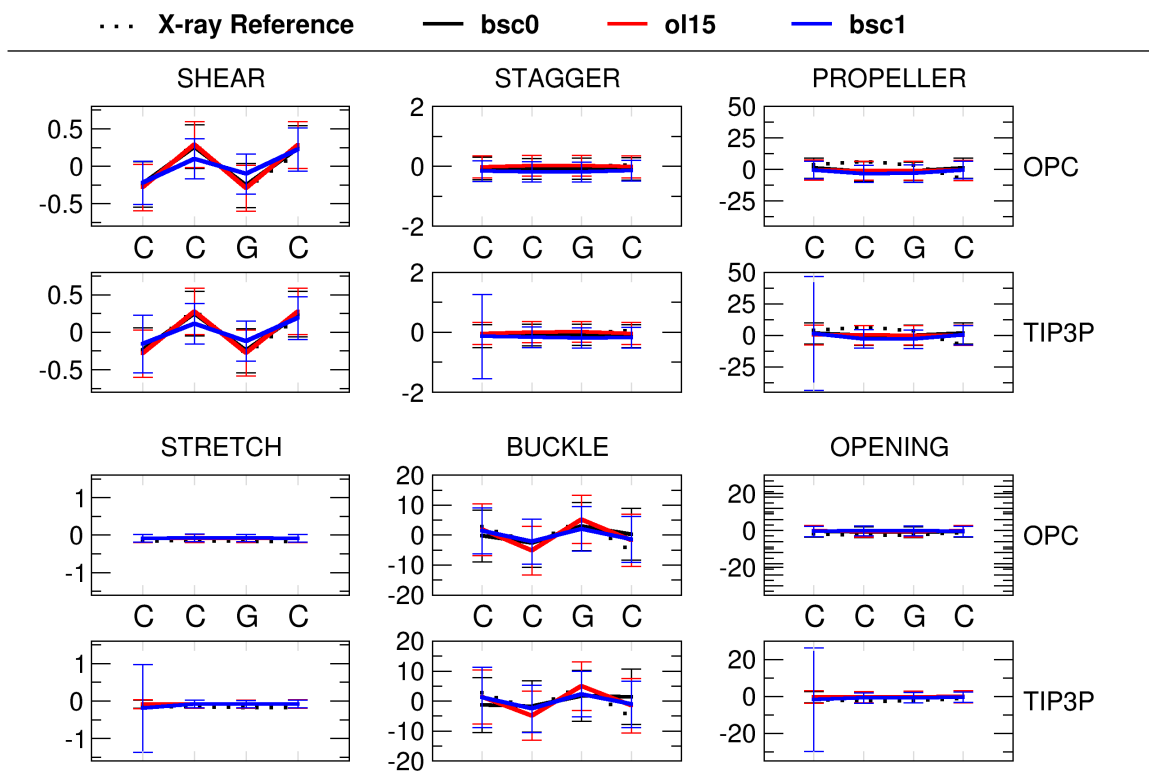


Figure 2.26: 110T base pair parameters. Each property was calculated using the CPP-TRAJ implementation of 3DNA.

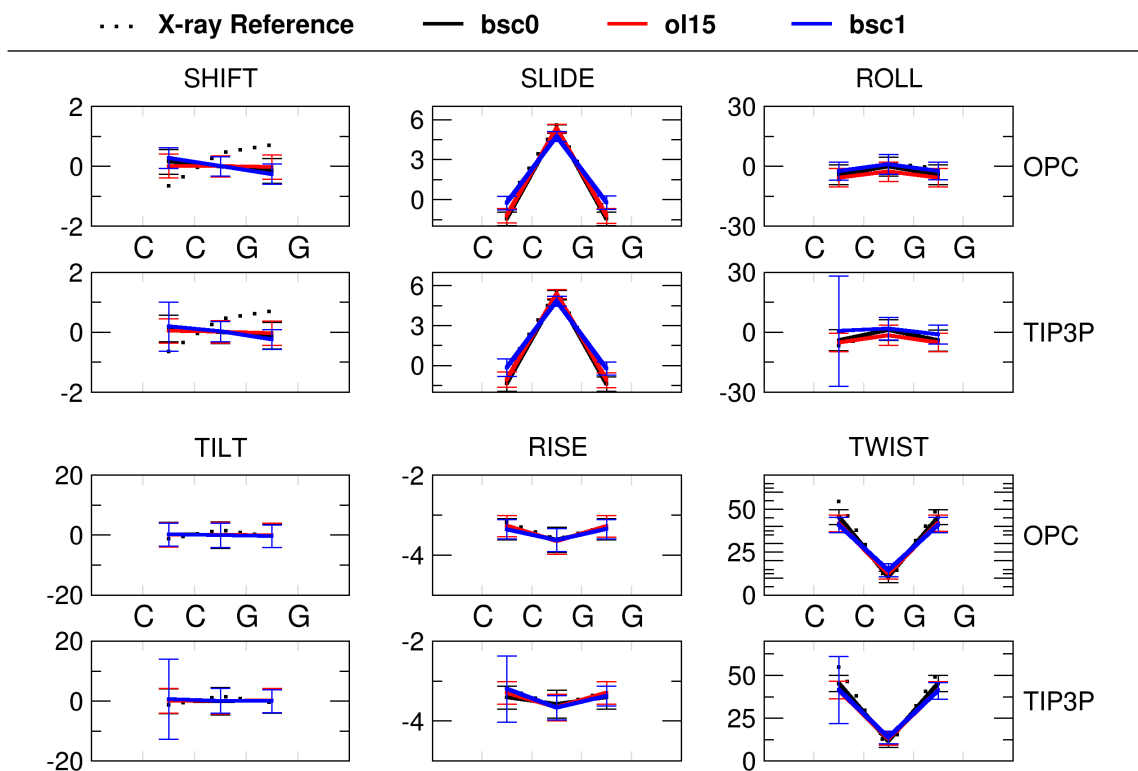


Figure 2.27: 110T base pair step parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

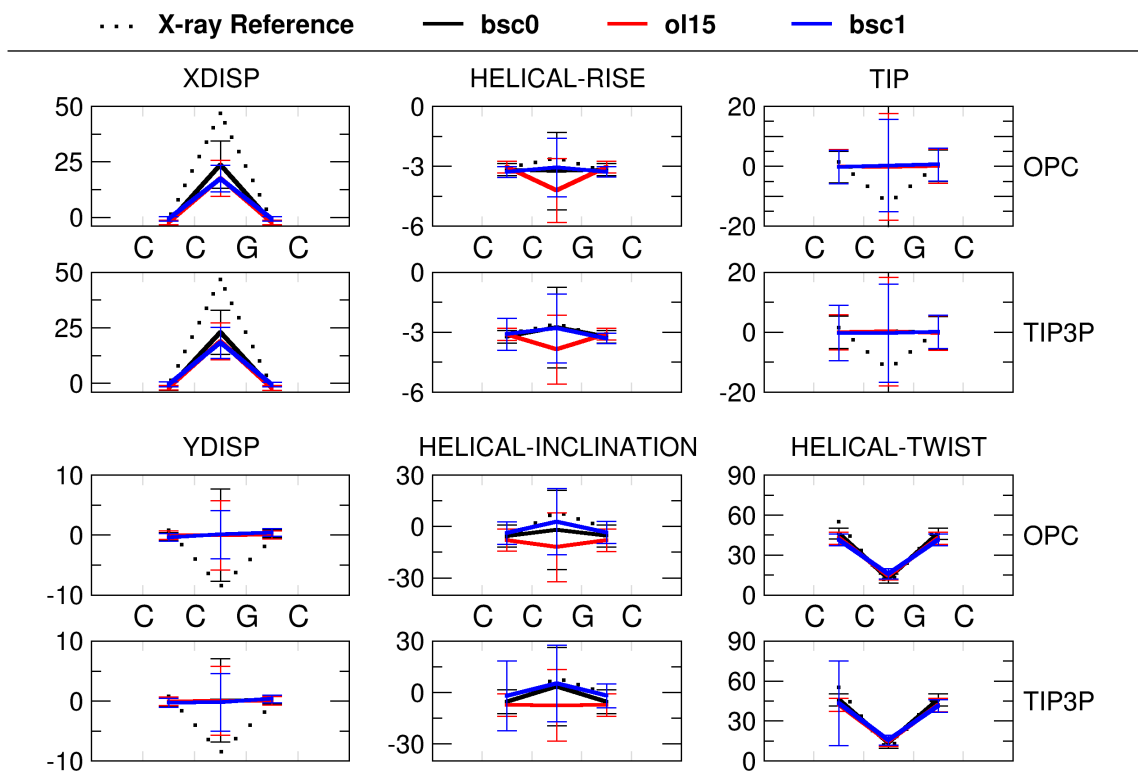


Figure 2.28: 1I0T helical parameters. Each property was calculated using the CPPTRAJ implementation of 3DNA.

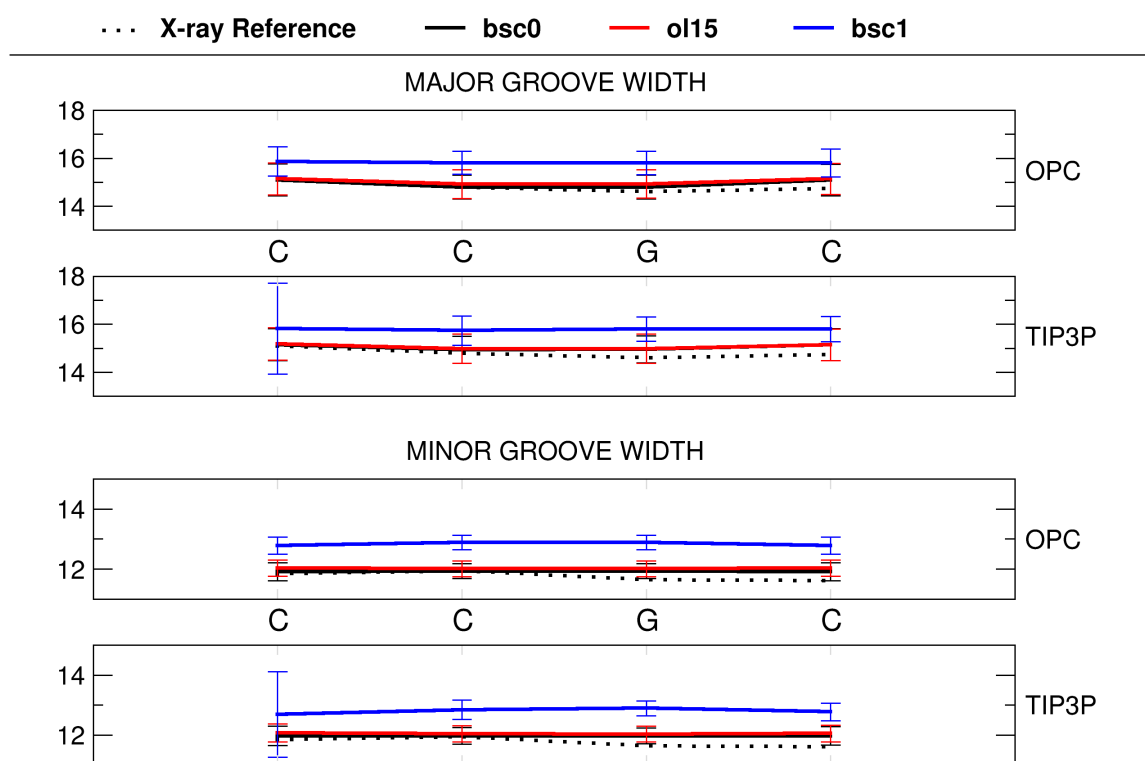


Figure 2.29: 1I0T groove widths. Each property was calculated using the CPPTRAJ implementation of 3DNA.

CHAPTER 3

DNA BACKBONE BI/BII DISTRIBUTION AND DYNAMICS IN E2 PROTEIN- BOUND ENVIRONMENT DE- TERMINED BY MOLECULAR DYNAMICS SIMULATIONS

3.1 Chapter Notes

This chapter was adapted with permission from:

James C. Robertson and Thomas E. Cheatham III. DNA Backbone BI/BII Distribution and Dynamics in E2 Protein-Bound Environment Determined by Molecular Dynamics Simulations *J. Phys. Chem. B*, **2015**, 119 (44) 14111-14119. Copyright 2015 American Chemical Society

J.C. Robertson and T.E. Cheatham III designed the experiments. J.C. Robertson ran the simulations, performed the analysis and wrote the manuscript. T.E. Cheatham III revised the manuscript.

3.2 Abstract

BI and BII conformational substates in the DNA backbone typify canonical B-form DNA. The BI and BII substates are important for structural variation of DNA and have been implicated in protein-nucleic acid recognition mechanisms. Recent refinements have been made to nucleic acid force fields employed in molecular dynamics simulations that demonstrate a better ability to model the BI and BII states, leading to overall improved modeling of DNA structure and dynamics. These force field improvements have yet to be significantly demonstrated in the context of a protein-DNA system extended to long time scales. Our plan was to run molecular dynamics simulations of a well-studied protein-DNA system (E2-DNA) into the microsecond time scale and determine the ability of the force field to populate BII states in the DNA backbone consistent with dinucleotide steps crystallized

in the BII conformation. The results showed that the dinucleotide steps in the E2-DNA complex with the highest BII populations from simulation trajectories corresponded to the dinucleotide steps crystallized in the BII state and that decoy BI and BII states converge to the same results within approximately one microsecond.

3.3 Introduction

Protein-DNA interactions are critical for the control and regulation of cellular function and are involved in many key aspects of biology, ranging from transcription and translation to genome repair and packaging. Despite considerable efforts in structural biology numerous details of the precise structural and dynamic interaction mechanisms between proteins and DNA remain poorly understood [84–86]. As an adjunct to high-resolution crystallography, cryo-electron microscopy, and nucleic magnetic resonance spectroscopy, molecular dynamics (MD) simulations can be a powerful tool to study these interactions. However, MD approaches have been limited by force fields that did not adequately represent nucleic acid structure and dynamics, and by insufficient sampling that restricted MD simulations to tens or hundreds of nanoseconds [87]. Recent improvements to the AMBER nucleic acid force field have overcome a number of the inadequacies exposed with previous force fields that, for example, underestimated helical twist, under populated BII backbone substates, and over-populated gamma = *trans*, the glycosidic chi in high *anti*, and particular epsilon/zeta backbone substates [20, 23, 88]. Now that the force field has been improved and advances in hardware and software have taken the state of the art in nucleic acid simulations to routine microsecond-length, [5–7, 16] it is important to determine how well simulations can reproduce protein-DNA interactions, particularly the DNA backbone BI and BII substates. As over half of protein-DNA interactions occur through side-chain interactions with the DNA backbone, and many involve selecting for or inducing different DNA backbone conformational substates, [89] this underscores the importance of accurate modeling of DNA backbone structure and dynamics.

The epsilon and zeta torsions in the DNA backbone undergo correlated transitions from t, g- (BI) to g-, t (BII); also defined as epsilon in the 120-210° range and zeta from 235-295° range for BI, and epsilon in the 210-300° range and zeta 150-210° range for BII [34, 35, 38, 90]. Alternatively, epsilon – zeta < 0 for BI and epsilon – zeta > 0 for BII. The BI/BII rearrangements are associated with changes in helical twist and roll, and changes in the geometry of the major groove [34, 35, 38, 87, 91–95]. Because of these dynamics and structural changes – further evidenced by the key role nonspecific contacts play in protein-

DNA interactions – the hypothesis has emerged that backbone structure and dynamics play as significant a role in protein-DNA interactions as base specific contacts [84, 85, 96–98]. Demonstrating that simulations with state of the art force fields, extended into the microsecond range with multiple copies to show statistical significance and reproducibility, can faithfully reproduce DNA backbone geometry is important to showing that MD is a valuable tool in studying protein-DNA interactions on the nanosecond to microsecond time scale. In this work we demonstrate the E2 protein interacting with its consensus DNA sequence and focus primarily on characterizing BI/BII substates through the course of the simulations.

The E2 protein from papillomaviruses contains a DNA binding domain (henceforth referred to simply as E2) that recognizes ACCN₍₆₎GGT consensus sequences (when N is any nucleotide) in the upstream regulatory region and acts as the master regulator of transcription [42, 44, 99–101]. The E2 protein forms a homodimer prior to DNA binding, and the recognition helices (helix 1) of each monomer insert into the major grooves at the conserved bases ACC and GGT of the consensus sequence, forming numerous specific interactions with bases and the DNA backbone [43]. Figure 3.1 shows the full DNA sequence. In the crystal structure, the noncontacted spacer sequence, GACGTC, lies between the two recognition half-sites on DNA. This sequence can vary, and there are believed to be 12 unique spacer sequences with various E2 binding affinities, but the GACGTC spacer sequence was included in the DNA crystallized with E2 by Hegde et al. in 1992. The E2-DNA system with GACGTC is well suited for studying BI/BII states in simulations for multiple reasons. A crystal structure [43] shows BII states at four locations (per strand), which allows us to see how well the force field reproduces this structural preference. The full DNA sequence is palindromic and the contacts with E2 are at the same locations in each of the E2 monomers so the results are expected to have symmetry, i.e. it is expected that BI/BII populations, and all other helical parameters and protein-DNA interactions, in one strand are identical, or very close to it, in the complimentary strand, providing additional evidence of reproducibility. Past studies have performed simulations studying BI/BII in E2-DNA but used force fields that underpopulated BII states, and the simulations were on time scales below what is required to converge BII populations, as determined in this study [102, 103]. The results of the previous studies allow us to show progress in the field due to both improvements in the force fields and also the computational methods that allow for long time scale sampling.

In this study we start with the native E2-DNA complex and run multiple copies of

microsecond length unbiased MD to determine whether BII states from the X-ray structure can be reproduced by simulation, and to determine the time scale required for the BII populations to converge. In addition, we manipulate the native state to two so-called decoy states that shift where the BI/BII states are located: decoy1 starts with the backbone entirely in the BI state, and decoy2 starts with the same number of BII states as the native state but in alternative positions on the backbone. The native and decoy states are found to converge to the same BII populations, those that match experiment, within $\sim 1 \mu\text{s}$, though even at $4 \mu\text{s}$ are not completely converged. DNA from the E2-DNA complex is also simulated in the absence of E2, hereafter referred to as free DNA, and the decoy states are also studied in free DNA. BII populations are found to converge on a shorter time scale and more completely than the E2-bound DNA, and the BII distributions in free DNA form a landscape different than that of E2-bound DNA. Essentially, we show that the initial BI/BII state of the DNA backbone in the protein-DNA complex does not matter, that with about $1 \mu\text{s}$ of sampling the BII populations converge to the same values, that the dinucleotide steps in the complex with the highest BII populations correspond to the steps crystallized in the BII state, and that free DNA and E2-bound DNA populate different BI/BII landscapes. The importance of BII substates in the E2-DNA recognition and binding mechanism is also discussed as well as differences in overall dynamics between the E2-bound DNA and free DNA.

3.4 Methods

3.4.1 Native State Molecular Dynamics Simulations

The coordinates for the E2-DNA complex were taken from PDB entry 2BOP, an X-ray crystal structure of the bovine papillomavirus E2 DNA-binding domain bound to a 17-mer duplex DNA at 1.7 \AA resolution [43]. BIOMT symmetry transformations were performed with the `mono2poly.tcl` script (http://www.ks.uiuc.edu/Research/vmd/script_library/scripts/mono2poly/mono2poly.tcl) to generate the biological state from the single strand structure present in the PDB. The X-ray structure had an unpaired base at the termini to assist with crystallization. To avoid end-effects which are common at the termini, especially for isolated bases [5, 16, 30], a guanine was built in at the 3' end of each DNA strand by editing the PDB file by hand to add the residue name, then the LEaP program from AmberTools [12] and AMBER residue templates were used to complete the DNA helix. For free DNA, the E2 protein coordinates were stripped away so that only DNA remained. The 18-mer DNA duplex sequences for both the free and complex DNA

was d(CCGACCGACGTCGGTCGG)₂.

The models were built with LEaP and parameterized using ff99bsc0 that contains the ff99 charge set and parmbsc0 α/γ dihedral modifications [20]. The χ OL4 chi dihedral modifications and ϵ/ζ OL1 with epsilon and zeta dihedral modifications were also used (the DNA parameter set currently recommended by AMBER developers) [22,23]. TIP3P explicit solvent [54] molecules were added to the periodic, truncated octahedron. Na⁺ ions with Joung/Cheatham [60] parameters were added to neutralize the system. The mass of all solute hydrogen atoms was shifted to 3.024 Da, and the mass of all heavy atoms attached to hydrogen atoms was adjusted so the total mass of each pair was unchanged. This was done using HMassRepartition in ParmEd from AmberTools, and this allowed for use of 4 fs integration time step during all dynamics [13].

Minimization, heating, and equilibration were run with PMEMD from AMBER 14 [67]. Structures were first minimized with 1000 steps of steepest descent followed by 1000 steps of conjugate gradient with 25 kcal/mol-Å² restraints on all solute atoms. The systems were then heated from 100 to 300 K over 200 ps, with a weak coupling thermostat using 1.0 ps coupling time [104] at constant pressure of 1 atm. Bonds involving hydrogen were constrained with SHAKE [105], a nonbonded cutoff of 9 Å was used, and PME was employed for long-range electrostatics with default parameters. Following heating, a series of minimization steps with decreasing restraints on solute atoms and equilibration steps with constant temperature and pressure were performed. Specifically, restraints were decreased from 5 kcal/mol-Å² to 1.0 kcal/mol-Å² over five minimizations, each with 500 steps of steepest descent followed by 500 steps of conjugate gradient. Restraints for equilibration steps followed the same scheme as the minimizations; each of these equilibration steps was 100 ps. Weak distance restraints were placed between terminal base pairs of DNA to prevent fraying and end effects. These were put in place for the final two equilibration steps: first with restraint energy of 0.5 kcal/mol for 1000 steps, the second with restraint energy of 3.0 kcal/mol for 5 ns. Restraints were between N3 of cytosine and the hydrogen of N1 on guanine, the lower and upper bounds of the square well were set to 1.0 and 3.0 Å respectively, with parabolic sides to 0 and 999 Å. The weak restraints at the termini were retained for production. Restraints were used for two reasons. First, dynamics that involve breaking of Watson-Crick base pairing at the termini are poorly understood in experiment and accurate modeling with force fields has proved difficult [30]. Second, the E2 protein binds to the DNA recognition site in a continuous strand of DNA, so we determined it was important to maintain the duplex, double-helical conformation of DNA as best we could,

even if it meant applying artificial restraints.

Prior to production, ion positions were randomly swapped with water molecules; ions were placed at least 4.0 Å from each other, and at least 6 Å from solute atoms. Randomization of ions was repeated for each system to create three unique copies. All production runs were on GPUs using PMEMD.cuda from AMBER 14 and AMBER 15 [12,67]. Production simulations were run at constant pressure using the Monte Carlo barostat and 1.0 ps pressure relaxation time. Constant temperature was maintained with Langevin dynamics [63] using collision frequency of 2 ps⁻¹; random starting velocities were used for all simulations. The nonbonded cutoff was set to 8.0 Å, and PME with grid spacing of 1 Å was used for calculating electrostatics. Coordinates were saved every 1000 steps.

3.4.2 Decoy State Molecular Dynamics Simulations

The procedure outlined above for native state systems is identical to that for decoy systems except additional steps were taken to manipulate the DNA backbone for the decoy systems prior to production. After equilibration, dihedral restraints were placed on specified epsilon and zeta dihedrals in DNA. For decoy1, the restraint energy was 1000 kcal/mol and a square well potential was used to keep all dihedrals within the BI range. The lower and upper bounds for epsilon were 120 and 210 °, while the lower and upper bounds for zeta were 235 and 295 °. Parabolic sides extended a single degree from each bound and were linear beyond that. The restraints were applied in conditions identical to the equilibration conditions above, but without positional restraints on the solute, and they were applied for 2500 steps.

Decoy2 was prepared in similar fashion to decoy1, but the restraints were applied so that epsilon and zeta were in the BII state at steps C₂pG₃, C₅pC₆, C₁₂pG₁₃, C₁₄pG₁₅. Lower and upper bounds for epsilon were 210 and 300 °, and 150 to 210 ° for zeta. Prior to production the ions were randomized as described above to create three copies. The decoy restraints were removed for production, and simulation conditions for both decoy states were identical to native state.

3.4.3 BI/BII Restraints at C₉pG₁₀pT₁₁ Steps

Restraints on epsilon and zeta dihedrals were placed at dinucleotide steps 9 and 10 in various combinations (discussed in Percent BII Distribution below). The restraints were placed in the same manner as described above for decoy1 and decoy2 except that the restraint energy was to 500 kcal/mol. These systems were prepared in similar fashion to the decoy systems, but the restraints were limited to dinucleotide steps 9 and/or 10. Each

system was reproduced to make three copies of each, and each copy was run in production MD for 80 ns per copy. The overall bend and percent BII were then calculated with Curves+.

3.4.4 Analysis of Trajectories and DNA Structures

CPPTRAJ [69] from AMBER 14 and AMBER 15 was used to analyze all trajectories. The DNA dihedral angles used to calculate BII substates were obtained using CPPTRAJ. Percent BII was calculated from dividing the number of frames with epsilon in the range 210-300 ° and zeta in the range 150-210 ° by the total number of frames, then multiplying by 100. Principal component analysis (PCA) was performed using CPPTRAJ; see refs [?, 69] for more on PCA. Groove widths, inter- and intrabase pair helical parameters, and bend were calculated with Curves+ [70]. DNA and E2-DNA structural representations were generated with VMD [71]. Porcupine plots were created with NMWiz [72], a VMD plugin for viewing vectors describing molecular motion.

3.5 Results and Discussion

3.5.1 Convergence of Percent BII for Native and Decoy Simulations

Regardless of the starting position of the BII state, whether it was native, decoy1 or decoy2, the percent BII at each dinucleotide step converged to the same value for that step in both free DNA and E2-bound DNA (Figure 3.1). For example, at G₃pA₄ of the protein-DNA complex, the native system started in the BII state while decoy1 and decoy2 started in the BI state, yet by 1 μ s all systems had reached 70% BII populations. Convergence on this time scale was observed across the DNA backbone for both free DNA and E2-bound DNA; however, the E2-bound DNA did converge slower than free DNA, especially at C₉pG₁₀ and G₁₀pT₁₁, which never converged as completely as free DNA. Figure 3.1B shows convergence for a single strand; however, the complimentary strand displayed similar trends, and by 1 μ s there were no differences between the two strands in free DNA. For E2-bound DNA at 1 μ s, there were differences less than 2 percentage points at 15 of 17 dinucleotide steps. Steps C₉pG₁₀ and G₁₀pT₁₁ differed by 10 and 8%, respectively, which was consistent with the overall poor convergence at these positions in the E2-bound DNA. The free DNA was extremely well converged by 1 μ s, but minor differences were present in the E2-bound DNA BII populations at C₉pG₁₀ and G₁₀pT₁₁ even at 4 μ s (Figure 3.2). The additional degrees of freedom in the protein influenced the dynamics of the DNA so that helical parameters, in addition to BII populations, were slower to converge (Figures 3.3 and 3.4). The heterogeneity of the DNA dynamics due to the presence of the protein was also apparent

by the differences in epsilon and zeta dihedrals measured on average structures of the native and decoy systems (Figure 3.1C, Tables 3.1 and 3.2). In contrast, there was nearly complete agreement of the epsilon and zeta values, not simply the percent BII, for free DNA.

The per step percent BII of native and decoys systems was converged by 1 μ s for free DNA and reasonably well converged by 1 μ s for E2-bound DNA, demonstrating that epsilon and zeta dihedrals artificially manipulated to BI or BII beginning states approach the same BI or BII conformation given sufficient sampling. The percent BII time series is now considered for native free DNA and E2-bound DNA. Not surprisingly, considering percent BII convergence of native and decoy systems, dinucleotide steps in free DNA quickly settle on percent BII values and show little fluctuation over time. This can be seen in Figure 3.5 for steps G₃pA₄, C₉pG₁₀, G₁₀pT₁₁, and C₁₆pG₁₇. E2-bound DNA, however, shows percent BII values fluctuate over time compared to free DNA. The four steps shown in Figure 3.5 are representative of the behavior of all dinucleotide steps and were chosen for the plot because of the relatively high percent BII at these positions compared to other steps, and these are the BII steps found in the reference structure. Regardless of the position of the backbone in E2-bound DNA, whether it was directly interacting with E2 or not, similar fluctuations in percent BII over time were observed. For example, step G₃pA₄ is in close proximity to the E2-DNA interface while step C₁₆pG₁₇ is on the opposite of the DNA as the E2-DNA interface. Both steps show percent BII values that fluctuate over time, suggesting that the overall dynamics of E2-bound DNA with respect to epsilon and zeta are influenced by E2, not just at the binding interface but at all steps. The E2-DNA complex alters DNA dynamics leading to longer convergence times for percent BII.

The root-mean-square-deviation (RMSD) of the simulations was quite stable over the 4 μ s of sampling per copy (Figure 3.6). This was true for both the free DNA and the protein-DNA complex. RMSD for the complex was tightly distributed about 2.3 Å, and RMSD for DNA hovered around 1.9 Å. The conformational variability of the complex due to E2 was evidenced by calculating RMSD of only the E2 component from the E2-DNA complex and observing larger jumps in RMSD and less agreement between copies compared to RMSD for the DNA component from the complex. Visual inspection of the trajectories confirmed stability and that the E2-DNA complex did not dissociate during the simulations.

3.5.2 Percent BII Distribution

Figure 3.7 shows that BII substates are most favorable in the E2-bound DNA at C₁pC₂, G₃pA₄, C₉pG₁₀, G₁₀pT₁₁, and C₁₆pG₁₇. The remaining steps have BII states accessed through the simulation but with generally less than 10% BII populations, and some with

effectively no BII population. The steps crystallized in the BII state had the largest BII populations in the simulations of the E2-DNA complex, implying accurate modeling of the epsilon and zeta torsions by the force field. The exception to agreement with the experimental structure is at C₁pC₂. In the X-ray structure C₁ is an unpaired base, a tactic commonly used in crystallizing protein-DNA complexes, but in simulations the DNA chains were extended to be fully base-paired duplex DNA for the length of the 18-mer. Though we cannot compare to experimental data at this step, relatively high BII populations were predicted.

The major grooves of free DNA that face the E2 binding site are lined with dinucleotide steps with high BII populations (Figure 3.7A). The G₃pA₄ step is required for binding, the C₁₂pG₁₃ step contains the preferred C₁₂ and required G₁₃ bases, and the G₁₃pG₁₄ step contains bases required for binding. These all have high BII populations in free DNA. Although these BII populations are shifted in the E2-bound DNA, the flexibility of the backbone in those positions may play a necessary role in the sequence specific recognition process. In addition, the elevated BII populations at these positions leads to increased exposure to the bases and the specific contact points in the major groove, possibly a key part of the mechanism for protein binding.

Steps C₁pC₂ and G₃pA₄ both have elevated BII populations in the E2-DNA complex relative to free DNA, and both have backbone interactions with conserved E2 residues. The oxygen OP2 of A4 hydrogen bonds with N336, a residue in the recognition helix shown to be required for DNA binding [42]. The amine group of N336 interacts with the phosphate and N7 of A4, while the carboxyl oxygen interacts with the amine group of C5. BII geometry at G₃pA₄ allows these simultaneous interactions while a shift to BI would alter the distances and angles of each interaction, possibly reducing binding affinity. At nearby step C₁pC₂ arginine 344 interacts with the OP2 of C2 through both the epsilon nitrogen and one from the guanidinium group. R344 mutants have been shown to be DNA binding defective due to the hydrogen bonding involved in the direct readout at this location. These conserved interactions serve as a potential explanation for increased BII at these steps in the E2-bound DNA relative to free DNA. The X-ray structure shows good agreement with average values from simulation for hydrogen bonding distances (heavy atom to heavy atom) and angles for ASN336 and ARG344 to DNA. The top 3 hydrogen bond interactions with DNA for residues ASN336 and ARG344 are shown in Table 3.3.

The increased BII populations of G₁₀pT₁₁, and the neighboring BII population at C₉pG₁₀, in the protein-DNA complex suggests that the capacity to assume BII geometry

in those positions could play an important role in the ability of DNA to adopt, or at least maintain, the 30° bend around E2, and possibly, the flexibility at those positions expressed as transitions between BI and BII offset the entropic penalty of complex formation by having relatively high atomic fluctuations (see DNA Dynamics in Free DNA vs. E2-bound DNA below). The BII population of $G_{10}pT_{11}$ in free DNA is only 1% while in the protein-DNA complex BII population at that same step increases to nearly 20%. To investigate the association between increased BII substates at the $G_{10}pT_{11}$ step of the protein-DNA complex and the increased bending of DNA, short simulations were run on free DNA with artificial restraints to impose BI or BII states at steps C_9pG_{10} and $G_{10}pT_{11}$, and the overall bend of DNA was calculated and compared between these systems (the full list of systems and results are in Table 3.4). It was found that when BII substates were induced and maintained at C_9pG_{10} and $G_{10}pT_{11}$, either on a single chain or both at the same time, the overall bend increased relative to unrestrained DNA. Free DNA without restraints had an average overall bend of $13.4 \pm 7.2^\circ$. With BII restraints at both C_9pG_{10} and $G_{10}pT_{11}$ on both strands, the average bend increased to $20.9 \pm 9.9^\circ$ and the maximum bend was nearly equal to the maximum bend observed in the E2-bound DNA. Restraints on only a single strand at C_9pG_{10} and $G_{10}pT_{11}$ also led to increased bending ($18.1 \pm 9.1^\circ$), while all other BI/BII restraint combinations at C_9pG_{10} and $G_{10}pT_{11}$ had bend values comparable to unrestrained, free DNA. Restraining BII at either step individually led to little increase in bending, though systems with BII restraints at $G_{10}pT_{11}$ had slightly higher average bends.

BII geometry facing BII geometry on opposite strands of DNA (denoted BII.BII) has been shown to be rare [35, 106], but the X-ray structure of the E2-DNA complex shows BII geometry at C_9pG_{10} , directly across from the complimentary CpG step, also with BII geometry. The percent of BII.BII occurrences in simulations were calculated to determine if this, or other, steps had relatively high BII.BII. It was determined that the terminal steps actually had the highest percent BII.BII in the E2-bound DNA, but the terminal steps and the central C_9pG_{10} step were nearly equivalent (Figure 3.8). All other steps in the E2-bound DNA had simultaneous occurrence of BII geometry in opposite facing steps less than 0.5% of the simulation. In contrast, the free DNA had a distinctly different landscape of BII.BII geometries. The internal C_9pG_{10} steps had nearly 5% population of BII.BII, more than double the same steps in the complex DNA. The relatively high levels of BII.BII percentage extended from the termini into the first two steps; in free DNA the two terminal steps each had nearly 2% BII.BII while only the first step in complex DNA had elevated BII.BII and the next step was nearly 0. Also interesting to note was that the recognition half-sites on the

free DNA had much higher occurrences of BII.BII compared to the complex. In E2-bound DNA the BII.BII percent at the half-sites was nearly zero while C₅pC₆, C₆pG₇, C₁₂pG₁₃, and G₁₃pG₁₄ in free DNA each had elevated BII.BII geometry.

Occurrences of BII in consecutive dinucleotide pairs were also found to be quite rare [106], but, in agreement with the X-ray structure that had BII states at C₉pG₁₀ and G₁₀pT₁₁, the most common dinucleotide neighbors in the complex simulations were at the same C₉pG₁₀ and G₁₀pT₁₁ steps. This implies that E2-bound DNA has increased flexibility and ability to adopt rare conformations, evidenced by the neighboring BII geometries. The ability to adopt this unique conformation may play a role in the E2-DNA recognition mechanism and/or differential binding affinity of this spacer sequence compared to other E2 consensus sequence sites. The other neighboring dinucleotide pair with elevated concurrent BII geometries in the complex was the pair at the 3' terminus, which has G18 built in for the simulations so, as mentioned earlier, it could not be compared to the X-ray structure because that final base was unpaired. Figure 3.8 shows the percentages of simultaneous BII geometries in consecutive dinucleotide steps. The neighboring BII states in the spacer region of the E2-bound DNA was clearly induced by the E2 interaction, as neighboring BII states in the C₉pG₁₀ - G₁₀pT₁₁ neighbors were extremely rare in free DNA. Occurrences of consecutive C₉pG₁₀pT₁₁ steps in the BII substate on both strands was completely absent in free DNA and extremely rare, only 0.01% of frames, in the protein-DNA complex. Also compelling in the free DNA is that the highest occurrences of neighboring BII states have a 3' pyrimidine while the lowest occurrences have a 3' purine. For example the top five highest are C₁pC₂pG₃, C₂pG₃pA₄, C₆pG₇pA₈, C₁₂pG₁₃pG₁₄, and C₁₆pG₁₇pG₁₈ while the five lowest are A₄pC₅pC₆, C₉pG₁₀pT₁₁, G₁₀pT₁₁C₁₂, G₁₃pG₁₄pT₁₅, and G₁₄pT₁₅pC₁₆.

3.5.3 Frequency and Lifetime of BII States

Lifetimes of BII states were calculated for all steps in both free DNA and E2-bound DNA. Snapshots from the MD trajectory were saved every 4 ps, so any BI/BII transitions that occurred under this time scale were not accounted for during analysis. With the snapshots that were collected, however, a BII lifetime was considered present for any occurrence of a single frame in the BII conformation, and the frames were then converted to time. For example, if five sequential frames went BI, BII, BI, BII, BII then there would be 2 BII lifetimes, the first lasting a single frame (4 ps) and the second lasting 2 frames (8 ps). The average lifetime of BII states varied from step to step but was short-lived, generally around 10 ps (Table 3.5). Long-lived BII states were generally limited to 400 ps with the exception of G₁₀pT₁₁ in E2-bound DNA, which averaged closer to 800 ps. The differences between

average and maximum lifetimes of the free DNA compared to E2-bound DNA was minimal with the exception of the high maximum lifetime for G₁₀pT₁₁ of complex.

The rate of transitions to the BII state strongly correlates to the overall BII population (Figure 3.9). This demonstrates that a dinucleotide step with high BII population does not have the high population due to relatively few, extended visits to BII. Rather, the high BII populations are due to a high frequency of transitions to the BII substate. The correlation between percent BII and rate of transitions was stronger in free DNA compared to E2-bound DNA.

Just as the BII populations were strongly influenced by the presence of the protein, the frequency of transitions was as well. It cannot be said that the protein environment reduced the frequency of BII transitions, nor can the opposite be concluded from the data. Rather, there was a position dependent effect. In some cases the frequency of transitions was increased in the presence of E2, such as steps C₁pC₂, G₃pA₄, G₁₀pT₁₁, and C₁₆pG₁₇. And in other steps, the frequency was reduced in presence of E2, such as in steps C₂pG₃, C₁₂pG₁₃, and G₁₃pG₁₄. In C₉pG₁₀, the frequency was nearly the same for free DNA and E2-bound DNA.

3.5.4 DNA Dynamics in Free DNA vs. E2-Bound DNA

The primary focus of this study was the ability of our simulations to reproduce high BII populations in the protein-DNA complex that corresponded to the dinucleotide steps crystallized in the BII substate, to determine whether artificial decoy states would converge to same BI/BII distribution as the native system, and to then compare the BI/BII properties between free DNA and the protein-DNA complex. Now, we examine overall dynamics of DNA in the presence and absence of the E2 protein partner. Principal component analysis (PCA) was performed on all heavy atoms of DNA to determine the dominant, slow modes of motion in DNA (protein atoms were excluded from the calculation on the protein-DNA complex).

The primary modes in free DNA were bending and twisting, as expected (Figure 3.10). In contrast, these modes of motion were absent in the top three principal components of the E2-bound DNA. Rather the first principal component (PC) in E2-bound DNA was centered about the backbone of the spacer region, the second PC was on the backbone opposite the E2-DNA interaction interface, and the third PC was also opposite the E2-DNA interface and corresponded to stretching that increases the major groove opposite the E2 protein and compresses the minor groove facing E2 (see Figure 3.10A). The first principal component

of E2-bound DNA is commensurate with the CGT sequence in the spacer region, which has elevated BII populations in the simulations.

Motivated by the apparent backbone motions in the spacer region of E2-bound DNA determined from PCA, epsilon and zeta dihedrals were calculated on the pseudotrajectory generated from PCA. It was found that motions of the E2-bound DNA spacer region shown in the porcupine plot for PC-1 of Figure 3.10A are from BI/BII transitions. Epsilon and zeta dihedral measurements on the pseudotrajectory for PC-1 are shown in Figure 3.10B for both free DNA and E2-bound DNA. Dinucleotide step C₉pG₁₀ shows a transition from BI to BII in both strands of DNA, while step G₁₀pT₁₁ undergoes transition from BII to BI. These motions underscore the different dynamics between free DNA and E2-bound DNA while highlighting the importance of BI/BII transitions in the spacer region. The significance of BI/BII transitions in E2-bound DNA is not well understood and requires further investigation. For example, it is not clear if these transitions as primary modes of motion in the E2-bound DNA are observable in other spacer DNA sequences and in E2 proteins from other papillomavirus proteins. Further investigation should be aimed at better understanding the importance of BI/BII transitions in E2-DNA binding and whether the simulations accurately capture biological motions in this system.

3.6 Conclusion

The improved force fields combined with the ability to run multiple simulations into the low microsecond time scale allowed us to show that the highest populations of BII geometry in simulations corresponded to the dinucleotide steps that were crystallized in the BII state; this was true for DNA in the E2-DNA complex. The same DNA sequence was simulated in the absence of E2 protein and displayed a BII landscape different than that of the complex, implying E2 induces backbone geometry favorable to the bound state and the ability of the sequence to adopt a specific BII landscape plays an important role in binding affinity. When decoy backbone states were created as starting structures for decoy1 and decoy2 simulations, the BII populations converged for both free DNA and E2-bound DNA, albeit more slowly in the E2-bound DNA. The dinucleotide steps in the spacer region between recognition half-sites were particularly slow to converge BII populations, and even at 4 μ s were not as well converged as any step in free DNA at 1 μ s. The influence of additional complexity of protein dynamics causes poor convergence for helical parameters, relative to free DNA. Interactions between the recognition helices of E2 and the half-sites of the DNA are maintained throughout the simulations, i.e., binding was never disrupted. The

spacer region, however, experienced higher fluctuations relative to half-sites compared to free DNA, and more conformational diversity as shown in the poor overlap between helical parameters of native and decoy systems.

C₉pG₁₀ and G₁₀pT₁₁ deserve special attention because they were crystallized with both BII.BII facing geometry and consecutive BII states, each a relatively rare occurrence yet both captured in the 2BOP crystal. These steps were also remarkable in the simulations. For the E2-bound DNA, they had the highest levels of BII.BII states as well as highest population of consecutive BII states. Even though the BII.BII and consecutive BII populations for each of the four steps (two per strand) were very low, the four steps each had about 20% overall BII. This means that although the probability of finding more than one BII state in those four dinucleotide steps was quite low, it was extremely high to find at least one BII state, meaning the backbone in that spacer region of the DNA was constantly transitioning from BII in one step to BII in another step whether it was directly across, across and down a step, or in the neighboring step. The primary mode of motion from the first principal component shows this motion in the E2-bound DNA. While the primary mode in free DNA was bending, the primary mode in E2-bound DNA was BI/BII transitions at steps C₉pG₁₀ and G₁₀pT₁₁.

The simulations show that microsecond time scales are required to converge BII populations and that the latest nucleic acid force fields (a full nucleic acid force field comparison is beyond the scope of this article, but a brief force field comparison regarding BI/BII populations is discussed in Figure 3.11) reproduce BI/BII backbone geometry in an E2-DNA crystal structure. Manipulating the epsilon and zeta dihedrals to decoy states does not hinder the ability of the backbone to converge BII populations within a microsecond for free DNA, and 4 μ s for E2-bound DNA. Overall the modes of motions for DNA from the protein-DNA complex were quite different from free DNA, and BI/BII substates required longer time scales for convergence.

Table 3.1: Epsilon and zeta from average structures: free DNA. Epsilon and zeta dihedrals for native, decoy1, and decoy2. Each is from the zeta-epsilon data in Figure 3.1, i.e. the epsilon and zeta were measured from an average structure. The epsilon for free DNA shows nearly complete convergence but the complex values differ. Data shown for W strand only.

	Epsilon			Zeta		
	Native	Decoy1	Decoy2	Native	Decoy1	Decoy2
CpC	192.8	192.6	192.9	255.2	255.6	255.6
CpG	200.3	200.2	201.0	251.3	251.9	250.2
GpA	221.7	221.7	220.8	208.0	208.2	209.4
ApC	180.5	180.5	180.3	264.3	264.6	264.3
CpC	187.6	187.5	187.6	260.4	260.6	260.3
CpG	199.7	199.6	199.8	258.4	258.1	257.5
GpA	229.6	228.6	229.1	198.5	199.7	199.1
ApC	181.9	182.0	182.0	268.1	267.9	268.2
CpG	203.9	203.4	203.9	231.2	231.8	231.0
GpT	180.4	180.5	180.5	264.2	264.1	264.2
TpC	187.5	187.5	187.7	267.4	267.5	267.1
CpG	205.5	206.1	206.0	234.0	232.9	233.4
GpG	196.5	196.3	196.2	238.3	238.4	238.9
GpT	180.8	180.7	180.8	265.4	265.3	265.5
TpC	188.4	188.4	188.3	267.4	267.3	267.3
CpG	206.9	207.2	206.4	232.5	232.1	234.2
GpG	195.9	196.3	196.8	243.6	243.1	243.8

Table 3.2: Epsilon and zeta from average structures: E2-bound DNA. Epsilon and zeta dihedrals for native, decoy1, and decoy2. Each is from the zeta-vs.-epsilon data in Figure 3.1, i.e. the epsilon and zeta were measured from an average structure. The epsilon for free DNA shows nearly complete convergence but the complex values differ. Data shown for W strand only.

	Epsilon			Zeta		
	Native	Decoy1	Decoy2	Native	Decoy1	Decoy2
CpC	260.1	251.7	253.0	170.0	182.2	179.1
CpG	185.9	192.4	192.2	279.4	268.2	270.0
GpA	236.5	236.2	234.1	183.5	185.1	186.5
ApC	180.1	179.7	179.9	272.7	272.6	273.0
CpC	197.0	196.7	196.8	244.1	244.3	244.1
CpG	198.5	198.4	198.7	270.2	268.8	268.3
GpA	172.6	172.4	176.4	258.8	258.9	253.8
ApC	180.6	180.9	181.3	268.1	268.4	267.7
CpG	217.8	223.9	221.3	219.2	210.3	214.0
GpT	207.6	194.4	195.2	229.8	243.1	242.5
TpC	172.6	176.8	174.0	265.2	262.5	264.6
CpG	198.5	197.7	198.9	282.2	279.3	278.9
GpG	179.5	179.0	178.1	272.0	272.3	272.1
GpT	177.0	177.4	177.3	257.0	256.7	257.1
TpC	188.7	188.7	188.7	271.6	271.1	271.0
CpG	220.7	219.3	218.4	208.6	209.9	211.4
GpG	184.7	184.3	185.0	260.7	261.4	260.1

Table 3.3: Hydrogen Bonding Interactions from E2-DNA Simulations and X-ray Reference Structure. Hydrogen bonds between two conserved residues in the recognition helix and DNA are shown. The H-bond fraction is based on the presence of hydrogen bond within angle cutoff of 135° and heavy atom distance of 3 Å.

Atoms E2	Atoms DNA	H-bond Frac. MD	Dist. X-ray	Dist. MD	Angle X-ray	Angle MD
ASN336 OD1	C5 N4	0.78	2.8	2.8	163.2	158.2
ASN336 ND2	A4 N7	0.41	2.9	2.9	159.1	155.0
ASN336 ND2	A4 O2P	0.39	3.0	2.8	154.0	158.1
ARG344 NH2	2C O2P	0.78	3.1	2.8	150.9	152.3
ARG344 NE	2C O2P	0.62	3.3	2.8	148.3	150.7
ARG344 NH1	13G N7	0.27	3.6	2.9	111.5	155.6

Table 3.4: Overall bend (in degrees) and percent BII from the internal 12 bp of unrestrained E2-bound DNA and free DNA, as well as free DNA with restraints at various combinations of steps 9 and 10. For example, CpGpT-BI means that both CpG and GpT steps on both strands were restrained to BI state, CpGpT_W-BI means that those steps were restrained on only a single strand etc.

System	Total Bend (°)					Overall Percent BII
	Avg.	Std.Dev.	Range	Min.	Max.	
E2-bound DNA	44.0	5.7	42.0	21.7	66.4	11
Free DNA	13.4	7.2	45.0	0.1	48.3	18
CpGpT-BI	13.1	6.9	48.0	0.1	51.8	14
CpGpT-BII	20.9	9.9	63.0	0.1	64.2	33
CpGpT_W-BI	13.5	7.1	49.0	0.0	49.3	16
CpGpT_W-BII	18.1	9.1	57.0	0.0	57.2	27
CpG-BI	13.3	6.9	47.0	0.1	48.8	15
CpG-BII	14.2	7.5	50.0	0.1	53.6	24
CpG_W-BI	13.4	7.1	49.0	0.1	52.3	17
CpG_W-BII	13.7	7.2	55.0	0.0	55.3	20
GpT-BI	13.4	7.1	48.0	0.1	49.2	18
GpT-BII	14.1	7.5	55.0	0.0	61.2	23
GpT-BI_W-BI	13.5	7.3	51.0	0.1	55.4	18
GpT-BI_W-BII	14.9	7.9	50.0	0.1	50.3	22

Table 3.5: Average lifetime of BII states per dinucleotide step, the maximum BII lifetime per step, and the frequency of BII states per step for free DNA and E2-bound DNA. Values are averages of three simulation copies with standard deviations.

	Avg Lifetime (ps)		Max Lifetime (ns)		BII Substates (ns ⁻¹)	
	DNA	E2-DNA	DNA	E2-DNA	DNA	E2-DNA
CpC	9.2 ± 0.0	8.9 ± 0.1	0.12 ± 0.12	0.18 ± 0.03	11.81 ± 0.06	38.46 ± 2.21
CpG	14.8 ± 0.1	9.2 ± 3.1	0.25 ± 0.08	0.22 ± 0.11	11.94 ± 0.19	0.41 ± 0.06
GpA	15.6 ± 0.0	18.5 ± 0.9	0.21 ± 0.02	0.40 ± 0.03	27.73 ± 0.18	38.09 ± 1.23
ApC	9.6 ± 0.1	7.8 ± 0.4	0.10 ± 0.01	0.03 ± 0.01	3.40 ± 0.04	0.02 ± 0.01
CpC	10.9 ± 0.1	9.3 ± 0.2	0.19 ± 0.03	0.28 ± 0.05	5.70 ± 0.03	16.50 ± 1.35
CpG	16.2 ± 0.1	9.8 ± 0.2	0.28 ± 0.10	0.19 ± 0.04	8.05 ± 0.02	2.85 ± 0.55
GpA	15.8 ± 0.0	8.0 ± 0.2	0.22 ± 0.00	0.12 ± 0.02	30.59 ± 0.02	8.42 ± 0.73
ApC	10.7 ± 0.2	9.4 ± 0.1	0.13 ± 0.02	0.10 ± 0.02	1.71 ± 0.04	2.48 ± 0.29
CpG	13.0 ± 0.1	11.7 ± 0.7	0.33 ± 0.07	0.28 ± 0.08	22.68 ± 0.08	22.82 ± 3.19
GpT	7.0 ± 0.1	11.0 ± 0.9	0.12 ± 0.09	0.90 ± 0.62	1.96 ± 0.03	16.60 ± 4.14
TpC	8.9 ± 0.1	6.4 ± 1.0	0.13 ± 0.02	0.30 ± 0.15	1.71 ± 0.02	0.56 ± 0.13
CpG	15.9 ± 0.1	7.9 ± 2.8	0.32 ± 0.06	0.14 ± 0.11	19.48 ± 0.05	0.14 ± 0.12
GpG	15.2 ± 0.1	15.5 ± 4.6	0.22 ± 0.02	0.09 ± 0.05	16.21 ± 0.13	0.01 ± 0.00
GpT	6.5 ± 0.1	6.2 ± 0.3	0.07 ± 0.01	0.08 ± 0.01	1.10 ± 0.02	4.88 ± 1.10
TpC	8.3 ± 0.1	7.3 ± 0.1	0.12 ± 0.02	0.07 ± 0.01	1.64 ± 0.04	0.10 ± 0.04
CpG	16.0 ± 0.1	16.3 ± 0.3	0.26 ± 0.07	0.30 ± 0.05	19.78 ± 0.17	30.63 ± 1.35
GpG	14.3 ± 0.0	12.4 ± 0.2	0.18 ± 0.04	0.16 ± 0.03	14.97 ± 0.26	6.39 ± 1.04

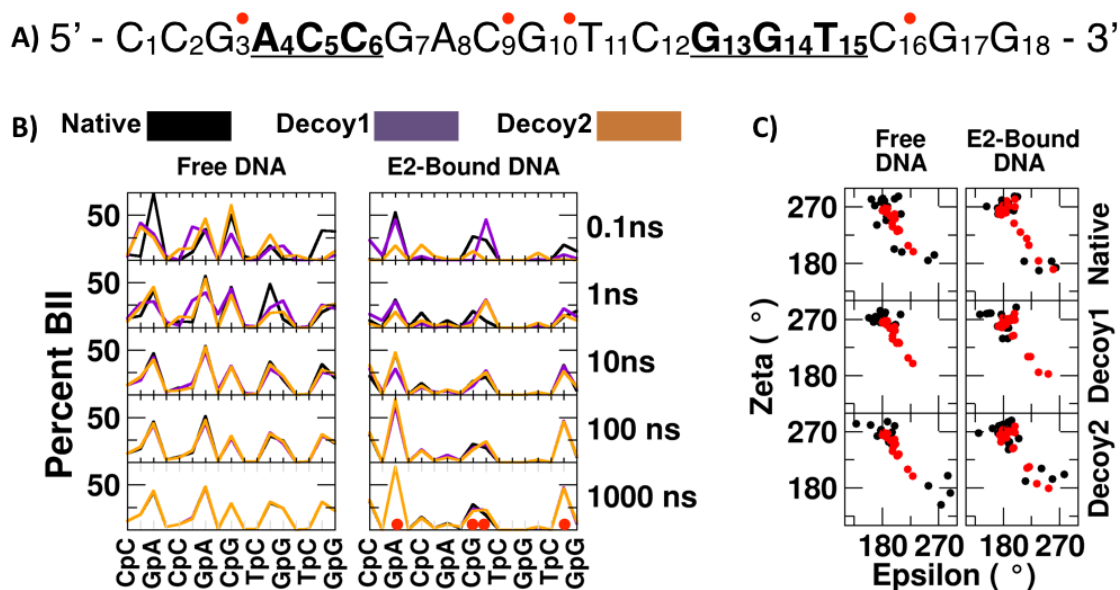


Figure 3.1: Convergence of percent BII over time. (A) The nucleotide sequence of DNA used in all free DNA and E2-bound DNA simulations. The numbering scheme shown is used throughout the manuscript to indicate position of each base in the sequence. The consensus half-sites are bolded, and the spacer region lies between the consensus half-sites. Red dots are used to indicate the location of steps crystallized in the BII conformation. Note the palindromic nature of the sequence. A single strand is shown, however, all simulations were of duplex DNA. (B) The convergence of percent BII per dinucleotide step is shown for the native (black), decoy1 (purple), and decoy2 (orange) systems. Percent BII was measured at each step for the indicated time interval. Every other step is labeled (e.g., CC is step 1, GA is step 3, etc.). The red dots at 1000 ns indicate BII positions in the X-ray structure. (C) Epsilon and zeta dihedral angles for each step were measured prior to production (black) and on an average structure obtained over the final microsecond of simulation (red). Epsilon and zeta values in the upper left of each plot are BI, while those in the lower right are BII.

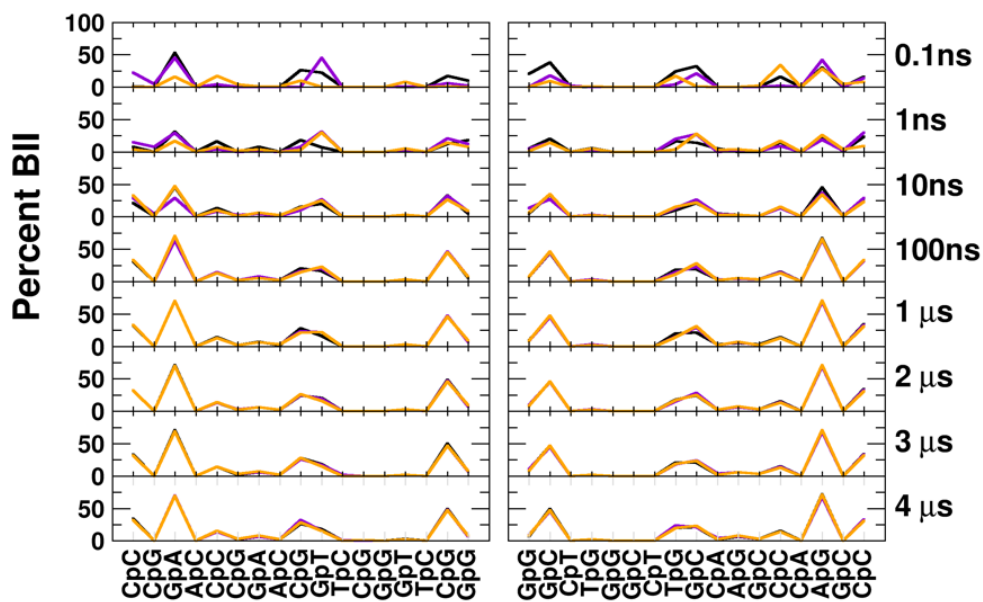


Figure 3.2: Convergence of percent BII per dinucleotide step over time for protein-DNA complex. The convergence of percent BII per dinucleotide step is shown for the E2-bound DNA native (black), decoy1 (purple), and decoy2 (orange) systems. Percent BII was measured at each step for the indicated time interval.

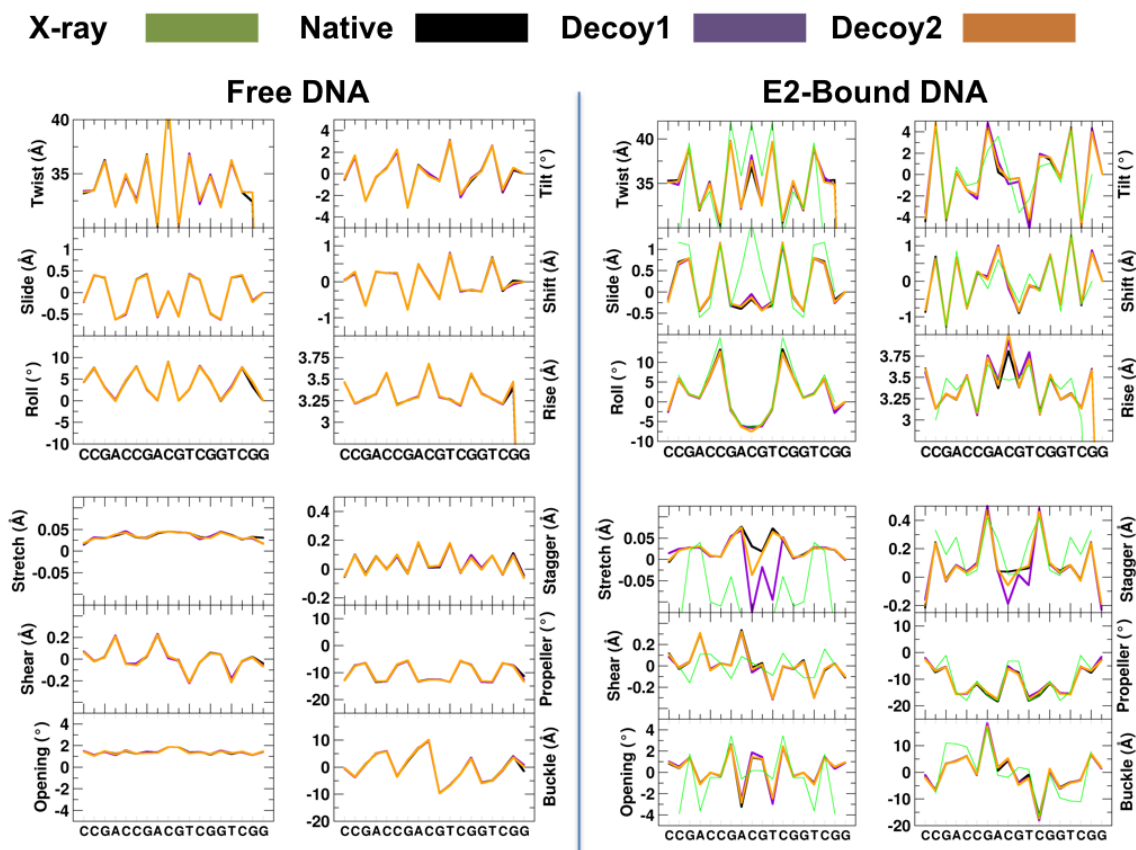


Figure 3.3: Helical parameters for free DNA and complex DNA: native and decoys. Helical parameters were calculated with Curves+ for free DNA and E2-bound DNA. In the case of the E2-bound DNA, the helical parameters were also calculated for the experimental X-ray structure (PDB ID 2BOP). The comparison to the experimental X-ray values for free DNA can be made by looking at the E2-bound reference values, but were excluded from the free DNA plots because they are not expected to be in agreement, since the free DNA and E2-bound DNA are expected to have different structures and helical parameters. A discussion on the agreement between helical parameters from experimental structures and averages from simulation for free DNA can be found in reference [23]. Helical parameters for native, decoy1, and decoy2 states for free DNA (left) and DNA from E2-bound DNA (right); the values for the X-ray structure are included for the E2-bound DNA. Helical parameters from simulations were calculated over all three copies for each system. Near complete overlap is observed for all helical parameters of the DNA systems, while differences in E2-bound DNA are apparent, especially at the middle bases. Data is shown for a single strand in all cases, though the same trends were observed in both strands.

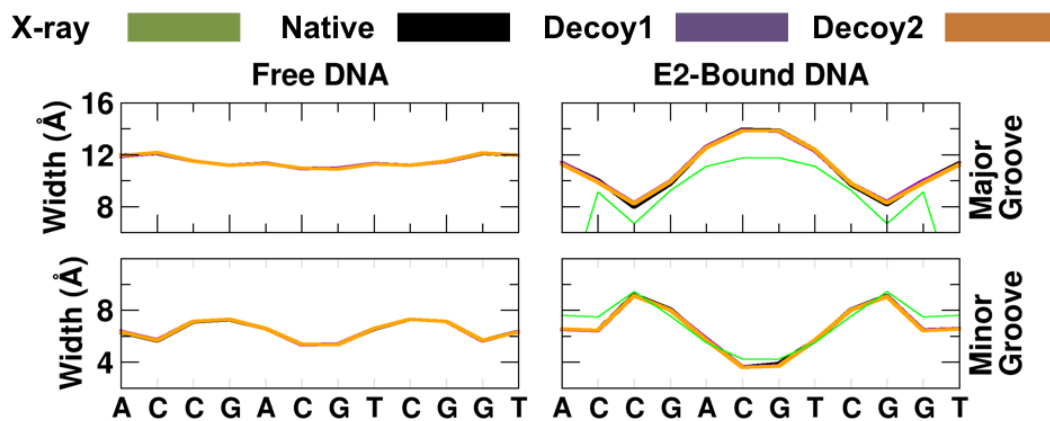


Figure 3.4: Groove widths for free DNA and complex DNA: native and decoys. the values for The X-ray structure are included for the E2-bound DNA (values are shown for the internal bases 4-15).

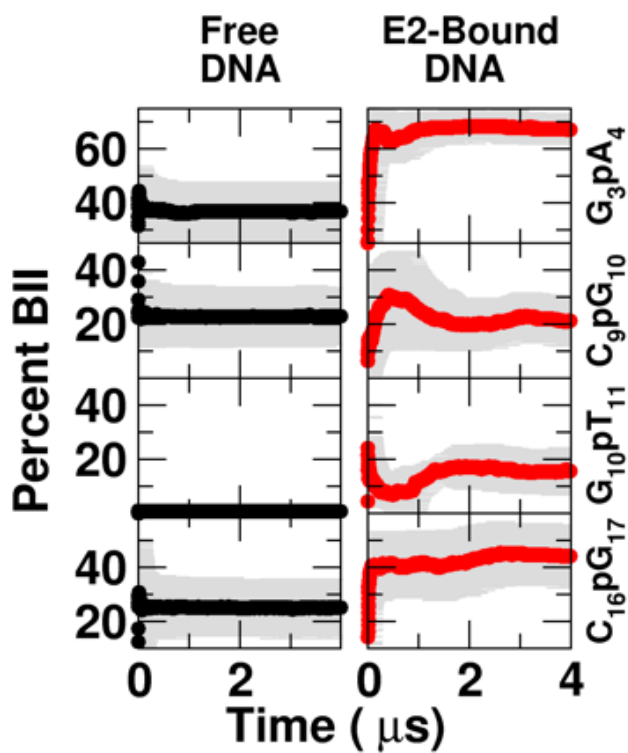


Figure 3.5: Convergence of percent BII for over time for 4 steps. Percent BII over time for steps 3 (GpA), 9 (CpG), 10 (GpT), and 16 (CpG) converges more slowly for E2-bound DNA compared to free DNA. The selected steps were chosen because they were crystallized in the BII state. Average percent BII (free DNA, black; E2-bound DNA; red) from 3 simulations copies with standard deviations in gray.

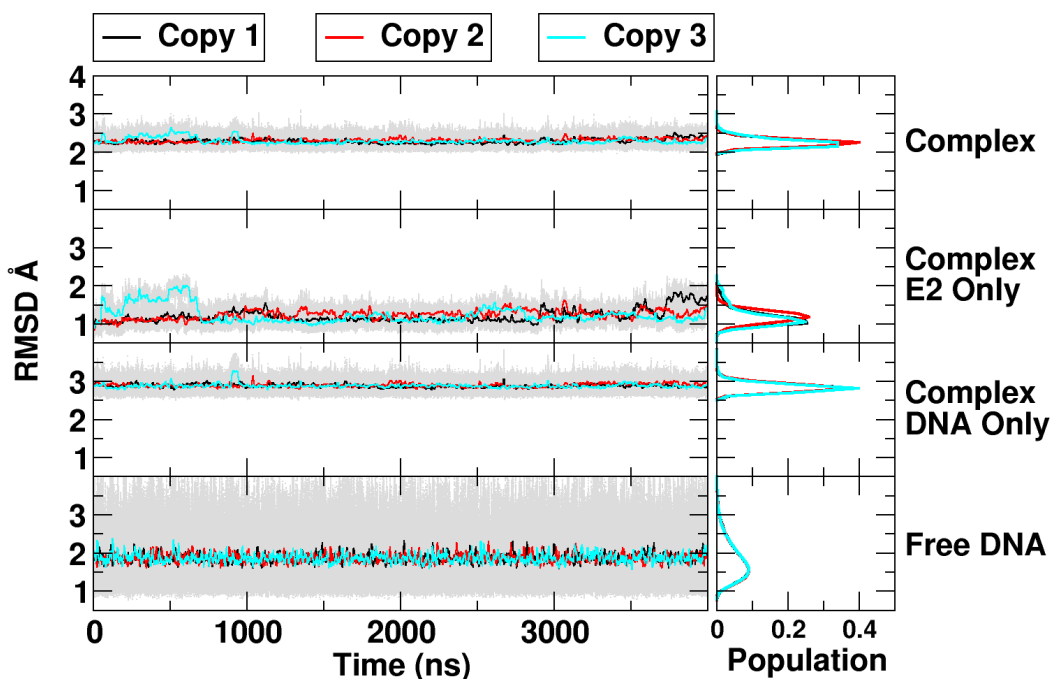


Figure 3.6: Root mean square deviations (RMSD) measurements. For the E2-DNA complex, RMSD was measured using the X-ray crystal structure as a reference. All DNA heavy atoms and E2 protein backbone atoms (N, C, CA, and O) were used to calculate RMSD. The RMSD of E2 (backbone atoms) from the complex was calculated separately, as well as the DNA (all heavy atoms), with the corresponding X-ray component used as a reference. An average structure was used as a reference for free DNA RMSD calculations, and all heavy atoms were used to determine RMSD. The average reference structure was an average of all frames from the three separate free DNA trajectories each run to μs . The RMSD for each system is quite stable over the course of the simulations. There are slight jumps in the RMSD of the complex, however, and by separating out the protein and DNA components it was possible to see how each contributed to overall RMSD. For example, copy 3 of the complex has higher RMSD in the first 750 ns and the E2 RMSD clearly contributes to this, while at about the 3750 ns mark of copy 1 there is a short jump in RMSD and appears to also be from deviations in the protein. These were minor (about 1 Å or less) jumps in RMSD, and as previously stated in the main body of the text, the E2-DNA complex did not dissociate during the simulations. However, it does support the idea that the added complexity of E2 contributes to greater conformational diversity and longer convergence times of properties in the DNA of the E2-DNA complex. RMSD over time for E2-DNA complex (top), the E2 component of the E2-DNA complex (second from top), the DNA component of the E2-DNA complex (third from top), and the free DNA (bottom). The X-ray structure was used as a reference for the RMSDs calculated on the complex. RMSD was calculated for protein backbone atoms (C, CA, N, and O) and all heavy atoms of DNA. An average structure from all copies (4 μs each) was used as a reference for free DNA. Population histograms are to the right of each RMSD time series.

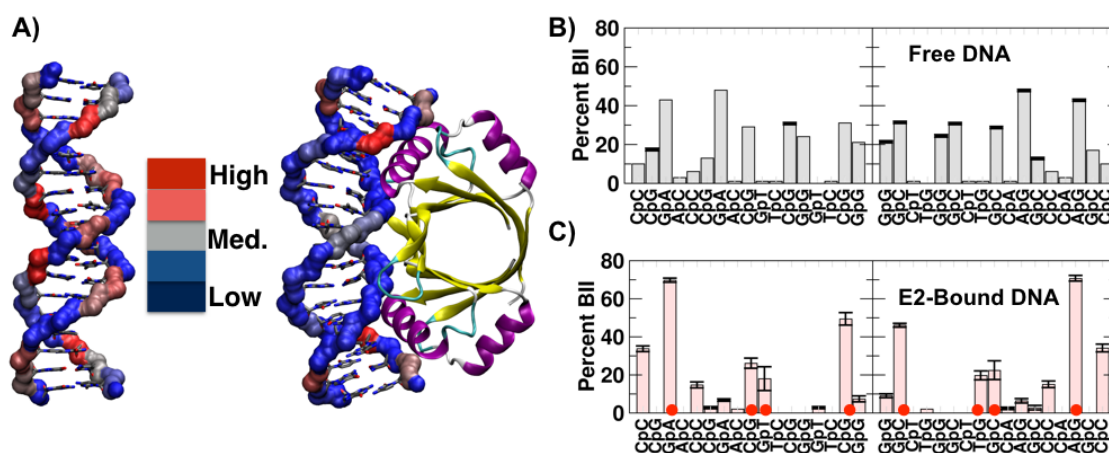


Figure 3.7: Percent BII distributions in free DNA and E2-bound DNA. (A) Free DNA is on the left, the E2-DNA complex on the right. They share the same orientation so that the major grooves facing E2 in the free DNA are facing the same direction as those bound to E2 in the complex. Percent DNA is mapped to the DNA backbone and colored blue for low BII population, gray for medium BII population, and red for high BII population. The protein is colored by secondary structure: yellow for beta sheets, purple for helices, cyan for turns, and white for loops. The structures are averages of three simulation copies, considering structures every 1 ns, and fit to the DNA backbone heavy atoms. (B) The percent BII at each dinucleotide step is shown for free DNA and (C) E2-bound DNA. The red dots in panel C indicate the location of steps crystallized in the BII conformation. The steps are labeled on the X-axis in the 5' to 3' direction of the left of the vertical line and labeled 3' to 5' on the right of the vertical line for both B and C. Percent BII was calculated at each step for each of three simulation copies. The average value for the three copies with standard deviation is shown for each step.

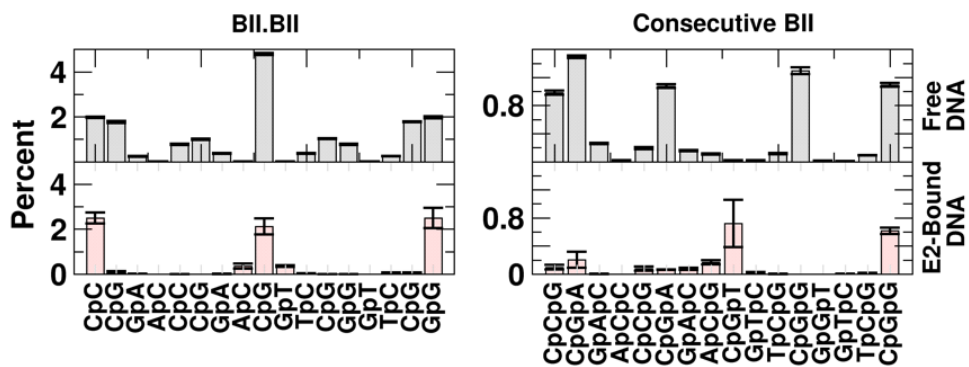


Figure 3.8: Percent BII in facing and consecutive steps. Percent of BII occurring in facing dinucleotide steps (BII.BII) is shown on the left for free DNA (top, black) and E2-bound DNA (bottom, red). On the right, the percent of BII occurring in consecutive steps is shown for free DNA (top, black) and E2-bound DNA (bottom, red). All measurements are an average of three copies with standard deviation error bars. Steps are labeled for a single strand, 5'–3', left to right.

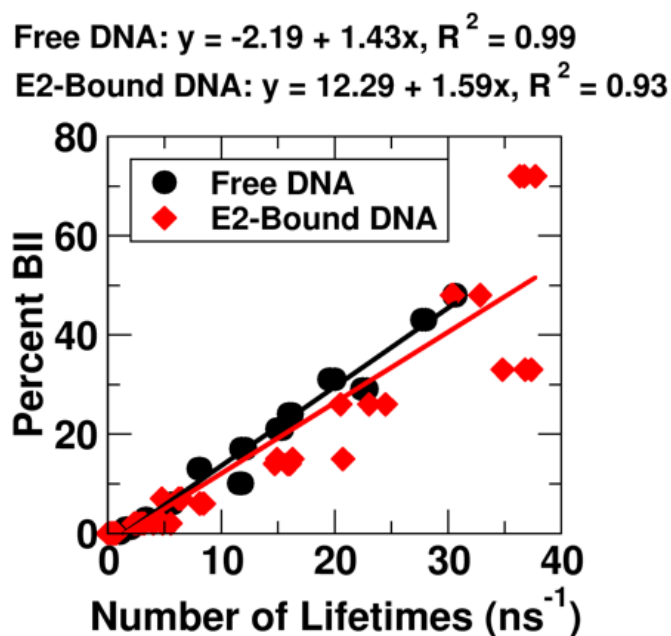


Figure 3.9: Correlation between percent BII and BII frequency. Percent BII over number of lifetimes per ns for free DNA (black) and DNA from E2-DNA complex (red). There are three copies of data for each system from each simulation copy. The outliers in the complex system are from steps 1 and 3. Step 3 has transitions to BII less than expected while step 1 has higher than expected BII frequency.

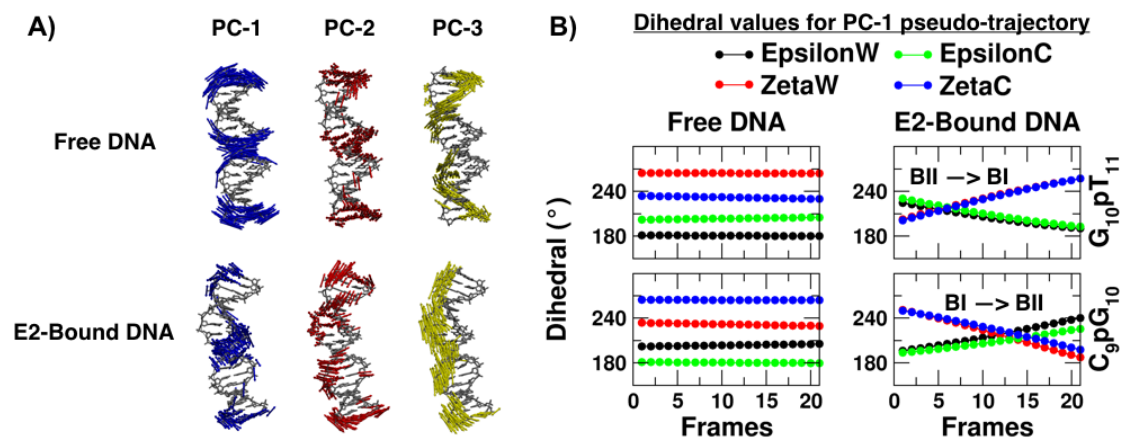


Figure 3.10: Principal component projections calculated from DNA heavy atoms. (A) Porcupine plots mapped onto free DNA (top) and E2-bound DNA (bottom) show motions of the first three primary modes. Arrows on the porcupine plots are colored blue (PC-1), red (PC-2), and yellow (PC-3), and are rendered to show motion in both directions. Nucleic acid heavy atoms are colored silver in all representations. The E2-DNA complex is bent due to the interaction with the E2 protein (not shown), but the protein interface is on the right side of the DNA, the same direction the DNA is bending. All representations of the DNA are aligned and are oriented the same for each PC. (B) Epsilon and zeta dihedral measurements for pseudotrajectories for PC-1 of free DNA and E2-bound DNA show BI/BII transitions for E2-bound DNA at steps 9 and 10.

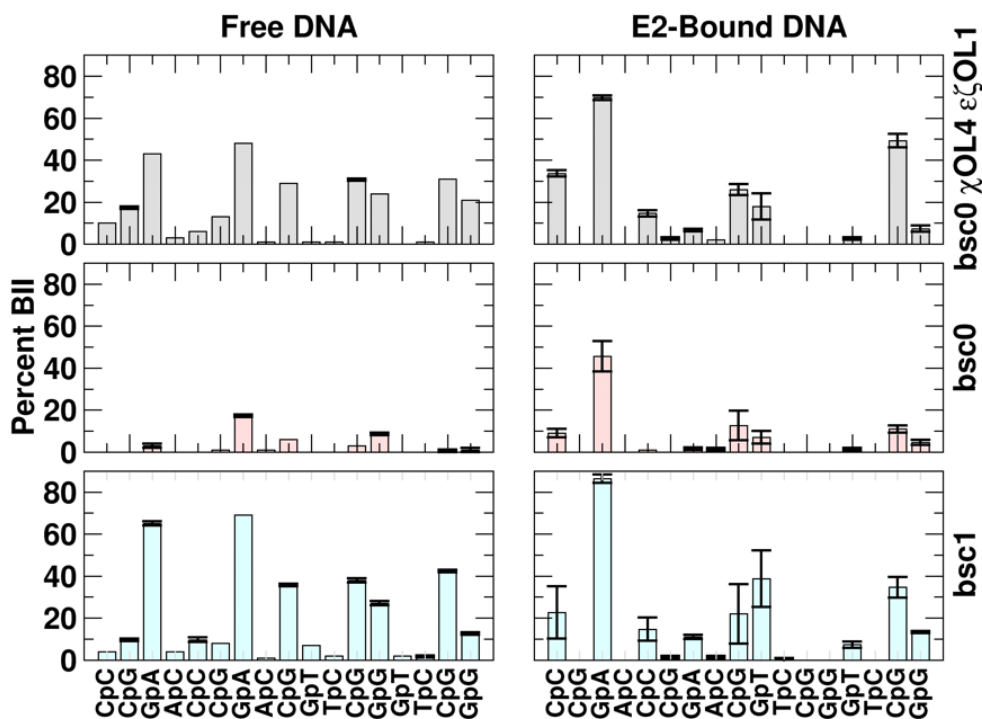


Figure 3.11: Force fields and BI/BII states. A full force field comparison was not discussed in the chapter but percent BII was calculated for the systems simulated with bsc0 [20] (no chi, epsilon, or zeta corrections) and bsc1 [21] force fields. It was interesting to see the BII landscape, or relative BII populations per dinucleotide, was fairly consistent across force fields, but that bsc0 clearly showed underpopulated BII relative to bsc1 and bsc0 χ OL4 ϵ ζ OL1. Percent BII for three different force fields are shown above. Free DNA (left) and E2-bound DNA (right) are shown for bsc0 χ OL4 ϵ ζ OL1 (the force field used in this chapter), bsc0, and bsc1. A single strand is shown for each but the symmetry shown in the main text is consistent for all force fields. The pattern of BII population along the backbone shares a similar pattern across the force fields.

CHAPTER 4

THE ROLE OF BI/BII TRANSITIONS IN E2-DNA RECOGNITION

4.1 Abstract

Molecular dynamics simulations of E2-bound DNA recently demonstrated dinucleotide steps with relatively high BII populations corresponded to steps in the X-ray structure crystallized in the BII conformation. Additionally, BII DNA backbone subtype populations in E2-bound DNA have been shown to converge on the low microsecond time scale. The biological implications of BI/BII substate dynamics have not been clearly elucidated for E2-DNA complexes, but transitions between BI and BII subtypes in BPV E2-bound DNA have been observed as dominant modes of collective motion from principal component analysis. We are interested in determining whether these motions were DNA sequence-dependent or completely induced by E2, and whether this behavior was conserved across multiple E2 types. Molecular dynamics simulations of BPV, HPV 6, and HPV 18 E2-DNA with duplex DNA of each of two different spacer sequences, AATT and ACGT, were run for 3 μ s, free DNA with each spacer sequence was also simulated. BI/BII substate populations and dynamics were found to depend on both the E2 binding partner and the DNA spacer sequence. E2 proteins from the human papillomaviruses preferentially bind DNA with the AATT spacer and our results suggest that DNA backbone substate dynamics may be implicated in differential binding of E2 partners.

4.2 Introduction

Sequence specific recognition of DNA by proteins has numerous biological implications and is important for functional regulation. Structural data of protein-DNA complexes has led to understanding these important interactions on many levels but incompletely [85,86]. The E2 proteins from papillomaviruses present an interesting challenge to determine the mechanism of sequence specific binding to DNA. The E2 proteins bind to DNA as homodimers by inserting α 1 helices into consecutive major grooves of two DNA consensus

half-sites [42, 44, 99–101]. The full recognition sequence is ACCGN₄CGGT, where N is any four bases in what is called the spacer region between the two half-sites. The E2 α 1 helices make direct contacts through conserved residues with the DNA bases in the major groove and the phosphate backbone. The spacer region is not directly contacted by E2 proteins in X-ray crystal structures of BPV E2-DNA [43], HPV 6 E2-DNA [45], and HPV 18 E2-DNA [46] complexes, yet DNA binding studies have shown that spacer sequence does matter, and that the HPV E2 proteins show preference for binding spacer DNA with the AATT sequence over ACGT spacers [46–48, 107, 108]. The BPV E2 was found to demonstrate no preference for binding spacer sequence. Because the spacer DNA does not have direct contacts with E2, the indirect readout mechanism has been implicated in the differential binding of E2 proteins to DNA. Indirect readout has been shown to be important for TNF κ B [39], when replacing DNA bases proximal to the consensus site with altered BI/BII backbone substate dynamics altered binding. The DNA backbone populates two substates, BI and BII, determined by ϵ and ζ dihedrals; when ϵ and ζ are in t/g-configuration DNA is BI, when ϵ and ζ are in g-/t configuration DNA is BII [34, 35, 38, 90]. The backbone substates populations are sequence-dependent in free DNA, and shift upon protein binding [109]. The X-ray crystal structure of BPV E2-DNA [43] was solved with two of the steps in the spacer sequence (ACGT) in the BII state, the HPV 6 and HPV 18 E2-DNA [45, 46] complexes with AATT spacer lacked the BII state in the spacer. During a recent molecular dynamics investigation of BII convergence in E2-bound DNA, transitions of BI/BII substates in the spacer region emerged as a dominant mode of motion [109].

The observation of BI/BII transitions ranked as the highest collective motion in the BPV E2-DNA complex led to a number of questions that we address in this work. Are the BI/BII substates dynamics completely induced by the E2 protein, or are the intrinsic structural properties of DNA sequence preserved in the E2-bound state? To study this we designed DNA oligomers in silico with the spacer region mutated to AATT, and compared the results to our simulations with ACGT spacer. There are no X-ray structures of BPV E2-DNA bound to DNA with the AATT spacer, but DNA binding studies have shown BPV E2 binds to AATT spacer with similar affinity as the ACGT spacer. Free DNA of both spacer sequences was also run so that we could investigate the influence of E2 binding on BI/BII substate populations. BI and BII states are dependent on sequence and the influence of the tetranucleotide environment [28, 36–38], so BII populations in the two sequences are expected to diverge in free DNA. But would E2-bound DNA, independent of spacer sequence, have the same backbone substate populations? From X-ray structures of

HPV 6 E2-DNA and HPV 18 E2-DNA, which are crystal structures with AATT spacers, and do not have any BII substates in the spacer region, the expectation would be that sequence-dependent BII persists in the E2-bound state and E2-bound A-tract DNA would lack significant BII populations. However, the HPV E2 proteins share $\sim 50\%$ identity with BPV E2, and it has not been determined how the different protein compositions influence DNA backbone substates. Would HPV E2-bound DNA with ACGT spacer have relatively high BII populations in the spacer region? To address this we mutated AATT spacer DNA in the two HPV E2-DNA complexes to ACGT. In total, molecular dynamics simulations were performed on BPV E2-DNA, HPV 6 E2-DNA, and HPV 18 E2-DNA, each with AATT and ACGT spacer sequences. Additionally, free DNA with the two spacer sequences were run, and free BPV E2 was also simulated. Simulations were run to achieve $3 \mu\text{s}$ sampling, based on E2-bound DNA requiring at least $1 \mu\text{s}$ to achieve BII population convergence.

The E2-bound DNA with ACGT and AATT spacers showed markedly different BII substate populations. DNA bound to the three E2 types, BPV, HPV 6, and HPV 18, showed similar BII substate populations in the consensus half-sites and spacer regions. These BII populations differed from free DNA at many steps, the GpA, CpG, and GpG steps in the consensus half-sites all had reduced BII populations when E2-bound. The CpG in the spacer region was similar in both free and E2-bound DNA, but the neighboring GpT step increased BII state when E2-bound. DNA with the AATT spacer underwent similar changes to ACGT when free and E2-bound DNA were compared. The BII states in consensus half-sites shifted similar to ACGT DNA. However, changes in the spacer region were different because AATT free DNA had modest BII population at the ApA step but BII was essentially absent in other spacer steps. In AATT E2-bound DNA, the ApA BII population was decreased and the BII populations at the other steps in the spacer region remained low. Principal component analysis showed that E2-bound DNA with ACGT spacer sequences had collective motions with BI/BII transitions that ranked highly over 20 calculated modes, this was true for all E2 types simulated with ACGT spacers. On the other hand, E2-bound DNA with AATT spacers almost completely lacked the BI/BII transitions, only a single low-ranking mode was observed in one of the E2 types and absent in the other two. DNA bending and minor groove widths were also investigated but showed similarities between AATT and ACGT E2-bound DNA. The contrasts between AATT and ACGT E2-bound DNA backbone substates opens the question of whether these differences contribute to preferential binding of AATT spacer over ACGT spacer DNA for, at least, HPV E2 proteins.

4.3 Methods

4.3.1 Preparing Systems for Simulations

The coordinates for the BPV E2-DNA complex were taken from PDB entry 2BOP, an X-ray crystal structure of the bovine papillomavirus E2 DNA-binding domain bound to a 17-mer duplex DNA at 1.7 Å resolution [43]. The setup for this system was previously described in Reference [109]. The HPV 18 E2-DNA model was built from the 2.4 Å X-ray crystal structure, PDB ID 1JJ4 [46]. In 1JJ4, residues G324, A325, G326, and N327 were missing from both monomers of E2, and residue E328 was missing from chain A. The missing residues were built in with LEaP from AMBER 14 [67]. PDB entry 2AYG, an X-ray structure at 3.1 Å was used to build the HPV 6 E2-DNA complex [45]. BPV E2-DNA was cocrystallized with DNA that contained the ACGT spacer sequence, while both 1JJ4 and 2AYG were crystallized with the AATT spacer sequence. To swap the spacer sequences and be able to simulate each system with both ACGT and AATT the PDB files were edited and the appropriate residues in the spacer region were modified to the desired sequence. The starting coordinates for free DNA and free E2 were stripped from the BPV E2-DNA system. An overview of systems prepared for molecular dynamics is presented in Table 4.1, and Figure 4.1 shows the starting structures for the three E2-DNA complexes and the accompanying DNA sequences.

All systems were built with LEaP, the ff12SB [110] parameter set was used for the E2 proteins, and the bsc0χOL4εζOL1 [20,22,23] parameter set was used for DNA. The systems were solvated with TIP3P [54] water molecules, and Na⁺ ions with Joung/Cheatham [60] parameters were added to neutralize the systems. The simulation cell was built as a truncated octahedron. The HMassRepartition function in AmberTools ParmEd was used to shift the mass of all solute hydrogen atoms to 3.024 Daltons and the mass of all heavy atoms attached to hydrogen atoms was adjusted so the overall mass of each pair was not affected. The simulation dynamics were then integrated with 4 fs time step [13].

PMEMD from AMBER 14 was used to minimize each system, and all were minimized following the same protocol. Ion positions were randomly swapped with water molecules; ions were placed at least 4.0 Å from each other, and at least 6 Å from solute atoms. Randomization of ions was repeated for each of three copies per system. Structures were then minimized with 1000 steps of steepest descent followed by 1000 steps of conjugate gradient with 25 kcal/mol-Å restraints on all solute atoms. The systems were then heated from 100 to 300 K over 200 ps, with a weak coupling thermostat and 1.0 ps coupling time [104] at 1 atm constant pressure. SHAKE was used to constrain bonds involving

hydrogen [105], a nonbonded cutoff of 9 Å was used, and PME was employed for long-range electrostatics with default parameters. Restraints were decreased over a series of minimization steps from 5 kcal/mol-Å² to 1.0 kcal/mol-Å², these minimizations were run with 500 steps of steepest descent followed by 500 steps of conjugate gradient. These minimization steps were interspersed with 100 ps equilibration steps, following the same scheme as the minimizations of reducing restraints on solute atoms. Terminal Watson-Crick base pairing restraints were applied to the DNA, as described in [109].

All production runs were on GPUs using PMEMD.cuda from AMBER 14 and AMBER 15 [12,67]. Production simulations were run at constant pressure using the Monte Carlo barostat and 1.0 ps pressure relaxation time. Constant temperature was maintained with Langevin dynamics [63] using collision frequency of 2 ps⁻¹; random starting velocities were used for all simulations. The nonbonded cutoff was set to 8.0 Å and PME with grid spacing of 1 Å was used for calculating electrostatics. Coordinates were saved every 5000 steps, except for BPV, which had coordinates saved every 1000 steps as previously described [109].

4.3.2 Analysis and Visualization

Trajectory analysis was carried with CPPTRAJ [69] from AMBER 14 and AMBER 15. Percent BII was calculated with CPPTRAJ and an example script is shown below. Essentially, epsilon and zeta dihedrals were calculated for all steps and BII was considered present when epsilon - zeta > 0. Principal component analysis (PCA) was performed using CPPTRAJ. Groove widths, inter-, and intra-base pair helical parameters were calculated with the CPPTRAJ implementation of 3dna. Bend was calculated with Curves+ [70]. Molecular graphics were generated with VMD [71] and porcupine plots generated with the NMWiz plugin for VMD [72].

```
# !/bin/bash
# Calculate percent BI/BII
# D. Roe 2016-01-06
#
SOURCEDIR= # set directory where system files are
SYS= # set name of system
N_SEGMENTS=3 # set to 1 for BII over entire traj or break into segments
BEGW=3
ENDW=10
BEGC=15
```

```

ENDC=22
WORKDIR='pwd'
INPUT=cpptraj.in
CPPTRAJ='which cpptraj'
# If DEBUG is set to 1 no runs, only first script will be generated
DEBUG=0

cd $WORKDIR # Return us to the original working directory
MYDIR=$SYS/percent-bii
if [[ ! -e $MYDIR ]] ; then
mkdir -p $MYDIR
fi
cd $MYDIR
# Set up files
TOP=$SOURCEDIR/$SYS/top.prmtop
CRD=$SOURCEDIR/$SYS/traj.nc
# Get total number of frames. Output of '-tl' is 'Frames: X'
TOTAL_FRAMES='$CPPTRAJ -p $TOP -y $CRD -tl | awk 'print $2;''
if [[ -z $TOTAL_FRAMES || $TOTAL_FRAMES -lt 1 ]] ; then
echo "Could not get number of frames." > /dev/stderr
exit 1
fi
echo " Trajectory has $TOTAL_FRAMES frames."
# Divide roughly into N_SEGMENTS segments
((DIV1 = $TOTAL_FRAMES / $N_SEGMENTS))
IDX=1
for ((START=1; START < $TOTAL_FRAMES; START += $DIV1)) ; do
if [[ $IDX -eq $N_SEGMENTS ]] ; then
END="last"
else
((END = $START + $DIV1 - 1))
fi
# Create output directory for segment
OUTDIR="Segment$IDX"

```

```

if [[ ! -e $OUTDIR ]] ; then
mkdir $OUTDIR
fi
# Create cpptraj input for calculating dihedrals
cat > $INPUT <<EOF
parm $SOURCEDIR/$SYS/top.prmtop
trajin $SOURCEDIR/$SYS/traj.nc $START $END
check @N1,N2,P,OP1,OP2,C1',C2',C3',C4',O3',O4' skipbadframes silent
multidihedral EZ epsilon zeta resrange $BEGW-$ENDW,$BEGC-$ENDC \
range360 out $OUTDIR/multidihedral.dat
run
EOF
# Generate cpptraj input for calculating epsilon - zeta for each base,
# as well as BI/BII (via 'datafilter') and percent BI/BII (via the
# avg() function).
for i in $(eval echo "$BEGW..$ENDW $BEGC..$ENDC") ; do
cat >> $INPUT <<EOF
EZdiff.$i = EZ[epsilon]:$i - EZ[zeta]:$i
datafilter EZdiff.$i min -360 max 0 name BI.$i
PctBI.$i = avg(BI.$i)
datafilter EZdiff.$i min 0 max 360 name BII.$i
PctBII.$i = avg(BII.$i)
EOF
done
# Generate cpptraj input for calculating overall %BII
cat >> $INPUT <<EOF
datafilter BII.* min 0 max 0 name overallPctBII
PctBIIoverall = avg(overallPctBII)
EOF
# Write out raw BI and BII as well as percent BI and BII. Use the
# 'invert' keyword for the latter so that the values are in a column
# instead of a row.
cat >> $INPUT <<EOF
writedata $OUTDIR/bi.dat BI.*

```

```

writedata $OUTDIR/bii.dat BII.*
writedata $OUTDIR/PctBi.dat invert PctBI.*
writedata $OUTDIR/PctBii.dat invert PctBII.*
writedata $OUTDIR/PctBii.overall.dat invert PctBIIoverall
EOF
# Run cpptraj
if [[ $DEBUG -ne 0 ]] ; then
echo "DEBUG is set, only script for first segment is generated."
else
($CPPTRAJ -i $INPUT 2>&1) | tee $OUTDIR/cpptraj.out
if [[ $? -ne 0 ]] ; then
echo "Cpptraj error." > /dev/stderr
exit 1
fi
fi
((IDX++))
if [[ $DEBUG -ne 0 ]] ; then
break
fi
done # END loop over segments
exit 0

```

4.4 Results and Discussion

4.4.1 Overview of Dynamics and E2-DNA Interactions

Each simulation was run in triplicate, and each copy had simulation times of 3000 ns, for an aggregated total of 9000 ns per system. Root mean square deviations (RMSD) of the trajectories indicated that the simulations were stable over that time scale (Figure 4.2), and relatively low RMSD values near 3 Å for the E2-DNA systems means that the simulations sampled conformational spaces close to the experimental X-ray structures. This was especially important to note for the systems with altered DNA spacer sequences, these systems were well-equilibrated and the base mutations did not disrupt the stability of the complexes. There was one exception, a single copy of HPV 6 E2-DNA with the ACGT spacer had relatively high RMSD due to one of the DNA consensus half-sites dissociating

from the corresponding E2 recognition helix, this event is evident in the RMSD plot as early as 1 μ s and especially at 2 μ s where there is a sharp jump in the RMSD and the DNA half-site loses contact with the E2 helix. Many of the following analyses were performed on aggregated trajectories, including HPV 6 E2-DNA ACGT, so many values reported reflect the inclusion of this trajectory. With the exception of the single HPV 6 ACGT copy just discussed, the RMSD values for systems with ACGT compared to AATT spacer sequences were very close, especially for BPV and HPV 6. Interestingly, the mutated ACGT sequence in HPV 18 E2-DNA showed lower RMSD values compared to the X-ray structure AATT sequence. Average structures from aggregated trajectories of each system also support the quality of the simulations and how closely the dynamics sampled space around the X-ray structures. The backbone RMSD of average structures to X-ray structures are reported in Table 4.2, and are remarkable for being close to 1 Å in all cases. The different DNA sequences in BPV E2-DNA each resulted in 1.2 Å RMSD, while both HPV 18 E2-DNA systems had RMSD values under 1 Å. Even with the inclusion of all three copies of HPV 6 ACGT, the RMSD was still low at 1.4 Å and only slightly higher than 1.1 Å for HPV 6 AATT. The relatively high RMSD of free DNA (Figure 4.2) results from the RMSD reference being E2-bound DNA from the X-ray structure. This DNA was smoothly bent around E2, and the free DNA simulations quickly lost the bend and sampled conformations more characteristic of typical B-form DNA.

Free E2 was also stable in simulations on the 3 μ s time scale (Figure 4.3). The RMSD values were mostly below 3 Å and the backbone RMSD of the average structure to the X-ray was 1 Å, with most of the deviations coming from the disordered regions between β sheets 2 and 3. The atomic fluctuations of free E2, shown in Figure 4.3B, indicated that the highest root mean square fluctuations (RMSF) were in the loops between β sheets 2 and 3. Atomic fluctuations in DNA-bound E2 is shown in Figure 4.4. The sharp peak in one of the β 2- β 3 loops is reduced compared to free E2, and the asymmetry suggests a lack of convergence of those dynamics. Comparison of RMSF across E2 types shows that the fluctuations of DNA-bound E2 are similar. The α 1 helices show an increase in fluctuations in the N- to C-terminal direction, while the α 2 helices have the highest RMSF, excluding termini. Free and E2-bound DNA RMSF are also shown in Figure 4.4 and differences between the spacer sequence, and the E2 type binding the DNA, are difficult to detect. An interesting feature, however, is that the ACGT spacer sequence has slightly elevated RMSF compared to AATT spacer DNA, the only region in each RMSF plot where differences can be observed between the two DNA sequences. This suggests increased atomic fluctuations for ACGT relative to

AATT.

Principal component analysis on free E2 heavy atoms provided insight into the collective motions sampled during the simulations (Figure 4.3D). The top mode shows the loops between $\beta 2$ and $\beta 3$ moving toward each other, then extending away from one another. At the same time the $\alpha 1$ helices move toward and away from each other. The helices separate and the loops come together, the loops separate and the helices move toward one another. The motion of the loops are indicated by arrows on the porcupine plots of Figure 4.3D. Principal components 2 and 3 are similar, but show motions primarily on a single monomer, while PC1 is both monomers dynamically coupled. It is important to note that the loop regions that are so dynamic in the principal component analysis associate with the DNA minor groove, and that the average structures of free BPV E2 and E2-DNA have similar conformations in these loops compared to the X-ray structure, as shown in Figure 4.5. However, the range of the loop differs from loops that travel relatively short distances, from 27.1 Å to 34.5 Å in the E2-DNA complex, to loops that cover distances from 19.8 Å to 35.7 Å apart in free E2 (Table 4.3). While the average structures are similar, and the extended distances between loops similar, the free E2 loops travel closer to each other, a conformation not allowed by the presence of DNA. This suggests a conformational capture mechanism as E2 is more likely to receive DNA with the loops extended, rather than rearrange the loops upon binding. The structural rearrangement of the $\alpha 1$ helices is also of interest because some of these residues are highly conserved, required for specific DNA recognition, and directly contact bases in the consensus half-sites. Hooley et al. [45] reported displacement of alpha helices on binding DNA for X-ray structures and we were interested in comparing results from simulations to experimental values. C α atoms from one of the recognition helices were aligned (RMSD 2.0 Å) and the RMSD between the other two recognition helices was measured to be 5.3 Å, the alignment based on one of the recognition helices is shown in Figure 4.6. The results demonstrate close agreement with HPV 16 NMR studies [47].

4.4.2 DNA Bending and Minor Groove Widths

The smoothly bent DNA around E2 is one of the most immediately noticeable and interesting features of the E2-DNA complexes. It has been suggested that the HPV E2 proteins preferentially bind AATT spacer sequences because these A-tracts tend to have increased bend compared to other sequences [111]. The simulations show that free DNA with the AATT spacer has higher bending (calculated over the internal 12 base pairs) compared to DNA with the ACGT spacer, 15.4° compared to 13.5° (see Table 4.4). DNA

with the AATT spacer sequence also has a higher range of bending motions, exploring bending up to 63.4° compared to 54.6° for DNA with the ACGT spacer. Additionally, the different E2 proteins influence the DNA bend, ranging from 39.9° in HPV 6 E2-bound DNA to 45.5° in HPV 18 E2-bound DNA. The low value of 32.4° for HPV 6 ACGT E2-bound DNA was lowest, but this results from one of the three copies of this system losing the connection between one of the half-sites and its corresponding recognition helix, the DNA in this system lost the characteristic bend of E2-bound DNA, which led to the low average bend for HPV 6 ACGT DNA (when the bending analysis was run on the two copies of HPV 6 ACGT, the average value was $36.9 \pm 5.7^\circ$). The type of E2 protein determines the amount of bending, but the difference in spacer sequence seems to have minimal effect on E2-induced bending. BPV E2-DNA bends DNA with the ACGT spacer 43.9° and DNA with AATT spacer 43.4° , the range of bending in BPV E2-bound DNA is also similar. Likewise, HPV 18 E2-bound DNA is bent 45.5° for ACGT spacer DNA, and 45.3° for AATT spacer DNA. The bending, therefore, is determined by the E2 binding partner, and is not a sequence-dependent property in E2-bound DNA. Free DNA, on the other hand, bends in a sequence-dependent manner; DNA with the AATT spacer shows increased bend compared to DNA with the ACGT spacer. This lends support to the idea that HPV E2 preferentially binds DNA with the AATT spacer because of the intrinsic property of increased bending in this sequence, however, the differences between average bending of AATT and ACGT are only $\sim 2^\circ$, and regardless of spacer sequence the DNA undergoes considerable deformation from the average free structure to the average E2-bound conformation. Also, both free DNA sequences have maximum bends that are beyond the average bends for the E2-bound structures.

The minor groove of E2-bound DNA between the two consensus half-sites does not directly contact E2, unlike the two consensus half-sites that make direct contacts with E2 from the major groove to the recognition helices of DNA. The minor groove has, however, been implicated in differential binding of E2 proteins to AATT and ACGT spacer DNA [45, 46, 112]. The increased deformability of A-tract DNA and compression of the minor groove width in AATT spacer has been proposed as an explanation for preferential binding of DNA with AATT spacers by E2 proteins. In the starting structures we used, we found that the ACGT spacer in BPV E2-bound DNA had the most narrow minor groove width averaged over the spacer, 12.0 \AA compared to 12.6 \AA and 12.3 \AA for the minor groove widths of HPV 6 and HPV 18 AATT spacer DNA, respectively (Table 4.5 and Figure 4.7). The minor groove widths from simulations of E2-bound DNA with AATT or ACGT spacers

were within 0.3 Å in all systems, and the ACGT spacer more narrow in each case. The free DNA minor groove values were 12.3 Å for ACGT spacer, and 12.4 Å for AATT spacer, both slightly decreased compared to E2-bound DNA. The minor groove width, therefore, appears to slightly expand, rather than contract, in E2-bound DNA. Solution NMR structures for these DNA sequences are not available, so the comparison to free DNA cannot be made, but recent improvements in force fields have shown close agreement with NMR minor groove widths. The compact minor groove in BPV E2-DNA X-ray structure may be influenced by the crystal packing environment. The per base pair analysis of groove widths shows close agreement across the types of E2 proteins. Compared to free DNA, the major groove compacts around the consensus half-sites due to direct contact with the $\alpha 1$ helices of E2. From the simulation data presented, it does not appear that E2 induces different minor groove widths based on spacer sequence, or that free DNA minor groove widths vary by spacer sequence.

4.4.3 BI/BII Backbone Substate Dynamics

Convergence of BII substates in BPV E2-DNA was recently demonstrated to be on the low microsecond time scale, and faster for free DNA [109]. In this work, the fraction BII was calculated at increasing time intervals from 1 ns to 3000 ns, to determine convergence in E2-bound and free DNA for DNA with AATT and ACGT spacers (Figure 4.8). By 3000 ns the BII fraction in all cases has completely converged or, in the case of BPV E2-bound DNA HPV 6 E2-bound DNA with ACGT spacers at the middle CpG step, nearly converged. This is in agreement with previous results on BPV E2-DNA with decoy and native BII populations, the CpG and GpT steps in the spacer region were slowest to converge. Here, we demonstrate that HPV 18 E2-bound DNA differs from the other E2-bound DNA with ACGT spacers by converging at 1 μ s. The E2-bound DNA with AATT spacers also converges quickly, and the low populations in the spacer region are apparent, this is subsequently discussed. The free DNA, as expected, converged on the time scale of ~ 100 ns. The protein clearly affects the time scale of BII convergence and the fraction BII at each base step.

The fraction BII from molecular dynamics in the CpG and GpT steps of the spacer region was recently shown to have the highest BII populations relative to other steps in the internal 12 base pairs of BPV E2-bound DNA, corresponding to steps that were crystallized in the BII state [109]. This is the best agreement we have so far that simulations of protein-bound DNA can reproduce BI/BII equilibrium from experiment. There is more evidence to support that simulations of free DNA, with the appropriate force fields, have

improved the agreement with predicted and experimentally determined BII fractions [21,24]. Simulations of free DNA have also shown strong sequence dependence for BII populations. Based on the X-ray crystal structures of HPV 6 and HPV 18 E2-DNA, it was not expected that the AATT spacer have high BII populations, but the X-ray structure is a snapshot of averaged motions, dependent on experimental conditions and crystal packing artifacts, so we were interested to see whether the MD simulations agreed more with the X-ray structures or with previous MD simulations of the ACGT spacer. We also wanted to explore whether E2 proteins induced common BI/BII backbone dynamics independent of spacer sequence, or whether these states were sequence-dependent in E2-bound DNA. Furthermore, we wanted to know whether these results differed depending on the type of E2 protein bound to DNA.

The BI/BII backbone dynamics in free DNA are clearly sequence-dependent, and base steps are influenced by, at least, the neighboring tetranucleotides. The fraction BII was calculated for each of the internal 12 DNA base pairs for free and E2-bound DNA, see Figure 4.9. For the two sequences of free DNA, one with the ACGT spacer, the other with AATT spacer, the fraction BII is in close agreement in the recognition half-sites, especially steps ApC, CpC, and CpG on the 5' end of the spacer region. The GpA step is elevated by $\sim 15\%$ in DNA with the ACGT spacer. Due to the proximity of this step to the spacer region, and the different DNA sequences, the neighboring bases of the spacer appear to influence the population at this step. Beyond the spacer region, all four steps are in close agreement, though it is interesting to note that the CpG and GpG steps appear to have flipped populations from one sequence to the next. The influence of neighboring bases in the spacer region could be playing a role, though this is two positions away from the spacer. Although populations differ slightly, the overall pattern of BII in the recognition half-sites matches between the two sequences. Within the spacer the sequence-dependent effects are clear. The ApC step has little to no BII population while the corresponding ApA is elevated to $\sim 20\%$. The next two steps in the AATT spacer have BII fractions approaching zero, but in the ACGT spacer the CpG step has fraction BII near 0.40. The GpT step is similar to TpT, with little to no sampling of BII states.

In DNA with the ACGT spacer, a change in BI/BII backbone substates were observed between free and E2-bound DNA. The patterns for each type of E2 are similar: the GpA step upstream of the spacer region is shifted down to 30% or less, the CpG step in the spacer region remains near 40-50%, the neighboring GpT step jumps dramatically from zero to about 40% (only $\sim 20\%$ for HPV 18), and the CpG and GpG steps in the second consensus half-site drop from $\sim 40\%$ to less than 10%. The most remarkable step here is GpT

in the spacer region that jumps from spending almost no time in the BII state to spending nearly half the simulation occupying the BII state. This also places two steps with $\sim 50\%$ BII population next to each other, though it was previously shown that the steps rarely spent time in the BII substate concomitantly, rather the steps were continuously moving in alternating fashion between BI and BII states [109]. The pattern of BII populations also shifts dramatically from free DNA to E2-bound DNA in the systems with AATT spacers. In this case, all BII fractions are knocked down below 0.10 except the CpC step in the first recognition half-site, which goes from $\sim 10\%$ population in free DNA to between 15-30% for E2-bound DNA, depending on the E2 type. In DNA, whether it had ACGT or AATT spacer, a clear shift occurred in BII fraction from free to E2-bound DNA. However, the DNA backbone substate dynamics are not completely E2-induced. The DNA sequence bound to E2 also determines substate dynamics, demonstrated here with two DNA spacer sequences and multiple E2 protein types.

To explore the BI/BII transitions as the dominant modes of motion in BPV E2-DNA, we calculated the top 20 modes for free and E2-bound DNA to determine how common the modes were in each system, to understand whether the modes were ranked higher in E2-bound compared to free DNA, to understand the DNA sequence dependence of the modes, and finally, to determine whether these modes were observed in all E2 types. Principal component analysis was calculated on all heavy atoms, pseudotrajectories were generated for each mode, and ϵ and ζ dihedrals calculated for each pseudotrajectory. BI and BII states were calculated for the first and last frame of each pseudotrajectory, and a transition was noted if there was a BI/BII switch between the first and final frames of the pseudotrajectory. Only the ACGT and AATT spacer regions were considered, because we were interested in the different dynamics between the two sequences, and the results are presented in Table 4.6. Comparing E2-bound DNA, the ACGT spacer BI/BII transitions are ranked highly; a transition was observed in the top mode for BPV E2-bound DNA, the sixth mode of HPV 6 E2-bound DNA had a transition, and the second mode of HPV 18 E2-bound DNA. In contrast, there were no BI/BII transitions for E2-bound DNA with the AATT spacer for BPV or HPV 6, while the 19th mode of HPV 18 had a BI/BII transition. Out of the 20 calculated modes, E2-bound DNA with the ACGT spacer there were BI/BII transitions in 100%, 65%, and 80% of the modes for BPV, HPV 6, and HPV 18, respectively. The only E2 type that had similar backbone dynamics was HPV 18, and only a single mode had a BI/BII transition. For free DNA, a BI/BII transition was observed in the 5th mode for the ACGT spacer and in the 14th mode for the AATT spacer. Out of the top 20 modes, 30%

of modes with the ACGT spacer had BI/BII transitions in free DNA, compared to only 5% for DNA with the A-tract spacer. This was not surprising based on the fraction BII analysis described earlier. It was interesting, however, to visualize the pseudotrajectories and observe similarity in overall motions between the different DNA spacer sequences. A representative example is presented in Figure 4.10. The top modes were visualized for BPV E2-DNA with each spacer sequence. The nucleic acids in the figure are colored according to the amount of motion at each position and the DNA appear nearly identical. However, when the ϵ and ζ dihedrals are calculated along the trajectory, clear differences are observed. The CpG and GpT steps of the ACGT spacer are both shifting BI/BII states; CpG starts in the BII state and moves to BI, while GpT starts in BI and moves to BII. Meanwhile, no backbone substate transitions were observed for the AATT spacer.

4.5 Conclusion

Molecular dynamics simulations of free E2, free DNA, and E2-bound DNA with both AATT and ACGT spacer sequences were run in triplicate with 3 μ s of sampling per copy to determine BI/BII substate dynamics in free and E2-bound DNA, and establish the influence of E2 binding partner on the BII populations. Average structures from simulations were in close agreement with X-ray reference structures, backbone RMSD was ~ 1 Å in all cases.

The overall DNA bend differed between the internal 12 base pairs when free DNA with AATT and ACGT spacers were compared. Sequence dependent bending was expected for free DNA and we observed increased bend for DNA with the AATT spacer relative to ACGT spacer DNA. The type of E2 protein bound to DNA was found to determine the bending in E2-bound DNA. Interestingly, differences in bending were not observed for the AATT and ACGT spacers. Both sequences had bends of 43.9° and 43.4° for BPV E2-DNA, 36.9° and 39.9° for HPV 6 E2-DNA, and 45.5° and 45.3° for HPV 18 E2-DNA. This indicated that the E2 binding partner determined DNA bending rather than sequence differences in the spacer. Minor groove widths in the spacer region were within 0.3 Å for all E2-bound DNA, and slightly expanded compared to free DNA. Again, the E2 binding partner influenced minor groove widths more than differences in spacer sequences.

Sequence dependent BII populations were observed in free DNA, as expected, and shifts in BII populations were observed between free and E2-bound DNA for both AATT and ACGT DNA. The resulting populations in each E2 system were similar, but differed between the spacer sequences. High BII populations were observed in the CpG and GpT steps in the spacer region of E2-bound ACGT DNA, while low BII populations were observed in

all spacer steps of AATT DNA. These results indicated that BII populations in E2-bound DNA are not determined by E2 alone, the populations depend on a combination of E2 binding partner and DNA spacer sequence. The differential backbone substate dynamics in the spacer region between AATT and ACGT spacers suggest backbone dynamics could play a role in the binding preference for E2 proteins.

Table 4.1: Descriptions of each system simulated. All molecular dynamics were run in net-neutralizing ion conditions with Joung-Cheatham ions [60] and fully solvated with TIP3P water model [54]. Three copies of each were run. Spacer sequences with an asterisk are the sequences that were co-crystallized with the E2.

System	PDB ID	X-Ray Resolution (\AA)	DNA Spacer Sequence	Simulation Time per Copy (ns)
BPV 1 E2-DNA	2BOP	1.7	ACGT*	3000
			AATT	
HPV 6 E2-DNA	2AYG	3.1	ACGT	3000
			AATT*	
HPV 18 E2-DNA	1JJ4	2.4	ACGT	3000
			AATT*	
Free DNA	2BOP	1.7	ACGT*	3000
			AATT	
Free E2	2BOP	1.7	N/A	3000

Table 4.2: RMSD of average structures from simulations with X-ray structures as reference (\AA). The average structure was calculated from an aggregated trajectory of three independent copies of each system, totaling 9000 ns of sampling per average structure calculation. RMSD was calculated over nucleic acid and protein backbone atoms.

System	ACGT Spacer	AATT Spacer
BPV 1 E2-DNA	1.2	1.2
HPV 6 E2-DNA	1.4	1.1
HPV 18 E2-DNA	0.9	0.8

Table 4.3: Center of mass distance between the disordered regions between $\beta 2$ and $\beta 3$ sheets (\AA).

System	Average	Std. Dev.	Minimum	Maximum
BPV 1 E2-DNA	29.7	0.6	27.1	34.5
Free E2	28.3	2.5	19.8	35.7

Table 4.4: Total helical bending in free and E2-bound DNA. Bending was calculated for the internal 12 base pairs of each system using Curves+ [70].

System	DNA Spacer	Average ($^{\circ}$)	Std. Dev.	Min	Max
BPV 1 E2-DNA	ACGT	43.9	5.7	15.2	74.5
	AATT	43.4	5.5	14.5	73.5
HPV 6 E2-DNA	ACGT	32.4	10.1	0.0	63.9
	AATT	39.9	5.3	11.7	63.6
HPV 18 E2-DNA	ACGT	45.5	5.7	18.5	78.3
	AATT	45.3	5.4	17.5	73.6
Free DNA	ACGT	13.5	7.2	0.0	54.6
	AATT	15.4	7.8	0.0	63.4

Table 4.5: Average minor groove widths for the spacer region of free and E2-bound DNA.

System	DNA Spacer	Average (Å)	Std. Dev.	X-ray Average (Å)
BPV 1 E2-DNA	ACGT	12.5	0.1	12.0
	AATT	12.7	0.0	-
HPV 6 E2-DNA	ACGT	12.5	0.0	-
	AATT	12.8	0.0	12.6
HPV 18 E2-DNA	ACGT	12.5	0.1	-
	AATT	12.6	0.0	12.3
Free DNA	ACGT	12.3	0.0	-
	AATT	12.4	0.1	-

Table 4.6: BI/BII transitions in the spacer region of free and E2-bound DNA from principal component analysis pseudotrajectories. Principal component analysis was calculated on all heavy atoms of DNA for 20 modes. n/d means that BI/BII transitions were not observed in any mode.

System	DNA Spacer	Highest PC with BI/BII Transition in Spacer	Fraction of PCs with BI/BII Transition in Spacer (N=20)
BPV 1 E2-DNA	ACGT	1	1.00
	AATT	n/d	0.00
HPV 6 E2-DNA	ACGT	6	0.65
	AATT	n/d	0.00
HPV 18 E2-DNA	ACGT	2	0.80
	AATT	19	0.05
Free DNA	ACGT	5	0.30
	AATT	14	0.05

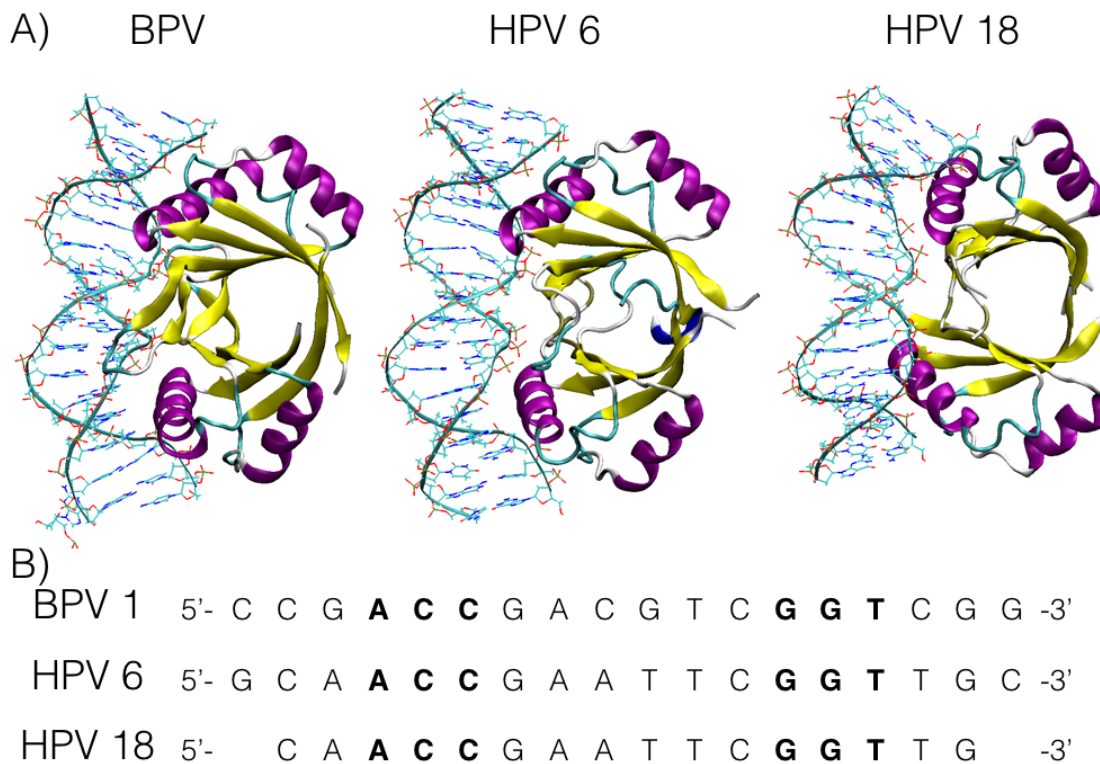


Figure 4.1: X-ray structures of E2-DNA complexes and DNA sequences. A) X-ray structures show the conserved E2 protein fold and DNA deformation in the E2-DNA complexes. The E2 proteins are homodimers with β barrel cores and α 1 helices that insert into consecutive DNA major grooves. DNA is bent toward E2 and the minor groove facing the protein is slightly compressed. Protein is colored yellow for β barrels, α helices are purple, and loops are cyan. The PDB IDs of the BPV 1, HPV 6, and HPV 18 structures are 2BOP, 2AYG, and 1JJ4, respectively. B) The DNA sequences from each crystal structure is shown. Consensus half-sites are bold, and in between the half-sites lies the spacer region.

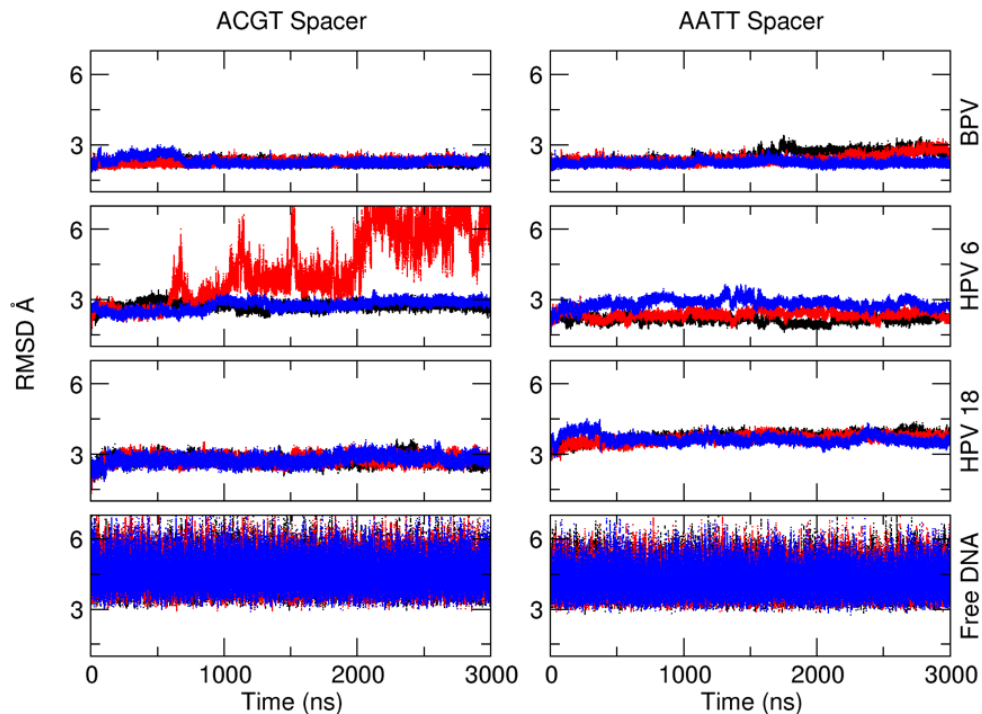


Figure 4.2: RMSD over time for molecular dynamics trajectories. Three copies are shown for each system, black, red, and blue. All systems with ACGT spacer DNA are on the left, and systems with AATT spacers are on the right. The RMSDs were calculated on all heavy atoms in reference to the X-ray structures, and indicate simulations were stable over 3000 ns, the exception being one copy of HPV 6 ACGT. Visual inspection of this trajectory showed that one of the DNA consensus sites became unbound from corresponding α helix, the other half-site remained bound to E2. The relatively high RMSD for free DNA occurred because the reference structure was bent DNA from the E2-DNA complex.

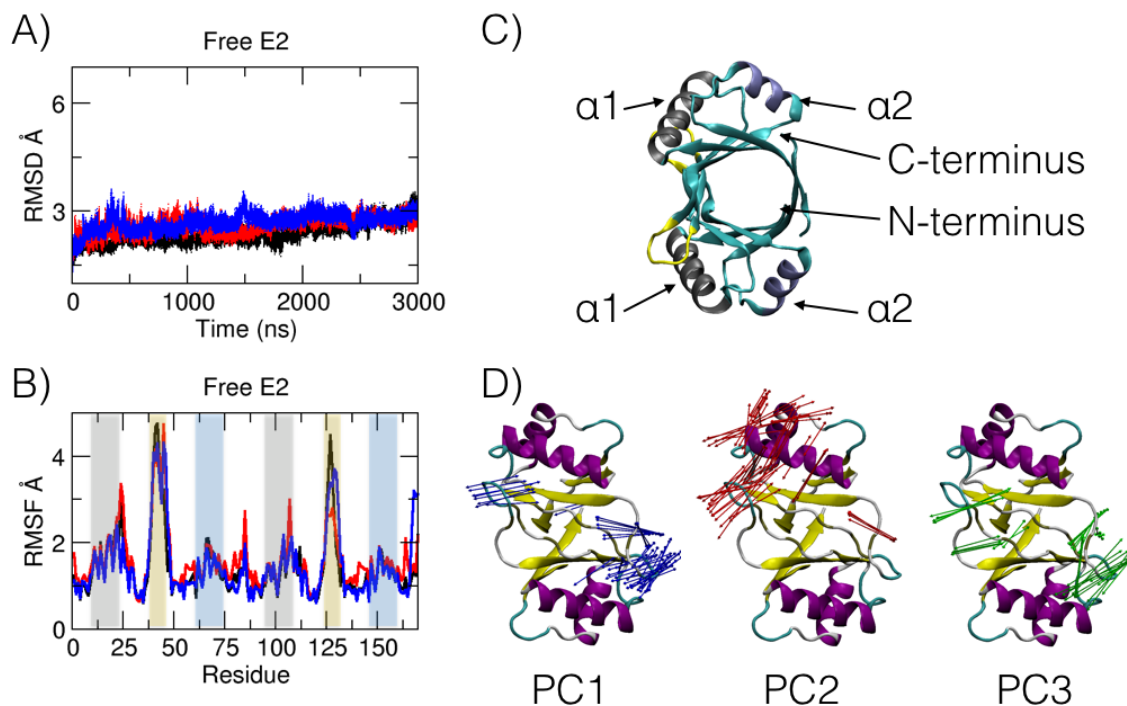


Figure 4.3: Free E2 RMSD, RMSF, and PCA from molecular dynamics trajectories. A) RMSD was determined for all heavy atoms with X-ray structure as reference. Three independent copies of the trajectories are shown in black, red, and blue. B) The atomic fluctuations were calculated by residue. Three copies are shown. Highlighted regions correspond to $\alpha1$ helices (gray), $\alpha2$ helices (blue), and the disordered region (yellow) between sheets $\beta2$ and 3. Residues were renumbered starting from 1 to 85, and both monomers are shown. C) An average structure from an aggregated trajectory of the three independent copies. Coloring matches the description in B, with remaining regions in cyan. α helices are labeled for both monomers but only one N-terminus and C-terminus are indicated with arrows. D) Principal component analysis was calculated and porcupine plots generated to indicate collective motion for the top three modes of free E2. The structures in show the DNA binding face and the $\alpha1$ helices, a 90° rotation from the view in C). Refer to main text for description of the modes.

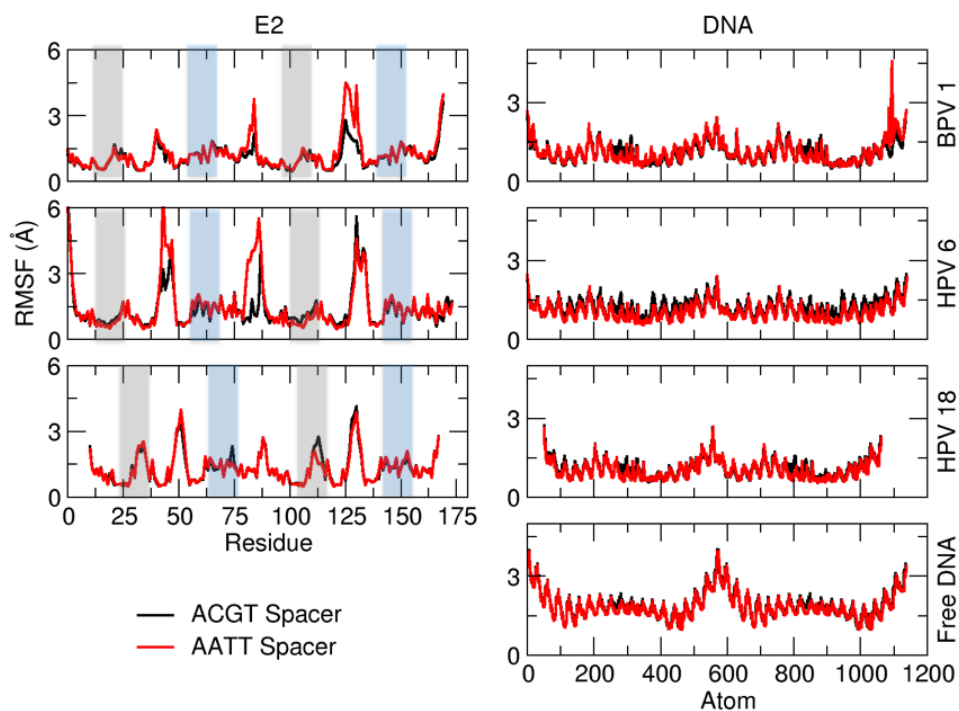


Figure 4.4: RMSF of free DNA, E2-bound DNA, and DNA-bound E2. Atomic fluctuations were determined relative to average structures, by residue for E2 and by atom for DNA. All RMSF calculations were done on aggregated trajectories, 9000 ns each. Protein residue numbers are numbered from zero. Both monomers for E2, and both strands of DNA, are shown. Highlighted regions on E2 correspond to $\alpha 1$ helices (gray) and $\alpha 2$ helices (blue).

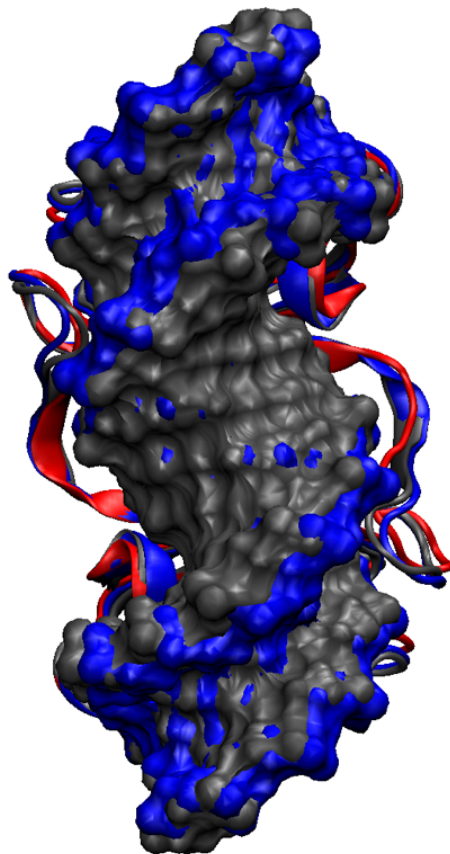


Figure 4.5: BPV E2 loop regions hug DNA minor grooves. The BPV E2-DNA complex from X-ray crystallography (gray) with the average structure from molecular dynamics of the same BPV E2-DNA complex (blue), and the average structure from molecular dynamics of free E2 (red). The conformation of the $\beta 2$ - $\beta 3$ loops between the three structures show similar conformation between X-ray crystal, E2-DNA from simulation, and free E2 from simulation. The two E2-DNA complexes show the loops have close association with the minor grooves, and the free E2 loops adopt a favorable conformation for the minor grooves even in the absence of DNA. The DNA sequences in the E2-bound DNA contain the ACGT spacer. DNA are represented as molecular surface and E2 as cartoons representations. The two average structures from MD were aggregated trajectories totaling 9000 ns each. Structures were aligned by all E2 heavy atoms.

BPV 1

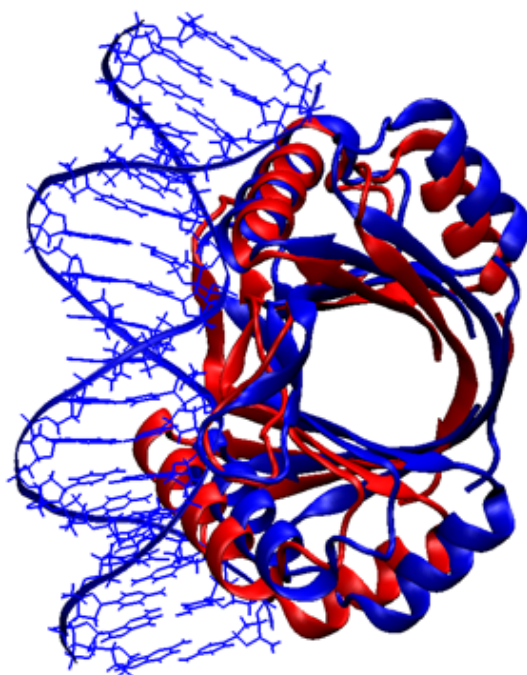


Figure 4.6: RMSD of $\alpha 1$ helix between free E2 and DNA-bound E2. The average structure of free BPV E2 was aligned to the average BPV E2-DNA complex based on the $\alpha 1$ helix from a single monomer. The RMSD was calculated between the $\alpha 1$ helices of the other monomer. Free E2 (red) has an RMSD of 5.3 Å to E2-DNA (blue) $\alpha 1$.

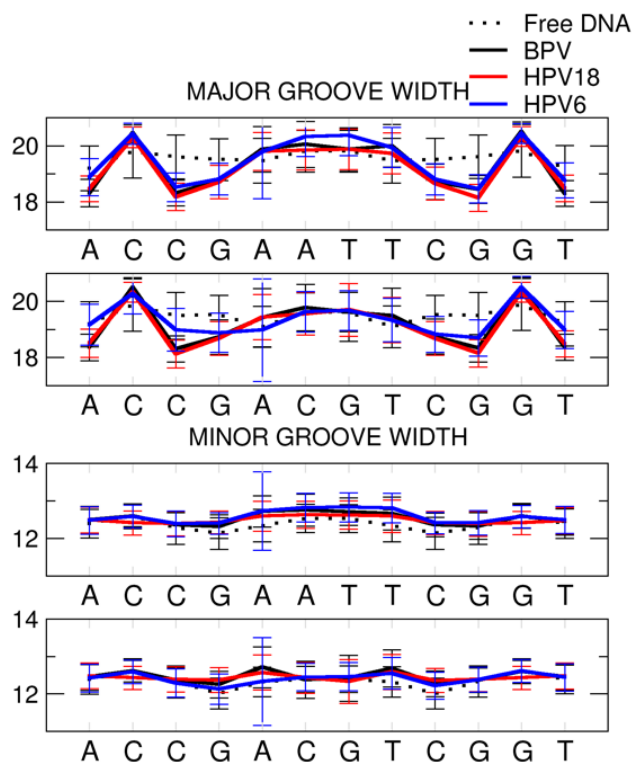


Figure 4.7: DNA major and minor groove widths. Groove widths for the internal 12 base pairs of all free and E2-bound DNA. Major groove widths for DNA with ACGT and AATT spacers are on top, while minor groove widths for the two spacer sequences are shown on the bottom. Groove widths were calculated on trajectories aggregated from three independent copies of each system, each aggregate totaling 9000 ns. Major groove widths are reduced in the consensus half-sites, compared to free DNA. Minor groove widths were not compressed in the spacer regions.

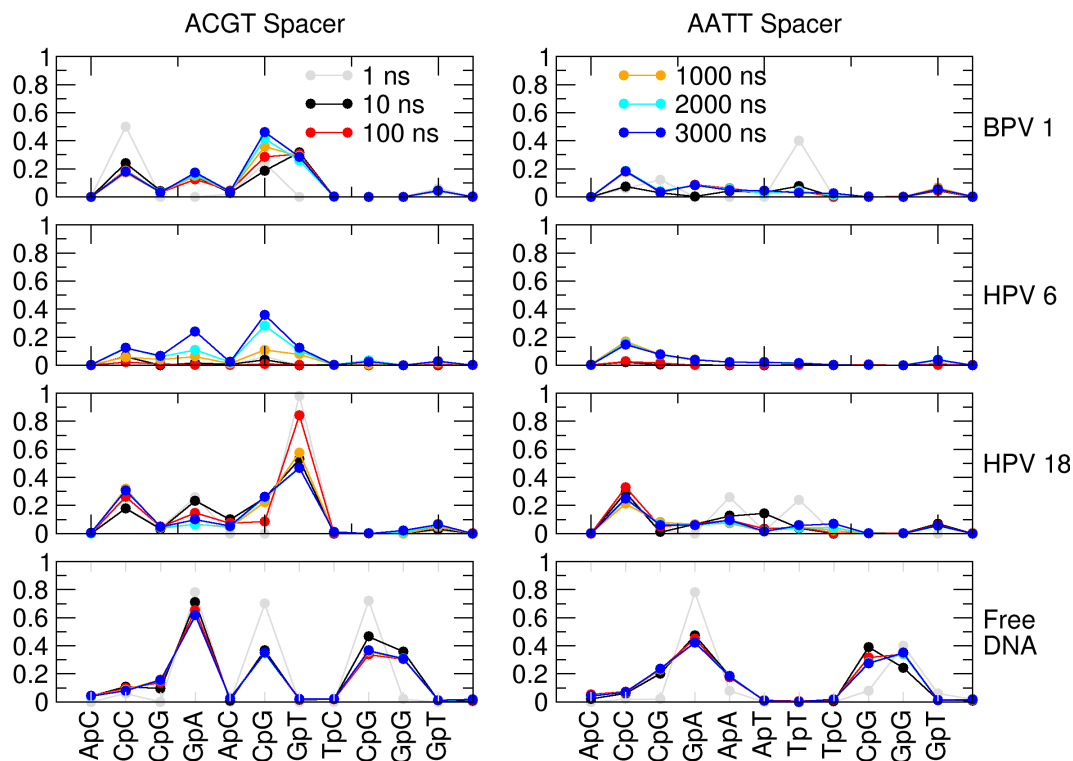


Figure 4.8: Convergence of fraction BII in free and E2-bound DNA. Fraction BII was measured at increasing time points to determine convergence; BII was measured at 1 ns, 10 ns, 100 ns, 1000 ns, 2000 ns, and 3000 ns. Systems with ACGT spacer DNA are on the left, and systems with AATT spacer DNA are on the right. Overall the free DNA converges must faster than E2-bound DNA, see text for detailed discussion. The populations of BII at each step differ between AATT and ACGT (see Figure 4.9 and the text for discussion).

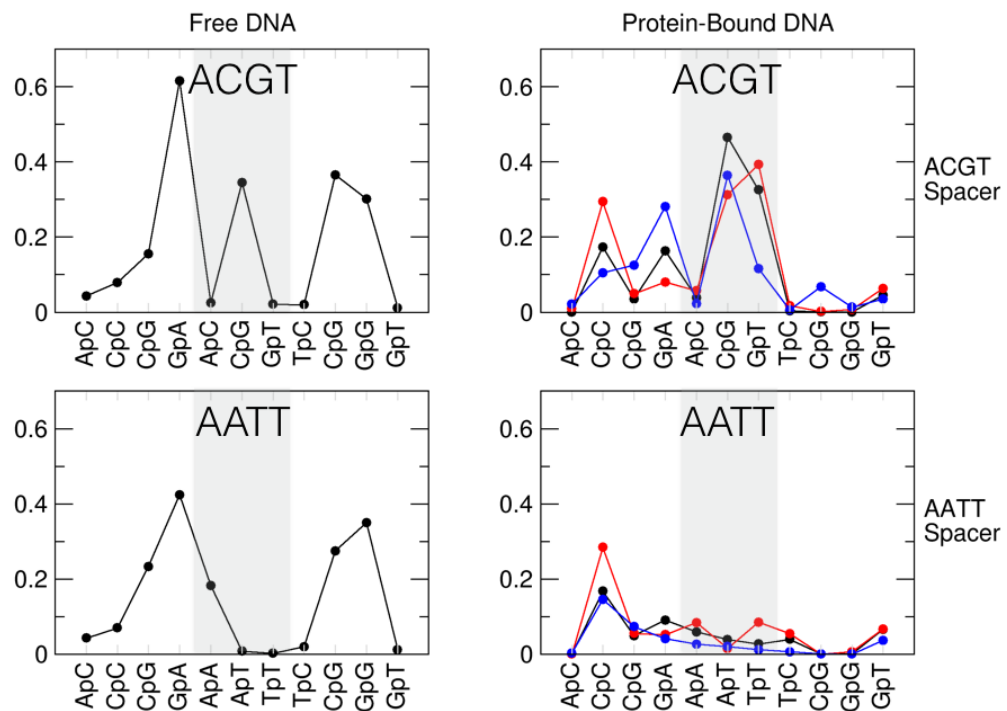


Figure 4.9: Fraction BII in free and E2-bound DNA. The fraction of BII from each base step over the internal 12 base pairs was calculated for free and E2-bound DNA. Sequence preference for BII is shown in free DNA (left column). In E2-bound DNA, both E2-induced and sequence-dependent BII populations are demonstrated. BPV E2-DNA is black, HPV 6 E2-DNA red, and HPV 18 E2-DNA blue. Gray shading was used to highlight the DNA spacer sequence.

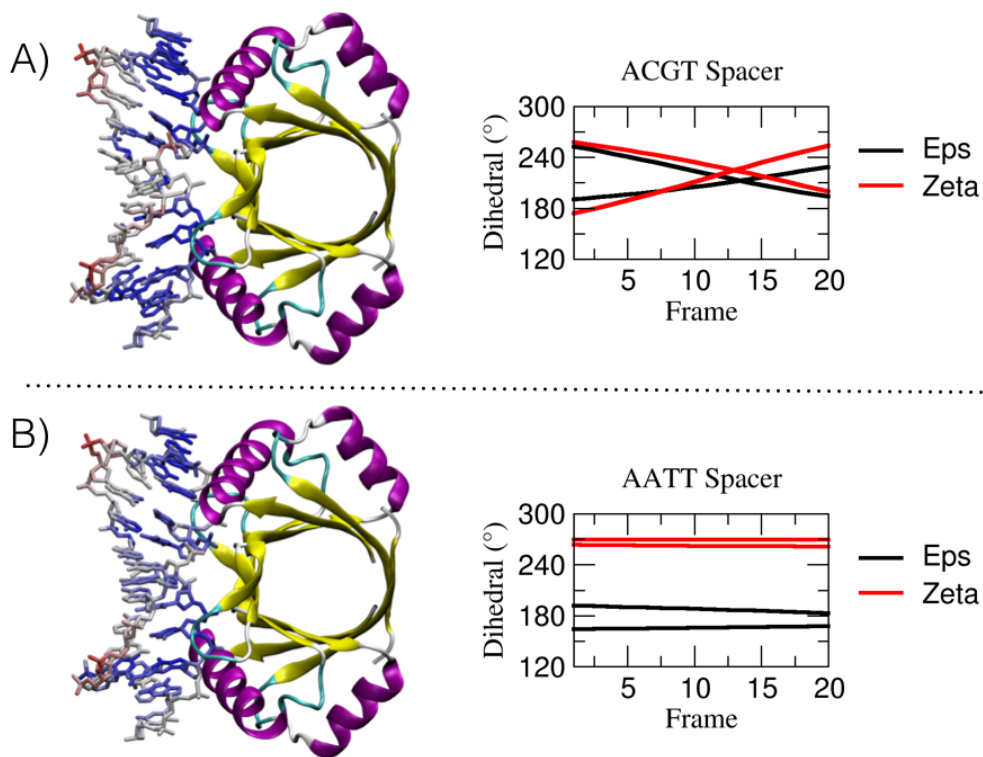


Figure 4.10: BI/BII dynamics in E2-bound DNA are sequence-dependent. The top modes of motion from principal component analysis are shown for BPV E2-bound DNA with A) DNA spacer ACGT and B) DNA spacer AATT. For both molecular graphics representations the protein is colored by secondary structure; α helices are purple, β sheets yellow, and remaining regions cyan. The DNA is colored by fluctuations of the pseudo trajectory generated by principle component analysis. Red regions have the highest fluctuations, followed by white, and then blue with the least. Based on visual inspection of the motions, A) and B) both show the major groove collapsing and expanding. The plots on the right show ϵ and ζ dihedrals for two steps in the spacer region, CpG and GpT for ACGT spacer, ApT and TpT for AATT spacer. For the ACGT spacer, CpG is observed transitioning from BII to BI, and GpT moves from BI to BII. No BI/BII transitions are observed for the AATT sequence. The top mode is shown in the figure, a full discussion of all modes can be found in the main text.

CHAPTER 5

CONCLUSION

5.1 Significance

The force field evaluation presented in Chapter 2 demonstrated that both bsc1 and OL15 parameter sets improve upon earlier versions of AMBER nucleic acid force fields for simulating DNA. Agreement with experiment was highlighted by $< 1 \text{ \AA}$ RMSD of average structures from simulations to NMR reference for the DDD system. The sampling times to evaluate the force fields with the DDD test system were remarkable, 100 copies of simulations each run to $10 \mu\text{s}$. Combined with total aggregated trajectories from all test systems, sampling was nearly 7 ms. These long time scales and multiple copies are required to uncover deficiencies in the force fields, the extensive sampling and agreement with experiment show the robustness of the most recent nucleic acid force fields. There are very few research groups capable of evaluating the force fields with the extensive sampling we employed, due to our relative wealth of computational resources, and we expect the community benefits from these efforts.

We demonstrated BII convergence in protein-bound DNA for the first time, that we are aware of. The native and decoy BII states converged by $\sim 1 \mu\text{s}$ for most bases, but the high population BII states in the spacer sequence were not fully converged by μs . Free DNA was shown to converge much faster than protein-bound DNA. In Chapter 3, it was demonstrated that high population BII states from simulations corresponded to BII states crystallized in BPV E2-DNA. The unexpected observation that BI/BII backbone transitions are a dominant mode of motion in E2-DNA led to investigating the role of sequence-dependent BII dynamics in E2-bound DNA. We found that E2 induces BII populations that differ from free DNA, but are conserved across three types of E2 protein, BPV, HPV 6, and HPV 18. The sequence of spacer in E2-bound DNA, did influence BII populations, supporting the idea that protein-bound BII populations are sequence-dependent, and not entirely determined by the protein-binding partner.

Understanding DNA structure and dynamics is important for determining mechanisms

of ligand binding to DNA, protein-DNA interactions, and DNA conformational changes. The work presented here demonstrated progress in the field by evaluating the state of the art nucleic acid force fields, showing convergence of BII populations in protein-bound DNA, and revealing that BI/BII transitions are dominant modes of motion in E2-bound DNA with ACGT spacers.

5.2 Future Directions

Nucleic acid simulations present many opportunities for future research. Although improvements, and strong agreement with experiment, were shown in Chapter 2, the nucleic acid force fields can continue to be improved. The RNA force field was not addressed in this work but that is an area of research behind DNA. For DNA force fields, however, there were still discrepancies in many helical parameters. Additionally, simulations will likely soon be extended to millisecond time scales and currently inaccessible events like base-pair opening in the internal helix will be observed and may reveal further deficiencies in the current force fields. This is also significant because millisecond simulations will also allow further testing the hypothesis that this is a gap in DNA dynamics on the μs - ms time scale.

Determining whether BI/BII dynamics are important for differential E2 binding to AATT and ACGT spacer sequences is an important and interesting question to pursue. A carefully constructed biasing potential could be used to simulate free energy of binding the different spacers to DNA using umbrella sampling. Investigating BI/BII dynamics can also be extended to larger systems like the nucleosome. Molecular dynamics simulations of nucleic acids will continue to be an important tool for investigating structure, energetics, and dynamics on the atomic level.

REFERENCES

- [1] Bai, X.; McMullan, G.; Scheres, H.W.S. How Cryo-EM is Revolutionizing Structural Biology. *Trends Biochem. Sci.* **2014**, *40*, 49–57.
- [2] Shi, Y. A Glimpse of Structural Biology Through X-Ray Crystallography. *Cell* **2014**, *159*, 995–1014.
- [3] Clore, M.G. NMR in Structural and Cell Biology. In *Encyclopedia of Cell Biology*; Bradshaw, R.A. Stahl, P.D., Eds.; Academic Press: Waltham, MA., 2016, Vol. 1, pp 98–107.
- [4] Karplus, M. McCammon, J.A. Molecular Dynamics Simulations of Biomolecules. *Nature Struct. Biol.* **2002**, *9*, 646–52.
- [5] Galindo-Murillo, R.; Roe, D. R.; Cheatham, T.E., III. On the Absence of Intrahelical DNA Dynamics on the μ s to ms Timescale. *Nat. Commun.* **2014**, *5*, 5152.
- [6] Salomon-Ferrer, R.; Götz, A.W.; Poole, D.; LeGrand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- [7] Götz, A.W.; Williamson, M.J.; Xu, Dong; Poole, D.; LeGrand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.
- [8] Pearlman, D.A.; D.A., Case; Caldwell, J.W.; Ross, W.S.; Cheatham, T.E., III; DeBolt, S.; Ferfuson, D. M.; Seibel, G.L.; Kollman, P.A. Molecular Dynamics Simulations of Biomolecules. *Nature Struct. Biol.* **2002**, *9*, 646–52.
- [9] Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
- [10] Galindo-Murillo, R.; Bergonzo, C.; Cheatham, T.E., III. Molecular Modeling of Nucleic Acid Structure. In *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc., 2001.
- [11] Bergonzo, C.; Galindo-Murillo, R.; Cheatham, T.E., III. Molecular Modeling of Nucleic Acid Structure: Energy and Sampling. In *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc., 2001.
- [12] Case, D.A.; Berryman, J.T.; Betz, R.M.; Cerutti, D.S.; Cheatham, T.E., III; Darden, T.A.; Duke, R.E.; Giese, T.J.; Golke, H.; Goetz, A.W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H.T.; Omelyan, A.; Onufriev, A.; Roe, D.R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling,

- C.L.; Smith, W.; Swails, J.; Walker, R.C.; Wang, J.; Wolf, R.M.; Wu, X.; York, D.M.; Kollman, P.A. *AMBER 2015*, University of California, San Francisco, 2015.
- [13] Hopkins, C.W.; LeGrand, S.; Walker, R.C.; Roitberg, A.E. Long Time Step Molecular Dynamics Through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.
- [14] Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Shaw, D.E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.
- [15] Simmerling, C.; Strockbine, B.; Roitberg, A.E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Amer. Chem. Soc.* **2002**, *124*, 11258–11259.
- [16] Galindo-Murillo, R.; Roe, D.R.; Cheatham, T.E., III Convergence and Reproducibility in Molecular Dynamics Simulations of the DNA Duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta, Gen. Subj.* **2014**, *1850*, 1041–1058.
- [17] Shan, Y.; Kim, E.T.; Eastwood, M.P.; Dror, R.O.; Seeliger, M.A.; Shaw, D.E. How Does a Drug Molecule Find Its Target Binding Site? *J. Amer. Chem. Soc.* **2011**, *133*, 9181–9183.
- [18] Cheatham, T.E., III; Miller, J.L.; Fox, T.; Darden, T.A.; Kollman, P.A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Amer. Chem. Soc.* **1995**, *117*, 4193–4194.
- [19] Pérez, A.; Luque, F.J.; Orozco, M. Dynamics of B-DNA on the Microsecond Time Scale. *J. Amer. Chem. Soc.* **2007**, *129*, 14739–14745.
- [20] Perez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T.E., III; Laughton, C.A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- [21] Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S.A.; Case, D.A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2015**, *13*, 55–58.
- [22] Zgarbova, M.; Otyepka, M.; J., Šponer ; Mládek, A.; Banáš, P.; Cheatham, T.E., III; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.
- [23] Zgarbova, M.; Luque, F.J.; Sponer, J.; Cheatham, T.E., III; Otyepka, M.; Jurecka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339–2354.
- [24] Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T.E., III; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, .

- [25] C., Liu; P.A., Janowski; Case, D.A. All-Atom Crystal Simulations of DNA and RNA Duplexes. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1059 – 1071.
- [26] Beveridge, D.L.; Barreiro, G.; Byun, K.S.; Case, D.A. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on d(CpG) Steps. *Biophys. J.* **2004**, *87*, 3799–3813.
- [27] Beveridge, D.L.; Cheatham, T.E., III; Mezie, M. The ABCs of Molecular Dynamics Simulations on B-DNA, circa 2012. *J. Biosci.* **2012**, *37*, 379–397.
- [28] Pasi, M.; Maddocks, J.H.; Beveridge, D.L.; Bishop, T.C.; Case, D.A.; Cheatham, T.E., III; Dans, P.D.; Jayaram, B.; Lanksa, F.; Laughton, C.; Mitchell, J.; Osman, R.; Orozco, M.; Perez, A.; Petkeviciute, D.; Spackova, N.; Sponer, J.; Zakrzewska, K.; Lavery, R. μ ABC: A Systematic Microsecond Molecular Dynamics Study of Tetranucleotide Sequence Effects in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 12272–83.
- [29] Dans, P.D.; Pérez, A.; Faustino, I.; Lavery, R.; Orozco, M. Exploring Polymorphisms in B-DNA Helical Conformations. *Nucleic Acids Res.* **2012**, *40*, 10668–10678.
- [30] Zgarbova, M.; Otyepka, M.; Šponer, J.; F., Lankaš; Jurečka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177–3189.
- [31] Reddy, S.Y.; Leclerc, F.; Karplus, M. DNA Polymorphism: A Comparison of Force Fields for Nucleic Acids. *Biophys. J.* **2003**, *84*, 1421–1449.
- [32] Dans, P.D.; Faustino, I.; Battistini, F.; Zakrzewska, K.; Lavery, R.; Orozco, M. Unraveling the Sequence-Dependent Polymorphic Behavior of d(CpG) Steps in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 11304–20.
- [33] Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. α/γ Transitions in the B-DNA Backbone. *Nucleic Acids Res.* **2002**, *30*, 5398–5406.
- [34] Hartmann, B.; Piazzola, D.; Lavery, R. BI-BII Transitions in B-DNA. *Nucleic Acids Res.* **1993**, *21*, 561–568.
- [35] Djuranovic, D.; Hartmann, B. DNA Fine Structure and Dynamics in Crystals and in Solution: The Impact of BI/BII Backbone Conformations. *Biopolymers* **2004**, *73*, 356–68.
- [36] Schwieters, C.D. Clore, G.M. A Physical Picture of Atomic Motions Within the Dickerson DNA Dodecamer in Solution Derived from Joint Ensemble Refinement Against NMR and Large-Angle X-Ray Scattering Data. *Biochemistry* **2007**, *46*, 1152–66.
- [37] Tian, Y.; Kayatta, M.; Shultis, K.; Gonzalez, A.; Mueller, L.J.; Hatcher, M.E. 31P NMR Investigation of Backbone Dynamics in DNA Binding Sites. *J. Phys. Chem. B* **2009**, *113*, 2596–603.
- [38] Heddi, B.; Foloppe, N.; Bouchemal, N.; Hantz, E.; Hartmann, B. Quantification of DNA BI/BII Backbone States in Solution. Implications for DNA Overall Structure and Recognition. *J. Am. Chem. Soc.* **2006**, *128*, 9170–9177.

- [39] Wecker, K.; Bonnet, M.C.; Meurs, E.F.; Delepierre, M. The Role of the Phosphorus BI-BII Transition in Protein-DNA Recognition: The NF κ B Complex. *Nucleic Acids Res.* **2002**, *30*, 4452–4459.
- [40] Steffen, N.R.; Murphy, S.D.; Toller, L.; Hatfield, G.W.; Lathrop, R.H. DNA Sequence and Structure: Direct and Indirect Recognition in Protein-DNA Binding. *Bioinformatics* **2002**, *18*, S22–S30.
- [41] Rohs, R.; West, S.M.; Sosinsky, A.; Liu, P.; Mann, R.S.; Honig, B. The Role of DNA Shape in Protein-DNA Recognition. *Nature* **2009**, *461*, 1248–53.
- [42] McBride, A.A. The Papillomavirus E2 Proteins. *Virology* **2013**, *445*, 57.
- [43] Hegde, R.S.; Grossman, S.R.; Laimins, L.A.; Sigler, P.B. Crystal Structure at 1.7 Å of the Bovine Papillomavirus-1 E2 DNA-Binding Domain Bound to its DNA Target. *Nature* **1992**, *359*, 505.
- [44] Rogers, A.; Waltke, M.; Angeletti, P.C. Evolutionary Variation of Papillomavirus E2 Protein and E2 Binding Sites. *Virol. J.* **2011**, *8*, 379.
- [45] Hooley, E.; Fairweather, V.; Clarke, A.R.; Gaston, K.; Bray, L.R. The Recognition of Local DNA Conformation by the Human Papillomavirus Type 6 E2 Protein. *Nucleic Acids Res.* **2006**, *34*, 3897–3908.
- [46] Kim, S.; Tam, J. K.; Wang, A.; Hegde, R.S. The Structural Basis of DNA Target Discrimination by Papillomavirus E2 Proteins. *J. Biol. Chem.* **2000**, *275*, 31245–31254.
- [47] Dell, D.; Wilkinson, R. K.W.and Tranter; Parish, J.; Brady, R.L.; Gaston, K. Comparison of the Structure and DNA-Binding Properties of the E2 Proteins from an Oncogenic and a Non-oncogenic Human Papillomavirus. *J. Mol. Biol.* **2003**, *334*, 979 – 991.
- [48] Hines, C.S.; Meghoo, C.; Shetty, S.; Biburger, M.; Brenowitz, M.; Hegde, R.S. DNA Structure and Flexibility in the Sequence-Specific Binding of Papillomavirus E2 Proteins. *J. Mol. Biol.* **1998**, *276*, 809 – 818.
- [49] 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.; Kollman, P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Amer. Chem. Soc.* **1995**, *117*, 5179–5197.
- [50] Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The Development/Application of a ‘Minimalist’ Organic/Biochemical Molecular Mechanic Force Field Using a Combination of ab initio Calculations and Experimental Data. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications* Gunsteren, W.F.; Weiner, P.K.; Wilkinson, A.J., Eds.; Springer Netherlands: Dordrecht, 1997, Vol. 3, 83–96.
- [51] Cheatham, T.E., III; Cieplak, P.; Kollman, P.A. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–62.

- [52] Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comp. Chem.* **2000**, *21*, 1049–1074.
- [53] Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T.E., III; Jurečka, P.; Spöner, J. Reference Simulations of Noncanonical Nucleic Acids with Different χ Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506–2520.
- [54] Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **1983**, *79*, 926.
- [55] Berendsen, H.J.C.; Grigera, J.R.; Straatsma, T.P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- [56] Horn, H.W.; Swope, W.C.; Madura, J.D.; Dick, T.L.; Hura, G.L.; Head-Gordon, T.L. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665.
- [57] Izadi, S.; Anandakrishnan, R.; Onufriev, A.V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.
- [58] Bergonzo, C.; Cheatham, T.E., III. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11*, 3969–3972, PMID: 26575892.
- [59] Mukhopadhyay, A.; Fenley, A.T.; Tolokh, I.S.; Onufriev, A.V. Charge Hydration Asymmetry: The Basic Principle and How to Use It to Test and Improve Water Models Charge Hydration Asymmetry. *J. Phys. Chem. B* **2012**, *116*, 9776.
- [60] Joung, I.S.; Cheatham, T.E., III. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020.
- [61] Joung, I.S.; Cheatham, T.E., III. Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2008**, *113*, 13279–13290.
- [62] Shih, C.C.; Georghiou, S. Large-Amplitude Fast Motions in Double-Stranded DNA Driven by Solvent Thermal Fluctuations. *Biopolymers* **2006**, *81*, 450–463.
- [63] Pastor, R.W.; Brooks, B.R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65*, 1409–1419.
- [64] Darden, T.A.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.
- [65] Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.
- [66] Case, D. A.; Cheatham, T.E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber Biomolecular Simulation Programs. *J. Comp. Chem.* **2005**, *26*, 1668–1688.

- [67] Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; Cheatham, T.E., III; Darden, T.A.; Duke, R.E.; Gohlke, H.; Goetz, A.W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H.T.; Omelyan, A.; Onufriev, A.; Roe, D.R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C.L.; Smith, W.; Swails, J.; Walker, R.C.; Wang, J.; Wolf, R.M.; Wu, X.; York, D.M.; Kollman, P.A. *AMBER 2014*, University of California, San Francisco, 2014.
- [68] Cheatham, T.E., III; Case, D.A. Using AMBER to Simulate DNA and RNA. In *Computational Studies of RNA and DNA* Sponer, J.; Lankaš, F., Eds.; Springer Netherlands, 2006, Vol. 2, 45–72.
- [69] Roe, D.R.; Cheatham, T.E., III PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data *J. Chem. Theory Comput.* **2013**, *9*, 3084.
- [70] Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929.
- [71] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- [72] Bakan, A.; Meireles, L.M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575.
- [73] Dršata, T.; Pérez, A.; Orozco, M.; Morozov, A.V.; Šponer, J.; Lankaš, F. Structure, Stiffness and Substates of the Dickerson-Drew Dodecamer. *J. Chem. Theory Comput.* **2012**, *9*, 707–721.
- [74] Pérez, A.; Lankas, F.; Luque, F. J.; Orozco, M. Towards a Molecular Dynamics Consensus View of B-DNA Flexibility. *Nucleic Acids Res.* **2008**, *36*, 2379–2394.
- [75] Čech, P.; Kukul, J.; Černý, J.; Schneider, B.; Svozil, D. Automatic Workflow for the Classification of Local DNA Conformations. *BMC Bioinformatics* **2013**, *14*, 205.
- [76] C., Bergonzo; N.M., Henriksen; D.R., Roe; J.M., Swails; A.E., Roitberg; Cheatham, T.E., III. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492.
- [77] Savelyev, A. MacKerell, A.D. All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J. Comp. Chem.* **2014**, *35*, 1219–39.
- [78] Wu, Z.; Delaglio, F.; Tjandra, N.; Zhurkin, V.B.; Bax, A. Overall Structure And Sugar Dynamics Of A DNA Dodecamer From Homo- and Heteronuclear Dipolar Couplings and ³¹P Chemical Shift Anisotropy. *J. Biomol. NMR* **2003**, *26*, 297–315.
- [79] Drew, H.R.; Wing, R.M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R.E. Structure of a B-DNA Dodecamer: Conformation and Dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179–83.
- [80] MacDonald, D.; Herbert, K.; Zhang, X.; Pologruto, T.; Lu, P.; Polgruto, T. Solution Structure of an A-Tract DNA Bend. *J. Mol. Biol.* **2001**, *306*, 1081–98.

- [81] Han, G.W.; Langs, D.; Kopka, M.L.; Dickerson, R.E., Unpublished work, 2004.
- [82] Maehigashi, T.; Hsiao, C.; Woods, K.K.; Moulaei, T.; Hud, N.V.; Williams, L.D. B-DNA Structure is Intrinsically Polymorphic: Even at the Level of Base Pair Positions. *Nucleic Acids Res.* **2012**, *40*, 3714–22.
- [83] Tereshko, V.; Wilds, C.J.; Minasov, G.; Prakash, T.P.; Maier, M.A.; Howard, A.; Wawrzak, Z.; Manoharan, M.; Egli, M. Detection of Alkali Metal Ions in DNA Crystals Using State-of-the-Art X-Ray Diffraction Experiments. *Nucleic Acids Res.* **2001**, *29*, 1208–15.
- [84] Strauch, M. A. ProteinDNA Complexes: Specific. In *eLS*; John Wiley & Sons, Ltd, 2001.
- [85] Garvie, C.W.; Wolberger, C. Recognition of Specific DNA Sequences. *Mol. Cell* **2001**, *8*, 937.
- [86] Redding, S.; Greene, E.C. How do Proteins Locate Specific Targets in DNA? *Chem. Phys. Lett.* **2013**, *570*, 1.
- [87] Heddi, B.; Foloppe, N.; Oguey, C.; Hartmann, B. Importance of Accurate DNA Structures in Solution: The Jun-Fos Model. *J. Mol. Biol.* **2008**, *382*, 956.
- [88] Svozil, D.; Šponer, J. E.; Marchán, I.; Pérez, A.; Cheatham, T.E., III; Forti, F.; Luque, F. J.; Orozco, M.; Šponer, J. Geometrical and Electronic Structure Variability of the Sugar-Phosphate Backbone in Nucleic Acids. *J. Phys. Chem. B* **2008**, *112*, 8188–8197.
- [89] Schneider, B.; Cerny, J.; Svozil, D.; Cech, P.; Gelly, J. C.; Brevern, de A. G. Bioinformatic Analysis of the Protein/DNA Interface. *Nucleic Acids Res.* **2014**, *42*, 3381–94.
- [90] Trieb, M.; Rauch, C.; Wellenzohn, B.; Wibowo, F.; Loerting, T.; Liedl, K.R. Dynamics of DNA: BI and BII Phosphate Backbone Transitions *J. Phys. Chem. B* **2004**, *108*, 2470.
- [91] Djuranovic, D.; Hartmann, B. Conformational Characteristics and Correlations in Crystal Structures of Nucleic Acid Oligonucleotides: Evidence for Sub-States. *J. Biomol. Struct. Dyn.* **2003**, *20*, 771.
- [92] Srinivasan, A.R.; Olson, W.K. Nucleic Acid Model Building: The Multiple Backbone Solutions Associated with a Given Base Morphology. *J. Biomol. Struct. Dyn.* **1987**, *4*, 895.
- [93] van Dam, L.; Levitt, M.H. Nucleotides in the B and C Forms of Natural-Sequence Polymeric DNA: A New Model for the C Form of DNA. *J. Mol. Biol.* **2000**, *304*, 541.
- [94] Winger, R.H.; Liedl, K.R.; Pichler, A.; Hallbrucker, A.; Mayer, E. Helix Morphology Changes in B-DNA Induced by Spontaneous BI-BII Substate Interconversion. *J. Biomol. Struct. Dyn.* **1999**, *17*, 223.
- [95] Lefebvre, A.; Femandjian, S.; Hartmann, B. Sensitivity of NMR Internucleotide Distances to B-DNA Conformation: Underlying Mechanics. *Nucleic Acids Res.* **1997**, *25*, 3855.

- [96] Kalodimos, C.G.; Biris, N.; Bonvin, A.M.; Levandoski, M.M.; Guennegues, M.; Boelens, R.; Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science* **2004**, *305*, 386.
- [97] Rozenberg, H.; Rabinovich, D.; Frolow, F.; Hegde, R. S.; Shakked, Z. Structural Code for DNA Recognition Revealed in Crystal Structures of Papillomavirus E2-DNA Targets. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 15194.
- [98] Zhang, Y.; Xi, Z.; Hegde, R.S.; Shakked, Z.; Crothers, D.M. Predicting Indirect Readout Effects in Protein-DNA Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8337.
- [99] Androphy, E.J.; Lowy, D.R.; Schiller, J.T. Bovine Papillomavirus E2 Trans-Activating Gene Product Binds to Specific Sites in Papillomavirus DNA. *Nature* **1987**, *325*, 70.
- [100] Hegde, R.S. The Papillomavirus E2 Proteins: Structure, Function, and Biology. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 343.
- [101] Hawley-Nelson, P.; Androphy, E.J.; Lowy, D.R.; Schiller, J.T. The Specific DNA Recognition Sequence of the Bovine Papillomavirus E2 Protein is an E2-Dependent Enhancer. *EMBO J.* **1988**, *7*, 525.
- [102] Djuranovic, D. Hartmann, B. Molecular Dynamics Studies on Free and Bound Targets of the Bovine Papillomavirus Type I E2 Protein: The Protein Binding Effect on DNA and the Recognition Mechanism. *Biophys. J.* **2005**, *89*, 2542.
- [103] Falconi, M.; Oteri, F.; Eliseo, T.; Cicero, D.O.; Desideri, A. MD Simulations of Papillomavirus DNA-E2 Protein Complexes Hints at a Protein Structural Code for DNA Deformation. *Biophys. J.* **2008**, *95*, 1108.
- [104] Berendsen, H.J.C.; Postma, J.P.M.; Gunsteren, van W.F.; DiNola, A.; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684.
- [105] Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- [106] Madhumalar, A.; Bansal, M. Sequence Preference for BI/BII Conformations in DNA: MD and Crystal Structure Data Analysis. *J. Biomol. Struct. Dyn.* **2005**, *23*, 13–27.
- [107] Steger, G. Corbach, S. Dose-Dependent Regulation of the Early Promoter of Human Papillomavirus Type 18 by the Viral E2 Protein. *J. Virol.* **1997**, *71*, 50–58.
- [108] Bedrosian, C.L.; Bastiav, D. The DNA-Binding Domain of HPV-16 E2 Protein Interaction with the Viral Enhancer: Protein-Induced DNA Bending and Role of the Nonconserved Core Sequence in Binding Site Affinity. *Virology* **1990**, *174*, 557 – 575.
- [109] Robertson, J.C.; Cheatham, T.E., III. DNA Backbone BI/BII Distribution and Dynamics in E2 Protein-Bound Environment Determined by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2015**, *119*, 14111–14119.

- [110] Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- [111] Hizver, J.; Rozenberg, H.; Frolow, F.; Rabinovich, D.; Shakked, Z. DNA Bending by an Adenine-Thymine Tract and its Role in Gene Regulation. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8490–8495.
- [112] Crothers, D.M.; Haran, T.E.; Nadeau, J.G. Intrinsically Bent DNA. *J. Biol. Chem.* **1990**, *265*, 7093–7096.

CURRICULUM VITAE

JAMES C. ROBERTSON

CONTACT INFORMATION

University of Utah College of Pharmacy, Department of Medicinal Chemistry
30 S 2000 E
Salt Lake City, UT 84112
(801) 585-3285
j.c.robertson@utah.edu
www.linkedin.com/in/jamescrobertson1

EDUCATION

Ph.D., Medicinal Chemistry, *Biomolecular Simulations* **May 2016**

University of Utah College of Pharmacy
Advisor: Thomas E. Cheatham, III

B.S. Chemistry with ACS Certificate in Biochemistry **2011**

Southern Oregon University, Ashland, OR
Graduated Cum Laude

TEACHING EXPERIENCE

Teaching Assistant **Spring 2014**

Physiological Chemistry II, University of Utah College of Pharmacy Salt Lake City, UT

- Prepared and taught 5 lectures on transcription, translation, DNA replication, and viruses
- Graded assignments and exams; helped write exam questions
- Instructed students on difficult concepts and course material during office hours

Guest Lecturer**Fall 2014**

Organic Medicinal Chemistry, University of Utah College of Pharmacy Salt Lake City, UT

- Delivered two lectures on cholinergics

Student Mentor**2009-2010**

Organic Chemistry, Southern Oregon University

Ashland, OR

- Led workshops to solve organic chemistry review problems
- Selected by faculty to lead the workshops

Workshop Participant**Fall 2014**Center for Teaching and Learning Excellence Annual Teaching Symposium, University of Utah
Salt Lake City, UT

- Active Learning • Classroom Civility • Multimedia in Canvas • Spicing Up Your Lecture

PUBLICATIONS

4. Galindo-Murillo R, Robertson JC, Zgarbová M, Šponer J, Otyepka M, Jurečka P, Cheatham III, TE. Assessing the Current State of AMBER Force Field Modifications for DNA. *J. Chem. Theory Comput.*, Submitted For Review
3. Robertson JC, Cheatham III, TE. DNA Backbone BI/BII Distribution and Dynamics in E2 Protein-Bound Environment Determined by Molecular Dynamics Simulations, **2015**, *J. Phys. Chem. B*, 119, 14111-14119.
2. Robertson JC, Hurley N, Tortorici M, Ciossani G, Borrello MT, Vellore NA, Ganesan A, Mattevi A, Baron R. Expanding the Druggable Space of the LSD1/CoREST Epigenetic Target: New Potential Binding Regions for Drug-Like Molecules, Peptides, Protein Partners, and Chromatin, **2013**, *PLoS Comp. Biol.*, 9(7):e1003158. doi:10.1371/journal.pcbi.1003158
1. Dixon AS, Miller GD, Bruno BJ, Constance JE, Woessner DW, Fidler TP, Robertson JC, Cheatham III TE, Lim CS. Improved Coiled-Coil Design Enhances Interaction with Bcr-Abl and Induces Apoptosis, **2012**, *Mol. Pharm.*, 9, 187-195.

NATIONAL PRESENTATIONS & POSTERS

1. Presentation: "BI/BII Backbone Substate Dynamics in Protein-bound DNA" *American Chemical Society Spring Meeting* San Diego, CA, **Spring 2016**
2. Poster: "Assessing the Current State of AMBER Force Field Modifications for DNA" *American Chemical Society Spring Meeting* San Diego, CA, **Spring 2016**
3. Poster: "Human Low Molecular Weight Protein Tyrosine Phosphatases: Molecular Dynamics of A and B Isoforms" *International Society of Quantum Biology and Pharmacology President's Meeting* Telluride, CO, **June 2014**
4. Poster: "Molecular Dynamics Generated Ensemble for Structure-Based Drug Design" *Biophysical Society 58th Annual Meeting* San Francisco, CA, **Feb 2014**
5. Presentation: "Using Configurational Ensembles to Expand LSD1/CoREST Druggability" *94th Annual AAAS Pacific Division Meeting* Las Vegas, NV, **June 2013**
6. Presentation and Poster: "Ensemble-Based Virtual Screening of LSD1/CoREST" *SC12 Early Research Showcase, SC12* Salt Lake City, UT, **Nov 2012**
7. Poster: "Ensemble-Based Virtual Screening of LSD1/CoREST" *Utah Bioscience Symposium* Salt Lake City, UT, **Sep 2012**
8. Poster: "Primer-Directed Biocement and Kinase Searches from *Phragmatopoma lapedosa* and *Pectinaria gouldii* cDNA" *American Chemical Society Spring Meeting* Anaheim, CA, **Spring 2011**

HONORS & AWARDS

- AAAS Pacific Division Student Travel Grant, *94th Annual AAAS Pacific Division Meeting*, Las Vegas, NV, **June 2013**
- Coyner Graf Memorial Scholarship, Southern Oregon University, **2010-2011**
- National Science Foundation Research Experience for Undergraduates Chemistry Leadership Group Travel Award, **Spring 2011**
- Southern Oregon University Department of Chemistry Outstanding Service Award, **2010-2011**
- Southern Oregon University Department of Chemistry Award in Inorganic Chemistry, **2010-2011**

LEADERSHIP & SERVICE

- Student Advisory Committee for Retention, Promotion, and Tenure; University of Utah College of Pharmacy, *Member and Chair* **2014-2016**
- Tutoring PharmD Students for Physiological Chemistry II (MDCH 5120); University of Utah, **Spring 2014**
- Biological Chemistry Student Retreat Organizing Committee; University of Utah, **2012-2013**
- Chemistry Club President; Southern Oregon University, **2010-2011** (*Member 2009-2011*)

TECHNICAL SKILLS & EXPERTISE

- Molecular dynamics simulations and data analysis primarily with the AMBER software suite; additional experience with NAMD and Gromacs
- Utilizing HPC resources including Blue Waters at NCSA, and the following through XSEDE: Stampede, Maverick, Gordon, Comet, Keeneland, and Kraken
- Proficient in bash, awk, VMD, xmgrace, LaTeX, and tcl
- Familiarity with C++, python, html, perl, R, SQL, git, and parallel programming

ADDITIONAL RESEARCH EXPERIENCE**Baron Lab** **2012-2014**University of Utah, Department of Medicinal Chemistry Salt Lake City, UT

Advisor: Riccardo Baron

- Biomolecular Simulations, Computational Drug Discovery

Biological Chemistry Program 1st Year Rotations **2011-2012**University of Utah Biological Chemistry Graduate Program Salt Lake City, UT

- Cheatham Lab • Rainier Lab • Sigman Lab • Ireland Lab • Baron Lab

National Science Foundation Research Experience for Undergraduates Summer 2010Coe College Chemistry Department Cedar Rapids, IA

Advisor: Maria A. Dean

- Investigating marine worm protein bio-cement using bioinformatics and scanning electron microscopy

Senior Capstone Research Project**2010-2011**

Southern Oregon University

Ashland, OR

Advisor: Hala G. Schepmann

- Natural product extraction and characterization from *Mahonia aquifolium*

Lab Assistant**2008-2009**

Southworth Mycorrhizae Lab, Southern Oregon University

Ashland, OR

Advisor: Darlene Southworth

- Sample collection, DNA amplification via PCR, DNA sequencing and analysis

MEMBERSHIPS & AFFILIATIONS

- American Chemical Society, **Since 2008**
- American Chemical Society Computers in Chemistry Division, **Since 2012**
- Biophysical Society, **Since 2013**
- American Association for the Advancement of Science, **Since 2013**