MOLECULAR MODELLING OF DNA TRANSLOCATION IN BIO-FUNCTIONALIZED NANOPORES WITH APPLICATIONS IN GENE SEQUENCING

by

ABHIJIT RAMACHANDRAN

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN BIOMEDICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON
December 2009

ACKNOWLEDGEMENTS

I would like to acknowledge and extend my heartfelt gratitude to all the people who have made the completion of this Thesis possible. First, I would like to thank and express my gratitude to my supervisor, Dr. Yaling Liu, for his guidance, comments, and encouragement during my research work. He refined in me the ability to plan and organize the studies along with being patient and persistent, especially when research becomes slow and frustrating. His willingness to discuss contrasting opinions and challenge intuitive thoughts helped me develop strong fundamentals and more clear way of expressing my ideas. Deepest gratitude is also due to Dr. Samir Iqbal, without whose knowledge and assistance this study would not have been successful. I would like to extend my thanks to Dr. Hanli Liu and Dr. Behbehani for being part of my thesis defense committee. During my Master's study at University of Texas at Arlington, I thoroughly enjoyed the courses I took, especially from Dr. Behbehani, Dr. Liu, Dr. Alexandrakis, Dr. Dave, Dr. Iqbal and Dr. Casa-Esperón. I also express my appreciation to all my lab mates for their suggestions, questions and brain storming with me for solutions. I thank the staff of Biomedical Engineering Department, Cindy, Suzanne, and Amanda for their administrative support. I express thanks to my friends for being there and helping me with timely help when needed. Last, but not least, I am grateful to my family: my parents, K. N. Ramachandran, and Susan Chandran and my sister, Abhilasha Ramachandran, for their endless love and support.

November 20, 2009

ABSTRACT

MOLECULAR MODELLING OF DNA TRANSLOCATION
IN BIO-FUNCTIONALIZED NANOPORES
WITH APPLICATIONS IN
GENE SEQUENCING

Abhijit Ramachandran, M.S.

The University of Texas at Arlington, 2009

Supervising Professor: Yaling Liu

The human genome project which was completed in 2003 took more than 12 years to complete at a cost of $2.7 billion. The advancement in technologies in the next five years led to the same feat being achieved in 5 months with a total cost of $1.5 million. Micro-fabrication technology has now made it possible miniaturize to micro- and nanofluidic devices for biochemical analysis. Solid-state nanopore based molecular analysis provides the potential to characterize and sequence DNA, based on the signature of the ionic current measured. A major challenge in using solid-state nanopores for DNA detection and sequencing is the molecular selectivity and sensitivity, and the control of DNA-nanopore interface. Presently, various approaches to modify nanopore surface properties and functionalized nanopores have been developed. However, the interaction between DNA and the bio-functionalized nanopores is still not well understood due to the small length scales of the DNA/nanopore and the dynamic nature of the translocation process.

The aim of this thesis is to understand the interaction between the DNA and chemically modified nanopore surfaces. The translocation process of a DNA is analyzed by probing the DNA-nanopore interaction mechanisms through full atomistic molecular dynamics and the DNA-functionalized nanopore interaction mechanisms through coarse-grained molecular dynamics modelling. The DNA translocation dynamics through nanopores of various diameters and under various applied bias voltages are characterized. A non-linear relationship between DNA translocation speed and applied voltage is revealed. The effect of nanopore coating on DNA translocation speed is also analyzed. In particular, DNA translocation in nanopores coated with hairpin loop DNAs (HPL) and single stranded DNAs (ssDNAs) are compared. It is observed that a small effective pore diameter (EPD) provides a high confinement where the DNA translocation speed is dependent on the interaction potential, type, density of the coatings at voltages lower than 100mv/nm. Also, DNA is found to translocate in a ssDNA coated nanopore 900% faster when compared with the HPL coated nanopore mainly at a bias voltage of 0.01V, due to the less stiffness of the ssDNA as compared to the HPL. Such observations partially explain the bio-functionalized nanopore molecular selectivity mechanism. Such surface property-translocation dynamics relationship can be used for the optimal designs of future lab-on-chip molecular diagnostic devices. Furthermore, this study has the potential to enhance the understanding of a variety of ligand-receptor combinations of significant importance.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION

1.1 Overview

Cell is the basic structural and functional unit of all known living organisms. It is the smallest unit of an organism that is classified as living, and is often called the building block of life[1]. The instructions needed to direct the activities of these cells are contained within the chemical deoxyribonucleic acid (DNA). DNA from all organisms is made up of the same chemical and physical components. The DNA sequence or genetic sequence is the particular side-by-side arrangement of bases along the DNA strand (e.g. AAGGCTGTAG). This order spells out the exact instructions required to create a particular organism with its own unique traits. The genome is an organism's complete set of DNA. It varies widely in size with smallest known genome being of a bacterium that contains about 600,000 DNA base pairs, while human and mouse genomes have some 3 billion DNA base pairs (see Fig. 1.1) [2].

Organism	Genome Size (Base Pairs)	Estimated Genes
Human (*Homo sapiens*)	3.2 billion	25,000
Laboratory mouse (*M. musculus*)	2.6 billion	25,000
Mustard weed (*A. thaliana*)	100 million	25,000
Roundworm (*C. elegans*)	97 million	19,000
Fruit fly (*D. melanogaster*)	137 million	13,000
Yeast (*S. cerevisiae*)	12.1 million	6,000
Bacterium (*E. coli*)	4.6 million	3,200
Human immunodeficiency virus (HIV)	9,700	9

Figure 1.1. Comparison of Genome size and number of genes in different organisms, Venter *et. al.*, 2001

In 1990, U.S. Department of Energy (DOE) and the National Institutes of Health developed a plan for sequencing of the entire human genome and thus began the Human Genome Project (HGP). The HGP's ultimate goal was to generate a high-quality reference DNA sequence for the human genome's 3 billion base pairs and to identify all human genes [3-5]. This would help us understand human evolution, enable personalized medicine and recognize the role of environment and heredity in defining the human condition [2]. The human genome project (completed 2003, cost $2.7 billion) took more than 12 years to complete [6]. With the advent of novel sequencing technologies and abundant computation power, the cost and time needed for sequencing a genome has reduced dramatically, now at ~3 cents/base [6]. Yet, this is an enormous amount for 3 billion bases to be sequenced. The challenge to reduce the cost and time further has pushed the development of sequencing technology. Developing from Sagner's revolutionizing dideoxy sequencing method in 1977 [7], many new technologies are now on the horizon that could yield massive increases in our capacity for de novo DNA sequencing. The major techniques are either electrophoretic method like the slab gel electrophoresis, capillary electrophoresis and microfabricated Capillary Arrays or non electrophoretic method like Pyrosequencing, Sequencing by Hybridization-Microarrays, Massively Parallel Signature sequencing and Single Molecule Methods- Atomic Force Microscopy, and Single-Channel-Current measurements. The Single-molecule method of our interest uses the principle of Ion channels using nanopores along with single channel current measurements to enable DNA sequencing. The nanopore based device provides a highly confined space within which the polymers like DNA and RNA (Ribonucleic Acid) can be electrophoretically driven and analyzed at high throughput. The unique analytical capability that kilobase (kb) length polymers can be identified and characterized without amplification or labeling makes rapid DNA sequencing an inexpensive possibility through nanopores [8]. Thus nanopores, in spite of their material, electrical and fabrication shortcomings and challenges they have the potential to achieve the $1000/genome target.

1.2 Principle of DNA Sequencing

The concept of nanopore-based sequencing of DNA is very simple. An electrochemical system is shown in Fig. 1.2, where an applied bias voltage will drive the ions to flow from anode (-) to cathode (+) through the nanopore. This gives the baseline current value. The DNA carries a '–e' charge due to its phosphate backbone and hence is electrophoretically driven through the nanopore towards the cathode. The side on which the DNA is added is termed as 'cis' and the other side to which it translocates is termed as 'trans'. As it flows through the nanopore it blocks the free passage of ions, which causes a dip in the ionic current that is being measured using the electrical circuit. The DNA is forced to pass through the nanopore of a comparable diameter, thus it has to pass in a linear mode on a base by base fashion instead of being coiled. The degree and duration of the ionic current dip is proportional to the molecules structure and length. The technique though simple, promises a range of analyses with high throughput than the current methods.

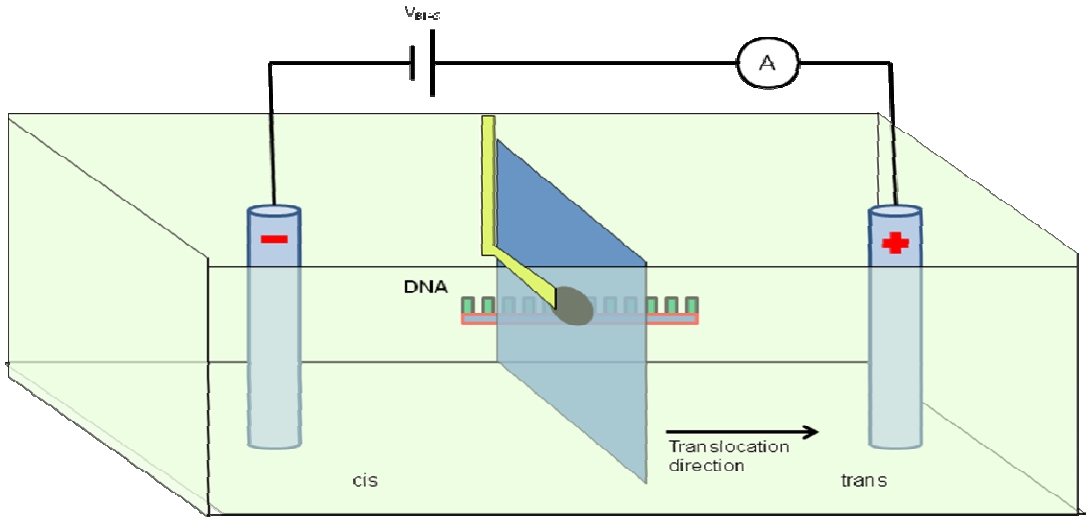


Figure 1.2. Electrophoretic translocation of ssDNA through the nanopore

Genome sequencing with such a simplified model is difficult to be achieved. Various attempts to modulate the existing system and obtain a de novo sequencing method are being

developed. One such method which successfully demonstrated gene sequencing is the motivation for this thesis and is described below.

1.3 Genome Sequencing Through Functionalized Nanopores

Iqbal *et. al.* developed single-molecule electrophoretic transport measurements of single strand of DNA (ssDNA) through Hairpin Loop DNA (HPL) functionalized nanopores which allowed gene sequencing[9]. They explained their observations using Bauer *et. al.*'s [10] theoretical predictions of channel-molecule interactions through nanopores.

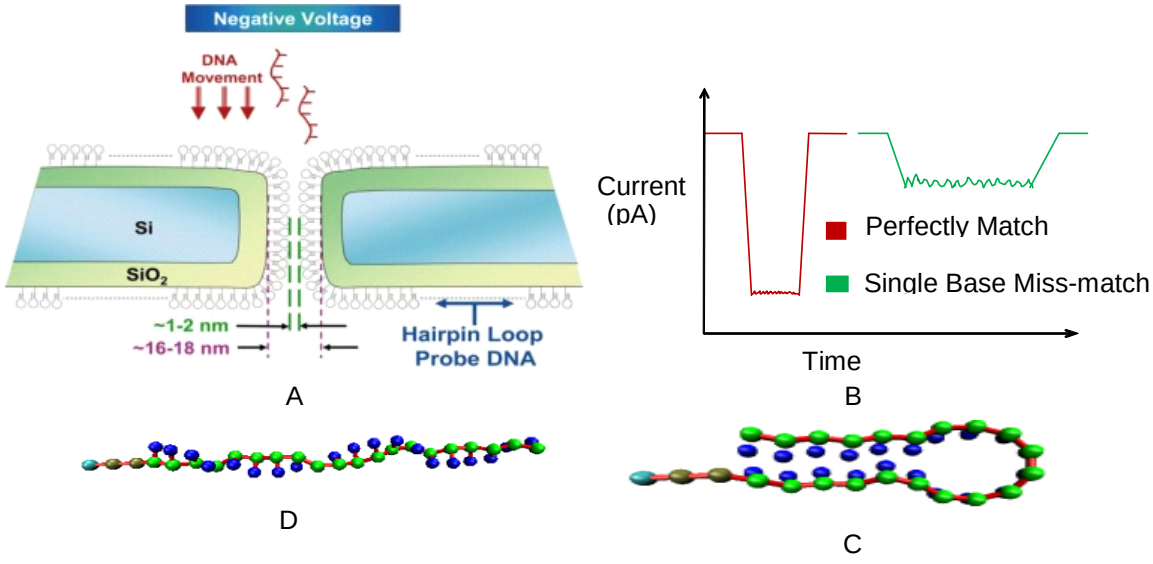


Figure 1.3. (A) Schematic of the bio-functionalized system (B) Schematic to show selectivity achieved due to the difference in duration of translocation times measured using the dip in ionic current of matched and mismatched DNA, (C) HPL DNA structure and (D) ssDNA structure.

The nanopores were coated with hairpin loop DNA molecules (Fig. 1.3.C). This coating made the nanopores selective towards short lengths of 'target' ssDNA (Fig. 1.3.D), transporting perfect complementary (PC) molecules faster than mismatched ones (MM). Even a single base mismatch between the probe and the target resulted in longer translocation pulses and a significantly reduced number of translocation events (Fig. 1.3.B), The pulses for perfect complementary ssDNA were narrower and deeper than those for mismatched ones, indicating

facilitated (and faster) transport of PC-DNA while interacting with surface bound hairpin molecules.

1.4 Aim of the Thesis

The aim of this thesis is to understand the mechanisms of DNA translocation in bio-functionalized nanopores which have enabled imparting selectivity to solid-state nanopores. For this purpose the studies were divided into three parts, first, we will characterize the interaction between a DNA and a bare Nanopore [11]. Second, the surface properties of a functionalized nanopore are studied. Finally we explore the interaction between a DNA and a functionalized nanopore. Our simulation system is similar to the schematic in Fig.1.3.A and an ideal selectivity achieved from the system is shown in Fig.1.3.B. In particular, we study the interaction between ssDNA (Fig.1.3.C) and hair-pin loop (HPL, Fig.1.3.D) coated nanopore. We compare the translocation processes of a ssDNA with sequence perfectly matching with the probe HPLs and that with a completely mismatched sequence, thus revealing the underlying molecular mechanisms and process of functionalized nanopore-based DNA sequencing. Such and similar functionalization schemes can be used for a variety of ligand-receptor combinations of significant importance, and the solid-state functionalized nanopore can serve as next generation of sequencing tools, whereas a pore functionalized with specific probe can be used as detector of specific nucleotide(s)/biomarkers [9].

CHAPTER 2

REVIEW OF CURRENT DNA SEQUENCING TECHNIQUES

2.1 Background

To understand DNA sequencing modalities, the concept on which these technologies are based on is necessary to be illustrated. The following section will highlight the basics of DNA and its useful properties.

2.1.1 Deoxyribonucleic acid (DNA)

DNA is a nucleic acid that contains the blueprints of genetic instructions used in the development and functioning of all known living organisms. The term nuclein was coined by Miescher in 1869 for substances isolated from the nucleus of a cell and with time it was modified as nucleic acid [12]. The DNA has a helical structure as identified by Watson and Crick in 1953 [13] and a diameter of 2nm. It is a polymer that consists of two long chains which are made of repeated sub-units, called nucleotides, arranged in linear fashion.

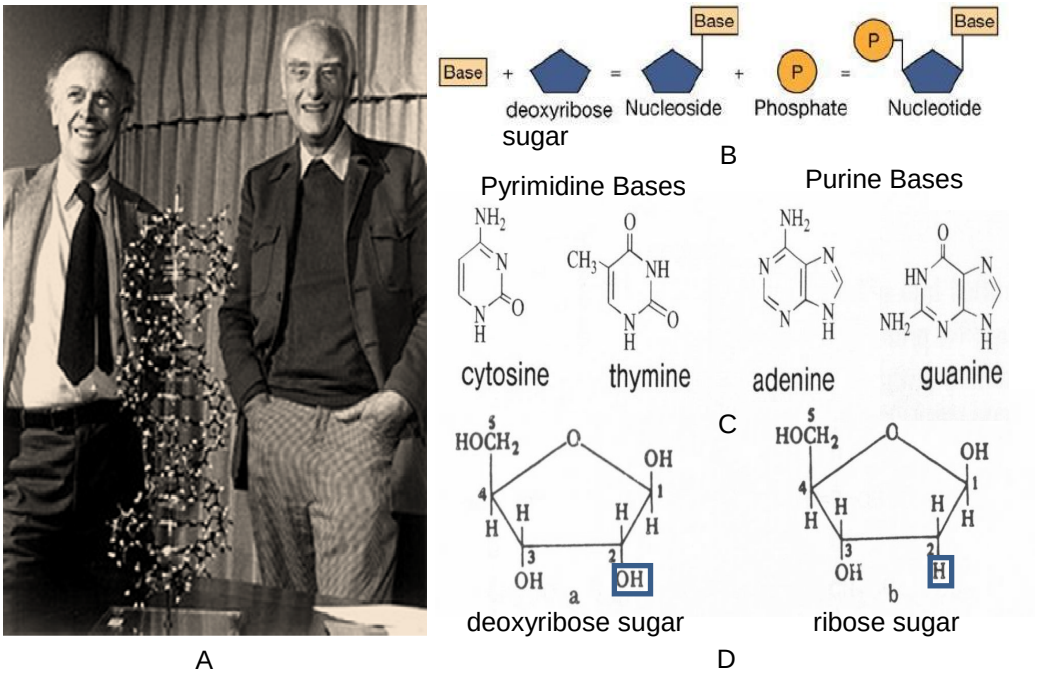


Figure 2.1. Watson and Crick with their double helix structure of DNA (A), basic components of a single DNA nucleotide (B), The different bases of DNA (C) and the difference between the sugar found in DNA and RNA respectively (D). Wink *et. al.*, 2006

A nucleotide consists of a phosphate group, a pentose sugar, and a base, as shown in the Fig. 2.1.B [14]. There are two classes of pentose sugar namely ribose (in ribonucleic acid, RNA) and deoxyribose (in DNA), the difference between the two is the presence of a hydrogen atom instead of a hydroxyl group on carbon atom 2 (see Fig. 2.1.D) for the deoxyribose. Also, there are two types of bases, a Pyrimidine (Thymine [T] and Cytosine [C]) which is a heterocyclic aromatic organic compound similar to benzene, containing two nitrogen atoms at positions 1 and 3 of the six-member ring and purines (Adenine [A] and Guanine [G]) which are more complex fused nitrogeneous rings consisting of pyrimidine and imidazole rings (see Fig. 2.1.C) [12].

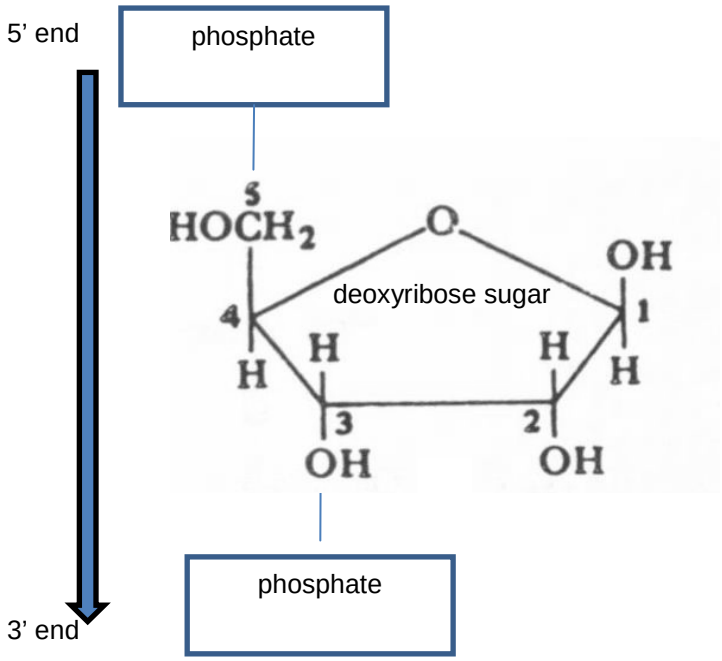


Figure 2.2. The direction given by the carbon atom at which the phosphate molecule binds, in this case shows the 5'➔ 3' directionality. Gilchrist *et. al.*, 1997

The long chains (strands) forming the double stranded DNA (dsDNA) are anti-parallel in nature i.e. they run in opposite direction to each other (5' – 3' and 3' – 5'). This is established since, as shown in Fig. 2.2, the phosphate attaches to the 5' carbon of the trailing pentose sugar (phophodiester bond) and the 3' carbon of the next, giving it a 5'➔3' direction, while its complimentary strand will have the 3'➔5' direction. The DNA has a negative charge (-e) due to its phosphate backbone, a free-floating phosphate has a charge of '-2e' and is represented as PO_4^{-2}. However, phosphate in DNA is not free phosphate due to its additional attachment with the sugar and thus the phosphate in DNA has a charge of '-1e', (PO_4^{-1}) [14].

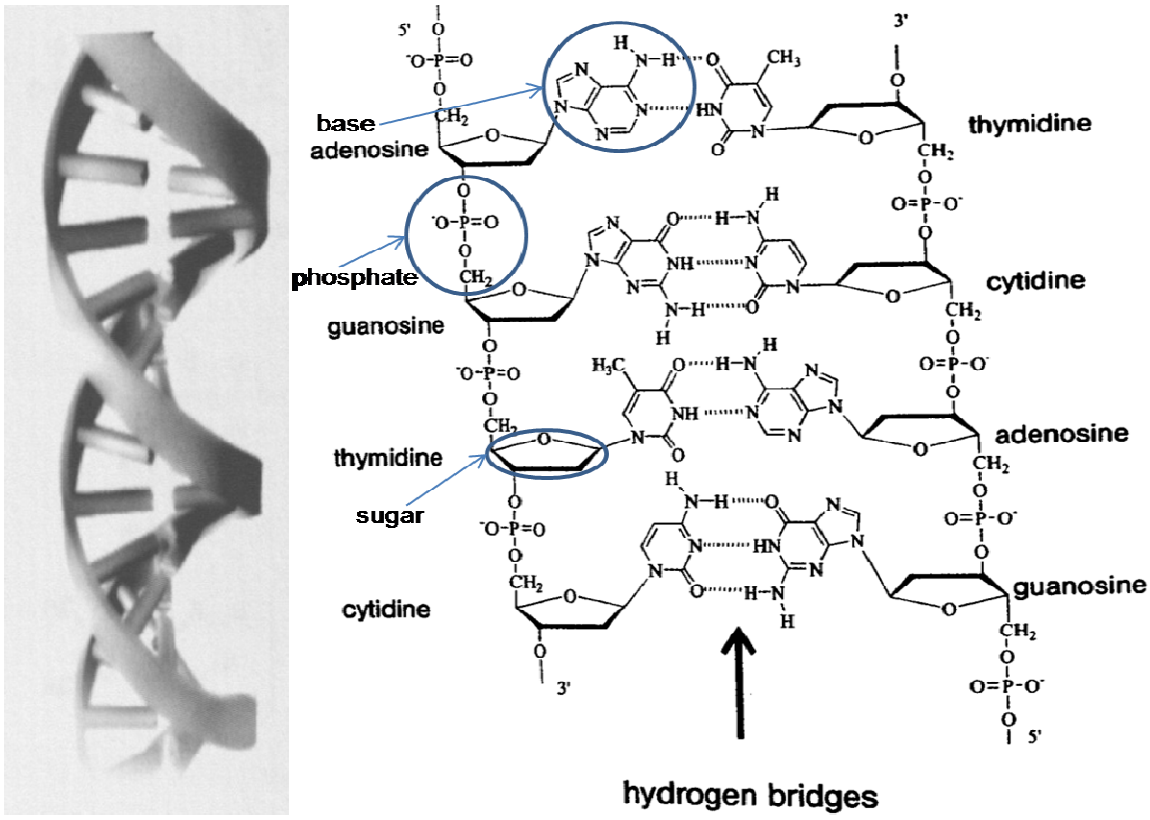


Figure 2.3. The regular double helix structure of DNA (A) and the molecular level detail of the DNA strand with the complimentary base pairing with the formation of the hydrogen bonds (B). Wink *et. al.*, 2006

Further, the bases of DNA have the property of complimentary base pairing which is achieved with the specific formation of two or three hydrogen bonds between A-T and G-C pairs, respectively (see Fig. 2.3.B). These hydrogen bonds enable molecular recognition but are weaker than the covalent bonds between sugar and phosphate of the DNA backbone. Base pairing occurs if the complimentary bases are in close proximity of each other. This gives them the ability to self-organize and form supramolecular structures without the requirement of energy or regulatory helpers. Also, if either strand from a dsDNA is removed, it can be recreated by using other as a template, which is an important requirement for basic genetic processers (replication, transcription and recombination). The processes by which the order of the nucleotide bases in a molecule of DNA can be determined are termed as DNA sequencing methods [14].

The DNA double helix is very stable owing to the base pairing between complementary strands and stacking between adjacent bases. The dsDNA can be separated into two single strands of DNA (ssDNA) either by high temperature or changing the salt concentration (pH) of the solution [15]. The process of denaturing the DNA to form single strands by increasing the temperature is termed as melting or de-hybridization of DNA. Then cooling the system such that the two single strands of DNA would re-form (renaturation) double helices at regions where sequence is complimentary is termed as DNA hybridization. The temperature at which the 50% of DNA is present as single strands is termed as T_m (melting temperature). T_m is dependent on the G-C content of DNA, higher the percentage of G-C, higher is the temperature needed for them to de-hybridize. This is because the G-C pairs have three hydrogen bonds as compared to two formed between A-T pairs (see Fig. 2.3.B). This technique is useful for determining sequence similarity among DNAs of different origin and the amount of sequence repetition within one DNA[15]. Another aspect of DNA which is exploited most is its ability to replicate. For the process of DNA replication the dsDNA separates into two ssDNA. These original ssDNA's now act as templates for constructing new dsDNA, utilizing the property of complimentary base-pairing. For this a primer (a strand of nucleic acid that serves as a starting point for DNA replication) attaches to the 3' end of the ssDNA template strand and the DNA polymerase (family of enzymes that carry out all forms of DNA replication by adding free nucleotides) adds new nucleotides to the existing template strand of DNA from the end where the primer has bound. The process is terminated once the new strand is completely synthesized from template strand, thus we obtain two dsDNA. This ascertains that if either strand from a dsDNA is removed, it can be recreated exactly by using the other as the template. Furthermore, this lead to the revolutionizing technique called PCR (Polymerase Chain Reaction). PCR is a process to amplify a DNA strand across several orders of magnitude and enabling generation of thousand copies. The process is elaborated in Fig. 2.4 [16].

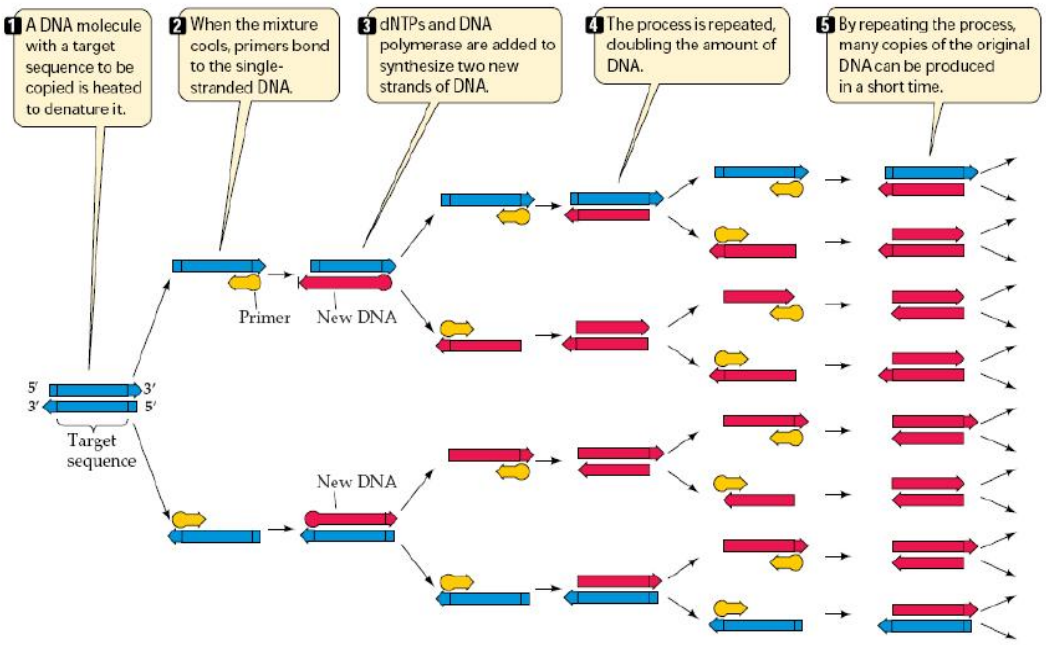


Figure 2.4. PCR (polymerase chain reaction) process to amplify a strand of DNA and obtain thousands of copies, Purves *et. al.*, 2003

2.2 DNA Sequencing Modalities

Human Genome Project is a 13 year effort for obtaining the entire Human genome for the very first time with a cost of $2.7billion and finished in 2003. The advancement in technologies in the next five years led to the same feat being achieved in a time-span of 5 months with a total cost of $1.5million [17]. The section below gives a general insight into the various current DNA sequencing modalities which are technically advanced and significantly reduce the cost.

2.2.1 Slab Gel Electrophoresis

The chain termination sequencing method developed by Sanger in 1971 utilizes the base specific chain terminations in four separate reactions ('A','T','G' and 'C') which correspond to the four different nucleotides in the DNA composition [7]. The principle of this method is explained in Fig. 2.5, it requires four separate DNA extension reactions. Each reaction contains a template DNA, a short primer, DNA polymerase and all four deoxynucleotides (dNTPs➔ dATP, dTTP,

dGTP and dCTP) of which one is radioactively labeled for each of the reactions represented by an asterisk (dATP*, for the current example). Also these reactions are spiked by a corresponding dideoxynucleotides **(**ddNTP ➔ ddATP, ddTTP, ddGTP and ddCTP). The dNTPs are added with the nucleotides and they bind with the complimentary base, but when the ddNTP is added in connectivity the growth of the chain is terminated. This is because the ddNTP lacks a 3` hydroxyl goup (OH) necessary to form the linkage with an incoming nucleotide. Thus the new strand will extend until a ddNTP is incorporated. Using a polyacrylamide gel, the radioactive products are separated through the four lanes and scored according to their molecular masses. The location of the strands in the four lanes gives us its nucleotide position while the lane dictates the type of base at that position [18].

The modified Sanger approach (see Fig. 2.5.B) utilizes four different fluorescent dyes to label the different ddNTPs instead of radioactively labeling a particular dNTP [19]. The reaction products are then separated electrophoretically in a single glass capillary filled with a polymer. The DNA bands move inside the capillary according to their lengths (masses). Lasers are used to excite the fluorophores at the end of the capillary. The sequence of DNA is determined based on the color that corresponds to a particular nucleotide. The Sanger method is still considered as the "gold standard" for sequencing but it has certain limitations. One of the limitation is the issue of tracking the different strands that terminate at different locations in the gel but the major limitation is that the cost of this method is $1 per kilobase (kb), thus it would cost $10,000,000 to sequence the human genome [20].

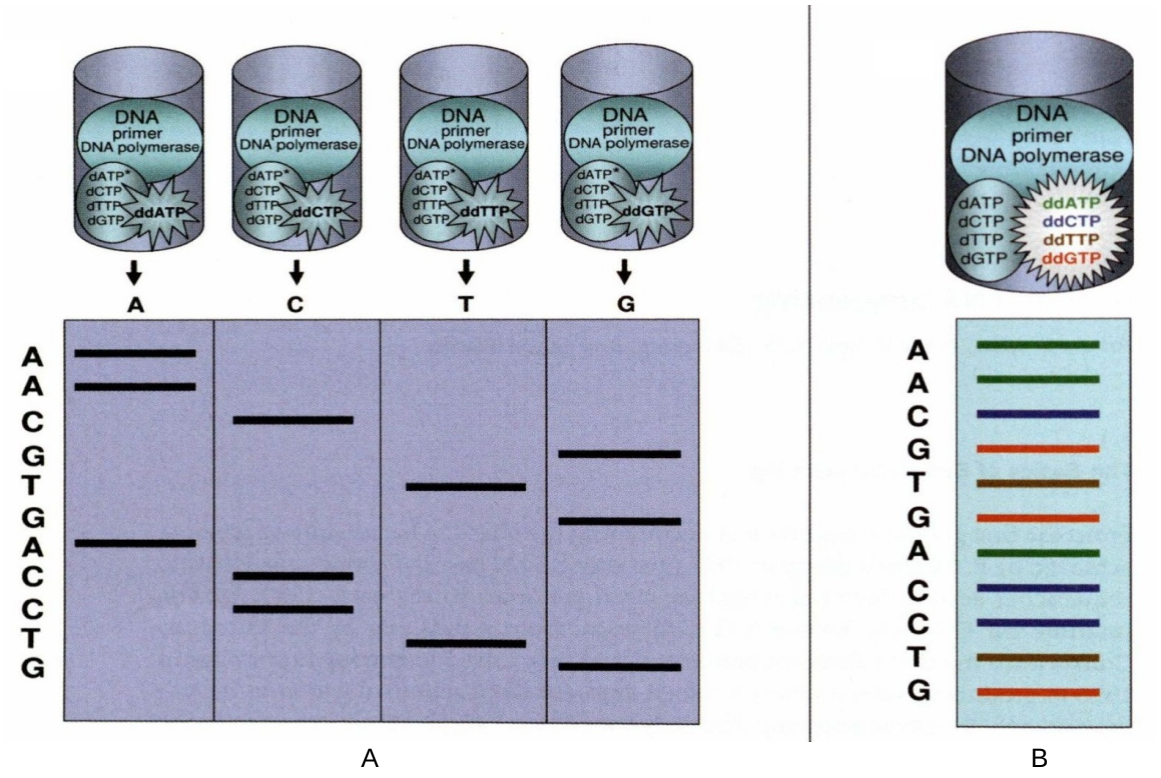


Figure 2.5. The Sanger technology which uses the slab gel electrophoresis in which the radioactive products are separated through the four lanes and scored according to their molecular masses (A). This technique was improvised and different radioactive label are used for to denote a particular dNTP thus enabling a single lane readout of the DNA sequence (B). Janitz *et. al.*, 2008

2.2.2 Capillary Electrophoresis (CAE)

The need for a faster sequencing device and higher throughput led to the development of Capillary Array Electrophoresis (CAE) [21]. It is designed to separate species based on their size to charge ratio in the interior of a small capillary filled with an electrolyte. The principle of CAE is similar to that of slab gel electrophoresis except that the Sanger DNA sequencing fragments are separated on an array of capillaries, the detection process remains the same. The advantage is that there is only a single sample and that eliminates tracking problems. In capillary electrophoresis the fragment migration time is directly related to the number of bases present. The resolution of this system is 3 base pairs [22].

2.2.3 Micro-fabricated Capillary Arrays

Performance improvements in the CAE are being achieved with miniaturization of the current system. This will reduce the cost of the devices, separation time and make more capillaries available in an instrument.

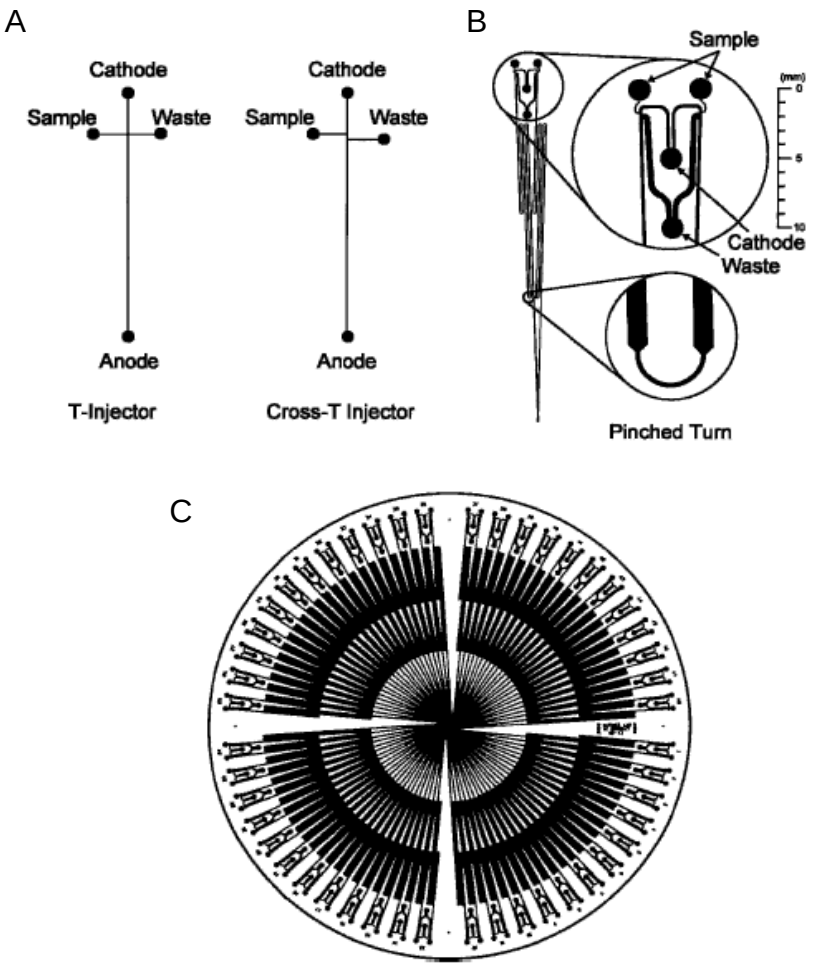


Figure 2.6. T-injector and cross T-injector geometry of microfabricated capillary electrophoretic devices (A) respectively. Enlarged views of sample injector and pinched turn (B), schematic of radial chip design with 96 channels (C). Metzeker *et. al.*, 2005.

A few such developmental designs are shown in Fig. 2.6, the T-injector design and cross-T design[23]. In this, the sample is electrophoresed from left to right such that it fills the T intersection. When current flow from the sample to waste is terminated and a potential is applied across the main cathode and anode, part of the injected sample, of length equal to the channel

overlap at the T intersection, is electrophoresed into the main capillary. This technique provides superior sample control, higher resolution and shorter separation lengths [24].

2.2.4 Fluorescent in Situ Sequencing (FISH)

Fluorescent in situ sequencing (sequencing by synthesis, *(SBS)*) is the most widely used method for four-color DNA sequencing and it was described by Smith *et. al.* 1986 [25] and Prober *et. al.* 1987 [26]. Resolution of the emission signal from a dye labeled nucleotide into color, with subsequent assignment in the DNA sequence is the basis for this method. It usually involves the following steps: the DNA to be sequenced is attached onto a solid surface, labeled nucleotides with cleavable chemical group to cap an $-OH^-$ group at 3'-position of the deoxyribose are added along with the DNA polymerase. The incorporation of these nucleotides will terminate the reaction and the sequence is read from the labels used for the nucleotides (see Fig. 2.7).

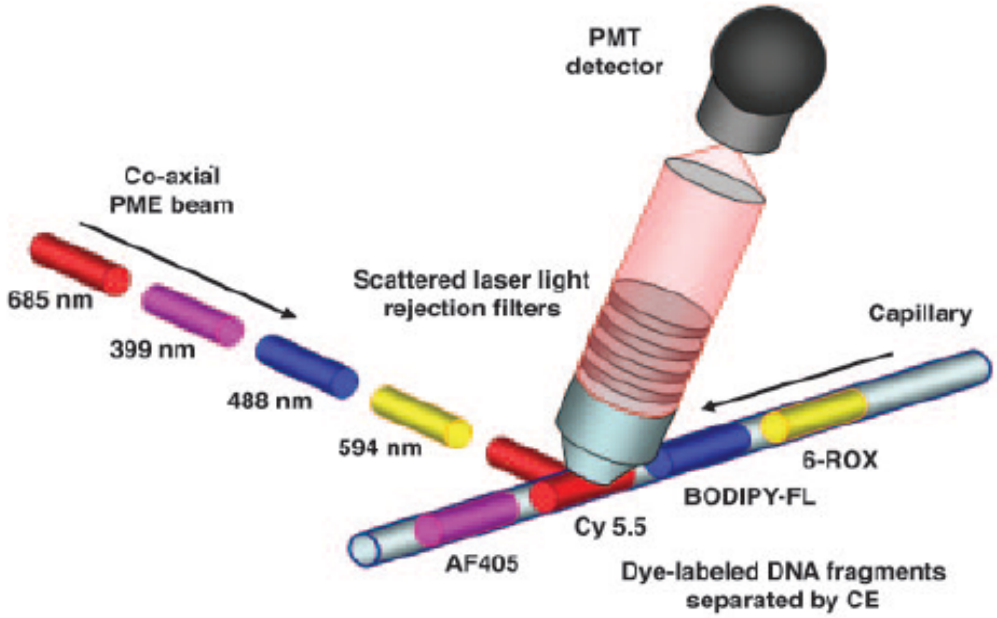
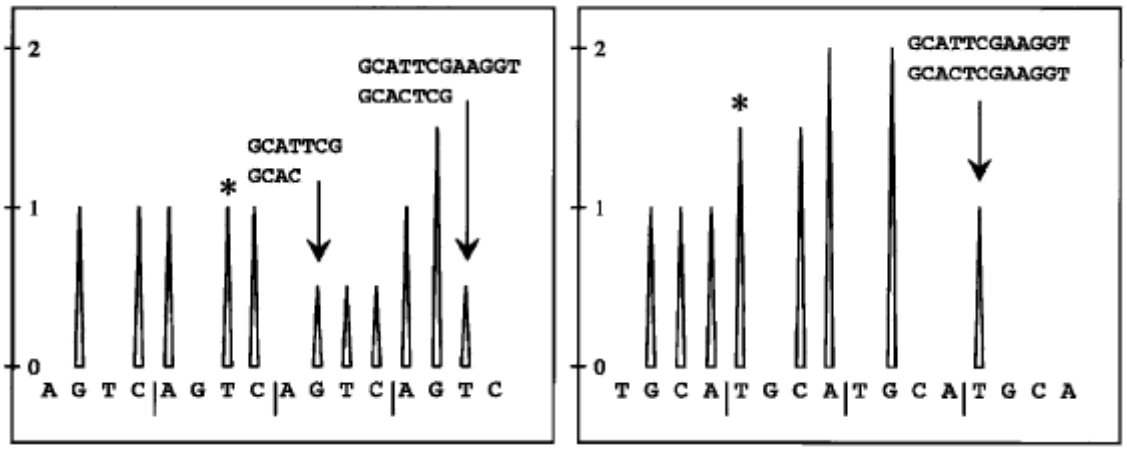


Figure 2.7. Pulsed multiline excitation (PME) technology an advanced technique of the fluorescence in-situ sequencing in which a single coaxial PME beam interrogates the fluorescently labeled DNA fragments which are already separated through capillary gel electrophoresis. The scattered laser light is detected using a Photo multiplier tube (PMT) and processed to obtain sequence of the DNA with respect to the different dye primers used for respective ddNtps. Lewis *et. al.*, 2005

Then the 3-OH^- is uncapped and washed from the 3'- position of the deoxyribose and the sequencing cycle is continued. This method provides a very high throughput since each base is detected while the DNA chain is growing [27, 28]. But it has certain limitations like inefficient excitation of the fluorescent dyes, significant spectral overlap, and inefficient collection of emission signals. The attempts to overcome these limitations are being made, the use of fluorescence resonance energy-transfer (FRET) partially addressees the inefficient excitation problem [29]. Further attempts to use the fluorescence life-time [30] and radio frequency modulation [31] to overcome the current deficiencies are being developed.

2.2.5 Pyrosequencing

Pyrosequencing is a single-nucleotide addition method (SNA) and its name is derived from the pyrophosphate (PPi) that is naturally released when the DNA polymerase incorporates a nucleotide into the duplicate strand. It is a non-fluorescence technique that measures the release of PPi. During DNA synthesis, cycles of four dNTPs are separately dispensed into the reaction mixture iteratively. After each dispensation, the DNA polymerase incorporates the dNTP into the duplicate strand. The PPi is released after each nucleotide is incorporated by the DNA polymerase. The released PPi is then converted into ATP by ATP sulfurylase. This ATP is used to convert luciferase reporter enzyme to oxyluciferin, this conversion causes emission of light which is detected using charge coupled (CCD) camera. The light signal is proportional to the number of nucleotides that are incorporated (T, AA, GGG etc) into the duplicate strand and is recorded as a series of perks called pyrogram (see Fig. 2.8). Each signal peak corresponds to the order of complimentary nucleotides incorporated and reveals the underlying sequence of DNA [32, 33]. The limitations of pyrosequecing are low throughout and high background noise. Attempts to improve this are made by introducing dATP-α-S SP isomer and to develop it into a massively parallel microfluidic sequencing platform (Array-based pyrosequencing) [22].



A B

Figure 2.8. The effect of two different dNTPs that are used which dispenses the order of the outcome in the pyrogram profile. Each peak obtained signifies presence of that base. Sussman *et. al.*, 2006

2.2.6 Cyclic Reversible Termination (CRT)

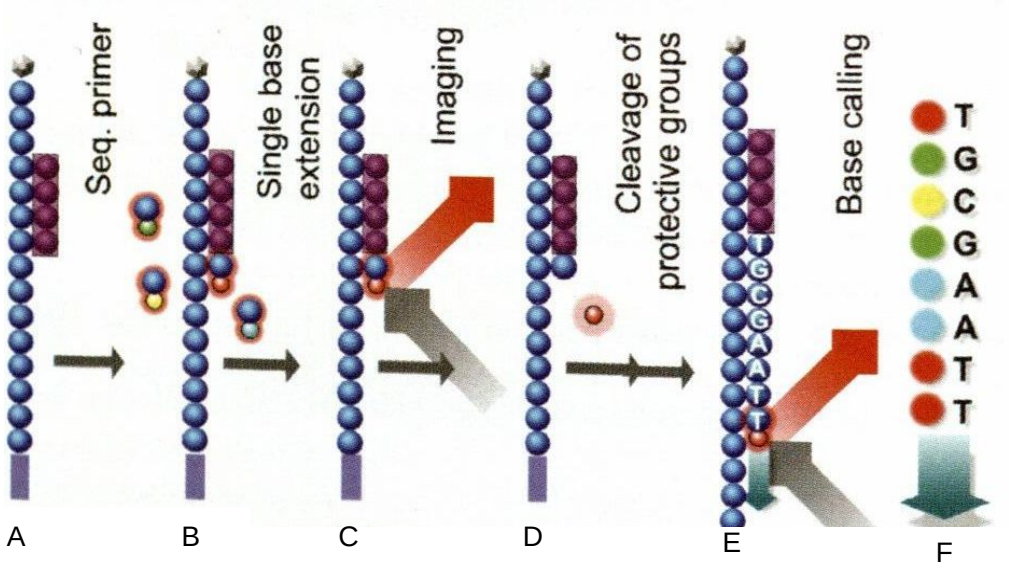


Figure 2.9. The target strands are cleaved with their 3' ends blocked and sequencing primers are hybridized to these strands (A). The DNA polymerase incorporates a reversible terminator (B), fluorescent base attached to the polymerase is imaged (C), this tag is then cleaved off and the terminator is unblocked (D) Process is repeated (E) until the entire sequence is obtained (F). Janitz *et. al.*, 2008

The cyclic reversible termination (CRT) consists of three stages: incorporation, imaging and cleavaging/unblocking. The reversible terminators are modified nucleotides that terminate the DNA strand formation after incorporation since they contain a blocking group at the 3'-end of the ribose group. These terminators are fluorescently labeled and after incorporation enable reading of the bases at each step of the process. The process of CRT is described in Fig. 2.9, the template strand of DNA is attached onto a substrate and is hybridized to sequencing primers (Fig. 2.9.A). Incorporation stage (Fig. 2.9.B), DNA polymerase incorporates the modified nucleotides (reversible terminators) to the template strand. Imaging stage, following the incorporation the DNA sequencing is terminated and the fluorescent nucleotide is imaged (Fig. 2.9.C). In the last stage, the fluorescent tag is cleaved off and the terminator is unblocked (Fig. 2.9.D), this allows the DNA sequencing to be continued. The process is repeated (Fig. 2.9.E), until the template strand is completely sequenced (Fig. 2.9.F) [20, 22].

2.2.7 Sequencing by Ligation

The approach for sequencing by ligation is very much similar to the CRT platform but the DNA polymerase is replaced by DNA ligase and the four nucleotides are substituted by fluorescently labeled oligonucleotide probes [34].

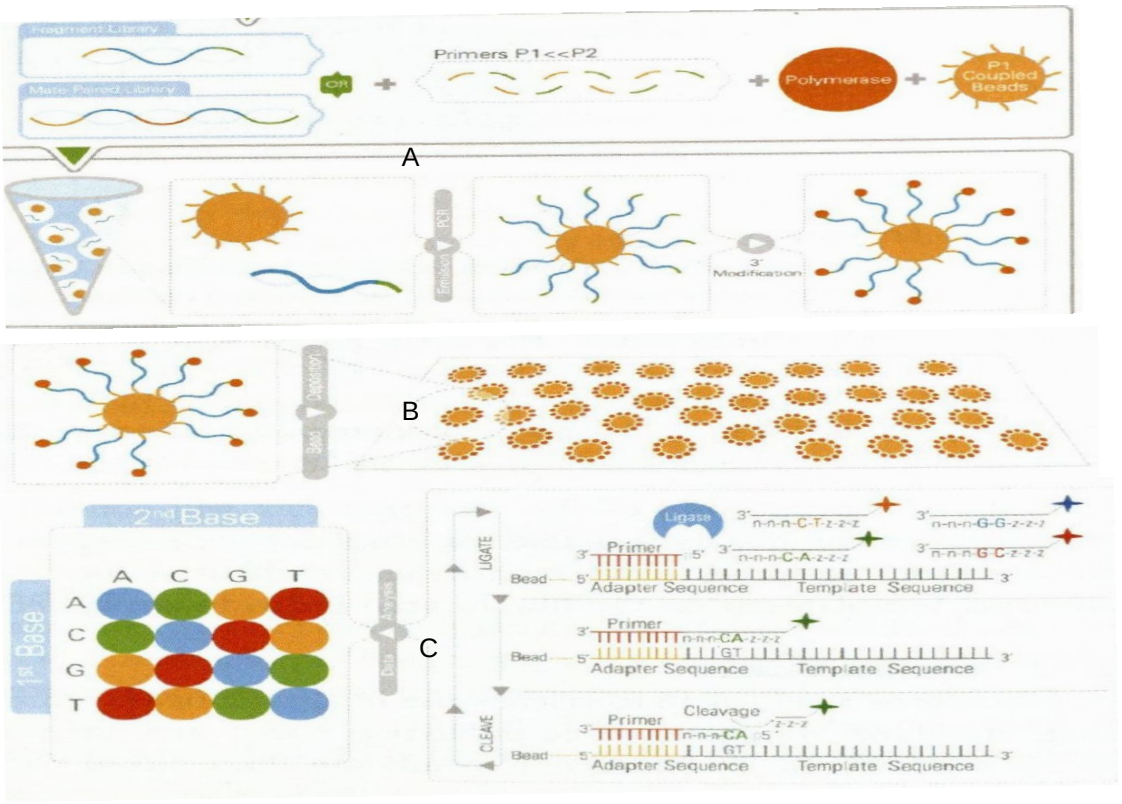


Figure 2.10. Colonal bead populations with 3' modification are prepared in microreactors that contain the template, polymerase and the primered beads (A). These modified beads are then tethered covalently onto a slide (B). Primer hybridizes to the adapter sequence on the beads followed by the attachment of the dye-labeled probes, which are optically read out (C). Janitz *et. al.*, 2008

The process of sequencing by ligase is shown in Fig. 2.10, the DNA fragments to be sequenced are amplified by emulsion polymerase chain reaction (PCR) and captured on beads (see Fig. 2.10.A). These modified beads are then deposited onto a glass slide to form a random array for sequencing (see Fig. 2.10.B). Primers are then added which hybridize to the adapter sequence on the bead. The four color dye-labeled oligo probes compete for ligation to the sequencing primer. Each oligo probe is eight-bases long and the two bases in the middle (4^{th} and 5^{th}) are encoded. The probe specificity for ligation is achieved by analyzing every fourth and fifth base during ligation. After hybridization, ligation and detection, the color of the 4^{th} and 5^{th} bases is recorded. Determination of di-nucleotide sequences every five bases (e.g. …4 5…9 10…14 15.. etc) is achieved by repeating this process. After a few cycles, the sequencing reaction is reset by

denaturing the DNAs and a new sequencing primer with a one base offset (n-1) is introduced. By repeating the ligation procedure we obtain di-nucleotide sequences every five bases but beginning from n-1 (e.g. ..3 4….8 9…13 14.. etc) (see Fig. 2.10.C). Again the cycle is repeated by providing offset to the sequencing primer being introduced by 2, 3 and 4 bases (n-2, n-3 and n-4), until the complete DNA sequencing is achieved[35].

2.2.8 Polony Sequencing

Polony sequencing is a method to obtain colony of DNA that is amplified from a single nucleic acid in situ on a thin polyacrylamide film. The term polony means polemerase colony in molecular biology [36]. The implementation of polony sequencing can be broadly divided into three steps, namely, library construction, template amplification and sequencing. The polonies are developed by diluting a library of DNA molecules into a mixture that contains PCR reagents and acrylamide monomer. Distinct spherical polonies are formed by the amplification of the dilute mixture of single–template molecules (Fig. 2.11). Thus, molecules in two distinct polonies are amplicons of different single molecules but all molecules within a given polony are amplicons of the same single molecule [20].

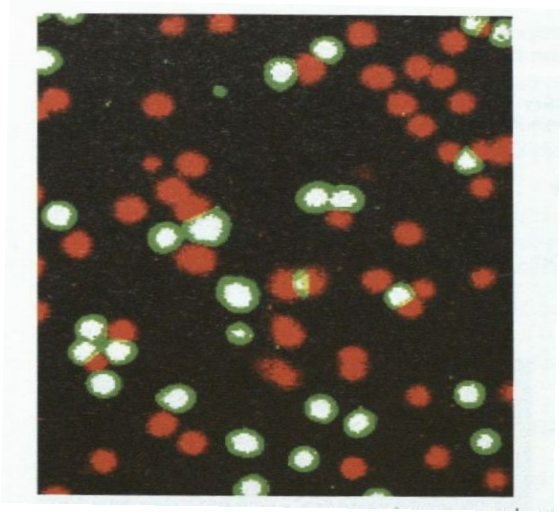


Figure 2.11. Red and green polonies are observed since the polonies are tagged with fluorescent probes. Each colony of DNA (polony) arises from an individual DNA molecule. Janitz *et. al.*, 2008

The amplification primers usually consist of a 5-acrydite modification which causes it to be covalently attached to the gel matrix. Thus, after PCR, the same strand of every double-stranded amplicon is tethered on to the gel. Now the gel is exposed to denaturing conditions which enables efficient removal of the strands that are unattached. The DNA sequencing is then done using the copy of the strands that are attached onto the gel matrix, thus the full set of amplified polonies is used in a highly parallel manner for DNA sequencing by synthesis as described earlier in section 2.2.4 [20].

2.2.9 Sequencing by Hybridization

Sequencing by hybridization is a well known concept and is a method to obtain the sequence of the target DNA strand by allowing it to hybridize to the perfectly complimentary probe. The microarray technology uses single-stranded DNAs of known genes and attaches them onto the silicon surface/substrate. Each well of this microarray has a unique sequence of the single-stranded probe DNAs attached. The strands that are to be tested is first fluorescently labeled and then applied to the DNA chip. The target DNAs attach to the complimentary tethered (template) strand sequences. The chip is then washed which causes all the unbound DNAs to be rinsed away while perfectly adhered target DNAs will continue to remain bonded [37] (Fig. 2.12).

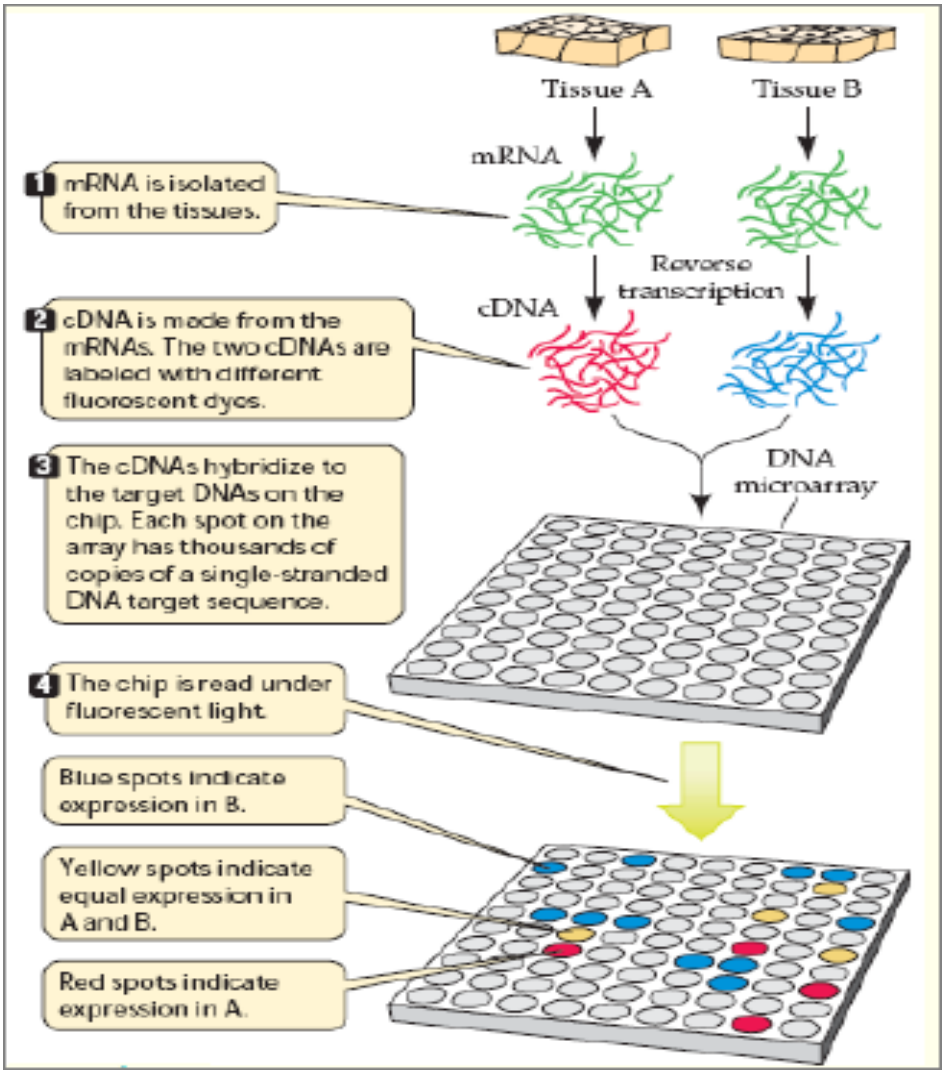


Figure 2.12. The DNA microarray technology is based on the principle of sequencing by hybridization and optical readout measurement. Purves *et. al.*, 2003

Finally, the chip is illuminated with UV light to excite the fluorescence tags and the resulting image captured is shown in Fig.2.12. The output indicated the presence/absence of the target sequences of probed sequences and the intensity of the fluorescence indicates the frequency of the probed sequences.

	Probe type	Probes/gene	Printing method	Detection method	One- or two-color system
Spotted cDNA arrays	cDNA	Investigator dependent	Standard robotic printing	Cy-3 and Cy-5	Two-color
Affymetrix	25-mer	11–20	Photolithography	Fluorescent	One-color
Agilent	60-mer	1(most genes)	*In situ* ink-jet printing and synthesis	Fluorescent	Two-color
ABI	60-mer	1	Printing of prefabricated oligos	Digoxygenin (DIG)	One-color
CodeLink	30-mer	1	Piezoelectric dispension	Strepavidin-Alexa Fluor 647	One-color
NimbleGen	24-mer	Investigator dependent	*In situ* photo-directed (micromirror) synthesis and arraying	Fluorescent	One- or two-color

Figure 2.13. The technology adapted by different companies to produce varies varieties of microarrays. Hofmann *et. al.*, 2006

The microarray technology which utilizes this technique is dominating the current market in terms of sequencing technologies. Fig 2.13, gives a summary of the variations in this technology adapted by the different commercial companies.

CHAPTER 3

DNA SEQUENCING USING NANOPORES

3.1 History

The concept of DNA translocation in nanopores and utilizing it for sequencing rely on a combination of many ideologies. The origins can be traced back to 1940 when Wallace Coulter was attempting to standardize the size of paint particles when he ran out of paint. So he used his own blood as a substitute and found that even the red blood cells could be detected as they pass through a narrow aperture driven by a pressure difference as they restrict the flow of ions. This technique was termed as “resistive-pulse technique” [38] and was used by Coulter Electronics Company to develop market instruments for blood cell counting and cell sizing. The use of nanopores began at General Electric by DeBlois and Bean, as they used track-etched nanopores allowing detection limit to 60nm. They also introduced the concept of electrophoretically driving charged particles through the nanopore [39]. On the organic front, biological nanopores and ion channels were being explored by Hodgkin, Huxely and Hille [40]. Ion channels are transmembrane pores that allow the passage of ions (charged particles) into and out of a cell down the electrochemical gradient [41]. The patch clamp technique developed on the basis of voltage clamp method by noble prize winners Neher and Sakaman in 1970s enabled to record currents from single ions channels [42].

These developments and the dreamed idea of David Deamer for analyzing DNA with nanopores in 1989 on a cross-country road trip was attained a few years later. In 1996, the concept for analyzing DNA in nanopores was realized by Deamer and Kasianowicz [43] and they patented technique with Baldarelli and Church who had independently developed a similar method [44].

3.2 Biological Nanopores

The organic pore of α-hemolysin is the most extensively used for detection of nucleic acids. The pore is formed by 33-kDa protein secreted by Staphylococcus aureus. Fig. 3.1 shows a cross sectional view of the α-hemolysin channel at 1.9A^{o} resolution as revealed by X-ray crystallography [45, 46]. From *cis* to *trans* side, the mouth of this channel is 2.6nm, then a vestibule of 4.6nm, followed by a limiting aperture of 1.5nm diameter with a long stem approximately 5nm in length and the exit diameter of 2nm.

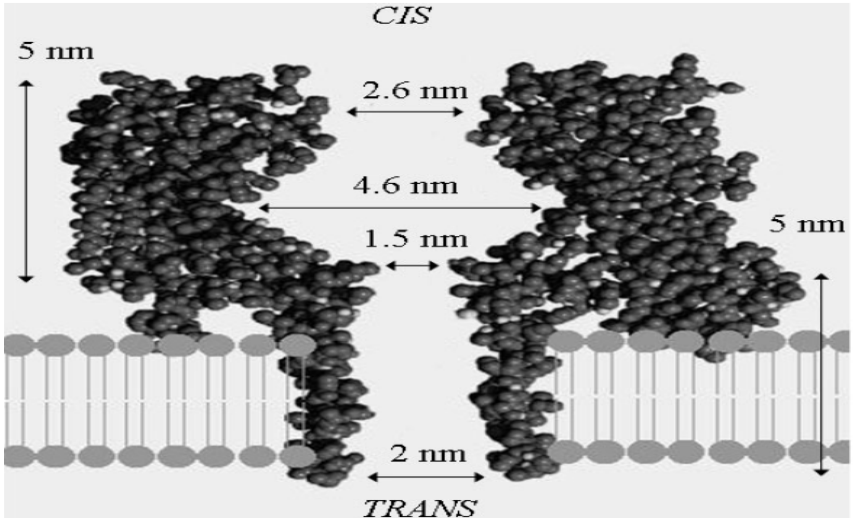


Figure 3.1. The cross section of an alpha-hemolysin channel which is embedded in a lipid bilayer with channel dimensions adapted from Deamer *et. al.*, 2000, Nakane *et. al.*, 2003.

Kasianowicz *et. al.* in 1996 demonstrated that the single-stranded RNA and DNA molecules can be driven through ion channels in a lipid bilayer membrane under the application of an electric field [43]. During translocation, nucleotides within the polynucleotide must pass through the α-hemolysin pore in sequential, single-file order because the limiting diameter of the pore can accommodate only a single strand of RNA or DNA and each polymer as it passes through the membrane as an extended chain, it partially blocks the channel. This causes a transient decrease of the ionic current whose duration is proportional to the length of the polymer translocating. Thus the channel blockades could be used as to measure the polynucleotide length and with further modifications this might lead to high-speed detection of the sequence of bases in

single molecules of DNA or RNA (see Fig. 3.2)[8]. The polymer translocation through the alpha hemolysin pores can be achieved under the influence of an applied voltage bias or free motion or getting a mechanical force to pull the DNA in a particular axial direction [47].

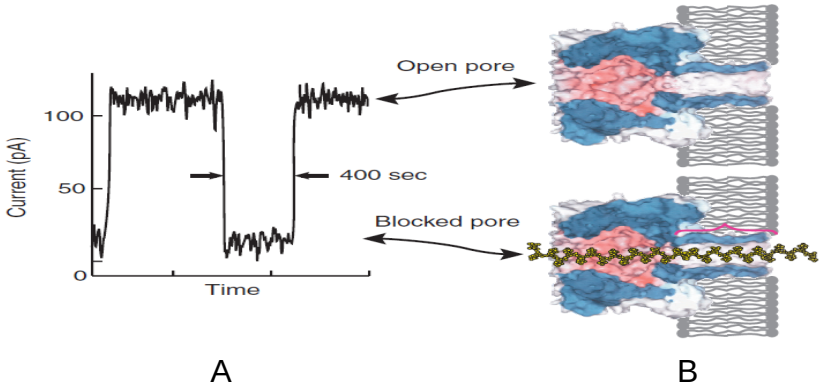


Figure 3.2. The schematic shows the principle by which sequencing of DNA is done using a biological nanopore, (a) shows the modulation in current measurement with the open pore current when there is no DNA (b-top) and the dip in the current due to the presence of DNA which blocks the pore (b-bottom) Branton *et. al.*, 2008

The experimental set-up used by Kasianowicz was modified by Akeson *et. al.* to allow lower noise levels and lower analysis volume, schematic is shown in Fig. 3.3 [48]. They used this set-up to compare the translocation of the RNA homopolymers [poly(C), poly(A) and poly(U)] and also a combination of poly(CA).

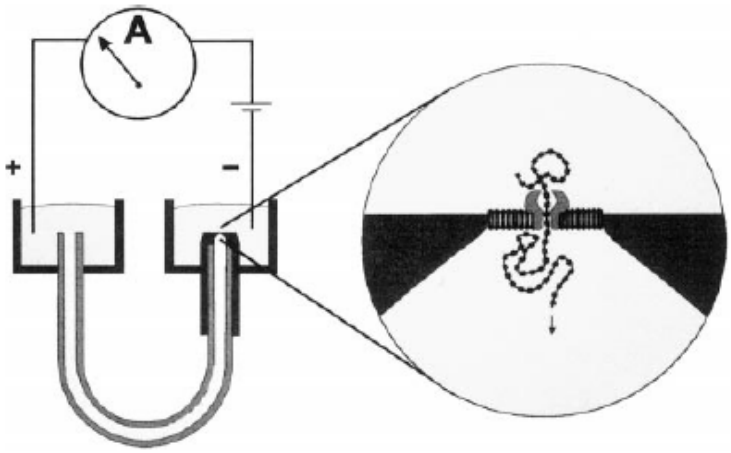


Figure 3.3. The improvised electrical set up provided by Akeson *et. al.*, DNA polymers are driven through the alpha hemolysin channel inserted into the lipid bilayer, by an applied voltage of 120 mV, adapted from Akeson *et. al.*,1999 .

They observed characteristic depths and durations for each of the homopolymers. Fig. 3.4 indicates the difference in the ionic current measurements for comparison between poly(A), poly(C) and poly(CA). The open pore (base line) ionic current at 120mV was 126pA and each of these homopolymers gave a trait significant to its length and composition. The poly(A) had blockades of 22±6μs per nucleotide and had reduced the current to ~20 pA (85% blockades) while on the other hand the poly(C) had blockades of 5±2μs per nucleotide and had reduced the current to ~5 pA (95% blockades). The 35%-55% blockage observed is termed to be due to partial entry into the pore and is not valid for the length dependent ionic current measurement. Further, for the combined poly(CA), we can see a sequence of 85% and 95% blockages indicating the combination of the nucleotides.

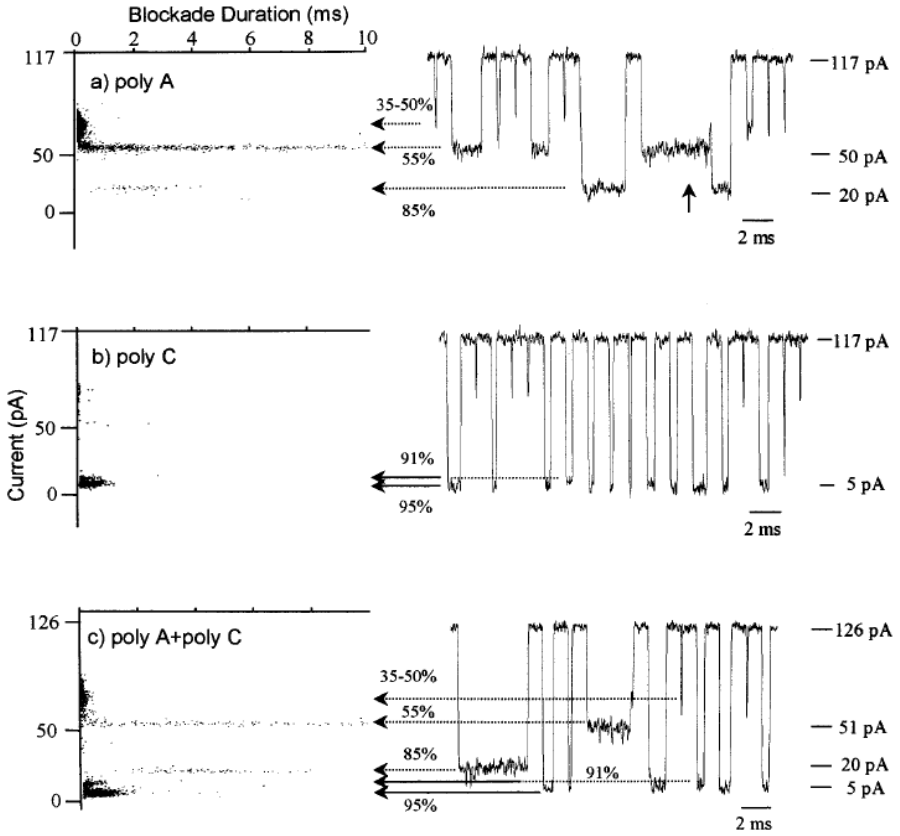


Figure 3.4. The distinguishing current measurements as obtained by Akeson *et. al.*, 1999 which helps detect the different homopolymers based on the depth of the current dip and its duration.

But these differences cannot be correlated directly with the nucleotide sequence of RNA as this can be also due to the modified difference in the secondary structures adopted by these homo-polymers as indicated in schematic in Fig 3.5.

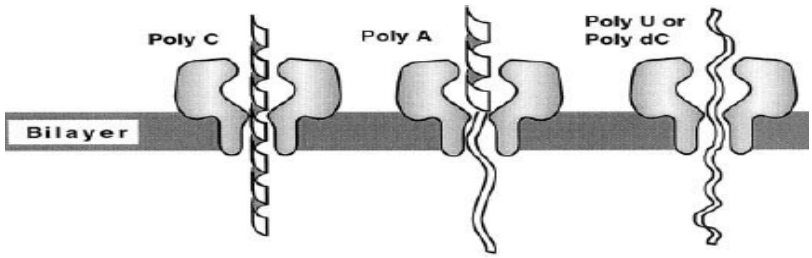


Figure 3.5. The different configurations adapted by the secondary structures of the translocating polymer as they translocate through the nanopore, adapted from Akeson *et. al.*,1999 .

Meller [49-51] extended this knowledge of DNA translocation by studying the translocation of homopolymers and diblock copolymers. The statistical summary of results obtained by Meller *et. al.* is shown in Fig. 3.6, and it helps us to derive the conclusion that on the basis of well defined statistical parameters in this case current peak value (I_P) and translocation duration peak (t_{p1}) help us discriminate between polynucleotides of similar length with the nanopores. Furthermore, the parameter 't_D' alone might help to discriminate between individual polynucleotides on an individual nuclotide basis in case of some favorable cases.

Polymer	I_{P1}	t_{P1}, μsec
$(dA)_{100}$	0.126 ± 0.012	192 ± 10
$(dC)_{100}$	0.134 ± 0.010	76 ± 4
$(dA)_{50}(dC)_{50}$	0.128 ± 0.010	136 ± 7
$(dAdC)_{50}$	0.141 ± 0.011	177 ± 9
$(dC)_{50}(dT)_{50}$	0.140 ± 0.011	137 ± 7
$(dCdT)_{50}$	0.144 ± 0.012	82 ± 4

Figure 3.6. Meller *et. al.*, 2000 showed the difference in the peak current (Ip) value and the translocation duration for the difference polymer strands considered.

Meller *et. al.* also concluded that there is a strong temperature dependence of the statistical parameters considered. The graph in Fig. 3.7 gives us a clear indication that the peak of the translocation time (t_p) of the homopolymers and the diblock copolymers varies significantly with temperature (T) with T^{-2} dependence. At lower temperatures the measurements are highly sensitive while at high temperatures peak of the translocation time converges. This strong T^{-2} dependence on temperature could lead to the possibility of controlling the translocation speed and enhancing the differences between the numerous types of polymers.

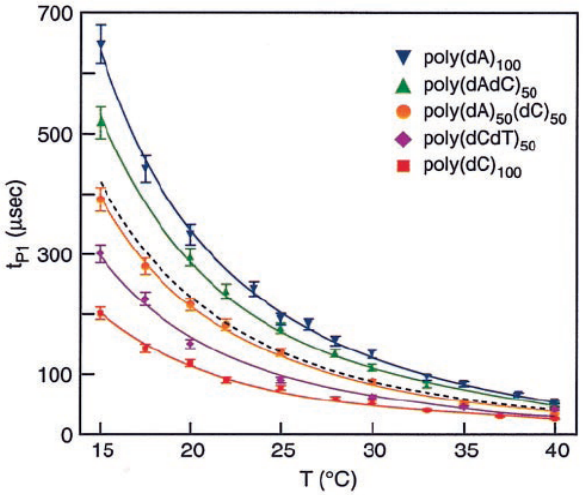


Figure 3.7. The effect of temperature on the translocation duration (tp1) of different polymers adapted from Meller *et. al.*, 2000.

The translocation process through the nanopores was identified to be more complex than mere passage of bases by Mathe *et. al.*, their molecular dynamics studies concluded that the global orientation of the DNA also have a great influence on the DNA-channel interaction. They had observed that the 3'-5' oriented DNA translocation (trans to cis direction) is two times slower than the same sequenced poly (dA) strand with orientation 5'-3' orientation. This shows that the DNA translocation in nanopores also involves tilting of the bases and stretching of the ssDNA within the nanopore [52]. Their view is also supported by other simulation studies by Butler *et. al.* and Wang *et. al.*[53, 54]. Wang *et. al.* had shown that it is possible to detect and differentiate between all combinations of phosphorylation that is present or absent at either end, due to their finding that phosphorylation affects the ease with which the relevant end of the molecule can enter the pore [54]. It has also been shown that in a α-hemolysin pore with 1M KCL solution and 120mv potential from cis to trans results in a current of approximately 90pA, while at the same potential from trans to cis the current is approximately 120pA [46].

The driving Voltage is a major influencing parameter in the translocation of DNA through nanopore and Henrickson *et. al.* examined its effect on the mechanism by which individual DNA molecules enter nanometer-scale pores [55]. An exponential relation was obtained between the

voltage applied and the event frequency. The blockade frequency was found to be proportional to the polymer concentration, which increases exponentially with the applied voltage, and the DNA enters the pore more readily through the entrance that has the larger vestibule. Also, a threshold voltage was obtained below which events did not occur due to the random diffusion and repulsive forces [55]. The dependence of event frequency on voltage was predicted by Nakane *et. al.* which holds true even at higher voltages, it has a derived capture rate of [46]

$$R = 4 \cdot C \cdot D \cdot a \cdot P(V) \qquad (3.1)$$

where *C* is the polynucleotide concentration, *D* is the diffusion coefficient, *a* is the pore radius, *P(V)* which is estimated based on Monte Carlo simulations is the probability that a polynucleotide which collides with the pore will then translocate through it [56]. Aksimentiev *et. al.* has summarized the characteristics of DNA translocation in the alpha hemolysin pores using molecular dynamics and have mapped the Ionic conductance, Osmotic permeability and the Electrostatic potential using molecular dynamics [57-59]. Furthermore, there have been attempts to make self assembled structures using these channels which act as a cavity to trap organic molecules and transfer from side to side [60].

3.3 Solid-State Nanopores

In spite of these extensive studies and proven to be useful for some interesting translocation experiments, biological pores are limited by their fixed size and sensitivity to environmental conditions such as pH, salt concentration, temperature, and limited lifetime. Fabrication of nanopores from solid-state materials presents obvious advantages over their biological counterpart such as high stability, control of diameter and channel length, adjustable surface properties and the potential for integration into devices and arrays [61].

It is only in recent years that advances in nanotechnology have enabled controlled fabrication of nanopores with diameters as small as 1nm. Pores with true nanometer dimensions have been fabricated in various ways such as etching a hole into an insulating material [62], using tightly focused e-beams [63], ions beam sculpting [64] or thermal shrinking [65]. Due the fabrication

process, most nanopores have a wide inlet/outlet and a narrow channel in the middle. Besides nanopores, slit like nanochannel with one side open has also been fabricated on substrates for DNA diffusion studies [66]. DNA diffusion in channels of a few hundred nanometers has been reported by Dekker *et. al.* [67], Matin *et. al.* [68], and Majumdar *et. al.* [69]. The state of art of the nanopore development has been summarized in Healy *et. al.* [70].

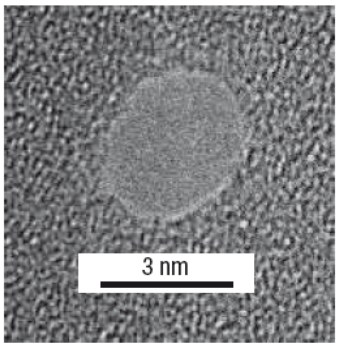


Figure 3.8. A TEM images of a 3nm nanopore fabricated using focused ion beam technique. Dekker *et. al.*, 2007

DNA detection by a synthetic nanopore was first reported by Li *et. al.* in 2001. They had observed blockade events for the translocation of 500bp dsDNA through the nanopore [64]. Most work to date has focused on dsDNA length measurement [71-74]. Each of them considered different polymer length and pore diameter and thus cannot be used for comparison with each other. Li *et. al.* had used a 10nm pore with 3kb and 10kb dsDNA strands for comparative study and had observed non overlapping peaks in the event duration histogram [71]. Mara *et. al.* did similar studies but with a 4nm pore and shorter dsDNA strands with length of 286bp, 974bp and 4126bp, they observed corresponding peaks but significant overlapping of the peaks [72]. Their observations were partially supported by Storm *et. al.* who had conducted studies with 6.6kb, 9.4kb, 11.5kb, 27.5kb ,48kb and 96 kb fragments and observed overlapping peaks only for 6.6kp with 9.9kp and 9.9kb with 11.5kb [74]. These observations and results obtained raise the concern over ability to resolve DNA length wise, since peaks due to shorter lengths are completely overlapped by those of longer lengths.

Heng *et. al.*, using molecular dynamics, studied the translocation behavior of DNA through the nanopore having diameter of 3nm and lesser. They found that ssDNA can be discriminated from dsDNA based on the duration and magnitude of the blocking current that is measured when these DNA translocate through the synthetic nanopores [73]. Researchers have made attempts to investigate the mechanical properties of DNA using an electric field-induced translocation of single molecules through a nanometer diameter pore [69, 75]. A threshold of 60pN is identified for translocating dsDNA through the nanopore [76]. Ho *et. al.* have measured the ionic conductance across nanopores as a function of time, bath concentration and pore diameter [63]. The modulation of pH and its effect on dsDNA translcoating through the nanopore has also been studied [62, 77, 78]. Fologea *et. al.* observed that increasing the pH beyond a certain threshold causes the dsDNA to denature into ssDNAs [77]. Molecular dynamics studies have enabled to get better insight into this organic nanoscale system, Muthukumar *et. al.* mentioned that the vestibule plays a significant role by providing an entropic trap and thereby limitedly slowed down the process. But interestingly their claim that they had observed no correlation between translocation time and blocked current which is converse to what has been found experimentally [79].

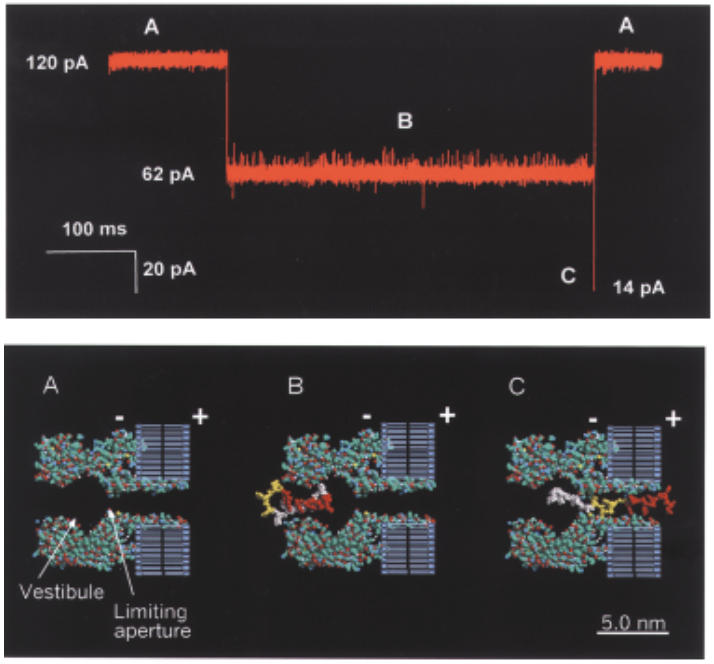


Figure 3.9. The characteristic dip in the current measurement as the HPL strand translocates through the biological nanopore. There is an initial dip the point at which the HPL blocks the channel, then it unzips which causes a deeper dip followed by a sharp rise to the original baseline, adapted from Vercoutere *et. al.*, 2001.

When a HPL DNA passes through these nanopores a characteristic pattern with its translocation is observed which provides nucleotide resolution. Vecoutere *et. al.* observed that when this technique is applied, the current from the baseline initially falls to an intermediate level (termed as the shoulder region) and remains there for some time (Fig. 3.9) [80],then briefly drops to a lower level before returning sharply to the initial level. Fig. 3.9 shows the predicted stages of this specific pattern. The hairpin first enters the wide vestibule of the pore, remains there for some time until random thermal fluctuations combines with the electric-field induced force, to unzip it, and then rapidly translocating through the pore.

The energy required to unzip a hairpin increases with its length and so does the durations of the shoulder regions and the degree of blockage. This is because for longer HPL DNAs, large and statistically less frequent, thermal fluctuations would be required to unzip the DNA. While larger HPL occupy more area of the vestibule thus displacing more ions. The shoulder region and

the blockage depths are highly sensitive to the HPL length and can readily allow single base-pair length difference to be identified between the event classes [80-82].

3.4 Bio-Functionalized Nanopores

Though "bare" nanopores have provided useful information about biophysical properties of DNA, the true potential of important developments like genome sequencing, protein detection, etc are difficult to be realized with bare nanopores. Speed of polynucleotide translocation through the nanopore is one of the major concerns since the resolution of today's electronic detection methods limits the amount of information gathered at these high speeds [83]. Reducing the speed of DNA translocation not only helps gather more information with reasonable bandwidth requirements on the measure systems [84], but also to detect the individual nucleotide thereby enabling gene sequencing. Researchers have also raised concern over difficulties in wetting bare nanopores manufactured by regular methods [63, 65].

This has prompted researchers to coat these bare nanopores with structures that would facilitate increased time of DNA within the pores. Siwy *et. al.* embedded gold nanotube within a mechanically and chemically modified polymeric membrane which allowed passing ionic current through the nanopore [85]. This system did not function on the transient current pulses due to the entry of the protein but by using the protein analyte binding tendency to the biochemical molecular recognition agent immobilized at the inlet opening of the nanotube. The protein molecule once attached at inlet of the pore, plugs the nanopore effectively due to comparable diameter and completely blocks the ion current, thus leading to a detection signal [85]. Kohli *et. al.* had shown that single-base mismatch selectivity can be achieved with functionalized nanotube membranes [86]. Selective permeation was achieved since the membranes recognize and transport only those DNA strands that are complementary to the transporter strand attached in the inner walls of these nanotubes [86]. They had measured flux in a large number of pores in the filter membrane after the simultaneous passage of molecules through it. However, this method failed to provide single molecule translocation signatures. Iqbal *et. al.* later developed single-molecule electrophoretic transport measurements of ssDNA through HPL functionalized nanopores [9] (see

Chapter 1). Such and similar functionalization schemes can be used for a variety of ligand-receptor combinations of significant importance, and the solid-state functionalized nanopore can serve as next generation of sequencing tools, whereas a pore functionalized with specific probe can be used as detector of specific nucleotide(s)/biomarkers [9]. Wanunu *et. al.* have developed a high-throughput single-molecule method for evaluating small molecule binding to DNA using nanopores of ~3nm. The regular method to measure the residual current of native DNA is modified and their label free method measures the shift in residual ion current which results from threading of the dye-intercalated DNA molecule [87].

3.5 Challenges in Nanopore Sequencing

Transition of nanopore sequencing from research level to practical molecular diagnostics will need improvements in technical procedures, robustness, accuracy, and cost [20]. The ionic current blockades caused by the translocation of the ssDNA are not yet sensitive to individual nucleotides. The minimum length of the pore in the experiments that reported length determination of ssDNAs on their passage through the nanopores is 5nm. This means around 10-15 bases are within the nanopore at a time and the resulting ionic blockade is a combined result of these nucleotides [8]. Thus for an ideal system, distinct electrical signals from the space between the bases is to be obtained. Another challenge is to minimize the errors in results attained even though the method is sensitive enough, they are known to be computationally intensive and chances are that false readouts are obtained. For 3-million data points, a 1% false negative can surely overwhelm the real matches. 25kb is the maximum length of the DNA that has been analyzed which shows it ability for a high throughput [50]. The key challenge for the Nanopore sequencing device is to reduce the speed of DNA translocation in these nanopores, controlling the translocation and reducing the fluctuations in translocation kinetics due to the pore-surface interactions [8]. Another challenge that might be an issue for the next generations sequencing device is the huge amount of data that is generated. It is predicted that in case of personalized medicine for 10% of US population, would generate 10^5 TB FASTA format data that will have to be searched [20].

3.6 Summary

Electrophoretic methods lead the way in comparison with any other genome sequencing methods, mainly because of its high throughput (longest read out length) ability. The integration of the technology advances in all the different fields including, instrumentation, fabrication, microfluidics, software control, automation and informatics will hold the key to a robust DNA sequencing platform. The cost per base is substantially lower for these current modalities that are being developed in comparison with current Sanger technology (Gold Standard). Furthermore, the ability for high throughput, decreased sequencing time and streamline sample preparation are the reasons for genome centers and commercial enterprises are readily adopting these new technologies. Most of these new technologies have a very short read out length. These technologies being developed could surely be modulated for other applications within the genomic world, like single nucleotide polymorphism (SNP) detection or expression analysis. The fabrication limitations of nanopores, which are being improved still remains a major bottleneck. The DNA sequencing in nanopores is a potential candidate for a major revolution in form of the de novo sequencing modality and brings the future of personalized medicine within our grasps.

CHAPTER 4

FUNDAMENTALS OF MOLECULAR DYNAMICS

4.1 Introduction

The nature of matter is found in the structure and function of its constituents building blocks. This has led to the questions that are the bulk properties of matter a consequence of underlying interactions among the constituent atoms or molecules? To understand these puzzles and in view of the complicated nature of molecular motions, molecular models are used rather than attempting to deduce microscopic behavior directly from experiments. Molecular Dynamics (MD) simulation provides a methodology for detailed microscopic modeling at molecular level. It enables capture of motion of these molecules thereby helping us understand how the positions, velocities, and orientations change with time. As an analogy, MD constitutes as a motion picture that follows molecules as they are moving to and fro, turning, twisting, colliding with one another or with their container.

The theoretical basis for MD embodies contributions from elite scientists namely Newton (Laws of motion), Laplace (solution to many-body problem), Babbage (concept of the computer), Euler (relationship between the trigonometric functions and the complex exponential function) and Hamilton (equations of motion). This chapter gives a brief overview of the fundamentals of MD simulation studies.

4.2 Statistical Mechanics

Statistical mechanics forms a link between the energy of a molecular system and its macroscopic thermodynamic functions of the N-body system through rigorous mathematical expressions. It enables us to examine the energetics and mechanisms of conformational changes that take place in proteins and polymers. It serves as a tool to predict macroscopic phenomena

from the properties of individual molecules making up the system through time averages over corresponding values in various microstates. For practical purpose, ensemble averages are considered instead of time averages, since for the time averages the molecular dynamics simulations must pass through all possible states corresponding to the particular thermodynamic constraints which make it very complex and time consuming. Thus statistical mechanics deals with ensemble averages,

$$G = \frac{\int G(r1,...r_{N_m}) e^{-\beta U(r_1,...r_{N_m})} dr_1...r_{N_m}}{e^{-\beta U(r_1,...r_{N_m})} dr_1...r_{N_m}} \qquad (4.1)$$

where *(r_i, i= 1,…N)* are the coordinates, $\beta = 1/k_B T$ **,** and k_B is the Boltzman constant and *T* is the temperature.

An ensemble is a collection of all possible systems which have the same macrostate but different microstates. The types of ensembles are,

Microcanonical ensemble (NVE): The thermodynamic state characterized by a fixed number of atoms, N, a fixed volume, V, and a fixed energy, E. This corresponds to an isolated system.

Canonical Ensemble (NVT): This is a collection of all systems whose thermodynamic state is characterized by a fixed number of atoms, N, a fixed volume, V, and a fixed temperature, T.

Isobaric-Isothermal Ensemble (NPT): This ensemble is characterized by a fixed number of atoms, N, a fixed pressure, P, and a fixed temperature, T.

Grand canonical Ensemble (mVT): The thermodynamic state for this ensemble is characterized by a fixed chemical potential, m, a fixed volume, V and a fixed temperature, T.

Using this fact about ensembles, the statistical mechanics is founded on two postulates about the properties of ensembles. The first postulate is that, the time average of a dynamical quantity in a macroscopic system equals its ensemble average.

$$x = \sum_i P_i x_i . \qquad (4.2)$$

where x_i is the value of x in quantum state i, and P_i is the relative probability of this state. The relative probability P_i equals the number of systems in the quantum state i divided by the total number of systems. Consider a closed box with 2 white balls and a black ball, if a person inserts his hand into this box and blindly removes a ball 999 times, the ratio of black to white balls will be 1:2 (333:666). On the other hand say 999 hands took a ball blindly from 999 boxes with same 2 white balls and a black ball, again the ratio of white balls will be 1:2 (333:666). The former was time averaging and corresponds to pressure measurement with a macroscopic measuring instrument while the latter is the ensemble averaging. Thus time averaging equals the ensemble averaging. The second postulate is that in a microcanonical ensemble all possible states are equally probable, but this postulate has been extended to canonical and grand canonical ensembles.

$$P_i = P_j \tag{4.3}$$

where *i* and *j* are the two microscopic states in a system, P_i and P_j are their associated probabilities. For two microscopic states *i* and *j* in a microcanonical system (fixed NVE), the two ensembles are the same, thus the associated probabilities $P_i = P_j$. Proof of this postulate can be found in statistical mechanics books. [88-90]

4.3 Classical Mechanics

Statistical mechanics gives us the understating that the macroscopic behavior of a system is determined by its partition function which is determined by the energy of the molecular system. The total energy of a system consists of kinetic and potential contributions $E_p + E_k = E_{total}$. Motions of molecules determine the kinetic energy which can be formulated independent of the configuration of the molecular environment. The potential energy is determined by the electrical force field in which the molecules move along with the other molecules of the system, thus potential energy depends on configuration of the molecular environment. Quantum mechanics is preferred over classical mechanics for describing the energy of objects as small as a few atoms.

But due to the simplified conceptual and computational nature of classical mechanics, it is preferred to base the molecular dynamics models using it. [89-92]

4.3.1 Newtonian Dynamics

The molecular dynamics simulation method is based on Newton's second law or the equation of motion which states that a body under a force **F** experiences acceleration **a** related to **F** by

$$F = ma \tag{4.4}$$

where *m* is the mass of the body. Alternatively, force is proportional to the time derivative of momentum

$$\frac{dp}{dt} = F(x) \tag{4.5}$$

where $p = m\frac{dx}{dt}$ is the momentum of the particle in x direction and x is a function of time t.

The acceleration is given by

$$a = \frac{d^2x}{dt^2} \tag{4.7}$$

Thus,

$$F(x) = m\frac{d^2x}{dt^2} \tag{4.8}$$

Analogous equations are valid of y and z direction. For conservative systems force $F(x)$ can always be expressed in terms of negative derivative of the potential energy $U(x)$ w.r.t the position coordinate *x*,

$$F(x) = -\frac{dU(x)}{dx} \tag{4.9}$$

where $U(x)$ results from the electrical interactions of molecular charges.

Multiplying both sides by $\frac{p}{m} = \frac{dx}{dt}$ and integrating over time, we obtain,

$$E_1^{kin} + U_1 = E_2^{kin} + U_2 = E^{kin} + U = cons. \quad (4.10)$$

From this energy conservation equation of classical mechanics, the kinetics energy can be obtained,

$$E^{kin} = \frac{p^2}{2m} = \frac{mv^2}{2}. \quad (4.11)$$

Thus the total energy (sum of kinetic and potential energy) of the system is constant i.e. independent of time *t*. The above formulation of classical mechanics is sufficient for a single particle moving along a trajectory described in a Cartesian coordinate system with conservative force field. [93]

4.3.2 Hamiltonian Dynamics

Equation 4.4, which is Newton's second law of motion is time independent. However, molecular position and forces change with time. Consequently we expect there to be some function of position and velocities whose value is constant in time, this function is called the Hamiltonian H

$$H(r^N, p^N) = const. \quad (4.12)$$

Where the momentum p_i is defined in terms of its velocity by

$$p_i = m\frac{dr_i}{dt} \quad (4.13)$$

Total energy (E) i.e. the combined kinetic and potential energies of molecules in an isolated system is conserved. Therefore, the total energy for an isolated system is considered as a Hamiltonian; then for N spherical molecules, H takes the form

$$H(r^N, p^N) = \frac{1}{2m}\sum_i p_i^2 + U(r^N) = E. \quad (4.14)$$

where the potential energy U results from intermolecular interactions.

By following a sequence of procedures, [90] we can obtain the Hamiltonian equations of motion, these equations are

$$\frac{\delta H}{\delta p_i} = \frac{p_i}{m} = \frac{dr}{dt} \text{ and } \frac{\delta H}{\delta p_i} = -\frac{dp_i}{dt} \quad (4.15)$$

For a system of N spherical molecules, the Hamiltonian equations of motion represent 6N first-order differential equations that are equivalent to Newton's 3N second-order equations.

For a comparison with Newton's second law of motion, the following result can be deduced,

$$F_i = -\frac{\delta H}{\delta r_i} = \frac{\delta U}{\delta r_i}. \quad (4.16)$$

The above equation gives us an understanding of the difference between the Newtonian and Hamiltonian dynamics. In Newtonian, the motion is a response to an applied force while for Hamiltonian, forces do not occur explicitly; instead, motion occurs in such a way that it preserves Hamiltonian function[93].

4.3.3 Phase Space Trajectories

The Newtonian and Hamiltonian dynamics has helped us realize the ability of MD to describe position trajectory. The same study can be extended to include molecular moment. A phase-space is a 6N-dimensional hyperspace used for plotting the positions and atoms of N-atoms. It is composed of 3N-dimensional configuration space in which the coordinate axes are components of position vectors and the 3N-dimensional momentum space in which the coordinate axes are components of momentum vectors.

An example to understand phase-space trajectory is shown in Fig. 4.1, which shows a mass-spring system that is isolated from its surroundings. The stiffness with which the spring resists the displacement due to its expansion or compression is measured by the constant γ. Further, when the spring is moved to a new position *r* from its equilibrium position r_o, the displacement of the mass is measured by $x = r - r_o$. The potential energy of this system is given by

$$U(x) = \frac{1}{2}\gamma x^2. \quad (4.17)$$

Applying Newton's Law of motion, we get

$$F(x) = \frac{dp}{dt} = m\frac{d^2x}{dt^2} = -\frac{du(x)}{dx} \tag{4.18}$$

Therefore, $m\frac{d^2x}{dt^2} = -\gamma x$ thus the force acting on the mass is linear to its displacement.

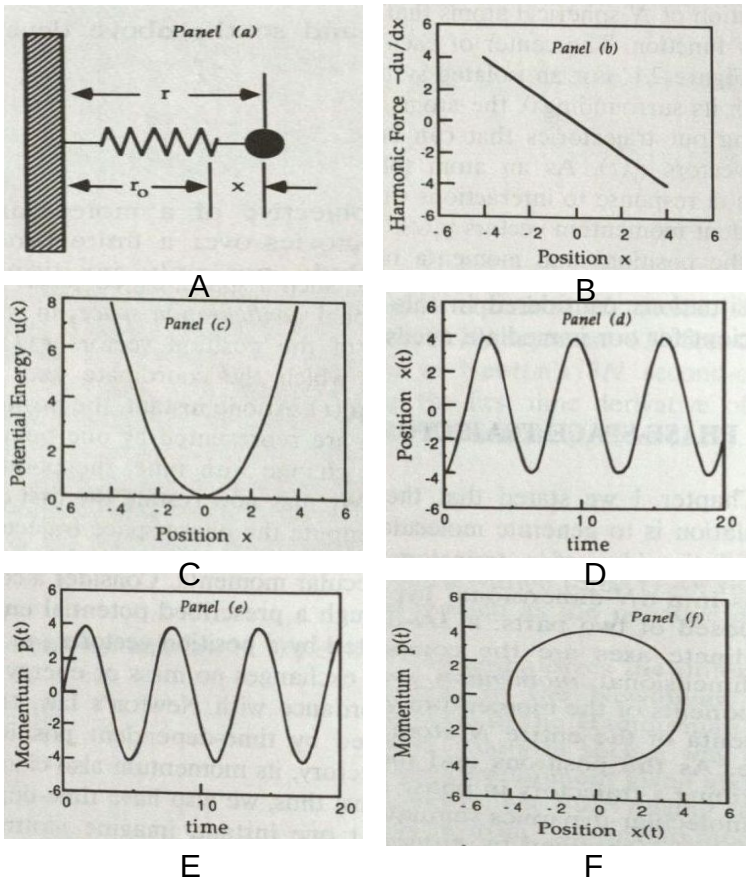


Figure 4.1. Motion of one dimensional Harmonic Oscillator, (a) Ball has mass *m*=1 and spring stiffness (γ) =1. The natural position of the spring places the ball at r_o and the displacement from r_o is measured by *x*. (b) Force exerted by the spring is linear in *x*; slope of this line is γ. (c) Potential energy is quadratic in *x* and symmetric about *x=0* (d,e) Time dependent position *x(t)* and momentum *p(t)*, both of them are sinusoidal but total energy is constant since they are out of phase with each other. (f) Phase-space trajectory obtained from (d) and (e). Haile *et. al.*, 1992

To determine the phase-space trajectory for this simple system, the equation denoting the total energy constant is used,

$$E = E_k + U = const \tag{4.19}$$

where *E* and *U* are kinetic and potential energy of the mass respectively

$$E = \frac{1}{2m} p^2 + \frac{1}{2}\gamma x^2 \tag{4.20}$$

The phase-space for this situation is two dimensional with a position coordinate x and a momentum coordinate p. Based on equation 4.20 we get an ellipse but since m= γ =1, this ellipse disintegrates into a circle. The same relation could be obtained using the definition of momentum and two forms of second law,

$$\frac{dx}{dt} = \frac{p}{m} \text{ And } \frac{dp}{dt} = -\gamma x \tag{4.21}$$

We combine the two equations of 4.21, to obtain the phase-space (phase-plane in this case) trajectory.

$$\frac{dx}{dp} = \frac{p}{m}\left(\frac{-1}{\gamma x}\right) \tag{4.22}$$

The integration of equation 4.22 will give us 4.20

Now that we have the understanding how the position and momentum trajectories are attained, which are governed by interaction potentials [93].

4.4 Force Fields (Interaction Potentials)

Force field provides a depiction of the relative energy or forces of the ensemble for any geometric arrangement of its constituent atoms. This depiction includes the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and electrostatic interactions between atoms The interaction potential can be broadly classified into two classes: intra-molecular and inter-molecular potentials. The former describes interactions that arise from bonded structures and the latter describes pair interaction between distant atoms. The total potential energy of the system can be given as,

$$U_{\text{Total}} = U_{\text{intra-molecular}} + U_{\text{inter-molecular}} \tag{4.23}$$

4.4.1 Intra-Molecular (Bonded) Potentials

The Intra-molecular (Bonded) Potentials describes the covalent energy of a molecule with the help of 2-, 3-, and 4-body interactions of the bonded atoms.

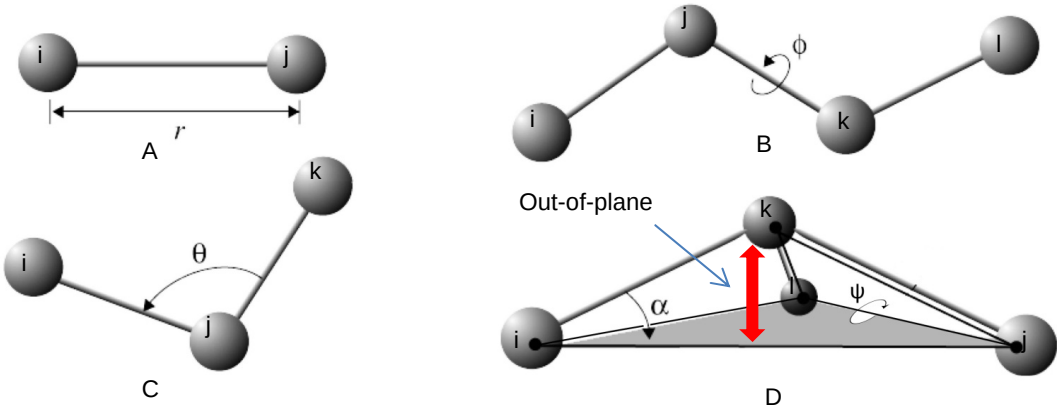


Figure 4.2. The (A) 2- (bond), (B) 3- (angle), and (C, D) 4- (dihedral and improper respectively) body interactions of covalently bonded atoms comprise of bonded potential terms. The bond stretching is governed by 'r', bond angle term is represented by θ, the dihedral angle is represented by φ, the "improper dihedral" angle ψ governs the out of plane angle α. [94]

It is given by

$$U_{\text{intra-molecular}} = U_{bond} + U_{angle} + U_{dihedral} + U_{improper} \qquad (4.24\text{-a})$$

Therefore,

$$U_{\text{intra-molecular}} = \sum k(r_{ij} - r_0)^2 + \sum k_{\theta_{ijk}}(\theta_{i\,jk} - \theta_{eq})^2 + \sum \frac{1}{2} k_{\phi}\left[1 + \cos\left(n\phi - \delta\right)\right] + \sum k_{\psi}\left(\psi - \psi_0\right)^2 \qquad (4.24\text{-b})$$

Two-Body Spring Bond Potential

The harmonic vibrational motion between pair (i^{th} and j^{th}) of covalently bonded atoms is described by 2-body spring bond potential, schematic is shown in Fig. 4.2.A [94]. The equation for it is given by,

$$U_{bond} = \sum k(r_{ij} - r_0)^2 \qquad (4.25)$$

where the distance between the i[th]and j[th] atoms is given by $r_{ij} = | r_i - r_j |$ while r_o is the equilibrium distance and *k* is the spring constant. Alternatively, Morse potential is

$$U(r_{ij}) = K_M \left(e^{-\beta\left(r_{ij} - r_0\right)} - 1 \right)^2 \tag{4.26}$$

is used which allows bond breaking. K_M and β are the strength and distance related parameters of the potential.

Three-Body Angular Bond Potential

The angular vibrational motion occurring between the i[th], j[th] and k[th] covalently bond atoms is described by 3-body angular potential, schematic is shown in Fig. 4.2.B. The angle term helps us understand how the energy of a bond angle changes when it is distorted away from its equilibrium position. The equation for it is given by,

$$U_{angle} = \sum k_{\theta_{ijk}} (\theta_{ijk} - \theta_{eq})^2 \tag{4.27}$$

where θ is the angle formed by the bonds extending between i[th], j[th] and k[th] atoms and θ_{eq} is the equilibrium angle.

Four-Body Angular Bond Potential

The 4-body torsion angle (dihedral angle) potential describes the angular spring between the planes formed by the first three and last three atoms of a consecutively bonded i[th], j[th], k[th] and l[th] atoms, schematic is shown in Fig. 4.2.C. Dihedral or torsion energy for the system describes how the energy of a molecule changes as it undergoes a rotation about one of its bonds. The equation for it is given by,

$$U_{dihedral} = \sum \frac{1}{2} k_\phi \left[1 + \cos\left(n\phi - \delta \right) \right] \tag{4.28}$$

Where n is the periodicity of the angle, ϕ is the dihedral angle, δ is the phase of the angle and k_ϕ is the force constant.

The last term “improper dihedral” describes the energy out-of-plane. Since the dihedral term alone is not sufficient to maintain the planarity of groups, such as sp^2 hybridized carbons in carbonyl groups and in aromatic systems, additional improper dihedral is used. The equation for it is given by,

$$U_{improper} = \sum k_{\psi} \left(\psi - \psi_0 \right)^2 \qquad (4.29)$$

where k_{ψ} and ψ_o are the force constant for the energy term and the equilibrium value of the improper dihedral angle, respectively. Another common angle used to define the distortion due to out-of-plane motions is the angle (α) shown in Fig. 4.2.D. [95-97]

4.4.2 Inter-Molecular (Non-Bonded) Potentials

The non-bonding energy terms help to define the interactions between the atoms of different molecules or between atoms that are not directly bonded together in the same molecule. The overall conformation of the molecular system is defined with the help of these interactions. They are broadly classified as the short range exchange-repulsion and long range dispersion interactions (LJ), electrostatic interactions (elect), and induced or polarized interactions (polar).

$$U_{\text{inter-molecular}} = U_{elect} + U_{LJ} + U_{polar} \qquad (4.30)$$

Therefore,

$$U_{\text{inter-molecular}} = \frac{1}{4\pi \in_0} \sum_{ij-pairs} \frac{q_i q_j}{r_{ij}} + 4 \in \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right] + \frac{1}{2} \sum \mu_i \mathrm{E}_i \qquad (4.31)$$

Electrostatic Potential

The electrostatic potential mimics the energy arising from the electrostatic interactions between two charge distributions. It is repulsive for atomic charges with the same sign and attractive for atomic charges with opposite signs. It is calculated using the Coulombs law,

$$U_{elect} = \frac{1}{4\pi \in_0} \sum_{ij-pairs} \frac{q_i q_j}{r_{ij}} \qquad (4.32)$$

where q_i and q_j are the fractional charges on atoms i and j and r_{ij} is the distance between the two particles, and ε_o is the permittivity of empty space.

Lennard-Jones Potential

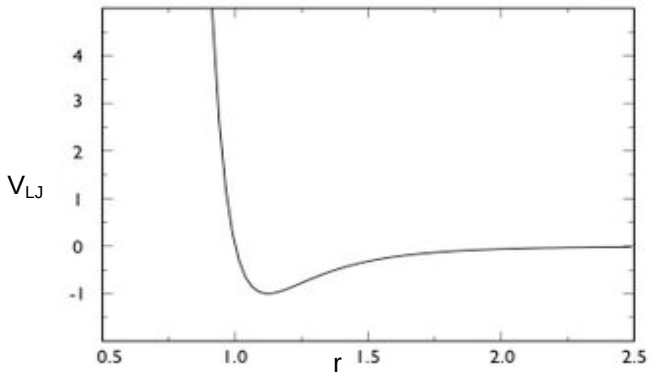


Figure 4.3. The Lennard-Jones (LJ) energy for a pair of atoms, Field *et. al.*, 2007

The Lennar-Jones (LJ) potential helps us define the van-der-Waals forces of interactions. It accounts for the weak dipole attraction between distant atoms (*dispersion* represented by 6^{th} order term) and the hard core repulsion (*short-range repulsion* represented by 12^{th} order term) as atoms become close, usually excluding pairs of atoms already involved in a bonded term.

$$U_{LJ} = 4 \in_{ij} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right] \tag{4.33}$$

where ε is the depth of the potential well, σ is the (finite) distance at which the inter-atom potential is zero and r is the distance between the atoms. The parameters σ and ε depend not only on two atoms, but the cross or unlike interactions i≠j between unlike pairs of molecules, thus are not so easily obtained. To enable their definitions from parameters of single atoms we could use certain combination rules such as Lorentz-Berthelot mixing rules[98], given by

$$\in_{ij} = \sqrt{\in_i \in_j} \quad \text{and} \quad \sigma_{ij} = \frac{1}{2} \sigma_j + \sigma_j \tag{4.34}$$

where i and j denote the i^{th} and j^{th} atomic species.

Polarization Potential

Polarizable force fields enable the charge distribution to respond to the dielectric environment and are of high importance in case of simulation studies involving biological systems. It arises from the fact that the charge distribution of a group of molecule is distorted by interactions with its neighbors. If the polaraziblity of an i^{th} atom denoted by α_i and the field at the atom is E_i (vector quantity) then the dipole induced at the atom μ_i, is

$$\mu_i = \alpha_i \mathrm{E}_i \tag{4.35}$$

The total polarization potential in the charged system is given by,

$$U_{polar} = \frac{1}{2}\sum \mu_i \mathrm{E}_i \tag{4.36}$$

where μ_i is the dipole moment associated with atom *i* and E_i is the electric field experienced at atom *i*. Many commercial software force-fields are developed that can be directly used for atomistic scale molecular dynamics studies, few of these are CHARMM [99], AMBER [100] and Cornell[101].

4.5 Periodic Boundary Condition

The MD systems are relatively small with the number of atoms being around few thousands. For such a small system surface effects like the interactions of the atoms with the container wall dominate. A simulation of such a system with solvent will not provide any information on the bulk characteristic of the solvent (liquid) but only on the interaction of liquid with the container walls. To eliminate the surface effect, periodic boundary conditions (PBC) are used.

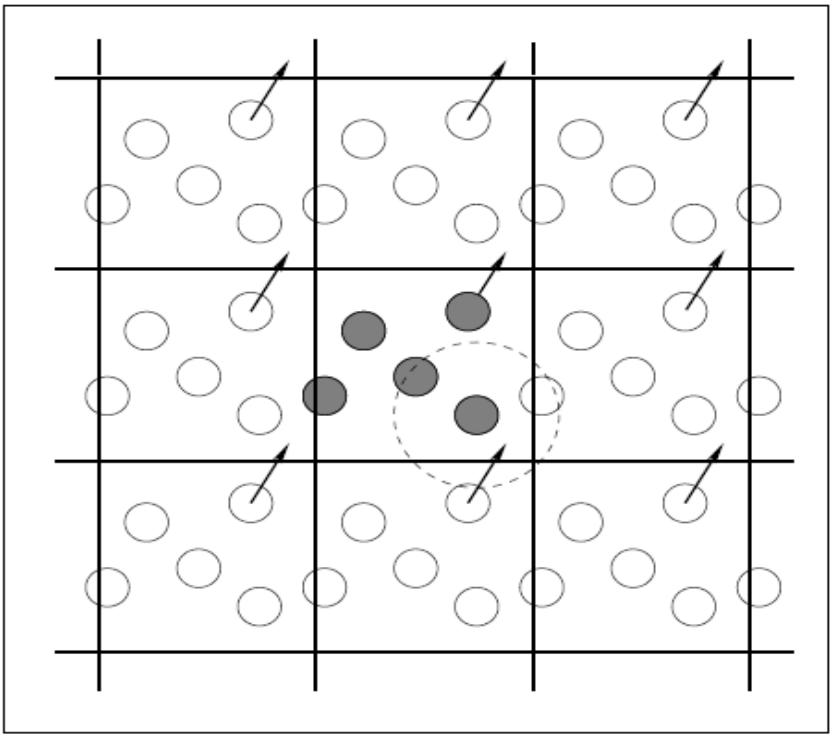


Figure 4.4. Periodic Boundary Condition (PBC), as a particle moves out of the simulation box through one of the bounding face an image particle moves in to replace it. Allen *et. al.*, 1989

The schematic shown in Fig. 4.4 is of a system that is bound free of physical walls. PBC creates an infinite, space-filling array of identical copies of the simulation region. Thus, an atom that leaves the simulation region through a particular bounding face reenters the region through the opposite bounding face. Further, an atom lying within the equilibrium distance from the boundary will interact with atoms in the adjacent copy of the system or equivalently with atoms near the opposite boundary. This is termed as the wraparound effect. The wraparound effect has to be accounted for during integrating the equations of motion and while analyzing the results[90].

4.6 Integration Equation

The atomic positions of all the atoms in the system define the potential energy of that system. Due to its complicated nature there is no analytical solution to the equations of motion and hence has to be solved numerically. There are many numerical algorithms that are available for integrating the equations of motion [89]. But there are certain criteria's to be satisfied such as

conservation of energy and momentum, allowing long time step for integration, and computational efficiency. Based on these, the two popular algorithms are the Leapfrog and Verlet methods since they are accurate and treat energy conservation better than higher-order methods; also their storage requirements are minimal [89].

All Integration methods assume the position, velocities and accelerations can be approximated by a Taylor series expansion,

$$r(t+\partial t) = r\left(t\right) + v(t)\partial t + \frac{1}{2}a\left(t\right)\partial t^2 + \ldots \quad (4.37)$$

$$v(t+\partial t) = v\left(t\right) + a(t)\partial t + \frac{1}{2}b\left(t\right)\partial t^2 + \ldots \quad (4.38)$$

$$a(t+\partial t) = a\left(t\right) + b(t)\partial t + \ldots \quad (4.39)$$

Where *r* is the position, *v* is the velocity (the first derivative with respect to time), *a* is the acceleration (the second derivative with respect to time), etc.

4.6.1 Verlet Algorithm

The Verlet algorithm uses positions and accelerations at time 't-δt' and time 't' to calculate the new positions at time 't+*dt'* and uses no explicit velocity. It has the advantage that it is simple and straightforward and has modest storage requirements. The disadvantage is that it has moderate precision.

Its derivation from Taylor expansion,

$$r(t+\partial t) = r\left(t\right) + v(t)\partial t + \frac{1}{2}a\left(t\right)\partial t^2 \quad (4.40)$$

$$r(t-\partial t) = r\left(t\right) - v(t)\partial t + \frac{1}{2}a\left(t\right)\partial t^2 \quad (4.41)$$

Summing the two equations 4.40 and 4.41, we get

$$r(t+\partial t) = 2r\left(t\right) - r(t-\partial t) + a\left(t\right)\partial t^2 \quad (4.42)$$

4.6.2 Leapfrog Algorithm

The Leapfrog algorithm uses the velocities at time 't+0.5δt' to calculate the positions, *r*, at time 't+δt'. Thus the velocities leap over the positions and then the positions leap over the velocities. It has the advantage that the velocities are explicitly calculated but the disadvantage is that the velocities are not calculated at the same time as the positions.

Its derivation from Taylor expansion,

$$r(t+\partial t) = r\left(t\right) + v\left(t+\frac{1}{2}\partial t\right)\partial t \tag{4.43}$$

$$v\left(t+\frac{1}{2}\partial t\right) = v\left(t-\frac{1}{2}\partial t\right) + a\left(t\right)\partial t \tag{4.44}$$

The equation for velocity will be,

$$v\left(t\right) = \frac{1}{2}\left[v\left(t-\frac{1}{2}\partial t\right) + v\left(t+\frac{1}{2}\partial t\right)\right] \tag{4.45}$$

4.7 Summary

The wealth of potentials available makes MD a useful tool and important method. The rapid growth in computing power and the even greater improvement in the cost –performance ratio has directly benefited MD simulations. Further, the efficient implementation of simulation algorithms on parallel computers has also improved the efficiency and accuracy of MD simulations. MD simulation studies can be applied to fundamental studies like understanding diffusion, transport properties, size dependence or for phase transition studies like phase coexistence, order parameters and to study collective behavior like that of translational and rotational motion, vibration, dielectric properties, fluid dynamics, Biomolecular behavior, structure, polymers transport process, etc. There are phenomena that require length and time scales beyond current capabilities of MD simulations but the field is still comparatively young and has bright future prospects [89, 92].

CHAPTER 5

ATOMISTIC MODEL SIMULATIONS

5.1 Introduction

Eminent modeling of biomolecular and condensed phased systems can be attained with the use of atomistic molecular dynamics simulation [90, 102, 103]. The ability to control the internal motion makes molecular dynamics an ideal tool to study the polymeric systems. There have been many theoretical models to explain the chain dynamics, primarily governed by segment chain interactions and connectivity [104-106]. MD allows one to visualize and obtain a detailed understanding of the polymers behavior, since these simulations have explicit solvents with counter-ions, reliable force-fields and proper representation of long range electrostatics. The all atom system has been successfully implemented to accurately reproduce many experimental results [107-112]. This prompted us to use this powerful tool to analyze and characterize DNA translocation in bare nanopores. The understanding of the dynamics of DNA in nanopores is important for the development of lab-on-chip devices for biomolecular analysis. However, the interaction between DNA and nanopores is still not well understood due to the small length scales of the DNA/nanopore and the dynamic nature of the translocation process.

Studies of electrophoretic transport of DNA in nanochannels have revealed that DNA-channel surface interaction leads to a diffusion rate much lower than predictions by traditional diffusion theory[113]. Extensive experiments and simulations have focused on understanding translocation of a long DNA strand by analyzing the dips in the ionic current due to the blockage of the channel and correlating the time of dip in the ionic current to the polymer length[49, 51, 57, 75, 114-116]. There has also been work done to understand the stretching of the DNA in the nanopore under the application of electric field[76, 79], as well as the effect of temperature and DNA sequence on the translocation kinetics[117].

These works have a common feature: they have considered a DNA with length greater than that of the nanopore and initially located outside the pore. The type and size of the pore influences the electrophoretic mobility of the DNA translcoating through it [72]. It is known that the translocation kinetics outside and inside the pore are different due to the ion accumulation near the pore entrances. Meller *et. al.*, had reported that in case of the biological nanopore, a DNA strand having the length shorter than the pore translocates faster than the one that translocates through a smaller pore [51]. Thus the electrophoretic mobility is higher in the case that length of the DNA strand and the pore is comparable. This can be explained on the basis that when the DNA is outside the nanopore the effective charge on it is significantly less due to the presence of counter-ion cloud that move with it. These counter-ion cloud charges can be disrupted due to sterric effect or charged inner surfaces of the nanopore [118] and this leads to a change of force experienced by the translocating DNA and thus its mobility. A conflicting explanation is the electro-osmotic [119] flow which also occurs as a consequence of surface charges in these narrow channels. Chang *et. al.*, proposed that the electro-osmotic flow will be towards the cathode, thus would oppose DNA translocation causing reduction in the velocity of translocating DNA due to hydrodynamic and electrostatic drag forces. Storm *et. al.* [74] qualitatively supported the power law [120] for polymer which states that the electrophoretic mobility decreases with the increase in DNA length, given by

$$\mu_e \propto L^{(1-\alpha)} \qquad (5.1)$$

where μ_e is the electrophoretic mobility of the translocating DNA, *L* is the length of the DNA and α =1.27 is the power law constant.

Moreover in the reported simulation studies, higher voltages (compared to experimental voltages) are applied to overcome the timescale limitations. This may prevent understanding the factual behavior of the system since the kinetics could be different at reduced bias voltages as compared to higher. This could be one of the reasons that Muthukumar *et. al.* [79] have mentioned

that they had observed no direct correlation between translocation time and blocked current which is surprisingly converse to the experimental results.

In the absence of surface friction, nanochannels would be a free solution environment in which DNA molecules would move with a length-independent mobility. Also, due to entropic effect, stretched DNA relaxes in a nanopore if electric field is turned off. The importance of channel size lies on two folds: first, it constrains the relaxation of the DNA molecule; second, it interacts with DNA bases upon contact. Of course, these two effects are coupled together on the fact that higher confinement leads to stronger interaction. Theoretical models of confined polymers predict hydrodynamic friction coefficients that are dependent on channel diameter and viscosity but not on the electric field [121]. The hydrodynamic friction force on the DNA is given as $f=gv$. The drag coefficient is assumed to be $g=el$ with the friction coefficient per unit length e and the contour length l. However, such friction coefficient ignores the contribution from DNA-nanopore surface interaction. The DNA bases confined in a very narrow channel (<2 nm) are found to be attracted toward the surface, tilting their orientation parallel to the channel surface [52], leading to a large surface friction. But for nanochannel above 3nm, DNA tends to fold in various forms and only a small portion of the DNA base pairs are in contact with the nanochannel surface. Thus, we characterize the translocation at different channel size regimes and determine the velocity-dependent friction coefficient. In this study, periodic boundary conditions are applied at the inlet and outlet of the pore. Thus, there is no ion depletion under higher voltages and DNA is always within the pore. This helps characterize translocation kinetics of DNA inside a nanopore and DNA-nanopore interaction without the entrance and exit effect, eliminating effects due to change of environment.

Nanopores with molecular selectivity are needed to determine DNA conformation and base pair level information simultaneously[122]. To provide such selectivity, simple chemical treatment such as silane coating will not be sufficient[62, 123-125]. Iqbal *et. al.*[9] have demonstrated that such selectivity can be imparted by using tethered DNA hair-pin loops coated on the inner surface of the nanopore. But little is known about properties of these coated DNA monolayers

which are influenced by size, geometry, and surface curvature of the pore[126]. Other factors that also modulate the tethered DNA orientation include initial orientation, attachments with the wall, applied electric field, and the hydrodynamic interactions[47, 117]. Thus, it is important to characterize the interaction between DNA and chemically modified nanopores.

The goal of this chapter is to develop an understanding of the interaction between DNA and nanopore surfaces and the translocation process of DNA by probing the DNA-nanopore interaction mechanisms through molecular dynamics simulations.

5.2 Method

The software used by us to perform full atomistic scale molecular dynamics studies is Nanoscale molecular dynamics (NAMD) [127]. Analysis and visualizations are done with Visual Molecular Dynamics (VMD)[128] . The system used for the simulation consists of a nanopore, a DNA, water molecules and ions. To perform the simulation studies each component of the system has to be built and then integrated to form the entire system.

5.2.1 Building the nanopore

We had decided to build a silicon nitride nanopore since these are the most commonly used nanopores in the experimental world. We have tried to match the experimental conditions in these simulations such that the results could provide better insights to the unknown kinetics of translocation. The pore was built using a single unit of silicon nitride as shown in Fig. 5.1.A. This unit was then replicated and placed side by side based on the equilibrium distance to generate a cubical membrane of the silicon nitride shown Fig. 5.1.B. The cubical membrane was then cut into a more convenient geometry to obtain a hexagonal membrane as shown in Fig. 5.1.C. A hole is drilled of desired radius in this membrane to obtain the final nanopore as shown in Fig. 5.1.D.[129]

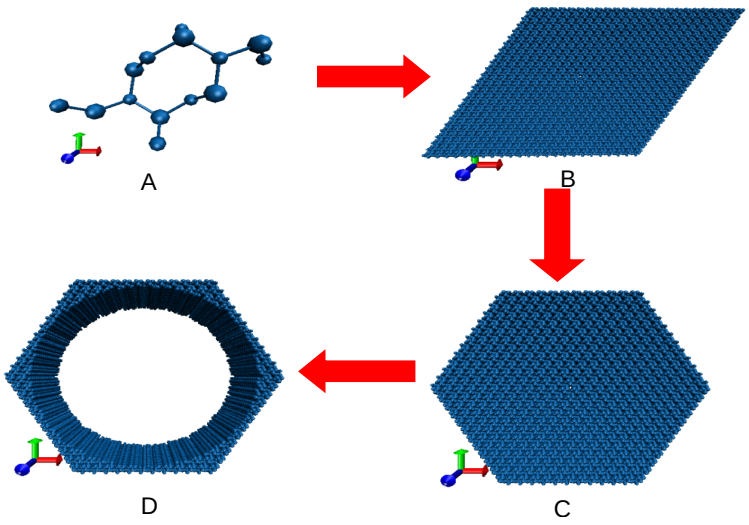


Figure 5.1. Nanopore membrane is formed by repeating the unit cell of silicon nitride (A) to form a cubical membrane (B), followed by obtaining a desired shape for the membrane (C) and finally the nanopore (D).

5.2.2 Building the DNA

In most experimental studies, kilo-bases of DNA are translocated. Due to computational limitations, for simulation studies usually 8-20 base pairs (bp) of DNA are considered.

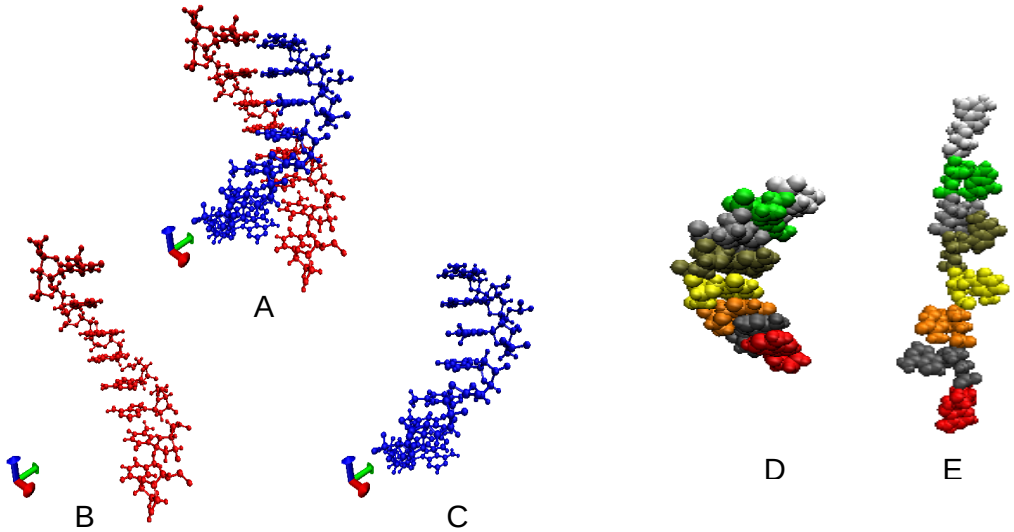
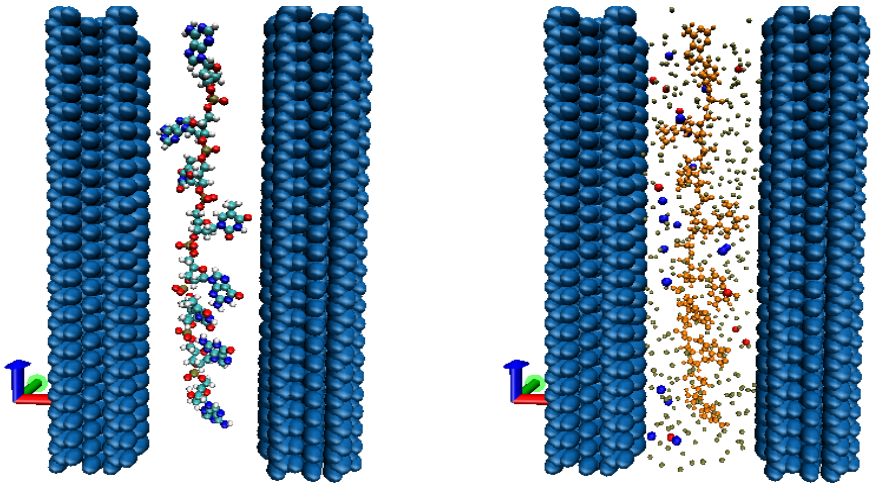


Figure 5.2. The figure shows the (A) double strand DNA obtained from the RCSB data bank for obtaining the pdb (coordinate) file. This dsDNA is then used to obtain two ssDNAs (B and C). The different configurations of the DNA i.e. curled DNA (D) and the straightened DNA (E).

We used a 8-bp dsDNA with randomly selected sequence. The pdb file for the double stranded dsDNA model was obtained from the RCSB protein data bank [130] (Fig. 5.2. A, B and C). The dsDNA obtained was then converted into two single strand DNAs, the ssDNA that we used had the sequence AATTGTGA. The DNA enters the nanopore in a sequential single file nucleotide by nucleotide thus it is straightened and hence the ssDNA obtained was stretched using Interactive molecular dynamics (see Fig. 5.2.D and Fig.5.2.E) before running the simulation.

The nanopore and the DNA obtained were then combined and the positon of the DNA was adjusted such that it is in the middle of the pore as shown in Fig. 5.3.A. This system was then solvated using the automated solvation function of VMD. The solvate function generates water molecules not only inside the nanopore but also outside it. The additional water molecules outside the nanopore and those within 1A inside the nanopore were deleted, as illustrated in Fig. 5.3.B. Ions were added to this system to obtain a Molarity of 0.1M. The .psf (bonding specifications) files needed for the system are constructed using autopsfgen funciton of VMD.



A B

Figure 5.3. The formation of the final system: (A) first the straightened DNA is combined with the nanopore and then solvated and ionized to obtain the final system (B).

The CHARMM[99] force field is used for the DNA, nanopore (silicon nitride), and their interactions. The total potential energy is composed of bonded energies and non-bonded pair interaction energies:

$$U = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{improper}} + E_{\text{elec}} + E_{\text{vdw}} + E_{\text{constrain}} + E_{\text{other}}$$
$$= \sum k_b \left(b - b_0\right)^2 + \sum k_\theta \left(\theta - \theta_0\right)^2 + \sum k_\chi \left(1 + \cos\left(n\left(\chi - \delta\right)\right)\right)$$
$$+ \sum k_\psi \left(\psi - \psi_0\right)^2 + \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{i<j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + E_{\text{constrain}} + E_{\text{other}} \quad (5.2)$$

where k_b is the bond force constant and $b - b_0$ is the distance from equilibrium that the atom has moved, k_θ is the angle force constant and $\theta - \theta_0$ is the angle from equilibrium between 3 bonded atoms, k_χ is the dihedral force constant, n is the multiplicity of the function, χ is the dihedral angle and δ is the phase shift, k_ψ is the force constant and $\psi - \psi_0$ is the out of plane angle, q_i and q_j are the fractional charges on atoms *i* and *j*, r_{ij} is the distance between the two particles, and ε_0 is the permittivity of empty space, A_{ij} and B_{ij} are terms where the Lennard-Jones potential crosses the x-axis i.e. it is zero and r_{ij} is the distance between the two particles.

Non-bonded interactions are cut off at 1.2nm and the Lennard–Jones potential is smoothly shifted to zero between 1nm and the cutoff distance. The pair list is updated every step using a 1.4nm cutoff. Particle mesh Ewald (PME) is applied for long range electrostatics. The temperature is set to be 295K and a time step of 1fs is used. Periodic boundary conditions are applied at the entrance and outlet of the pore. All simulations are performed on the supercomputer cluster POPEL at Pittsburgh Supercomputing Center (PSC)[131]. A 2ns simulation on 8 processors takes around 24 hours. The number of atoms in the system varies with the size of the nanopore. A typical simulated system has 14814 atoms, consisting of 8000 for nanopore, 6383 for water, 258 for DNA and 200 for ions. The timescale allowed by molecular dynamics

simulations is currently limited to few nanoseconds[58]. To accelerate the translocation events that normally take milliseconds, the MD simulations are performed at a higher applied voltage than 100–200 mV, applied typically in experiments.

5.3 Results

5.3.1 Benchmark Problem

The full atomistic Molecular dynamics simulations usually do not need a benchmark since even the slightest information of the system developed is preserved. But we wanted to first reproduce the results obtained by other research groups experimentally. This was to observe the dip in ionic current when the DNA enters the pore and thus blocks the ionic flux. In our system the DNA is already within the pore and the length of the DNA is equivalent to the length of the pore. The periodic boundary conditions are applied at the inlet and outlet of the pore, thus the process of dip in ionic current will not be observed for our system results mentioned latter in the chapter.

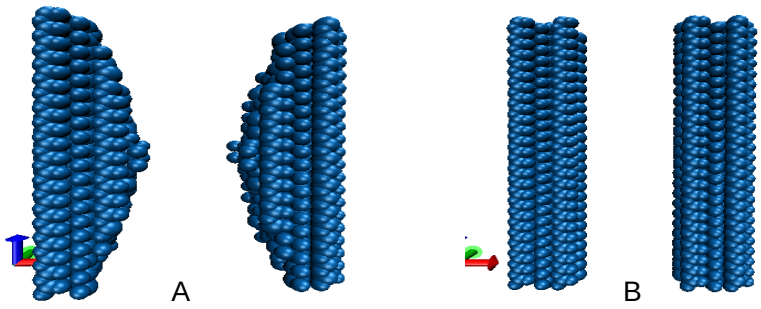


Figure 5.4. Comparison of the different nanopore configurations, (A) allows us to observe the dip in ionic current while (B) allows us to run simulations at a higher E-field.

The cross sectional view the two systems under discussion are shown in Fig. 5.4.A and 5.4.B. To validate the system we tried to obtain the results previously achieved with the system in Fig. 5.4.A.

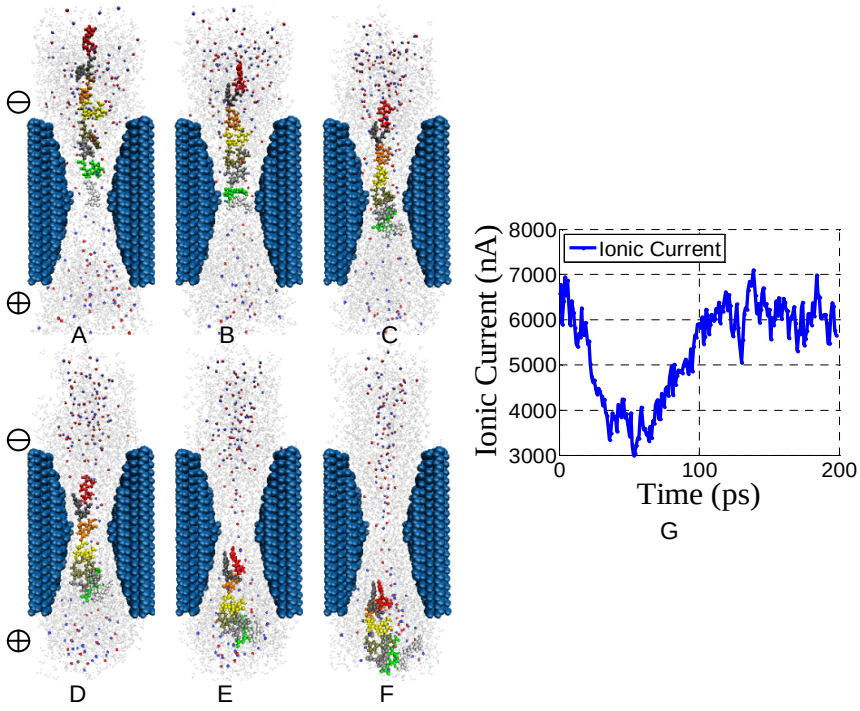


Figure 5.5. The DNA translocation through the nanopore shown through the sequence of images in (A-F), the dip in the ionic current measured as the DNA enters the nanopore and then when it exits the current raises back to the baseline level (G).

Fig. 5.5 (A-F) shows the sequence of images as the DNA translocates through the nanopore and the corresponding dip in the ionic current that is being measured as seen in Fig. 5.5.G. We can clearly see the dip in the ionic current due to the blocking effect of the DNA as it enters the nanopore. The ionic current is measured using the following formula,[129]

$$I\left(t+\frac{\Delta t}{2}\right)=\frac{1}{\Delta t\cdot l_z}\sum_{i}^{N}q_i\left(z_i\left(t+\Delta t\right)-z_i\left(\Delta t\right)\right) \qquad (5.3)$$

where z_i and q_i are the z-coordinate and charge of the ion i respectively, l_z is the length of the nanopores in the z-direction and Δt is the simulation time. The rest of the results are only considering the system shown in Fig. 5.4.B and its analysis to understand the interaction mechanism of DNA with the nanopore.

5.3.2 Nanopore Size-Dependent DNA-Nanopore Interaction

In the absence of surface friction, nanochannels would be a free solution environment in which DNA molecules would move with a length-independent mobility. However, electrophoretic mobility of DNA molecules in slit-like nanochannels has been observed to be length dependent[113]. This clearly indicates that molecular interactions with the confining walls are significant, and the notion of free solution electrophoresis breaks down. To understand the size-dependent translocation, DNA interaction with pores of various sizes is studied. A short single strand DNA (ss-DNA) of 8 bases (AATTGTGA) is driven by electrophoresis through nanopore with diameters ranging from 1.5nm to 4nm, as shown in Fig. 5.6 (A: Adenosine, T: Thymine, G: Guanine, C: Cytosine). The van der Waals (vdW) force between the DNA and nanopore, the translocation velocity of DNA, and the ionic current as function of pore size are plotted in Fig. 5.7 (A), (B), (C), respectively. The water molecules and ions are not shown in the figure for clarity.

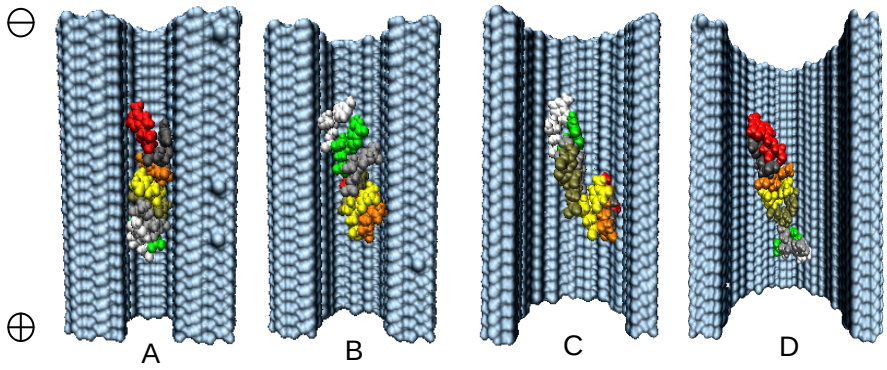


Figure 5.6. DNA translocation in pores of sizes: (A) 1.5nm; (B) 2nm; (C) 3nm; (D) 4nm.

From Fig. 5.7.A, DNA-nanopore interaction force decreases as the pore diameter increases, due to the reduced confinement effect. Accordingly, DNA translocation velocity increases as the pore size increases. However, both the translocation velocity and van der Waals force do not change linearly with size change, but satisfy an exponential curve. There is a gradual increase in translocation velocity as nanopore diameter increases at the beginning. It then reaches a constant limit indicating that an increase in the pore diameter beyond 3nm will not

affect the ss-DNA translocation process anymore. Establishing such nanopore size – DNA translocation velocity relationship can help design nanopore size for optimized signal yield. The ionic current increases with the increase of pore diameter, satisfying a parabolic curve.

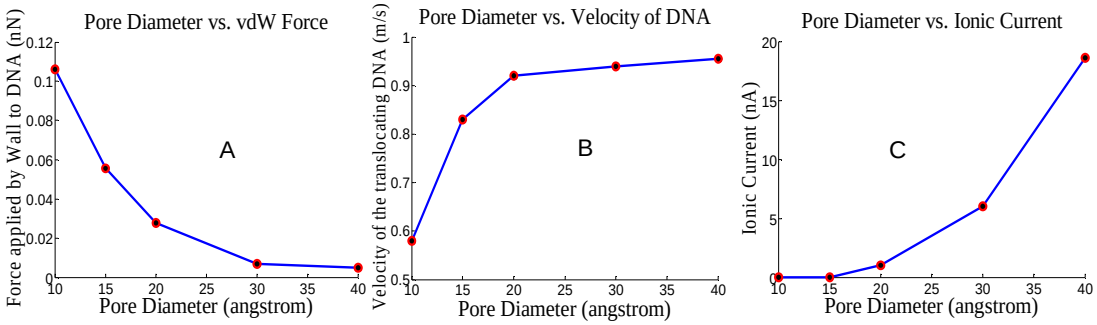


Figure 5.7. DNA translocation in pores of different sizes (A) van der Waals force; (B) Velocity; (C) Ionic current.

It should be noted that periodic boundary conditions are applied at the inlet and outlet of the pore, thus there is no depletion of ions as is usually observed in nanopore experiments when higher voltage is applied. Although experimental data suggests that the translocation time is different for long DNA strands with different sequences[58,50], we did not observe significant difference in translocation dynamics for homopolymers poly-dA or poly-dT through the nanopores and believe that it might be due to the small strand (8 base) ssDNA used in the simulation. Furthermore, it is expected that there is an optimal pore diameter for a particular voltage bias applied, which would provide sufficient confinement to enhance the molecule detection yet large enough for DNA to pass through. Such optimal designs will be explored in the future work.

5.3.3 Voltage-Dependent DNA-Nanopore Interaction

Besides nanopore diameter, the applied voltage also largely influences the DNA translocation speed. Considering a balance between confinement and translocation, a nanopore with a diameter of 2nm is chosen and is applied with various voltages.

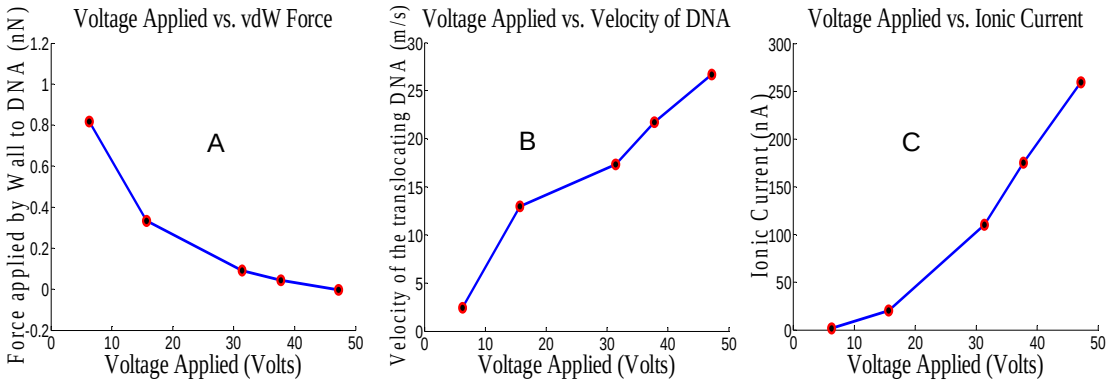


Figure 5.8. DNA translocation under different applied voltages (A) van der Waals force; (B) Velocity; (C) Ionic current.

From Fig. 5.8, a nearly linear trend can be observed between applied voltage and translocation velocity (Fig. 5.8.B), while a parabolic trend is observed between applied voltage and vdW force or ionic current (Fig. 5.8.A and Fig. 5.8.C). Ideally, ionic current should increase linearly with the increase of applied voltage for free ions passing through a nanopore. The nonlinear relationship observed in Fig. 5.8.C is due to the blockage effect from the existence of DNA. It should be noted that a high voltage is applied to observe the translocation in few nanoseconds. To study the translocation at lower voltages a coarse grained DNA model is needed, which will be included in our future study.

5.3.4 Nanopores with Surface Functionalizations

Coating nanopores with DNA or other organic molecules like silanes can make nanopores more biologically friendly and provide control over surface charges, hydrophobicity, and chemical functionality. It is important to characterize the surface property, size, and orientation of the modified surface. The coatings used in our case are 8-base long single strands of DNA. The tethered DNAs carry a negative backbone, thus these re-orient and are stretched under the applied bias voltage. Characterization of the coating polymer re-orientation is essential to predict the effective pore diameter under a particular applied voltage, which influences DNA translocation process. The spacing between the attached ss-DNA molecules on nanopore surface is determined by their radius of gyration. To avoid tethered DNA from peeling off,

nucleotide of DNA strand closest to the wall are fixed on the nanopore surface. The charge on each tethered DNA is assumed "–*e*", due to the phosphate backbone.

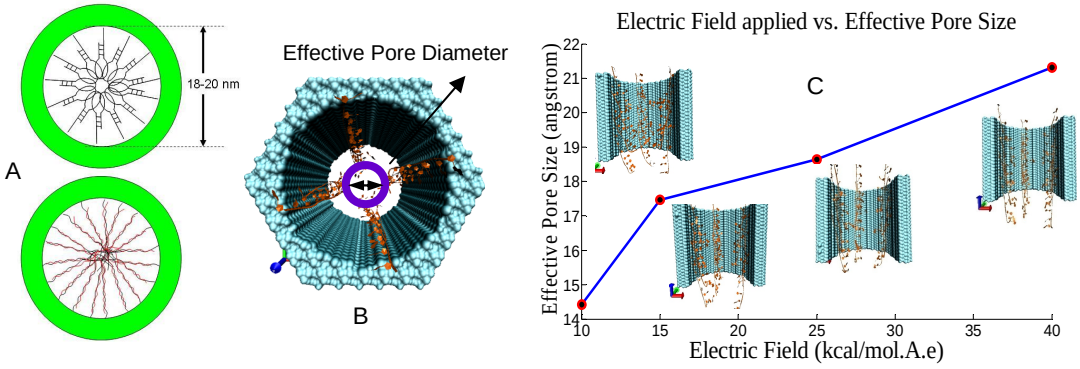


Figure 5.9. ss-DNAs coated on nanopore surface. (A) Radii of gyration for same length ss-DNA in hairpin and linear conformation, coated inside a pore; (B) Effective pore diameter; (C) DNA re-orient under applied Electric Field of various strengths and effective pore diameter as a function of Electric Field strength.

Under an applied voltage bias, translocation of ions towards the opposite electrodes is observed. Since the coated ss-DNA carries a –e charge, these curled tethered ss-DNA begin to get straightened in the direction of the oppositely charged electrode. The straightened DNA increases the effective pore diameter, which is measured as the empty space in the nanopore not occupied by the coated ss-DNA. After the system reaches equilibrium, the distances between opposing ss-DNA molecules are calculated and averaged to get the effective pore diameter. There are many factors that contribute to the effective pore diameter such as the flow of ions, the ionic loop around the DNA, voltage bias applied, etc. The effective pore diameters are plotted as a function of the applied voltage bias in Fig. 5.9. The effective pore diameter increases with the increase of applied electric field strength. The kinetics of DNA translocation in a polymer coated pore differs from that in a bare nanopore. In a bare nanopore, ions inside the nanopore form a double layer which prevents DNA translocation at a faster speed within the pore. With a tethered DNA coating on the surface, the ion layers disappear and different translocation speeds are expected.

5.4 Conclusion

The translocation speed of DNA in a nanopore depends on physical parameters such as nanopore diameter, electrophoretic bias, and surface coatings. The orientation of the tethered DNA in the nanopore changes the effective nanopore diameter. The effective nanopore diameter is found to be controllable in two ways: the bare nanopore size during fabrication and the strength of the applied electric field. This model-based system can be used to optimize parameters in the design of nanopore systems for DNA/gene sequencing. The major drawback of full atomistic molecular simulation is the limitation of timescale of a few nanoseconds. In the next chapter, a coarse grained DNA model is used to simulate low-voltage DNA translocation with longer time-span.

CHAPTER 6

COARSE GRAINED MODELING

6.1 Introduction

Lots of biological phenomena occur over time scales that are well beyond the current capabilities of the atomic-level simulation[132-136]. This challenge lead to the development of Coarse Grained (CG) models to overcome the time and space scale limitation. Coarse Graining is an approach applied in Molecular Dynamics (MD) to bridge the gap between the atomistic and mesoscale levels. An all-atom system (AA) in which each atom is explicitly represented has Angstrom-level detail and femto-second resolution which allows direct investigation of molecular structure and dynamics such as protein fluctuations. But the major disadvantages are the timescale limitations of a few nanoseconds (ns) and the lack of capability to study the group effect. A mesoscale model on the other hand allows us to observe the average density, charge or other characteristics of a material/structure like observing the self-assembly in biomolecular systems[137]. The schematic in Fig. 6.1 gives a general overview of the temporal and spatial scales that are accessible by the current simulation techniques[133].

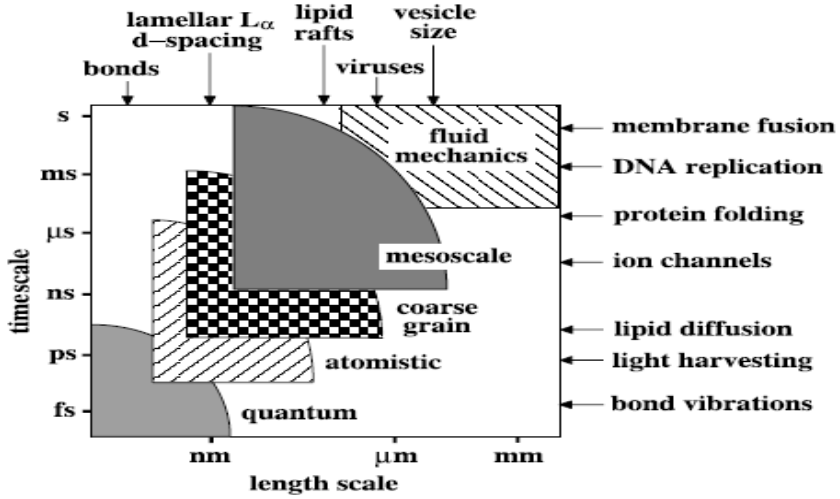


Figure 6.1. The temporal and spatial scales that are accessible by current simulation techniques are shown in the schematic, Nielsen *et. al.*, 2004.

Thus, coarse graining is an approach in which atoms are grouped together into new CG sites, which would reduce the computational complexity with fewer atoms and allow longer simulation time. The major challenge of a CG model is that it must be able to reserve the underlying AA system properties.

6.1.1 Background

The formula that describes CG system in terms of statistical mechanics is given by[138]

$$\exp\left(-F/k_B\mathrm{T}\right)=C\int dx\left(\exp\left[-V\left(x\right)/k_B\mathrm{T}\right]\right) \tag{6.1}$$

$$\exp\left(-F/k_B\mathrm{T}\right);\ C\int dx_{CG}\left(\exp\left[-V_{CG}\left(x_{CG}\right)/k_B\mathrm{T}\right]\right) \tag{6.2}$$

Equation (6.1) is for the all-atom simulation, where *F* is the free energy of the system, $V\left(x\right)$ is the system potential energy as a function of *x* of all atoms of the system, *T* is the thermodynamic temperature, k_B is the Boltzman constant and *C* is the constant of normalization. Free energy for the CG system shown in Equation (6.2) is represented by x_{CG} and V_{CG} (effective potential). This

equation forms the basis for the various distribution functions, equilibrium averages and properties, etc[138].

6.1.2 Approaches to obtain Coarse Grained Model

The first step in obtaining a CG model is to convert the all-atom representation into a model with less interaction sites. This is usually done either by residue based or the shape based approach. In a residue based method a single bead is placed at the center of mass of a group of atoms in the atomistic model. The CG site formed from the group of atoms (I_i) in the atomistic scale is denoted as 'R_{Ii}' which is obtained using a linear mapping function M_{R_I} [139],

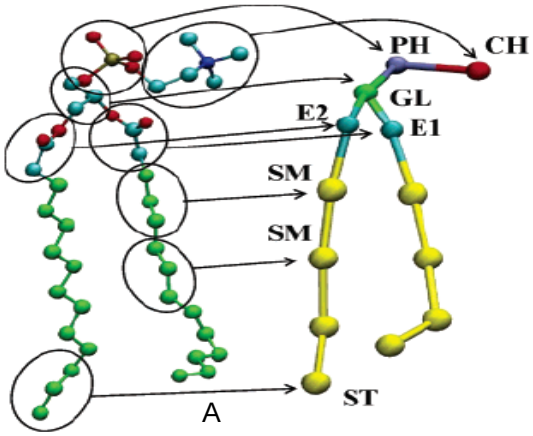
$$M_{R_I} = \sum_{i=1}^{n} c_{Ii} r_i \qquad (6.3)$$

where c_{Ii} *is* the constant for the group of atomistic scale atoms that are selected and $c_{Ii} \neq 0$, while r_i is the coordinates of the group of selected atomistic scale atoms(in most cases CG-site is at the center of mass of the selected group of atoms). The major advantage of this method is that it enables reconstruction of all atom scale from the coarse grained model obtained. This helps when we are trying to test a new system since we can compare back and forth the behavior of the system and understand the dynamics better at least for a prodigy system and then expand it to a large system. This method has been successfully used in study of lipids and proteins simulations[140-142].

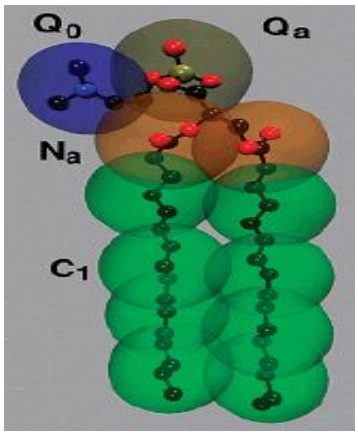
In case of the shape based system, a neural network learning algorithm is used to determine the placement of CG beads. Masses are correlated to the clusters of atoms which the CG beads are representing. This method doesn't enable exact reconstruction of the all atom system but it helps in understanding targeted properties of structures of interest. This can be used further to get coarse grain model using the electron density maps incase the all-atom model is unavailable. This method has been successfully used in study of stability and dynamics of viruses[143, 144] and rotating bacterium flagellum [145, 146].

6.1.3 Development of Coarse-Grained Potentials

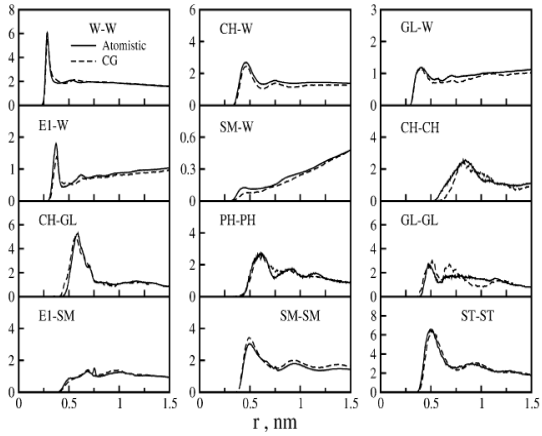
Popular approaches for obtaining coarse-grained model potentials are described in the following section. One of the oldest known method is the inverse Monte Carlo technique, the models of which reveal the essential physics of a given class of system and provides qualitative information about the system, but fail, to give quantitative accurate predictions[147]



A



B



C

		P	N				C	Q			
group	subtype		0	d	a	da		0	d	a	da
P		I	IV	III	III	II	V	I	I	I	I
N	0	IV	III	III	III	III	III	III	III	III	III
	d	III	III	II	II	II	IV	III	III	II	II
	a	III	III	II	II	II	IV	III	II	III	II
	da	II	III	II	II	I	V	III	II	II	I
C		V	III	IV	IV	V	III	V	V	V	V
Q	0	I	III	III	III	III	V	III	III	III	II
	d	I	III	III	II	II	V	III	III	II	I
	a	I	III	II	III	II	V	III	II	III	I
	da	I	III	II	II	I	V	II	I	I	I

D

Figure 6.2. The DMPC coarse grained model used for multi-scale coarse grain modeling, Izvekov *et. al.*, 2005 (a) and the Martini force-field coarse grain model, Marrink *et. al.*, 2007 (b) The comparison of the force values for the atomistic and CG force data at different intersite potentials, Izvekov *et. al.*, 2005 (c) The interaction potentials used in case of Martini force-field method, Marrink *et. al.*, 2007 (d).

The “multiscale” approach is a bottom-up method which incorporates force data from atomistic MD simulations for the CG system construction. The multi-scale coarse grained method

is established by Izvekov *et. al.* [148, 149], in which the inter-particle force field is accomplished by the application of force matching procedure to the force data obtained from the fully atomistic scale simulation of the biomolecular system of interest. The canonical (NVT) equilibrium coordinate distribution function for an atomistic model is given by, [150]

$$p_r(r^n) = \frac{1}{z_n} e^{-U(r^n)/k_B T} \tag{6.4}$$

where k_b is the Boltzman's constant, T is the temperature, $z_n = z(N,V,T)$ is the canonical configuration integral, $r^n = [r_1,\ldots,r_n]$ is the Cartesian coordinates for the 'n' atoms and $U(r^n)$ is the potential energy function of these n-particles. For the canonical equilibrium coordinate function for a coarse grained model obtained from the atomistic model denoted above is,

$$P_R(R^n) = \frac{1}{Z_n} e^{-U(R^n)/k_B T} \tag{6.5}$$

where $Z_n = Z(N,V,T)$ is the canonical configuration integral, $R^n = [R_1,\ldots,R_n]$ is the Cartesian coordinates for the 'n' atoms of CG and $U(R^n)$ is the potential energy function of these n-particles. Each CG site R_n is defined using mapping Equation (6.3). The detailed notes of this can be obtained from Noid *et. al.* 2008[139]

This multiscale approach has been successfully implemented for accurate CG models of carbohydrates[151], simple and ionic fluids[152], bilayers simple and mixed[153], and peptides [138 154]. This method also takes into account implicit solvent model which matches with the explicit solvent molecules in the atomistic model and is in accordance with Equation (6.1)[138]. Fig. 6.2.A shows the coarse grained model obtained by the multi-scale coarse grain approach. The image on the left in Fig. 6.2.A indicates the MD model while the right is its CG model with the different sites named[149]. Fig. 6.2.C shows the force matching between the inter-sites for comparison between the atomistic and the CG model obtained [149]. The algorithm details for these methods are provided in references [155, 156].

The “inversion” coarse graining approach which is well established by Marrink *et. al.* coined as MARTINI force field[157] makes use of the experimental, thermodynamic and/or average structural properties of the system all-atomistic system to obtain the CG system with similar properties. An example of this can be the model to reproduce the thermodynamic data between organic and aqueous phases in comparison with the experimentally available data. The free energy of partition between an organic (oil) and aqueous (water) phases ($\Delta G^{oil/aq}$) is given as[157],

$$\Delta G^{oil/aq} = k_b T \ln(\rho_{oil} / \rho_{aq}) \qquad (6.6)$$

where ρ_{oil} is the equilibrium density of organic (oil) phase and ρ_{aq} is the equilibrium density of aqueous phase of CG particles respectively. The ρ values are obtained by long full-atomistic scale simulations of the two-phase system with small amounts (0.01M) of target substance dissolved. The major advantage of this approach is that instead of concentrating on the structural details, the target is broader range without the need for re-parameterization of the model each time. The system obtained is simple, computationally fast and flexible enough to study the desired attribute of the system[138]. The MARTINI force-field is successfully used to study different types of bilayers elasticity[134], rupture tension, lipid lateral diffusion rates[158], liquid densities[159], lipid conformations[134]. Fig. 6.2.B shows the coarse grained model obtained by the Martini force-field which has the CG atom site placed at the center of mass of each subset of four atoms[157]. The Fig. 6.2.D, gives the general martini force-field potentials were the interactions potentials are classified as attractive (I), semi-attractive (II), neutral (III), semi-repulsive (IV), repulsive (V) and the groups Polar (P), Non-polar (N), a-polar (C) and charged (Q) with sub-types ‘0’ for no hydrogen bonding capabilities present, ‘d’ for groups acting as hydrogen bond donor, ‘a’ for groups acting as hydrogen bond acceptor, and ‘da’ for groups with both donor and acceptor options[157].

Basically the method chosen depends on the system under study or the goal of the study. If the goal is to preserve the general property of the entity of interest, a multi-scale approach is preferred. However, if the aim is to study a particular attribute, there is no need to characterize

the general attributes, and instead, the attribute under interest can be characterized by comparing with experimental results.

6.1.4 Challenges and Limitations of Coarse Grained (CG) System

Coarse graining is a promising tool for the future of Molecular Dynamics as it allows understanding the aimed attributes of a system by allowing longer simulation time in comparison with all-atom models. It is still in its developing stages and there are many concerns and challenges, few of which are mentioned here. Coarse graining is still not predictive, the point is that the systems are usually built with a goal with what can happen or an intuitive thought of what should happen, which may cause the potentials used to be biased without the knowledge. The other is the degree of transferability of these coarse grained models to other systems and the modifications needed in the thermodynamic conditions. Not all characteristics of the CG model can be transferred but currently there is no literature which specifies the properties that can be transferred or not. Further, with the development in computational efficiency with better MD algorithms and high performance CPUs, the time required for full-atomistic scale simulations is being reduced and there is a worry that the work done in the CG system might become obsolete if the all-atom system can be used to do the same. But the bright side is that the coarse grained length scale can surely be extended[138].

The CG models are usually built to capture a particular attribute of a system or a particular mechanism. In our case, we need a CG system of DNA to capture the hybridization property and a proper bare pore-DNA interaction potential. Our system is developed following the residue based approach and combined with MARTINI force-field for the bare nanopore-DNA interaction potential.

6.2 Coarse Grained Method

The purpose of CG method is to obtain a system that is simple, yet sufficiently detailed to allow study of real-time dynamics. The most important component of our system is to characterize the interaction potential between DNA strands. Two of the successfully constructed CG DNA models available in literature are of Knotts *et. al.* (three-site DNA model) and Schatz *et.*

al. (two-site DNA model). Description of these models along with a brief outline of our DNA model is explained in the sections below.

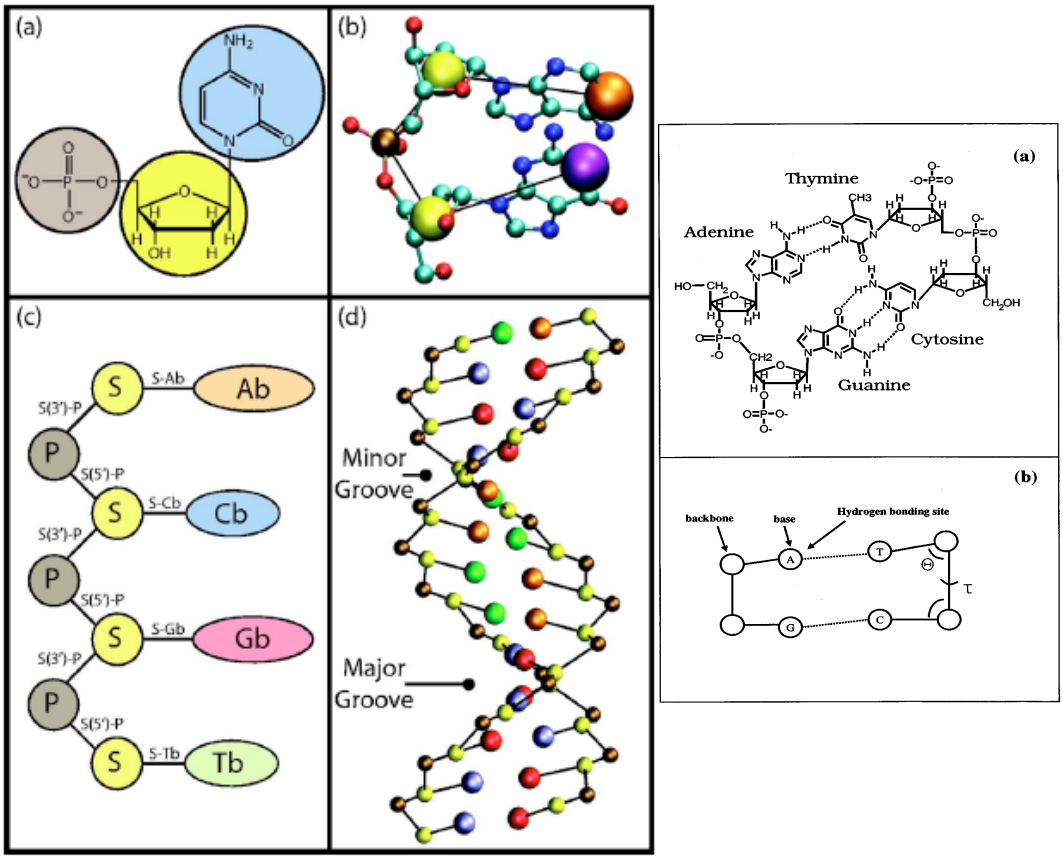


Figure 6.3. The 3-site DNA model by Knotts *et. al.*, 2007 (a) and the 2-site DNA model by Schatz *et. al.*, 2000 (b) is shown for comparison purpose.

6.2.1 Three-Site DNA Model

Knotts *et. al.*[160] developed a three site model of DNA. They presented a mesoscale molecular model of DNA which describes aspects of melting, hybridization, salt effects, the major and minor grooving of DNA, and the mechanical properties of DNA. The groups of atoms replaced by each site of the coarse grain model for a cytosine base are shown in Fig. 6.3.a. the CG site is the center of mass of the sugar and phosphate respectively while for the bases, the CG site was N1 for purines (adenine and guanine) while N3 for the pyrimidines (thymine and

cytosine). The DNA system was developed such that for an atom (x, y, z) the corresponding atom on the other side is at (x, -y, -z). Screw symmetry was used to place successive residues of the DNA with 10 residues per turn and an axial rise of $3.38a^{o}$; thus an atom at (r, φ, z) places the next at ($r, \varphi+36^{o}, z+3.38a^{o}$), thus the dna is aligned and centered with the z-axis[160].

6.2.2 Two-Site DNA Model

Schatz *et. al.*[161] (see Fig. 6.3.B) had developed the two-site model of DNA based on a simplified backbone-base structure where each nucleotide is symbolized by two sites. One site represents the backbone i.e. placed at the combined center of mass of sugar and phosphate while the other represents the center of mass of the base. Their model was able to capture interactions such as hydrogen bonding formation and reformation between the inter strands of DNA, bending and torsions, the melting transition for DNA decamers containing only A/T, or only C/G, base pairs and assessed the importance of the helical structure. The distance between two atoms in the two-site DNA was 1.7 A^{o} and screw symmetry was used to place successive residues of the DNA with 10 residues per turn. The hydrogen bonding sites were at the center of mass of the base and were divided into acceptor and donor sites. The hydrogen bonds were formed only if the backbone-donor-acceptor angle is within a defined range[161].

In their system, the charges were not included and even the model parameters were chosen to give reasonable melting behavior only at salt concentration of 0.1M. This made their system irrelevant under different salt concentrations. Also, in order to mimic the correct melting behavior modifying the model parameters for each different salt concentrations considered will be impractical. The possible extension to this system was to incorporate charges to the atoms.

6.2.3 Coarse-Grained DNA Model with Charges

We followed the two-site model and incorporated the charges following the CG model of Marrink *et. al.*[134, 157-159]. The system is tested at 0.1M salt concentration for hybridization and the melting behavior, extending the work to different salt concentrations is not part of this thesis but an aspect of future work.

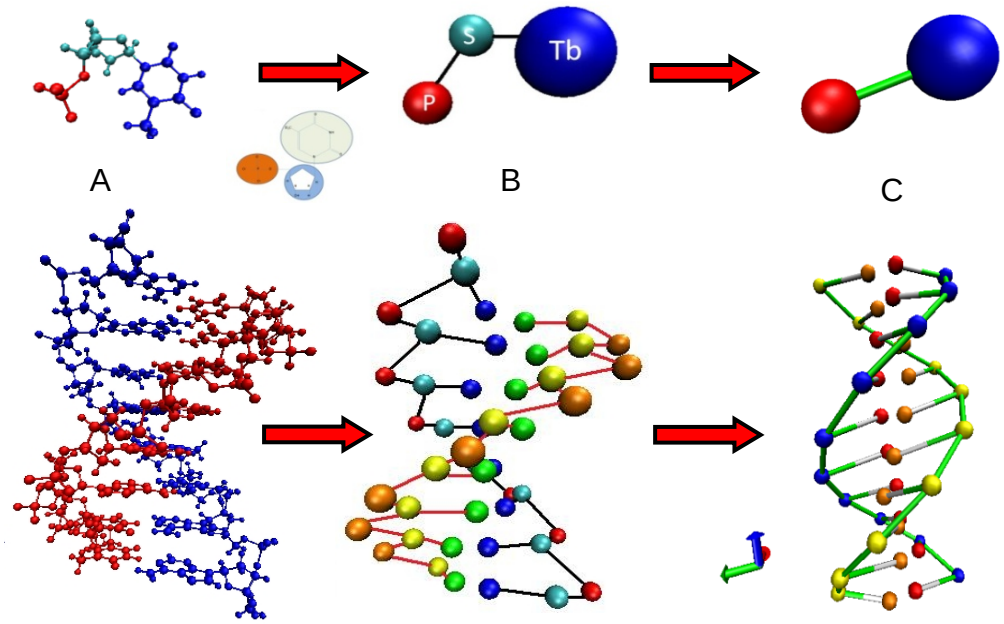


Figure 6.4. Coarse Grained DNA model (a) the full atomistic scale dsDNA (30 atoms form a nucleotide) (b) the 3-site dsDNA model (3 atoms form a nucleotide) developed by us following Knotts *et. al.* (c) the 2-site DNA model (2 atoms form a nucleotide) of Drukker *et. al.* developed by us. The first row shows a single nucleotide while the second column shows the entire DNA.

A system similar to atomistic scale with the silicon nitride pore, tethered DNAs, and translocating DNA is built. All the necessary input files (.gro, .itp, .top, .index, etc.) for the simulation are obtained using scripts in Matlab for the different systems that are built.

The non-bonded interactions between particles are described by Lennard-Jones potential $E_{nbpair} = \sum \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)$. Coulombic potential $E_{\text{coulombic}} = k \frac{(q_1 * q_2)}{r^2}$ is used to define charged particle interactions.

The total potential energy is composed of bonded energies and non-bonded pair interaction energies:

$$E_{total} = E_{bond} + E_{angle} + E_{diheral} + E_{improper} + E_{nbpair} + E_{\text{coulombic}} + E_{stack} + E_{bp} + E_{ex}$$

$$= \sum k(r_{ij} - r_0)^2 + \sum k_{\theta_{ijk}} (\theta_{ijk} - \theta_{eq})^2 + \sum \frac{1}{2} k_\phi \left[1 + \cos(n\phi - \delta)\right] + \sum k_\psi (\psi - \psi_0)^2$$
$$+ \sum \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + k_q \frac{(q_1 * q_2)}{r^2} + E_{stack} + E_{bp} + E_{ex} \qquad (6.7)$$

where the first four terms describe the bonded inter-site interactions including bond stretching, bending, and torsion, and the last two terms are the pairwise potential which describes the non-bonded (nbpair) interactions and the columbic potential that describes the electrostatic potential. The non-bonded interactions include E_{stack} accounting for the base stacking phenomena, E_{bp} (which also has a columbic term) to describe bonding between complementary base pairs, E_{ex} to describe an excluded volume interactions among A, T, C, and G bases. All the *k* subscript parameters are the constants for each term respectively. The E_{bp} term is critical for modeling the hybridization of an ss-DNA with a complimentary ss-DNA or hairpin loop and is parameterized using thermal denaturation experimental data at a fixed salt concentration ~0.1 M. All parameters used can be found in reference[161].

The bond distance between the atoms is 0.47nm and the angle between backbone sites is 150^o while the base and the phosphate are kept at 180 ^{o} [126]. The schematic of the different models of DNA developed by us is shown below in Fig. 6.4.

6.2.4 Pore-DNA Interaction Potential

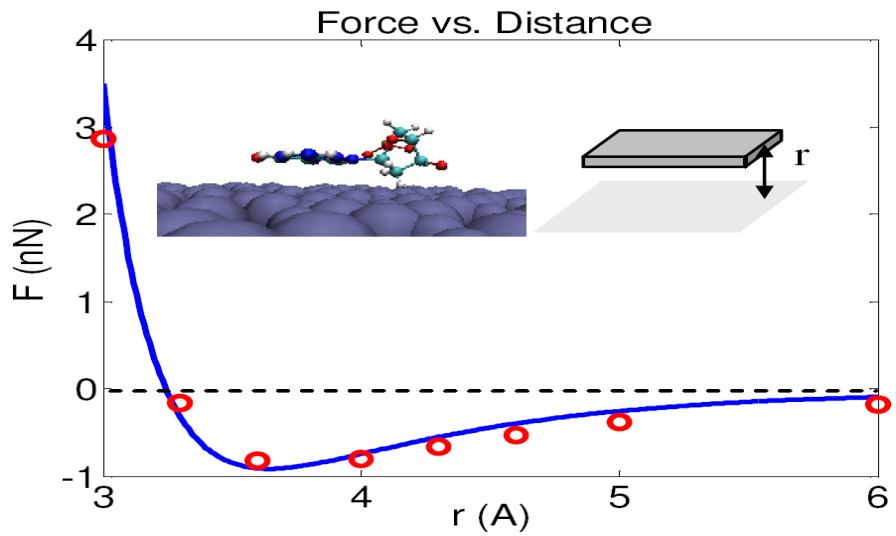


Figure 6.5. DNA-base-surface interaction force-distance functions from all-atomistic model and from coarse-grained model. Inset: a DNA base placed parallel to the surface

As the DNA translocates through a bio-functionalized nanopore its interaction with bare nanopore has to be defined and characterized. This interaction potential will not only be applied between the translocating DNA (t-DNA) and the nanopore but also between the coated DNAs (c-DNAs) and the nanopore. The interaction potential between the nanopore and the DNA is mapped from the atomistic model into the CG model. Interaction between a DNA base and bare nanopore surface is described by a Leonard-Jones (LJ) potential as,

$$V_{DN} = 4\varepsilon_{DN}\left(\left(\frac{\sigma_{DN}}{r}\right)^{12} - \left(\frac{\sigma_{DN}}{r}\right)^{6}\right) \tag{6.8}$$

where ε_{DN} is the depth of the potential well, σ_{DN} is the (finite) distance between the DNA base and the nanopore surface at which the potential is zero and *r* is the distance between the center of mass of the DNA base and fixed nanopore surface.

The vdW energy density function is given as,

$$w_{DN} = \frac{2}{N_0} V_{DN}(r) \tag{6.9}$$

where N_0 is the unit cell of Nanopore surface and $V_{DN}(r)$ is the LJ potential function. The total interaction energy (E_{DN}) between center of mass of DNA nucleotide (q_0) and total surface of silicon nitride nanopore (Ω) can be given as the integral over the surface,

$$E_{DN} = \int_{\Omega} w_{DN}\left[\left\|q_0 - x\right\|\right] d\Omega \tag{6.10}$$

where x is a site on the nanopore surface.

Thus, the total vdW interaction energy density is a function of the two LJ parameters (ε_{DN}, σ_{DN}) and can be controlled by modulating them to match with the interaction energy from the full-atomistic scale molecular dynamics simulation. The CHARMM force field is used for the DNA, nanopore (silicon nitride), and their interactions in the full atomistic scale simulation. [99] . The times step is 1fs with the temperature at 300k and molarity of 0.1M is used. The discrete force-distance data from the full-atomistic simulation and from the coarse-grained simulation are plotted in Fig. 4.5. This process leads to an effective interaction between the CG sites as is present in the underlying all-atom simulation.

If an explicit solvent (water) is used it increases the number of atoms dramatically and thus the computational demand increases. For this reason solvent effects are included implicitly into the system by using Langevin dynamics. The solvent bath is represented by Langevin equations of motion, stochastic frictional forces and the velocity Verlet algorithm [162]. Friction constant (ξ) is given as, [161]

$$\xi = \frac{4\pi}{m}\eta \cdot r_{eff} \tag{6.11}$$

where m is the mass of the coarse grained DNA particles, r_{eff} is the solute's effective hydrodynamic radius considered 0.5nm and η is the solvent viscosity given as,

$$\eta = \eta_{20} \cdot \exp(-A/B)$$

$$A = 1.37023(T-20) + 8.36\times10^{-4}(T-20)^2$$

$$B = 109 + T \qquad (6.12)$$

where T is the temperature, η_{20} = 0.93975*10^{-3} is the viscosity at 20^{o}C for water.

6.2.5 Simulation Details

The coarse-grained MD simulations are run in GROMACS 4.0 [163]. The Electric Field (E-field) applied is in the range of 5mV/nm – 500mV/nm. The temperature is set at 300k with molarity of the system being 0.1M and a time step of 0.01ps is used. Non-bonded interactions are cut off at 1.2nm and the Lennard–Jones potential is smoothly shifted to zero between 1nm and the cutoff distance. The pair list is updated every step using a 1.4nm cutoff distance. Particle mesh Ewald (PME) is applied for long range electrostatics [164]. For simplicity of the system homo-polymers are used, coatings in our system are poly-dG while the translocating DNA is poly-dC or poly-dT depending on the attractive or neutral case respectively. The tethered and the translocating DNA each consisted of 20 bases which helped us get a better comparative study. Periodic boundary conditions are applied at the entrance and outlet of the pore. Analysis and visualizations are done with Visual Molecular Dynamics (VMD). All simulations are performed on the supercomputer cluster Ranger at TACC which is a part of the Teragrid resources. A 200ns simulation on 2-processors takes around 24-hours. The number of atoms in the system varies with the size of the configuration considered. A typical simulated system has 4714 atoms with 2610 atoms for the silicon nitride pore, 1960 atoms of DNA coatings, 40 Atoms of DNA (t-DNA) and 144 ions.

6.3 Benchmark Problems

The system is first validated by performing three benchmark runs to ensure that CG model match the pure atomistic model.

6.3.1 Melting Behavior of DNA

The E_{bp} term which determines the interaction potential for hybridization between ssDNAs (homo-polymers in this case) is parameterized using thermal denaturation experimental data at a fixed salt concentration of 0.1M. A 20 base-pair dsDNA with poly (dGC) is placed in an implicit solvent bath with temperature of 300k. A simulation of 100ns is run to test if the strands separate at a temperature of 300k, this is done to confirm that the inter strand potential (E_{bp}) used will maintain the hybridized state of the dsDNA.

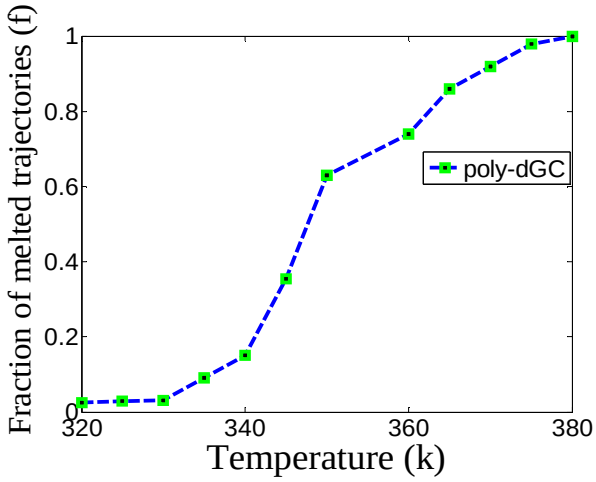


Figure 6.6. Fraction of melted trajectories for a 20bp dsDNA with gradual increase of temperature in the system.

It is observed that the dsDNA continued to be double stranded entity and there is no fraction of melted/separated DNA at 300k. The next step is to gradually increase the temperature by 5k after every 10ns and this procedure is continued until 400k, thus the total simulation time is of 200ns. A script in the tcl (scripting language) [165] is written to measure the distance between the opposite base atoms of the complimentary strands to measure the distance between them. The threshold is kept at 4A° (two-site hydrogen bonds) such that if the separation between the opposite base atoms of the complimentary strands is more than 4A° then it would be considered in the melted fraction and not if otherwise. The graph in Fig. 6.6 shows the melting fraction of

DNA with gradual increase in temperature. During the initial temperature increments there is no separation between the two ssDNAs forming the dsDNA and thus the fraction of melted/separated DNA bases are close to zero. But then the fraction of melted DNA bases increases as the temperature increases, reaching the value 1 close to 380k. A reassumed form gives us the value for E_{bp} and the parameters are determined to match with the experimental denaturation data [166].

6.3.2 Hybridization of DNA

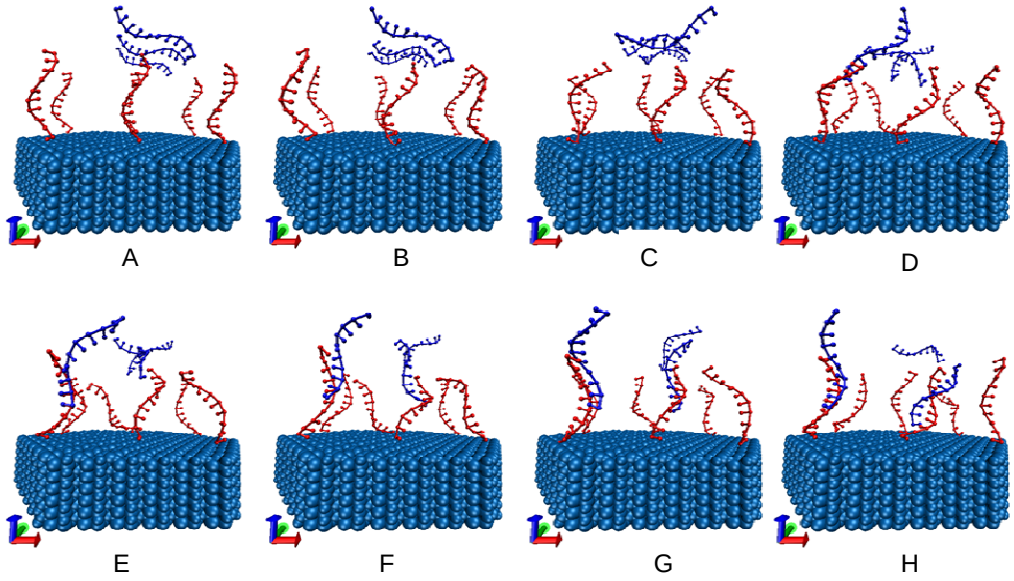


Figure 6.7. Brownian hybridization of free floating poly (dC) (blue) with a base tethered poly dG (red)

To further test if the inter-strand potential (E_{bp}), we test the hybridization of ss-DNAs. The system is set-up as shown in Fig. 6.7 with 3-ssDNAs poly(dG) having one of their ends attached on the surface while the other 3-ssDNAs poly(dC) are placed in the solvent bath above the substrate. The solvent is implicit with the temperature set at 300k and the salt concentration of 0.1M. The fixed base of poly(dG) is constrained on to the pore by applying a force of 10000kJ/mol. The system is first equilibrated for 10ns to ensure that the strands are in minimized energy formations. After that, a real run of 200ns is performed and the image sequence in Fig.

6.7 shows 2 out of the 3 poly(dC) ssDNAs have formed hybridized pairs with two of the poly(dG). The results support that the E_{bp} considered is satisfactory.

6.3.3 Comparison of MD system with the CG system

Since the inter-strand potentials of the DNA is established, we now need to compare if the coarse grained system developed satisfactorily represents the underlying full atomistic scale system. To assure that the coarse-grained model accurately describes the molecular motion, a benchmark case is performed by comparing the coarse-grained and all-atom molecular dynamics results. To test this we considered the same formation for both the systems as shown in Fig. 6.8. Each system consisted of two straightened strands of ssDNAs with 20bp each and a circular pore.

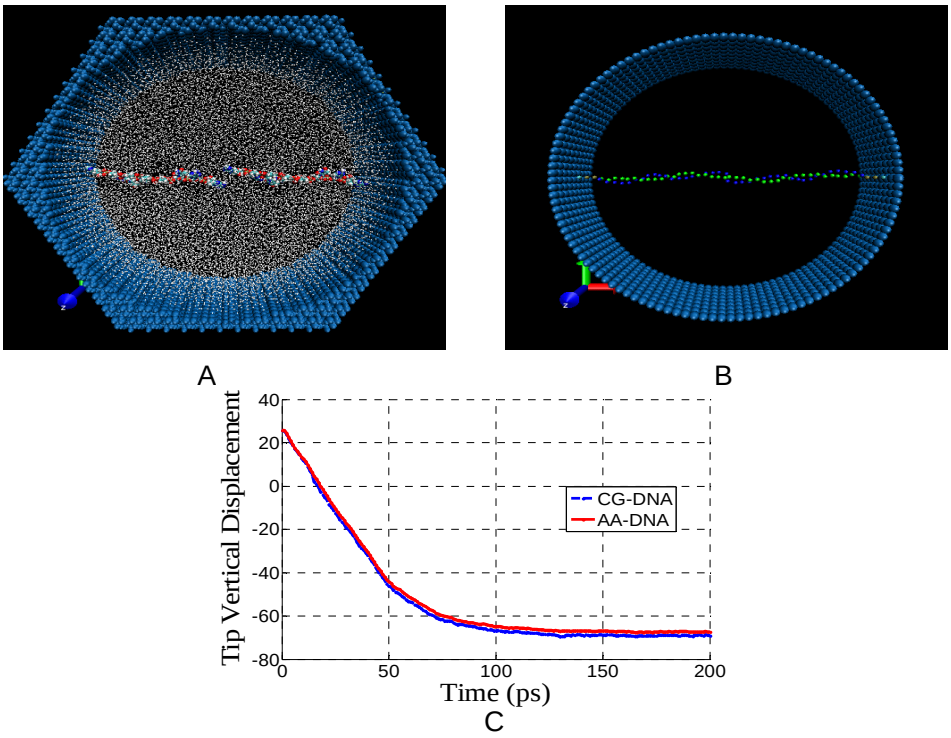


Figure 6.8. Comparison between the all-atom system (A) and its coarse grained model (B), the graph (C) indicates the similarity between the two systems since similar trends are observed in the vertical tip displacement measured for the two systems.

The CG system is solvated implicitly through the langevin dynamics while the fully atomistic scale system is explicitly solvated. A 200ps simulation is run with the temperature set at 300k and the planar E-field applied is 0.1mV/nm. As shown in Fig. 6.8, the DNA displacement time history under a given E-field is quantitatively compared between the coarse-grained and MD system. The result compares the displacement of the tethered ssDNA tip between the CG and the fully atomistic scale. We can clearly see that both the systems have similar trend and are almost identical with each other in terms of the bending of the DNA as indicated by the tips of the two systems.

6.4 Results for Bio-Functionalized Nanopores

Bio-functionalization is an essential way to impart selectivity to the nanopores. However, this process needs careful characterization to achieve the desired effective pore diameter and translocation kinetics. To understand the effect of bio-functionalization on nanopore, the following parameters are to be analyzed: the type of coating, density of the coating, applied bias voltage, and the effective pore diameter. The coating DNA strands have a charge of -e, thus they reorient themselves under the applied bias voltage. The re-orientation of the coating molecules leads to an Effective Pore Diameter (EPD) different from the original bare pore diameter (Ramachandran *et. al.*, 2009) [11]. A set of simulations are done to understand the behavior of these DNA coatings based on their type, circumferential density and longitudinal density.

6.4.1 Effect of Different Coatings

The selectivity desired to be imparted to the nanopore will be defined by the coatings used and its properties under the application of the Electric field (E-field) applied. Behavior of these coated structures is measured in terms of the Effective Pore Diameter (EPD) which is the diameter of the hollow region not occupied by the coating molecules in the center of the pore as shown in Fig 6.9.

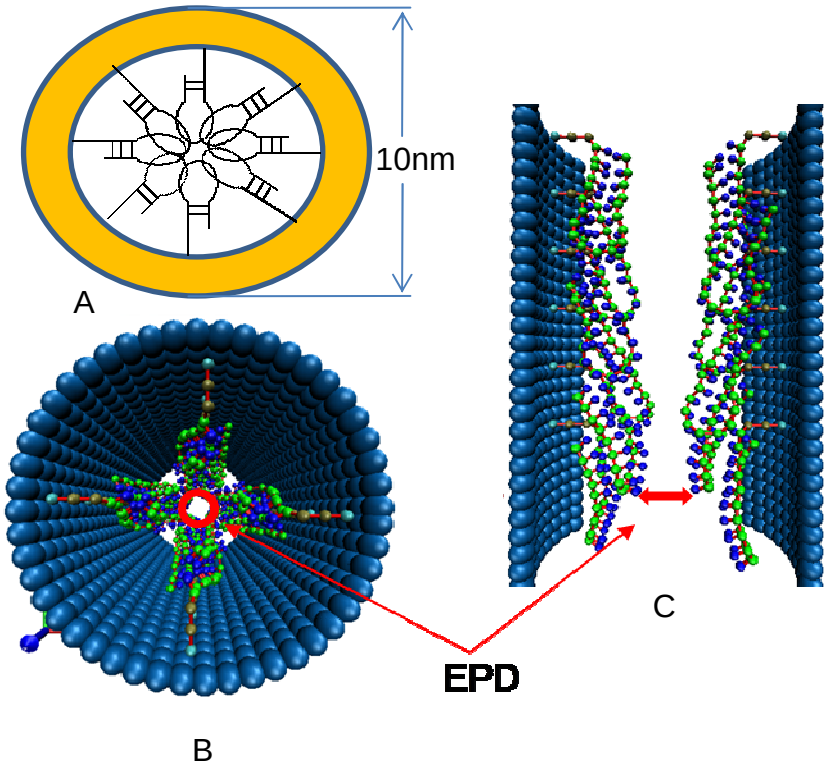


Figure 6.9. The effective pore diameter (EPD), schematic shown in (A), Top view of the 1nm EPD (B) and the cross-sectional view of 1nm EPD for the CG model developed (C).

Under an applied bias voltage, the coated DNAs with the –e charge re-orientate themselves along the E-field direction. These coated structures then begin to bend and this bending depends on their rigidity which is due to their shape, structure, intra-strand bonds, etc. The EPD (shown in Fig. 6.9) would influence the interaction of the translocating DNA (t-DNA) with the DNA coatings and the net translocation velocity. To understand the effect of type of coating, we compared the HPL coatings with ssDNA coatings for a pore with actual pore diameter of 10nm. The two coatings considered are the HPL and the ssDNA. The reason for choosing these two types is that when a perfectly-matched complimentary ss-DNA is within a close range, the HPL opens up to form an ssDNA to hybridize with the complimentary t-DNA. There is difference in the rigidity of these two structures with the former (HPL) being more rigid while the latter (ssDNA) being highly flexible. To characterize their effect on the EPDs obtained, we performed a

sequence of runs at different E-fields. The probe molecules are made to be uniformly coated on the pore surface. The results obtained in Fig. 6.10 indicate that the HPL (more rigid) coated nanopore has a smaller EPD compared to the ssDNA (less rigid) coated nanopore. This leads to the conclusion that the t-DNA translocating through an HPL coated nanopore will have more interactions when compared with the ssDNA coated nanopore. Thus, t-DNA translocates faster in ssDNA coated pores compared to HPL coated pores.

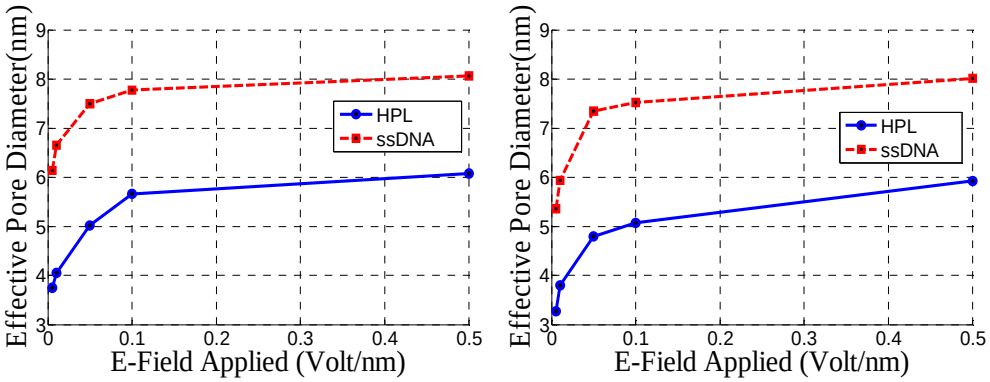
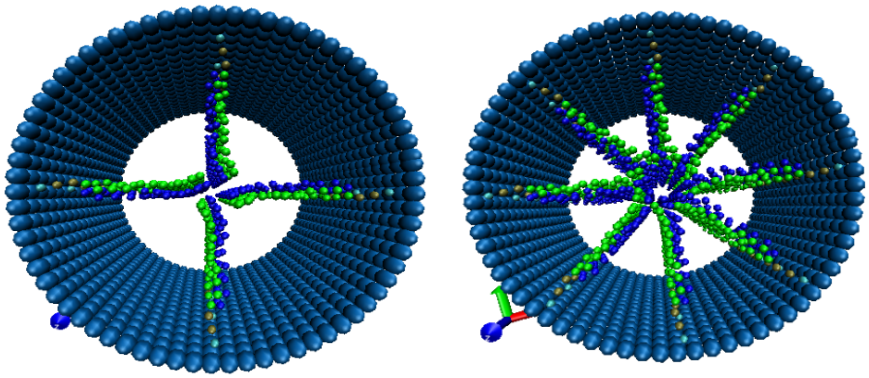


Figure 6.10. Effective pore diameter (EPD) for two different type of coatings, (a) shows the EPD obtained when 4-strands are used for coating while (b) shows the epd obtained with when 8-strands are used for coating. Both the results indicate that the epd obtained from the HPL coatings are much smaller compared to the EPD of the ssDNA for the nanopores with original bare pore diameter of 10nm.

6.4.2 Effect of Circumferential Density

The number of strands coated circumferentially defines the circumferential density of the coatings. The denser a system better is the interaction of the t-DNA with the coatings. To understand the circumferential density effect, we considered four strands forming a circumferential coating with that of eight strands. These studies are done to see the lateral motion of the coated DNAs and to understand if there is difference in the EPDs obtained for the two systems under consideration. The top views of the two systems are shown in Fig. 6.11. The densities of these two systems are one strand per 5.0 nm and one strand per 2.5nm for the 4-strands circumferentially coated and 8-strands circumferentially coated in the pores respectively.



A B

Figure 6.11. Two systems considered to understand the effect of circumferential density with density of one strand per 5.0 nm and one strand per 2.5nm for the 4-strands circumferentially coated (A) and 8-strands circumferentially coated (B) pores respectively.

The results shown in Fig. 6.12 indicate a lower EPD for the system with higher circumferential density. All the three graphs show a parabolic trend for the EPDs obtained as the E-field applied is gradually increased. The comparison of results of Fig. 6.12.C with other two indicates that the effect of circumferential density is more evidently seen in the coated nanopore with the least longitudinal density. The first two data points show a large difference in the EPDs obtained with the different circumferential density. The difference however is reasonable for Fig. 6.12.A and Fig. 6.12.B where consistent parabolic trend is observed.

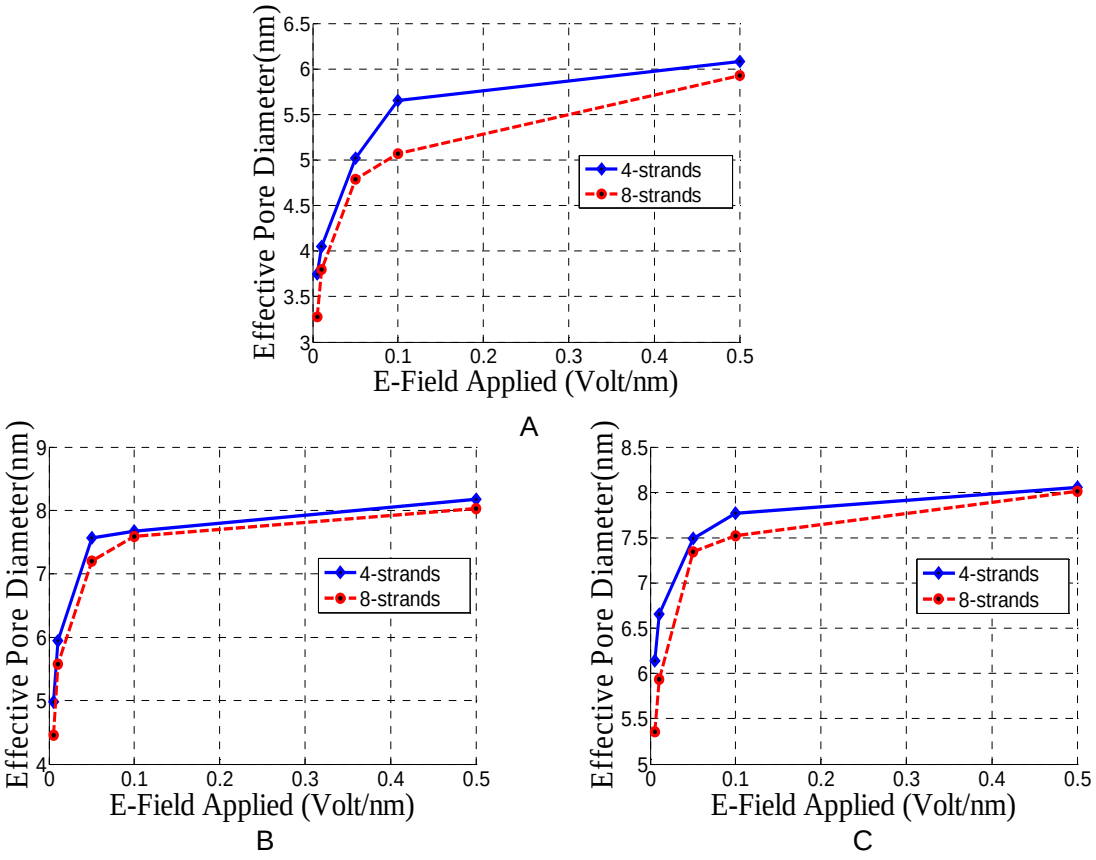


Figure 6.12. Result of circumferential density on the effective pore diameter (EPD), each of the graphs gives a comparative analysis of the EPD obtained with the four strands coated circumferentially to that with eight strands coated. Different coatings considered are (a) HPL, (b) ssDNA 1nm apart and (c) ssDNA 2nm apart.

In this case, the translocation velocity of the t-DNA will be higher in the less densely coated nanopore compared to the densely coated nanopore. Also, more circumferential density reduces the erroneous results that can be obtained since the t-DNA may translocate through the least resistance path with no coatings.

6.4.3 Effect of Longitudinal Density

Besides circumferential density, longitudinal density of coatings also influences. To understand this effect, we compared two ssDNA cases, in which the coatings are placed 1nm apart compared to 2nm apart strands as indicated by Fig. 6.13.A and Fig. 6.13.B respectively. The importance of longitudinal density is that the inter strand interactions between the coatings

increases with the increase in the longitudinal density. The lower the longitudinal density, the less is the inter strand interaction of the coatings and more the coatings stretch under the application of the E-field, thereby increasing the EPD. The cross sectional views of the two systems are shown in Fig. 6.13. The densities of these two systems are one strand per 8.3nm and one strand per 16.7nm for the ssDNA coated 1nm apart longitudinally and ssDNA coated 2nm apart longitudinally in the pores respectively.

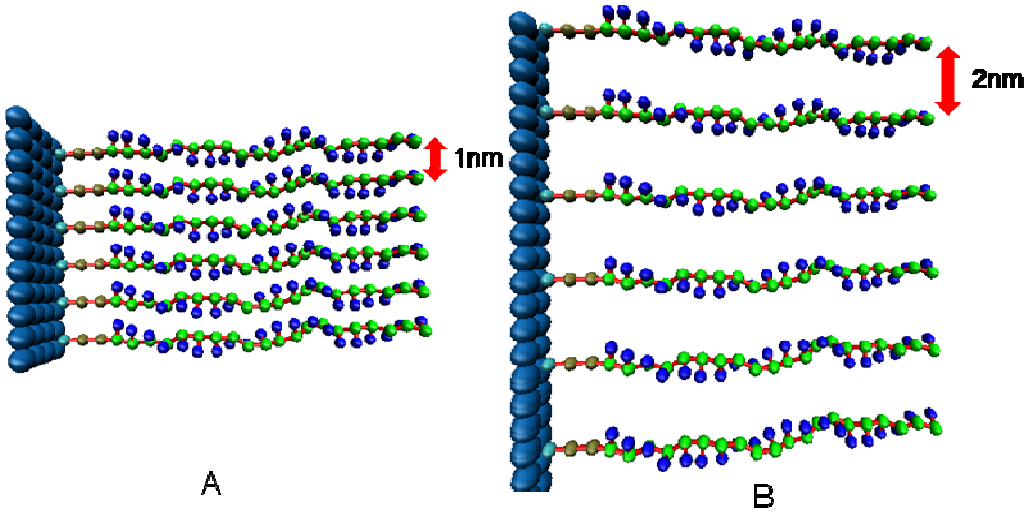


Figure 6.13. Variation in the longitudinal density of the systems considered, the strands in (A) are placed 1nm apart while in (B) the strands are placed 2nm apart.

The results shown in Fig. 6.14.A and Fig. 6.14.B indicate that at low E-fields there is significant difference in the EPD obtained d ue to the difference in the longitudinal densities of the two systems. As the applied bias voltage is increases the influence of the E-field is more apparent compared to the inter strand interactions which causes the coatings to stretch to a large extent, overlaying the EPD obtained with different densities values at very high E-fields.

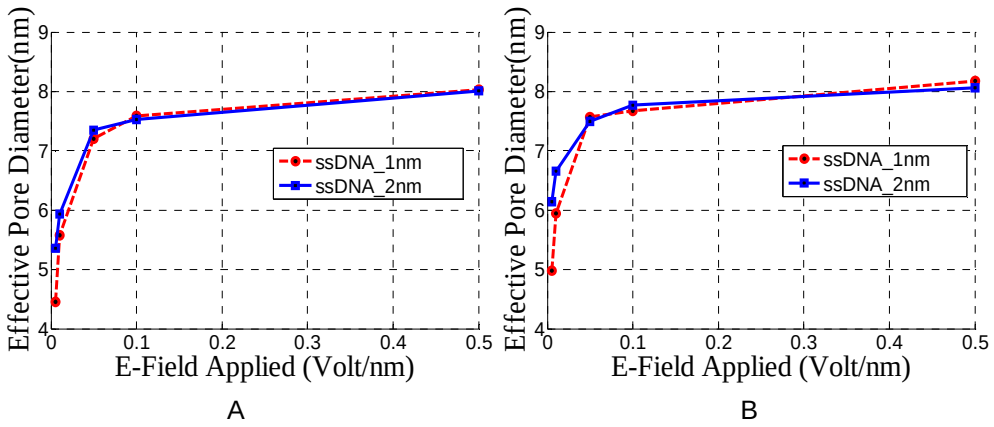


Figure 6.14. Effect of longitudinal density on the effective pore diameter (EPD), each of the graphs gives a comparative analysis of the epd obtained with the ssDNA strands coated 1nm apart longitudinally to that coated 2nm apart. Different coatings considered are (a) 4-strand circumferential coating and (b) 8-strand circumferential coating.

The velocity of the t-DNA will be larger in the less dense system compared to the denser system only for the low E-fields. At higher E-fields, there will be no significant difference in the velocity of the t-DNA in the two systems under consideration.

6.4.4 Effect of Interaction Potential

Kaisnowicz *et. al.* [43] had mentioned that a mere potential difference within the nanopores will not create the selectivity needed to get the gene sequencing done. There is much more complex procedures needed to attain it. The interaction potential of the coatings will surely have a significant influence on the velocity of the t-DNA.

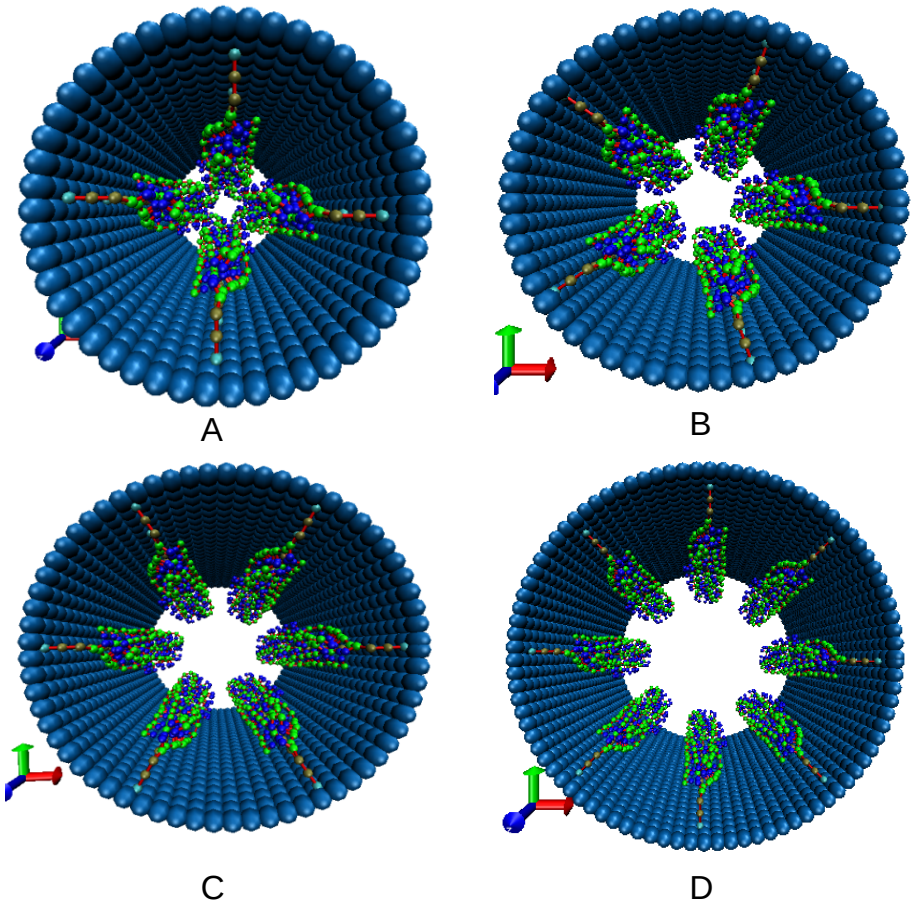


Figure 6.15. Different effective pore diameters considered, with EPD of 1nm (A), 2.8nm (B) 4.6nm (C) 81nm (D). These are used to test the potential effect on translocating DNA (t-DNA), with attractive and neutral potential.

The system for understanding the effect of t-DNA potential is developed by considering the final configuration of the coated DNAs in past runs and recoating them to obtain the respective diameters as shown in Fig. 6.15. Thus, the stretched coatings are recoated onto the nanopores to obtain an EPD of 1nm. Also, it is ensured that irrespective of the types of coating, the initial EPDs will be 1nm (Fig. 6.15.A) for the three different coating systems that are considered. The EPD is also gradually increased by keeping a constant density of the coatings. The different cases considered are shown in Fig. 6.15. For simplicity of the system, homo-polymers are considered. In case of the system to test effect of attractive potential, the coatings

are poly (dC) while the translcoating DNA is poly (dG), while for the intermediate potential system; the coatings are poly (dC) while the translcoating DNA is poly (dT).

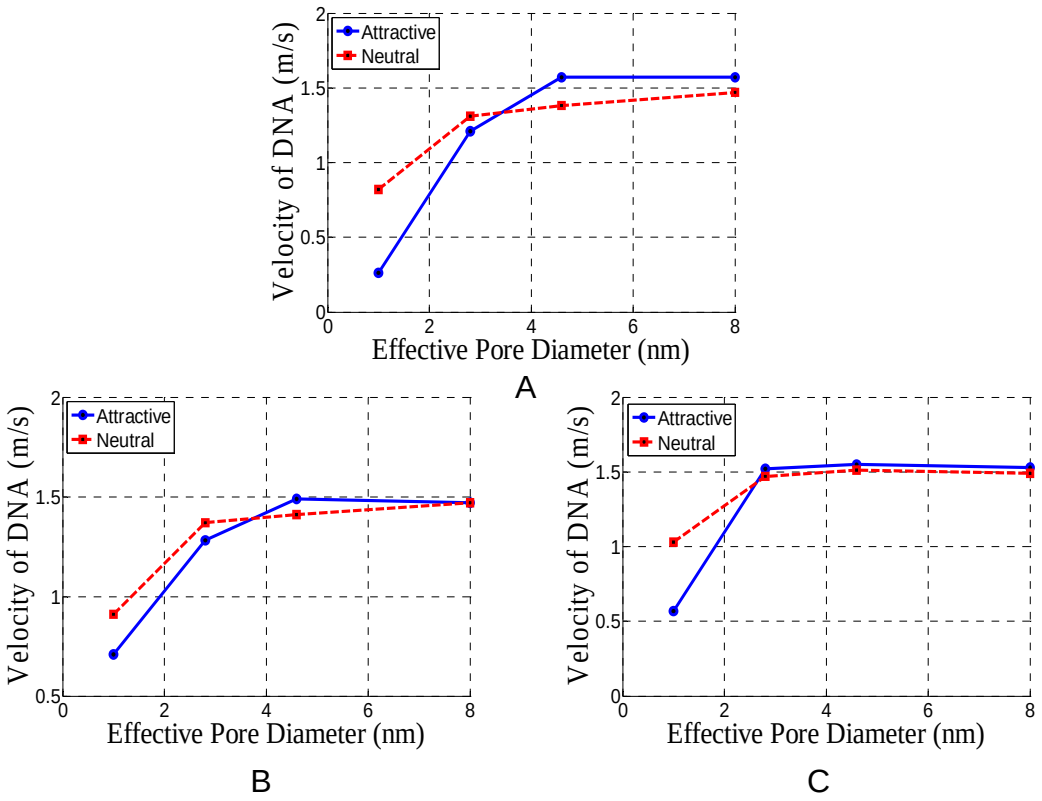


Figure 6.16. Influence of interaction potential between the translocating DNA and coating on its translocating velocity, each of the graphs gives a comparative analysis of the velocity of the translocating DNA with an attractive potential to that with a neutral potential with an increase in the EPD. Different coatings considered are (a) HPL, (b) ssDNA 1nm apart and (c) ssDNA 2nm apart.

The results obtained are shown in Fig. 6.16, which indicate significant difference in the velocity of the t-DNA in the smallest 1nm EPD configuration while in other systems velocities are approximately equal. This is mainly due to the large interaction between the t-DNA and the coatings in the smallest system (EPD = 1nm). As the EPD increases the interaction between the t-DNA and the coatings reduce which thereby decreases influence of t-DNA potential. The significant difference in the velocity obtained between the neutral potential t-DNA compared to the

attractive potential t-DNA for the 1nm EPD shows the impact of the potential in the system if sufficient interactions take place.

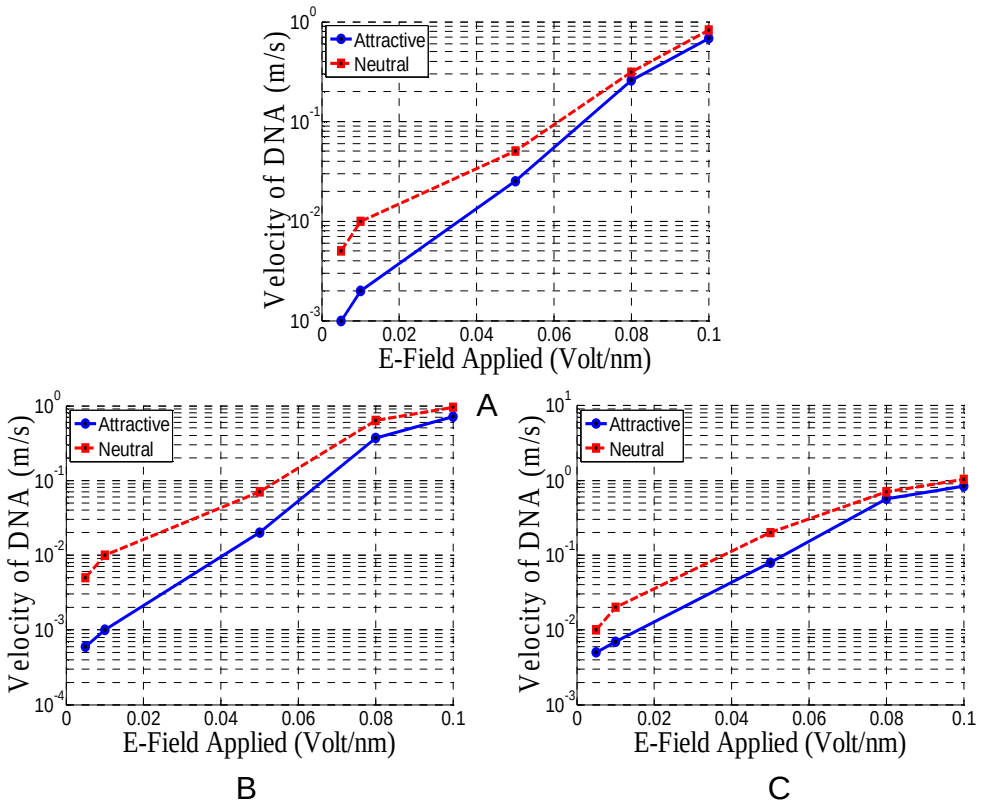


Figure 6.17. Influence of interaction potential on DNA translocating velocity, each graph gives a comparative analysis of the velocity of the translocating DNA with an attractive potential to that with a neutral potential with an increase in the voltage bias applied. Different coatings considered are (a) HPL, (b) ssDNA 1nm apart and (c) ssDNA 2nm apart.

To affirm the conclusion from the Fig. 6.16, we considered the case of smallest EPD with 1nm and observed the velocity of the t-DNA under the influence of gradually increasing E-field. The results are shown in Fig. 6.17, which indicates that there is significant difference in the translocation velocities and that t-DNA translocates faster under the neutral potential than that under attractive potential. The interaction potential becomes less dominant due to the increase in the applied bias voltage. Therefore, reduced difference of velocities between the t-DNA with neutral potential compared with the attractive potential is observed. The significant differences in the velocities purely based on potentials at the lower E-fields suggest that the modulation of

potential can also be of significant purpose in the future nanotechnology applications. This study could also be applied to see the effect of a t-DNA with a repulsive potential and something else, but those aspects are not in the scope of this paper.

6.4.5 DNA Sequence Selectivity

The main purpose of this research is to understand the kinetics with which the selectivity can be imparted to the solid-state nanopores. Having analyzed the different parameters in individual cases separately, we attempted to mimic the actual process of opening up of HPL structures into ssDNA and its effect on the t-DNA translocation velocity.

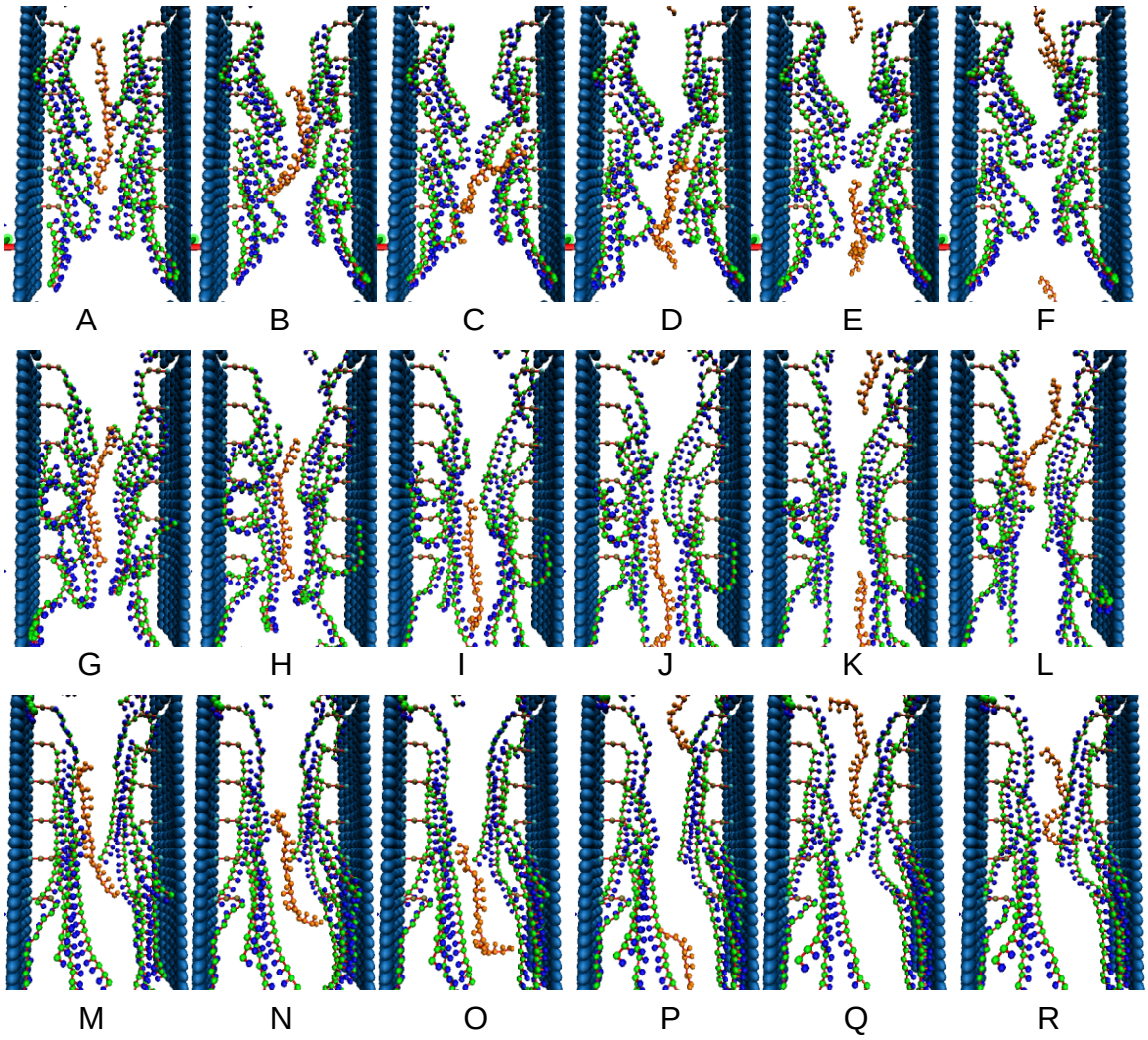


Figure 6.18. The process of t-DNA (translocating DNA) translocation in a HPL coated nanopore (A-F), t-DNA translocation in a HPL coated nanopore which opens up into ssDNA (G-L), t-DNA translocation in ssDNA (M-R).

Since the hybridization process and the opening of the HPL due to the presence of a matched DNA is a complex process we decided to run a similar simpler case, which compares a HPL coated pore with another HPL coated pore that opens into ssDNA under the application of the voltage bias. The images in Fig. 6.18 (a-f) show the translocation process of a mismatched t-DNA that has to translocate through the HPL coated nanopores. The images in Fig.6.18 (g-l) show the translocation process of a matched t-DNA that has to translocate through the HPL

coated nanopores which open into the ssDNA and thereby increase the EPD and facilitate the translocation process.

This is to understand, how the perfectly matched DNA takes lesser time to translocate through the HPL coated nanoproes compared to the mismatched one. Based on the analysis results from Fig.6.19, we can see that there is significant difference in the translocation times of the mismatched and the matched DNA. The matched DNA will translocate much faster compared to the mis-matched one, since the HPL opens up into the ssDNA and under the influence of the bias applied this increases the EPD and the overall effect causes the reduced time of translocation for the matched DNA.

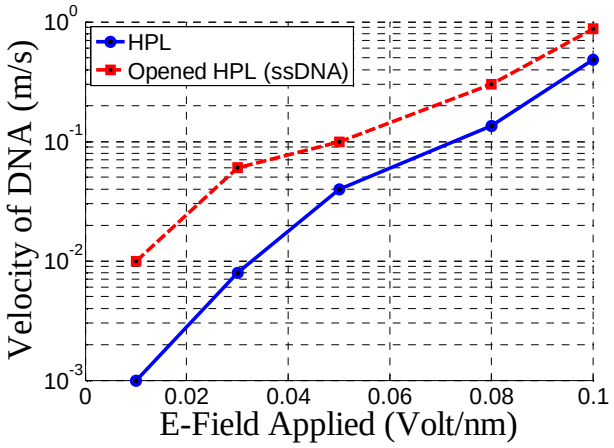


Figure 6.19. Effect of Bio-functionalization: This graph shows a mimic of the actual process by giving a comparison with the HPL coatings and what happens when they open up into ssDNA. The results clearly indicate that the main reason for the reduced translocation time for the matched DNA is opening of the HPL structures into the ssDNA, which increases the EPD and facilitates the translocation.

Based on the results obtained from the current set of simulations (Fig.6.19 and Table 6.2), we clearly see that there is a notable difference in the translocation velocity. This underscores the fact that the opening up of the HPL structures due to the presence of the matched DNA plays a very significant role in facilitating the transport of the matched DNA. While the mis-matched DNA has to force its way through the HPL coatings which are more rigid and it takes a longer translocation time.

Table 6.1. Velocity of the DNA when it translocates in a HPL coated nanopore in comparison with the HPL that opens into the ssDNA.

E-field (V/nm)	Velocity of DNA in HPL coated Nanopores (m/s)	Velocity of DNA in HPL opening to ssDNAs (m/s)	Percentage of increase
0.01	0.001	0.01	900.00%
0.03	0.008	0.06	650.00%
0.05	0.04	0.1	150.00%
0.08	0.135	0.3	122.22%
0.1	0.48	0.88	83.33%
0.3	3.48	4.33	24.42%

Furthermore, the percentage increase in the velocity of the DNA shown in table 6.1 is much larger at smaller bias voltages applied. In previous results, it is indicated that the density and coatings play a more influential role in the lower bias voltage range.

6.5 Conclusion

A CG system is developed to understand the kinetics of translocation of DNA in a bio-functionalized nanopore. This CG system is tested with three benchmark cases and ensured that it satisfactorily represents the underlying full atomistic scale system. The results are divided into two major parts, the effect bio-functionalization and the effect of the interaction potential as the E-field applied increases gradually. The parameters under consideration are the type of coatings applied, circumferential density, and the longitudinal density of the coatings. The EPD increases as the E-field applied increases and shows a parabolic trend which dictates the interaction between the t-DNA and the coatings. EPD for a HPL coated nanopore is significantly smaller than the nanopore coated ssDNA with the same density due to the differences in their stiffness with

the HPL being more rigid compared to the ssDNA. In case of densities, the eight strands circumferential coated nanopores has smaller EPD compared with the four strands coated nanopore. Further, even in case of longitudinal comparative study between ssDNAs placed 1nm apart and the ssDNAs placed 2nm apart, it is observed that the EPD is smaller for the denser ssDNA coated nanopore placed 1nm apart. This leads to the conclusion that a prominent way to control the EPD obtained will be to have a more rigid coating (dsDNA, HPL, ssDNA) or have a higher density of coatings. The next set of results are to understand the effect of interaction potential between the t-DNA and the coatings, the type and density are kept constant for different systems considered. It is observed that a t-DNA with an attractive potential to the coatings takes a longer translocation time as compared to that with a neutral interaction potential. However, as the EPD increases or the E-field applied increases, the difference in velocity becomes negligible and both t-DNAs have almost equal velocities.

The final section is an attempt to mimic the actual process by which selectivity can be imparted to the nanopores. The original process is that when a perfectly matched t-DNA translocates through the HPL coated nanopore, these coatings open up to hybridize with the translocating t-DNA. In case of a mismatch t-DNA, the coatings continue to remain as HPL, this results in the perfectly matched t-DNA translocating faster than the mismatched one. In our case, comparative study is carried between a t-DNA translocating in an HPL coated nanopore and the system in which this coated HPL opens up into ssDNA as the t-DNA translocates through bio-functionalized nanopore. The results indicate that there is significant increase in velocity of the t-DNA once the HPL opens up into the ssDNA as compared to the system in which the coatings continue to remain as HPL. These observations indicate that opening of HPL into the ssDNA facilitates the t-DNA translocation and thus helps imparting selectivity into the solid-state nanopore devices for DNA analysis.

CHAPTER 7

FUTURE WORK

High speed DNA translocation in bare nanopore is a limitation for nanopore based DNA analysis. The dependence of t-DNA translocation velocity on the potential of the coatings at lower voltages suggests that this can be used to control speed of the translocation process. Coatings with different potential within the nanopores modulate the t-DNA translocation velocity. In this thesis we characterized the attractive and neutral potential of coatings, in future this study can be extended to coatings with repulsive potential or mixed potentials of coatings which will help us optimize the system and obtain ideal parameters for reducing the speed of translocation of t-DNA.

Nanopore based DNA analysis is a promising next generation sequencing tool. A key aspect to be studied for perfectly replicating the actual process of gene sequencing in bio-functionalized nanopore is elaborated in this thesis. This last phase will involve characterizing the hybridization process with the interaction potential between the DNA and the HPL. This will help mimic the exact translocation process of the t-DNA in the bio-functionalized nanopore with the coatings opening up and hybridizing with the t-DNA as it is sufficiently close to them. This suggests that the parameters needed for obtaining the property-function relationship can be used for the optimal designs of the lab-on-chip DNA diagnostic devices. The targeted application schematics are shown in the Fig. 7.1. The understanding of this system could then be integrated with the real solid-state nanopore devices and bring us more close in achieving the ultimate goal of genome sequencing within $1000.

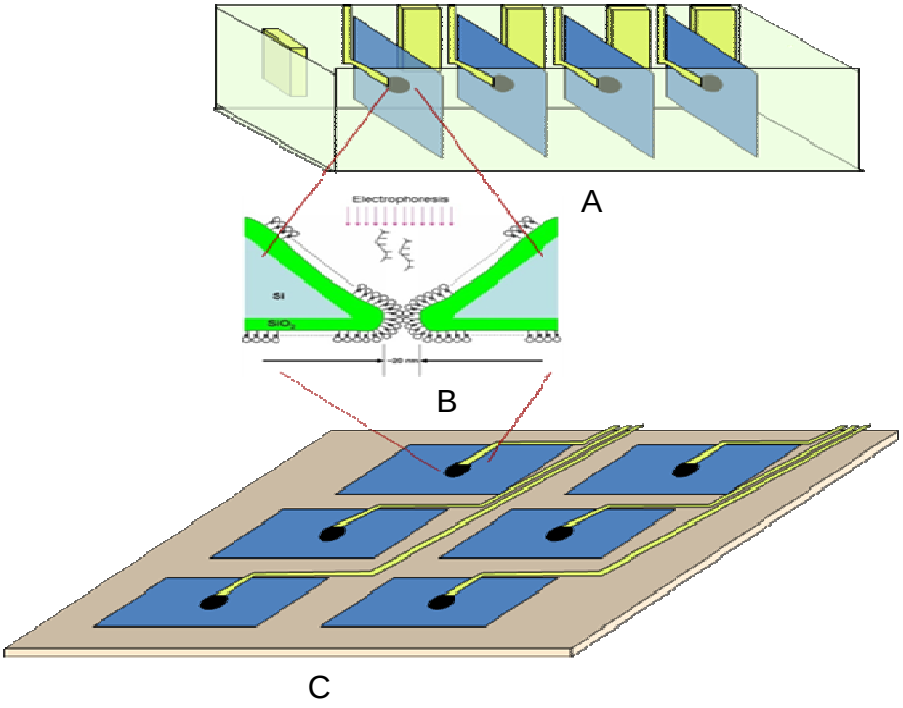


Figure 7.1. Targeted applications for bio-functionalized nanopore based DNA analysis, the serial arrangement (A) of bio-functionalized nanopores (B) while consists of the parallel arrangement (C). The analysis will be done with electrophoresis translocation of the DNA through the coated nanopores.

A major limitation of these bio-functionalized nanopores for gene sequencing is that once the coated HPLs open up for hybridizing with the perfectly matched t-DNA they do not form the HPL again. This means that the nanopore can be used only once. This limitation can be overcome, if by using molecular dynamics we can determine the conditions in which the ssDNA (opened HPL) can be made to form an HPL (original HPL) again within the nanopore.

Furthermore, similar concept can be extended to other biological system and similar functionalization schemes can be used for a variety of ligand-receptor combinations of significant importance. The solid-state functionalized nanopore can serve as next generation of sequencing tools, with the pore functionalized with specific probe can be used as detector for specific nucleotide(s)/biomarkers. [87, 167-169]

APPENDIX A

PROCEDURE AND SCRIPTS

All the files (mentioned below) are obtained using Matlab scripts but only the scripts to obtain the coordinate (pdb or gro) file is mentioned in this appendix.

File Types:

- .gro File: Files with the gro file extension contain information of molecular structure, like atom name, atom number, position, velocity and so on, in a fixed-column file format.
- .pdb File: Files with the pdb extension are molecular structure files in the protein databank file format, which describes the positions of atoms in a molecular structure.
- .itp File: Files with the itp extension define the topology for a single specific molecule type. The topology contains a complete description of all the interactions within and between atoms/particles in the system.
- .ndx File: A ndx file is an index file, which is required only when special groups are required. Deafult index groups include entire system, 9 groups for proteins, and one for every other residue name.
- .top File: The top file extension stands for topology, which contains a complete description of all the interactions within and between atoms/particles in the system.

Procedure to obtain CG model of “ssDNA within a nanopore”

1. Use Script-1, to obtain the gro file for ssDNA which is straight in length and minimize this structure to obtain the ssDNA with the helical nature, using the itp file and top file with correct parameters.
2. Use Script-2, to obtain the gro file for the nanopore with sufficient radius and length.
3. Load the two files and place the DNA in the center of the nanopore using VMDs tcl scripting.
4. Shift this whole system into the positive quadrant, keeping in mind the PBC conditions. Save the new set of coordinates which will give a pdb file since VMD is used.
5. Convert these ssDNA.pdb and pore.pdb files into separate gro files using Script-4 and Script-5 respectively. Recombine the file latter.

6. This should give a system coordinate file with ssDNA placed in the center of the nanopore.
7. Obtain the other required files for running the simulation.
8. Run the simulation in Gromacs to understand DNA translocation in a bare nanopore using a Coarse Grained Model.

The scripts to obtain pdb (coordinate) file of the nanopore, the two-site DNA strand, the three-site DNA strand, script for conversion of pdb to gro for ssDNA and script for conversion of pdb to gro for nanopore are enclosed in this section.

1. Two-site DNA model:

```
% Code to built Coarse Grained Two-site DNA by Abhijit Ramachandran.

clc;

clear all;

fprintf('\n\nThis code by Abhijit Ramachandran is to built a coarse
grained two-site single strand of DNA\n');
i=1;
j=1;
x1=1.5;
for varib =j:1:30
    x(i)=x1;
    x1=x1+0.5
    i=i+1;
end

% Reinitialization for Output
Abhijit=20;
j=1;
k=4;
i=1;
z=0;
y=0;
y1=0.5;
% writing outputs in gromacs file format

file_2 = fopen('TwoSiteDNA.txt','w');
fprintf(file_2,'Abhijit coordinate generation code used\n');
fprintf(file_2,'  %3i\n',Abhijit);

for vari = j:1:Abhijit
%     if ((vari > 0 && vari < 10))
    fprintf(file_2,'           1plan1           P%1i                  %4d
%8.3f%8.3f%8.3f\n',k,k,x(i),y,z);
```

```
    k=k+1;
    fprintf(file_2,'           1plan1          P%1i                %4d
%8.3f%8.3f%8.3f\n',k,k,x(i),y1,z);

    k=k+1;
    i=i+1;
end

fprintf(file_2,'2.50000  2.7800  2.0000');
```

2. Coarse Grained Nanopore:

```
% Code to built Coarse Grained Pore by Abhijit Ramachandran.

clc;

clear all;

fprintf('\n\nThis code by Abhijit Ramachandran is to built a coarse
grained cylindrical pore\n');

% %input from the user
% radius = input('\nRadius of the cylindrical pore, kindly input an
integer value: ');
% length = input('\nLength of the cylindrical pore, kindly input an
even integer value: ');
% bondlength = input('\nBond Length for the cylindrical pore in unit
nm; default is 0.523 which is perfect so kindly put that if not sure:
');

%file_2 = fopen('HPL_6sd.txt','w'); %Output filenames depending on the
case
file_2 = fopen('St_1nm_8sd.txt','w');
%file_2 = fopen('St_2nm_8sd.txt','w');

%Initialization
rho = radius;
i=1;
l = 2*pi; %total angle of a circle
ll=length/2;
bl = bondlength;
arcl = bl/rho; %arc length for uniform placement of atoms on the cirlce

    %Extended circle to form a cylinder in Z
    for z=-ll:bl:ll

        %X Y - circle
        for theta=0:arcl:l
            [X,Y,Z] = pol2cart(theta,rho,z);
            cartcoordx(i) = X;
            cartcoordy(i) = Y;
            cartcoordz(i) = Z;
            i=i+1;
```

```
        end

    end

% column to row conversion
xf = cartcoordx';
yf = cartcoordy';
zf = cartcoordz';

% Reinitialization for Output
Abhijit=i-1;
j=1;
k=1;
i=1;

% writing outputs in gromacs file format
%file_2 = fopen('TwoSiteDNA.txt','w');
fprintf(file_2,'Abhijit coordinate generation code used\n');
fprintf(file_2,'  %3i\n',Abhijit);

for vari = j:1:Abhijit
    if ((vari > 0 && vari < 10))
    fprintf(file_2,'                     1plan1                     P%1i
%5d%8.3f%8.3f%8.3f\n',k,k,xf(i),yf(i),zf(i));
    elseif (vari > 9 && vari < 100)
    fprintf(file_2,'                     1plan1                     P%2i
%5d%8.3f%8.3f%8.3f\n',k,k,xf(i),yf(i),zf(i));
    elseif (vari > 99 && vari < 1000)
    fprintf(file_2,'                     1plan1                     P%3i
%5d%8.3f%8.3f%8.3f\n',k,k,xf(i),yf(i),zf(i));
    else
    fprintf(file_2,'                                                1plan1
P%4i%5d%8.3f%8.3f%8.3f\n',k,k,xf(i),yf(i),zf(i));
    end
  %General GRO format 1WATER  OW1    1   0.126   1.624   1.679
    k=k+1;
    i=i+1;
end

fprintf(file_2,'2.50000  2.7800  2.0000');
```

3. Three-site DNA model:

```
% Code to built Coarse Grained Three-site DNA by Abhijit Ramachandran.

% Script for getting the coarse grain coordiante structures
% Code is Built by Abhijit Ramachandran
% Contact: abhijit.ramachandran@mavs.uta.edu

clc;
clear all;
%_____________________________________________________________________

%------Specify First Coordinates for sugar, phosphate & BASE ---------
```

```
%------Also the number of base pairs to be generated-----------------

fprintf('\n\nThis code is limited by 33 base-pairs\n');
Nb = input('\nCoarseGrain PDB file is to be built for how many base-
pairs, multiple of 3 needed: ');

Xb = 1.472;
Yb = 1.243;
Zb = 1.119;

Xs = 1.063;
Ys = 2.054;
Zs = 1.005;

Xp = 1.237;
Yp = 2.657;
Zp = 0.872;

% _____________________________________________SUGAR_____________________

i=1;
j=1;
L=Nb;


X = Xs;
Y = Ys;
Z = Zs;

[THETA,RHO,Z] = cart2pol(X,Y,Z);

theta1 = THETA;
rho1 = RHO;
z1 = Z;

for vari = j:1:L

[X,Y,Z] = pol2cart(theta1,rho1,z1);

cartcoordx1(i) = X;
cartcoordy1(i) = Y;
cartcoordz1(i) = Z;

% while i == 99
%     break;
% end

i = i+1;
%m = length(cartcoordx)
theta1 = theta1 + 0.6283; %Conversion of Theta increment in radians
z1 = z1+3.38;

end
```

```
xs = cartcoordx1';
ys = cartcoordy1';
zs = cartcoordz1';
%-------------------------------------------------------------------------
%_____________________________________PHOSPHATE_____________________________

X = Xp;
Y = Yp;
Z = Zp;

i=1;

[THETA,RHO,Z] = cart2pol(X,Y,Z);

theta2 = THETA;
rho2 = RHO;
z2 = Z;

for vari = j:1:L

[X,Y,Z] = pol2cart(theta2,rho2,z2);

cartcoordx2(i) = X;
cartcoordy2(i) = Y;
cartcoordz2(i) = Z;

i = i+1;

theta2 = theta2 + 0.6283; %Conversion of Theta increment in radians
z2 = z2+3.38;

end

xp = cartcoordx2';
yp = cartcoordy2';
zp = cartcoordz2';

%-------------------------------------------------------------------------
% Code is Built by Abhijit Ramachandran
% Contact: abhijit.ramachandran@mavs.uta.edu
%_____________________________________BASE_____________________________

X = Xb;
Y = Yb;
Z = Zb;
i=1;

[THETA,RHO,Z] = cart2pol(X,Y,Z);

theta3 = THETA;
rho3 = RHO;
z3 = Z;

for vari = j:1:L
```

```
[X,Y,Z] = pol2cart(theta3,rho3,z3);

cartcoordx3(i) = X;
cartcoordy3(i) = Y;
cartcoordz3(i) = Z;

i = i+1;

theta3 = theta3 + 0.6283; %Conversion of Theta increment in radians
z3 = z3+3.38;

end

xb = cartcoordx3';
yb = cartcoordy3';
zb = cartcoordz3';

%-------------------------------------------------------------------------
%_____________________________________PRINT_____________________________________


file_2 = fopen('CG_dsdna_33bp.pdb','w');
i = 1;
k=1;

for vari = j:1:L

    fprintf(file_2, 'ATOM     %3.0i AD1   ADE A%3.0i      %8.3f%8.3f%8.3f
1.00  1.00      ADNA\n',k,i,xb(i),yb(i),zb(i));
    k=k+1;

    fprintf(file_2,    'ATOM            %3.0i    S%2.0i       SUG    A%3.0i
%8.3f%8.3f%8.3f  1.00  1.00      ADNA\n',k,i,i,xs(i),ys(i),zs(i));
    k=k+1;

    fprintf(file_2,    'ATOM            %3.0i    P%2.0i       PHO    A%3.0i
%8.3f%8.3f%8.3f  1.00  1.00      ADNA\n',k,i,i,xp(i),yp(i),zp(i));
    k=k+1;
    Abhijit=i;
    i=i+1;
end
%_____________________________________________END_____________________________


% Code is Built by Abhijit Ramachandran
% Contact: abhijit.ramachandran@mavs.uta.edu

%_____________________________________STRAND TWO_____________________


%_____________________________________SUGAR_____________________
LL=L+1;
```

```
i=501;
j=1;

X = Xs;
Y = -Ys;
Z = -Zs;

[THETA,RHO,Z] = cart2pol(X,Y,Z);

theta1 = THETA;
rho1 = RHO;
z1 = Z;

for vari = j:1:LL

[X,Y,Z] = pol2cart(theta1,rho1,z1);

cartcoordx1(i) = X;
cartcoordy1(i) = Y;
cartcoordz1(i) = Z;

i = i+1;

theta1 = theta1 + 0.6283; %Conversion of Theta increment in radians
z1 = z1+3.38;

end

xs = cartcoordx1';
ys = cartcoordy1';
zs = cartcoordz1';
%-------------------------------------------------------------------------
----
%_____________________________________PHOSPHATE_____________________________
____

X = Xp;
Y = -Yp;
Z = -Zp;

i=501;

[THETA,RHO,Z] = cart2pol(X,Y,Z);

theta2 = THETA;
rho2 = RHO;
z2 = Z;

for vari = j:1:LL

[X,Y,Z] = pol2cart(theta2,rho2,z2);

cartcoordx2(i) = X;
cartcoordy2(i) = Y;
```

```
cartcoordz2(i) = Z;

i = i+1;

theta2 = theta2 + 0.6283; %Conversion of Theta increment in radians
z2 = z2+3.38;

end

xp = cartcoordx2';
yp = cartcoordy2';
zp = cartcoordz2';

%-------------------------------------------------------------------------
----
% Code is Built by Abhijit Ramachandran
% Contact: abhijit.ramachandran@mavs.uta.edu
%_____________________________________BASE_____________________________
____


X = Xb;
Y = -Yb;
Z = -Zb;

i=501;

[THETA,RHO,Z] = cart2pol(X,Y,Z);

theta3 = THETA;
rho3 = RHO;
z3 = Z;

for vari = j:1:LL

[X,Y,Z] = pol2cart(theta3,rho3,z3);

cartcoordx3(i) = X;
cartcoordy3(i) = Y;
cartcoordz3(i) = Z;

i = i+1;

theta3 = theta3 + 0.6283; %Conversion of Theta increment in radians
z3 = z3+3.38;

end

xb = cartcoordx3';
yb = cartcoordy3';
zb = cartcoordz3';

%-------------------------------------------------------------------------
----
```

```
% Code is Built by Abhijit Ramachandran
% Contact: abhijit.ramachandran@mavs.uta.edu
%_____________________________________PRINT_____________________________
____


i = 501;
k=501;

for vari = j:1:L

     fprintf(file_2, 'ATOM    %3.0i AD1  ADE B%3.0i     %8.3f%8.3f%8.3f
1.00  1.00      BDNA\n',k,i,xb(i+1),yb(i+1),zb(i+1));
    k=k+1;

    fprintf(file_2,   'ATOM            %3.0i   S%2.0i      SUG   B%3.0i
%8.3f%8.3f%8.3f                        1.00                        1.00
BDNA\n',k,Abhijit,i,xs(i+1),ys(i+1),zs(i+1));
    k=k+1;

    fprintf(file_2,   'ATOM            %3.0i   P%2.0i      PHO   B%3.0i
%8.3f%8.3f%8.3f                        1.00                        1.00
BDNA\n',k,Abhijit,i,xp(i+1),yp(i+1),zp(i+1));
    k=k+1;
    Abhijit=Abhijit+1;
    i=i+1;

end

fprintf('\nThe CoarseGrain PDB file of %4.0f base-pairs is successfully
completed \n',Nb);
fprintf('\nI, Abhijit Ramachandran is Glad that this code worked for
you\n\n');

% Code is Built by Abhijit Ramachandran
% Contact: abhijit.ramachandran@mavs.uta.edu
```

4. PDB to Gromacs file conversion script (e.g. ssDNA)

```
% Script to convert PDB to GRO file, by Abhijit Ramachandran
%This code has a small error, since there are multiple statements to be
%printed within the if loop, it might be that the number exceeds in the
%second print out line, so kindly sheck the output for such an error,
it
%should only be at 2 places numbered 10 and 100 in the output Names

clear all;
clc;

%Reading the pdb file named strand2.txt in this case.
[index, id, name, grp, tp, resid, x1, y1, z1, beta, occ] =
textread('1.txt', '%s %d %s %s %s %d %f %f %f %f %f', 9999);
  %ATOM  15 N760 plan1 X 01    -16.761   0.000   0.000  0.00  0.00
```

```
%An output file is selected
file_2 = fopen('cg_dna30A.txt','w');

%Initialization
k=1;
length=length(x1);
abhijit=length-1;
strands=length/43;
bases=20;
begin=1681;
begin=1;%This is used only for veriyinf the output
j=1;
i=1;

x=x1/10;
y=y1/10;
z=z1/10;

%Writing the GRO file.
fprintf(file_2,'MD coordinate generation\n'); %First Line
fprintf(file_2,'  1408\n'); %Second Line

for vari = k:1:strands
    if ((j > 0 && j < 10))
        fprintf(file_2,'  2TET    C%1i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET    P%1i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET    P%1i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
    elseif (j > 9 && j < 100)
        fprintf(file_2,'  2TET    C%2i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET    P%2i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET    P%2i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
```

```
        i=i+1;
        begin=begin+1;
    elseif (j > 99 && j < 1000)
        fprintf(file_2,'  2TET    C%3i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET    P%3i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET    P%3i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
     else
        fprintf(file_2,'  2TET
C%4i%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET
P%4i%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
        fprintf(file_2,'  2TET
P%4i%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;
    end
end

%PArt 2
ending=bases*strands;
j=1;
for vari = k:1:ending

        if ((j > 0 && j < 10))
        fprintf(file_2,'  2TET    Q%1i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;

        fprintf(file_2,'  2TET    N%1i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
```

```
        begin=begin+1;
        elseif (j > 9 && j < 100)
        fprintf(file_2,'  2TET    Q%2i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;

        fprintf(file_2,'  2TET    N%2i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;

        elseif (j > 99 && j < 1000)

        fprintf(file_2,'  2TET    Q%3i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        i=i+1;
        j=j+1;
        begin=begin+1;

        fprintf(file_2,'  2TET    N%3i
%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        j=j+1;
        i=i+1;
        begin=begin+1;

        else
        fprintf(file_2,'  2TET
Q%4i%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));
        i=i+1;
        j=j+1;
        begin=begin+1;

        fprintf(file_2,'  2TET
N%4i%5d%8.3f%8.3f%8.3f\n',j,begin,x(i),y(i),z(i));

        j=j+1;
        i=i+1;
        begin=begin+1;
        end
end

fprintf(file_2,'10.00000  10.0000  10.0000');%lastline
```

5. PDB to Gromacs file conversion script (e.g. cylindrical pore)

```
% Script to convert PDB to GRO file, by Abhijit Ramachandran.

clear all;
```

```
clc;

%Delete the first line of the pdb file and the last line.
%Reading the pdb file named strand2.txt in this case.
[index, id, name, grp, tp, resid, x1, y1, z1, beta, occ] =
textread('pore.pdb', '%s %d %s %s %s %d %f %f %f %f %f', 9999);

%Beware for the X, in some pdb files it is not present.
%ATOM    310 OT31 TRL X   1      20.159   9.066   3.457  1.00  0.00

file_2 = fopen('PoreCylindrica.txt','w');%An output file is selected

%Initialization
k=1;
abhijit=length(index);

x=x1/10;
y=y1/10;
z=z1/10;

%Writing the GRO file.

j=1;%just an increment variable

%write the GRO file complete


fprintf(file_2,'Abhijit coordinate generation code used\n');%First Line
fprintf(file_2,'  880\n');%Second Line

for vari = k:1:abhijit
    if ((vari > 0 && vari < 10))
    fprintf(file_2,'  1plan1  P%1i
%5d%8.3f%8.3f%8.3f\n',j,j,x(j),y(j),z(j));
    elseif (vari > 9 && vari < 100)
    fprintf(file_2,'  1plan1  P%2i
%5d%8.3f%8.3f%8.3f\n',j,j,x(j),y(j),z(j));
    elseif (vari > 99 && vari < 1000)
    fprintf(file_2,'  1plan1  P%3i
%5d%8.3f%8.3f%8.3f\n',j,j,x(j),y(j),z(j));
    else
    fprintf(file_2,'  1plan1
P%4i%5d%8.3f%8.3f%8.3f\n',j,j,x(j),y(j),z(j));
    end
    j=j+1;
end

fprintf(file_2,'2.50000  2.7800  2.0000');
```

REFERENCES

1. Alberts, B., Molecular Biology of the Cell. *Garland Science- 4th edition* **2002**.
2. Venter, J. C., The sequence of the human genome (vol 292, pg 1304, 2001). *Science* **2001,** 292, (5523), 1838-1838.
3. Watson, J. D., The Human Genome Project - Past, Present, and Future. *Science* **1990,** 248, (4951), 44-49.
4. Cantor, C. R., Orchestrating the Human Genome Project. *Science* **1990,** 248, (4951), 49-51.
5. Cantor, C. R.; Tang, K.; Graber, J. H.; Maloney, M.; Fu, D. J.; Broude, N. E.; Siddiqi, F.; Koester, H.; Smith, C. L., DNA sequencing after the Human Genome Project. *Nucleosides & Nucleotides* **1997,** 16, (5-6), 591-598.
6. Service, R. F., Gene sequencing - The race for the $1000 Genome. *Science* **2006,** 311, (5767), 1544-1546.
7. Sanger, F.; Nicklen, S.; Coulson, A. R., DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **1977,** 74, (12), 5463-5467.
8. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X. H.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A., The potential and challenges of nanopore sequencing. *Nature Biotechnology* **2008,** 26, (10), 1146-1153.
9. Iqbal, S. M.; Akin, D.; Bashir, R., Solid-state nanopore channels with DNA selectivity. *Nature Nanotechnology* **2007,** 2, (4), 243-248.
10. Bauer, W. R.; Nadler, W., Molecular transport through channels and pores: Effects of in-channel interactions and blocking. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** 103, (31), 11446-11451.
11. Ramachandran, A.; Liu, Y.; Asghar, W.; Iqbal, S. M., Characterization of DNA-Nanopore Interactions by Molecular Dynamics. *American Journal of Biomedical Sciences* **2009**.
12. Gilchrist, T. L., Heterocyclic Chemistry. *Prentice Hall* **1997**.
13. Watson, J. D.; Crick, F. H. C., Molecular-Structure of Nucleic-Acids - a Structure for Deoxyribose Nucleic-Acid. *Nature* **1953,** 171, 737 - 738.
14. Wink, M., An Introduction to Molecular Biotechnology: Molecular Fundamentals, Methods and Applications in Modern Biotechnology *Wiley-VCH* **2006**.
15. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D., Base-stacking and base-pairing contributions into thermal stability of the DNA double helix (vol 34, pg 564, 2006). *Nucleic Acids Research* **2006,** 34, (3), 1082-+.
16. Purves, W. K.; Sadava, D.; Orians, G. H.; Heller, H. C., Life: The Science of Biology, 7th Edition. *Sinauer Associates and W. H. Freeman* **2003**.
17. Wheeler, D. A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y. J.; Makhijani, V.; Roth, G. T.; Gomes, X.; Tartaro, K.; Niazi, F.; Turcotte, C. L.; Irzyk, G. P.; Lupski, J. R.; Chinault, C.; Song, X. Z.; Liu, Y.; Yuan, Y.; Nazareth, L.; Qin, X.; Muzny, D. M.; Margulies, M.; Weinstock, G. M.; Gibbs, R. A.; Rothberg, J. M., The complete genome of an individual by massively parallel DNA sequencing. *Nature* **2008,** 452, (7189), 872-U5.
18. Kim, S.; Tang, H.; Mardis, E. R., Genome Sequencing Technology and Algorithms. *Artech House Publishers* **2007**.

19. Swerdlow, H.; Gesteland, R., Capillary Gel-Electrophoresis for Rapid, High-Resolution DNA Sequencing. *Nucleic Acids Research* **1990,** 18, (6), 1415-1419.
20. Janitz, M., Next-Generation Genome Sequencing: Towards Personalized Medicine. *Wiley-VCH* **2008**.
21. Huang, X. H. C.; Quesada, M. A.; Mathies, R. A., DNA Sequencing Using Capillary Array Electrophoresis. *Analytical Chemistry* **1992,** 64, (18), 2149-2154.
22. Sussman, H. E.; Smit, M. A., Genomes. *Cold Spring Harbor Laboratory Press* **2006**.
23. Metzker, M. L., Emerging technologies in DNA sequencing. *Genome Research* **2005,** 15, (12), 1767-1776.
24. Xu, F.; Baba, Y., DNA Sequencing Studies by Capillary Electrophoresis *Encyclopedia of Chromatography* **2005**.
25. Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E., Fluorescence Detection in Automated DNA-Sequence Analysis. *Nature* **1986,** 321, (6071), 674-679.
26. Prober, J. M.; Trainor, G. L.; Dam, R. J.; Hobbs, F. W.; Robertson, C. W.; Zagursky, R. J.; Cocuzza, A. J.; Jensen, M. A.; Baumeister, K., A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. *Science* **1987,** 238, (4825), 336-341.
27. Lewis, E. K.; Haaland, W. C.; Nguyen, F.; Heller, D. A.; Allen, M. J.; MacGregor, R. R.; Berger, C. S.; Willingham, B.; Burns, L. A.; Scott, G. B. I.; Kittrell, C.; Johnson, B. R.; Curl, R. F.; Metzker, M. L., Color-blind fluorescence detection for four-color DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (15), 5346-5351.
28. Karger, A. E.; Harris, J. M.; Gesteland, R. F., Multiwavelength Fluorescence Detection for DNA Sequencing Using Capillary Electrophoresis. *Nucleic Acids Research* **1991,** 19, (18), 4955-4962.
29. Lee, L. G.; Spurgeon, S. L.; Heiner, C. R.; Benson, S. C.; Rosenblum, B. B.; Menchen, S. M.; Graham, R. J.; Constantinescu, A.; Upadhya, K. G.; Cassel, J. M., New energy transfer dyes for DNA sequencing. *Nucleic Acids Research* **1997,** 25, (14), 2816-2822.
30. Zhu, L.; Stryjewski, W. J.; Soper, S. A., Multiplexed fluorescence detection in microfabricated devices with both time-resolved and spectral-discrimination capabilities using near-infrared fluorescence. *Analytical Biochemistry* **2004,** 330, (2), 206-218.
31. Alaverdian, L.; Alaverdian, S.; Bilenko, O.; Bogdanov, I.; Filippova, E.; Gavrilov, D.; Gorbovitski, B.; Gouzman, M.; Gudkov, G.; Domratchev, S.; Kosobokova, O.; Lifshitz, N.; Luryi, S.; Ruskovoloshin, V.; Stepoukhovitch, A.; Tcherevishnick, M.; Tyshko, G.; Gorfinkel, V., A family of novel DNA sequencing instruments based on single-photon detection. *Electrophoresis* **2002,** 23, (16), 2804-2817.
32. Ronaghi, M.; Uhlen, M.; Nyren, P., A sequencing method based on real-time pyrophosphate. *Science* **1998,** 281, (5375), 363-+.
33. Langaee, T.; Ronaghi, M., Genetic variation analyses by Pyrosequencing. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **2005,** 573, (1-2), 96-102.
34. Brown, R. J., Comparative Genomics: Basic and Applied Research. *CRC press* **2007**.
35. Voelkerding, K. V.; Dames, S. A.; Durtschi, J. D., Next-Generation Sequencing: From Basic Research to Diagnostics. *Clinical Chemistry* **2009,** 55, (4), 641-658.
36. Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X. X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M., Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **2005,** 309, (5741), 1728-1732.
37. Hofmann, W. K., Gene Expression Profiling by Microarrays: Clinical Implications *Cambridge University Press* **2006**.
38. Coulter, W. H., "Means for counting particles suspended in a fluid." U.S. Patent Number 2656508. **1953**.
39. Deblois, R. W.; Bean, C. P.; Wesley, R. K. A., Electrokinetic Measurements with Submicron Particles and Pores by Resistive Pulse Technique. *Journal of Colloid and Interface Science* **1977,** 61, (2), 323-335.

40. Hille, B., Ion Channels of Excitable Membranes. *Sinauer Associates 3rd Casebound edition* **2001**.
41. Healy, K., Nanopore-based DNA analysis. *PhD Thesis, Dept. of Electrical and Electronic Engineering, University College Cork,* **2006**.
42. Neher, E.; Sakmann, B., Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **1976,** 260, (5554), 799–802.
43. Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W., Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **1996,** 93, (24), 13770-13773.
44. Church, G. M.; Deamer, D. W.; Branton, D.; Baldarelli, R.; Kasianowicz, J. J., Characterization of individual polymer molecules based on monomer-interface interactions. *US Patent 5795782* **1998**.
45. Deamer, D. W.; Akeson, M., Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends in Biotechnology* **2000,** 18, (4), 147-151.
46. Nakane, J. J.; Akeson, M.; Marziali, A., Nanopore sensors for nucleic acid analysis. *Journal of Physics-Condensed Matter* **2003,** 15, (32), R1365-R1393.
47. Wang, G. M.; Sandberg, A. C., Non-equilibrium all-atom molecular dynamics simulations of free and tethered DNA molecules in nanochannel shear flows. *Nanotechnology* **2007,** 18, (13), 135702-135710.
48. Akeson, M.; Branton, D.; Kasianowicz, J. J.; Brandin, E.; Deamer, D. W., Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophysical Journal* **1999,** 77, (6), 3227-3233.
49. Meller, A., Dynamics of polynucleotide transport through nanometre-scale pores. *Journal of Physics-Condensed Matter* **2003,** 15, (17), R581-R607.
50. Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D., Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** 97, (3), 1079-1084.
51. Meller, A.; Nivon, L.; Branton, D., Voltage-driven DNA translocations through a nanopore. *Physical Review Letters* **2001,** 86, (15), 3435-3438.
52. Mathe, J.; Aksimentiev, A.; Nelson, D. R.; Schulten, K.; Meller, A., Orientation discrimination of single-stranded DNA inside the alpha-hemolysin membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (35), 12377-12382.
53. Butler, T. Z.; Gundlach, J. H.; Troll, M. A., Determination of RNA orientation during translocation through a biological nanopore. *Biophysical Journal* **2006,** 90, (1), 190-199.
54. Wang, H.; Dunning, J. E.; Huang, A. P. H.; Nyamwanda, J. A.; Branton, D., DNA heterogeneity and phosphorylation unveiled by single-molecule electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** 101, (37), 13472-13477.
55. Henrickson, S. E.; Misakian, M.; Robertson, B.; Kasianowicz, J. J., Driven DNA transport into an asymmetric nanometer-scale pore. *Physical Review Letters* **2000,** 85, (14), 3057-3060.
56. Nakane, J.; Akeson, M.; Marziali, A., Evaluation of nanopores as candidates for electronic analyte detection. *Electrophoresis* **2002,** 23, (16), 2592-2601.
57. Aksimentiev, A.; Heng, J. B.; Cruz-Chu, E. R.; Timp, G.; Schulten, K., Microscopic kinetics of DNA translocation through synthetic nanopores. *Biophysical Journal* **2005,** 88, (1), 352a-352a.
58. Aksimentiev, A.; Heng, J. B.; Timp, G.; Schulten, K., Microscopic kinetics of DNA translocation through synthetic nanopores. *Biophysical Journal* **2004,** 87, (3), 2086-2097.
59. Aksimentiev, A.; Schulten, K., Imaging alpha-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophysical Journal* **2005,** 88, (6), 3745-3761.
60. Gu, L. Q.; Cheley, S.; Bayley, H., Capture of a single molecule in a nanocavity. *Science* **2001,** 291, (5504), 636-640.

61. Dekker, C., Solid-state nanopores. *Nature Nanotechnology* **2007,** 2, (4), 209-215.
62. Smeets, R. M. M.; Keyser, U. F.; Krapf, D.; Wu, M. Y.; Dekker, N. H.; Dekker, C., Salt dependence of ion transport and DNA translocation through solid-state nanopores. *Nano Letters* **2006,** 6, (1), 89-95.
63. Ho, C.; Qiao, R.; Heng, J. B.; Chatterjee, A.; Timp, R. J.; Aluru, N. R.; Timp, G., Electrolytic transport through a synthetic nanometer-diameter pore. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (30), 10445-10450.
64. Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M. J.; Golovchenko, J. A., Ion-beam sculpting at nanometre length scales. *Nature* **2001,** 412, (6843), 166-169.
65. Chang, H.; Kosari, F.; Andreadakis, G.; Alam, M. A.; Vasmatzis, G.; Bashir, R., DNA-mediated fluctuations in ionic current through silicon oxide nanopore channels. *Nano Letters* **2004,** 4, (8), 1551-1556.
66. Bai, J. G.; Chang, C. L.; Chung, J. H.; Lee, K. H., Shadow edge lithography for nanoscale patterning and manufacturing. *Nanotechnology* **2007,** 18, (40), -.
67. Stein, D.; Kruithof, M.; Dekker, C., Surface-charge-governed ion transport in nanofluidic channels. *Physical Review Letters* **2004,** 93, (3), -.
68. Siwy, Z.; Heins, E.; Harrell, C. C.; Kohli, P.; Martin, C. R., Conical-nanotube ion-current rectifiers: The role of surface charge. *Journal of the American Chemical Society* **2004,** 126, (35), 10850-10851.
69. Fan, R.; Karnik, R.; Yue, M.; Li, D. Y.; Majumdar, A.; Yang, P. D., DNA translocation in inorganic nanotubes. *Nano Letters* **2005,** 5, (9), 1633-1637.
70. Healy, K.; Schiedt, B.; Morrison, A. P., Solid-state nanopore technologies for nanopore-based DNA analysis. *Nanomedicine* **2007,** 2, (6), 875-897.
71. Li, J. L.; Gershow, M.; Stein, D.; Brandin, E.; Golovchenko, J. A., DNA molecules and configurations in a solid-state nanopore microscope. *Nature Materials* **2003,** 2, (9), 611-615.
72. Mara, A.; Siwy, Z.; Trautmann, C.; Wan, J.; Kamme, F., An asymmetric polymer nanopore for single molecule detection. *Nano Letters* **2004,** 4, (3), 497-501.
73. Heng, J. B.; Ho, C.; Kim, T.; Timp, R.; Aksimentiev, A.; Grinkova, Y. V.; Sligar, S.; Schulten, K.; Timp, G., Sizing DNA using a nanometer-diameter pore. *Biophysical Journal* **2004,** 87, (4), 2905-2911.
74. Storm, A. J.; Storm, C.; Chen, J. H.; Zandbergen, H.; Joanny, J. F.; Dekker, C., Fast DNA translocation through a solid-state nanopore. *Nano Letters* **2005,** 5, (7), 1193-1197.
75. Heng, J. B.; Aksimentiev, A.; Ho, C.; Marks, P.; Grinkova, Y. V.; Sligar, S.; Schulten, K.; Timp, G., The electromechanics of DNA in a synthetic nanopore. *Biophysical Journal* **2006,** 90, (3), 1098-1106.
76. Heng, J. B.; Aksimentiev, A.; Ho, C.; Marks, P.; Grinkova, Y. V.; Sligar, S.; Schulten, K.; Timp, G., Stretching DNA using the electric field in a synthetic nanopore. *Nano Letters* **2005,** 5, (10), 1883-1888.
77. Fologea, D.; Gershow, M.; Ledden, B.; McNabb, D. S.; Golovchenko, J. A.; Li, J. L., Detecting single stranded DNA with a solid state nanopore. *Nano Letters* **2005,** 5, (10), 1905-1909.
78. Yan, H.; Xu, B. Q., Towards rapid DNA sequencing: Detecting single-stranded DNA with a solid-state nanopore. *Small* **2006,** 2, (3), 310-312.
79. Muthukumar, M.; Kong, C. Y., Simulation of polymer translocation through protein channels. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** 103, (14), 5273-5278.
80. Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M., Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nature Biotechnology* **2001,** 19, (3), 248-252.
81. Mathe, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A., Nanopore unzipping of individual DNA hairpin molecules. *Biophysical Journal* **2004,** 87, (5), 3205-3212.
82. Meller, A.; Mathe, J.; Viasnoff, V.; Rabin, Y., Nanopore unzipping of individual DNA hairpin molecules. *Abstracts of Papers of the American Chemical Society* **2004,** 228, U301-U301.

83. Healy, K., Nanopore-based single-molecule DNA analysis. *Nanomedicine* **2007,** 2, (4), 459-481.
84. Fologea, D.; Uplinger, J.; Thomas, B.; McNabb, D. S.; Li, J. L., Slowing DNA translocation in a solid-state nanopore. *Nano Letters* **2005,** 5, (9), 1734-1737.
85. Siwy, Z.; Trofin, L.; Kohli, P.; Baker, L. A.; Trautmann, C.; Martin, C. R., Protein biosensors based on biofunctionalized conical gold nanotubes. *Journal of the American Chemical Society* **2005,** 127, (14), 5000-5001.
86. Kohli, P.; Harrell, C. C.; Cao, Z. H.; Gasparac, R.; Tan, W. H.; Martin, C. R., DNA-functionalized nanotube membranes with single-base mismatch selectivity. *Science* **2004,** 305, (5686), 984-986.
87. Wanunu, M.; Sutin, J.; Meller, A., DNA Profiling Using Solid-State Nanopores: Detection of DNA-Binding Molecules. *Nano Letters* **2009,** 9, (10), 3498-3502.
88. McQuarrie, D. A., Statistical Mechanics. *University Science Books* **2000**.
89. Rapaport, D. C., The Art of Molecular Dynamics Simulation. *Cambridge University Press* **1996**.
90. Allen, M. P.; Tildesley, D. J., Computer simulation of liquids. *Clarendon Press* **1989**.
91. Field, M. J., A Practical Introduction to the Simulation of Molecular Systems. *Cambridge University Press* **2007**.
92. Lucas, K., Molecular Models for Fluids. *Cambridge University Press* **2007**.
93. Haile, J. M., Molecular Dynamics Simulation: Elementary Methods. *Wiley Professional* **1997**.
94. http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node23.html.
95. Lakhtakia, A., Handbook of Nanotechnology: Nanometer Structure Theory, Modeling, and Simulation. *SPIE Publications* **2004**.
96. Mark, J. E., Physical Properties of Polymers Handbook. *Springer* **2007**.
97. Tersoff, J., Modeling Solid-State Chemistry - Interatomic Potentials for Multicomponent Systems. *Physical Review B* **1989,** 39, (8), 5566-5568.
98. Boda, D.; Henderson, D., The effects of deviations from Lorentz-Berthelot rules on the properties of a simple mixture. *Molecular Physics* **2008,** 106, (20), 2367-2370.
99. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998,** 102, (18), 3586-3616.
100. Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Protein Simulations* **2003,** 66, 27-+.
101. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *Journal of the American Chemical Society* **1996,** 118, (9), 2309-2309.
102. Haile, J. M., Molecular Dynamics Simulation: Elementary Methods. *Wiley Professional* **1992**.
103. Rapaport, D. C., The Art of Molecular Dynamics Simulation. *Cambridge University Press* **2004**.
104. Kosztin, D.; Gumport, R. I.; Schulten, K., Probing the role of structural water in a duplex oligodeoxyribonucleotide containing a water-mimicking base analog. *Nucleic Acids Research* **1999,** 27, (17), 3550-3556.
105. Sung, W.; Park, P. J., Polymer translocation through a pore in a membrane. *Physical Review Letters* **1996,** 77, (4), 783-786.
106. Lubensky, D. K.; Nelson, D. R., Driven polymer translocation through a narrow pore. *Biophysical Journal* **1999,** 77, (4), 1824-1838.

107. Gnanakaran, S.; Hochstrasser, R. M., Vibrational relaxation of HgI in ethanol: Equilibrium molecular dynamics simulations. *Journal of Chemical Physics* **1996,** 105, (9), 3486-3496.
108. Hofsass, C.; Lindahl, E.; Edholm, O., Molecular dynamics simulations of phospholipid bilayers with cholesterol. *Biophysical Journal* **2003,** 84, (4), 2192-2206.
109. Amisaki, T.; Hakoshima, T.; Tomita, K.; Nishiawa, S.; Uesugi, S.; Ohtsuka, E.; Ikehara, M.; Yoneda, S.; Kitamura, K., Molecular-Dynamics and Free-Energy Perturbation Calculations on the Mutation of Tyrosine-45 to Tryptophan in Ribonuclease-T-1. *Chemical & Pharmaceutical Bulletin* **1992,** 40, (5), 1303-1308.
110. Axelsen, P. H.; Prendergast, F. G., Molecular-Dynamics of Tryptophan in Ribonuclease-T1 .2. Correlations with Fluorescence. *Biophysical Journal* **1989,** 56, (1), 43-66.
111. Sands, Z. A.; Laughton, C. A., Molecular dynamics simulations of DNA using the generalized born solvation model: Quantitative comparisons with explicit solvation results. *Journal of Physical Chemistry B* **2004,** 108, (28), 10113-10119.
112. Stefl, R.; Spackova, N.; Berger, I.; Koca, J.; Sponer, J., Molecular dynamics of DNA quadruplex molecules containing inosine, 6-thioguanine and 6-thiopurine. *Biophysical Journal* **2001,** 80, (1), 455-468.
113. Cross, J. D.; Strychalski, E. A.; Craighead, H. G., Size-dependent DNA mobility in nanochannels. *Journal of Applied Physics* **2007,** 102, (2), 024701-024705
114. Zhao, X. C.; Payne, C. M.; Cummings, P. T.; Lee, J. W., Single-strand DNA molecule translocation through nanoelectrode gaps. *Nanotechnology* **2007,** 18, (42), 424018-424024.
115. Wanunu, M.; Sutin, J.; McNally, B.; Chow, A.; Meller, A., DNA Translocation Governed by Interactions with Solid-State Nanopores. *Biophysical Journal* **2008,** 95, (10), 4716-4725.
116. Chen, P.; Gu, J. J.; Brandin, E.; Kim, Y. R.; Wang, Q.; Branton, D., Probing single DNA molecule transport using fabricated nanopores. *Nano Letters* **2004,** 4, (11), 2293-2298.
117. Fyta, M. G.; Melchionna, S.; Kaxiras, E.; Succi, S., Multiscale coupling of molecular dynamics and hydrodynamics: Application to DNA translocation through a nanopore. *Multiscale Modeling & Simulation* **2006,** 5, (4), 1156-1173.
118. Sauer-Budge, A. F.; Nyamwanda, J. A.; Lubensky, D. K.; Branton, D., Unzipping kinetics of double-stranded DNA in a nanopore. *Physical Review Letters* **2003,** 90, (23), -.
119. Apel, P. Y.; Korchev, Y. E.; Siwy, Z.; Spohr, R.; Yoshida, M., Diode-like single-ion track membrane prepared by electro-stopping. *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms* **2001,** 184, (3), 337-346.
120. Clauset, A.; Shalizi, C. R.; Newman, M. E. J., Power-law distributions in empirical data. *SIAM Review* **2009**.
121. Mannion, J. T.; Reccius, C. H.; Cross, J. D.; Craighead, H. G., Conformational analysis of single DNA molecules undergoing entropically induced motion in nanochannels. *Biophysical Journal* **2006,** 90, (12), 4538-4545.
122. Fologea, D.; Brandin, E.; Uplinger, J.; Branton, D.; Li, J., DNA conformation and base number simultaneously determined in a nanopore. *Electrophoresis* **2007,** 28, (18), 3186-3192.
123. Kasianowicz, J. J.; Nguyen, T. L.; Stanford, V. M., Enhancing molecular flux through nanopores by means of attractive interactions. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** 103, (31), 11431-11432.
124. Hogg, T.; Zhang, M.; Yang, R., Modeling and Analysis of DNA Hybridization Dynamics at Microarray Surface in Moving Fluid. *IEEE International Conference on Robotics and Automation* **2008**, 3419-3424.
125. Chan, V.; Graves, D. J.; McKenzie, S. E., The biophysics of DNA hybridization with immobilized oligonucleotide probes. *Biophysical Journal* **1995,** 69, (6), 2243-2255.
126. Hofler, L.; Gyurcsanyi, R. E., Coarse grained molecular dynamics simulation of electromechanically-gated DNA modified conical nanopores. *Electroanalysis* **2008,** 20, (3), 301-307.
127. Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K., NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics* **1999,** 151, (1), 283-312.

128. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** 14, (1), 33-38.
129. http://www.ks.uiuc.edu/Training/Tutorials/science/bionano/bionano-tutorial-html/.
130. http://beta.rcsb.org/pdb/home/home.do.
131. Catlett, C.; Allcock, W. E.; Andrews, P.; Aydt, R.; Bair, R.; Balac, N.; Banister, B.; Barker, T.; Bartelt, M.; Beckman, P. e. a., TeraGrid: Analysis of Organization, System Architecture, and Middleware Enabling New Types of Applications. HPC and Grids in Action. *High Performance Computing and Grids in Action* **2008,** 16, Edited by: Grandinetti L. Amsterdam: IOS Press.
132. Muller, M.; Katsov, K.; Schick, M., Coarse-grained models and collective phenomena in membranes: Computer simulation of membrane fusion. *Journal of Polymer Science Part B-Polymer Physics* **2003,** 41, (13), 1441-1450.
133. Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L., Coarse grain models and the computer simulation of soft materials. *Journal of Physics-Condensed Matter* **2004,** 16, (15), R481-R512.
134. Marrink, S. J.; de Vries, A. H.; Mark, A. E., Coarse grained model for semiquantitative lipid simulations. *Journal of Physical Chemistry B* **2004,** 108, (2), 750-760.
135. Simons, K.; Ikonen, E., Functional rafts in cell membranes. *Nature* **1997,** 387, (6633), 569-572.
136. Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Moore, P. B.; Klein, M. L., Simulations of phospholipids using a coarse grain model. *Journal of Physical Chemistry B* **2001,** 105, (40), 9785-9792.
137. Sales-Pardo, M.; Guimera, R.; Moreira, A. A.; Widom, J.; Amaral, L. A. N., Mesoscopic modeling for nucleic acid chain dynamics. *Physical Review E* **2005,** 71, (5), -.
138. Voth, G. A., Coarse-Graining of Condensed Phase and Biomolecular Systems. *CRC- 1 edition* **2008**.
139. Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A., The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *Journal of Chemical Physics* **2008,** 128, (24), -.
140. Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K., Coarse grained protein-lipid model with application to lipoprotein particles. *Journal of Physical Chemistry B* **2006,** 110, (8), 3674-3684.
141. Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Sligar, S. G.; Schulten, K., Assembly of lipids and proteins into lipoprotein particles. *Journal of Physical Chemistry B* **2007,** 111, (38), 11095-11104.
142. Shih, A. Y.; Freddolino, P. L.; Arkhipov, A.; Schulten, K., Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations. *Journal of Structural Biology* **2007,** 157, (3), 579-592.
143. Arkhipov, A.; Freddolino, P. L.; Schulten, K., Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **2006,** 14, (12), 1767-1777.
144. Arkhipov, A. S.; Freddolino, P. L.; Schulten, K., Stability and dynamics of virus capsids described by coarse-grained modeling. *Biophysical Journal* **2007**, 540a-540a.
145. Arkhipov, A.; Freddolino, P. L.; Imada, K.; Namba, K.; Schulten, K., Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum. *Biophysical Journal* **2006,** 91, (12), 4589-4597.
146. Yin, Y.; Arkhipov, A.; Schulten, K., Simulations of Membrane Tubulation by Lattices of Amphiphysin N-BAR Domains. *Structure* **2009,** 17, (6), 882-892.
147. Stevens, M. J.; Hoh, J. H.; Woolf, T. B., Insights into the molecular mechanism of membrane fusion from simulation: Evidence for the association of splayed tails. *Physical Review Letters* **2003,** 91, (18), -.
148. Izvekov, S.; Violi, A.; Voth, G. A., Systematic coarse-graining of nanoparticle interactions in molecular dynamics simulation. *Journal of Physical Chemistry B* **2005,** 109, (36), 17019-17024.
149. Izvekov, S.; Voth, G. A., A multiscale coarse-graining method for biomolecular systems. *Journal of Physical Chemistry B* **2005,** 109, (7), 2469-2473.

150. Hansen, J. P.; McDonald, I. R., Theory of Simple Liquids. *Elsevier- 3 edition* **1986**.
151. Liu, P.; Izvekov, S.; Voth, G. A., Multiscale coarse-graining of monosaccharides. *Journal of Physical Chemistry B* **2007,** 111, (39), 11566-11575.
152. Wang, J. Y.; Deng, Y. Q.; Roux, B., Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical Journal* **2006,** 91, (8), 2798-2814.
153. Izvekov, S.; Voth, G. A., Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *Journal of Chemical Theory and Computation* **2006,** 2, (3), 637-648.
154. Liu, P.; Voth, G. A., Smart resolution replica exchange: An efficient algorithm for exploring complex energy landscapes. *Journal of Chemical Physics* **2007,** 126, (4), -.
155. Girifalco, L. A.; Hodak, M.; Lee, R. S., Carbon nanotubes, buckyballs, ropes, and a universal graphitic potential. *Physical Review B* **2000,** 62, (19), 13104-13110.
156. Hone, T. D.; Izvekov, S.; Voth, G. A., Fast centroid molecular dynamics: A force-matching approach for the predetermination of the effective centroid forces. *Journal of Chemical Physics* **2005,** 122, (5), -.
157. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., The MARTINI force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B* **2007,** 111, (27), 7812-7824.
158. Marrink, S. J.; Risselada, J.; Mark, A. E., Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. *Chemistry and Physics of Lipids* **2005,** 135, (2), 223-244.
159. Marrink, S. J.; Mark, A. E., Molecular view of hexagonal phase formation in phospholipid membranes. *Biophysical Journal* **2004,** 87, (6), 3894-3900.
160. Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J., A coarse grain model for DNA. *Journal of Chemical Physics* **2007,** 126, (8), -.
161. Schatz, G. C.; Drukker, K., Model for simulating dynamics of DNA denaturation. *Abstracts of Papers of the American Chemical Society* **2000,** 220, U176-U177.
162. Verlet, L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev* **1967,** 159, (1), 98.
163. Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R., Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Computer Physics Communications* **1995,** 91, (1-3), 43-56.
164. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* **1995,** 103, (19), 8577-8593.
165. Ousterhout, J. K., Tcl and the Tk Toolkit *Addison-Wesley Professional* **1994**.
166. Drukker, K.; Schatz, G. C., A model for simulating dynamics of DNA denaturation. *Journal of Physical Chemistry B* **2000,** 104, (26), 6108-6111.
167. Butler, T. Z.; Pavlenok, M.; Derrington, I. M.; Niederweis, M.; Gundlach, J. H., Single-molecule DNA detection with an engineered MspA protein nanopore. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, (52), 20647-20652.
168. Rong, G.; Weiss, S. M., Biomolecule size-dependent sensitivity of porous silicon sensors. *Physica Status Solidi a-Applications and Materials Science* **2009,** 206, (6), 1365-1368.
169. He, H.; Scheicher, R. H.; Pandey, R.; Rocha, A. R.; Sanvito, S.; Grigoriev, A.; Ahuja, R.; Karna, S. P., Functionalized nanopore-embedded electrodes for rapid DNA sequencing. *Journal of Physical Chemistry C* **2008,** 112, (10), 3456-3459.

BIOGRAPHICAL INFORMATION

Abhijit Ramachandran was born on May 18th, 1985 in Mumbai, India. He has an adoring family. He is the son of Mr. K. N. Ramachandran and Mrs. Susan Chandran, his dad works as an accountant and mom is a high school teacher, they both are the co-founders of R. C. Commerce Pvt. Tuitions. He is thankful to them for allowing him to pursue his interest be it in cultural or scientific aspects; they have always supported and encouraged him to attain the heights. He has a sister, Abhilasha Ramachandran, the most pampered member of the family who has just completed her schooling and is very caring and understands that a Brother is supposed to be a prankster and always will be. He is grateful that the lord has his hand upon him. His friends mean a lot to him and he is blessed to be in the company of good and helpful people.

He completed his primary school and secondary school education from Saint Francis D'Assisi High School, in Mumbai, India. His bachelor's degree was in Biomedical Engineering at Mahatma Gandhi Mission College of Engineering Technology, in Navi Mumbai, India from where he graduated in June 2007. In fall 2007, Abhijit began his graduate studies in Biomedical Engineering at the University of Texas in Arlington (UTA) under the Joint Program of Biomedical Engineering at the University of Texas in Arlington and University of Texas Southwestern Medical Centre at Dallas. Abhijit is currently involved in using Molecular Dynamics as a tool to understand the underlying physics of DNA translocation in Nanopores. This can enable Gene Sequencing at a very low cost and short time with the bio-functionalized Nanopore-based DNA analysis. He has published a journal paper and is in progress of publishing another one. He won the Graduate Deans award for his research poster presentation at the Annual Celebration of Excellence by Students (ACES) symposium in March 2009 at UTA. He was also the elected vice president of Biomedical Engineering Students Society (BMESS) at UTA in 2008-09. He is thankful to the almighty for a blessed life and hopes it continues in the same manner.