Molecular Dynamics Simulation Study of Single DNA Nucleotides Transport Through Nanoslits

Kai Xia
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Mechanical Engineering Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/2852

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
MOLECULAR DYNAMICS SIMULATION STUDY OF SINGLE DNA NUCLEOTIDES TRANSPORT THROUGH NANOSLITS

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

The Department of Mechanical and Industrial Engineering

by

Kai Xia
BS., University of Science and Technology of China, 2004
M.S., Fudan University, 2008
M.S., Louisiana State University, 2014
August, 2015
ACKNOWLEDGEMENTS

This research project is supported by National Science Foundation under EPSCoR LA-SiGMA grant EPS-1003897 and by Mechanical and Industrial Engineering Department at Louisiana State University. High performance computational resources for this research were provided by High Performance Computing at Louisiana State University (http://www.hpc.lsu.edu), and the Louisiana Optical Network Initiative (http://www.loni.org).

I would like to express my sincere gratitude and thanks to my research advisor, Prof. Dorel Moldovan, who gave me the opportunity to work on such an exciting and challenging research project. I am also very grateful to my colleagues Dr. Brian Novak and Dr. Jieqiong Lin for their help and support.

Last but not least I would like to thank my parents for their support and encouragement for me to pursue my doctoral degree. I am also grateful to my beloved wife, Xu Huang, for her support and understanding during difficult times.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS................................................................................................. ii

ABSTRACT....................................................................................................................... v

CHAPTER 1 INTRODUCTION AND GENERAL OBJECTIVES........................................ 1
  1.1 Structural Characteristics of DNA............................................................................. 2
  1.2 Current Sequencing Methods................................................................................... 5
  1.3 Nanopores and DNA Translocation........................................................................... 6
    1.3.1 Biological Nanopores......................................................................................... 8
    1.3.2 Synthetic Nanopores......................................................................................... 9
  1.4 Detection and Sequencing......................................................................................... 10
    1.4.1 Electronic Detection......................................................................................... 10
    1.4.2 Optical Detection.............................................................................................. 12

CHAPTER 2 MOLECULAR DYNAMICS METHOD......................................................... 14
  2.1 Interatomic Potential Energy.................................................................................... 17
  2.2 Time Integration Algorithms.................................................................................... 18
    2.2.1 Verlet Algorithm............................................................................................. 18
    2.2.2 Leap-frog Algorithm...................................................................................... 19
  2.3 Boundary Conditions.............................................................................................. 19
    2.3.1 Periodic Boundary Condition.......................................................................... 20
    2.3.2 Fixed Boundary Condition.............................................................................. 21
    2.3.3 Free Boundary Condition.............................................................................. 21
  2.4 Temperature Control............................................................................................... 21
  2.5 Time Step and Nearest Neighbor List...................................................................... 23

CHAPTER 3 ELECTROPHORETIC TRANSPORT OF SINGLE DNA NUCLEOTIDES
THROUGH NANOSLITS: A MOLECULAR DYNAMICS SIMULATION STUDY........... 25
  3.1 Introduction............................................................................................................. 25
  3.2 Methodology.......................................................................................................... 31
  3.3 Results and Discussion........................................................................................... 36
    3.3.1 dNMP Adsorption and Desorption.................................................................. 39
    3.3.2 dNMP Phosphate Association with Sodium Ions........................................... 54
    3.3.3 dNMP Velocities............................................................................................ 63
    3.3.4 Distance to Separate the dNMPs and Analysis of the Time of Flight per dNMP .............................................................. 65
  3.4 Conclusion............................................................................................................. 71

CHAPTER 4 MOLECULAR DYNAMICS SIMULATION OF ELECTRICALLY DRIVEN
DNA TRANSPORT THROUGH PMMA NANOSLITS.................................................. 75
  4.1 Introduction............................................................................................................. 75
  4.2 Methodology.......................................................................................................... 75
    4.2.1 Preparation of force field............................................................................... 76
    4.2.2 Verification of force field.............................................................................. 77
    4.2.3 Construction of PMMA slab.......................................................................... 78
ABSTRACT

There is potential for flight time based DNA sequencing involving disassembly into individual nucleotides which would pass through a nanochannel with 2 or more detectors. Molecular dynamics simulation of electrophoretic motion of single DNA nucleotides through 3 nm wide hydrophobic slits was performed. Electric field strength (E) varied from 0.0 to 0.6 V/nm. Slit walls were smooth or had a roughness similar to nucleotide size. Multiple nucleotide-wall adsorptions occurred. The electric field did not influence the nucleotide adsorption and desorption mechanism for $E \leq 0.1$ V/nm, but influenced nucleotide orientation relative to the field direction. The nucleotide-wall interactions differed due to nucleotide hydrophobicities and wall roughness, and determined duration and frequency of nucleotide adsorptions and their velocities while adsorbed. Transient association of nucleotides with 1 to 3 sodium ions occurred but the mean association numbers (AN) were weak functions of nucleotide type. ANs for pyrimidine nucleotides were slightly lower than for purine nucleotides. Nucleotide-wall interactions contributed more to separation of nucleotide flight time distributions than ion association.

A PMMA slab was built and a CHARMM force field file modified from the force field for a PMMA trimer was verified and then utilized to study the transport of dNMPs through PMMA nanoslits. The simulation studies show that, while moving along the PMMA nanoslit the mononucleotides are adsorbed and desorbed from the walls multiple times. Due to their strong interaction with the PMMA walls the
mononucleotides can be trapped in adsorbed state for hundreds of nanoseconds. When dNMPs are in the desorbed state, their traveling velocity along the axis of the nanochannel is mainly affected by the association between Na\(^+\) and the phosphate group.

The Brownian MD simulation studies show that, the main characteristics of the mononucleotides through a nanochannel can be obtained by performing simulations of the dNMPs-PMMA wall system using a coarse-grained representation of the system. The accuracy of this method depends on the accuracy of the potential of mean force used to describe the interaction between dNMP and the PMMA wall.
CHAPTER 1  INTRODUCTION AND GENERAL OBJECTIVES

In the mid of the last century scientists found that the genetic information in cells is stored in a linear heteropolymer called Deoxyribonucleic acid (DNA) and deciphering this information has become a challenging task in biology ever since. DNA encodes the genetic information used in the development and functioning of all organisms and many viruses. Typically, a DNA molecule consists of two strands coiled around each other in a double helix structure. Each DNA strand is made up of nucleotides joined together by covalent bonds. Each nucleotide is composed of one of the four nitrogenous nucleobases [adenine (A), cytosine (C), guanine (G), and thymine (T)], a sugar group called deoxyribose, and a phosphate group.

Development of a low-cost and fast DNA sequencing method has the potential to radically change the medical field and enable us to study biological science more efficiently. The currently used sequencing techniques cost millions of dollars and may take several months to sequence a single human genome\(^1\). In the last decades since the development of various DNA sequencing techniques\(^2\), \(^3\), genome-based medicine has come closer to reality. Many novel ideas have been proposed and implemented. To speed up the process of DNA sequencing new methodologies have been investigated. From the multitude of the proposed methods those based on nanopores and nanochannels show great promise. Specifically, in these approaches nanopores or nanochannels are used either as a housing for the nanoscale probes or as a restriction that causes difference in
the signal between the bases. In this research we investigate a novel method for DNA sequencing based on measuring the time of transport (flight) of mononucleotides through a nanochannel. Using molecular dynamics simulations (MD) we investigate the mechanism of transport of the four mononucleotides in solution through a nanoslit driven by an electric field. We examine the effect of several experimental and material parameters including wall and solvent characteristics on the overall mononucleotide transport through nanochannels. Specifically, we investigate:

- How different are the physically measurable properties between the bases?
- How do the mononucleotide structure affects the time of flight?
- How do the different bases interact with the wall of the nanochannels?
- How fast does the DNA translocate through the nanochannels?
- What is the role of the surrounding fluid and ions?

Answering these questions is the main focus of our research. Moreover throughout the research we stress their importance for a proposed experimental sequencing method and point out possible future research work to explore them in more depth.

1.1 Structural Characteristics of DNA

DNA is heteropolymer that is made up of different bases attached to a sugar-phosphate backbone. The four kinds of bases that are attached to DNA backbone are shown in Figure 1.1. Based on the chemical structure they can be classified into two categories: the purine bases (A and G) and the pyrimidine bases (C and T). The purine bases consist of a six-member and a five-member
ring with a common edge, while the pyrimidine bases have only a six-member ring. It is obvious that the purines are larger and thus it may be expected that they will have different interactions with surfaces in nanochannels.

Figure 1.1: Atomistic structure of the four bases found in DNA. Each of them are often abbreviated as a single letter: adenine(A), guanine(G), thymine(T) and cytosine(C). A and G are purine bases, and T and C are pyrimidine bases.

The atomistic structure of a single-strand DNA is shown in Figure 1.2. Each monomer unit in the polynucleotide consists of sugar, phosphate group and one of the four kinds of bases. An important property of polynucleotide is that the backbone is charged in solution. The pKa of the phosphate group is near 1. Thus under most ionic solution such as physiological saline, each nucleotide unit on the backbone contain a negative charge. However a part of this charge may be neutralized by the nearby counterions such as Na\(^+\), K\(^+\) or Mg\(^{2+}\). It is interesting that polynucleotide has a global orientation, with one end as 5' and the other end 3'. This property is important for its translocation through a nanopore.
Figure 1.2: Atomistic structure of a single DNA strand in which nucleotides are linked together by the sugar-phosphate groups. The numbering system is also shown. The bases are attached to the 1’ carbon atom of the sugar part. The 3’ and 5’ carbon atoms of the sugar are used to name the two ends of the single strand. The phosphate group is negatively charged in solution.

In some recently proposed sequencing approaches an enzyme, called exonuclease, can be used to cleave (cut) the DNA into individual nucleotides. The exonuclease enzyme acts as a catalyst and facilitates polynucleotide separation from the DNA strand. Basically, it grabs the DNA strand and cleaves one nucleotide at a time by hydrolyzing the linking oxygen atoms between nucleotide repeating units to two OH groups. The resulting mononucleotides differ between themselves by their bases. Thus the size of the bases and of the nucleotides are important properties and are shown in Table 1.1. The basic idea behind the computation of the size of the nucleotides and the bases is to consider each atom as a sphere with radius equal to its van der Waals radius.
The total volume and area of these structures being given by the sum of the volume and the areas of all atoms comprising each structure (see Table 1.1).

Table 1.1: Sizes of the DNA bases ($V_B$) and nucleotides ($V_N$) in Å³. Also shown are the surface areas $A_B$ and $A_N$ in Å².

<table>
<thead>
<tr>
<th>Base</th>
<th>$V_B$ ($A_B$)</th>
<th>$V_N$ ($A_N$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>157 (166)</td>
<td>349 (340)</td>
</tr>
<tr>
<td>G</td>
<td>168 (177)</td>
<td>359 (351)</td>
</tr>
<tr>
<td>C</td>
<td>133 (147)</td>
<td>324 (319)</td>
</tr>
<tr>
<td>T</td>
<td>150 (163)</td>
<td>339 (331)</td>
</tr>
</tbody>
</table>

1.2 Current Sequencing Methods

Since developed by Sanger and his colleagues in 1977\textsuperscript{6}, Sanger sequencing method has been widely used for more than 20 years and is the prototype of many other sequencing methods used nowadays\textsuperscript{1, 2, 6}. For most of these methods, the sequencing process includes four steps:

Step 1: isolate the DNA strands that will be sequenced.

Step 2: replicate the DNA strands and break them into shorter strands.

Step 3: at first a primer polynucleotide is added and elongated to single DNA strands of different lengths corresponding to the DNA strand. Secondly these elongated strands of different lengths are spatially separated using capillary electrophoresis technique. At last the primers with fluorescent tags at their ends are detected and the last bases on the primers can be indicated.
Step 4: postprocess the data obtained from the step 3 and assemble the short sequences to get the original DNA strand. Thus it is necessary to make sure that the short sequences overlap with each other so that they can match up.

The process discussed above indicates that the current sequencing methods involve complicated preparation of DNA samples and data postprocessing. An important reason for this complexity is that the maximum length of DNA strand used in the capillary electrophoresis is limited. The DNA strand length is limited because the separation ability of electrophoresis analysis is dependent on the physical difference between the strands of different lengths. Therefore as expected it becomes more difficult to distinguish a strand with N nucleotides and one with N+1 nucleotides when N is a large number because the difference caused by the last nucleotide in the larger strand is relatively small comparing to the similarity of the two strands. Furthermore, the capillary electrophoresis process is slow and expensive.

1.3 Nanopores and DNA Translocation

To avoid the problem of limited read length of the DNA strand mentioned in the section 1.2, some novel methods were proposed to improve the step 3 in the sequencing process. The improvement can also decrease the time and work in other steps and thus reduce the cost of the whole sequencing process.

Nanopore has been proposed to be a promising tool for DNA sequencing. In 1996 Kasianowicz and his colleagues fabricated a biological nanopore and pulled a single-stranded(ss) DNA through the nanopore by exerting a voltage. In this
experiment the translocation of the ss-DNA through the nanopore was detected by measuring the blockade current. Since this innovative work, nanopores have attracted researchers’ interest to explore the possibility to utilize it in DNA sequencing\textsuperscript{8-13}. In addition to biological nanopore, in DNA translocation experiments synthetic nanopores were also fabricated with ion beam or self-assembly methods\textsuperscript{14-16} in which it is convenient to control the size, stability and other properties. Besides these experiments, computer simulation is also utilized to study the DNA translocation through nanopores. The phenomenological models are a simplified approach of polymer dynamics but still can indicate the effect of the factors in nanopore sequencing such as DNA length and pore dimensions\textsuperscript{17-20}. To study the phenomenon of polynucleotide dynamics in atomistic scale such as the structural difference between the nucleobases, molecular dynamics simulation is a powerful tool\textsuperscript{21-24}.

A schematic of a typical nanopore-polynucleotide experiment is shown in Figure 1.3\textsuperscript{25}. This shows a nanopore in a thin layer which separates a solution into two parts. A bias electric field is applied across the nanopore to drive ions through the nanopore and thus an ionic current $I_0$ is induced in the open channel. However, the polynucleotide with negative charge can also be pulled into the nanopore and consequently the pathway in the nanopore is partially blocked by the polynucleotide. Thus the ionic current decreases from $I_0$ to $I_b$ when the polynucleotide is transporting through the nanopore. This block event enables us to obtain the length of the polynucleotide and possibly other information like the sequence of the bases.
1.3.1 Biological Nanopores

Generally the biological pores are formed by α-hemolysin assembled in a lipid layer\textsuperscript{26}. A schematic of its molecular model is shown in Figure 1.4\textsuperscript{25}. The narrowest diameter in the channel is about 14Å which allows ss-DNA to transport through but not double-strand (ds) DNA. The channel of the pore has the adequate diameter and length to force the ss-DNA to unravel and extend and this is beneficial for discriminate the bases. The nanopore is fabricated by adding hemolysin subunits to the lipid bilayer where they self-assemble to form the pore\textsuperscript{7}. 

Figure 1.3: Schematic of a nanopore-polynucleotide experiment. (a) A bias voltage is applied to drive ions to transport through a nanopore. (b) A ss-DNA with negative charge on the backbone is moving through the nanopore driven by an electric field.
Figure 1.4: Schematic of the atomistic model of an α-hemolysin pore. The gray area is the lipid bilayer which the nanopore forms into. The wider vestibule is the entry for ss-DNA.

1.3.2 Synthetic Nanopores

Another type of nanopore is the synthetic pores created using solid-state fabrication methods. Compared to biological nanopores, it has several advantages. For example, the pore size and other properties can be adjusted to satisfy different experiment condition. In contrast, biological pores can only work under certain voltages and temperatures. Another advantage is that it allows the integration of sensors or probes like transverse electrodes.

The synthetic nanopore fabrication can be achieved with two techniques. One was developed by Li using low-energy ion beam to drill a nanopore in Si$_3$N$_4$\textsuperscript{14}. The method starts with using a focused ion beam to drill a 60nm hole in a Si$_3$N$_4$ membrane. The hole is then exposed to an Ar$^+$ beam which induces a diffusion process to shrink the hole instead of impacting atoms out of the membrane. The
process can be monitored to control the size of the pore. The images of an experiment of are shown in Figure 1.514. The other method to fabricate the synthetic nanopore is similar except that high-energy electron beam was used to shrink the large pore created previously27. The rate of reducing pore size in this method is relatively slow and thus it is easier to control the size of the pore. However the limit of the pore size is about 1nm dimensions due to the rough surface. Another interesting property of synthetic pores is that the surfaces can have negative charges which may cause more complications28.

Figure 1.5: The experiment images of fabricating a solid-state nanopore. (b) A large pore was created. (c) The pore shrinks after being exposed to Ar+ beam.

1.4 Detection and Sequencing

In this section the physical mechanism of DNA detection and sequencing will be discussed. There are mainly two categories used in DNA detection: the electronic and optical approach.

1.4.1 Electronic Detection

The electronic detection method is achieved by measuring the ionic blockade current through the nanopore7,10. This is achieved by embedding electrodes in
the nanopore and measuring the current flowing through the pore when the DNA is located inside the pore\textsuperscript{4, 29}.

**Figure 1.6:** (a) The current of polynucleotide Poly A, Poly C and Poly U when transporting through a biological nanopore. (b) Scatter plot in which the points represent the amplitude and duration of ionic blockade caused by Poly A or Poly U.

Since the first successful result of DNA sequencing by measuring the ionic blockade in biological nanopores under certain conditions\textsuperscript{7}, this detection method has attract great interest to investigate what kind of information can be examined to distinguish DNA bases\textsuperscript{10, 30-32}. Figure 1.6 shows an example indicating that the current amplitude and blockade duration together can discriminate Poly A, Poly C and Poly U from each other\textsuperscript{8}. The current amplitude of Poly A is obviously different from that of Poly C, and is close to that of Poly U. However, the blockade duration is shorter than that of Poly C.

Embedding electrodes in a nanopore to measure the transverse current across ss-DNA when this is threaded through the nanopore provides also a promising way to distinguish its base sequence\textsuperscript{4}. The essential of this method is to directly detect the electronic structure of the bases using the specific
electrodes. Measuring transverse currents has been applied to examine base differences in other fields\textsuperscript{33, 34}. A schematic is shown in Figure 1.7\textsuperscript{35}.

![Figure 1.7: Schematic of the experiment of measuring transverse current across polynucleotides. $E_\parallel$ pulls the DNA through the pore and $E_\perp$ is perpendicular to the electrode surface and generated to measure the current.]

\textbf{1.4.2 Optical Detection}

Another method used for detection of DNA in nanopores is by reading the optical signal generated by activating fluorescent tags attached on a DNA strand. A schematic of the process is shown in Figure 1.8\textsuperscript{36}. The first step of this method is to amplify the ds-DNA so that each based is represented by a unique sequence composed of about 20 bases. When the ds-DNA is pulled to the nanopore, the double strand is unzipped to single strands and the fluorescent tag is activated which can be detected by optical device. This method has shown a promising potential for DNA sequencing.
Figure 1.8: Schematic of the experiment of pulling a double strand of DNA with fluorescent tags through a nanopore.
CHAPTER 2  MOLECULAR DYNAMICS METHOD

The physical properties of materials are closely related to their microscopic structures. Due to complex interactions among particles in natural systems, analytical solutions are not available for most situations with traditional methods. The appearance of computer simulation provides a new method to study these complex phenomena. Utilizing the knowledge about the system of interest the general approach consists in building a model of the physical system and then utilizing computers to perform the appropriate calculations. Thus the challenges are in both providing a model which is as accurate as the experiment scientist expect and in computing the evolution of the model using the available high performance computers. The validation of both the model and the computations are done by comparing the computation results with the experiments. With the computational model tested and validated one can use it to further investigate and predict the behavior of the system under various conditions thus leading to important savings in resources consisting in expensive experimental equipment and materials as well as shortening the time for product development.

The earliest computer simulations were performed with a method named Monte Carlo, which is a stochastic simulation method. Shortly thereafter, a deterministic method called molecular dynamics was developed. Alder and Wainwright performed the first molecular dynamics (MD) simulation of liquid argon by numerically solving a classical equation of motion for each atom comprising the system$^{37}$. Rahman applied the MD simulation in a Lennard-Jones system with a step-by-step method in which the force exerted on every particle
was computed at every small time step to study the time evolution of the system. Currently most of MD simulations adopt this step-by-step procedure. The main drawback is that the simulation time and length scales are limited and the systems investigated cannot be expanded to mesoscale or macroscale due to excessive number of degrees of freedom in such systems. In addition there are significant limitation in terms of time over the system can be followed in such simulations. This limitation is mainly due to the fact that the simulation time step in MD simulations must be less than the vibration time period of atoms, usually less than one hundredth of it, and therefore is of the order of a just few femtoseconds ($10^{-15}$ sec). Molecular dynamics is a deterministic method in statistical physics and widely used in the research of multiple-particle systems. It first builds a set of Newton equations for each atom comprising the system and then numerically solves these equations of motion to obtain the coordinates and momentum of the atoms as function of time (trajectories). Then statistical methods are used to compute the macroscopic properties of the system, such as density, energy, pressure, transport properties such as diffusion coefficients, etc.

In atomic scale simulations, atoms are the fundamental particles of the system and thus the Hamiltonian of the system is the function of atomic coordinates and momentum. Based on classical mechanics, the Hamiltonian can be described as the sum of the kinetic energy of all atoms and the potential energy of interaction among the atoms. The sets of coordinates, velocities and momentum of all atoms can conveniently be referred to by using the following notation
\vec{r}^N = (\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N), \quad \vec{v}^N = (\vec{v}_1, \vec{v}_2, \cdots \vec{v}_N), \quad \vec{p}^N = (\vec{p}_1, \vec{p}_2, \cdots \vec{p}_N)

Then the Hamiltonian of the system is given by

\[ H(\vec{r}^N, \vec{p}^N) = K(\vec{p}^N) + \Phi(\vec{r}^N), \quad (2.1) \]

where

\[ K(\vec{p}^N) = \sum_{i=1}^{N} \frac{\vec{p}_i \cdot \vec{p}_i}{2m_i} \quad (2.2) \]

is the total kinetic energy.

The Lagrange equations are

\[ \frac{d}{dt} \frac{\partial L}{\partial \vec{v}_i} - \frac{\partial L}{\partial \vec{r}_i} = 0 \quad (2.3) \]

Where \( L(\vec{r}^N, \vec{v}^N) \) is Lagrange function, which is related to Hamiltonian as

\[ L(\vec{r}^N, \vec{v}^N) = \sum_{i=1}^{N} \vec{v}_i \cdot \vec{p}_i - H(\vec{r}^N, \vec{p}^N) \quad (2.4) \]

Substituting it to the Lagrange equation, we can obtain the equation set:

\[
\begin{cases}
\frac{d\vec{r}_i}{dt} = \frac{\partial H}{\partial \vec{p}_i} = \frac{\vec{p}_i}{m_i} \\
\frac{d\vec{p}_i}{dt} = -\nabla_i \Phi(\vec{r}^N)
\end{cases}
\quad (2.5)
\]
Because \( \frac{dp_i}{dt} = \vec{f}_i \), the term \( -\nabla_i \Phi(\vec{r}^N) \) is the total force exerted on the particle \( i \) by all other particles. The main focus in MD method is to solve numerically the equation set above and to obtain \( \vec{r}^N(t) \) and \( \vec{p}^N(t) \). One of the main ingredients critical for a MD simulation is ability to define a potential function to appropriately describe the interactions between atoms. The key is whether an appropriate force field \( \Phi(\vec{r}^N) \) can be obtained.

2.1 Interatomic Potential Energy

The potential energy function is described as a sum of potentials of atom interactions, which can be divided to bonded interaction and non-bonded interaction.

The bonded potential energy describes the interaction between atoms within a molecule and consists of three parts: bond \( (V_b) \), angle \( (V_a) \) and bond rotation \( (V_d) \). In most force fields they can be written as below:

\[
V_b(r_{ij}) = \frac{1}{2} k_{ij} (r_{ij} - b_{ij})^2 \tag{2.6}
\]

\[
V_a(\theta_{ijk}) = \frac{1}{2} k_{ijk} (\theta_{ijk} - \theta_{ijk}^0)^2 \tag{2.7}
\]

\[
V_d(\phi_{ijkl}) = k_\phi (1 + \cos(n\phi - \phi_0)) \tag{2.8}
\]

The bond energy \( V_b \) is a harmonic potential between two bonded atoms \( i \) and \( j \), where \( k_{ij} \) is the constant coefficient that describes the strength of the bond, \( r_{ij} \) is the distance between two atoms and \( b_{ij} \) is the equilibrium distance. The angle energy \( V_a \) describes an angle vibration and is a harmonic function of the angle \( \theta_{ijk} \). The angle is formed by three atoms \( i, j \) and \( k \), where \( j \) is the atom in the middle and \( i \) and \( k \) are connected to \( j \). Similar to the bond potential, \( k_{ijk} \) gives the
constant coefficient and \( \theta_{ijk}^0 \) is the equilibrium angle. The bond rotation energy \( V_d \) occurs between four consecutively bonded atoms \( i, j, k \) and \( l \), and is a function of the angle between two planes which are determined by \( i, j, k \) and \( j, k, l \) respectively. As previous, \( k_4 \) is the force constant and \( n \) is a nonnegative integer and indicates periodicity.

2.2 Time Integration Algorithms

Many numerical methods have been developed to solve the differential equation obtained in previous section. These include: central difference scheme\(^{39} \), predictor-corrector scheme\(^{40} \) and Verlet scheme\(^{41} \). The choice of the integration algorithm depends mainly on whether or not the integration can achieve steady convergence, and also on the complexity of the computation.

2.2.1 Verlet Algorithm

One of the most commonly used time integration methods is the Verlet algorithm\(^{41} \). This method is based on expanding the positions of atoms \( r(t) \) to Taylor series as a function of time. The positions at \( (t + \Delta t) \) and \( (t - \Delta t) \) could be expanded as follows:

\[
r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \frac{1}{6}b(t)\Delta t^3 + O(\Delta t^4) \quad (2.9)
\]

\[
r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \frac{1}{6}b(t)\Delta t^3 + O(\Delta t^4) \quad (2.10)
\]

Adding the two equations together, we get the following equation:

\[
r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 + O(\Delta t^4) \quad (2.11)
\]
Then substitute \( a \) with \( \frac{F}{m} \)

\[
r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{F}{m} \Delta t^2 + O(\Delta t^4) \tag{2.12}
\]

The equation 2.12 indicates that the Verlet method has a fourth-order accuracy. The position at the time \((t+\Delta t)\) can be calculated with the positions at two backward steps.

A similar and more commonly used algorithm is the so-called velocity Verlet algorithm, which has the advantage that solves explicitly for velocity as well. The velocity Verlet equations are shown below:

\[
r(t + \Delta t) = r(t) + v(t) \Delta t + \frac{1}{2} a(t) \Delta t^2 \tag{2.13}
\]

\[
v(t + \Delta t) = v(t) + \frac{1}{2} [a(t) + a(t + \Delta t)] \Delta t \tag{2.14}
\]

### 2.2.2 Leap-frog Algorithm

Leap-frog is slightly different from the Verlet algorithm and uses acceleration at time \( t \), velocity at \( t + \Delta t/2 \) and position at \( t \) to predict the position at \( t + \Delta t \). The equations for updating position and velocity are

\[
r(t + \Delta t) = r(t) + v \left( t + \frac{1}{2} \Delta t \right) \Delta t \tag{2.15}
\]

\[
v \left( t + \frac{1}{2} \Delta t \right) = v \left( t - \frac{1}{2} \Delta t \right) + a(t) \Delta t \tag{2.16}
\]

### 2.3 Boundary Conditions

The macroscopic properties of materials are statistical behavior of a large number of particles. Thus the simulation model must be large enough to correctly
simulate the behavior of realistic system. However, the number of particles that molecular dynamics can deal with is still limited due to the restriction of computer memory. In the MD approach usually the focus is on a small cube extracted from the macroscopic realistic system. Naturally, the atoms on the boundary are subject to the force from external atoms which is balanced by the force exerted by the atoms inside the box. Moreover, the atoms on the boundary exchange energy from the outside environment when they are subject to disturbance. Therefore, it is necessary to choose an appropriate model for the boundary of the simulation system. Generally there are three kinds of boundary conditions.

2.3.1 Periodic Boundary Condition

In most molecular dynamics simulations periodic boundary condition is applied. Generally the simulation box is a cuboid in which the number of particles can range from several hundred to several millions. If the macroscopic system to be investigated is a bulk, then the center unit box is replicated in three dimensions. If a surface is the simulation system, then periodic boundary condition is applied in two dimensions and the simulation box will be repeated in the two dimensions. In these systems the images of the particles appear periodically in three or two dimensions, which are called pseudo-particles. The position of the image of the particle $i$ is:

$$(x_i+lL_x, y_i+mL_y, z_i+nL_z) \quad l,m,n=0, \pm 1, \pm 2, \ldots$$

Where $L_x$, $L_y$ and $L_z$ are the lengths of the three sides of the unit cell.

Under periodic boundary condition, there are infinite pseudo-particles in the system. However, if a particle leaves from the unit cell in the molecular dynamic,
one of its image particles enters the unit cell from the other side of the cell. Thus the number of particles in unit cell remains the same. One of the conditions to apply periodic boundary condition is that the space correlation between two particles with a distance of $L$ is negligible, which means that the cutoff distance of interaction between two particles should be less than the lengths of the three sides of the simulation box.

2.3.2 Fixed Boundary Condition

The molecular dynamics simulation of canonical system (NVT) often uses fixed boundary condition. In these simulations the system boundaries are provided by so called walls or surfaces consisting of several atomic layers which are held at fixed predefined locations. For example, when simulating the interaction between various particles and a solid surface the surface is provided by a slab several atomic layers thick that held at fixed location in space.

2.3.3 Free Boundary Condition

If the unit cell is large enough to contain all of the particles, or avoid the influence of periodic or fixed boundary conditions, molecular dynamics simulation can use free boundary condition which means that the particles on the boundary are not subject to any restriction.

2.4 Temperature Control

In statistical mechanics canonical ensemble is commonly used where the total particle number ($N$), volume ($V$) and Temperature ($T$) are constant. In this ensemble the simulation system can be regarded to be coupled to a heat bath
which can adsorb from or release to a large amount of energy to the simulation system. Thus the fluctuation of the kinetic energy of the system is suppressed and the system temperature can be kept constant at the desired value. There are several thermostat methods such as Andersen thermostat, Nose-Hoover thermostat, Langevin thermostat and Berendsen thermostat. In our molecular dynamics simulation Berendsen thermostat is adopted to generate the statistical data for analysis because it has a higher computational and allows temperature fluctuation in the canonical ensemble. The drawback of Berendsen method is that the ensemble it generates is unknown.

In Berendsen thermostat the temperature is updated at each time step so that the change of temperature is proportional to the difference between the system temperature and the desired temperature:

\[
\frac{dT}{dt} = \frac{1}{\tau} (T_0 - T)
\]  \hspace{1cm} (2.17)

where \(\tau\) is the coupling parameter which determines how tightly the heat bath and the simulation system are coupled, \(T_0\) is the desired temperature. This means that the system temperature decays to the desired value at an exponential rate:

\[
T = T_0 - Ce^{-t/\tau}
\]  \hspace{1cm} (2.18)

The discrete form of 2.17 is

\[
\Delta T = \frac{\Delta t}{\tau} (T_0 - T)
\]  \hspace{1cm} (2.19)
Because the temperature is proportional to the square of velocity, the scaling factor \( \lambda \) such that \( \vec{v}_i \rightarrow \lambda \vec{v}_i \) is determined:

\[
\lambda^2 = 1 + \frac{\Delta t}{\tau} \left( \frac{T_0}{T} - 1 \right)
\]  

(2.20)

where \( \Delta t \) is the integration time step, \( \tau \) is the coupling time constant which determines how fast the desired temperature is reached, \( T_0 \) is the desired temperature and \( T \) is the current temperature.

### 2.5 Time Step and Nearest Neighbor List

For numerical integration to the equations of motion in molecular dynamics, a smaller time step \( \Delta t \) is better in the perspective of computational accuracy. However the computational efficiency determines that the time step \( \Delta t \) can’t be too small. The choice of the time step is a compromise between the computational accuracy and efficiency. A basic principle to set the time step is that the fastest particle in the system at any moment can’t move more than 2% of the size of a unit cell. The percentage above is adjustable to achieve the highest efficiency. A typical integration time step is from 0.1 fs to 10 fs.

The most time-consuming part in molecular dynamics simulation is to compute the force on the particles exerted by other particles in the system at every time step. To increase the computational efficiency, Verlet\(^{41}\) developed a method in which the nearest neighbor list is created for every particle and updated every \( n \) time step. The nearest neighbor list of an arbitrary particle \( i \) includes the particles within a distance of \( r_m \) from the particle \( i \). The value of \( r_m \) should be larger than the maximum cutoff distance of the potential functions.
among the particles and assure that none of the particles outside of the nearest neighbor list can enter the range within which the particle $i$ has interaction.
3.1 Introduction

Automated DNA sequencing has attracted significant interest since the Human Genome Project began. A major goal is to develop a high-throughput and low-cost method to identify each DNA nucleotide in the correct sequence as it passes a sensor(s). There are three major sequencing approaches of this type under consideration. In some of these approaches intact DNA strands are used,\textsuperscript{42} and in others individual nucleotides are first cleaved sequentially from DNA by an exonuclease.\textsuperscript{43, 44} The first approach involves modification of the nucleotides so optical detection methods can be used. In the second approach, a DNA strand or single nucleotides from a disassembled strand are passed through a nanopore and the identity of each nucleotide is determined as it passes.\textsuperscript{45-49} The final approach involves passing single nucleotides from disassembled strands through a nanochannel containing multiple detectors. In this approach the flight times of the nucleotides between detectors are used to identify them.\textsuperscript{50}

There are also several methods of detection under consideration for use with the above approaches. In optical methods, fluorescence is used.\textsuperscript{51} If an electric field is used to drive DNA or nucleotides through a nanopore, a current associated with the flow of other ions through the pore exists. Part of this current is blocked when a nucleotide is in the pore,\textsuperscript{10, 52} and the magnitude of the
blockage depends upon the nucleotide type. Electrodes might be placed in a nanopore to measure the transverse conductance or transverse differential conductance associated with each nucleotide as it passes. Sequence specific hysteresis effects were observed when using an AC field with a nanopore. Graphene nanoribbons have been used with nanopores since the currents induced in the nanoribbons due to their interaction with each passing nucleotide are orders of magnitude higher than ion blockage currents. For flight time based sequencing, the detectors will likely have to involve a restriction similar to a nanopore. The detection method will also likely be similar to those being considered for the nanopore based methods. The difference is that in the flight time based method, only the presence of a nucleotide needs to be detected, not its identity; a noisier signal can be tolerated.

Optical methods rely on fluorescent labeling to distinguish the nucleotides. Color discrimination has been used to distinguish each nucleotide of an enzymatically disassembled DNA strand with fluorescently labeled nucleotides. In another method, each nucleotide is substituted for by a unique group of 16 nucleotides, then each of these oligonucleotides are hybridized to a fluorescently labeled strand, and finally a nanopore is used to remove the fluorescent strand and detect the oligonucleotide type and therefore the identity of the original nucleotide. A third method follows the incorporation of fluorescently labeled nucleotide triphosphates as they are added to a growing DNA strand by polymerases.
In nanopore based sequencing methods, the nanopores act to help constrain the configurations of DNA, as the housing for electrodes, or as a constriction to create a blockage in ion current. The reliability and cost for the manufacturing of devices containing nanopores will play an important role in whether these types of DNA sequencing will be successful. Nanopores may be biological or synthetic. Biological pores include the membrane proteins α-hemolysis\textsuperscript{10, 43, 52, 69-72} or porin A.\textsuperscript{73} Pores composed of synthetic DNA nanostructures\textsuperscript{74, 75} or carbon nanotubes\textsuperscript{76} spanning a lipid bilayer have been studied. Synthetic pores constructed in silicon nitride\textsuperscript{10, 77, 78} or hafnium oxide,\textsuperscript{79} or in 2D materials such as boron nitride\textsuperscript{80, 81}, graphene,\textsuperscript{82-88} or molybdenum sulfide\textsuperscript{89} provide additional flexibility including the ability to adjust the pore size and chemically modify the nanopore surfaces or entrap other structures such as synthetic DNA nanostructures\textsuperscript{75} within them, and improved mechanical stability compared to membrane-bound systems.

Nanopore sequencing has limitations. For intact DNA, the measured signal may be a convolution of multiple adjacent nucleobases.\textsuperscript{90} One advantage of using nanopores composed of 2D materials is that they can potentially overcome this problem since a single or a few sheets are of a similar or smaller depth than the depth of a single nucleotide. Another challenge is to improve the resolution to make detection more reliable. It only takes several microseconds for each nucleotide to move through a nanopore and it could be in a variety of different conformations, thus it is difficult to extract useful signals from background noise, particularly when ion blockage currents are used due to their small magnitude.
Solutions to this problem include reducing the traveling velocity of the DNA or single nucleotides or controlling the conformations of DNA while in the pores using various means\textsuperscript{43, 72, 91-96} or creating two detection sites within nanopores.\textsuperscript{97} Other detection methods such as transverse conductance and graphene nanoribbons discussed above are less susceptible to this problem due to the larger currents compared to ion blockage currents.

In this research we focus on the flight time based approach for DNA sequencing which involves sequential, enzymatic DNA disassembly into single nucleotides. These nucleotides are driven through a nanochannel with detectors placed at multiple locations. The enzymes used would be λ–exonucleases covalently attached to pillars. λ–exonucleases have been attached to poly(methyl methacrylate) (PMMA) pillars and the activity of the attached enzymes was slightly higher than for free enzymes.\textsuperscript{98} λ–exonucleases disassemble one strand of double stranded DNA into single nucleotides with phosphate groups on their 5’ ends, deoxynucleotide 5’-monophosphates (dNMPs). The identity of each different type of dNMP; deoxyadenosine 5’-monophosphate (dAMP), deoxycytidine 5’-monophosphate (dCMP), deoxyguanosine 5’-monophosphate (dGMP), deoxythymidine 5’-monophosphate (dTMP), and the epigenetically modified dNMP deoxy-5-methylcytidine 5’-monophosphate (dMCMP) might be determined by measuring the time(s) taken for it to travel between the different detectors (flight time(s) or time(s) of flight). These flight times will have a specific distribution for each type of dNMP.
The following is a discussion of the advantages and disadvantages of the flight time based approach to sequencing. The main advantage of this method is that the detectors only need to determine the presence of a dNMP rather than its identity. The problem of trying to distinguish closely spaced nucleotides in intact DNA does not exist when the DNA is disassembled. However, disassembling the DNA also has disadvantages. The first is simply the complication of having to introduce immobilized λ-exonuclease enzymes into the system. The use of free dNMPs instead of intact DNA allows for the possibility of misordering due to one dNMP passing another one. This can be minimized by changing the magnitude of the driving force and the rate at which the enzyme disassembles the DNA by altering solutions conditions such as temperature and pH. Another issue is that diffusion broadens the flight time distributions leading to longer channel lengths and analysis times per dNMP.\textsuperscript{50} To reduce this, it is desirable for the magnitude of the driving force to be as large as is practical. If the nucleotide-wall interactions are relied upon to separate the flight time distributions, then the channels should be as narrow as possible so that the dNMPs are in contact with them for a large fraction of the time. The channel walls should also be as homogeneous as possible both chemically and physically since heterogeneities will lead to varying dNMP-wall adsorption energies and broadening of the flight time distributions. The adsorption energies should also be small enough that the dNMPs do not become stuck on the walls for long periods of time since that could lead to misordering.
Since all the nucleotides are anionic for a pH above about 3.2, the simplest way to drive DNA or single dNMPs through nanopores or nanochannels is by using an electric field to cause electrophoresis and, with charged channel walls, an electrokinetic flow.

In this study we investigated the electrophoretically driven transport of the four major DNA nucleotide monophosphates or dNMPs (dAMP, dCMP, dGMP, and dTMP) through nanoslits composed of disordered carbon atoms using all atom, explicit solvent molecular dynamics (MD) simulations. The main goals include development of fundamental understanding of the mechanism of the dNMP transport and assessment of the likelihood of this type of hydrophobic surface, with no specific dNMP-surface interactions, for being suitable for discriminating the flight time distributions of the different types of dNMPs and therefore being useful for flight time based DNA sequencing. The variables considered were the electric field strength and wall roughness.

Various factors influencing the transport of the dNMPs through nanoslits were examined. The strength of the interaction of the dNMPs with the slit walls and their adsorption and desorption behavior is an important factor determining the channel length required to distinguish the flight time distributions of the four dNMP types. Association of counter ions with the phosphate group of the dNMPs can lead to a significant change in dNMP velocity along a channel. An understanding of these factors can provide guidance for the design of nanochannel surfaces.
3.2 Methodology

The simulation system consisted of a dNMP and sodium chloride in water confined between two slit walls. Periodic boundary conditions were used in the directions tangential to the walls. The wall slabs had dimensions of 5 X 5 nm in the tangential directions. Smooth wall slabs had a depth of 1.2 nm (see Figure 3.1).

Figure 3.1: Simulation system with smooth walls. Carbon atoms (gray), sodium ions (yellow), chloride ion (green), and dNMP atoms are represented as large spheres. The water molecules are in a ball and stick representation. The electric field is applied in the positive x direction causing the dNMP to move in the negative x direction on average.

For smooth walls, the atoms in the slit walls were placed outside two planes parallel to the xy plane, located at \( z = \pm \frac{h_{\text{slit}}}{2} \) nm. The centers of the wall atoms were located at \( z \leq -\frac{h_{\text{slit}}}{2} \) nm and \( z \geq \frac{h_{\text{slit}}}{2} \) nm. Slit walls were composed of atoms with Lennard-Jones 12-6 parameters for a carbonyl carbon atom (\( \epsilon_{\text{carbon-carbon}} = 0.11 \) kcal/mol, \( \sigma_{\text{carbon-carbon}} = 0.4/2^{1/6} \) nm). The mass of the wall atoms
was increased from 12.011 to 14.30226 amu, which is the average mass of the atoms in a united atom (no hydrogen atoms, but increased mass of atoms that would have hydrogen atoms bonded to them) representation of PMMA. The details on constructing the smooth slit walls are given in our previous work.\textsuperscript{50}

The slit width of 3.0 nm was chosen to maximize the contact of the dNMPs with the slit walls while still having bulk solvent near the slit center plane. Real channels with dimensions smaller than 5.0 nm have been fabricated.\textsuperscript{102} Note that small, hydrophobic nano-scale geometries are not practical due to the difficulty for aqueous solvent to enter them. The nature of the slit walls was chosen for simplicity for our initial studies. Future work will focus on more realistic walls composed of PMMA or modified PMMA.

10 slits with rough walls were constructed with an average RMS roughness of 2.2708 nm. The spacing between the slit walls was chosen so that the volume accessible by the dNMPs was approximately the same as in the smooth wall cases.

Once the walls were constructed, dNMPs were placed between the slit walls and water and ions were added using the Visual Molecular Dynamics (VMD) software.\textsuperscript{103} Constant pressure simulations could not be performed in LAMMPS using long range electrostatics in the slab geometry, so the amount of water was determined by trial and error and interpolation. The \textit{Solvate} function in VMD was used to add water. The \textit{Solvate} settings used were: boundary = 2.4, x and y bounds = ±25.65 Å, z bounds = ±15 Å for the smooth wall cases, and boundary =
2.4, x and y bounds = ±25.44 Å, z bounds sufficient to be at or beyond the rough part of the slabs for the rough wall cases. These settings gave a reasonable equilibrium bulk density of water for the model that was used (1.015 g/cm³ at the center of the slit) in equilibrium simulations containing only water and the slit walls. For the trial and error density calculation for the rough wall case, the slit with walls with an RMS roughness of 2.2747 nm was used. After using Solvate, water molecules were replaced with two sodium ions to neutralize the system, and an extra sodium and chloride ion using the Autoionize function in VMD. Final configurations for the smooth wall cases contained between 2199 and 2233 water molecules. Final configurations for the rough wall cases contained between 2264 and 2356 water molecules. Ion concentrations ([Na⁺] + [Cl⁻]) were between 100.7 and 102.3 mM for smooth wall cases and between 95.5 and 99.4 mM for rough wall cases based on bulk water density to estimate the volume of solvent.

The CHARMM27 force field¹⁰⁴ was used for the dNMP and ion parameters. The rigid CHARMM TIP3P model was used for water. dNMPs with the phosphate group on their 5' end can be produced by cutting up double-stranded DNA using a λ-exonuclease enzyme. The λ-exonuclease enzyme that is used will likely have the highest activity near the physiological pH of around 7.4. The pKₐ for the first protonation of the phosphate groups of the dNMPs is around 6.8,⁹⁹ so the phosphate groups were simulated as non-protonated which gave the dNMPs a net charge of -2e where e is the electron charge. The CHARMM27 topology file did not contain a terminal segment for DNA with a non-protonated phosphate on the 5' end. Construction of the required patch for a non-protonated phosphate on
the 5' end is explained in previous work\textsuperscript{50} and the patch is included in the Supporting Information for that work. The chemical structure of the simulated dNMPs is shown in Figure 3.2.

![Diagram of nucleotides](image)

Figure 3.2: The structure of nucleotides with an un-protonated phosphate group attached to the 5' atoms of their sugars. Axis 1 is defined by atoms in the nucleobase rings, and some analysis of nucleobase configuration is based on it. If Axis 1 passes through an atom in the picture, then the position of that atom is one of the points needed to define the axis. With dAMP and dGMP, the axis passes between two pairs of atoms. The geometric center of the pair furthest from the end of the nucleobase where the sugar is attached is used for the second point needed to define the axis.

Non-bonded interactions were treated as follows. The Lennard-Jones 12-6 interactions were switched to zero between 0.8 and 1.0 nm using the CHARMM switching function. The CHARMM force field uses Lorentz-Berthelot mixing rules.
to determine Lennard-Jones 12-6 parameters between atoms of different types. The short-range cutoff for electrostatic interactions was 1.0 nm. The 3-D particle-particle-particle-mesh (PPPM) method corrected for a slab geometry was used with the minimum accuracy in the PPPM forces set to $10^{-5}$. The box length for the PPPM in the non-periodic direction was three times the simulation box length in that direction.

Wall atoms were at a much higher density than Lennard-Jones particles would be at under the simulated conditions. Since a thermostat was applied to the wall atoms (see below), their positions could not be fixed. Therefore, to keep the walls intact, the wall atoms were attached to their initial positions by springs with force constants of 83,860 kCal/mol-nm$^2$.

Since the dNMPs have a net charge, they can be driven electrophoretically. For the non-equilibrium simulations, electric fields with strengths ranging from 0.0144 to 0.6 V/nm were applied to all atoms. The average steady state dNMP velocities in the direction of the applied field ranged from about 0.18 to 31 m/s. Simulation at high velocities relative to velocities typical of nano-scale flows is necessary in MD simulations to obtain statistically meaningful results in the relatively short time that can be simulated, nanoseconds to microseconds.

During non-equilibrium simulations, energy is continuously added to the system so a thermostat should be used to remove it. To avoid any artifacts caused by thermostatting the flowing solvent and dNMP, only the wall atoms were thermostatted at 300 K using a Berendsen thermostat with a time constant
of 0.1 ps. The fluid temperatures at steady state were no more than 10 K higher than the temperature of the thermostatted walls due to viscous heating of the fluid. Equilibrium simulations were thermostatted in the same way as the non-equilibrium simulations.

The following cases were simulated. Simulations with smooth walls were run for all four dNMPs with the electric field strengths of 0.0, 0.0144, and 0.1 V/nm. Five simulations with different starting configurations for the wall atoms were run with 70 ns of production time for each. Additional simulations with smooth walls were run for dTMP only with the electric field strengths of 0.3 and 0.6 V/nm using 5 different wall configurations for each electric field strength and 40 ns of production time per simulation. Finally, non-equilibrium simulations with rough walls were run for all mononucleotides with an electric field of 0.1 V/nm using 10 different wall configurations and with production times per simulation of 40 to 62 ns. The RMS roughnesses of the walls were 1.9799, 2.1131, 2.1967, 2.2085, 2.2747, 2.2926, 2.3326, 2.3412, 2.4411, and 2.5272 nm with an average of 2.2708 nm. 10 ns of each simulation was discarded to allow the simulation to reach a steady state or equilibrium state.

3.3 Results and Discussion

The primary variables for flight time based sequencing include the rate at which the enzyme disassembles a DNA strand, the nanochannel length, the number of detectors, the nanochannel dimensions, the magnitude and type of the driving force pushing the dNMPs through the nanochannel, and the interactions
of the dNMPs with the channel walls and other species in solution. A secondary variable is the redundancy; how many times each DNA sequence is analyzed, either in serial or in parallel. A given identification accuracy can be obtained by modifying both the primary variables and the secondary variable; more redundancy means the time of flight distributions in each nanochannel do not have to be as well separated.

The goal is to reduce the time required to analyze each dNMP below some acceptable maximum value and to minimize the cost to analyze each dNMP. For a given set of flight time distributions over some length, the following conditions have to be met for sequencing to be successful: 1) the enzyme disassembly rate must be set so that the time between dNMPs being cleaved is at least equal to the time between the upper edge of the flight time distribution over the distance between the first and last detectors of the fastest dNMP type and the lower edge of the flight time distribution over the distance between the first and last detectors of the slowest dNMP type to avoid the problem of misordering caused by dNMPs passing each other in the nanochannel, 2) either the nanochannel length must be long enough for the flight time distributions to be adequately separated for reliable identification of the dNMPs or the redundancy must be large enough for the chosen nanochannel length to obtain reliable identification of the dNMPs. In addition, the number of detectors should be as large as practical or cost effective since increasing the number of detectors improves the ability to identify each dNMP allowing for shorter nanochannel lengths and analysis times.
The flight time distributions are determined by the magnitude and type of the driving force, and by the interactions of the dNMPs with the rest of the system. The standard deviations of the flight time distributions decrease with increasing nanochannel length so assuming the distributions for each type of nucleotide are not exactly the same for a given length, they will eventually not overlap for long enough lengths. To obtain reasonable nanochannel lengths and analysis times, the flight time distributions for a given length need to be sufficiently separated. Since the dNMPs are charged, they can be driven electrophoretically. In that case, each nucleotide will have a different electrophoretic mobility. Unfortunately the different dNMP types are of similar size and have the same charge, so their mobilities are similar. Relying only on mobility differences would not provide good separation of the flight time distributions. Although the different dNMP types are similar in size, their nucleobase groups are chemically different which allows the interactions of the dNMPs with other parts of the system to be exploited to separate the flight time distributions. This is the factor exploited by chromatography which can be used to separate or analyze different dNMP types from a solution containing multiple types.\textsuperscript{106, 107} The flight time based approach is essentially a single molecule version of chromatography.

Since this study only involves the motion of dNMPs through a nanoslit and does not involve the exonuclease enzyme or detectors, the factors of interest are the dNMP-wall and dNMP-counter ion interactions. Statistics and dynamics of the dNMP adsorption to and desorption from the slit walls are reported. dNMP orientation during adsorption and desorption; fraction of time adsorbed;
frequency of adsorption and desorption events; and the mean times, distances traveled by the dNMPs in the direction of the driving force, and velocities in the direction of the driving force while adsorbed and desorbed were calculated. The mean association numbers of the sodium ions with the dNMP phosphate groups were calculated as well as the characteristic relaxation times for states where 1 or 2 sodium ions were associated with the phosphate group. From the dNMP velocity distributions, the required channel lengths and minimum analysis times per dNMP were estimated. There were two parameters of interest; the strength of the electric field driving the dNMPs and the roughness of the slit walls. Uncertainties in the form of error bars or in tables are 2 times the standard error unless otherwise indicated.

3.3.1 dNMP Adsorption and Desorption

In nanoscale geometries, the interaction of the solvent and the solutes in solution with the walls becomes more important since the volume of the solution-wall interfacial region becomes a significant fraction of the total solution volume. As mentioned above in the context of flight time based sequencing, the interaction of dNMPs with the nanochannel walls is important for distinguishing the flight time distributions of the different dNMP types. In nanopore based sequencing, the interaction of DNA or single dNMPs with the nanopore walls could also be important. The effect of binding of DNA to the polymer layer anchored on a nanochannel or directly on the surface of a nanochannel has been examined.\textsuperscript{108-110} During the translocation process through nanochannels, DNA is subjected to a series of adsorptions and desorptions to and from the wall surfaces. These transient adsorption events result in an effective mobility
decrease. This adsorption and desorption phenomenon is also observed in the electrically facilitated transportation of proteins through large nanopores. Our results presented below show that the dNMPs in our systems also undergo adsorption and desorption from the walls.

Figure 3.3: The position of the center of mass of the dNMPs driven by an electric field of 0.1 V/nm in the wall normal direction (Z) over 70 ns. The center of the slit is at Z = 0. Red arrows in the dAMP and dGMP trajectories indicate examples of periods when the dNMP is adsorbed to a slit wall.

Figure 3.3 depicts the time dependence of the Z coordinate (wall normal direction) of the four dNMPs driven by an electric field of 0.1 V/nm. Those plots show that all the dNMPs stay adsorbed to the walls for periods of time between about 1 ns to 20 ns, then desorb again. Thus based on the Z coordinate of the center of mass of the dNMPs, a trajectory can be decomposed into adsorbed
states and desorbed states which alternate with each other. The method of
determining these states is described in the Supporting Information for our
previous work\textsuperscript{50} for the smooth slit walls.

In the adsorbed state, the hydrophobic nucleobase parts of the dNMPs tend
to sit on the surface while the hydrophilic phosphate group points away from the
surface. Figure 3.4(a) shows a snapshot of dAMP with its nucleobase fully
adsorbed on a wall. The end of the nucleobase furthest from the hydrophilic
phosphate group is more apt to stick on the wall and thus the nucleobases are
inclined to the wall surface when they are in the process of adsorbing to or
desorbing from the walls as in Figure 3.4(b) which shows a snapshot of dAMP
which is being adsorbed to a wall.

Figure 3.4: Snapshots of dAMP near a wall. (a) The nucleobase and
sugar of dAMP are adsorbed. (b) The sugar is detached and the
nucleobase is inclined with Axis 1 (see Figure 3.2) forming an angle $\theta$
with the wall plane during adsorption.
Figure 3.5: The angle between Axis 1 (see Figure 2) of the nucleobase and the wall surface as a function of $d_w$ while the dNMP was adsorbing (red) and while the dNMP was desorbing (blue). $E = 0.1$ V/nm.

For the smooth wall cases the orientations of the dNMP nucleobases relative to the wall surface planes and relative to the direction of the electric field were calculated during adsorption and desorption. Specifically, the angles between Axis 1 defined in Figure 3.2 with the wall surface planes and with the electric field direction were calculated. The periods for adsorption were defined to be from midway between the previous desorption time and an adsorption time until midway between that adsorption time and the next desorption time. The periods for desorption were defined to be from midway between the previous adsorption
time and a desorption time until midway between that desorption time and the next adsorption time.

Figure 3.6: The angle between Axis 1 of the nucleobase and the wall surface as a function of \(d_w\) while the dNMP was adsorbing (red) and while the dNMP was desorbing (blue). \(E = 0\text{V/nm}\).

The average angles (\(\theta\)) of Axis 1 (see Figure 3.2) of the nucleobase parts of the dNMPs with the wall surfaces are plotted in Figure 3.5 as a function of distance from the nearest wall surface (\(d_w\)) for \(E = 0.1\ \text{V/nm}\). The behavior is similar for the electric field strengths of 0.0 and 0.0144 V/nm as shown in Figure 3.6 and Figure 3.7, indicating that the electric field does not affect the mechanism.
of adsorption and desorption, at least at the lower field strengths that were studied. Negative values mean that the end of the nucleobase attached to the sugar is further away from the wall than the opposite end; Axis 1 points towards the wall plane. For $d_w > 0.9$ nm, the average angles for the four dNMPs are nearly 0. For $d_w < 0.9$ nm, dAMP, dCMP, and dTMP have minima in $\theta$ for both adsorption and desorption, indicating that the sugar adsorbs after the nucleobase and desorbs before the nucleobase.

Figure 3.7: The angle between Axis 1 of the nucleobase and the wall surface as a function of $d_w$ while the dNMP was adsorbing (red) and while the dNMP was desorbing (blue). $E = 0.0144$ V/nm.
The magnitude of the minima follows the order dCMP < dAMP < dTMP, which is consistent with the order of hydrophobicity of their nucleobases.\textsuperscript{112, 113} For $d_w < 0.9$ nm, dGMP has a maximum in $\theta$ for both adsorption and desorption, indicating that the sugar adsorbs before the nucleobase and desorbs after the nucleobase. The different behavior of dGMP is likely due to the fact that its nucleobase is the least hydrophobic.\textsuperscript{112, 113} The adsorption and desorption curves for each type of dNMP are similar to each other which indicates that adsorption and desorption occur in a similar way.

![Figure 3.8](image)

Figure 3.8: The average angles ($\psi$) between Axis 1 of the nucleobases (see Figure 3.2) of dNMPs with the electric field direction as a function of $d_w$ during dNMP adsorption.
Figure 3.9: The average angles ($\psi$) between Axis 1 of the nucleobases of dNMPs with the electric field direction as a function of $d_w$ during dNMP desorption.

Figure 3.8 shows the average angle ($\psi$) between Axis 1 (see Figure 3.2) of the nucleobases of dNMPs with the electric field direction during adsorption as a function of $d_w$. The plots for all dNMPs for desorption are shown in Figure 3.9. For an electric field of 0.0 V/nm far from the wall, $\psi$ is nearly 90° since there is no reason for any orientation to be favored and $\psi$ is defined between 0° and 180°. The dNMPs have a net charge of -2e on the phosphate group, therefore for nonzero electric fields the phosphate group is pulled on average in the opposite direction compared to the electric field direction. This means that Axis 1 is more likely to point in the direction of the electric field, and $\psi$ is expected to be less
than 90° far from the wall. This is not noticeable in most of the plots for $E = 0.0144$ V/nm, but the average Axis 1 angle is clearly less than 90° for $E = 0.1$ V/nm far from the wall. This effect is smallest for the dNMPs with the largest nucleobases, dAMP and dGMP, and largest for the dNMP with the smallest nucleobase, dCMP. During adsorption and desorption when the nucleobases are very close to the wall, the nucleobase-wall interactions can make $\psi$ far from 90°. The direction of this deviation depends on the identity of the dNMP, whether it is adsorbing or desorbing, and even on the magnitude of the electric field.

Various statistics were calculated related to the adsorption of the dNMPs to the slit walls. The fraction of the total time that the dNMPs spent adsorbed, the frequency of dNMP adsorption events are shown in Figure 3.10 and the mean times per adsorption event, the mean distances the dNMPs traveled in direction of the driving force per adsorption event, and the mean dNMP velocities in the direction of the driving force while adsorbed are shown in Figure 3.11. Mean times per desorption event, mean distances the dNMPs traveled in the direction of the driving force per desorption event, and the mean dNMP velocities in the direction of the driving force while desorbed are shown in Figure 3.12.

Fractions of the total time that the dNMPs were adsorbed to the slit walls ($f_{t,ads}$) are shown in Figure 3.10(a). The fraction of time adsorbed is larger for the more hydrophobic nucleotides (dAMP and dTMP) and generally decreases with increasing electric field strength. Rough walls reduce the fraction of the time adsorbed for dAMP, dCMP, and dGMP compared to the smooth walls. However, the rough walls have little effect on $f_{t,ads}$ for dTMP. The dNMPs with the smaller
Figure 3.10: Statistics for the dNMPs while adsorbed. (a) Fractions of the total time that the dNMPs were adsorbed to the slit walls. (b) The frequency of adsorption events.

Pyrimidine nucleobases (dCMP and dTMP) are affected less by the roughness than the dNMPs with the larger purine nucleobases (dAMP and dGMP). As discussed previously, the nucleobase parts of the dNMPs tend to sit almost flat on the wall. When the wall is rough, the nucleobases have fewer locations where that is possible leading to more frequent desorption from the wall. This effect is
larger for dNMPs with larger nucleobases. The dTMP nucleobase is the most hydrophobic of the nucleobases. Therefore even if it desorbs easier from the rough walls than the smooth walls it will still re-adsorb very quickly so its fraction of time adsorbed is nearly the same as with smooth walls. This behavior is reflected in its much higher frequency of adsorption on rough walls compared to smooth walls which is shown in Figure 3.10(b) and discussed below.

The frequencies of adsorption events \( (F_{\text{ads}}) \) are shown in Figure 3.10(b). For the smooth walls, higher frequency is correlated with a lower fraction of time adsorbed (see Figure 3.10(a)) except for dGMP which has about the same frequency as dTMP. \( F_{\text{ads}} \) is generally higher for the rough walls compared to the smooth walls with the same electric field strength while \( f_{\text{t,ads}} \) is generally smaller for the rough walls. For dTMP, the adsorption frequency is significantly higher with the rough walls compared to the smooth walls which nearly overcome the fact that it is easier to desorb from the walls in the rough wall case so dTMP has almost the same fraction of time adsorbed with rough and smooth walls. Frequency increases with increasing electric field strength; the frequencies for \( E = 0.3 \) and 0.6 V/nm are significantly higher than the frequencies for lower field strengths.

Mean times per adsorption event \( (t_{\text{ads}}) \) are shown in Figure 3.11(a). The trends are the same as for the fractions of time adsorbed. The mean time per adsorption event generally decreases with increasing electric field strength. Adding roughness to the slit walls caused a decrease in \( t_{\text{ads}} \) consistent with the increase in \( F_{\text{ads}} \).
The mean velocities of the dNMPs along the direction of the driving force on them (-x) while adsorbed to the slit walls (v_{ads}) are shown in Figure 3.11(b). The velocities of the different dNMPs at a given electric field strength are similar. This means that despite the differences in the dNMP-wall interactions in the wall normal direction,^50 the interactions in the wall tangential directions are nearly the same for all dNMPs. The differences in the dNMP nucleobase sizes and in the mean sodium-dNMP association numbers (discussed later) also affect the velocities, but these effects are small. The velocities increase with increasing electric field strength. Making the slit walls rough reduces v_{ads} by a factor of about 1.5 to 2.

Mean distances traveled by the dNMPs in the direction of the driving force on them (-x) while adsorbed to the slit walls (d_{ads}) are shown in Figure 3.11(c). The effect of the electric field strength on this distance is influenced by the competing effects of the increasing dNMP velocity while adsorbed and the decreasing time per adsorption event with increasing electric field strength. Mean distances per adsorption event of the dNMPs at E = 0.1 V/nm are higher than at E = 0.0144 V/nm. However, the distances at E = 0.3 V/nm and 0.6 V/nm are not much larger than at E = 0.1 V/nm; the increasing dNMP velocity while adsorbed is offset by the decreasing time per adsorption event. As with the time per adsorption event, the rough walls decrease d_{ads} for dAMP, dCMP, and dGMP by much more than for dTMP.

The mean times per desorption event (t_{des}) are shown in Figure 3.12(a). The mean time per desorption event is not significantly affected by changing the
electric field strength since it depends on the motion of the dNMPs in the wall normal direction and their adsorption and desorption dynamics, which as mentioned previously are not affected much by the electric field strength. Adding roughness to the slit walls also generally had little effect on $t_{\text{des}}$. The order for $t_{\text{des}}$ is usually $\text{dTMP} < \text{dAMP} < \text{dCMP} < \text{dGMP}$, which is the reverse of the order for the hydrophobicity of the nucleobases, $\text{G} < \text{C} < \text{A} < \text{T}$.\textsuperscript{112,113} It is likely that greater hydrophobicity leads to a greater capture probability, the probability that when a dNMP comes near a wall that it will adsorb instead of bouncing off. A greater capture probability corresponds to a lower mean time per desorption event.

The mean distances traveled by the dNMPs in the direction of the driving force on them ($-x$) while desorbed from the slit walls ($d_{\text{des}}$) are shown in Figure 3.12(b). The distances increase with increasing electric field strength. As with the time while desorbed, $d_{\text{des}}$ is affected by the hydrophobicity of the dNMPs.

The mean velocities of the dNMPs along the direction of the driving force on them ($-x$) while desorbed from the slit walls ($v_{\text{des}}$) are shown in Figure 3.12(c). As expected, the velocities are nearly the same for all the dNMPs at a given electric field strength. Any differences are due to the differing size of the dNMP nucleobases and the differences in the mean sodium-dNMP association numbers which are discussed later. The velocities of the dNMPs increase with increasing electric field strength. The walls have little effect on the dNMPs when they are desorbed, so $v_{\text{des}}$ for the rough and smooth walls at $E = 0.1$ V/nm is approximately the same as expected.
Figure 3.11: Statistics for the dNMPs while adsorbed. (a) The mean times per adsorption event. (b) Mean velocities of the dNMPs along the direction of the driving force on them (-x) while adsorbed to the slit walls. Note that the vertical axis is logarithmic. (c) Mean distances traveled by the dNMPs in the direction of the driving force on them (-x) while adsorbed to the slit walls.
Figure 3.12: Statistics for the dNMPs while desorbed. (a) The mean times per desorption event. (b) Mean distances traveled by the dNMPs in the direction of the driving force on them (-x) while desorbed from the slit walls. (c) Mean velocities of the dNMPs along the direction of the driving force on them (-x) while desorbed from the slit walls. Note that the vertical axis is logarithmic.
3.3.2 dNMP Phosphate Association with Sodium Ions

Negatively charged DNA strongly interacts with positively charged counter ions which can form a stable layer near DNA in solution and play an important role in altering DNA structure and modulating the interaction between DNA and other molecules. The counter ion layer is too stable to be removed by conventional deionization methods. However recent research found that an electric field can promote the dissociation of the DNA-ion complex during capillary electrophoresis and as an irregular DNA migration velocity arises. The interaction of cationic counter ions with the dNMPs is important since tightly bound ions can effectively reduce or even reverse the sign of the charge on the nucleic acid – counter ion complex which reduces or reverses the sign of the force on the complex due to the electric field. For example, smaller alkali metal cations bind to DNA more strongly; Li$^+$ > Na$^+$ > K$^+$. Kowalczyk et al.\textsuperscript{93} exploited this by using Li$^+$ instead of Na$^+$ or K$^+$ to slow down the motion of single stranded DNA through a nanopore. Our results presented below show that transient binding of sodium ions to the dNMPs leads to irregular dNMP velocities.

The phosphate groups of the dNMPs strongly interact with sodium cations in solution and form associations with 1, 2, and sometimes 3 sodium ions. This strong association is shown in Figure 3.13, which is a plot of the radial distribution function, g(r), between the sodium cations and the phosphorus atom on the phosphate group of dAMP. The four dNMPs have similar distribution curves, indicating that their different nucleobase parts have little effect on the association. All the distributions have major peaks with magnitudes between
about 415 and 610, meaning that a very stable Na\(^+\) shell is formed around the phosphate group. All the major peaks occur at around \(r = 0.27\) nm. There are small secondary peaks around \(r = 0.5\) nm. The major peak widths are not affected by the electric field at all, but the peak height generally decreases with increasing electric field strength; the peaks are flattened slightly.

![Figure 3.13: Radial distribution functions between the sodium cations and the phosphorus atom in the phosphate group of dAMP. The abscissa \(r\) is the distance between the phosphorus atom and a sodium cation. Black solid lines, red dashed lines and blue dotted lines are, respectively, for the electric fields of 0.0 V/nm, 0.0144 V/nm and 0.1 V/nm.](image)

To capture the entire major peak but not the secondary peak of \(g(r)\), the cutoff distance to decide whether a Na\(^+\) ion was associated with the phosphate group was chosen to be 0.4 nm. Use of a single cutoff distance leads to some very short association events because a sodium ion can cross outside the cutoff distance and then right back inside it, or the reverse. This was avoided by only counting association states which lasted longer than a minimum time of 0.4 ns. This time removed the noise due to the single cutoff distance, yet was still less
than the mean time after a sodium dissociated from a dNMP until it passed the dNMP again through the periodic boundaries for electric field strengths of 0.0144 and 0.1 V/nm; use of this minimum time was unlikely to cause two association events to be counted as one. Use of a minimum time of 0.4 ns is equivalent to doing a 0.4 ns moving average on the association number trajectory followed by rounding to the nearest integer. Application of the minimum time has little effect on the mean association number, but affects dynamical quantities such as the mean relaxation time for association.

Figure 3.14: Association number of sodium cations with the dNMP phosphorus atom (red), and the -x coordinate of the center of mass of dNMPs (blue) as a function of time. The electric field is 0.1 V/nm.

The association number versus time for one simulation for each dNMP with $E = 0.1$ V/nm is shown in Figure 3.14. The coordinates of the center of mass of the dNMPs in the direction of the driving force (-x) are also plotted in the same
figures. The figures show that the traveling velocity of dNMPs (i.e., the slope of x-coordinate versus time curve) is strongly influenced by the association formed between Na$^+$ and the dNMPs. Take dGMP for example and compare three time periods: from 10 ns to 20 ns the x coordinate increases linearly with time and the association number is 1; from 20 ns to 36 ns the x coordinate remains nearly the same and the association number is 2; from 50 ns to 60 ns the x coordinate increases in time with a steeper slope than the first time period and the association number is generally 0, occasionally 1. The dNMPs were driven by the electric field so the driving force was proportional to the net charge on them. With no Na$^+$ bound, the net charge of the phosphate group was \(-2e\). With one Na$^+$ ion bound to the phosphate group, the net charge of the aggregate was \(-1e\) and thus the driving force was reduced to half. If two Na$^+$ ions were associated simultaneously, the aggregate was electrically neutral and the driving force became zero, thus the x coordinate remained the same in that time period.

![Figure 3.15: Mean association numbers for the dNMPs.](image)
Despite the effect of the individual sodium-phosphate association events on the dNMP velocity, overall the association does not have a significant effect since the mean association numbers are not affected much by the electric field strength or the roughness of the walls as shown in Figure 3.15. Even after averaging over all dNMPs for each case, the error bars still overlap for all cases; none of the differences are significant. There appears to be a small decrease in the association number averaged over all dNMPs when increasing the electric field strength from 0.0 to 0.1 V/nm. This is to be expected since the electric field pulls the sodium ions and dNMPs in opposite directions which is not favorable for association. The association numbers for dTMP at electric fields $E = 0.3$ and 0.6 V/nm appear to be slightly higher than for $E = 0.1$ V/nm, but this may be an artifact of the periodic boundary condition; once a sodium ion dissociates from a dNMP, it does not take long before they pass each other again since they are moving at relatively high velocities in opposite directions which gives them little time to diffuse out of the path of each other and possibly increasing the probability of re-association.

Figure 3.16(a) shows the distribution of the $\text{Na}^+\text{-P}$ association numbers (AN) averaged over all four dNMPs for $E = 0.0$, 0.0144, and 0.1 V/nm with smooth walls. The rough wall case with $E = 0.1$ V/nm is nearly identical to the smooth wall case with $E = 0.1$ V/nm, so the results for the rough wall are not shown. For all cases the order of the probabilities is $\text{AN}=1 > \text{AN}=0 > \text{AN}=2 > \text{AN}=3$. The probability of $\text{AN}=3$ is much lower than the other three states and doesn’t have a significant influence to the traveling velocity of the dNMPs. The probability of the
AN=0 increases slightly and the probability of AN=1 decreases slightly with increasing E leading to the overall decrease in association number with increasing E.

Figure 3.16: The distribution of the association numbers for (a) all of the dNMPs for the smooth wall cases with different electric fields, and for (b) the averages over the dNMPs with purine bases (dCMP and dTMP) and over the dNMPs with pyrimidine bases (dAMP and dTMP) with smooth walls and E = 0.1 V/nm.

Figure 3.16(b) shows the distribution of the Na\(^+\)-P association numbers averaged over the dNMPs with purine and pyrimidine nucleobases in the case of
smooth walls with $E = 0.1$ V/nm. The dNMPs with pyrimidine nucleobases (dCMP and dTMP) have a higher probability of $AN=0$ and lower probabilities of $AN=1$ and $AN=2$ compared to the dNMPs with purine nucleobases (dAMP and dGMP) leading to the overall higher association number for dAMP and dGMP which can also be seen in Figure 3.15. Other cases show similar behavior.

The relaxation times for the association were estimated by integration of an aggregate existence autocorrelation function (AEACF) which is a generalization of the dimer existence autocorrelation function (DACF) described by Brehm and Kirchner. Since there is only one dNMP in our system, and therefore only one possible aggregate, the AEACF can be written in a simpler form.

$$
AEACF(\Delta t) = \frac{\sum_{i=0}^{T_{\text{max}}-\Delta t} \beta(t+\Delta t)\beta(t)}{\sum_{i=0}^{T_{\text{max}}} \beta(t)}
$$

(3.1)

The definition of $\beta$ was also slightly different. For an association number of 1, the value of $\beta(t)$ was 1 if the association number was $\geq 1$ at time $t$, and $\beta(t+\Delta t)$ became 0 once the association number became 0 and remained 0 for all subsequent $\tau$. For an association number of 2, the value of $\beta(t)$ was 1 if the association number was $\geq 2$ at time $t$, and $\beta(t+\Delta t)$ became zero once the association number became 1 and remained 0 for all subsequent $\tau$. For each association event, the numerator in Equation (1) contributes $N_i - \Delta t / \Delta t_{\text{traj}}$ to the sum, where $N_i$ is the duration of event $i$ in number of trajectory time steps of length $\Delta t_{\text{traj}}$. The denominator is equivalent to $\sum_{i=1}^{r_{\text{max}}} N_i$, which is just the sum of the
duration of all events and is needed to normalize so that $AEACF(0) = 1$. The
AEACF can therefore be thought of in terms of an average over events. For an
event for an association number of 1 or 2, defined between the times that the
association number changed from 0 to 1 or from 1 to 2 and the next time that the
association number changed from 1 to 0 or from 2 to 1, the contribution to the
AEACF was a line starting at a value of 1 at $\tau = 0$ and ending at 0 at the end of
the event. The AEACF was the average of all the lines corresponding to each
event.

$$AEACF(\Delta t) = \frac{1}{n_{\text{events}}} \sum_{i=1}^{n_{\text{events}}} \left( -\frac{\Delta t}{T_i} + 1 \right) H \left[ -\frac{\Delta t}{T_i} + 1 \right]$$ \hspace{1cm} (3.2)

$$AEACF_{\text{fit}}(t) = A_1 \exp \left[ -t/T_1 \right] + A_2 \exp \left[ -t/T_2 \right] + A_3 \exp \left[ -t/T_3 \right]$$ \hspace{1cm} (3.3)

$$t_r = \int_0^{\infty} AEACF_{\text{fit}}(t) dt = A_1 T_1 + A_2 T_2 + A_3 T_3$$ \hspace{1cm} (3.4)

In Equation (3.2), $T_i$ is the time duration of event $i$ and $H$ is the Heaviside step
function. This allowed the events from multiple simulations to be combined to
calculate single AEACFs for each dNMP. Once an AEACF was obtained it was fit
to a function as shown in Equation (3.3). The relaxation time is obtained by
integrating the fitted AEACF function in Equation (3.4). Using the fitted function
instead of the original AEACF function is because that it might underestimate the
value of the integration due to its early decay to zero at the longest association
event. Figure 3.17 shows AEACF curves for all the dNMPs for $E = 0.1$ V/nm and
an association number of 1.
Figure 3.17: Aggregate existence autocorrelation functions (AEACFs) for the dNMPs with $E = 0.1$ V/nm and an association number of 1.

Figure 3.18: The mean relaxation times for a Na$^+$-P association number of (a) 1 and (b) 2.
Figure 3.18(a) shows the relaxation times for an association number of 1. The relaxation time generally decreases with increasing E because the electric field tends to pull the oppositely charged dNMPs and sodium ions apart. The rough wall case is similar to the smooth wall case at E = 0.1 V/nm except with dTMP. Figure 3.18(b) shows the relaxation times for an association number of 2. There is not much difference in the relaxation times for E = 0.1 V/nm and lower for both the rough and smooth walls. The relaxation time decreases for E = 0.3 and 0.6 V/nm.

3.3.3 dNMP Velocities

The mean velocities of the dNMPs give an indication of how well separated the flight time distributions will be. The mean velocities of the dNMPs in the nanoslit as a function of electric field strength and wall roughness are shown in Figure 3.19. At the lowest electric field of E = 0.0144V/nm, the velocities of dNMPs are nearly equal except for dGMP although the dGMP velocity is still within the uncertainties of the other velocities. For dTMP, four different electric field strengths were used and the results show that the velocities increase with electric field strength. The relationship between the average velocity and electric field is not linear, although the driving force on the dNMPs is linearly proportional to electric field. As discussed above, both adsorption to the walls and association between phosphate group and Na\(^+\) ions also affect the traveling velocity of dNMPs. The super-linear increase in the velocity with increasing electric field strength is primarily due to reduction in the fraction of time adsorbed (see Figure
3.10(a). A lower fraction of time adsorbed means a larger velocity since the dNMP is slowed down less due to contact with the walls.

![Graph](image)

Figure 3.19: The average overall velocities of the dNMPs. Note that the vertical axis is logarithmic.

The large uncertainties in the velocities arise for several reasons. The first is that the velocities are much smaller than the thermal velocities especially for the lower electric fields. Simulation times for the smallest electric field would have to be much longer to reduce the uncertainties to the same relative uncertainty of around 10 percent or less obtained for the larger electric fields. A second factor contributing to the variability in the velocities is the variability in the time of each association event between the dNMP phosphate group and Na\(^+\) cations. From Figure 3.17 we can see that the time of one association event ranges from less than 1 ns to 20 ns. The variability in the adsorption times and in the velocity while adsorbed due to the heterogeneity of the wall surface also contributes to variability in the velocities. The last two factors could also lead to biases as well if there are not enough association events or not enough adsorption events observed in the simulations. Of particular concern are that there are only a
limited number of time periods during which the association number is equal to 2, and that especially for dTMP there are only a limited number of adsorption events.

3.3.4 Distance to Separate the dNMPs and Analysis of the Time of Flight per dNMP

In our previous work, the required channel length to achieve reliable separation of the time of flight distributions of the dNMPs was calculated from the distributions of the times of flight over 0.5 nm segments. These times of flight were calculated for each simulation trajectory by starting from the first time step in the trajectory after equilibration, calculating the first time that a dNMP had advanced 0.5 nm beyond its initial position in the direction of the driving force, then repeating the process starting from the next time step in the trajectory after the end of the previous 0.5 nm segment until the end of the trajectory was reached. However, it is just as valid to start at the second, third, etc. time steps after equilibration. This would lead to different sets of times of flight and different estimates of the required channel length. In addition, using only the single sets of flight times determined from the simulation are not the best estimates of the distributions that would be obtained if in fact the simulations could be extended until the actual distributions no longer overlapped. To fix the problems mentioned above, some modifications to the calculation were made. The distributions of the simulation mean dNMP velocities estimated using the moving block bootstrap method were used instead of the time of flight distributions. Since the distributions of the mean velocities were nearly normal, a multiple of the standard deviation could be used as an estimate of the distribution widths instead of the
distance between where the cumulative distribution function was equal to \((1-\text{separation efficiency})/2\) and \((1+\text{separation efficiency})/2\). The required distance for separating any pair of dNMPs \((\alpha \text{ and } \beta)\) is \(N_\alpha d_\alpha\) or \(N_\beta d_\beta\). \(N_\alpha\) can be determined using the Equation 3.5:

\[
N_\alpha = Z^2 \frac{s_{v_\alpha} + s_{v_\beta} \sqrt{\frac{(v_\beta)\Delta T_\beta}{(v_\alpha)\Delta T_\alpha}}}{\left(\langle v_\alpha \rangle - \langle v_\beta \rangle\right)^2}
\]  

(3.5)

Table 3.1 shows the required distances for separation of the dNMP mean velocity distributions to 3 standard deviations. The maximum value is the minimum required channel length for that level of accuracy. These minimum channel lengths are about 166 μm, 107 μm, and 242 μm for the smooth wall case with \(E = 0.0144\) V/nm, smooth wall case with \(E = 0.1\) V/nm, and rough wall case with \(E = 0.1\) V/nm, respectively. However, these values are very sensitive to the difference in the estimated overall mean velocities for the pair of dNMPs that determine them.

Since the estimated distribution of the mean velocities can be approximated as normal, the flight time distributions can be easily derived from the velocity distributions. The flight time probability density functions for smooth walls with \(E = 0.0144\) V/nm and \(E = 0.1\) V/nm and rough walls with \(E = 0.1\) V/nm are shown in Figure 3.20.

The minimum required analysis times per dNMP can be estimated from the distance between the lower edge of the distribution with the shortest flight times.
and the upper edge of the distribution with the longest flight times. These edges are determined by the accuracy required. The reason that this determines the minimum analysis time is that if a dNMP with the longest flight times pass through the channel followed by one with the shortest flight times, it cannot be allowed to pass the previous dNMP; the distributions of the times when the two dNMPs pass the last sensor measured from the beginning of the analysis must be sufficiently separated. The required analysis times are shown in Figure 3.20.

Table 3.1: Required distances (μm) to separate the time of flight distributions of the dNMP pair types to Z = 3 standard deviations from the means of the distributions for smooth walls with E = 0.1 V/nm and E = 0.0144 V/nm, and rough walls with E = 0.1 V/nm. The longest distances which are in bold are the minimum required channel lengths.

<table>
<thead>
<tr>
<th>E = 0.0144 V/nm</th>
<th>dAMP</th>
<th>dCMP</th>
<th>dGMP</th>
<th>dTMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>dAMP</td>
<td></td>
<td>165.8</td>
<td>1.2</td>
<td>0.2</td>
</tr>
<tr>
<td>dCMP</td>
<td>165.8</td>
<td></td>
<td>1.3</td>
<td>52.2</td>
</tr>
<tr>
<td>dGMP</td>
<td>1.2</td>
<td>1.3</td>
<td></td>
<td>0.4</td>
</tr>
<tr>
<td>dTMP</td>
<td>0.2</td>
<td>52.2</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E = 0.1 V/nm (smooth)</th>
<th>dAMP</th>
<th>dCMP</th>
<th>dGMP</th>
<th>dTMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>dAMP</td>
<td></td>
<td>9.6</td>
<td>105.2</td>
<td>49.1</td>
</tr>
<tr>
<td>dCMP</td>
<td>9.6</td>
<td>6.6</td>
<td></td>
<td>26.1</td>
</tr>
<tr>
<td>dGMP</td>
<td>105.2</td>
<td>6.6</td>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td>dTMP</td>
<td>49.1</td>
<td>26.1</td>
<td>12.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E = 0.1 V/nm (rough)</th>
<th>dAMP</th>
<th>dCMP</th>
<th>dGMP</th>
<th>dTMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>dAMP</td>
<td></td>
<td>9.5</td>
<td>16.9</td>
<td>1.4</td>
</tr>
<tr>
<td>dCMP</td>
<td>9.5</td>
<td></td>
<td>241.9</td>
<td>3.7</td>
</tr>
<tr>
<td>dGMP</td>
<td>16.9</td>
<td>241.9</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>dTMP</td>
<td>1.4</td>
<td>3.7</td>
<td>3.1</td>
<td></td>
</tr>
</tbody>
</table>
In Figure 3.20, the order of the distributions is different in all three cases. For smooth walls with $E = 0.1$ V/nm, the dNMPs with the smaller pyrimidine nucleobases (dCMP and dTMP) move faster than the dNMPs with the larger purine nucleobases (dAMP and dGMP). The order of the flight time distributions, \( \text{dCMP} < \text{dTMP} < \text{dAMP} < \text{dGMP} \), is also the order of the surface area of the nucleobase part of the dNMPs and the order of the magnitude of the potential energy between the dNMPs and the wall calculated in previous simulations. This would suggest that the order is simply determined by the interactions of the dNMPs with the walls, those with stronger interactions move slower when adsorbed leading to slower overall velocities. The trend in the velocities while adsorbed is \( \text{dTMP} > \text{dCMP} \approx \text{dGMP} > \text{dAMP} \) (see Figure 3.11(a)), but if this simple explanation were true, then the expected order would be \( \text{dCMP} > \text{dTMP} > \text{dAMP} > \text{dGMP} \). The most important additional factor is the fraction of time adsorbed (see Figure 3.10(a)) which is related to the hydrophobicity of the dNMP nucleobases.\(^{112, 113} \) dTMP is more hydrophobic than dCMP and spends a much larger fraction of time adsorbed than dCMP, so its overall velocity is lower than for dCMP even though it moves faster than dCMP while adsorbed. A similar argument holds for dGMP and dAMP; dAMP is more hydrophobic leading to a swapping in the order of their overall velocities compared to their velocities while adsorbed. For smooth walls with $E = 0.0144$ V/nm, the statistics are too poor and the order of the dNMPs may not be accurate. It may be that dGMP gets stopped on the wall more often while adsorbed instead of sliding which could lead to its much longer flight times than the other dNMPs. However, it is unclear why
dAMP would have the shortest flight times. For rough walls with \( E = 0.1 \) V/\( \text{nm} \), the order of the dNMPs with purine nucleobases and pyrimidine nucleobases is switched compared to the case of smooth walls with \( E = 0.1 \) V/\( \text{nm} \); the order of the flight times is \( \text{dAMP} < \text{dGMP} < \text{dCMP} < \text{dTMP} \). This switching in the order is likely because the change in the fraction of time adsorbed (see Figure 3.10(a)) decreases more in the rough wall case compared to the smooth wall case for dNMPs with purine nucleobases than for the dNMPs with pyrimidine nucleobases. The likely reason for this is simply geometric; the nucleobases have fewer locations where they can adsorb strongly to the rough walls and this effect is greater for larger nucleobases. This is discussed earlier in the context of the fraction of time adsorbed.

It is desirable to have all the time of flight distributions about the same distance from each other to keep the minimum analysis time per dNMP small. Figure 3.20 shows that this is the case for a smooth wall with \( E = 0.1 \) V/\( \text{nm} \) and the analysis time is about 9 \( \mu \text{s} \). For the rough walls with \( E = 0.1 \) V/\( \text{nm} \), the distributions for dAMP and dTMP are not close to those for dGMP and dCMP and the analysis time is longer (about 46 \( \mu \text{s} \)). For a smooth wall with \( E = 0.0144 \) V/\( \text{nm} \), the distribution for dGMP is very far from the other three, and the analysis time is about 582 \( \mu \text{s} \). As with the channel length, the minimum analysis time per dNMP is also sensitive to the difference in the estimated mean velocities for the pair of dNMPs that are hardest to separate. Using the analysis mentioned previously in the context of channel length for the rough wall case gives a minimum value of about 15 \( \mu \text{s} \), a maximum value of about 341 \( \mu \text{s} \), and a median
value of about 45 μs. There may be a significant difference between the E = 0.0144 V/nm case and the cases with E = 0.1 V/nm.

Figure 3.20: Probability density functions (PDFs) of the times of flight for each dNMP for (a) E = 0.0144 V/nm with smooth walls, (b) E = 0.1 V/nm with smooth walls, and (c) E = 0.1 V/nm with rough walls. The minimum required analysis times per dNMP are also shown on the plots.
3.4 Conclusion

The electrophoretic transport of dNMPs in 3 nm wide slits composed of Lennard-Jones carbon atoms was studied using molecular dynamics simulations. The electric field strength (E) was varied, E = 0.0, 0.0144, 0.1, 0.3, or 0.6 V/nm, with atomically smooth, but disordered slit walls. In one case with E = 0.1 V/nm, slit walls with an RMS roughness on the order of the size of the dNMPs were also used. Quantities of interest related to the interactions of the dNMPs with the slit walls and the sodium ions in solution. Also of interest were the minimum channel lengths and analysis times per dNMP required to separate the time of flight distributions of the dNMPs to obtain a desired error rate for sequencing.

The dNMP trajectories consisted of multiple adsorptions and desorptions to and from the slit walls with the dNMPs tending to adsorb with their nucleobase groups nearly flat on the surface. The orientations of the dNMP nucleobase groups relative to the wall surfaces as a function of distance from a slit wall during the adsorption and desorption processes were similar. The orientations were also similar as a function of E for E = 0.0, 0.0144, and 0.1 V/nm. This indicates that the mechanism of adsorption and desorption is not affected by the electric field if E is small enough. The orientation of the dNMPs relative to the direction of the electric field was influenced by the electric field. The dNMP is dragged by the negatively charged phosphate group, and with a large enough E, 0.1 V/nm, this effect is significant since the force due to the electric field is significant compared to thermal fluctuations.
Statistics related to adsorption and desorption were computed. The increasing electric field with smooth walls decreased the fraction of the total time that the dNMPs were adsorbed to the walls, decreased the mean time per adsorption event, increased the frequency of adsorption events, and increased the velocity of the dNMPs while adsorbed. Increased driving force made desorption more likely and sped up the sliding of the dNMPs on the slit walls. The mean distance traveled by the dNMPs in the direction of the driving force while adsorbed increased with increasing E up to E = 0.3 V/nm, but was about the same for E = 0.3 and 0.6 V/nm due to the competing effects of increasing velocity while adsorbed and decreasing time per adsorption event with increasing E. Using rough walls generally made desorption more likely (decreased mean times and distances per adsorption event) and slowed down the sliding of the dNMPs on the walls (decreased velocity while adsorbed). The fractions of the total time that the dNMPs were adsorbed decreased for all dNMPs, but not significantly for dTMP. This was due to the large increase in the frequency of adsorption events for dTMP; its high hydrophobicity made re-adsorption after desorption very fast. In general, the frequency of adsorption increased when using rough walls. The mean time per adsorption event for dTMP also decreased less than for the other dNMPs when comparing smooth and rough walls due to its high hydrophobicity. The velocity of the dNMPs in the direction of the driving force while desorbed was not a function of dNMP or wall roughness as expected. The mean time and distance per desorption event were not a function of E or wall roughness, but decreased with increasing hydrophobicity of the nucleobase parts of the dNMPs.
due to faster re-adsorption for more hydrophobic dNMPs. The dNMP-wall interactions affected the frequency and duration of the dNMP adsorption and desorption periods as well as the dNMP velocities during those periods which helped to separate the dNMP time of flight distributions.

Transient ion association between the anionic phosphate group of the dNMPs and the sodium cations in solution was observed with the number of associated sodium ions varying from 0 to 3. The average number of associated sodium ions was around 1. Due to the transient nature of the association, the force on the dNMP-sodium complex due the electric field also varied leading to changes in the dNMP velocity ranging from very fast with an association number of 0, approximately zero with an association number of 2, and even having the opposite sign with an association number of 3 (rare). The mean association numbers appeared to be a function of the electric field strength, but the differences were not significant considering the uncertainties. The dNMPs with pyrimidine nucleobases (dCMP and dTMP) had slightly fewer associated sodium ions on average compared to the dNMPs with purine nucleobases (dGMP and dTMP). Sodium association had a large effect on the instantaneous dNMP velocities, but little effect on the mean velocities and therefore helped little with separating their time of flight distributions.

The order of the mean velocities or times of flight is affected by the rough walls compared to the smooth walls. For smooth walls, the order of the times of flight is dCMP < dTMP < dAMP < dGMP; the larger dAMP and dGMP are slowed down more when adsorbed. For rough walls, the order of the times of flight is
dAMP < dGMP < dCMP < dTMP; the dAMP and dGMP do not stay adsorbed as much and therefore move faster overall since there are fewer favorable adsorption sites for them on rough walls. Rough walls, or physically or chemically structured walls might be useful for improving the separation of the dNMP time of flight distributions.
CHAPTER 4  MOLECULAR DYNAMICS SIMULATION OF ELECTRICALLY DRIVEN DNA TRANSPORT THROUGH PMMA NANOSLITS

4.1 Introduction

In many DNA separation experiments using fluorescence lifetime discrimination techniques, the microchip electrophoresis devices are fabricated in poly-(methylmethacrylate) (PMMA). PMMA is a suitable substrate for fluorescence detection due to its low autofluorescence level. With the microchip devices fabricated in PMMA researchers have successfully separated a couple of biomolecules such as proteins, peptides and DNA\textsuperscript{120-122}. PMMA has been widely used in automobile, aerospace, coating and packaging due to its outstanding mechanical and thermal properties. The PMMA monomer is shown in Figure 4.1.

![PMMA monomer structure](image)

Figure 4.1: Structure of a PMMA monomer

4.2 Methodology

In this section we outline the system preparation and the simulation methodology. The first part describes the methodology used to generate the
PMMA slit system by creating two PMMA slabs. Finally the simulation details will be presented.

Figure 4.2: Molecular structure of a PMMA trimer. The group in green circle is regarded as the monomer which is replicated to build a PMMA chain. The two groups circled in red represent the head and tail groups.

4.2.1 Preparation of force field

As in the previous simulations the CHARMM27 force field was used for describing the interactions of dNMP and ions. The CHARMM TIP3P model was used for water. The parameters and partial charges of PMMA chains are derived from the parameters for a PMMA trimer. The molecule structure of a PMMA trimer is shown in Figure 4.2. In the newly created force field, rather than a single molecule like trimer, a PMMA molecule is considered as a polymer such as DNA or protein and consists of a certain number of residues and two ends. The residue is a monomer of PMMA which is circled in green in figure 4.2 and will be replicated a certain times and connected to each other to form the PMMA chain.
Then the two hydrogen atoms on the CH$_2$ group at one end of the chain will be deleted and three hydrogen atoms of different type are connected to the carbon atom. On the other end a methyl group is attached to the carbon atom on the backbone. The final structure of the PMMA molecule is composed of multiple monomers and two ends which are the same with that of the trimer. The topology and parameter files of PMMA can be found in the appendix.

4.2.2 Verification of force field

To verify the PMMA force field files we created based on the force field files for a trimer, we put a trimer in a unit cell and ran molecular dynamics with the original trimer force field in a NVT ensemble for a couple of time steps. Then the force field was changed to developed PMMA force field and other settings remained the same. Table 4.1 shows the potential energies obtained from these two simulations at the first 3 time steps. The values calculated using the two force fields are exactly the same. This means that the format of force field files we created is correct and can reproduce the same result as the original trimer force field.

![Figure 4.3: Molecular structure of a PMMA trimer in a NVE ensemble.](image)
Table 4.1: The potential energies of a trimer (Figure 4.3) computed using the original trimer force field and the PMMA force field.

<table>
<thead>
<tr>
<th>time (fs)</th>
<th>potential energy (Kcal/mole)</th>
<th>original</th>
<th>ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>630.3732</td>
<td>630.3732</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>461.17131</td>
<td>461.17131</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>402.65046</td>
<td>402.65046</td>
<td></td>
</tr>
</tbody>
</table>

4.2.3 Construction of PMMA slab

In a typical simulation, the computations extend over many molecules so as to average the computational results over a realistically large space domain. However the computer memory and speed are limited and therefore the simulated system have to be limited in size.

In our study the amorphous PMMA wall generated contains 5285 atoms and consists of 7 polymer chains with fifty monomer units per chain. The periodic boundary conditions are applied so that the unit cell is repeated in two dimensions to simulate the infinite slab geometry. At the beginning a sample of amorphous PMMA of a low density is constructed which is shown in figure 4.4(a). To construct PMMA sample as correctly and representatively as possible in a reasonable time, the dihedral angles formed by the atoms on the backbone are assigned randomly. The samples were then subjected to an energy minimization, followed by a “temperature annealing cycle” during which the temperature was raised from 300K to 600K and then kept for a short duration of 1ns.
Figure 4.4: Molecule structure of a PMMA trimer. Atomic snapshot from the molecular dynamics simulation used to compress the PMMA slabs at different levels. (a) Initially constructed to a low density of 0.3 g/cm$^3$ (b) compressed to the density of 0.6 g/cm$^3$ (c) compressed to the desired density of 1.2 g/cm$^3$

The annealing process is necessary because it can prevent the system from being trapped in a meta-stable state with local minimum of energy. The PMMA sample was then compressed from the top of the unit cell by a virtual reflecting wall while keeping the cell lengths in x and y directions constant. A reflecting wall will reflect particles when they try to move through the wall. The other reflecting
wall is placed at the bottom of the unit cell and fixed. At the beginning of the deformation the compression rate is 5m/s until the density of PMMA reaches 0.6 g/cm³. Then the compression rate is slowed down to 0.3 m/s to make sure the system can be relaxed and the simulation doesn’t crash. The snapshots in figure 4.4 show the deformation process.

4.2.4 Simulation details

Once the PMMA wall was constructed, the dNMPs, water and ions were added between two PMMA walls using VMD. The distance between the two walls is 4.5nm. A schematic of the simulation system is shown in Figure 4.5. The dimensions in x and y directions are 5nm and 5nm. The thickness of a wall is 2.3nm and the distance between the walls is 4.5nm. The solution consists of 1 dNMP, 4 sodium, 2 chloride ions and water molecules. The ions are driven by an electric field along the x direction. The carbon atoms in PMMA walls which are less than 0.4nm away from the top or bottom of the simulation box are constrained and the other atoms in PMMA walls move freely. In the simulation the walls atoms are thermalized at 300K using the velocity rescaling method. For each dNMP, 5 simulations were run with different PMMA wall configurations with the expectation that they represent a reasonable average for a larger molecular model. The three-dimensional particle mesh Ewald method corrected for slab geometry was used for calculating the long-range electrostatic interactions.
4.3 Results

4.3.1 Roughness of the PMMA walls

To identify the surfaces of the PMMA walls, a molecular-probe scanning method is used. A schematic of this method is shown in Figure 4.6. In this method, surface (solid line) is determined from the trajectory of the molecular probe (dashed line). This trajectory is obtained from the contact points between the molecular probe and polymers, which are specified by the calculation of the LJ interaction.
In the simulation system, the bottom surface of the upper PMMA wall and the top surface of the lower PMMA wall are exposed to water molecules and part of the surface atoms will stretch out of the surface. Thus the surfaces contacted with water become rough. The profile of the top surface of the lower PMMA wall is shown in Figure 4.7. The initial position of the surface is at -15Å. The height profile of the PMMA surface shows that the highest peak is about 10Å and a valley of about 5 Å deep also exists. The valleys can play an important role in the transportation of mononucleotides through nanochannels. The roughness of the walls doesn’t change much with time after it reaches equilibrium. Figure 4.8 shows the standard deviation of the height of the points on the surfaces of the two walls. The standard deviation of the height increases rapidly in 0.5 ns and then doesn’t change much. It means that the walls become rough in a short time and then the expansion ceases. Some water molecules can enter PMMA slabs due to the expansion of the surface.
Figure 4.7: A height profile of the top surface of the lower wall which is exposed to water molecules. The initial position of the surface is at -15 Angstroms.

Figure 4.8: The standard deviation of the height of the points on the surfaces of the two walls.
4.3.2 Adsorption on Walls and Association with Na$^+$

The x component, z component of the trajectories of the center of mass of dNMPs, and the association number between Na$^+$ and the phosphate group is shown in Figure 4.5. The curves of the z component of the trajectories shows that dNMPs are adsorbed and desorbed from the PMM walls and obviously have difference in the adsorption behavior: dCMP is hardly adsorbed and the duration of each adsorption event is very short, and dTMP is adsorbed on the walls for most of the time. The adsorption on PMMA walls plays an important role in the transport of dNMPs because due to the strong interactions with the walls the dNMPs will stop moving in x direction even if the driving force exists. For example at 400ns of dAMP, 130ns and 450ns of dGMP and the first 100ns of dTMP. The different z position of the adsorption events is due to the rough surface of the PMM walls. When dNMPs are in the state of desorption, the traveling velocity is mainly affected by the association.

The average association number over the 500ns simulation and the traveling distance of the dNMPs is shown in Table 4.2. The traveling distance of dCMP is much longer than the other three and can be easily separated, however its association number is in the middle of the dNMPs. Thus the separation of dCMP is mainly attributed to its weak adsorption on the walls.

<table>
<thead>
<tr>
<th></th>
<th>dAMP</th>
<th>dCMP</th>
<th>dGMP</th>
<th>dTMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asso. Num</td>
<td>1.546</td>
<td>1.368</td>
<td>1.49</td>
<td>1.297</td>
</tr>
<tr>
<td>Travel Dist (nm)</td>
<td>376</td>
<td>937</td>
<td>540</td>
<td>367</td>
</tr>
</tbody>
</table>
Figure 4.9: Transportation properties of dNMPs through the PMMA nanoslit as a function of time. Blue line represents the traveling distance of dNMPs in the direction of electric fields; green line represents the position of the center of mass of dNMP in the direction perpendicular to the wall surface; red line represents the number of sodium ions associated with the phosphate group on dNMPs. The position of the slit in z direction is from 2nm to 6.5nm.
4.4 Conclusion

Our MD simulation studies show that, while moving along the PMMA nanoslit the mononucleotides are adsorbed and desorbed from the walls multiple times. Due to their strong interaction with the PMMA walls the mononucleotides can be trapped in adsorbed state for hundreds of nanoseconds. When dNMPs are in the desorbed state, their traveling velocity along the axis of the nanochannel is mainly affected by the association between Na$^+$ and the phosphate group. The MD study suggests that with careful control devices build on PMMA nanochannels could be used for biopolymer sequencing.
CHAPTER 5  BROWNIAN DYNAMICS SIMULATION OF DNA TRANSPORT THROUGH PMMA NANOCHANNELS

5.1 Introduction

The advances in nanofluidic technology have enabled its application in low-cost DNA sequencing. However as the size of the nanofluidic systems decreases and the ratio of surface to volume increases, the interaction between solute and surface cannot be ignored. For example, although in the regions of fluid near surfaces, the atomic-scale properties of solutes and surfaces play an important role. Classical all-atom molecular dynamics is an appropriate method to study these atomic-scale phenomena on the surface of nanofluidic device. However, the high computational cost of molecular dynamics to achieve the atomistic accuracy makes it impractical to simulate the full-scale model of nanofluidic systems. Some proposed particle-based simulation methods increased the computation efficiency, however some precision in the evaluation of inter-atom forces is lost. In 2011 a Brownian dynamics method parameterized through all-atom molecular dynamics simulations was proposed to simulate the dynamics of a small solute in nanofluidic systems. This method is efficient to simulate experimental-scale nanochannels while retaining atomic-scale precision.

In this work Brownian dynamics is used to simulate the transport of a nucleotide driven by an electric field through nanoslits with surfaces of different characteristics. As a first step the free energy profile of nucleotide walls interaction is assumed to be described by a one-dimensional profile. As more information becomes available the methodology can be extended to account for
more complex and realistic representation of the nucleotide-wall interaction. The three-dimensional potential of mean force between the dTMP and the surface of the PMMA wall is the focus of our future research.

### 5.2 Brownian Dynamics Method

The transport behavior of a small solute through nanoslits includes the diffusion of the solute in the solvent, the drift of the solute with the solvent flow, the adsorption or desorption on/from the walls of the nanochannel, and the random motion caused by the collision with surrounding solvent molecules.

![Figure 5.1: An example of the models used in molecular dynamics (left) and Brownian dynamics (right)\textsuperscript{125}](image)

Different from all-atom molecular dynamics methods, the solvent and wall material of the walls are not simulated explicitly in Brownian dynamics. A comparison of the two models is shown in Figure 5.1\textsuperscript{125}. Thus the computational cost is greatly reduced and it is feasible to expand the simulation temporally from several microseconds to milliseconds and spatially from tens of nanometers to microns. However, some of the atomic-level information is included in the
Brownian dynamics model via sets of parameters such as potential of mean force (PMF), diffusion coefficients, etc which are obtained from specifically designed all-atom MD simulations or experiments. The trajectory of the solute particle (DNA mononucleotide in this case) is calculated by solving the Langevin equation and is given the following equation:

\[ r(t + \Delta t) = r(t) + v_{\text{flow}}(r) \cdot \Delta t + \frac{D}{k_B T} \cdot F(r(t)) \cdot \Delta t + \sqrt{2 \cdot D \cdot \Delta t} \cdot R(t) \]  

(5.1)

Here \( r(t) \) is the position of the solute at time \( t \), \( D \) is the diffusion coefficient of the solute, \( v_{\text{flow}} \) is the local velocity of the solvent, \( F(r(t)) \) is the force exerted on the solute and \( R(t) \) is a vector of independent normal deviates with a mean of zero and a standard deviation of one. The diffusion coefficient \( D \) is calculated from molecular dynamics simulations, and the force on the solute \( F(r(t)) \) consists of the electrostatic force caused by the external electric field \( E \) and the interaction with the nanochannel walls, solute-surface potential of mean force \( W_{ps} \). The three dimensional potential of mean force \( W_{ps} \) is usually calculated from all-atom molecular dynamics simulations.

\[ F(r(t)) = E \cdot q - \nabla_r (W_{ps}) \]  

(5.2)

### 5.3 Simulation Results

The simulation model includes an explicit particle and two implicit walls. The particle is driven by an electric field of 0.1V/nm and the charge of the particle is 1e. The diffusion constant of the particle in the solution is 0.91nm\(^2\)/ns. The temperature of the system is 300K. The size of the simulation box is 5nm by 5nm by 9nm. The thickness of one wall is 2nm and therefore the particle is restricted
within the range of $z = \pm 2.2\text{nm}$. Since the 3 dimensional PMF for the PMMA wall are not currently available we considered instead a simplified representation of the interactions with the wall via an averaged one-dimensional PMF function which in principle can be obtained from all-atom MD simulation. Various one-dimensional PMF(z) functions considered are given in Figure 5.2. The usage of one-dimensional PMF means that the surface of the wall is homogeneous and thus the force exerted on the particle by the wall is always perpendicular to the wall. Because the PMF calculated from molecular dynamics simulations is always discrete spatially, three-dimensional spline interpolation is applied to compute the continuous potential of the particle. The Brownian dynamics simulations described above are performed by a computer code developed in-house.

The Brownian dynamics simulations were run for different PMFs with potential wells of the depth of $4kT$, $2kT$ and $0kT$, and the simulation time is 50ns. The PMF curves are shown in the left plots in Figure 5.2 and the z position of the particle is plotted versus time and shown on the right figures in Figure 5.2. It can be seen that the interaction between the particle and the walls has an important influence on the trajectory of the particle. With the potential well of $4kT$, the particle is adsorbed on the walls most of the time and can desorb from the wall occasionally, however the wall is not sticky at all when the PMF doesn’t have a potential well and the particle can easily desorb from the walls.
Figure 5.2: The left figure shows the one-dimensional PMFs of the particle as a function of the distance from the walls; the right figure shows the z position of the particle as a function of time. The first row is for the PMF with a potential well of a depth of 4kT, the second row is 2kT and the third row has no potential well.
Figure 5.3: The distribution of the time of flight over 20nm segments. From top to bottom are 4kT, 2kT and 0kT
Then the simulation time for the three cases is extended to 8 microseconds. The transportation process over the 8µs was divided into thousands segments of the length of 20nm and the particle’s time of flight over these 20nm segments is computed. Figure 5.3 shows the distribution of the flight time. The red curves in the plots are the Gaussian fitting curves. It can be seen that the mean time of flight without potential is slightly less than the other two. However, to better distinguish solute particles in the solution, a heterogeneous wall surface might be more helpful. Thus in the future it is necessary to develop the three-dimensional potential of mean force based on molecular dynamics simulations of PMMA walls.

5.4 Conclusion

The Brownian MD simulation studies show that, the main characteristics of the mononucleotides through a nanochannel can be obtained by performing simulations of the dNMPs- PMMA wall system using a coarse-grained representation of the system. The accuracy of this method depends on the accuracy of the potential of mean force used to describe the interaction between dNMP and the PMMA wall.
REFERENCES


APPENDIX: THE CHARMM FORCE FIELD FILES OF PMMA

The topology file:

* Charmm rtf modified from trimer_pmma.rtf
*
 22 0
!MASS

MASS  298 C202  12.01100  C
MASS  299 C301  12.01100  C
MASS  300 C321  12.01100  C
MASS  301 C331  12.01100  C
MASS  302 HGA2   1.00800  H
MASS  303 HGA3   1.00800  H
MASS  304 O2D1  15.99900  O
MASS  305 O302  15.99900  O
MASS     1 HT     1.00800  H
MASS     70 OT    15.99940  O
MASS    195 CLA   35.45000  CL
MASS    190 SOD   22.98977  NA

DECL  -C1
DECL  +C2X

DEFA FIRS NONE LAST NONE
AUTO ANGLES DIHE

RESI  PMM   -0. 0
GROUP
ATOM  C4    C331 -0.2700
ATOM  H1C4  HGA3  0.0900
ATOM  H2C4  HGA3  0.0900
ATOM  H3C4  HGA3  0.0900
GROUP
ATOM  C2X   C301 -0.1000
ATOM  C3    C202  0.8300
ATOM  C1    C301  0.3400
ATOM  O1    O2D1  0.5200
ATOM  C5    C331 -0.1400
ATOM  H1C5  HGA3  0.0900
ATOM  H2C5  HGA3  0.0900
ATOM  H3C5  HGA3  0.0900
GROUP
ATOM  C1    C321 -0.1800
ATOM  H1C1  HGA2  0.0900
ATOM  H2C1  HGA2  0.0900
BOND  C4    H1C4
BOND  C4    H2C4
BOND C4  H3C4
BOND C4  C2X
BOND C2X  C1
BOND C2X  C3
BOND C3  O1
BOND O1  C5
BOND C5  H1C5
BOND C5  H2C5
BOND C5  H3C5
BOND C1  H1C1
BOND C1  H2C1
BOND +C1  C2X
DOUBLE C3  02
IMPR C3  O1  C2X  02

PRES PMMT 0.0    ! C13–TERMINUS
ATOM C1  C331  -0.2700
ATOM HTT1  HGA3  0.0900
ATOM HTT2  HGA3  0.0900
ATOM HTT3  HGA3  0.0900
!
DELETE ATOM H1C1
DELETE ATOM H2C1
!
BOND C1  HTT1
BOND C1  HTT2
BOND C1  HTT3

PRES PMMH 0.0    ! C5–TERMINUS
ATOM C5AT  C331  -0.2700
ATOM H5T1  HGA3  0.0900
ATOM H5T2  HGA3  0.0900
ATOM H5T3  HGA3  0.0900
BOND C5AT  C2X
BOND C5AT  H5T1
BOND C5AT  H5T2
BOND C5AT  H5T3

END
The parameter file:

*Generated PARAM file for ligand
*

ANGLES

HGA3 C331 C301 33.43  110.10  22.53  2.17900  ! SEARCHED: HGA3 CG331 CG301 / ACTUAL: CG301 CG331 HGA3
HGA3 C331 HGA3 35.50  108.40  5.40   1.80200  ! SEARCHED: HGA3 CG331 HGA3 / ACTUAL: CG301 CG331 HGA3
C331 C301 C202 52.00  108.00  11.16   2.561  ! SEARCHED: CG331 CG301 CG202 / ACTUAL: CG202 CG301 CG331
C331 C301 C331 58.35  113.50  11.16   2.561  ! SEARCHED: CG331 CG301 CG331 / ACTUAL: CG301 CG331 CG331
C331 C301 C321 58.35  113.50  11.16   2.561  ! SEARCHED: CG331 CG301 CG321 / ACTUAL: CG321 CG301 CG331
C301 C202 C321 52.00  108.00  11.16   2.561  ! SEARCHED: CG301 CG321 CG321 / ACTUAL: CG321 CG301 CG321
C301 C202 C321 55.00  109.00  20.00   2.3260  ! SEARCHED: CG301 CG321 CG321 / ACTUAL: CG321 CG331 HGA2
C301 C202 C321 40.00  116.00  50.00   2.353  ! SEARCHED: CG301 CG321 CG321 / ACTUAL: CG301 CG321 CG321
C202 C202 322.00  1.1110  ! SEARCHED: CG202 CG321 CG321 / ACTUAL: CG321 HGA2
C301 C321 C301 58.35  113.50  11.16   2.561  ! SEARCHED: CG301 CG321 CG301 / ACTUAL: CG301 CG321 CG301
C301 C321 HGA2 26.50  110.10  22.53   2.179  ! SEARCHED: CG301 CG321 HGA2 / ACTUAL: CG301 CG321 HGA2
HGA2 C321 HGA2 35.50  109.00  5.40   1.802  ! SEARCHED: HGA2 CG321 HGA2 / ACTUAL: HGA2 CG321 HGA2
C321 C301 C321 58.35  113.50  11.16   2.561  ! SEARCHED: CG321 CG301 CG321 / ACTUAL: CG321 CG301 CG321

BONDS

HGA3 C331 322.00  1.1110  ! SEARCHED: HGA3 CG331 / ACTUAL: CG331 HGA3
C331 C301 222.50  1.5380  ! SEARCHED: CG331 CG301 / ACTUAL: CG301 CG331
C301 C202 200.00  1.5220  ! SEARCHED: CG301 CG202 / ACTUAL: CG202 CG301
C301 C321 222.50  1.5380  ! SEARCHED: CG301 CG321 / ACTUAL: CG301 CG331
C202 C331 150.00  1.3340  ! SEARCHED: CG202 CG331 / ACTUAL: CG301 CG321
C202 C202 750.00  1.2200  ! SEARCHED: CG202 CG202 / ACTUAL: CG202 CG202
C301 C321 340.00  1.4300  ! SEARCHED: CG301 CG331 / ACTUAL: CG331 CG202
C321 HGA2 309.00  1.1110  ! SEARCHED: CG321 HGA2 / ACTUAL: CG321 HGA2

DIHEDRALS

HGA3 C331 C301 C202 0.2000  3  0.00  ! SEARCHED: HGA3 CG331 CG301 CG202 / ACTUAL: CG202 CG301 CG331 HGA3
<table>
<thead>
<tr>
<th>HGA3</th>
<th>C331</th>
<th>C301</th>
<th>C331</th>
<th>0.1600</th>
<th>3</th>
<th>0.00</th>
<th>SEARCHED: HGA3 CG331 CG301 CG331 / ACTUAL: CG331 CG301 CG331 HGA3</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGA3</td>
<td>C331</td>
<td>C301</td>
<td>C321</td>
<td>0.1600</td>
<td>3</td>
<td>0.00</td>
<td>SEARCHED: HGA3 CG331 CG301 CG321 / ACTUAL: CG321 CG301 CG331 HGA3</td>
</tr>
<tr>
<td>C331</td>
<td>C301</td>
<td>C2O2</td>
<td>O302</td>
<td>0.0500</td>
<td>6</td>
<td>180.00</td>
<td>SEARCHED: CG331 CG301 CG2O2 OG302 / ACTUAL: OG302 CG2O2 CG311 CG331</td>
</tr>
<tr>
<td>C331</td>
<td>C301</td>
<td>C2O2</td>
<td>O2D1</td>
<td>0.0500</td>
<td>6</td>
<td>180.00</td>
<td>SEARCHED: CG331 CG301 CG2O2 OG2D1 / ACTUAL: OG2D2 CG2O3 CG301 CG331</td>
</tr>
<tr>
<td>C331</td>
<td>C301</td>
<td>C321</td>
<td>C301</td>
<td>0.2000</td>
<td>3</td>
<td>0.00</td>
<td>SEARCHED: CG331 CG301 CG321 CG301 / ACTUAL: CG331 CG301 CG321 CG321</td>
</tr>
<tr>
<td>C331</td>
<td>C301</td>
<td>C321</td>
<td>HGA2</td>
<td>0.1900</td>
<td>3</td>
<td>0.00</td>
<td>SEARCHED: CG331 CG301 CG321 HGA2 / ACTUAL: CG331 CG301 CG321 HGA2</td>
</tr>
<tr>
<td>C301</td>
<td>C2O2</td>
<td>O302</td>
<td>C331</td>
<td>2.0500</td>
<td>2</td>
<td>180.00</td>
<td>SEARCHED: CG301 CG2O2 OG302 CG331 / ACTUAL: CG311 CG2O2 OG302 CG331</td>
</tr>
<tr>
<td>C301</td>
<td>C321</td>
<td>C301</td>
<td>C2O2</td>
<td>0.2000</td>
<td>3</td>
<td>0.00</td>
<td>SEARCHED: CG301 CG321 CG301 CG2O2 / ACTUAL: CG2O2 CG311 CG321 CG311</td>
</tr>
<tr>
<td>C301</td>
<td>C321</td>
<td>C301</td>
<td>C321</td>
<td>0.1950</td>
<td>3</td>
<td>0.00</td>
<td>SEARCHED: CG301 CG321 CG301 CG321 / ACTUAL: CG301 CG321</td>
</tr>
<tr>
<td>C331</td>
<td>O302</td>
<td>C2O2</td>
<td>O2D1</td>
<td>0.9650</td>
<td>1</td>
<td>180.00</td>
<td>SEARCHED: CG331 OG302 CG2O2 OG2D1 / ACTUAL: OG2D1 CG2O2 OG302 CG331</td>
</tr>
<tr>
<td>O2D1</td>
<td>C2O2</td>
<td>C301</td>
<td>C321</td>
<td>0.0500</td>
<td>6</td>
<td>180.00</td>
<td>SEARCHED: OG2D1 CG2O2 CG301 CG321 / ACTUAL: OG2D1 CG2O2 CG311 CG321</td>
</tr>
<tr>
<td>C321</td>
<td>C301</td>
<td>C321</td>
<td>HGA2</td>
<td>0.1580</td>
<td>3</td>
<td>0.00</td>
<td>SEARCHED: CG321 CG301 CG321 HGA2 / ACTUAL: CG311 CG301 CG321 HGA2</td>
</tr>
</tbody>
</table>

IMPROPER

C2O2 C301 C321 HGA2 0.1580 3 0.00 ! SEARCHED: CG321 CG301 CG321 HGA2 / ACTUAL: CG311 CG301 CG321 HGA2 |

NONBONDED nbxmod 5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctodb 12.0 ctom 10.0 eps 1.0 el4fac 1.0 wmin 1.5
C2O2 0.0 -0.0980 1.7000 ! methyl acetate update viv 12/29/06
C301 0.0 -0.0320 2.0000 0.0 -0.01 1.9 ! alkane (CT0), neopentane, from CT1, viv
C321 0.0 -0.0560 2.0100 0.0 -0.01 1.9 ! alkane (CT2), 4/98, yin, adm jr, also used by viv
C331 0.0 -0.0780 2.0500 0.0 -0.01 1.9 ! alkane (CT3), 4/98, yin, adm jr; Rmin/2 modified from 2.04 to 2.05
HGA2 0.0 -0.0350 1.3400 ! alkane, igor, 6/05
HGA3 0.0 -0.0240 1.3400 ! alkane, yin and mackerell, 4/98
O2D1 0.0 -0.1200 1.7000 0.0 -0.12 1.40 ! carbonyl
O302 0.0 -0.1000 1.6500 ! ester; LJ from THP, sng 1/06
VITA

Kai Xia was born in Hefei, China, on August, 1983. He entered University of Science and Technology of China and earned a Bachelor of Science degree in Physics in 2004. He joined the Institute of Modern Physics at Fudan University and obtained a Master of Science in Physics in 2008.

He was accepted in the PhD program in the department of Mechanical Engineering at Louisiana State University in 2009, where he began the research of molecular simulation of a novel DNA sequencing technique under the supervision of Dr. Dorel Moldovan. In addition, in 2009 he enrolled in the department of Petroleum Engineering at Louisiana State University to pursue for a dual degree. In 2014 he received the master degree in Petroleum Engineering.