Application and Development of Computational Methods in Conformational Studies of Bio-molecules

Aleksandra Karolak  
*University of South Florida*, aleksandrak@mail.usf.edu

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Chemistry Commons

Scholar Commons Citation

https://digitalcommons.usf.edu/etd/5520

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.
Application and Development of Computational Methods in Conformational Studies of Bio-molecules

by

Aleksandra Karolak

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Department of Chemistry
College of Arts and Sciences
University of South Florida

Major Professor: Arjan van der Vaart, Ph.D.
Wayne Guida, Ph.D.
Randy Larsen, Ph.D.
Inna Ponomareva, Ph.D.
Brian Space, Ph.D.

Date of Approval:
April 10th, 2015

Keywords: Proteins, DNA, Conformational Changes, Molecular Dynamics

Copyright © 2015, Aleksandra Karolak
DEDICATION

To my parents and brother in picturesque Poland, to my husband. For making me a better person and unconditional love.
ACKNOWLEDGMENTS

I specially recognize my major research advisor, Dr. Arjan van der Vaart, for his continuous efforts to make me look further and accomplish more as a researcher. I truly thank Dr. Randy Larsen for his utter support and opening the door to his lab and beyond. I thank my committee members Dr. Wayne Guida, Dr. Inna Ponomareva and Dr. Brian Space for their guidance from my first steps taken at USF; the Martin Family for their goodwill; Dr. Dean Martin for his sense of humor. Thanks to all my colleagues, scientists and non-scientists from USF and institutions all over the world. I acknowledge my current and former lab mates for scientific discussions, problems solving and for a good company.

Dr. Arjan van der Vaart’s NSF CAREER Award CHE-1007816 supported this work; the computer time was provided by USF Research Computing and XSEDE.
TABLE OF CONTENTS

List of Tables iii
List of Figures iv
Abstract vi

Chapter One: Introduction and Methods 1
  1.1 Importance of Conformational Changes 1
  1.2 Methods employed in Conformational Studies of Bio-molecules 3
  1.3 Computational Techniques 4
  1.4 Fundamentals of Molecular Dynamics Simulations 5
  1.5 Force Field 9
  1.6 Treatment of Solvent 11
  1.7 Conformational Free Energy Calculations 15
    1.7.1 Replica Exchange Molecular Dynamics 16
    1.7.2 Umbrella Sampling Molecular Dynamics 17
    1.7.3 Weighted Histogram Analysis Method 18
  1.8 Time Scales and Calculations Efficiency in the Simulations 19

Chapter Two: Importance of Local Interactions for the Stability of Inhibitory Helix 1 in Apo ETS-1 21
  2.1 Abstract 21
  2.2 Introduction 22
  2.3 Materials and Methods 23
  2.4 Results 26
  2.5 Discussion and Conclusions 31

Chapter Three: Enhanced Sampling Simulations of DNA Step Parameters 33
  3.1 Abstract 33
  3.2 Introduction 34
  3.3 Methods 36
  3.4 Simulation Setup 44
  3.5 Results 46
  3.6 Method Refinement 53
  3.7 Discussion and Conclusions 54

Chapter Four: BII Stability and Base Step Flexibility of N6-Adenine Methylated GATC Motifs 56
4.1 Abstract 56
4.2 Introduction 56
4.3 Results and Discussion 59
4.4 Conclusions 65
4.5 Computational Methods 65

Chapter Five: Conformational Effects of Cytosine Oxidative Damage 68
5.1 Abstract 68
5.2 Introduction 69
5.3 Computational Setup 71
5.4 Results and Discussion 72
5.5 Conclusions 79

References 80

Appendix A: Permission from the Publisher for reprinting of the manuscript in chapter 2 98

Appendix B: Permission from the Publisher for reprinting of the manuscript in chapter 3 99
LIST OF TABLES

| Table 2.1: | Free energies of the HI-1 folded state relative to the unfolded state. | 27 |
| Table 3.1: | List of 235 pdb files used in correlation tests. | 40 |
| Table 3.2: | Correlation of DNA step parameters with 3DNA for representative sets of atomic selections. | 41 |
| Table 3.3: | Average step parameters in umbrella sampling simulations of unmethylated DNA calculated by the introduced method and 3DNA. | 49 |
| Table 3.4: | Values of the step parameters at the free energy minima. | 51 |
| Table 3.5: | Stiffness constants of the step parameters. | 52 |
| Table 4.1: | Population (%) of BII forms in the GATC motif. | 59 |
| Table 4.2: | Populations (%) of mixed BI and BII states on DNA main and complementary strands at A_{6}T_{7}/A_{18}T_{19} and T_{7}C_{8}/G_{17}A_{18} steps. | 61 |
| Table 4.3: | Twist and roll covariance matrices. | 64 |
| Table 5.1: | Average values of DNA step parameters and corresponding standard deviations at the steps G_{6}C_{7}/G(A)_{22}C_{23} and C_{7}A_{8}/T_{21}G(A)_{22}. | 76 |
| Table 5.2: | Population (%) of BI states on DNA strands. | 76 |
| Table 5.3: | Sugar puckering (deg) at the bases around C_{7} on DNA main strand. | 78 |
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Partial unfolding of Ets-1 transcription factor upon DNA binding.</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Illustration of DNA modifications discussed in chapters 4 and 5.</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Structure of apo ETS-1.</td>
<td>24</td>
</tr>
<tr>
<td>2.2</td>
<td>Systems studied with REX.</td>
<td>25</td>
</tr>
<tr>
<td>2.3</td>
<td>Contacts observed in the folded state: F304 + V308, K305 + R309. In the unfolded state: F304 + R309.</td>
<td>27</td>
</tr>
<tr>
<td>2.4</td>
<td>Contacts between F304/Y307/Y329 in the HI-1+HI-2 construct in the folded state.</td>
<td>29</td>
</tr>
<tr>
<td>3.1</td>
<td>DNA step parameters.</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>Local and median vectors for the step parameters calculations.</td>
<td>37</td>
</tr>
<tr>
<td>3.3</td>
<td>Atom selection for purine and pyrimidine bases.</td>
<td>41</td>
</tr>
<tr>
<td>3.4</td>
<td>Correlation of calculated step parameters with 3DNA values.</td>
<td>43</td>
</tr>
<tr>
<td>3.5</td>
<td>Contribution of rise to the overall deformation energy.</td>
<td>43</td>
</tr>
<tr>
<td>3.6</td>
<td>Convergence of the free energy as a function of the central twist angle.</td>
<td>47</td>
</tr>
<tr>
<td>3.7</td>
<td>DNA conformational free energy as a function of the central step parameter for the unmethylated DNA calculated with method introduced and 3DNA.</td>
<td>48</td>
</tr>
<tr>
<td>3.8</td>
<td>DNA conformational free energy as a function of the central step parameter for the unmethylated and methylated system.</td>
<td>51</td>
</tr>
<tr>
<td>3.9</td>
<td>Refined correlation of calculated step parameters twist and rise with 3DNA values.</td>
<td>54</td>
</tr>
<tr>
<td>4.1</td>
<td>Simulated sequences containing GATC motif.</td>
<td>59</td>
</tr>
</tbody>
</table>
Figure 4.2: Correlation between stacking interactions and $\varepsilon - \zeta$ values. 60
Figure 4.3: Correlation between slide and twist and $\varepsilon - \zeta$ values. 62
Figure 4.4: Free energy as a function of the N4-O6 distance in the C8/G17 base pair. 63
Figure 4.5: Free energy surfaces calculated as a function of twist and roll. 64
Figure 5.1: Undamaged and hydroxyl damaged cytosine. 72
Figure 5.2: DNA base pair parameters. 73
Figure 5.3: Hydrogen bonding patterns observed in MD simulations. 73
Figure 5.4: DNA base pair step parameters. 74
Figure 5.5: Stacking interactions in the average structures of C7/G22 and C7OH/A22. 75
Figure 5.6: Free energy curves calculated as a function of twist steps flanking C7. 77
ABSTRACT

The work presented in my dissertation focuses on the conformational studies of biomolecules including proteins and DNA using computational approaches. Conformational changes are important in numerous molecular bioprocesses such as recognition, transcription, replication and repair, etc. Proteins recognize specific DNA sequences and upon binding undergo partial or complete folding or partial unfolding in order to find the optimal conformational fit between molecules involved in the complex. In addition to sequence specific recognition, proteins are able to distinguish between subtle differences in local geometry and flexibility associated with DNA that may further affect their binding affinities. Experimental techniques provide high-resolution details to the static structures but the structural dynamics are often not accessible with these methods; but can be probed using computational tools. Various well-established molecular dynamics methods are used in this work to study differences in geometry and mechanical properties of specific systems under unmodified and modified conditions. Briefly, the studies of several protein and DNA systems investigated the importance of local interactions and modifications for the stability, geometry and mechanical properties using standard and enhanced molecular dynamics simulations. In addition to the conformational studies, the development of a new method for enhanced sampling of DNA step parameters and its application to DNA systems is discussed.
Chapter 1 reviews the importance of the conformational changes in bioprocesses and the theory behind the computational methods used in this work. In the project presented in chapter 2 unbiased molecular dynamics and replica exchange molecular dynamics are employed to identify the specific local contacts within the inhibitory module of ETS-1. ETS-1 is a human transcription factor important for normal but also malignant cell growth. An increased concentration of this protein is related to a negative prognosis in many cancers. A part of the inhibitory module, inhibitory helix 1 (HI-1) is located on the site of the protein opposite to the DNA binding site and although loosely packed, stays folded in the apo state and unfolds upon ETS-1 binding to DNA. Our study investigated the character and importance of contacts between HI-1 and neighboring helices of the inhibitory module: HI-2 and H4. We also identified a mutant of HI-1, which possessed the higher helical propensity than the original construct. This study supported the experimental findings and enhanced the field by the identification of new potential target for experimental tests of the system, which plausibly inhibits binding to DNA.

In the studies discussed in chapters 3-5 the conformational dynamics of DNA under normal conditions and upon specific epigenetic modifications are presented. Since DNA conformation can be accurately described by six base pair step parameters: twist, tilt, roll, shift, slide and rise, these were extensively analyzed and the results elucidated insights into the properties of the systems. In order to enhance unbiased simulations and allow for easier crossing of the energy barriers, we developed and implemented a novel method to control DNA base pair step parameters. With this approach we obtained the free energy estimates of e.g. DNA rearrangements in a more efficient manner. This advanced computational method, supported by standard and additional enhanced techniques, was then applied in the studies of DNA methylation on cytosine or adenine bases and oxidative damage of cytosine.
CHAPTER ONE: 
INTRODUCTION AND METHODS

1.1 Importance of Conformational Changes

The interactions between proteins and DNA drive a majority of biological processes including recognition, repair, and transcription. Proteins that bind to specific DNA sequences often undergo conformational changes, including the folding of specific domains or, in one particular case partial unfolding (Figure 1.1, chapter 2). In addition to recognition of specific sequences, proteins are able to detect subtle changes in the spatial organization of DNA through indirect readout, which is closely linked to the orientation of DNA bases, sugar-phosphate backbone and DNA distortion.

![Figure 1.1. Partial unfolding of Ets-1 transcription factor upon DNA binding. In the apo protein, inhibitory helix 1 (in black) is folded (a), in the bound state it unfolds (b).](image)
To provide the most stable structure of the protein-DNA complex, the process of recognition and binding is not only accompanied by the conformational rearrangements of proteins, but involves structural changes in DNA as well. Binding often induces alterations in local geometry, flexibility or overall bending of DNA.\(^\text{10}\) DNA flexibility thus plays an important role this process; for example, in DNA mismatch repair where bases are flipped, in DNA looping which regulates transcription and replication, and in DNA wrapping around histones in nucleosomes.\(^\text{11-16}\)

Conformational changes convey to the binding free energy through the entropy and enthalpy terms. Stability of the protein-DNA complex comes from the favorable contribution of enthalpy term related to direct protein-DNA interactions and entropy part related to release of water molecules and ions. These factors make binding a highly favorable process; in fact, in some cases, these two factors would make binding too favorable and outside the physiological range. In those cases, binding is brought to the physiological range by the introduction of conformational changes involving entropically unfavorable protein folding and/or enthalpically unfavorable DNA distortion. The ease of DNA distortion may be modulated by chemical modifications of DNA\(^\text{7,10}\) for example methylation or oxidation\(^\text{17-21}\), which can further change its flexibility, shape or curvature. In human cells, cytosine methylation is the most common epigenetic modification, which is important for normal conditions but also in many diseases.\(^\text{18,22-25}\) Methylation of an adenine base is the second most common epigenetic modification (Figure 1.2), but it is limited to prokaryotes and lower eukaryotes.\(^\text{26-29}\) The influence of methylation on DNA structure depends on DNA sequence, type of modified base, and position within a base, and this will be discussed in detail in chapters 3 and 4. Similar to methylation, oxidation of DNA bases may alter binding of proteins and influence biological processes (Figure 1.2). Unlike DNA
methylation however, oxidation is a highly undesirable modification due to its mutagenic character. Understanding the effects of this modification on DNA local geometry and stiffness is incomplete but important; fatal consequences can occur if this damage remains unrecognized. A detailed description of the conformational rearrangements taking place upon the oxidative damage of cytosine will be discussed in chapter 5.

![DNA modifications](image)

**Figure 1.2.** Illustration of DNA modifications discussed in chapters 4 and 5: thymine paired with N\textsubscript{6}-methyl adenine (a) and C\textsubscript{5}-hydroxyl cytosine paired with guanine (b). Modified groups are shown in black. Atoms involved in hydrogen bonding are highlighted as spheres with the increasing size in order: hydrogen, nitrogen and oxygen.

1.2 Methods employed in Conformational Studies of Bio-molecules

Many techniques can be used to study the structure and conformational motion of bio-molecules. For example, fluorescence resonance energy transfer (FRET) and fluorescence correlation spectroscopy provide information about intra or intermolecular distances. FRET has applications in many studies, from protein folding, to the unwrapping of DNA from the histone core. Ultraviolet-visible absorption spectroscopy uses the absorption of visible light leading to electronic transitions that detects changes in structure or follow the progress of many reactions. Information about the secondary structure can also be obtained with circular
dichroism spectroscopy, particularly for proteins. Other methods include mass spectroscopy, atomic force microscopy, photoacoustic calorimetry, nuclear magnetic resonance (NMR) spectroscopy, electron paramagnetic resonance, and X-ray crystallography. Infrared and Raman spectroscopy provide details about system's vibrational modes. The most common techniques to produce high-resolution structures of bio-molecules are NMR spectroscopy and X-ray crystallography. With NMR insights beyond static structures may be obtained, including the differences in chemical shifts and secondary structure, exchange rates and equilibrium constant calculated through measured nuclei relaxation times. In X-ray crystallography the diffraction produces a spatial electron density map of the crystal fit into the atomic model, or the crystal structure of the system. Noteworthy, fluctuations in atom positions around their average value, expressed through so called B factors, can provide certain information about system dynamics or indicate for which atoms the presence of fitting errors may be more important. While more methods capable of providing detailed information about the systems structure exist, NMR spectroscopy and X-ray crystallography offer the highest spatial resolution information up to date. The structures of bio-molecules obtained with either NMR or X-ray techniques are stored and explicitly available in the databases such as the Protein Data Bank (PDB), and these provide starting information for modeling and computational studies.

1.3. Computational Techniques

To fully understand the link between the system’s conformation and biological function, investigation of structural changes taking place is beneficial. Application of computer simulations may supplement the results of experimental studies on the high-resolution level. Simulations allow for the studies of systems behavior on a very short time scales, often
inaccessible with experiments. Until now, application of computer simulations rationalized experimental findings and bridged macroscopic properties with interactions on the molecular level for many systems and biological events.\textsuperscript{54-56} Numerous problems in the fields of medicine, chemistry, biology and physics have been addressed using computational methods coupled with experimental results.\textsuperscript{54} In addition to detailed structural studies computer simulations allowed for more efficient and less costly approach towards the development of new compounds, not limited to the field of medical or pharmaceutical applications. Simulations have been described as experiments on a computer helping predict system’s behavior.\textsuperscript{54} Within the computational methods most suitable for the conformational studies of bio-molecules, two fundamental, widely used techniques are Monte Carlo\textsuperscript{57} and Molecular Dynamics (MD),\textsuperscript{54} the latter was employed in the work presented here.

\textbf{1.4. Fundamentals of Molecular Dynamics Simulations}

The foundations of MD simulations lie in the integration of Newton's equations of motions and provide details about structure of bio-systems as a function of time. In classical mechanics, Newton's equations of motion are expressed as:

\[ F = ma, \]  

(1.1)

where \( F \) is the force acting on a particle or atom of mass \( m \), and \( a \) is acceleration. This equation can be recast to show the relationship between atomic coordinates, mass and potential energy:

\[ -\frac{dU}{dx} = m \frac{d^2x}{dt^2}, \]  

(1.2)

where \( U \) is the system's potential energy, \( x \) is the position and \( t \) is time. Newton's equations allow for the calculation of the motion of atoms characterized by certain masses under the forces acting on them. In MD simulations the starting structures and positions of atoms usually come from
NMR or X-ray structures as mentioned earlier, or from modeling if no data is available. The integration of system equations of motion is then performed by propagation algorithms. Various integrators are available for the propagation of positions and velocities in simulations; major differences between them lie in the time points at which evaluations are performed and the associated errors. An essential property of the integrator is the ability to preserve energy on long time scales and a commonly used integrator fulfilling this condition is the Verlet integrator. All of the work in this dissertation was based on Verlet integrators, which are obtained from Taylor expansions of the atomic positions $x$ around time $t$ keeping up to the third term in further calculations:

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{F(t)\Delta t^2}{2m} + \frac{\Delta t^3}{3!} \ddot{x} + O(\Delta t^4).$$  \hspace{1cm} (1.3)$$

Velocities $v$ corresponding to these positions, are simply derivates of the positions with respect to time. Addition of the third order Taylor expansion for $x(t + \Delta t)$ and $x(t - \Delta t)$ leads to the Verlet algorithm, where the new positions and new velocities are calculated by:

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{F(t)\Delta t^2}{m} + O(\Delta t^4),$$

$$v(t) = \frac{x(t + \Delta t) - x(t - \Delta t)}{2\Delta t} + O(\Delta t^2).$$ \hspace{1cm} (1.4)$$

One disadvantage of the Verlet algorithm rests in the introduction of numerical errors by summation of a term $2x(t) - x(t - \Delta t)$ which is linear in $\Delta t$, and $F(t)\Delta t^2/m$, which is quadratic in $\Delta t$. Derived from Verlet, with velocities defined at half-integer points, the Leap-frog integrator produces mathematically equivalent results using a velocity term in the position calculation:

$$x(t + \Delta t) = x(t) + v(t + \Delta t / 2)\Delta t + O(\Delta t^4),$$

$$v(t + \Delta t / 2) = v(t - \Delta t / 2) + \frac{F(t)\Delta t}{m} + O(\Delta t^2).$$ \hspace{1cm} (1.5)$$
In Leap-frog the magnitudes of terms in the position calculation are similar and the numerical errors introduced are smaller. However, the updated positions and velocities are estimated at different points in Leap-frog. The velocity-Verlet algorithm allowed for calculations of new positions and velocities at the same point:

\[
\begin{align*}
x(t + \Delta t) &= x(t) + v(t)\Delta t + \frac{F(t)\Delta t^2}{2m} + O(\Delta t^3), \\
v(t + \Delta t) &= v(t) + \frac{[F(t) + F(t + \Delta t)]}{2m} \Delta t + O(\Delta t^3).
\end{align*}
\]

(1.6)

The properties of MD integrators include e.g. efficiency, time reversibility and accuracy.\textsuperscript{54} The calculation of exact trajectories is not feasible with these integrators; however, as long as the integrator is volume or area preserving i.e. the determinant of Jacobian matrix for the transformation in time equals 1,\textsuperscript{54} the results produced are considered satisfactory.

In MD simulations, the newly generated ensemble of configurations corresponds to the micro canonical ensemble characterized by the constant number of particles \(N\), constant volume \(V\) and energy \(E\) (\(NVE\) ensemble).\textsuperscript{54,58} Constant temperature is commonly achieved with thermostats, such as one introduced by Nosé\textsuperscript{54,59} and modified by Hoover.\textsuperscript{60} Starting from the Lagrangian equations of motion:

\[
\begin{align*}
L &= K - U, \\
p &= \frac{\partial L}{\partial q}, \\
\dot{p} &= \frac{\partial L}{\partial q},
\end{align*}
\]

(1.7)

where \(K\) is kinetic energy, \(q\) and \(p\) correspond to generalized coordinates and momenta, the Nosé-Hoover thermostat generates the \(NVT\) ensemble by introducing an additional degree of freedom to the system:
\[ L_{\text{new}} = \sum_i \frac{m_i}{2} s^2 \dot{x}_i^2 - U(x^N) + \frac{Q}{2} s^2 - \frac{C}{\beta} \ln s. \]  

(1.8)

In this extended Lagrangian, \( s \) denotes the coordinates of the extra particle, \( Q \) corresponds to its effective mass, the third term on the right side corresponds to its kinetic energy and the fourth to its potential energy. When \( C \) is chosen as \( C = 3N + 1 \), with \( N \) the number of particles, thermodynamic averages correspond to the \( NVT \) ensemble.

Constant temperature in MD simulations can also be preserved through the Langevin equations of motions. In the Langevin equation, system dynamics are solved in the presence of friction, which represent collisions with the solvent that are present under real conditions.\(^{61}\) In Langevin dynamics the equations of motion take the form of:

\[-\frac{dU}{dx} - \lambda \frac{dx}{dt} + \eta(t) = m \frac{d^2x}{dt^2}.\]  

(1.9)

The second term on the left-hand side of Eq. 1.9 (\( \lambda \)) corresponds to the frictional force proportional to the atom's velocity and the third term corresponds to the stochastic "noise" or collision effect with a solvent (\( \eta \)).\(^{61}\) This approach allows for sampling of \( NVT \) ensemble but the dynamics generated are not deterministic or time-reversible.

MD can also be subjected to pressure control in order to generate the isothermal-isobaric (\( NPT \)) ensemble. Commonly a Langevin piston is used to apply external pressure to the system and to control the volume of simulation box,\(^ {62}\) allowing for fluctuations of the system density.

The vibrational frequencies of the covalent bonds with hydrogen atoms limit the simulation time steps to 1 fs. In order to increase this time step, the SHAKE\(^ {63}\) algorithm is applied during the simulations, which fixes the length of bonds with hydrogen atoms through a holonomic constraint and allows for a two-fold increase of the time step length.\(^ {64}\) The equations of motion are modified and the Lagrangian takes the extended form:
\[ -\frac{dU}{dx} - \sum \lambda_i \frac{d\sigma_i}{dx} = m \frac{d^2x}{dt^2}, \]  

(1.10)

where \( \lambda \) corresponds to Lagrange multipliers, and \( \sigma \) is the geometric constraint: \( \sigma_i(q) = (\zeta(q) - \zeta_i) \) with the parameter \( \zeta_i \) being set to a constant value during the simulation.

The ensemble average of \( A \) can be formally computed as follows:

\[
\langle A \rangle_{\text{ensemble}} = \frac{\int Ae^{-\frac{E}{k_B T}} dq}{\int e^{-\frac{E}{k_B T}} dq},
\]

(1.11)

where \( k_B \) is Boltzmann constant and \( T \) is temperature. The denominator in Eq. 1.11 equals the partition function \( Z \), the integral of the Boltzmann factor going over all positions and momenta.

Direct calculation of \( Z \) is impossible in MD simulations, because the integral is over all of space. According to the ergodic hypothesis, one can instead use time-averaging and calculate the average value of the property of interest over all frames. This is based on the assumption that in the long simulations all possible states will be visited and ensemble average equals time average. \( Z \) is fundamental in the calculations of thermodynamics quantities such as energy, entropy, pressure, and free energy. For example, the free energy is given by:

\[
f = -k_B T \ln Z.
\]

(1.12)

In free energy calculations, to avoid direct calculations of partition function, the relative free energy or free energy differences are calculated from the ratio of partition functions. Details will be presented when discussing conformational free energy.

1.5. Force Field

Calculation of the system energy in MD simulations is based on the summation of multiple terms representing interactions between atoms or particles, called the force field. In the CHARMM force fields, these terms are given by:
\[ U_{\text{Bonded}} = \sum_{\text{Bonds}} k_b (b - b_i)^2 + \sum_{\text{Angles}} k_\theta (\theta - \theta_i)^2 + \] 
\[ \sum_{\text{Dihedrals}} k_d [1 + \cos(n \varphi - \delta)] + \sum_{\text{Improper}} k_\gamma (\gamma - \gamma_i)^2 + \sum_{\text{Urey–Bradley}} k_{ab} (s - s_i)^2 + \sum_{\text{CMAP}} f(\varphi, \psi), \]

Each term of the force field corresponds to the specific type of interaction, and can be divided into the bonded and non-bonded terms. Bonded terms are calculated between atoms or particles that are connected through covalent bonds or angles, while non-bonded terms correspond to the interactions between atoms separated by at least four interaction sites or three bonds. In CHARMM the bond and angle terms are characterized by harmonic potentials representing vibrations, dihedral terms account for the periodic character of torsion angles and are given by a Fourier expansion, and improper dihedrals control out of plane bending.\(^{65-67}\) CHARMM is one of the very few force fields that possesses the Urey-Bradley term, which improves calculated vibrational spectra. The CMAP\(^{69}\) term is also specific for CHARMM force field and accounts for improved protein secondary structure by parameterization of cross terms between protein dihedral angles \(\varphi\) and \(\psi\):

\[ f(\varphi, \psi) = \sum_{j=1}^{4} \sum_{j=1}^{4} c_{ij} \left( \frac{\varphi - \varphi_{\text{ref}}}{\Delta \varphi} \right)^{\frac{1}{2}} \left( \frac{\psi - \psi_{\text{ref}}}{\Delta \psi} \right)^{\frac{1}{2}}. \] (1.14)

The non-bonded force field term consists of the van der Waals interaction, modeled by a Lennard-Jones potential, and the electrostatic interaction according to Coulomb's law. In the van der Waals non-bonded term, the \(\epsilon\) parameter corresponds to the minimum of the potential between two atoms at equilibrium distance \(R_{ij}\), while \(r_{ij}\) is a distance at given step; \(q_i\) and \(q_j\) are the atomic charges and \(\varepsilon_0\) is the dielectric constant. Equilibrium values of the parameters are calculated from quantum mechanics, gas phase structures, and X-ray crystallography, and force constants often come from \textit{ab initio} quantum calculations, IR or Raman spectroscopy.\(^{70}\) Calculation of the non-
bonded interactions is the most costly step of MD. To avoid counting for the interactions that have a small contribution to the overall energy of the system, pre defined distances (cutoffs) are introduced; beyond these the non-bonded interactions are not considered. This approximation introduces discontinuities into the energy. To circumvent numerical instabilities SWITCH or SHIFT functions are used at the sites of applied cutoffs,\textsuperscript{71} and the electrostatic energies is multiplied by the value of these functions. In case of SWITCH function, minimum and maximum boundaries ($R_{on}$ and $R_{off}$) define a range of distances within which the electrostatic energy is modified according to:

\[
SWITCH(r_{ij}, R_{on}, R_{off}) = \begin{cases} 
1 & r_{ij} < R_{on} \\
\frac{(R_{off} - r_{ij})^2 (R_{off} + 2r_{ij} - 3R_{on})}{(R_{off} - R_{on})^2} & R_{on} < r_{ij} < R_{off} \\
0 & r_{ij} > R_{off}
\end{cases}
\]

(1.15)

While the SWITCH function adjusts the non-bonded interactions only within the $R_{on}$ and $R_{off}$ range, the SHIFT function affects the energy at all distances shorter than $R_{cutoff}$:

\[
SHIFT(r_{ij}, R_{cutoff}) = \begin{cases} 
1 - \frac{2r_{ij}^2 + r_{ij}^4}{R_{cutoff}^2 R_{cutoff}} & r_{ij} < R_{cutoff} \\
0 & r_{ij} > R_{cutoff}
\end{cases}
\]

(1.16)

Since SHIFT function introduces lower perturbation to the forces, it is more frequently applied than SWITCH.

1.6. Treatment of Solvent

Definition of the solvent is of high importance for accurate representation of the system and depending on the systems of interest, either explicit or implicit solvation models are employed.
When the solvation effects and interactions with individual or groups of water molecules are of greater importance than the cost of calculations, the solvent is explicitly represented in the simulations. A straightforward and efficient rigid TIP3 model is often used in MD simulations and was applied in the DNA studies discussed of chapters 3 – 5. As the name suggests, the three sites assigned in TIP3 correspond to the water oxygen and hydrogen atoms. Bond lengths are fixed, and an angle of 104.5° is applied to mimic the imperfect tetrahedral geometry of water molecules. The point charges are assigned to each atom and in the non-bonded interactions are calculated between oxygen (in CHARMM also hydrogen) atoms of neighboring water molecules. Such representation of water molecules is multiplied in space to generate the simulation box. To limit computational costs, the simulation box has a limited size, which does not correspond to true bulk; the particles close to the box boundary may experience different effect of neighboring atoms in comparison to those in the box center. As a result, periodic boundary conditions (PBC) are applied to count for the effect of bulk water. In PBC the original cell is replicated in space through its images; if a particle crosses the original cell boundary, it takes a space in one of the cell images. Proper settings of PBC in simulations are essential for keeping the number of particles in the main cell unchanged.

Electrostatic interactions between each pair of atoms are by far the most expensive part of energy computation and require efficient algorithms. The most common method, the Ewald summation, more accurately estimates long-range interactions in periodic systems than the application of cutoffs. In Ewald summation, short- and the long-range contributions are considered in the long-ranged electrostatic interactions; the first term is calculated in real space, the latter with a Fourier transform.  

\[ \phi(x) = \phi_{\text{rec}}(x) + \phi_{\text{dir}}(x), \]  

(1.17)
These calculations scale with $N^2$ (N-number of particles), which does not reduce the computational bottleneck, but allows for mimicking the bulk phase. A more efficient solution is the particle mesh Ewald approach that scales with $N \log N$ where one the Ewald summation terms is calculated on the grid with the particle mesh Ewald (PME) approach using Fast Fourier transform. In the work presented here each atom has a fixed-point charge. However, current applications move towards polarizable force fields where the polarizability of each atom is evolving in response to environment dynamics. Examples include Drude oscillators and AMOEBA potentials.

In the implicit solvation models atomistic representation of the solute and implicit representation of the solvent allow for quicker calculations, while retaining accuracy. With such approach, the limitations in the time scales can be overcome. The total energy of the system is defined as the sum of the energy of the molecule in vacuum and a free energy term corresponding to the effect of placing the molecule in the solvent ($\Delta G_{solv}$). The latter term corresponds to solute-solvent and solvent-solvent interactions:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{nonp},$$  \hspace{1cm} (1.18)

where $\Delta G_{elec}$ is the free energy related to removing the charges from vacuum and placing them in the solvent, and $\Delta G_{nonp}$ corresponds to hydrophobic effect. $\Delta G_{nonp}$ is proportional to the solvent accessible surface area. Calculation of $\Delta G_{elec}$ forms a computational bottleneck due to the long-range electrostatic calculations that must be counted for. This difficulty leads to several approximations in the calculations of free energy of solvation in computer simulations. Accurate but expensive computations of the electrostatic interactions are based on the Poisson equation, which accurately treats the continuum electrostatics in the system:

$$\nabla[\epsilon(x)\nabla \psi(x)] = -4\pi \rho_{solute}(x),$$  \hspace{1cm} (1.19)
where $\varepsilon(x)$ is dielectric constant depending on position, $\psi(x)$ is the electrostatic potential at given point $x$, and $\rho_{solute}$ is a solute constant charge density. With ions present, the interactions between them significantly increase the cost of calculations. This was overcome by Poisson-Boltzmann approximation (PB), where the density of ions is represented by a Boltzmann distribution. The accuracy of the method was preserved nonetheless the method was still very costly. Another additional approximation in order to apply PB in computer simulations more efficiently, was to linearize the exponential related to concentration of ions under the approximation that concentration of ions is low; this produced the linearized PB version applied in Generalized Born model$^{83}$ (GB). GB is one of the most accurate approximations commonly used in the CHARMM program. In GB the specific analytical form is assigned to the dielectric boundary and the free energy of solvation is given by:

$$
\Delta G_{elec} = \frac{-k}{\varepsilon_{solute} - \varepsilon_{solvent}} \sum_{i,j} \frac{q_i q_2}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{-\frac{r_{ij}}{\varepsilon \alpha_i \alpha_j}}}}.
$$

(1.20)

where $\varepsilon$ – dielectric constant, $q$ – electrostatic charge, $r$ – distance between atoms $i$ and $j$, and $\alpha$ is the Born radius. Calculation of the Born radius for every atom is the main difficulty and varies through available GB methods.

In chapter 2, the Fast Analytical Continuum Treatment of Solvation method (FACTS)$^{84}$ was used to estimate $\Delta G_{elec}$. This method analytically estimates the volume and geometrical symmetry of the solvent displaced by neighboring atoms and allows for fast and accurate radii calculations in the protein systems. Although there are no dielectric boundaries nor a Coulomb field, the agreement with GB methods is good and speeds up of around 10 times are achieved.$^{84}$ The formula to calculate the electrostatic solvation energy with FACTS is given by:
\[ \Delta G_{\text{elec}}^i = a_0 + \frac{a_1}{1 + e^{-a_2(C_i - a_3)}} \quad (1.21) \]

\[ C_i = A_i + b_1 B_i + b_2 A_i B_i. \]

where \( C_i \) is a function of sigmoidal shape and accounts for the displacement of a solvent around solvated particle \( i \); \( a_0, a_1, a_2 \) and \( a_3 \) parameters are determined by consideration of exposed or buried atom.\(^{84} \) \( A_i \) and \( B_i \) are related to the volume and symmetry of the space occupied by each atom, which efficiently replaces Born radii calculations.

### 1.7 Conformational Free Energy Calculations

The calculation of free energy can provide very useful information about the conformational transitions in molecules. As discussed earlier, the calculation of partition function in MD simulations is a hard task and instead, the differences in free energy can be obtained from the ratio of partition functions between two states. Such approach leads to the ensemble average for which statistics are collected in the simulations:

\[ \Delta f = -kT \ln \frac{Z_1}{Z_0} = -kT \ln \frac{\int e^{-U_i/\beta} dq}{\int e^{-U_0/\beta} dq} = -kT \ln \left\langle e^{-\beta U} \right\rangle. \quad (1.22) \]

The ratio of partition functions can be rewritten as a function of the probability distribution or histograms as follows:

\[ \frac{Z_1}{Z_0} = \int e^{-\beta U} d\Delta U \int \delta(\Delta U - \Delta U(q)) e^{-U_i/\beta} dq = \int e^{-\beta U} P_0(\Delta U) d\Delta U. \quad (1.23) \]

If, however, there is no or poor overlap between these two states, additional simulations need to be introduced in order to enhance the sampling of the conformational space. Standard MD simulations have provided important insights into the dynamics of many bio-molecules however,
if the energy barriers associated with particular transitions that separate conformational states are on order of $10 k_B T$, they will not be easily sampled in current MD with timescales of $100 \text{ ns} - 1 \mu\text{s}$\textsuperscript{85}. Hence, various enhanced sampling methods have been developed to help overcome the sampling problems. The choice of the technique depends on the system properties and question at hand. For example, sampling of the conformational space of the inhibitory module of ETS-1 (chapter 2) was improved by employment of the replica exchange MD\textsuperscript{86} simulations (REMD), where multiple replicas of the systems are simulated in parallel at different temperatures. Another example is the umbrella sampling method\textsuperscript{87} (US), which allows for enhanced sampling of a preselected reaction coordinate by adding a biasing potential to the force field. This method was applied to DNA systems studied in chapters 3 – 5. REMD and US methods are discussed in more details below.

### 1.7.1. Replica Exchange Molecular Dynamics

To enhance sampling of the conformational space, in REMD multiple copies of the same system are run in parallel. Each replica is in the canonical ensemble but has a different temperature. At predefined simulation steps, the coordinates of the replicas are swapped, allowing the low-temperature and high-temperature systems to mix; as a result the sampling of the space is greatly improved. The swapping of replicas follows the acceptance rule according to the detailed balance condition:

$$\frac{\text{acc}(i,\beta_i),(j,\beta_j) \rightarrow (i,\beta_i),(j,\beta_j)}{\text{acc}(i,\beta_j),(j,\beta_i) \rightarrow (i,\beta_i),(j,\beta_j)} = e^{\beta_j U(j) - \beta_i U(i)} e^{-\beta_j U(j) + \beta_i U(i)},$$

(1.24)

where $\beta_i$ corresponds to $1/k_B T_i$, $U_i$ is the potential energy. In detailed balance each move from one state (replica) to another should be balanced by the opposite move to maintain equilibrium.
While the low temperature replicas are usually trapped in local minima, the high temperature replicas easily go over the barriers. Swapping of the coordinates allows for the efficient sampling of the conformational space. An important condition for successful swapping is sufficient overlap of the energy distribution of the replicas otherwise the probabilities of exchange will be zero. In REMD simulations of protein folding, the range of temperatures is distributed around the melting temperature, and usually replicas are more closely spaced near this temperature. Since the number of replicas scales with the square root of the number of degrees of freedom in the system, REMD is often applied in implicit solvent simulations.

1.7.2. Umbrella Sampling Molecular Dynamics

The US method is another enhanced sampling approach. In addition to the normal unbiased potential, a biasing potential is applied along the reaction coordinate that closely corresponds to the property of interest. This facilitates transitions between states. US can be performed in a one or multi-dimensional manner. The biasing potential drives the system to sample the pre-selected parameter over the area of significance but generates a non-Boltzmann distribution. In US, the energy of the system \( U \):

\[
U = U_0 + W_{\text{bias}},
\]

is modified by adding the biasing potential \( W_{\text{bias}} \) to the normal potential \( U_0 \). \( W_{\text{bias}} \) is most commonly in the harmonic form:

\[
W_{\text{bias}}^{i} = \sum_{i} k(\theta - \theta^{'})^2,
\]

where \( \theta \) is the reaction coordinate which can be a distance, angle, dihedral, or, in case of chapter 3, DNA step parameter. In the simulations \( \theta \) fluctuates around pre-defined value \( \theta^{'}) \) and the distribution spread is controlled by force constant \( k \). Multiple windows are constructed for a
particular parameter in such way that the range of interest is covered with sufficient overlap between adjacent windows. Unlike REMD each of these can be running independently from one another, which is a great computational advantage. Two-dimensional (2D) US simulations are of the interest for the parameters that typically correlate (as twist and roll DNA step parameters in chapter 4) and may provide more detailed insights into the conformational preferences of the system.

Since the sampling in US simulations is enhanced through the additional biasing potential, the average distribution must be reweighted. Reweighting of property $A$ can be performed as follows:

$$
\langle A \rangle = \frac{\langle Ae^{\beta W_{bias}} \rangle}{\langle e^{\beta W_{bias}} \rangle},
$$

(1.27)

To decrease error of the calculations, statistics from all simulations are combined during the reweighting process using weighted histogram analysis method\textsuperscript{54} (WHAM) discussed next.

### 1.7.3. Weighted Histogram Analysis Method

The time average of the property of interest from the biased simulations can be expressed as follows:

$$
\langle A \rangle = \frac{\int A e^{-\beta U} dq}{Z_{bias}} = \frac{\int A e^{-\beta U_0} e^{-\beta W_{bias}} dq}{Z_{bias}} = \frac{\langle Ae^{\beta W_{bias}} \rangle Z_0}{Z_{bias}},
$$

(1.28)

that leads to the calculation of reweighted probability distribution; in practice this calculation is based on constructing the histograms $h_i$ using statistics from all simulations:

$$
P(q) = e^{\beta W_{bias}(q)} \frac{h_i(q) Z_{bias}}{n_i Z_0},
$$

(1.29)
This estimation of probability density may provide a weak statistical estimate because the large part of the distribution may come from the area poorly sampled by the reaction coordinate. To avoid that, multiple simulations with either biasing potentials or range of temperatures are performed. The reweighting is then performed with WHAM where the estimate of \( P_0^{\text{est}} \) is calculated from linear combination of these multiple statistics using a weight function \( w_i \) minimizing the estimated variance of \( P_0^{\text{est}} \):

\[
P_0^{\text{est}} = \sum_{i=1}^{n} w_i e^{\beta w_i} P_i \frac{Z_i}{Z_0}.
\]  

(1.30)

The free energy differences are then calculated using the ratio of partition functions given by:

\[
\frac{Z_i}{Z_0} = \int e^{-\beta W_i} dq \frac{\sum_{j=1}^{n} h_j}{\sum_{k=1}^{n} e^{-\beta W_k} M_k Z_0 \frac{Z_k}{Z_0}}.
\]  

(1.31)

Since the ratio of partition functions is calculated instead of the absolute value, first \( Z_i \) (e.g. \( Z_i \)), is fixed to a constant value\(^{54} \) and the free energy difference between the states 1 and \( n \) is equals:

\[
\Delta f = -kT \ln \frac{Z_n}{Z_1}.
\]  

(1.32)

1.8. Time Scales and Calculations Efficiency in the Simulations

In MD, the equations of motion are numerically integrated using 1 – 2 fs time steps corresponding to the shortest vibration times of the molecule and in order to achieve energy preservation. Since the conformational rearrangements of proteins and/or DNA often take place at timescales reaching \( \mu \)s to ms, efficient, parallel calculations are a must. In this work the efficiency has gained by:

1) Application of SHAKE\(^{63} \) algorithm discussed earlier which allows for longer time steps;

2) Employment of implicit solvent models, where the efficiency highly depends on the system
size. Such approach was applied in REMD simulations of inhibitory module in ETS-1 in chapter 2 where FACTS implicit solvent model was used.\textsuperscript{90} It is important to underline that employment of enhanced sampling methods themselves (REMD, US) had the purpose of reducing the computational cost in comparison to standard unbiased simulations in the first place;

3) Parallel computing with CHARMM program instead of serial jobs running, where parts of the system are simulated using multiple CPUs (Central Processing Units or processors); all simulations discussed in this work run using parallel versions of CHARMM program;

4) Additional enhancement in parallelization is by the introduction of domain decomposition (domdec) parallelization.\textsuperscript{91} In domdec the simulation box is divided into multiple smaller boxes, each with a CPU assigned to it. This approach was applied in chapters 4 and 5;

5) Application of Graphics Processing Units (GPUs) instead of standard CPUs.\textsuperscript{92} One GPU may consist of hundreds of cores while each CPU can only be made of several cores. If combined, CPU and GPU can provide the most efficient solution to the performance up to date. This approach has been applied to certain simulations of the systems studied in chapters 4 and 5; standard parallel as well as domdec version of CHARMM can be simulated this way;

6) Use of Open Molecular Mechanics (OpenMM) which is a highly scalable toolkit for molecular dynamics simulations and available freely.\textsuperscript{93} The design of the algorithms in OpenMM allows for very fast force calculations and efficient integration, particularly if combined with GPUs.\textsuperscript{92} This approach was applied in chapter 4.
CHAPTER TWO:

IMPORTANCE OF LOCAL INTERACTIONS FOR THE STABILITY OF INHIBITORY HELIX 1 IN APO ETS-1

Note to Reader

Reprinted and adapted with permission from Aleksandra Karolak and Arjan van der Vaart Biophysical Chemistry. See Appendix A.

2.1 Abstract

Inhibitory helix 1 (HI-1) of the Ets-1 human transcription factor unfolds upon binding the target DNA sequence. To identify the interactions that stabilize HI-1 in the apo state, we performed replica exchange and molecular dynamics simulations of various apo Ets-1 constructs. The simulations indicate the importance of local interactions for the stability of HI-1. The HI-2 and H4 helices stabilize the helical state of HI-1 through specific residue-residue contacts and macrodipolar interactions. The amount of stabilization in small length HI-1+H2 and HI-1+H4 constructs was similar to that in the protein. The studies suggest that the partial unfolding of Ets-1 upon DNA binding can be achieved by the removal of just a few specific local contacts.
2.2 Introduction

The human Ets-1 transcription factor is important for embryonic development,\textsuperscript{94} apoptosis,\textsuperscript{95} and angiogenesis\textsuperscript{96} in normal and pathological growth. Ets-1 is also involved in cancer metastasis and tumor progression. High expression levels in breast,\textsuperscript{97,98} ovary,\textsuperscript{99-101} and cervix\textsuperscript{102} tumors correlate strongly with bad prognosis, while elevated expression is relevant for lung,\textsuperscript{103} colon,\textsuperscript{104} pancreatic,\textsuperscript{105,106} thyroid,\textsuperscript{106,107} and oral\textsuperscript{108} cancers. In addition, Ets-1 plays a role in immunity and autoimmune diseases.\textsuperscript{109,110} The protein consists of six domains.\textsuperscript{111} The N-terminal domain contains a RAS-responsive phosphorylation site,\textsuperscript{112,113} which regulates the transcriptional activity of Ets-1. This domain is followed by the pointed domain, important for protein-protein interactions,\textsuperscript{114} the transactivation domain, important for transcription activation,\textsuperscript{115} and the D, ETS, and F domains which regulate DNA-binding.\textsuperscript{111,116}

DNA is bound by a winged helix-turn-helix motif in the ETS-domain (residues 331-415).\textsuperscript{111} This highly conserved domain binds the GGAA/T sequence in the major groove of purine-rich DNA by insertion of the recognition helix (H3). The minor groove is bound by a loop between $\beta$-sheets S3 and S4 and the turn between $\alpha$-helices H2 and H3. The binding affinity for an auto-inhibitory module, which flanks the ETS domain and decreases the binding affinity of DNA 10 to 20 fold, modulates DNA compared to the bare ETS domain.\textsuperscript{117} The auto-inhibitory module consists of residues 301-330 of the D domain and residues 415-440 of the F domain.\textsuperscript{118-120} These residues are folded into four $\alpha$-helices: inhibitory helix 1 and 2 of the D domain (HI-1 and HI-2, respectively), and H4 and H5 of the F domain. The DNA-binding affinity is further regulated by calcium-dependent phosphorylation of an unstructured serine-rich region of the D domain (residues 243-300),\textsuperscript{121} and by binding of protein partners like run-related transcription factor.\textsuperscript{116}
Autoinhibition is achieved by a highly unusual mechanism that involves the unfolding of HI-1.\textsuperscript{111,116,120,122} Previous studies in our laboratory have shown that the unfolding is due to a change in correlated motions between H4 and HI-1.\textsuperscript{123} In the apo protein, HI-1 and H4 move in a correlated (in-phase) fashion, and HI-1 is stabilized by hydrogen bonding and macrodipolar interactions with H4. In the DNA bound state, the motion between HI-1 and H4 is anti-correlated (out-of-phase), which disrupts the macrodipolar and hydrogen bonding interactions. Computer simulations have shown that the change in correlated motions is due to hydrogen bonding between the amide backbone of Leu337 of H1 and the DNA.\textsuperscript{123} This hydrogen bond was shown to act as a conformational switch in biochemical experiments as well.\textsuperscript{124,125} Simulations showed that the conformational switch transfers the information that DNA is present to HI-1 by a network of correlated motions between H1, H4 and HI-1.\textsuperscript{126}

Given the central role of HI-1 in the autoinhibition mechanism of Ets-1, we performed computer simulations to further investigate how HI-1 is stabilized in the apo state. Structural analyses show that HI-1 is loosely packed in the protein, making contacts with HI-2, H5 and H4 only. Moreover, experiments showed that HI-1 is marginally stable in the apo protein, and conformationally dynamic on the milli to microsecond time scale.\textsuperscript{124} Taken with our previous simulation data, these observations suggest that HI-1 is mostly stabilized by a few local contacts. The aim of our study was to establish the identity of these contacts, and to quantify which contribute most to the stabilization.

\textbf{2.3 Materials and Methods}

Since no full-length structure of the Ets-1 protein exists, we used the NMR structure of the apo construct Δ301 (Ets-1 residues 301-440; Protein Data Bank entry 1R36,\textsuperscript{124} Figure 2.1) to
generate all of our initial coordinates. This construct contains the ETS domain and the autoinhibitory module. Biochemical experiments have shown that the binding behavior of Δ301 is similar to the full-length protein, with unfolding of HI-1 upon specific DNA binding.\textsuperscript{120,124,127} Several constructs were simulated: HI-1 (residues 301-314, Figure 2.2a), HI-1+HI-2 (residues 301-334, Figure 2.2b), HI-1+H4 (residues 304-310 and 418-422, Figure 2c), and the Δ301 construct. HI-1 and H4 were fused in the HI-1+H4 construct by a GT linker in order to mimic the nearly continuous HI-1-H4 helix that is observed in Δ301.\textsuperscript{124} The HI-1+H4 construct was simulated in two ways: a simulation in which H4 was restrained to be α-helical, and a simulation in which no such restraints were used. These harmonic restraints were applied to the backbone of H4 with a force constant of 1.01 kcal/(mol Å\textsuperscript{2}). Several mutant constructs were simulated as well.

\textbf{Figure 2.1.} Structure of apo ETS-1.
Figure 2.2. Systems studied with REX. (a) HI-1. (b) HI-1 + HI-2. (c) HI-1 + H4. HI-1 is shown in dark grey, and the arrow points to the GT linker (see text).

The simulations were performed using the CHARMM 19\textsuperscript{128} force field and FACTS implicit solvent model\textsuperscript{90} as implemented in the CHARMM program.\textsuperscript{90} Langevin dynamics with a time step of 2 fs was used, and SHAKE\textsuperscript{129} was applied to constrain bonds involving hydrogen atoms. To fully sample the conformational space, all constructs except Δ301 were simulated with temperature replica exchange (REX).\textsuperscript{86} In this method, identical copies of the system (replicas) are run at different temperatures. By frequently swapping the configurations based on a criterion that maintains detailed balance, much more configurational space can be sampled than in normal molecular dynamics (MD) simulations while maintaining thermodynamic equilibrium. Each REX simulation used 16 replicas at 200, 209, 219, 230, 241, 252, 264, 276, 289, 303, 317, 333, 348, 365, 382, 400 K. Selection of these temperatures was based on the calculated folding temperature (based on the calculated heat capacity of trial runs), and chosen to well cover the folded state and folding/unfolding transition and also to ensure good exchange between replicas. Since the simulations are performed in an implicit solvent model using a force field parameterized at room temperature, the folding temperature does not necessarily correspond to a physiological temperature. Coordinate swaps were attempted every 5 ps, and a total simulation time of 250 ns was used per replica. Since we were interested in the folding/unfolding of HI-1, and not in the folding/unfolding of the entire protein, REX could not be used for Δ301. Instead,
we used four independent normal, unbiased MD runs for Δ301; each of 150 ns production length. All REX and MD systems were first heated for 500 ps with weak harmonic restraints on the backbone, followed by 500 ps of equilibration during which the harmonic restraints were gradually removed. For the fused HI-1+H4 system in which H4 was restrained to be helical, only the restraints on HI-1 were gradually removed.

The secondary structure assignment was determined using program STRIDE, and the output was used for the helical fraction (HF) calculations. HI-1 was considered folded when HF>0.5. Free energy surfaces at 300 K were calculated using the weighted histogram analysis method (WHAM) using the data from all replicas. These surfaces were based on the HF; surfaces for root mean square deviation, number of native contacts and radius of gyration gave similar results. WORDOM was used for contact analysis. Dipoles were constructed using the N, H, CA, HA, C and O backbone atoms, and dipole-dipole interactions were calculated using standard methods. All figures were generated using VMD.

2.4 Results

Replica exchange simulations of the HI-1 construct showed a significant preference for the unfolded state, with the free energy of the unfolded state 0.74 kcal/mol lower than that of the folded state (Table 2.1). The higher stability of the unfolded state was not surprising, since the construct is short (14 residues) and helix formation generally takes more residues. In addition to α-helical structures, the folded state structural ensemble also showed π-helical structures. Prevalent hydrophobic contacts were made between F304 and V308, and R309 with K305 (Figure 2.3a and 2.3b). These contacts are also prevalent in Δ301, and help to stabilize the helix.
Table 2.1. Free energies of the HI-1 folded state relative to the unfolded state.\textsuperscript{a}

<table>
<thead>
<tr>
<th>System</th>
<th>Wild-type</th>
<th>F304V+R309L</th>
<th>F304A+Y307A</th>
<th>F304V+Y307V</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI-1\textsuperscript{b}</td>
<td>0.74±0.07</td>
<td>-0.29±0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HI-1+HI-2\textsuperscript{b}</td>
<td>-0.91±0.07</td>
<td></td>
<td>-0.12±0.05</td>
<td>-0.38±0.07</td>
</tr>
<tr>
<td>HI-1+H4 (unrestr.)\textsuperscript{b,c}</td>
<td>-0.14±0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HI-1+H4 (restr.)\textsuperscript{b,d}</td>
<td>-0.69±0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ301\textsuperscript{c}</td>
<td>-1.00±0.02</td>
<td>-1.89±0.02</td>
<td>-0.69±0.02</td>
<td>-0.60±0.02</td>
</tr>
</tbody>
</table>

\textsuperscript{a} In kcal/mol. The free energies of all unfolded states are at 0.0 kcal/mol; free energies were calculated using the helical fraction as reaction coordinate.
\textsuperscript{b} Standard deviations calculated from WHAM.\textsuperscript{132}
\textsuperscript{c} Using no restraints on H4.
\textsuperscript{d} Using backbone restraints on H4 in order to keep it helical.
\textsuperscript{e} Standard deviations calculated from bootstrapping.\textsuperscript{136}

The HI-1 unfolded state consisted of a large ensemble of partially structured loops. Many of these structures had one helical turn or incomplete helical turns, while partial sheet structures were also observed. Structural analyses of the unfolded state showed the prevalence of a hydrophobic contact between F304 and R309 that did not occur in the folded state (Figure 2.3c).

Figure 2.3. Contacts observed in the folded state: (a) F304 + V308. (b) K305 + R309. In the unfolded state: (c) F304 + R309.
In the unfolded state this contact was mostly observed in structures with residues 306 to 310 in a helical conformation, but also in structures with a turn between residues 304 to 307. In the folded state, F304 and R309 form contacts with V308 and K305, respectively, which stabilize the helix.

To investigate the importance of the F304-R309 contact in stabilizing the unfolded state, we decided to perform simulations of the F304V+R309L double mutant. This mutant was chosen in order to preserve the overall hydrophobicity of the residues, and to retain the overall helical propensity of the construct (the helical propensities of F and V are 0.54 and 0.61, respectively, while the helical propensity of R and L is 0.21\textsuperscript{137}); therefore, the mutation only introduced geometrical effects. The simulations showed that the unfolded state is indeed destabilized in the mutant. Contacts between residues 304 and 309 occurred less frequently in the unfolded state, and the free energy of the unfolded state in terms of the helical fraction was decreased by 1.03 kcal/mol, making the mutant folded state slightly more favorable than the unfolded state. The mutant unfolded state favored partial $\pi$-helical structures over sheet or loop structures. Contacts between residues 304 and 308 were diminished in the folded state of the mutant. In the folded state, F304V was more frequently observed to interact with R309L, stabilizing $\pi$-helical conformers.

The importance of the F304-R309 contact for the unfolded state, and our strategy to stabilize the folded state by destabilizing the unfolded state was verified in simulations of apo Ets-1 \textDelta301. In accordance with experiments, in the wild-type HI-1 was preferentially folded. The free energy of the partially unfolded state of the protein (with HI-1 unfolded) was 1.01 kcal/mol higher than that of the folded state. In the F304V+R309L mutant, the free energy of the unfolded state was increased by 0.89 kcal/mol, and contacts between residues 304 and 309 were less frequently observed in the partially unfolded state.
The presence of HI-2 significantly increased the stability of the folded state of HI-1, with the folded state 0.91 kcal/mol more stable than the unfolded state. In the folded state, significant hydrophobic contacts are made between F304, Y307, and Y329 (Figure 2.4), while in the unfolded state frequent contacts are observed between Y307 and K316, and F304 and K318. To test the importance of the F304/Y307/Y329 contacts for the folded state, simulations were performed on a F304A+Y307A mutant of the HI-1+HI-2 construct, as well as a F304V+Y307V mutant. The latter mutant has similar helical propensities as the wild type (the helical propensities of F, Y, and V are 0.54, 0.53, and 0.61, respectively) and retains the hydrophobic character of the residues, so the mutation only introduced geometrical effects. Simulations of the mutant showed that the folded state was destabilized by 0.79 kcal/mol in the F304A+Y307A mutant, and by 0.53 kcal/mol in the F304V+Y307V mutant, although in both constructs the folded state remained the most favorable. Simulations of the apo Ets-1 Δ301 showed that the folded state became destabilized by 0.31 and 0.41 kcal/mol for F304A+Y307A and F304V+Y307V mutants respectively.

**Figure 2.4.** Contacts between F304/Y307/Y329 in the HI-1+HI-2 construct in the folded state.

While both mutants destabilize the folded state, there are marked differences in the mechanism by which this destabilization is achieved. Consistent with the fact that alanine has the
highest helix propensity of all amino acids, the F304A+Y307A mutant showed more helical structure than the F304V+Y307V mutant. However, due to the small size of Ala, the F304A+Y307A mutant had a diminished ability to interact with other side chains, leading to a decrease in 304/307/329 interactions and an overall destabilization of the F304A+Y307A mutant compared to the wild type. Valine on the other hand, has a larger hydrophobic side chain, and the F304V+Y307V mutant interacted with the other side chains much more frequently than F304A+Y307A. Due to the lack of aromatic side chains however, interactions between residues 304/307/329 were less frequently observed than in the wild type. Therefore, consistent with its lower helical propensity, the helical state of HI-1 was less populated than the wild type. The destabilizing effect of both mutants was lower in Δ301 than in the HI-1+HI-2 constructs. In Δ301 stabilizing interactions between HI-1 and H4 (see below) partly counteracted the mutations, leading to a weaker overall destabilization of the mutants in the protein.

Simulations of the HI-1+H4 construct showed that H4 had a positive effect on the stability of the folded state of HI-1. The folded state became slightly more stable than the unfolded state (by 0.14 kcal/mol). In the construct, the same F304-L421 and K305-L422 hydrogen bonds form as in the apo Δ301. We observed that whenever these hydrogen bonds broke, the alignment between the H4 and HI-1 helices was disrupted, and HI-1 unfolded. The C terminus of HI-1 generally unfolded the fastest, while residues next to the linker remained helical through most of the simulations. From the simulations, we calculated that the likelihood of H4 and HI-1 folded at the same time is six times higher than the likelihood of HI-1 folded while H4 is unfolded. Our results indicate that the equilibrium is shifted towards structures with both helices HI-1 and H4 folded, and that the folded state of H4 stabilizes the folded state of HI-1.
To further verify the influence of H4 on the stability of HI-1 we performed simulations of the HI-1+H4 construct in which the backbone of H4 was restrained to be helical. In these simulations the folded state was significantly more stable than the unfolded state, by 0.69 kcal/mol. The helix macrodipoles of HI-1 and H4 interacted favorably when both helices were folded, with an average interaction energy of -4.76 kcal/mol. Whenever the F304-L421 and K305-L422 hydrogen bonds between the helices broke, the alignment of helices was lost, leading to a loss of macrodipolar interaction (average of -1.10 kcal/mol), and the unfolding of HI-1. These observations confirm the importance of macrodipolar and hydrogen bonding interactions between HI-1 and H4 for the stability of HI-1, and form additional evidence that disruptions of these interactions may lead to the unfolding of HI-1 in the protein.123

2.5 Discussion and Conclusions

Simulations of various Ets-1 HI-1 constructs showed the importance of local interactions for the stabilization of HI-1 in the apo state. The addition of HI-2 or H4 stabilized HI-1 by a similar amount of energy as observed in full-length Δ301. The stabilization by HI-2 results from specific hydrophobic contacts between F304, Y307 of HI-1, and Y329 of HI-2. The importance of these residues is in agreement with experimental mutation studies that showed a large disruption in autoinhibition for F304A and Y307A mutations.124 Of particular interest is the F304V+R309L mutant, which stabilized the folded state of the protein. This stabilization is achieved through a destabilization of the unfolded state, and results from the removal of a favorable hydrophobic interaction between F304 and R309 in the unfolded state. Stabilization of H4 is through macrodipolar and hydrogen bonding interactions with HI-1.
The simulation studies elucidated the factors that stabilize HI-1 in the apo state and help explain the loose packing of the helix in the protein. Our results suggest that the unfolding of Ets-1 upon DNA binding can be achieved by the removal of just a few local contacts, and does not require major rearrangements of the protein. Indeed, the backbone RMSD of Ets-1 residues 318-440 between the trimolecular Ets-1Δ280—Pax-5—DNA complex and apo Ets-1Δ300 was only 0.68 Å;\textsuperscript{124} our data suggests that similar structural agreement will exist for the Ets-1Δ280—DNA complex not stabilized by Pax-5. The simulations provide further support for the central role of H4 for the autoinhibition/unfolding mechanism.\textsuperscript{116,122-124,138}
CHAPTER THREE:
ENHANCED SAMPLING SIMULATIONS OF DNA STEP PARAMETERS

Note to Reader
Reprinted and adapted with permission from Aleksandra Karolak and Arjan van der Vaart Journal of Computational Chemistry. See Appendix B.

3.1 Abstract
A novel approach for the selection of step parameters as reaction coordinates in enhanced sampling simulations of DNA is presented. The method uses three atoms per base and does not require coordinate overlays or idealized base pairs. This allowed for a highly efficient implementation of the calculation of all step parameters and their Cartesian derivatives in molecular dynamics simulations. Good correlation between the calculated and actual twist, roll, tilt, shift and slide parameters is obtained, while the correlation with rise is modest. The method is illustrated by its application to the methylated and unmethylated 5’-CATGTGACGTCACATG-3’ double stranded DNA sequence. One-dimensional umbrella simulations indicate that the flexibility of the central CG step is only marginally affected by methylation.
3.2 Introduction

The mechanical properties of DNA play an important role in protein-DNA binding and sequence recognition. For example, nuclear receptors exploit subtle differences in local DNA flexibility for sequence recognition,\textsuperscript{139,140} while many transcription factors bend DNA in order to modulate their binding affinities.\textsuperscript{10,141} DNA bending is also crucial for its packaging in the cell, notably the packaging of eukaryotic DNA into nucleosomes and higher order structures.\textsuperscript{142,143} Atomistic molecular dynamics (MD) simulations can aid the understanding of these processes by providing valuable insights into the structure, conformational dynamics, and mechanical properties of DNA.\textsuperscript{18,144} While detailed information can be obtained from long unbiased simulations,\textsuperscript{145-147} it is often more efficient to use biasing or enhanced sampling simulations\textsuperscript{148,149} like umbrella sampling.\textsuperscript{87} With enhanced sampling methods a larger amount of the relevant configurational space can be sampled, and free energy surfaces as a function of a low dimensional order parameter can be calculated. These surfaces can be used to help rationalize molecular properties in terms of structure and energetics. In most enhanced sampling methods biasing potentials operate on a predefined reaction coordinate, which is selected to best represent the property of interest. Here we introduce a simplified method to select the DNA step parameters as reaction coordinates in enhanced sampling methods, and illustrate its application by studying the effect of methylation on the flexibility of DNA.

Since the DNA bases are rigid, the orientation of the DNA bases are well described by the DNA step parameters twist, roll, tilt, rise, slide and shift that describe the rigid body translation and rotation of adjacent base pairs (Figure 3.1).\textsuperscript{150-153} Statistical analyses have shown that overall shape of DNA is mostly determined by the step parameters, especially roll and twist.\textsuperscript{151,154} In addition, a set of rigid body translations and rotations can be used to describe the
orientation of individual bases within a base pair.\textsuperscript{150-153} Step parameters have been used to study the effect of sequence,\textsuperscript{155-157} and chemical modifications,\textsuperscript{19} on DNA geometry and flexibility. Given the importance of step parameters on DNA structure, roll and twist-like reaction coordinates have previously been used in umbrella sampling simulations. We have employed a pseudo-roll angle to study DNA bending\textsuperscript{158,159} and coupled protein binding and DNA bending,\textsuperscript{160,161} while Zacharias \textit{et al.} used a twist-like reaction coordinate to study the over and untwisting of various DNA sequences.\textsuperscript{162} Here we introduce a comprehensive approach to enhance the sampling of all step parameters in enhanced sampling simulations. A key aspect of the method is the use of three atoms per base in the definition of the step parameters. This reduced representation significantly simplifies and speeds up the calculation of the derivatives that are needed for the force calculation, and avoids costly overlays that are used in other methods,\textsuperscript{150-153} while retaining high accuracy for twist, roll, tilt, slide and shift and modest accuracy for rise.

\textbf{Figure 3.1.} DNA step parameters. DNA bases are indicated by the rectangles.

The method is illustrated by a study of the effect of methylation on the flexibility of DNA. Methylation of cytosine is an epigenetic modification essential for the regulation of many biological processes,\textsuperscript{25,163-165} and DNA methylation affects nucleosome shape and dynamics. Single molecule FRET,\textsuperscript{166} solid state NMR,\textsuperscript{167} and AFM studies\textsuperscript{168} have indicated that the
nucleosome rigidifies upon DNA methylation. An increase in the stiffness of bare DNA upon methylation was seen in DNA circularization experiments,\textsuperscript{17,18} gel studies,\textsuperscript{169} and FTIR experiments,\textsuperscript{170} and methylation was shown to alter the curvature of DNA in electrophoresis.\textsuperscript{171} Other experimental work showed no effect of methylation on the stiffness of bare DNA,\textsuperscript{168,172} and a slight increase in breathing of the nucleosome upon methylation,\textsuperscript{173} however. Unbiased molecular dynamics simulations showed an increase in bare DNA stiffness upon CG step methylation,\textsuperscript{18,174} but in some studies the effects were marginal and the overall shape of DNA was minimally affected,\textsuperscript{174-176} while \textit{ab initio} calculations predicted larger flexibility of the cytosine base upon methylation.\textsuperscript{177} Here we will illustrate our method by performing one dimensional umbrella sampling simulations of the central CG step of the 5’-CATGTGACCGTCACATG-3’ double stranded DNA sequence in the methylated and unmethylated forms. The structural properties of these systems will be compared and discussed.

### 3.3 Methods

DNA step parameters have long been used for the conformational analysis of DNA.\textsuperscript{150-153,178,179} The standard definition requires a root mean square (rmsd) overlay with an idealized base pair.\textsuperscript{150} In order to avoid the computational cost of performing these overlays at each MD simulation step and to significantly simplify the atomic derivatives of the step parameters, we based our definition on local coordinates. This definition is similar to the FREEHELIX algorithm,\textsuperscript{151} but computationally simpler since it does not require a least squares fit for the vectors normal to the base pair plane. The method is illustrated in Figure 3.2. Indicating the base pairs by \(i\), the coordinates of three purine atoms \(r_{i,1}, r_{i,2}, r_{i,3}\) per base are used to calculate the vector \(\vec{R}_i\) that is normal to the purine base plane:
\[ v_i^R = r_{3,i} - r_{2,i}, \]
\[ w_i^R = \begin{cases} r_{i,i} - r_{2,i}, & \text{if } \{r_{i,i}, r_{2,i}, r_{3,i}\} \in \text{main strand} \\ r_{2,i} - r_{1,i}, & \text{otherwise} \end{cases}, \]
\[ R_i = v_i^R \times w_i^R. \]  

(3.1)

In a similar manner, the coordinates of three pyrimidine atoms \(j_{1,i}, j_{2,i}\) and \(j_{3,i}\) are used to construct the vector normal to the pyrimidine base plane \(Y_i\). The \(R_i\) and \(Y_i\) vectors are used to construct the "perpendicular" vector \(P_i\) that is orthogonal to the base plane (Figure 3.2a):

\[ P_i = \frac{R_i + Y_i}{|R_i + Y_i|}. \]  

(3.2)

Figure 3.2. Local (a) and median (b) vectors for the step parameters calculations illustrated for the GG step.
The "long" vector $L_i$ is placed along the long axis of the base pairs:

$$L_i = \begin{cases} 
\frac{r_{ij} - y_{ij}}{|r_{ij} - y_{ij}|} & \text{if } r_i \in \text{main strand} \\
\frac{y_{ij} - r_{ij}}{|y_{ij} - r_{ij}|} & \text{otherwise}
\end{cases}.$$  \hspace{1cm} (3.3)

The "short" vector $S_i$ is the cross product of $P_i$ and $L_i$:

$$S_i = L_i \times P_i \hspace{1cm} (3.4)$$

$P_i$, $L_i$ and $S_i$ form a local coordinate system for each base pair. These vectors are subsequently used to construct a set of orthogonal median vectors $P_{i,i+1}$, $L_{i,i+1}$, and $S_{i,i+1}$ between two adjacent base pairs (Figure 3.2b):

$$P_{i,i+1} = \frac{P_i + P_{i+1}}{|P_i + P_{i+1}|}$$

$$L_{i,i+1} = \frac{L_i + L_{i+1}}{|L_i + L_{i+1}|}$$

$$S_{i,i+1} = L_{i,i+1} \times P_{i,i+1}.$$  \hspace{1cm} (3.5)

The position of the center points of the base pairs are along the $L_i$ vectors and given by:

$$c_i = \begin{cases} 
\frac{1}{2}(r_{ij} - y_{ij}) & \text{if } r_i \in \text{main strand} \\
\frac{1}{2}(y_{ij} - r_{ij}) & \text{otherwise}
\end{cases}.$$ \hspace{1cm} (3.6)

Within a step, the movement of the base pairs is specified by the shift in position of the center points:

$$\Delta_i = c_{i+1} - c_i.$$  \hspace{1cm} (3.7)
The rise, slide and shift translational step parameters are then defined as the projections of this
movement onto the $L_{i,i+1}$, $P_{i,i+1}$, and $S_{i,i+1}$ vectors:

\[
\begin{align*}
\text{rise}' &= \Delta_i \cdot P_{i,i+1} \\
\text{slide} &= \Delta_i \cdot L_{i,i+1} \\
\text{shift} &= \Delta_i \cdot S_{i,i+1} 
\end{align*}
\] (3.8)

Here rise is indicated by a prime, since its value will be modified below. The twist rotational step
parameter is given by the angle between $L_i$ and $L_{i+1}$ as projected on the plane perpendicular to
$P_{i,i+1}$, while roll and tilt are given by the angle between $P_i$ and $P_{i+1}$ as projected onto planes
perpendicular to $L_{i,i+1}$ and $S_{i,i+1}$ respectively. This calculation requires the cross products of the
local and median vectors. For twist they take the form of:

\[
\begin{align*}
\alpha_{i}^{\text{twist}} &= \frac{L_i \times P_{i,i+1}}{|L_i \times P_{i,i+1}|}, \\
\alpha_{i+1}^{\text{twist}} &= \frac{L_{i+1} \times P_{i,i+1}}{|L_{i+1} \times P_{i,i+1}|}.
\end{align*}
\] (3.9)

While for roll:

\[
\begin{align*}
\alpha_{i}^{\text{roll}} &= \frac{P_i \times L_{i,i+1}}{|P_i \times L_{i,i+1}|}, \\
\alpha_{i+1}^{\text{roll}} &= \frac{P_{i+1} \times L_{i,i+1}}{|P_{i+1} \times L_{i,i+1}|}.
\end{align*}
\] (3.10)

and tilt:

\[
\begin{align*}
\alpha_{i}^{\text{tilt}} &= \frac{P_i \times S_{i,i+1}}{|P_i \times S_{i,i+1}|}, \\
\alpha_{i+1}^{\text{tilt}} &= \frac{P_{i+1} \times S_{i,i+1}}{|P_{i+1} \times S_{i,i+1}|}.
\end{align*}
\] (3.11)
With these cross products, the rotational parameters are given by:

\[
\begin{align*}
\text{twist} &= \cos^{-1}\left(\alpha_{\text{twist}}^\text{twist} \cdot \alpha_{\text{twist}}^\text{twist}\right) \\
\text{roll} &= \cos^{-1}\left(\alpha_{\text{roll}}^\text{roll} \cdot \alpha_{\text{roll}}^\text{roll}\right) \\
\text{tilt} &= \cos^{-1}\left(\alpha_{\text{tilt}}^\text{tilt} \cdot \alpha_{\text{tilt}}^\text{tilt}\right)
\end{align*}
\]  

(3.12)

Table 3.1. List of 235 pdb files used in correlation tests.

<table>
<thead>
<tr>
<th>1ah</th>
<th>1du0</th>
<th>1ig7</th>
<th>1l5u</th>
<th>1nk6</th>
<th>1skw</th>
<th>1y6f</th>
<th>2bq3</th>
<th>2evg</th>
<th>2hot</th>
<th>2r8i</th>
<th>3cmv</th>
<th>3g6u</th>
<th>3hdd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ah</td>
<td>1du0</td>
<td>1ig7</td>
<td>1l5u</td>
<td>1nk6</td>
<td>1skw</td>
<td>1y6f</td>
<td>2bq3</td>
<td>2evg</td>
<td>2hot</td>
<td>2r8i</td>
<td>3cmv</td>
<td>3g6u</td>
<td>3hdd</td>
</tr>
<tr>
<td>1ah</td>
<td>1du0</td>
<td>1ig7</td>
<td>1l5u</td>
<td>1nk6</td>
<td>1skw</td>
<td>1y6f</td>
<td>2bq3</td>
<td>2evg</td>
<td>2hot</td>
<td>2r8i</td>
<td>3cmv</td>
<td>3g6u</td>
<td>3hdd</td>
</tr>
<tr>
<td>1ah</td>
<td>1du0</td>
<td>1ig7</td>
<td>1l5u</td>
<td>1nk6</td>
<td>1skw</td>
<td>1y6f</td>
<td>2bq3</td>
<td>2evg</td>
<td>2hot</td>
<td>2r8i</td>
<td>3cmv</td>
<td>3g6u</td>
<td>3hdd</td>
</tr>
</tbody>
</table>

We based the selection of the three base atoms that define the step parameters on correlations of the calculated step parameters and DNA deformation energies\textsuperscript{155} with those obtained from 3DNA.\textsuperscript{153,179} For this goal, a previously constructed database of DNA structures was used,\textsuperscript{159} from which systems with ill-defined step parameters (such as systems with flipped bases or broken strands), Z-DNA and intercalator-bound DNA were removed. This led to a database of 235 structures (Table 3.1) for which a considerable number of possible atom selections for the step parameter calculations were tested. Representative results are listed in Table 3.2. Based on
these tests, the purine C₈, N₃ and C₆ atoms and pyrimidine C₆, C₂ and C₄ atoms were chosen for \( r_{i,j} - r_{3,i} \) and \( y_{1,i} - y_{3,i} \), respectively (Figure 3.3). The correlation coefficients for this selection were 0.997 for tilt, 0.891 for twist, 0.998 for roll, 0.943 for shift and 0.978 for slide and −0.535 for rise (Table 3.2).

![Figure 3.3. Atom selection for purine and pyrimidine bases.](image)

<table>
<thead>
<tr>
<th>Atoms Selections&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Tilt</th>
<th>Twist</th>
<th>Roll</th>
<th>Shift</th>
<th>Slide</th>
<th>Rise</th>
</tr>
</thead>
<tbody>
<tr>
<td>C8-N3-C6-C6-C2-C4</td>
<td>0.997</td>
<td>0.891</td>
<td>0.998</td>
<td>0.943</td>
<td>0.978</td>
<td>0.535</td>
</tr>
<tr>
<td>C8-N3-C6-C5-C2-N3</td>
<td>0.985</td>
<td>0.551</td>
<td>0.997</td>
<td>0.922</td>
<td>0.891</td>
<td>0.480</td>
</tr>
<tr>
<td>C8-N3-C6-C6-N1-C4</td>
<td>0.992</td>
<td>0.891</td>
<td>0.996</td>
<td>0.943</td>
<td>0.978</td>
<td>0.534</td>
</tr>
<tr>
<td>C8-N3-C6-C5-N1-C4</td>
<td>0.982</td>
<td>0.551</td>
<td>0.996</td>
<td>0.922</td>
<td>0.891</td>
<td>0.479</td>
</tr>
<tr>
<td>C8-N3-C6-C6-N1-N3</td>
<td>0.997</td>
<td>0.891</td>
<td>0.994</td>
<td>0.943</td>
<td>0.978</td>
<td>0.534</td>
</tr>
<tr>
<td>C8-N3-C6-C5-N1-N3</td>
<td>0.987</td>
<td>0.551</td>
<td>0.996</td>
<td>0.922</td>
<td>0.891</td>
<td>0.480</td>
</tr>
<tr>
<td>C8-N3-N1-C6-C2-C4</td>
<td>0.997</td>
<td>0.891</td>
<td>0.998</td>
<td>0.943</td>
<td>0.978</td>
<td>0.534</td>
</tr>
<tr>
<td>N7-N3-C6-C6-N3-C2</td>
<td>0.988</td>
<td>0.249</td>
<td>0.997</td>
<td>0.884</td>
<td>0.440</td>
<td>0.425</td>
</tr>
<tr>
<td>N7-N3-C6-N1-C4-C6</td>
<td>0.985</td>
<td>0.249</td>
<td>0.996</td>
<td>0.884</td>
<td>0.440</td>
<td>0.426</td>
</tr>
<tr>
<td>N7-N3-C6-N1-C4-C5</td>
<td>0.990</td>
<td>0.782</td>
<td>0.997</td>
<td>0.954</td>
<td>0.915</td>
<td>0.377</td>
</tr>
<tr>
<td>N7-C4-C5-C6-C2-C4</td>
<td>0.984</td>
<td>0.249</td>
<td>0.997</td>
<td>0.883</td>
<td>0.440</td>
<td>0.424</td>
</tr>
<tr>
<td>N7-C4-C5-C6-C4-C2</td>
<td>0.986</td>
<td>0.782</td>
<td>0.998</td>
<td>0.954</td>
<td>0.915</td>
<td>0.376</td>
</tr>
<tr>
<td>N7-C4-C5-C2-N3-C6</td>
<td>0.985</td>
<td>0.249</td>
<td>0.996</td>
<td>0.884</td>
<td>0.440</td>
<td>0.424</td>
</tr>
</tbody>
</table>

<sup>a</sup> In the order \( r_{1,i} \), \( r_{2,i} \), \( r_{3,i} \), \( y_{1,i} \), \( y_{2,i} \) and \( y_{3,i} \) respectively.

Table 3.2. Correlation of DNA step parameters with 3DNA for representative sets of atomic selections. Examples of high correlations are shown on top, low in the bottom.
The correlation of rise with 3DNA values was improved by using a fitting function:

\[
\text{rise} = 4.97 - 0.50 \cdot \text{rise}'
\]  

(3.13)

After the fitting function was applied, the correlation between values of rise obtained with the current method and 3DNA became 0.535.

While correlations for tilt, twist, roll, shift and slide were in good agreement with 3DNA, the correlation for rise was modest. The discrepancy in the rise values presented here arose from subtle differences between our and 3DNA's definition of the step parameters. While we used the purine C\(_8\) and the pyrimidine C\(_6\), 3DNA uses the sugar C\(_1'\) atoms. Since there is a rotatable bond between the sugar and base, small differences between our and the 3DNA values may occur, especially for the \(L_r\) and \(L_{r+1}\) vectors. These differences will be amplified by the use of idealized base pairs in 3DNA. While one could in principle use an atom selection for rise that is different from the atom selection of the other parameters, tests showed that this did not lead to increased correlations. A large variance in calculated rise values across various methods has been observed before\(^\text{180}\) and fitting functions were also used to match the rise from other programs with 3DNA.\(^\text{180}\) Since the simulations can always be reweighted to 3DNA's rise values, the modest correlation with rise is not a problem in practice. Moreover, rise is arguably the least interesting step parameter for enhanced sampling simulations, since its value does not vary much across straight, bent or distorted DNA (Figure 3.4). For example, in the database of 235 DNA structures, 3DNA rise values varied between 2.67 and 3.99 Å with only one data point over 4 Å (4.47 Å), and on average the rise component of the DNA deformation energy contributed only 8.2 ± 3.4% to its overall value (Figure 3.5). Correlation plots for the calculated DNA step parameters and 3DNA are shown in Figure 3.4.
Figure 3.4. Correlation of calculated step parameters with 3DNA values. Correlation coefficients are reported in the text.

Figure 3.5. Contribution of rise to the overall deformation energy calculated for a database of 235 DNA structures.
To enhance the sampling of the DNA step parameters in MD, a harmonic umbrella potential in the form of:

$$E_{\text{bias}} = \frac{k}{2}(\theta - \theta_{\text{desired}})^2$$  \hspace{1cm} (3.14)

was applied, where $k$ is a force constant, $\theta$ corresponds to the instantaneous step parameter value in the simulation and $\theta_{\text{desired}}$ is the desired step parameter value. The analytical forms of the step parameter derivatives with respect to atomic coordinates, which are needed for the force calculation, were obtained from Mathematica,\textsuperscript{181} and verified by finite difference methods. These are far from trivial, but not presented here due to their complex and lengthy forms. The calculation of the step parameters, umbrella potentials and their derivatives were implemented in the CHARMM simulation program,\textsuperscript{182} tested for accuracy, and optimized for speed and memory consumption. Analysis of the trajectories was performed with WORDOM,\textsuperscript{133} and VMD.\textsuperscript{134} Stiffness force constants (FCs) were calculated by applying Mathematica's nonlinear (quadratic) fitting function to the parabolic regions of free energy minima.

### 3.4 Simulation Setup

Umbrella sampling simulations were performed on the central, underlined base pair step of the 5’-CATGTGACGTCACATG-3’ double stranded DNA sequence, Protein Data Bank entry 1SAA.\textsuperscript{53,183} Simulations were performed for the unmethylated and methylated system; for the latter, the C\textsubscript{5} position on the cytosine of the central CG step was methylated on both strands. The DNA strands were solvated in a rectangular box of 150 mM NaCl solution of TIP3 water,\textsuperscript{184} with a minimum distance of 12 Å between DNA and the edge of the box. After minimization the systems were gradually heated from 120 K to 300 K over a period of 1 ns and equilibrated for
1.4 ns. During heating and equilibration, harmonic restraints with a force constant of 1 kcal mol$^{-1}$ Å$^{-2}$ were used on the heavy atoms to maintain the structure of DNA close to the starting value. These restraints were subsequently released in steps of 500 ps each, during which the force constants were decreased from 1 to 0.5, to 0.1 kcal mol$^{-1}$ Å$^{-2}$. A subsequent final unrestrained equilibration was performed for 1 ns. After equilibration, one-dimensional umbrella sampling simulation of the twist, roll, tilt, shift, slide and rise parameters of the central base step were performed. Each umbrella window was simulated for 5 ns, using a force constant of 1 kcal mol$^{-1}$ deg$^{-2}$ for the rotational and 1 kcal mol$^{-1}$ Å$^{-2}$ for the translational step parameters. The umbrella windows were distributed between $+10^\circ$ and $+60^\circ$ for twist, $-20^\circ$ and $+40^\circ$ for roll, $-25^\circ$ and $+25^\circ$ for tilt, $-2.5$ Å and $+2.5$ Å for shift, $-2.5$ Å and $+3.5$ Å for slide, and $+1.5$ Å and $+4.5$ Å for rise, using a window at every 2.5° for twist, roll and tilt, every 0.5 Å for shift, slide and rise. This setup resulted in a total of 98 simulations for each unmethylated and methylated system, for a total production time of 980 ns. Overlap of distributions was checked visually and free energy surfaces were calculated using the weighted histogram analysis method,$^{131,185,186}$ using a bin size of 0.1 Å for translational and 1° for rotational step parameters. Reweighting was also performed with respect to the 3DNA step parameter values. Error bars and convergence were assessed from block averaging. All simulations were performed with the CHARMM program,$^{182}$ using the CHARMM 36 force field.$^{187}$ A time step of 2 fs was used, SHAKE was applied to constrain the bonds with hydrogen atoms,$^{63}$ the temperature was controlled with the Nosé-Hoover thermostat,$^{188}$ and long-range electrostatic interactions were handled by the particle mesh Ewald method.$^{74}$
3.5 Results

Umbrella sampling simulations of the central CG step of the methylated and unmethylated 5’-CATGTGACGTCACATG-3’ DNA duplex were performed. Values of the step parameters calculated by the introduced method were compared to 3DNA for each of the trajectories. Table 3.3 shows the average values and standard deviations for representative unmethylated DNA simulations; similar values were obtained for methylated DNA. In general, the calculated twist, roll, tilt, shift and slide step parameters agreed well with 3DNA. Average values of twist, roll and tilt closely correspond to the desired values, and generally closely matched the 3DNA values. Significant differences with 3DNA only occurred when DNA was severely under or overtwisted, or at very large tilt values. For example, the average deviation with 3DNA was about 2.5° for a twist of 10°, 4.2° for a twist of 50°, 12.7° for a twist of 60°, and 4.2° for a tilt of 25°, while for all other twist and tilt values the average deviation was ~1° or less. To illustrate the rarity of such twist and tilt values, in the database of 235 structures with a total of 2769 base pair steps, no steps had twist values ≤ 10°, 2 steps had twist values ≥ 50°, no steps had twist values ≥ 60°, and no steps had tilt values ≥ 25°. These deviations stemmed from the differences in atom selection when calculating step parameters. The biasing potential acts on three atoms of each base, but since the bases are rigid, all base atoms will be pulled along. In addition to the base atoms, 3DNA also uses the C1’ sugar atoms in calculating the step parameters. This atom is not subjected to the biasing potential, and since it is not part of the base, it will adjust to distortions introduced by the biasing potential. As a result the twist and tilt values in the 3DNA calculations may vary somewhat from our values, especially when very large distortions are introduced. The shift and slide parameters corresponded well with 3DNA values. Values closer to desired could have been obtained by using larger force constants for the biasing
potential; since well-converged free energy surfaces were obtained, this was not pursued here. As expected, rise showed modest agreement with 3DNA. Good agreement was obtained for rise values above 3.0 Å, but below this value poor agreement was obtained.

Figure 3.6. Convergence of the free energy as a function of the central twist angle: after 2 ns (grey) and 5 ns (black) of simulation per umbrella window. Error bars were obtained from block analyses. Twist values as calculated by the introduced method (Eq. 3.12); results are shown for the unmethylated system.

Convergence of the free energy simulations with the central twist angle as reaction coordinate is shown for the unmethylated system in Figure 3.6; analogous plots were obtained for the other step parameters. The twist simulations converged within 0.062 kcal/mol after 2 ns of sampling per window, and within 0.033 kcal/mol after 5 ns/window. Figure 3.7 shows a comparison of the free energy surfaces of the unmethylated system. Black dashed curves show the surfaces with respect to the step parameters calculated from Eq. 3.8, 3.12, and 3.13, while the grey curves show the surfaces with respect to 3DNA step parameter values. In agreement with Table 3.3, the curves of shift, roll, slide and tilt largely overlap, while small deviations up to 1 kcal/mol are observed for overtwisted and undertwisted DNA. As expected from the poorer
correlation between calculated and 3DNA rise values, the free energy curve for rise shows larger deviations.

**Figure 3.7.** DNA conformational free energy as a function of the central step parameter. For the black dashed curves step parameter values were calculated by the introduced method (Eq. 3.8, 3.12-3.13); for the grey curves by 3DNA. Results are shown for the unmethylated system; error bars were obtained from block analyses.

A comparison of the free energy curves of the unmethylated and methylated systems as a function of the step parameters calculated from Eq. 3.8, 3.12-3.13 is shown in Figure 3.8. To ease the comparison, error bars are not shown; these were similar for the methylated and unmethylated system (shown in Figure 3.8). The locations of the free energy minima are reported in Table 3.4. The free energy minimum occurred at a twist value of $+38.5^\circ$ for the unmethylated and at $+41.5^\circ$ for the methylated system.
Table 3.3. Average step parameters in umbrella sampling simulations of unmethylated DNA calculated by the introduced method and 3DNA.

<table>
<thead>
<tr>
<th>Step parameter</th>
<th>$\theta_{\text{desired}}$</th>
<th>Calculated</th>
<th>3DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>St. Dev.</td>
</tr>
<tr>
<td>Twist (deg)</td>
<td>10.00</td>
<td>10.06</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>20.00</td>
<td>20.09</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>30.00</td>
<td>30.10</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>40.00</td>
<td>40.06</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>50.00</td>
<td>49.53</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>60.00</td>
<td>59.13</td>
<td>0.76</td>
</tr>
<tr>
<td>Roll (deg)</td>
<td>-20.00</td>
<td>-19.79</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>-10.00</td>
<td>-9.80</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.14</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>10.00</td>
<td>10.02</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>20.00</td>
<td>19.91</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>30.00</td>
<td>29.80</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>40.00</td>
<td>39.84</td>
<td>0.77</td>
</tr>
<tr>
<td>Tilt (deg)</td>
<td>-25.00</td>
<td>-24.59</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>-15.00</td>
<td>-14.75</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>-5.00</td>
<td>-4.87</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>5.00</td>
<td>4.97</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>15.00</td>
<td>14.79</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>25.00</td>
<td>24.71</td>
<td>0.76</td>
</tr>
<tr>
<td>Shift (Å)</td>
<td>-2.50</td>
<td>-0.94</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>-1.50</td>
<td>-0.65</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>-0.50</td>
<td>-0.16</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.27</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>0.54</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>0.92</td>
<td>0.43</td>
</tr>
<tr>
<td>Slide (Å)</td>
<td>-2.50</td>
<td>-0.80</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>-1.50</td>
<td>-0.16</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>-0.50</td>
<td>0.21</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.60</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>0.99</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>1.12</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>3.50</td>
<td>1.52</td>
<td>0.60</td>
</tr>
<tr>
<td>Rise (Å)</td>
<td>1.50</td>
<td>1.77</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.95</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>2.96</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>3.09</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>3.50</td>
<td>3.30</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>4.41</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>4.50</td>
<td>4.67</td>
<td>0.45</td>
</tr>
</tbody>
</table>
For twist, the free energy curve is not symmetric around the minimum, and the untwisting of DNA is less costly than overtwisting. This effect has been observed by others as well, but is generally not included in elastic models of DNA deformation. The free energy surfaces of roll, tilt, slide shift are symmetric with respect to each minimum, while the free energy surface for rise is unsymmetrical.

Universal trends in the behavior of twist and roll step parameters, where the increase in twist is accompanied by the decrease in roll value and vice versa, is reflected in the values of the twist and roll free energy minima (Table 3.4). In the methylated system the cytosine methyl groups are on the same side of the DNA. Due to hydrophobic interactions, this leads to a slight increase in twist, which is accompanied by a slight decrease in roll. While our simulations showed a slight increase in twist, unbiased MD simulations of different DNA strands using the AMBER force field saw a slight reduction of twist upon methylation, while a crystallographic study saw a slight increase in twist. Table 3.4 shows that the equilibrium tilt, shift, slide and rise values also changed upon methylation.

Methylation increases slide and rise, and decreases tilt and shift, but most of the changes are subtle. Methylation preserves the positive correlation between tilt and shift (simultaneous increase or decrease in values) that was observed for unmethylated DNA sequences in the literature. Coupling of twist and roll to translational step parameters is not commonly discussed in the literature, but we observed a negative correlation between roll and slide and a positive correlation between twist and slide. In principle, correlations could be systematically studied by extending our umbrella sampling simulations into two dimensions, but due to the high cost, this analysis is beyond the scope of the current study.
Figure 3.8. DNA conformational free energy as a function of the central step parameter. Unmethylated system shown by dashed black line, methylated system by continued grey line, and step parameters as calculated by the introduced method (Eq. 3.8, 3.12-3.13). The curves were constructed using 5 ns of simulation time per window.

Table 3.4. Values of the step parameters at the free energy minima.

<table>
<thead>
<tr>
<th></th>
<th>Twist (deg)</th>
<th>Roll (deg)</th>
<th>Tilt (deg)</th>
<th>Shift (Å)</th>
<th>Slide (Å)</th>
<th>Rise (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmethylated</td>
<td>38.50</td>
<td>11.50</td>
<td>1.50</td>
<td>0.15</td>
<td>0.25</td>
<td>3.35</td>
</tr>
<tr>
<td>Methylated</td>
<td>41.50</td>
<td>10.50</td>
<td>-1.50</td>
<td>0.05</td>
<td>0.75</td>
<td>3.45</td>
</tr>
</tbody>
</table>

The width of the free energy curves around the minima is slightly broader for unmethylated DNA, suggesting slightly higher flexibility of the unmethylated system. The difference is very subtle, especially for roll and tilt, but consistent with the smaller values of the FCs for the unmethylated systems (Table 3.5). These FCs were obtained by fitting the free energy surfaces to quadratic functions, and the values for the unmethylated systems are similar to literature values (obtained either from DNA structures or unbiased molecular dynamics).
simulations\textsuperscript{157,162}). Taken together, the free energy surfaces of Figure 3.8 and the FCs of Table 3.5 suggest that there is hardly any difference in DNA flexibility or mechanical properties between the methylated and unmethylated CG steps at equilibrium. Differences could possibly occur at high bending angles, for example in the nucleosome, when various step parameters for multiple DNA steps have values that differ greatly from their equilibrium values. In such instances, the slightly higher FCs of methylated DNA may make the methylated DNA observably less flexible than unmethylated DNA.

Table 3.5. Stiffness constants of the step parameters (in kcal mol\(^{-1}\) deg\(^{-2}\) for rotational and kcal mol\(^{-1}\) Å\(^{-2}\) for translational).

<table>
<thead>
<tr>
<th></th>
<th>FC(_{\text{twist}})</th>
<th>FC(_{\text{roll}})</th>
<th>FC(_{\text{tilt}})</th>
<th>FC(_{\text{shift}})</th>
<th>FC(_{\text{slide}})</th>
<th>FC(_{\text{rise}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmethylated</td>
<td>0.034</td>
<td>0.010</td>
<td>0.016</td>
<td>1.616</td>
<td>1.254</td>
<td>13.572</td>
</tr>
<tr>
<td>Methylated</td>
<td>0.042</td>
<td>0.012</td>
<td>0.017</td>
<td>1.774</td>
<td>1.268</td>
<td>14.228</td>
</tr>
</tbody>
</table>

The distance between the two cytosine C\(_5\) atoms of the central DNA step was calculated to test if the introduction of methyl groups leads to any spatial rearrangements around the bases. Frames were extracted from the twist biasing trajectories around each free energy minimum (twist of 38.5° ± 0.1° for the unmethylated strand, and 41.5° ± 0.1° for the methylated strand), and the average C\(_5\)–C\(_5\) distance was calculated. This average was 7.50 ± 0.43 Å for the unmethylated, and 6.92 ± 0.40 Å for the methylated system. The result indicates that methyl groups move slightly away from the DNA backbone and favor positions where C\(_5\) atoms on the central cytosines are closer to each other. This effect is likely due to the increased hydrophobic character of the methylated bases, but steric effects might play a role as well. As a consequence (and due to the rigidity of the cytosine base), the configuration of the entire base changes subtly, and so do the step parameters.
3.6 Method Refinement

In order to improve correlations between calculated and 3DNA twist and rise parameters, a known mathematical optimization procedure was employed. The simplex algorithm introduces constraints and iteratively operates on the set of variables used to calculate the parameter of interest (twist, rise) in such a way that the correlation with a comparison set (3DNA) is maximized. To apply the simplex algorithm and calculate refined values of twist and rise, several modifications in definitions of these step parameters were introduced. First, instead of original form of Eq. 3.3, the new definition of $L_i$ vector along the long axis of the base pairs is given by:

$$L_i = \begin{cases} \frac{r_i^{cent} - y_i^{cent}}{y_i^{cent} - r_i^{cent}} & \text{if } r \in \text{main strand} \\ \frac{r_i^{cent} - y_i^{cent}}{y_i^{cent} - r_i^{cent}} & \text{otherwise} \end{cases},$$

where $r_i^{cent}$ and $y_i^{cent}$ are the weighted center points of each base calculated using:

$$r_i^{cent} = q_{1,i}r_{1,i} + q_{2,i}r_{2,i} + q_{3,i}r_{3,i},$$
$$y_i^{cent} = q_{1,i}y_{1,i} + q_{2,i}y_{2,i} + q_{3,i}y_{3,i},$$

From the Eq. 3.15 the $L_i$ vector is calculated between center points of purine and pyrimidine base ($r_i^{cent}$ and $y_i^{cent}$) instead of atoms C8 on purine and C6 on pyrimidine as in Eq. 3.3. In addition, the position of each of the three atoms on each base does not contribute to the center point equally but is weighted by the constraints $q_{1,i}^R, q_{2,i}^R, q_{3,i}^R, q_{1,i}^Y, q_{2,i}^Y, q_{3,i}^Y$ provided by simplex optimization algorithm. Values of these constraints applied in twist calculations are 1.96, -1.51, -0.44, 2.66, -2.47, -0.18, in rise calculations are -0.44, 0.15, 1.32, -1.36, 0.90, 1.43 for atoms $r_{1,i}, r_{2,i}, r_{3,i}, y_{1,i}, y_{2,i}$ and $y_{3,i}$ respectively. All other calculations remain the same. The resulting
correlations of these step parameters with 3DNA are 0.998 and 0.997 for twist and rise as illustrated on Figure 3.9.

![Figure 3.9.](image)

**Figure 3.9.** Refined correlation of calculated step parameters twist and rise with 3DNA values. Correlation coefficients are reported in the text.

### 3.7 Discussion and Conclusions

We introduced a simplified method to select DNA step parameters in enhanced sampling simulations and illustrated its application by investigating the effect of methylation on DNA flexibility. The method uses three atoms per base in calculating the step parameters and its Cartesian derivatives, and does not require coordinate overlays or idealized base pairs. In general, excellent correlations between calculated twist, roll, tilt, shift and slide and 3DNA values were obtained, while the correlation with rise was modest. In umbrella sampling simulations, deviations with 3DNA values only occurred for rise, for severely under or over twisted DNA and for very large tilt values. These deficiencies were due to the exclusion of the C1' atom and absence of idealized base pairs in calculating the step parameters, and can be mitigated by calculating the free energy surfaces as a function of the 3DNA step parameter values from biasing simulations in the simplified coordinate. Overall, the method is highly efficient for use in molecular dynamics simulations.

Applications to an unmethylated and methylated DNA strand showed minor changes in DNA conformation and stiffness upon methylation, and suggest that the mechanical properties
are not changed upon methylation unless the DNA is severely bent. The main contribution to these subtle changes comes from the hydrophobic effect, which favors positions that bring the C5 atoms of the central cytosines closer together upon methylation. Negative twist-roll and positive tilt-shift correlations upon methylation were observed, as well as anharmonicity in twist and rise in the methylated and unmethylated strands.
CHAPTER FOUR:
BII STABILITY AND BASE STEP FLEXIBILITY OF N$_6$-ADENINE METHYLATED GATC MOTIFS

4.1 Abstract

The effect of N$_6$-adenine methylation on the flexibility and shape of palindromic GATC sequences has been investigated by molecular dynamics simulations. Variations in DNA backbone geometry were observed, which were dependent on the degree of methylation and the identity of the bases. While the effect was small, more frequent BI to BII conversions were observed in the GA step of hemimethylated DNA. The increased BII population of the hemimethylated system positively correlated with increased stacking interactions between methylated adenine and guanine, while stacking interactions decreased at the TC step for the fully methylated strand. The flexibility of the AT and TC steps was marginally affected by methylation, in a fashion that was correlated with stacking interactions. The facilitated BI to BII conversion in hemimethylated strands might be of importance for SeqA selectivity and binding.

4.2 Introduction

A common epigenetic modification in prokaryotes and lower eukaryotes is N$_6$-methylation of adenine, which is important for transcription regulation, replication and repair.26-
In *Escherichia Coli*, DNA adenine methyltransferase (Dam) methylates the N\textsubscript{6} position of adenine in GATC sequences.\textsuperscript{191-193} Eleven of the \textasciitilde19,000 GATC sites are clustered at the *E. Coli* replication origin (*oriC*).\textsuperscript{194} While the adenines in most GATC sites are normally fully methylated (FMe), after replication and before re-methylation by Dam DNA exists in the hemimethylated form (HMe).\textsuperscript{27,195} This HMe form is recognized by SeqA, which negatively regulates replication initiation and ensures that replication occurs only once during the cell cycle by binding HMe GATC at *oriC*.\textsuperscript{27,196,197}

Why SeqA preferentially binds HMe DNA has been studied in some detail. In addition to using electrostatic and hydrophobic interactions, DNA-binding proteins exploit subtle changes in local conformation and flexibility to recognize specific DNA sequences.\textsuperscript{8,9,14,198,199} These properties are also used to identify epigenetic modifications, since methylation increases local DNA hydrophobic character and may also affect the local conformation and elasticity of DNA.\textsuperscript{17,200} In particular, N\textsubscript{6}-adenine methylation decreases the melting temperature of DNA,\textsuperscript{200,201} may modulate DNA curvature,\textsuperscript{202} and induces undertwisting of the AT and overtwisting of the TC step in FMe and HMe GATC motifs.\textsuperscript{201,203} It also affects DNA hydration,\textsuperscript{204} and stabilizes the BI conformation in TA repeats.\textsuperscript{205} Structural studies of the SeqA-HMe complex showed that the only sequence specific contacts present are between the protein and the AT base pair within the GATC site.\textsuperscript{206} Other studies implied the existence of additional factors that contribute to binding specificity, since the mutation of C or G in the CG base pair of the GATC motif negatively influenced recognition.\textsuperscript{207} NMR studies of unbound HMe DNA showed a compression of the major groove around the site of methylation,\textsuperscript{201} which was not observed in unbound FMe or UMe DNA;\textsuperscript{203} this compression was similar to that observed in the SeqA-HMe complex.\textsuperscript{206} Moreover, in the SeqA-HMe complex, the GC base pairs of the GATC
motif are slightly opened, and unbound HMe DNA displayed faster base-pair opening and closing rates of these base pairs than unbound FMe DNA. While the difference in estimated barrier for opening between HMe and FMe is small (1.4±0.4 kcal/mol), the observation that somewhat less energy is required for the opening of the GC base pairs in HMe than FMe suggests that this opening might be another driving force in the selective recognition of HMe.

Here we used molecular dynamics (MD) simulations to further investigate the structural and mechanical properties of the GATC motif of unbound UMe, HMe, and FMe DNA. Our study focused on two aspects. The first was the DNA backbone configuration, since crystal structures of SeqA-DNA complexes indicate the occurrence of the BII conformation at the AT step of the unmethylated chain and the GA step of the methylated chain. The DNA BI and BII form describe the relative position of the O3' atom, which points towards the outside of the helix in BI and towards the inside of the helix in the BII form, and are defined with respect to the ε (C4'-C3'-O3'-P) and ζ (C3'-O3'-P-O5') dihedral angles. BI is present when ε − ζ < 0 and BII when ε − ζ > 0. BI is the common configuration for B-DNA, but BII is important for DNA-protein recognition, since BII enhances exposure of DNA bases in the major groove. The second aspect of our study focused on the flexibility of DNA, which is important for indirect readout. Given that the overall shape of DNA is mostly determined by roll and twist angles, we focused on the flexibility of roll and twist at the AT and TC steps, and determined how this flexibility changed upon methylation. We also confirmed the effect of methylation on GC base pair opening by free energy simulations. The sequence of the simulated strands is shown in Figure 4.1; in HMe A_{18} of the complementary strand is methylated.
4.3 Results and Discussion

The BII populations of all steps of the central GATC motif of the UMe, HMe and FMe strands are shown in Table 4.1. Both the GA and TC steps showed significant amounts of BII conformations, while the AT step was almost purely BI for all chains. Steps involving two pyrimidines are thought to visit the BII conformation rarely, but here the TC step of all systems populated the BII form by more than 10%, with a slightly higher percentage in the UMe and FMe systems on the complementary strand. The most significant variations in BII populations were observed at the GA step, where the HMe system showed the largest BII population on the methylated strand. This population for the methylated strand was 7% higher than UMe and FMe, and 6% higher than in the unmethylated HMe strand.

<table>
<thead>
<tr>
<th>Step</th>
<th>UMe</th>
<th>HMe unmethylated</th>
<th>HMe methylated</th>
<th>FMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>27.5±3.7</td>
<td>28.8±5.9</td>
<td>34.5±3.9</td>
<td>27.6±4.5</td>
</tr>
<tr>
<td>AT</td>
<td>1.4±0.3</td>
<td>2.6±1.4</td>
<td>1.4±0.4</td>
<td>1.6±0.8</td>
</tr>
<tr>
<td>TC</td>
<td>14.5±1.6</td>
<td>11.8±3.7</td>
<td>11.1±2.5</td>
<td>14.4±1.2</td>
</tr>
</tbody>
</table>

Averaged values over both DNA strands.

Further analyses revealed a correlation between the BI/BII populations and stacking interactions. In Figure 4.2 the distance between A₁₈ and G₁₇ is graphed versus the $e - \zeta$ value of
the G_{17}A_{18} step. For UMe this distance corresponds to the H_6 atom of A_{18} and the 5-membered ring of G_{17}, while for HMe and FMe the distance is between the carbon methyl atom of A_{18} and the 5-membered ring of G_{17}. The figure shows that the BII conformation is accompanied by much tighter stacking in both HMe and FMe. While the A_{18}-G_{17} distance fluctuates strongly in the BII conformation of UMe, the distance is locked to smaller values in BII of HMe and FMe. The strongest stacking interactions in BII are observed in HMe, which helps explain that the largest fraction of BII was observed for the GA step of HMe.

![Figure 4.2](image)

**Figure 4.2.** Correlation between stacking interactions and $\varepsilon - \zeta$ value for UMe (a), HMe (b) and FMe (c). Distance is measured between the carbon atom of methyl group on A_{18} (hydrogen in UMe) and the 5-membered ring of G_{17}.

In the BI form the A_{18}-G_{17} distance increases in FMe (Figure 4.2c), which is due to hydrophobic interactions between the methyl groups of A_6 and A_{18}. Our data indicates that hydrophobic interactions might also be responsible for the low occurrence of BII at the AT step. Structural analyses showed that interactions between the methyl groups of mA_6 and mA_{18} and the methyl groups of the adjacent thymine within the AT step, weaken stacking interactions in the BII form. The thymine and adenine methyl groups are both located in the major groove, and the decrease of TA stacking within the step might disfavor the BII transition.
The occurrences of mixed BI/BII states for the A\textsubscript{6}T\textsubscript{7}/A\textsubscript{18}T\textsubscript{19} and T\textsubscript{7}C\textsubscript{8}/G\textsubscript{17}A\textsubscript{18} steps are listed in Table 4.2. Since the AT steps sampled BII only marginally, BI/BI was predominant and there was a low fraction of BI/BII and BII/BI. At this step, twist values were around 30°. The populations of BI/BII and BII/BI increased for all systems at the T\textsubscript{7}C\textsubscript{8}/G\textsubscript{17}A\textsubscript{18} step, which was accompanied by elevated values of twist in all systems (41° for UMe, 42° for HMe, 40° for FMe). A correlation between twist and mixed state populations has been observed in other sequences before.\textsuperscript{19,201,212} HMe had the highest value of twist and the highest population of mixed BI/BII states (39%) at the T\textsubscript{7}C\textsubscript{8}/G\textsubscript{17}A\textsubscript{18} step.

Table 4.2. Populations (%) of mixed BI and BII states on DNA main and complementary strands at A\textsubscript{6}T\textsubscript{7}/A\textsubscript{18}T\textsubscript{19} and T\textsubscript{7}C\textsubscript{8}/G\textsubscript{17}A\textsubscript{18} steps.

<table>
<thead>
<tr>
<th>BI / BI</th>
<th>BI / BII</th>
<th>BII / BI</th>
<th>BII / BII</th>
</tr>
</thead>
<tbody>
<tr>
<td>A\textsubscript{6}T\textsubscript{7} / A\textsubscript{18}T\textsubscript{19}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMe</td>
<td>98</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HMe</td>
<td>97</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>FMe</td>
<td>95</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>T\textsubscript{7}C\textsubscript{8} / G\textsubscript{17}A\textsubscript{18}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMe</td>
<td>68</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>HMe</td>
<td>56</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>FMe</td>
<td>63</td>
<td>25</td>
<td>6</td>
</tr>
</tbody>
</table>

In addition to the correlation with twist, the BII population at the T\textsubscript{7}C\textsubscript{8}/G\textsubscript{17}A\textsubscript{18} steps was strongly correlated with slide (Figure 4.3). More weakly positive correlations were observed with shift, and to a lesser degree tilt, while the correlation with roll was weak and negative. Whereas slide for UMe system was ~0.23 Å in BI and ~0.24 Å in BII form, methylation of adenine altered slide in a strongly correlated fashion. The higher the degree of methylation, the lower the value of slide in BI form and the higher its value in BII. For FMe, this led to an average negative slide
(-0.09 Å) in the BI form, and large positive (0.93 Å) in BII. Since nonoptimal values of slide may lead to the loss of stacking interactions, this change in slide may help explain the decreased stability of FMe.

![Figure 4.3](image)

**Figure 4.3.** Correlation between slide and twist and and ε – ζ values at the T7C8/G17A18 steps for UMe (a), HMe (b) and FMe (c).

One-dimensional umbrella sampling simulations of the N4–O6 distance at base pair C8/G17 confirmed that base pair opening is slightly less energy costly for HMe. The free energy of opening is shown as a function of distance in Figure 4.4. None of the curves is symmetric around the minima, because of steric clashes at close proximity of the bases. In HMe the equilibrium N4–O6 distance is slightly higher (3.05 Å) than in UMe and FMe (2.95 Å). Moreover, while the cost of opening is similar in UMe and FMe, this energy is nearly halved in HMe. The similar behavior of FMe and UMe implies that presence of second methyl group counteracts the
effect of hemimethylation. Although the effect is subtle, HMe will sample opening distances close to that observed in the SeqA-DNA complex more frequently than UMe and FMe.

**Figure 4.4.** Free energy as a function of the N4-O6 distance in the C8/G17 base pair for UMe (solid line), HMe (dashed line), and FMe (dotted line).

Two-dimensional free energy surfaces as a function of twist and roll for the A6T7/A18T19 and T7C8/G17A18 steps are shown in Figure 4.5; the twist-roll covariance matrices are shown in Table 4.3. The free energy surfaces show undertwisting of the A6T7/A18T19 step and overtwisting of the T7C8/G17A18 steps. The basins stretch along the anti-diagonal direction; this negative correlation between changes in twist and roll is commonly observed in DNA.18,151,155,162 The HMe A6T7/A18T19 step is slightly more flexible in twist and roll than UMe and FMe, while FMe is somewhat more flexible in twist and stiffer in roll than UMe. The roll flexibility of the T7C8/G17A18 step strongly correlates with G17A18 stacking (Figure 4.2). The highest flexibility is observed for UMe, which had the least amount of G17A18 stacking, while HMe had the largest stacking interactions and is least flexible. In a similar manner, the twist flexibility at the T7C8/G17A18 step is also correlated with the stacking flexibility, but this correlation is less
pronounced than for roll. The methylated systems show slightly stiffer twists than the unmethylated systems, and the twist of HMe is the least flexible.

**Figure 4.5.** Free energy surfaces (in kcal mol$^{-1}$) calculated as a function of twist and roll at $A_6T_7/A_{18}T_{19}$ step for UMe (a), HMe (b) and FMe (c) and for $T_7C_8/G_{17}A_{18}$ step for UMe (d), HMe (e) and FMe (f).

**Table 4.3.** Twist-roll covariance matrices from data points up to 4 kcal/(mol deg$^2$) on the free energy surfaces.

<table>
<thead>
<tr>
<th>Step</th>
<th>UMe</th>
<th>HMe</th>
<th>FMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_6T_7/A_{18}T_{19}$</td>
<td>56 -24</td>
<td>62 -46</td>
<td>43 -24</td>
</tr>
<tr>
<td></td>
<td>-24 92</td>
<td>-46 128</td>
<td>-24 102</td>
</tr>
<tr>
<td>$T_7C_8/G_{17}A_{18}$</td>
<td>51 -25</td>
<td>44 -17</td>
<td>46 -17</td>
</tr>
<tr>
<td></td>
<td>-25 115</td>
<td>-17 82</td>
<td>-17 99</td>
</tr>
</tbody>
</table>
4.4 Conclusions

In conclusion, MD simulations of UMe, HMe, and FMe GATC sequences showed differences in BI/BII populations. The methylated strand of HMe showed larger BII populations of the GA step than UMe and FMe. The small increase in BII population was correlated with an increase in stacking interactions in HMe. In FMe, a large decrease in slide was observed when the TC step was BI and a large increase in slide when it was BII, indicating a decrease of stacking interactions. Given that the GA step of the methylated strands is in BII form in SeqA-DNA complexes, the small difference in BII populations might be exploited by SeqA to facilitate the recognition of the hemimethylated strand. Simulations confirmed that the opening of the GC base pair is more facile in HMe, and showed slight differences in the ease of twist and roll deformations at the AT and TC step of the UMe, HMe, and FMe systems. These differences in stiffness were well correlated with relative stacking interactions.

4.5 Computational Methods

MD simulations of 5’-GCGAGATCTGCG-3’ double stranded DNA were performed, with adenine in the central GATC site in unmethylated and N$_6$-methylated forms. Both HMe, with methylation of the complementary strand, and FMe, with methylation of both the main and complementary strands were simulated. Initial coordinates for the strands were obtained from Protein Data Bank (PDB), entries 1OPQ (UMe), 1UAB (HMe), and 2KAL (FMe). In the HMe and FMe PDB files and the simulations, the N$_6$-adenine methyl groups were oriented trans to the adenine N$_1$ atom, this configuration also corresponds to that observed in SeqA-DNA complexes. The DNA strands were solvated in a rectangular box of 150 mM NaCl solution of TIP3 water, with a minimum distance of 12 Å between DNA and the edge of the
box. After minimization the systems were gradually heated from 120 K to 300 K over a period of 1 ns and equilibrated for 1.5 ns. During heating and equilibration, harmonic restraints with a force constant (FC) of 1 kcal mol$^{-1}$ Å$^{-2}$ were used on the DNA heavy atoms. These restraints were subsequently released in steps of 500 ps each, using FCs of 0.5 and 0.1 kcal mol$^{-1}$ Å$^{-2}$. Production runs started after a final unrestrained equilibration of 1 ns. These production runs consisted of 100 ns normal, unbiased MD; a total of three independent unbiased MD simulations starting from different random seeds for the heating were performed per system. In addition, one and two-dimensional umbrella sampling simulations were performed. In one-dimensional umbrella sampling simulations, the N$_4$−O$_6$ distance between the C$_8$/G$_{17}$ base pair were restrained to values between 1.5 and 5 Å, using windows of 0.5 Å and a FC of 1 kcal mol$^{-1}$ Å$^{-2}$; each window was simulated for 5 ns. Two-dimensional umbrella sampling simulations were performed of the twist and roll angles of the central A$_6$T$_7$/A$_{18}$T$_{19}$ and T$_7$C$_8$/G$_{17}$A$_{18}$ base steps. In these simulations the base step parameter calculations are similar to those done previously for roll with a highly efficient approach to obtain analytical derivatives. The simulations were performed in 99 windows that were distributed between 20° and 60° for twist and −20° and 30° for roll, using a step size of 5° and a FC of 0.5 kcal mol$^{-1}$ deg$^{-2}$. After restrained heating and 1ns equilibration each window was sampled for 2 ns. Overlap of distributions in the umbrella sampling simulations was verified visually, and free energy surfaces were calculated using the weighted histogram analysis method$^{131,185}$ using a bin size of 0.1 Å for the N$_4$−O$_6$ distance and 1° for twist and roll. Total production times were 0.9 µs for the unbiased MD, 120 ns for the one-dimensional and 1.2 µs for the two-dimensional umbrella sampling simulations. All simulations were performed with the CHARMM program$^{182}$ and the CHARMM 36 force field$^{187,214}$ which was optimized to reproduce BI/BII population. The simulations used a time step of 2 fs, SHAKE
to constrain bonds with hydrogen atoms, the Nosé-Hoover thermostat for temperature control, and the particle mesh Ewald method for long-range electrostatic interactions. Trajectories were analyzed with VMD, PyMol, 3DNA and CHARMM, error analyses were performed by block averaging, and free energy surfaces were obtained with Mathematica.
5.1 Abstract

A high level of oxidative damage of cytosine has long been related to various types of disease. The observed lower repair efficiency in comparison to oxidative damage of other DNA bases suggested conformational similarity of DNA with undamaged and damaged cytosine. The relatively stable conformation of oxidized cytosine mismatched with adenine revealed its additional mutagenic character; nevertheless the structural rearrangements of DNA upon hydroxylation have remained elusive. We used molecular dynamics simulations to study the influence of cytosine hydroxylation on DNA conformation in normal and adenine mismatched base pairs in DNA sequence identified in prostate cancer cell lines. Results uncovered high spatial similarity between undamaged and damaged DNA systems without mismatch on the level of base pair and base pair step parameters, which might plausibly mitigate the recognition and repair. Comparable geometry of DNA backbone and sugar puckering in these systems supports the idea. On the contrary, the systems containing cytosine/adenine mismatch show disrupted hydrogen-bonding patterns, which are correlated with the shifts in values of base pair parameters and DNA backbone geometry. Still, the stacking interactions of mismatching adenine with the adjacent bases allowed maintain the overall stability of DNA sequence, which is higher for damaged-mismatched than mismatched DNA.
5.2 Introduction

The oxidative damage on carbon 5 atom of cytosine (C\text{OH}) occurs mainly in mitochondria and is caused by multiple external and internal factors such as ionization, life style or aging.\textsuperscript{20,31,216-221} Once oxidized, DNA looses its ability to perform fundamental biological functions and in case of the sufficient repair failure, the oxidative damage of DNA leads to fatal consequences. In rapidly proliferating and having high metabolism rates cancer cells the elevated concentration of C\text{OH} has been commonly detected.\textsuperscript{222} It was reported that the repair of C\text{OH} occurs but is insufficient, likely due to the significantly increased level of this modification.\textsuperscript{222} However, different rates for incision repair depending on the character of damaged base, location in the body or the enzymes targeting damage\textsuperscript{222} suggest that the low C\text{OH} repair efficiency could be related to the similarity between undamaged and damaged DNA in the first place.

Unfortunately the fatal consequences of cytosine oxidation go beyond the accumulation levels or the potential conformational similarities discussed; the unrepaiired C\text{OH} leads to mutations to thymine\textsuperscript{223,224} because deamination of C\text{OH} yields hydroxyl uracil that can lead to GC to AT mutations.\textsuperscript{225-227} C\text{OH} can additionally directly pair with adenine base (C\text{OH}/A) with a higher affinity in comparison to the regular mismatch (C/A), directly involving C\text{OH} in mutagenic mechanism.\textsuperscript{30} Since the C\text{OH}/A mismatch serves as a target for only limited number of repair enzymes, its conformational stability is worrisome for the mutagenic and carcinogenic consequences it leads to.\textsuperscript{226} Mutagenic character of the damage and increased contribution to non-carcinogenic diseases such as Parkinson’s and Alzheimer’s make this modification an essential target for accurate and early detection.\textsuperscript{30}

The characterization of structural differences in DNA that might occur upon cytosine hydroxylation can be performed with molecular dynamics simulations (MD). The overall
geometry of DNA and relative position of DNA bases within the base pair step have been accurately resolved by the computations of DNA six base pair step parameters: twist, roll, tilt, shift, slide and rise;\textsuperscript{150-153} our recent study allowed for their enhanced sampling in MD simulations.\textsuperscript{89} Up to date, the fluctuations of DNA step parameters due the modifications of DNA sequences or covalent modifications of the bases have been studied.\{Olson, 1998 #4418;Beveridge, 2004 #8868;Lankas, 2003 #4536;Drsata, 2013 #9073;\} In addition to DNA base pair step parameters, six DNA base pair parameters: opening, propeller, buckle, shear, stretch and stagger, provide detailed analysis of DNA spatial organization at the level of single base pair.\textsuperscript{153} These allow for a detection of shifts in hydrogen bonding patterns or relative orientation of the bases. Moreover, the cross-correlations between DNA base pair and base pair step parameters offer additional tool in the research of DNA properties.\textsuperscript{89,150,228}

The primary interests investigated in this study were twofold. First, the effect of cytosine oxidative damage on the local DNA geometry was investigated. Results indicated that the geometries of DNA containing undamaged (C/G) and damaged (C\textsuperscript{OH}/G) base pairs stay indifferent, which may contribute to already abated recognition by repair enzymes. Second, the influence of the mutagenic mismatch between damaged cytosine and adenine (C\textsuperscript{OH}/A) on DNA geometry was investigated. The outcomes implied strong alteration of DNA base pair geometry; plausibly this construct could serve as a better template for detection in a drug discovery studies towards molecules recognizing the damage with mismatch present rather then without. Nonetheless, less significant reorganization of C\textsuperscript{OH}/A base pair in comparison to the regular C/A mismatch suggested the more stabile structure in C\textsuperscript{OH}/A system; this effect may play an additional role in the alleviated recognition and increased accumulation of DNA mutations.
5.3 Computational Setup

MD simulations of 5'-CGCCA-G\textsubscript{6}-C\textsubscript{7}-A\textsubscript{8}-GCGACC-3' double stranded DNA in four forms: undamaged (C\textsubscript{7}/G\textsubscript{22}), damaged (C\textsubscript{7}OH/G\textsubscript{22}), undamaged-mismatch (C\textsubscript{7}/A\textsubscript{22}) and damaged-mismatch (C\textsubscript{7}OH/A\textsubscript{22}) were performed. Since there were no structures available in the PDB database,\textsuperscript{53} the systems were built with 3DNA software\textsuperscript{153} in B DNA form. The parameters for hydroxyl cytosine (Figure 5.1) were obtained using Paramchem\textsuperscript{229-232} and applied to C\textsubscript{7} on the main DNA strand. The DNA systems were solvated in a rectangular box of TIP3 water\textsuperscript{72} and 150 mM concentration of NaCl, with a distance of 12 Å between DNA and the box boundary. All systems were minimized and progressively heated for 1ns from 120 K to 300 K with 1.5 ns equilibration after each temperature level was achieved. Harmonic restraints with a force constant (FC) of 1 kcal mol\textsuperscript{-1} Å\textsuperscript{-2} were used on the heavy atoms during heating and equilibration in three independent simulation runs. Every 500 ps the restraints were released and the FCs were decreased from 1 to 0.5 and to 0.1 kcal mol\textsuperscript{-1} Å\textsuperscript{-2}. The unrestrained 1 ns equilibration was pursued and all systems were submitted to 100 ns long unbiased MD simulations. All simulations were performed with the CHARMM program,\textsuperscript{182} using the CHARMM 36 force field.\textsuperscript{187,201} Unbiased MD simulations run according to different parallelization protocol: standard parallel CHARMM version with CPUs,\textsuperscript{182} parallel CHARMM with domdec\textsuperscript{91} using CPUs and parallel CHARMM with domdec using GPUs.\textsuperscript{92} Enhanced one-dimensional (1D) umbrella sampling (US) simulations of twist at the central G\textsubscript{6}C\textsubscript{7}/G\textsubscript{22}C\textsubscript{23} and C\textsubscript{7}A\textsubscript{8}/T\textsubscript{21}G\textsubscript{22} steps were performed with our recently implemented in CHARMM subroutine.\textsuperscript{89} The calculations of twist step parameter in this method were improved by the re-defining the points on the DNA bases that are used in the calculations of certain vectors discussed in chapter 3. In all simulations a time step of 2 fs was used, SHAKE was applied to constrain the bonds with hydrogen atoms,\textsuperscript{129} the
temperature was controlled with the Nosé-Hoover thermostat,\(^6_0\) and long-range electrostatic interactions were handled by the particle mesh Ewald method.\(^7_4\) In 1D US of twist the umbrella windows were distributed between 20° and 60° at two base pair steps flanking the damage site, using umbrella window at every 2.5° and FC of 1 kcal mol\(^{-1}\) deg\(^{-2}\). This resulted in 102 simulations of twist each 2 ns long. The sufficiency in overlap of distributions in all US simulations was verified visually. Free energy surfaces were calculated using the weighted histogram analysis method\(^{13_1,18_5}\) using a bin size of 1° in each 1D US simulation.

![Figure 5.1. Undamaged and hydroxyl damaged cytosine.](image)

### 5.4 Results and Discussion

Trajectories from the independent unbiased MD runs were post-processed with 3DNA software and the average values of DNA base pair parameters were estimated.\(^{15_3}\) For C\(_7\)/G\(_{22}\) and C\(_7\)\(^{\text{OH}}\)/G\(_{22}\) systems no differences in buckle, stagger, stretch and shear were observed, subtle differences were detected for opening and propeller. This suggested that no disruption in the hydrogen bonding patterns nor in the geometry of base pair occurs due to the damage (Figure 5.2).
Mismatched undamaged and damaged DNA systems expressed significant changes in the values of base pair parameters and hydrogen bonding pattern; here only one nonstandard hydrogen bond was present between the bases (Figure 5.3). Despite this destabilization, the overall mismatched structures remained stable during the simulations and no flipping out of the bases occurred. Greater disruption in the geometry of base pair parameters in the C$_7$/A$_{22}$ in comparison to C$_7^{\text{OH}}$/A$_{22}$ system supports the higher affinity of pairing in C$_7^{\text{OH}}$/A$_{22}$.

Figure 5.2. DNA base pair parameters in C$_7$/G$_{22}$ (solid line), C$_7^{\text{OH}}$/G$_{22}$ (dashed line), C$_7^{\text{OH}}$/A$_{22}$ (dotted dark line), C$_7^{\text{OH}}$/A (dotted light line).

Figure 5.3. Hydrogen bonding patterns observed in MD simulations.
The average values for DNA base pair step parameters (Figure 5.4) only marginally differ between C7/G22 and C7OH/G22 as in the case of DNA base pair parameters. Introduction of the mismatch in damaged C7OH/A22 highly affected values of twist where the undertwisting at step G6C7/A22C23 and overtwisting at step C7A8/T21A22 were observed, as opposite to the systems without mismatch. Values of slide at step C7A8/T21A22 were also significantly altered by the mismatch. These observed changes of twist and slide are of particular interest due to their connection to the stacking interactions and geometry of the DNA sugar-phosphate backbone. Although the stabilization of base pair geometry through the hydrogen bonding is lost in C7OH/A22 and C7/A22, the stacking interactions between adenine and adjacent bases stabilize the mismatched sequence. On the contrary, in the C7/G22 and C7OH/G22 systems, the hydrogen bonding predominantly contributes to the base pair stability. Stacking between bases on the complementary DNA strand within the G6-C7-A8 sequence is illustrated in Figure 5.5.

Figure 5.4. DNA base pair step parameters in C7/G22 (solid line), C7OH/G22 (dashed line), C7OH/A22 (dotted dark line), C7/A22 (dotted light line).
Figure 5.5. Stacking interactions in the average structures of C7/G22 and C7\textsuperscript{OH}/A22.

The average values of step parameters and their standard deviations are summarized in Table 5.1. Rise at the G6C7/G(A)\textsubscript{22}C23 step has smaller values for both mismatched sequences in comparison to the C7/G22 and C7\textsuperscript{OH}/G22 systems. This is likely a result of the stronger stacking interactions observed in the presence of adenine within GAC sequence on the complementary strand where C\textsubscript{23} and A\textsubscript{22} bases come closer together. The standard deviations of the step parameters indicate some differences in local flexibility; this flexibility at G\textsubscript{6}C\textsubscript{7}/A\textsubscript{22}C\textsubscript{23} step is slightly higher for twist and lower for roll in C7\textsuperscript{OH}/G22 unlike in the C7/G22. Upon the mismatch introduction, additional differences in the local flexibility occur, with both mismatched structures being in general more flexible. Disruption in hydrogen bonding and increased flexibility suggest that DNA bases in the systems containing mismatch may become more exposed. Indeed, the level of hydration around cytosine base remained indifferent between C7/G22 and C7\textsuperscript{OH}/G22.
systems where 8 water molecules were located in the closest proximity (less than 2.6Å) to cytosine backbone on average. In the mismatched systems however, this number increased to 10 what supports a greater exposure of cytosine base to the solvent.

Table 5.1. Average values of DNA step parameters and corresponding standard deviations at the steps G₆C₇/G(A)₂₂C₂₃ and C₇A₈/T₂₁G(A)₂₂.

<table>
<thead>
<tr>
<th>Step</th>
<th>System</th>
<th>C₇/G₂₂</th>
<th>C₇OH/G₂₂</th>
<th>C₇OH/A₂₂</th>
<th>C₇/A₂₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twist (deg)</td>
<td>38.64 ± 4.52</td>
<td>38.16 ± 6.18</td>
<td>31.16 ± 4.16</td>
<td>24.86 ± 4.83</td>
<td></td>
</tr>
<tr>
<td>Roll (deg)</td>
<td>1.38 ± 5.67</td>
<td>1.21 ± 4.96</td>
<td>0.12 ± 5.38</td>
<td>7.50 ± 8.11</td>
<td></td>
</tr>
<tr>
<td>Tilt (deg)</td>
<td>1.99 ± 5.41</td>
<td>0.59 ± 4.94</td>
<td>2.76 ± 4.79</td>
<td>-5.75 ± 6.15</td>
<td></td>
</tr>
<tr>
<td>Rise (Å)</td>
<td>3.44 ± 0.32</td>
<td>3.39 ± 0.30</td>
<td>3.29 ± 0.38</td>
<td>3.19 ± 0.37</td>
<td></td>
</tr>
<tr>
<td>Slide (Å)</td>
<td>-0.20 ± 0.46</td>
<td>-0.25 ± 0.38</td>
<td>-0.10 ± 0.56</td>
<td>-0.12 ± 0.43</td>
<td></td>
</tr>
<tr>
<td>Shift (Å)</td>
<td>0.06 ± 0.67</td>
<td>0.13 ± 0.62</td>
<td>-0.23 ± 0.84</td>
<td>0.08 ± 0.74</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>System</th>
<th>C₇/G₂₂</th>
<th>C₇OH/G₂₂</th>
<th>C₇OH/A₂₂</th>
<th>C₇/A₂₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twist (deg)</td>
<td>32.43 ± 4.30</td>
<td>31.49 ± 5.41</td>
<td>41.46 ± 5.42</td>
<td>43.11 ± 5.49</td>
<td></td>
</tr>
<tr>
<td>Roll (deg)</td>
<td>8.75 ± 7.33</td>
<td>9.44 ± 6.94</td>
<td>6.26 ± 5.98</td>
<td>5.51 ± 7.54</td>
<td></td>
</tr>
<tr>
<td>Tilt (deg)</td>
<td>0.46 ± 5.02</td>
<td>-4.18 ± 4.06</td>
<td>-4.06 ± 5.55</td>
<td>-2.02 ± 5.79</td>
<td></td>
</tr>
<tr>
<td>Rise (Å)</td>
<td>3.41 ± 0.34</td>
<td>3.42 ± 0.32</td>
<td>3.48 ± 0.36</td>
<td>3.53 ± 0.34</td>
<td></td>
</tr>
<tr>
<td>Slide (Å)</td>
<td>-0.17 ± 0.58</td>
<td>0.03 ± 0.60</td>
<td>1.06 ± 0.90</td>
<td>0.55 ± 0.66</td>
<td></td>
</tr>
<tr>
<td>Shift (Å)</td>
<td>-0.37 ± 0.59</td>
<td>-0.32 ± 0.66</td>
<td>-0.27 ± 0.60</td>
<td>-0.61 ± 0.51</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2. Population (%) of BI states on DNA strands.

| %          | BI       |  | BI       |  |
|------------|----------|  |----------|  |
| DNA system | Main strand |  | Complementary strand |  |
| G₆C₇       | 91 ± 2   |  | G(A)₂₂C₂₃ | 88 ± 2 |
| C₇/G₂₂     | 96 ± 1   |  | C₇OH/G₂₂ | 86 ± 0 |
| C₇OH/G₂₂  | 81 ± 6   |  | C₇OH/A₂₂ | 83 ± 5 |
| C₇/A₂₂     | 67 ± 4   |  | C₇A₈     | 99 ± 3 |
| C₇A₈       | 87 ± 1   |  | T₂₁G(A)₂₂ | 95 ± 1 |
| C₇/G₂₂     | 86 ± 1   |  | C₇OH/G₂₂ | 92 ± 3 |
| C₇OH/G₂₂  | 96 ± 2   |  | C₇OH/A₂₂ | 80 ± 4 |
| C₇/A₂₂     | 95 ± 2   |  | C₇/A₂₂   | 86 ± 2 |
Variations of slide are typically linked to the changes in DNA backbone geometry reflected in the populations of BI and BII forms. Data shows that systems with $C_7/G_{22}$ and $C_7^{\text{OH}}/G_{22}$ pairs have similar geometry of DNA sugar-phosphate backbone at all steps flanking $C_7$ (Table 5.2); here the BI form is highly stabilized at $G_6/C_7$ and $T_{21}/G_{22}$ steps (over 90%) and subtly destabilized at the $C_7/A_8$ and $G_{22}/C_{23}$ steps (over 80%). The presence of mismatch tends to fairly stabilize BII form at all steps but $C_7/A_8$ and the disruption in BI and BII patterns in comparison to non-mismatched base pairs is evident. The variations in BI and BII forms also correlate with the changes in twist in standard and mismatched systems where the opposite patterns in BI and BII populations occur. The strongest deviations in BI and BII transitions are for $C_7/A_{22}$ system at $G_6/C_7$ step on the main, and $A_{22}/C_{23}$ on the complementary DNA strand; for this system the variations in values of rotational step parameters (twist, roll and tilt) and disruption in base pair geometry and hydrogen bonding (opening, propeller and buckle) most significantly deviated from the systems with standard base pairing or $C_7^{\text{OH}}/A_{22}$ system (Figures 5.2 and 5.4).

Figure 5.6. Free energy curves for $G_6/C_7/G_{22}/C_{23}$ (a) and $C_7/A_8/T_{21}/A_{22}$ (b) in $C_7/G_{22}$ (solid line), $C_7^{\text{OH}}/G_{22}$ (dashed line) and $C_7^{\text{OH}}/A_{22}$ (dotted line). Error bars obtained with block analysis.
The free energy curves from 1D US simulations are illustrated in Figure 5.6. Similarly to unbiased MD simulation, the overtwisting for the systems without mismatch occurs at the step G₆C₇/G₂₂C₃ and undertwisting is observed at C₇A₈/T₂₁A₂₂. The increased twist flexibility in the C₇/OH/G₂₂ system at G₆C₇/G₂₂C₃ step (Figure 5.6a) stays in agreement with previous observations and so far represents the only distinctive factor between C₇/G₂₂ and C₇/OH/G₂₂.

Changes of sugar-phosphate backbone geometry directly relate to the puckering of DNA sugar groups. In DNA the 5-membered sugar rings are puckered i.e. two neighboring atoms are situated on the reverse sites of the sugar plane while the remaining three atoms stay in its plane. Although multiple conformations are possible, C3'-endo and C2'-endo dominate on the so called “pseudo-rotation cycle” and correspond to A or B DNA forms respectively. In this study all trajectories were analyzed using the built-in puckering command in CHARMM. The average values were close for C₇/G₂₂ and C₇/OH/G₂₂, suggesting that the spatial organization in undamaged and damaged DNA on the level of DNA backbone is also similar (Table 5.3). As in the case of previous analysis, introduction of the mismatch disrupted these similarities yet C₇/OH/A₂₂ sugar puckers are less altered than in C₇/A₂₂ system. Interestingly the influence of the mismatch on the complementary strand extends to the main DNA strand.

<table>
<thead>
<tr>
<th></th>
<th>C₇/G₂₂</th>
<th>C₇/OH/G₂₂</th>
<th>C₇/OH/A₂₂</th>
<th>C₇/A₂₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>G₆</td>
<td>151 ± 27</td>
<td>143 ± 46</td>
<td>142 ± 43</td>
<td>149 ± 25</td>
</tr>
<tr>
<td>C₇</td>
<td>114 ± 60</td>
<td>115 ± 59</td>
<td>136 ± 48</td>
<td>123 ± 61</td>
</tr>
<tr>
<td>A₈</td>
<td>153 ± 38</td>
<td>156 ± 36</td>
<td>138 ± 52</td>
<td>134 ± 54</td>
</tr>
</tbody>
</table>

Table 5.3. Sugar puckering (deg) at the bases around C₇ on DNA main strand.
5.5 Conclusions

The impact of oxidative damage of cytosine on the geometry and mechanistic properties of DNA has been investigated using MD simulations. Lack of conformational differences in DNA systems with and without damage was detected and suggested that this similar spatial organization of DNA bases may be an important factor contributing to the inefficient recognition of oxidative damage by repair enzymes. Introduction of adenine mismatch to the sequence of DNA complementary strand lead to disruption in the standard bonding patterns and local geometry of DNA bases; nevertheless the stronger stacking interactions of adenine allowed for stabilization of the structure. Results indicate the fatality of C^OH damage coinciding with a poor prognosis for repair and may guide the future research towards the targeting of mismatched DNA systems instead.
REFERENCES


24 P. A. Jones, Progressive DNA methylation changes as potential risk factors for human cancer., *Cancer Epidemiology Biomarkers & Prevention* 2002, 11, 1262S.


34. B. Schuler, and W. A. Eaton, Protein folding studied by single-molecule FRET, *Current Opinion in Structural Biology* 2008, 18, 16.


45 G. Rhodes, *Crystallography made crystal clear* 2006).


58 S. Nose, A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods, *Journal of Chemical Physics* 1984, 81, 511.


112 C. Wasylyk, A. P. Bradford, A. GutierrezHartmann, and B. Wasylyk, Conserved mechanisms of Ras regulation of evolutionary related transcription factors, Ets1 and Pointed P2, Oncogene 1997, 14, 899.


121 B. Rabault, and J. Ghysdael, Calcium-Induced Phosphorylation of Ets1 Inhibits Its Specific DNA-Binding Activity, *Journal of Biological Chemistry* 1994, 269, 28143.


123 H. Kamberaj, and A. van der Vaart, Correlated motions and interactions at the onset of the DNA-induced partial unfolding of Ets-1, *Biophys J* 2009, 96, 1307.


126 H. Kamberaj, and A. van der Vaart, Extracting the causality of correlated motions from molecular dynamics simulations, *Biophys J* 2009, 97, 1747.


160 J. Spiriti, and A. van der Vaart, DNA Binding and Bending by Sac7d is Stepwise, *ChemBioChem* 2013, 14, 1434.


Y. Shimooka, J. Nishikawa, and T. Ohyama, Most Methylation-Susceptible DNA Sequences in Human Embryonic Stem Cells Undergo a Change in Conformation or Flexibility upon Methylation, *Biochemistry-Us* 2013, 52, 1344.

M. Banyay, and A. Graslund, Structural effects of cytosine methylation on DNA sugar pucker studied by FTIR, *Journal of Molecular Biology* 2002, 324, 667.


199 R. E. Harrington, DNA Curving and Bending in Protein DNA Recognition, *Molecular Microbiology* 1992, 6, 2549.


S. Shanak, and V. Helms, Hydration properties of natural and synthetic DNA sequences with methylated adenine or cytosine bases in the R.DpnI target and BDNF promoter studied by molecular dynamics simulations, *Journal of Chemical Physics* 2014, 141.


T. Brendler, and S. Austin, Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences, *Embo Journal* 1999, 18, 2304.


Y. S. Chung, T. Brendler, S. Austin, and A. Guarne, Structural insights into the cooperative binding of SeqA to a tandem GATC repeat, *Nucleic Acids Res* 2009, 37, 3143.


227 S. Tremblay, and J. R. Wagner, Dehydration, deamination and enzymatic repair of cytosine glycols from oxidized poly(dG-dC) and poly(dI-dC), *Nucleic Acids Res* 2008, 36, 284.


APPENDIX A:

PERMISSION FROM THE PUBLISHER FOR REPRINTING OF THE MANUSCRIPT

IN CHAPTER 2

RightsLink

Thank You For Your Order!

Dear Mrs. Aleksandra Karolak,

Thank you for placing your order through Copyright Clearance Center's RightsLink service. Elsevier has partnered with RightsLink to license its content. This notice is a confirmation that your order was successful.

Your order details and publisher terms and conditions are available by clicking the link below:
http://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=43363963-a270-474a-9998-23a9560e644

Order Details
Licensee: Aleksandra Karolak
License Date: Jan 9, 2015
License Number: 3544790180600
Publication: Biophysical Journal
Title: Importance of Local Interactions for the Stability of Inhibitory Helix 1 of Ets-1 in the apo State
Type Of Use: reuse in a thesis/dissertation
Total: 0.00 USD

To access your account, please visit https://myaccount.copyright.com.

Please note: Online payments are charged immediately after order confirmation; invoices are issued daily and are payable immediately upon receipt.

To ensure that we are continuously improving our services, please take a moment to complete our customer satisfaction survey.

B.1:v4.2

+1-855-239-3415 / Tel: +1-978-846-2777
customercare@copyright.com
http://www.copyright.com

98
APPENDIX B:

PERMISSION FROM THE PUBLISHER FOR REPRINTING OF THE MANUSCRIPT
IN CHAPTER 3

Thank You For Your Order!

Dear Mrs. Aleksandra Karolak,

Thank you for placing your order through Copyright Clearance Center's RightsLink service. John Wiley and Sons has partnered with RightsLink to license its content. This notice is a confirmation that your order was successful.

Your order details and publisher terms and conditions are available by clicking the link below:
http://a100.copyright.com/CustomerAdmin/PLF.jsp?ref=b69ed6c1-1db2-4a96-8c1e-675e902be9dd

Order Details
Licensee: Aleksandra Karolak
License Date: Jan 7, 2015
License Number: 3543780429238
Publication: Journal of Computational Chemistry
Title: Enhanced sampling simulations of DNA step parameters
Type Of Use: Dissertation/Thesis
Total: 0.00 USD

To access your account, please visit https://myaccount.copyright.com.

Please note: Online payments are charged immediately after order confirmation; Invoices are issued daily and are payable immediately upon receipt.

To ensure that we are continuously improving our services, please take a moment to complete our customer satisfaction survey.

B.1.x4.2

+1-855-239-3415 / Tel: +1-978-646-2777
customercare@copyright.com
http://www.copyright.com