Computer simulation of viral-assembly and translocation

Jyoti Prakash Mahalik
University of Massachusetts - Amherst, jmahalik@polysci.umass.edu

Follow this and additional works at: http://scholarworks.umass.edu/open_access_dissertations

Recommended Citation
COMPUTER SIMULATION OF VIRAL-ASSEMBLY AND TRANSLOCATION

A Dissertation Presented
by
JYOTI PRAKASH MAHALIK

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2013

Polymer Science and Engineering
© Copyright by Jyoti Prakash Mahalik 2013
All Rights Reserved
COMPUTER SIMULATION OF VIRAL-ASSEMBLY AND TRANSLOCATION

A Dissertation Presented

by

JYOTI PRAKASH MAHALIK

Approved as to style and content by:

________________________
Murugappan Muthukumar, Chair

________________________
Jon Machta, Member

________________________
Harry Bermudez, Member

________________________
Gregory Grason, Member

________________________
David Hoagland, Department Chair
Polymer Science and Engineering
PhD research often appears a solitary undertaking. However, it is impossible to maintain the degree of focus and dedication required for its completion without the help and support of many people.

First and foremost, I would like to thank Prof. Muthukumar for being my supervisor and teacher. Without his constructive critiques and recommendations this thesis would not be the same. Thank you so much for your unfailing support and guidance while challenging me to move beyond my intellectual comfort zones. Thank you so much for letting me work on so many interesting problems during my PhD.

I would like to thank my committee members, Prof. Machta, Prof. Grason, and Prof. Bermudez for their unfailing support and guidance throughout my PhD. Their encouragement, insightful comments and difficult questions have helped me shape up my thesis as well as my scientific thinking and writing style.

I would also like to thank all the PSE faculties for being such great teachers. I learned a lot from the lectures and the laboratories. I would also like to thank all the PSE staff members, especially Vivien, Lisa, Jessica, and Maria for their kind support throughout my PhD.

I would also like to thank my fellow labmates at Umass Amherst: Chris, Zhaoyang, Ryan, Joanna, Jessica, Clay, Mithun, Byuon, Steve, Arnout, Deniz, Andrew, Jessica, Rajeev, Homin, Harsh, Ben, Hamid, Sveta, Anand, Bo, and Ning for the passionate discussions - work related or not, and for all the fun we had in the last couple of years.

I would also like to thank my friends: Supratim, Ajanta, Ameet, Aishwarya, Naveen, Andy, Vimal, Anurag, Akshaya, Lokesh, and Abhisekh, Andrew, Nick, Scott, Semra, Chelsea, Jeff, Raj, Shilpi, and Jing for their unconditional love and support.
Last but not the least, I would like to thank my parents and my elder sisters for their unconditional love and support in pursuing my goals.
We investigated four different problems using coarse grained computational models: self-assembly of single stranded (ss) DNA virus, ejection dynamics of double stranded (ds) DNA from phages, translocation of ssDNA through MspA protein pore, and segmental dynamics of a polymer translocating through a synthetic nanopore. In the first part of the project, we investigated the self-assembly of a virus with and without its genome. A coarse-grained model was proposed for the viral subunit proteins and its genome (ssDNA). Langevin dynamics simulation, and replica exchange method were used to determine the kinetics and energetics of the self-assembly process, respectively. The self-assembly follows a nucleation-growth kind of mechanism. The ssDNA plays a crucial role in the self-assembly by acting as a template and enhancing the local concentration of the subunits. The presence of the genome does not
changes the mechanism of the self-assembly but it reduces the nucleation time and enhances the growth rate by almost an order of magnitude. The second part of the project involves the investigation of the dynamics of the ejection of dsDNA from phages. A coarse-grained model was used for the phage and dsDNA. Langevin dynamics simulation was used to investigate the kinetics of the ejection. The ejection is a stochastic process and a slow intermediate rate kinetics was observed for most ejection trajectories. We discovered that the jamming of the DNA at the pore mouth at high packing fraction and for a disordered system is the reason for the intermediate slow kinetics. The third part of the project involves translocation of ssDNA through MspA protein pore. MspA protein pore has the potential for genome sequencing because of its ability to clearly distinguish the four different nucleotides based on their blockade current, but it is a challenge to use this pore for any practical application because of the very fast traslocation time. We resolved the state of DNA translocation reported in the experimental work [1]. We also investigated two methods for slowing down the translocation process: pore mutation and use of alternating voltage. Langevin dynamics simulation and Poisson Nernst Planck solver were used for the investigation. We demonstrated that mutation of the protein pore or applying alternating voltage is not a perfect solution for increasing translocation time deterministically. Both strategies resulted in enhanced average translocation time as well as the width of the translocation time distribution. The increase in the width of the translocation time distribution is undesired. In the last part of the project, we investigated the applicability of the polyelectrolyte theory in the computer simulation of polyelectrolyte translocation through nanopores. We determined that the Debye Hückel approximation is acceptable for most translocation simulations as long as the coarse grained polymer bead size is comparable or larger than the Debye length. We also determined that the equilibrium translocation theory is applicable to the polyelectrolyte translocation through a nanopore under biasing condition. The unbiased
translocation behavior of a polyelectrolyte chain is qualitatively different from the Rouse model predictions, except for the case where the polyelectrolyte is very small compared to the nanopore.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACKNOWLEDGEMENTS</strong></td>
<td>iv</td>
</tr>
<tr>
<td><strong>ABSTRACT</strong></td>
<td>vi</td>
</tr>
<tr>
<td><strong>LIST OF TABLES</strong></td>
<td>xii</td>
</tr>
<tr>
<td><strong>LIST OF FIGURES</strong></td>
<td>xiii</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td><strong>1. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>2. SELF-ASSEMBLY OF VIRUS</strong></td>
<td>5</td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2.2.1 Mechanism of self-assembly of virus like particles (VLP)</td>
<td>6</td>
</tr>
<tr>
<td>2.2.2 Mechanism of self-assembly of virus</td>
<td>9</td>
</tr>
<tr>
<td>2.2.3 Modeling strategy</td>
<td>11</td>
</tr>
<tr>
<td>2.3 Method</td>
<td>11</td>
</tr>
<tr>
<td>2.3.1 Modeling Protein Subunit</td>
<td>12</td>
</tr>
<tr>
<td>2.3.2 Modeling Polymer</td>
<td>13</td>
</tr>
<tr>
<td>2.3.3 Simulation details</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Results and Discussion</td>
<td>20</td>
</tr>
<tr>
<td>2.4.1 Polymer-free assembly</td>
<td>20</td>
</tr>
<tr>
<td>2.4.2 Polymer-assisted assembly</td>
<td>32</td>
</tr>
<tr>
<td>2.5 Conclusions</td>
<td>39</td>
</tr>
</tbody>
</table>
3. EJECTION DYNAMICS OF DOUBLE STRANDED DNA FROM PHAGE

3.1 Abstract ................................................................................. 42
3.2 Introduction ........................................................................... 43
3.3 Method .................................................................................. 46
  3.3.1 Modeling Capsid .............................................................. 46
  3.3.2 Modeling Motor Protein ................................................... 47
  3.3.3 Modeling Double-stranded DNA/Small persistence length polymer ........................................... 47
  3.3.4 Simulation Technique ...................................................... 48
  3.3.5 Initial Configuration ........................................................ 50
  3.3.6 Persistence length determination ..................................... 51
  3.3.7 DNA Packing ................................................................. 52
  3.3.8 DNA Ejection ................................................................. 52
  3.3.9 Parameter evaluation ....................................................... 53
3.4 Results and Discussion ........................................................... 55
  3.4.1 Packing dynamics ............................................................ 55
  3.4.2 Ejection Dynamics ........................................................... 60
  3.4.3 Hysteresis .................................................................... 66
3.5 Conclusions ........................................................................... 71

4. TRANSLOCATION OF SINGLE STRANDED DNA THROUGH MSPA PROTEIN PORE

4.1 Abstract ................................................................................ 73
4.2 Background .......................................................................... 74
4.3 Method ................................................................................ 78
  4.3.1 Modeling Protein Pore ..................................................... 78
  4.3.2 Modeling Membrane ........................................................ 79
  4.3.3 Modeling ssDNA ............................................................. 79
  4.3.4 Simulation Summary ....................................................... 81
    4.3.4.1 Poisson-Nernst-Planck equation .................................. 82
    4.3.4.2 Langevin dynamics simulation details ......................... 84
  4.3.5 Obtaining Electric field from the Voltage ......................... 86
  4.3.6 Alternating voltage Application ........................................ 87
  4.3.7 Dielectric Mismatch Modeling .......................................... 88
  4.3.8 Initial Configuration ....................................................... 88
  4.3.9 Definitions .................................................................... 89
4.3.10 Free Energy computation and Radial density measurement........... 89

4.4 Results and Discussion .............................................. 90
  4.4.1 Voltage drop across the pore ..................................... 90
  4.4.2 M1MspA vs. M2MspA ............................................. 91
  4.4.3 Effect of the size of DNA and voltage on translocation
       through M1MspA .................................................. 92
  4.4.4 M3MspA .......................................................... 93
  4.4.5 Effect of Alternating Voltage .................................... 99

4.5 Conclusions ........................................................... 102

5. SEGMENTAL DYNAMICS OF A POLYMER
   TRANSLOCATING THROUGH A SYNTHETIC
   NANOPORE .......................................................... 105

  5.1 Abstract ............................................................. 105
  5.2 Background .......................................................... 106
  5.3 Method ............................................................... 108
     5.3.1 Poisson-Boltzmann Equation (PBE) .............................. 108
     5.3.2 Langevin dynamics simulation details ......................... 111
  5.4 Results and Discussion .............................................. 115
     5.4.1 Poisson-Boltzmann equation .................................... 115
     5.4.2 Rouse Model Verification ...................................... 116
        5.4.2.1 Neutral chain in free solution ........................... 117
        5.4.2.2 Charged Chain in free solution ........................... 119
        5.4.2.3 Charged Chain in a long cylindrical confinement ....... 119
     5.4.3 Charged Chain in a small cylindrical confinement:
        Unbiased ......................................................... 119
     5.4.4 Charged chain in a small cylindrical confinement: External
        bias .............................................................. 122
  5.5 Conclusions ........................................................... 126

6. CONCLUSIONS ............................................................ 129

BIBLIOGRAPHY ............................................................. 132
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Coarse grain radius and charge of the backbone and side chains of MspA pore</td>
<td>81</td>
</tr>
<tr>
<td>4.2 Dwell time and Translocation time for dT50 at 180 mV through the three MspA pores</td>
<td>96</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure | Page
--- | ---
2.1 Lifecycle of a single stranded genome virus[2] | 6
2.2 Wireframe representation of the trimer subunit of MVM with purple patches denoting the sticky subdomains as established by Reguera et al [3] | 7
2.3 Schematic of self-assembly pathway for VLP formation in Literature | 9
2.4 Schematic of self-assembly pathway for Virus formation in Literature | 11
2.5 Stable assembly of the icosahedron occurs only at intermediate reduced temperatures. (a) Starting configuration with 20 subunits; (b)-(d) are snapshots at $t = 10^5, 5 \times 10^5$ and $1.75 \times 10^6$, respectively at $T^* = 0.5$; (e) monster-like particle at $t = 1.75 \times 10^6$ and $T^* = 0.45$; (f) no assembly at $t = 1.75 \times 10^6$ and $T^* = 0.6$. | 22
2.6 Starting configuration with 100 subunits ($c=0.5188 \text{ mM}$); (b)-(d) final structures at $T^*=0.45$, 0.5, and 0.6, respectively, at $t=1.75 \times 10^6$. | 23
2.7 Time-evolution of the distribution functions of mass fraction of k-mers in the population. | 24
2.8 Growth kinetics of average aggregate size $n_{av}$. (a) Effect of temperature with $c=0.5188 \text{ mM}$ (no polymer). (b) Effect of subunit concentration at $T^*=0.476$ (no polymer). (c) Effect of the polymer length at $T^*=0.526$ and $c=0.5188 \text{ mM}$. | 26
2.9 Free energy as a function of aggregate size. (a) $T^*=0.5266$ and (b) $T^*=0.631$. | 27
2.10 Average time $\tau_k$ for the formation of k-mer from (k-1)-mer at $c=0.5188 \text{ mM}$. The reduced temperatures are (a)0.476, (b)0.5, (c) 0.526, and (d) 0.541. | 28
2.11 Nucleation time follows classical nucleation theory. (a) Plot of ln($\tau_{nuc}$) against 1/(Tm*-T*)T* is linear with the best fit value Tm*=0.581 (and regression coefficient=0.9989). (b) Plot of ln($\tau_{nuc}$) against 1/ln(c/cm) is linear with the best fit value cm=0.001 mM (and regression coefficient =0.9982).

2.12 Dependence of linear growth rate on temperature and subunit concentration. (a) Adsorption/desorption model is not obeyed as indicated by the poor fir with Eq. (2.14); the line is with $k_c=0.1$ and regression coefficient =0.8094. (b) The linear fir (with TM*=0.581 and regression coefficient =0.9997) for ln(G2) vs. 1/(Tm*-T*)T* supports the secondary nucleation model. (c) G2 is linear in subunit concentration.

2.13 Snapshots of assembly in the presence of the polymer (T*=0.5266, c=0.5188 mM, N=130).

2.14 Snapshot of the polymer after co-assembly (N=130). (b) Radially averaged monomer density distribution (N=130). (c) The chain spills out of the icosahedron for large values of N (N=200).

2.15 Plot of ln($\tau_{nuc}$) vs. 1/ln((c+αN)/cm) is linear supporting the nucleation mechanism in the presence of polymer. (b) Linear growth rate is approximately linear with the polymer length. In both (a) and (b) the reduced temperatures are T*=0.556 (circles), 0.526 (squares), and 0.5 (triangles).

3.1 Schematic of the steps involved in the viral ejection during in-vitro experiments.


3.3 Coarse grain model (a) Capsid (b) Capsid with dsDNA (c) close-up of initial configuration of DNA near the capsid mouth.

3.4 Persistence length as a function of $K_{angle}$

3.5 Packaging versus time (a) Effect of different motor force (b) Effect of persistence length of dsDNA

3.6 Pressure buildup during packing for different motor force (Average value). The total length packed is shown in figure 3.5.
3.7 Contribution of different component of energy to the packing pressure (Average value) ........................................... 59

3.8 Projected bead density on the xy plane .................................. 61

3.9 Radial probability density of dsDNA in capsid for different motor force ......................................................... 61

3.10 Axial order of the dsDNA in capsid for different motor force ............... 62

3.11 DNA arrangement within the capsid after complete packing (after packing timestep of about 2750 LJ Units) (a) 5 pN : Better axial order (b) 55 pN: Random order. .................................................. 62

3.12 Individual rate kinetics of ejection for the following packed conditions: (a) Force 55 pN, PL 60 nm (b) Force 55 pN, PL 20 nm (c) Force 20 pN, PL 60 nm (d) Force 5 pN, PL 60 nm. .............. 63

3.13 Trajectory of three beads along the chain during ejection (lower value of bead represents the bead that comes out last) (a) Persistence length = 60 nm (b) Persistence length = 20 nm. .................. 67

3.14 Displacement of three beads along the chain during ejection (lower value of bead represents the bead that comes out last) (a) Persistence length = 60 nm (b) Persistence length = 20 nm. ............... 68

3.15 (a) The angle between the chain near the mouth with the pore axis (b) corresponding ejection kinetics (corresponding colors) ............... 69

3.16 An example of hysteresis in the packaging-ejection cycle for dsDNA (55 pN) ................................................................. 70

3.17 Hysteresis for different motor force and persistence length ............... 71

4.1 Hairpin experimental results [7] demonstrating the distinct blockade current for different bases with MspA. ........................................... 75

4.2 MspA structures and mutants: Original protein structure belongs to WTMspA, M1MspA is engineered from WTMspA by replacing the negatively charged residues (shown in red) to neutral residues (shown in grey) at the narrow end of the pore. M2MspA is created from M1MspA by replacing the negatively charged residues by positive residues (shown in blue). Replacement of residues are done by site directed mutagenesis. [1] .................. 77
5.1 2D representation of the schematic of the system for PB calculation. .......................................................... 110

5.2 solution of the NPBE, PBE, and DH equation for different position of the fixed point charge (1e) along the pore, blue color symbol represents NPBE solution, red color symbol represents linear PBE solution, and green color symbol represents the solution of the DH equation (a) Outside the confinement (x=4, y=0, z=16), (b) Inside the confinement (x=4, y=0, z=0). ........................................ 116

5.3 (a)Flory exponent determination for the neutral chain (dashed green line represents the linear fit of average value of log($R_g$) versus log($N$), a slope of 0.622 is obtained, (b) MSD of center of mass of chain versus time: the slope determines the diffusivity of the polymer chain, four different values of N were used (30, 50, 70, and 100) (c) Plotting log(D) versus log(N) to verify Rouse model: A slope of -0.98 is obtained which is very close to the expected slope of -1. Averaged over 500 independent runs. ....................... 117

5.4 MSD versus time for the central monomer for (a)N=30: a slope of 0.66 is observed below $\tau_R$ and a slope of 0.99 is observed above $\tau_R$, $\tau_R$ is about 100 LJ time units. (b)N=100: a slope of 0.55 (expected slope=0.55) is observed below $\tau_R$ and a slope of 0.97 (expected slope=1) is observed above $\tau_R$, $\tau_R$ is somewhere between 100 to 1000 LJ time units. Averaged over 500 independent runs. ................................. 119

5.5 (a)Flory exponent determination for the charged chain (dashed green line represents the linear fit of average value of log($R_g$) versus log($N$), a slope of 0.622 is obtained, (b) MSD of center of mass of chain versus time: the slope determines the diffusivity of the polymer chain, four different values of N were used (30, 50, 70, and 100), (c) Plotting log(D) versus log(N) to verify Rouse model: A slope of -0.96 is obtained which is very close to the expected slope of -1 (d) MSD versus time for the central monomer for chain length of 100: a slope of 0.55 (expected slope=0.56) is observed below $\tau_R$ and a slope of 0.97 (expected slope=1) is observed above $\tau_R$, $\tau_R$ is somewhere between 100 to 1000 LJ time units. Averaged over 500 independent runs. ........................................ 120
5.6 (a) Flory exponent determination for the charged chain confined in a long nanopore (dashed green line represents the linear fit of average value of $\log(R_g)$ versus $\log(N)$, a slope of 1.0 is obtained, (b) MSD of center of mass of chain versus time: the slope determines the diffusivity of the polymer chain, four different values of $N$ were used (30, 50, 70, and 100), (c) Plotting $\log(D)$ versus $\log(N)$ to verify Rouse model: A slope of -1.06 is obtained which is very close to the expected slope of -1 (d) MSD versus time for the central monomer for chain length of 100: a slope of 0.70 (expected slope=0.67) is observed below $\tau_R$ and a slope of 1.0 (expected slope=1) is observed above $\tau_R$, $\tau_R$ is about 500 LJ time units. Averaged over 50 independent runs.

5.7 MSD of the central bead as a function of time for different $N$ values (20, 40, 60, 100, 220) (averaged over 150 independent runs).

5.8 MSD of the central bead of a polymer chain ($N=20$) as a function of time for different applied voltage (averaged over 150 independent runs).

5.9 MSD of the central bead of a polymer chain ($N=220$) as a function of time for different applied voltage (averaged over 150 independent runs).

5.10 Snapshots at different stages of a long polymer chain ($N=220$) undergoing translocation through a synthetic nanopore ($L=20$).
CHAPTER 1
INTRODUCTION

Polypelectrolytes are important for many processes in biology and biomedicine. In this thesis we investigate the interaction of the polyelectrolytes with viral capsids proteins and nanopores.

The first study will investigate how the polyelectrolytes interact with the protein subunits during the self-assembly of the virus. Understanding the mechanism of the self-assembly of the virus with or without the genome is important in optimal design of antivirals[8, 9] and novel drug delivery vehicles[10, 11]. In order to understand the mechanism of the self-assembly of virus, we have constructed a coarse grained model for parovirus Minute virus of Mice (MVM) and the polyelectrolyte. The kinetics of the self-assembly of the virus was investigated using Langevin dynamics simulation and the thermodynamics of the self-assembly pathway was investigated using replica exchange method. Proper assembly (formation of closed capsid structure) happens in a narrow interaction potential range. Low interaction strength (between subunits) does not allows the formation of any stable assembled structure whereas high interaction strength leads to the formation of irregular structures. We have investigated the mechanism of the self-assembly of a virus in this narrow interaction range. By investigating the effect of the temperature and the concentration on the kinetics of the self-assembly of the viral subunits, we have determined that the kinetics of self-assembly of virus like particle (without genome) is a nucleation-growth process. The growth process happens in two steps, initial linear growth regime followed by a slow growth regime. The calculated average nucleation time obeys the laws expected from
the classical nucleation theory. The linear growth rate is found to obey the laws of secondary nucleation as in the case of lamellar growth in polymer crystallization. The same mechanism is seen in the simulations of the assembly of capsid in the presence of the polyelectrolyte. The polyelectrolyte reduces the nucleation barrier significantly by enhancing the local concentration of subunits via adsorbing them on their backbone. The details of growth in the presence of the polyelectrolyte are also found to be consistent with the classical nucleation theory, despite the smallness of the assembled structures.

The second study will investigate the dynamics of the ejection of a double stranded DNA from phages. The kinetics of the DNA ejection from phages have been reported in the literature to be a stochastic process, with most of the ejection happening with an intermediate pause [5]. The intermediate pause was reported to happen anywhere within 10 % and 60 % ejection fraction. Some of the reasons speculated in the literature for the intermediate pause are: bad quality solvent induced jamming [12], single site interruptions [4], and local phase transitions and defects underpinning at high packing fractions [5]. Langevin dynamics simulation was performed on a coarse grained model of dsDNA and phi29 phage to investigate the ejection dynamics of the dsDNA from the phage. Our simulation results show significant variations in the local ejection speed, consistent with experimental observations reported in the literature for both in-vivo and in-vitro systems. In efforts to understand the origin of such variations in the local speed of ejection, we have investigated the correlations between the local ejection kinetics and the packaged structures created at various motor forces and chain flexibility. At lower motor forces, the packaged DNA length is shorter with better organization. On the other hand, at higher motor forces typical of realistic situations, the DNA organization inside the capsid suffers from significant orientational disorder, but yet with long orientational correlation times. This in turn leads to lack of registry between the direction of the DNA segments just to be ejected.
and the direction of exit. As a result, a significant amount of momentum transfer is required locally for successful exit. Consequently, the DNA ejection temporarily slows down exhibiting pauses. This slowing down occurs at random times during the ejection process, completely determined by the particular starting conformation created by prescribed motor forces. In order to augment our inference, we have additionally investigated the ejection of chains with deliberately changed persistence length. For less inflexible chains, the demand on the occurrence of large momentum transfer for successful ejection is weaker resulting in more uniform ejection kinetics.

The third study will investigate the translocation of single stranded DNA through MspA protein pore. Currently MspA is the best available pore for sequencing because of its inherent ability to clearly distinguish between the four different nucleotides [1]. But the DNA translocation through MspA is too fast, hence the position of the DNA is unresolved after the blockade current experiments [1] (whether the DNA is successfully translocating or it is bouncing back to the cis chamber). We have demonstrated that the translocation events are successful most of the time. Secondly, several methods have been suggested in the literature to slow down the translocation process. We investigate two such methods: mutation of the protein pore and the use of alternating voltage. We used Poisson Nernst Planck solver and Langevin dynamics simulation on a coarse grained model of MspA and DNA for our investigation. Use of both the techniques, mutation as well as the alternating voltage approach resulted in an enhanced average translocation time, but the translocation time distribution became wider. For genome sequencing a deterministic process is required so that each nucleotide along the DNA strand can be detected only once, diffusion may lead to erroneous sequencing. Recent effort in the literature has been directed towards using polymerase [13] to slow down the translocation process as well as to make the translocation process deterministic.
The fourth study will investigate the validity of the polyelectrolyte theory in the translocation simulation of polyelectrolyte through nanopores. The first part of the project investigates the validity of the mean field Debye Hückel approximation in the confinement of nanopores and the second part of the project investigates the validity of the Rouse model for a polyelectrolyte chain translocating in an unbiased/biased condition through a nanopore. For the first part of the project, the potential drop around a point charge (across various location along the nanopore) was evaluated by three different techniques: analytical calculation by using Debye Hückel potential equation, numerically solving linearized Poisson Boltzmann equation, and numerically solving the non-linear Poisson Boltzmann equation (NPBE). The NPBE solution is most accurate but the computation time is high. By comparing the potential evaluated from the three different techniques, we demonstrated that the Debye Hückel approximation is applicable as long as the coarse grain bead is larger than the Debye length, at least for the case of 1 M KCl solution (typical translocation experiments are performed at this salt concentration). For the second part of the project, effect of the voltage and the size of the polyelectrolyte chain on the segmental dynamics of a translocating polyelectrolyte chain (through a nanopore) was determined. Langevin dynamics simulation was used to determine the segmental dynamics of the polyelectrolyte chain undergoing translocation through a nanopore. The qualitative behavior of the dynamics of the segments was compared with the prediction of the Rouse model. We demonstrated that the qualitative behavior of the segmental dynamics agrees well with the Rouse model when the polyelectrolyte is translocating under a biased field. The Rouse model fails to predict the segmental dynamics of a polyelectrolyte chain undergoing translocation in unbiased condition, except for the case when the polyelectrolyte is much smaller than the nanopore in terms of length.
CHAPTER 2

SELF-ASSEMBLY OF VIRUS

2.1 Abstract

We investigated the polymer-assisted virus-like assembly using Langevin dynamics simulation and Replica exchange method. Coarse-grained models were constructed for the minute virus of mice (MVM) and the polyelectrolyte chain. Effect of the temperature and the concentration on the kinetics of the self-assembly of virus like particle (without polyelectrolyte) was investigated. Closed assembled structure is formed only in a narrow range of temperature. The kinetics/energetics of the self-assembly of virus was investigated in this narrow temperature range. After investigating the effects of temperature and concentration on the self assembly of virus in this narrow temperature range, we found that the kinetics of the self-assembly of virus follows three regimes: nucleation, linear fast growth, and slow growth. The nucleation and linear growth regime follows classical nucleation-growth theory of polymer crystallization. The presence of a polyelectrolyte does not changes the mechanism of self-assembly but it reduces the nucleation time and enhances the growth rate by almost an order of magnitude by adsorbing the subunits to its backbone leading to localized increase in the subunit concentration.

2.2 Introduction

Viruses are the simplest organism consisting of a genome protected by a protein layer and, optionally, a membrane outside of the protein layer. A virus replicates by hijacking the machinery of the host cell as shown in figure 2.1. The limited size of the
viral genome does not allow it to store a lot of information. The viral genome contains information for only a part of its capsid (subunit) besides other functionalities. Once the virus hijacks a host cell, it regulates the host cell machinery to produce multiple copies of its protein subunits from the information stored in its genome. These protein subunits (usually multiples of 60) self-assemble to form a complete viral unit which is an exact replica of the original virus. We will investigate the mechanism of the self-assembly of the virus. We will determine the conditions at which the subunits self-assemble into a replica of the original virus and the conditions in which they do not form the replica. We will also investigate the role played by the genome during the self-assembly process.

![Figure 2.1. Lifecycle of a single stranded genome virus][2]

### 2.2.1 Mechanism of self-assembly of virus like particles (VLP)

The self-assembly of the virus inside the host cell is a very complex process as it involves a lot of components of the host cell as well as the assembling component of the virus. Without going into the complexity of an in-vivo process we will fo-
Figure 2.2. Wireframe representation of the trimer subunit of MVM with purple patches denoting the sticky subdomains as established by Reguera et al [3]

(b) Side-view of the truncated prism used as a coarse-grained model in our simulations. (c)-(f) are the top-views of the first, second, third, and the fourth layers in (b). The white, green, and red colored beads denote the sticky hydrophobic, repulsive excluded volume, and positively charged subdomains, respectively. (g) Coarse-grained united atom model of the flexible polyelectrolyte chain.
Focus on the self-assembly of the virus like particle (without genome) and the virus (with genome) in an in-vitro environment. We started our investigation with a very simple system: in-vitro self assembly of VLP. It has been demonstrated that the self-assembly of the virus is possible without the genome but the solution conditions were far different from the physiological conditions\cite{8, 14, 15, 16} namely high salt concentration or low pH condition. In one of the investigation\cite{17} it was demonstrated that the self-assembled icosahedral virus like particles are formed under specific solution conditions. At other conditions larger aggregate structures ("monster-like-particle") were formed. These experiments demonstrated that the self-assembly pathway can be perturbed by altering the interaction between the protein subunits by changing the solution conditions. In order to determine the exact nature of the interaction, individual amino acids have been targeted. In these experiments, site directed mutagenesis was performed, whereby a protein was synthesized with one or more altered residues compared to the wild type protein. The newly synthesized protein was then allowed to associate. The newly synthesized protein subunits may or may not self-assemble depending on whether the mutated residues are critical for the self-assembly. Even if the self-assembly happens, the formed capsid may or may not be stable at a slightly higher temperature depending on the criticality of the mutated residue. One such experiment was performed on paraovirus, minute virus of mice (MVM)\cite{3}. It is the simplest of the icosahedral virus family. It has a T1 symmetry with 60 identical protein subunits making the whole capsid. More complex viruses are made up of 60*3, 60*5, ... number of protein subunits. The subunits for MVM are known to exist in a stable trimeric form in the solution. It was determined that the residues responsible for the self-assembly and stability of the virus lie regularly on the edges of the protein trimeric subunit, as shown in figure 2.2a and those critical residues are mostly hydrophobic in nature. Other viral systems are also known where the interaction between the subunits are hydrophobically dominated, hence disrupting the self-assembly
was attempted with highly hydrophobic compounds or other small compounds which alter the interaction potential between the hydrophobic domains. These options are being explored for potential antiviral application[8, 9]. Although experimental trials have been done on disrupting the virus self-assembly, but a full understanding of the kinetics is lacking. Understanding the kinetics of the self-assembly pathway will be helpful in optimal design of the antiviral load. There is conflicting evidence of the kinetics of the self-assembly kinetics in the literature. It is contended whether the kinetics is nucleation-growth [8, 14] or spontaneous process [18] (figure 2.3). We will investigate the kinetics of the self-assembly process by Langevin dynamics simulation of coarse grained models.

Figure 2.3. Schematic of self-assembly pathway for VLP formation in Literature

2.2.2 Mechanism of self-assembly of virus

Next, the role played by the genome in the self-assembly process was investigated. It has been demonstrated by X-ray crystallography[19] and cryo-electron experiments[20, 21] that the single stranded genome positions itself near the capsid walls within the capsid, and the symmetry of the capsid is partially transmitted to the
genome. Due to the close association of the negatively charged phosphate groups of genome and mostly positive interior surface of the capsid it was speculated that non-specific electrostatic interaction between the negatively charged phosphate groups on the genome and the positively charged arginine and lysine on the interior of the capsid protein drives the association of the capsid proteins and ss genome. Further evidence in the literature suggests non-specific electrostatic interaction between the genome-subunit. It has been demonstrated that the genome can be substituted with synthetic polyelectrolytes[10, 11] or positively charged nanoparticles[22]. These experiments indicate that the interaction between the protein capsid and the genome may not be dependent on the specific sequence of the genome, but predominantly non-specific electrostatic interaction. Although the synthetic polyelectrolytes are somewhat similar to the ss genome but the ss genome is known to form very compact structure due the intra-strand secondary interactions. Based on these experiments we propose a pseudo coarse-grain model for the ssDNA. Only non-specific electrostatic interactions will be considered in our model. The secondary interaction between the ssDNA strands are beyond the scope of the current model and is being considered for the future work. The ss genome model most likely resembles a synthetic polyelectrolyte.

The mechanism of the self-assembly of a virus has been speculated by several groups in the literature. We will present a schematic of the speculated pathways of the self-assembly of virus in figure 2.4. The first mechanism is proposed by Bruinsma [23] and Hu et al [10], the second mechanism is proposed by Hu et al[24] and the third mechanism is proposed by Freddolino et al[25]. All the mechanisms are very different and are merely speculations. One of the goal of this research is to demonstrate the mechanism of self-assembly of a virus.
2.2.3 Modeling strategy

Modeling and theoretical attempts have been reported in the literature on the kinetics of the self-assembly of the virus. All atomic model molecular dynamic simulation has been attempted on satellite tobacco mosaic virus[25, 26]. But the large number of atoms involved in all atomic simulation has limited the simulation time for tens of nanoseconds on a supercomputer; assembly kinetics studies with all atomic model is beyond the computational capability of powerful supercomputers within reasonable time. Hence a lot of attempts have been made to build coarse-grain models for the protein subunits[27, 28, 18, 29, 30]. Our coarse-grained model is based on the experiment[3] but is influenced by one of the models[18].

2.3 Method

Coarse-grained models were built for the protein subunits and the polyelectrolyte. Although the coarse-grain model for the protein subunit was built in the context of
the parovirus minute virus of mice (MVM), due to the experimentally available rich details [3], but the features are generic for non-enveloped T1 viruses. The kinetics of the self-assembly pathway was monitored using Langevin dynamics simulation and the free energy landscape of the self-assembly pathway was monitored using parallel tempering method. The interaction potentials, computational procedures, and methods of data analysis are briefly given below.

2.3.1 Modeling Protein Subunit

The coarse-grained model of the subunit is based on the extensive mutagenesis investigation on parovirus MVM by Reguera et al [3]. They showed that the significant intermediate is a trimer of the proteins which geometrically resembles a thick equilateral triangle as shown in figure 2.2a. By performing site directed mutagenesis they established that some residues are critical for the self-assembly and stability of the capsid. Those residues are capable of forming multiple hydrophobic contacts and/or hydrogen bonds and/or salt bridges with the same set of residues on the other trimers and they lie regularly on a thin equatorial belt around the trimer. Based on the realistic description of the trimer by Reguera et al. [3] figure 2.2a, we built a coarse-grained model as depicted in figure 2.2b. In constructing our model, we took into consideration several aspects of the virus assembly. First, there are a few attractive groups regularly placed on the perimeter of a triangular frame and there is asymmetric excluded volume interactions to allow curved assembly towards the interior of the virus. Finally, there must be positively charged domains on the interior side of the trimer. There are four layers of beads in our model. The edge length of the subunit is roughly the same as the real protein trimeric subunit (10 nm). The middle two layers consists of the excluded volume beads and the sticky hydrophobic beads at regular intervals. The sticky groups (white colored beads in figure 2.2b) lying at regular intervals on the second and third layer provide the glue for self-assembly.
The positions of the white beads on the second and third layers are chosen to closely resemble the experimentally observed locations of the critical residues required for self-assembly. In order to bring about the required wedge from the trimer building blocks in the formation of final icosahedral structure, the third layer is placed at an angle of 69° uniformly from all edges of the second layer. These two layers are necessary to form the icosahedra, but two more layers are created in order to prevent the stacking of the subunits on top of each other. The first layer is made up of excluded volume beads only, which is necessary to prevent the inter-subunit stacking. The fourth layer consists of positively charged beads which serves dual purpose of preventing the inter-subunit stacking as well as attaching to the negatively charged polymer. The positively charged beads are assigned +1e charge, where e represents the electronic charge. This top and bottom layers allow only sidewise attraction between the subunits. The whole construction is then taken as a rigid body in the simulations. It must be remarked that the composition of the above coarse-graining procedure is motivated by the need to explore the physical phenomena associated with the co-assembly kinetics of interacting proteins and polyelectrolytes and not to focus on the specific details on individual viruses. In fact, the crucial entropic roles [31] played by the cationic loops and tails emanating from the interior wall of capsids in real viruses are suppressed and simply represented by a certain net positive charge on the interior surface. Similarly all conformational fluctuations of various units in the virus are ignored. As a result, explicit accounting of angular and bond potentials are unnecessary facilitating the monitoring of assembly kinetics at very long times pertinent to experiments.

2.3.2 Modeling Polymer

The single stranded DNA is represented in the form of a bead spring model. The diameter of the coarse-grained bead is taken as 1 nm, which is same as the persistence
length of single-stranded polynucleic acid or polystyrene sulfonate at physiological conditions. By choosing such coarse grained diameter we are relieved from considering angular and dihedral potentials. Each coarse grained bead roughly consists of three bases, hence a charge of -3e is assigned to a coarse grained bead. In reality this charge could be much lower due to counterion adsorption. Furthermore, the genome could readily assume secondary structures [32] which in turn might affect the assembly kinetics. In our model, this feature is not included. Our results are perhaps more relevant to simpler polymers such as polystyrene sulfonate. The interactions between the united atoms and the forces for their connectivity into a chain are given below.

2.3.3 Simulation details

Langevin dynamics simulation was used for computing the trajectory and velocity of the beads, the computational procedure of which is well documented in the literature[33, 34]. Briefly, the trajectory of the ith bead in the system is computed by using the Langevin equation for the j component of the position vector of the ith bead (rij),

\[ m_i \frac{d^2 r_{ij}}{dt^2} = -\zeta_i \frac{dr_{ij}}{dt} - \nabla_j U_i + f_{ij} \]  

(2.1)

where \( t \) is the time. \( m_i \) and \( \zeta_i \) are, respectively, the mass and friction coefficient of the ith bead. \( U_i \) is the net potential acting on the ith bead, as given below. \( f_{ij} \) is the j component of the random force acting on the ith bead obeying the fluctuation-dissipation theorem with its magnitude given by \( \sqrt{k_B T \zeta_i / dt} \) (\( k_B T \) being the Boltzmann constant times the absolute temperature). There are three kinds of non-bonded potentials acting on each bead, namely the repulsive excluded volume potential, short-ranged Lennard-Jones-like potential, and the screened electrostatic interaction potential. In addition, for the polymer chain, a harmonic bead-spring
potential is used to account for the chain connectivity. Both the excluded volume
and short-ranged attractive interaction potentials are modeled as

\[ U_{COMPASS} = \epsilon \left[ 2 \left( \frac{\sigma}{r} \right)^9 - 3 \left( \frac{\sigma}{r} \right)^6 \right] + \epsilon_c, r < r_c \quad (2.2) \]

where

\[ \epsilon_c = -\epsilon \left[ 2 \left( \frac{\sigma}{r_c} \right)^9 - 3 \left( \frac{\sigma}{r_c} \right)^6 \right] \quad (2.3) \]

The range \( \sigma \) and strength \( \epsilon \) are parameters. We have chosen the cut-off \( r_c \) to be
\( \sigma \) for the repulsive excluded volume potential, and \( r_c = 2.5\sigma \) for the short-ranged
attractive hydrophobic interactions. The pairwise electrostatic interaction potential
between the \( i \)th bead of charge \( q_i \) and the \( j \)th bead of charge \( q_j \) separated by a distance
\( r \) is assumed to be the Debye-Hückel potential,

\[ U_{DH} = \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r} \exp\left(\frac{-\kappa r}{r}\right) \quad (2.4) \]

where \( \varepsilon_0 \) is the permittivity of the vacuum, \( \varepsilon_r \) is the dielectric constant of the
solution, and \( \kappa \) is the inverse Debye length. The connectivity of any two adjacent
beads of the polymer is taken to be Hookean,

\[ U_{bond} = K (r' - r_0)^2 \quad (2.5) \]

where \( r \) is the bond length, \( r_0 \) is the equilibrium bond length, and \( K \) is the force
constant for the bond. In solving the above equation, we have used the methodology of
LAMMPS[33, 34]. The velocities and positions of the beads belonging to the trimers
of the capsid protein are updated by the Richardson iterations. In this technique,
the force and torque on individual beads are computed at every time step. Based on
these inputs, the net force and torque acting on the trimer subunit are computed. The
centers of mass of individual subunits are then translated and the beads belonging
to individual subunits are rotated in the new coordinate frames of the respective
subunits. Hence the subunits move as single units keeping their shape intact at every
time step. Due to this rigid-body motion, it is unnecessary to compute the interaction
between the beads within a subunit. The velocities and positions of the polymer beads
are updated by the velocity Verlet algorithm.

\[
\frac{\partial r(t + \delta t)}{\partial t} = \frac{\partial r(t)}{\partial t} + \frac{1}{2}\delta t \left[ \frac{\partial^2 r(t)}{\partial t^2} + \frac{\partial^2 r(t + \delta t)}{\partial t^2} \right]
\]

(2.6)

\[
r(t + \delta t) = r(t) + \delta t \frac{\partial r(t)}{\partial t} + \frac{1}{2}\delta t^2 \frac{\partial^2 r(t)}{\partial t^2}
\]

(2.7)

where \( r \) is the position vector, \( t \) is the time and \( \delta t \) is the time step.

All variables in our simulations are expressed in dimensionless Lennard-Jones (LJ)
units, fully consistent with LAMMPS. The LJ units of length, mass, time, energy,
charge, and temperature are, respectively, \( \sigma \) (taken to be 1nm), \( m_0 \) (taken to be
1kg/mole corresponding roughly to one united atom for a polynucleotide-like poly-
mer), \( \sqrt{\frac{\sigma^2 m_0}{\epsilon}} \), \( \epsilon \), \( \sqrt{\frac{4\pi \epsilon_0 \epsilon \sigma}{\epsilon}} \), and \( \epsilon/k_B \). The dielectric constant \( \epsilon_r \) is taken as 80
and \( 1e = 7.5 \) LJ units. The Debye length \( \kappa^{-1} \) is 1 LJ unit, corresponding roughly
to 100mM monovalent strong electrolyte solution at room temperature. The force
constant \( K \) is given the value of 40 LJ units which is sufficient to keep the bond
length within 5% of the equilibrium value of \( r_0 = 1 \) LJ unit. The ratio of \( m_i \) to
the friction coefficient \( \zeta_i \) is taken as 100 LJ time units. \( \epsilon \) is taken as unity for the
excluded volume interaction and \( \epsilon \) is taken in the range from 1 to 5 for hydrophobic
interactions. The reduced temperature in our simulations is \( T^* = k_B T/\epsilon \). We have
monitored the effect of temperature on the assembly, by changing \( \epsilon \). As will be dis-
cussed below, successful assembly occurs only in the range of \( 1.8 \leq \epsilon \leq 2.1 \), and our
choice of \( \epsilon \) within this range enables a reliable exploration of the assembly kinetics.
The time step in our simulations of proteins is 0.05 LJ units in the absence of the
polymer and the time step is 0.02 LJ units in the presence of the polymer. The sim-
ulation box is a cube of side $L$ with periodic boundary conditions. Each simulation is carried out with multiples of 20 subunits (with and without the polymer) and the box length $L$ is chosen to reflect the protein concentration. $L$ is varied from 40 to 100 LJ units. In the beginning of the simulation, the capsid subunits are placed in randomly chosen locations with random orientations inside the box. For the system containing the polymer, the polymer chain is first equilibrated inside the box before inserting the capsid subunits randomly with random orientations. The coordinates of the beads and thermodynamic quantities such as the total potential energy are collected at regular intervals of the simulation for further analysis. We have monitored the size of the assembly and the orientation of the subunits with respect to each other in the assembled structures. The size of the assembly is determined by monitoring the distance between the centers of mass of the subunits. If the distance between the centers of mass is within 4.5 and 6.0 LJ units then the subunits are considered to be attached. This choice of the cut-off distances is motivated by the economy of the computation time and by the requirement for a successfully assembled structure that there should be an acute angle between the neighboring subunits. The magnitude of the range allows fluctuations in the angle between the inward normals from the subunits. We take the inward normal as the normal from the face of the subunits towards the charged side of the subunit. If the subunits are attached to each other, then the orientation of the subunits with respect to each other is computed by evaluating the dot product of the inward normals of all the adjacent pairs of subunits. Although the angle between the normals in an ideal icosahedron is close to $36^\circ$, we have set an upper bound of $60^\circ$ to allow sufficient fluctuations due to thermal forces. Only if the angle between the neighboring normals is within $60^\circ$, the configuration is accepted as a part of the assembling structure. Defining an orientation parameter for each pair of adjacent subunits, we assign a value of unity if the members of the pair have normals within $60^\circ$. If the sum of the orientation parameter values for all pairs
of subunits in the assembled structure is less than the total number of adjacent pairs in the structure, then the structure is labeled as incorrectly assembled. In a properly assembling structure, the total value of the orientation parameter is identical to the number of adjacent pairs. We have used the orientation parameter to assess the suitable conditions for the formation of proper capsids. At an intermediate time during the simulation, the trimers constitute a population of partially assembled structures. Let \( k \) be the number of the trimers assembled into a correctly assembled structure as outlined above. \( k = 1 \) represents individual subunits. We then count the number of such \( k \)-mers in the population as \( n_k \). The average size of the aggregate is defined as

\[
n_{av} = \frac{\sum_{k=1}^{20} k n_k}{\sum_{k=1}^{20} n_k} \tag{2.8}
\]

where the angular brackets denote the average over 200 independent simulations.

In addition to the kinetic pathways of assembly of the virus-like particles, we have used the parallel tempering method [35, 36] to calculate the free energy landscape for the assembly process, both in the presence and absence of the polymer. In the parallel tempering method, there are a certain number of replicas of the system. Each replica is in a canonical ensemble with a prescribed temperature. First, the system is allowed to evolve in one replica usually with a higher temperature. Then the information on the configurational details in the first replica is exchanged into another replica with an acceptance probability and the system would evolve in the new replica at the new temperature. After an elapse of certain time, the swapping of configurational information is performed into another replica, and this process is continued. The acceptance probability \( p_{\alpha \rightarrow \beta} \) in going from the \( \alpha \)th replica to the \( \beta \)th replica is

\[
p_{\alpha \rightarrow \beta} = \min\left[1, \exp\left(\frac{1}{k_B T_\alpha} - \frac{1}{k_B T_\beta}\right)(U_\alpha - U_\beta)\right] \tag{2.9}
\]
where $U_\alpha$ and $T_\alpha$ are, respectively, the total potential energy and the temperature of the system in the $\alpha$th replica. In order to ensure that the average kinetic energy of the system per particle remains as $(3/2)kBT$, the momentum of the $i$th particle is scaled to a new value during the swapping between replicas according to

$$p_{\text{new}}^{(i)} = \sqrt{\frac{T_{\text{new}}}{T_{\text{old}}}} p_{\text{old}}^{(i)}$$  \hspace{1cm} (2.10)$$

For the case of capsid assembly in the absence of the polymer, the highest temperature at which the system stays predominantly in the free subunit state is $k_B T = 1.3$. In the free energy calculations, the reduced temperature is obtained by fixing $\epsilon$ to be 1.9 and changing the temperature. In order to make sure that the energy histograms of the replicas overlap sufficiently to allow the use of Eq. (2.9), we have used four replicas at temperatures $k_B T = 1.0, 1.1, 1.2, \text{and} 1.3$ in constructing the free energy landscape. In the presence of the polymer, the temperatures of the four replicas are $k_B T = 1.0, 1.2, 1.4, \text{and} 1.6$. Instead of equal spacing between successive temperatures, other optimal strategies may be adopted to improve the rate of convergence. Since convergence was attained with the present choice for the current problem, we did not explore other procedures to improve the rate of convergence. The total number of simulation steps (in LJ units) were $5 \times 10^8$ and $9 \times 10^8$ for the system without and with the polymer, respectively. In the absence of the polymer, the swapping interval and the time step were, respectively, $1 \times 10^7$ and 0.04 in LJ units. In the presence of the polymer, the swapping time and the time step were, respectively, $8 \times 10^6$ and 0.01 in LJ units. The free energy of the $k$-mer $F(k)$ is obtained from

$$F(k) = -k_B T \ln\left(\frac{\sum_{k=1}^{20} k n_k}{\sum_{k=1}^{20} k n_k}\right)$$  \hspace{1cm} (2.11)$$

with the reference state of $F(1) = 0$. The average and standard deviation values of free energy were computed from three independent replica exchange simulations.
2.4 Results and Discussion

This section is the same as it appears in our publication [37].

2.4.1 Polymer-free assembly

The successful formation of icosahedral shape and the growth kinetics depend delicately on the energetics of the association of the trimeric subunits, which appear through the reduced temperature $T^*$ in our simulations. There emerges a narrow temperature range ($0.476 < T^* < 0.581$) in which successful assembly occurs. No assembly is observed at higher temperatures and monster-like particles without the icosahedral symmetry form at lower temperatures. This is illustrated in figure 2.5. Starting from the initial condition (figure 2.5a) of 20 randomly distributed subunits with random orientations inside a cubic box of length 40 LJ units (corresponding to the subunit concentration of $c = 0.5188\text{mM}$), figure 2.5b, figure 2.5c and figure 2.5d are the snapshots at $t = 10^5$, $5 \times 10^5$, and $1.75 \times 10^6$, respectively, at $T^* = 0.5$. At this temperature, the final assembled structure is the virus-like particle with full icosahedral symmetry. On the other hand, for the same elapse of time ($t = 1.75 \times 10^6$), monster-like structures form at $T^* = 0.45$ (figure 2.5e) and no assembly occurs at $T^* = 0.6$ (figure 2.5f). It is thus clear that for the formation of virus-like particles with correct morphological symmetry to occur, the interactions between the subunits need to be sufficiently weak to enable correction of errors during assembly and at the same time sufficiently strong for the assembly to proceed. This scenario is also observed when multiple virus-like particles assemble simultaneously. As an example, the starting configuration (figure 2.6a), with 100 subunits at the subunit concentration of $c = 0.5188\text{mM}$, leads to monster-like particles (figure 2.6b) at $T^* = 0.45$, multiple virus-like-particles (figure 2.6c) at $T^* = 0.5$, and no assembly (figure 2.6d) at $T^* = 0.6$. Based on such simulations, it is in principle possible to determine the ranges of the reduced temperature for different subunit concentrations, which would
allow the formation of correct icosahedral morphology. The determination of such ranges must inevitably depend on the cut-off time set in the simulations to monitor whether or not correct structures formed. This time-consuming exercise might not be that useful in itself as we do not address the specificities of particular viruses in our coarse-grained modeling. Instead, we are satisfied by observing that there is a narrow range of parameters where successful assembly would proceed and then focus on the general mechanisms of assembly in this range of parameters. The time-evolution of the population of the k-mers in the system shows that the initial population of unassociated subunits progressively generates larger and larger k-mers until all subunits are incorporated in the final icosahedron. The distribution functions of the mass fraction of k-mers in the entire system are given in figure 2.7a as functions of time. These distribution functions are constructed from 200 independent simulations at T* = 0.526 and c = 0.5188mM. As expected, there is a cascade whereby smaller ones feed into bigger ones which in turn feed into even bigger ones. The height, width and the characteristic time for the peak height increase as the size of the k-mer increases. Similar time evolution has been observed also in the presence of the polymer.

In addition, we have monitored the averaged growth kinetics of individual assemblies by monitoring the average number of subunits, at a given time, that are participants of partially assembled structures but with the correct orientation parameter. The time dependence of the average size of the assembling structure is given in figure 2.8a at different temperatures (c = 0.5188mM). The representative structures are also included in the figure for T* = 0.526. The effect of subunit concentration on the growth kinetics is illustrated in figure 2.8b at T* = 0.476. The role of polymer in the assembly kinetics is shown in figure 2.8c at T* = 0.5266 and c = 0.5188mM. It is evident from figure 2.8 that there are generically three regimes. A slower kinetics in the very early stage, a linear growth rate in the intermediate stage, and a slowing down of growth in the final stage. These three features are typical of the phenomenon
Figure 2.5. Stable assembly of the icosahedron occurs only at intermediate reduced temperatures. (a) Starting configuration with 20 subunits; (b)-(d) are snapshots at $t = 10^5, 5 \times 10^5$ and $1.75 \times 10^6$, respectively at $T^*=0.5$; (e) monster-like particle at $t = 1.75 \times 10^6$ and $T^*=0.45$; (f) no assembly at $t = 1.75 \times 10^6$ and $T^*=0.6$. 
Figure 2.6. Starting configuration with 100 subunits (c=0.5188 mM); (b)-(d) final structures at $T^*=0.45$, 0.5, and 0.6, respectively, at $t=1.75\times10^6$. 
Figure 2.7. Time-evolution of the distribution functions of mass fraction of k-mers in the population.
of crystallization, where the general mechanism of crystallization is nucleation and growth. This difficulty arises mainly from the smallness of the assembled structures and the fluctuations that are averaged out in constructing figure 2.8. In view of this we computed the free energy landscape to assess the possible existence of nucleation barriers for assembly. In addition, we analyzed the simulation data for individual events during the assembly process (as described below). As discussed in previous section, we computed the free energy landscape by using the parallel tempering method. By taking the free energy of the free unassociated subunits as zero, the free energy landscape for the formation of k-mers is given in figure 2.9. The temperatures are $T^* = 0.5266$ and $T^* = 0.631$ in figure 2.9a and figure 2.9b, respectively. The corresponding free energy landscapes in the presence of the polymer are included in these figures for further discussion below. These figures clearly show that there is a free energy barrier for assembly. At $T^* = 0.5266$, the barrier is about $5k_BT$ occurring at about $k = 3$. It is thus necessary for the formation of trimers of subunits to occur first before the subsequent growth into the fully assembled structure would occur. Also, at this temperature, the final assembled structure is more stable than the unassociated subunits by about $3k_BT$. In contrast, at the higher temperature of $T^* = 0.631$, the barrier is much larger (about $11k_BT$) and the assembly is an unfavorable process by about $4k_BT$. As will be discussed below, the presence of polymer has a profound effect on the free energy landscape and the nucleation barrier.

The existence of the barrier for the formation of three associated subunits allows us to delineate the initial nucleation regime with a lag time and the growth regime. These are marked as G1 and G2 in figure 2.8a. The third regime of G3 is the usual slowing down stage in any growth process with continuous depletion of building units in the reservoir. Based on the free energy landscape, the nucleation time is associated with the average time required for the formation of 3-mer. To lend additional support for identifying the size of the critical nucleus, we have monitored the time taken by a
Figure 2.8. Growth kinetics of average aggregate size $n_{av}$. (a) Effect of temperature with $c=0.5188$ mM (no polymer). (b) Effect of subunit concentration at $T^*=0.476$ (no polymer). (c) Effect of the polymer length at $T^*=0.526$ and $c=0.5188$ mM.
Figure 2.9. Free energy as a function of aggregate size. (a) $T^*=0.5266$ and (b) $T^*=0.631$. 
(k - 1)-mer to form k-mer, as given in figure 2.10 at different temperatures. The time for the formation of 2-mer from the 1-mer is significantly larger than the subsequent steps. The time for addition of one more unit beyond the 2-mer is roughly the same, except at later stages beyond 11-mers. The $\tau$ values for $k = 12$ onwards keep increasing with $k$. Therefore, these data indicate that G1 phase corresponds to the nucleation of 2-mers, G2 phase corresponds to the linear growth rate for $k = 3 - 11$, and G3 corresponds to the last slowing down stage. The same features are also seen when we analyzed the data at different subunit concentrations (not shown). It is to be noted that the free energy calculations suggest that the critical nucleus size is three whereas the kinetics data suggest that it is two. Therefore we have computed the average times taken in the formation 3-mers and 2-mers and analyzed the data within the framework of classical nucleation theory.

![Figure 2.10. Average time $\tau_k$ for the formation of k-mer from (k-1)-mer at c=0.5188 mM. The reduced temperatures are (a)0.476, (b)0.5, (c) 0.526, and (d) 0.541.](image)

Figure 2.10. Average time $\tau_k$ for the formation of k-mer from (k-1)-mer at c=0.5188 mM. The reduced temperatures are (a)0.476, (b)0.5, (c) 0.526, and (d) 0.541.
According to the classical nucleation theory, the nucleation time depends exponentially on the inverse of the quench depth in temperature $T_{m}^* - T^*$, where $T_{m}^*$ is the melting temperature. By taking the disassembly temperature as $T_{m}^*$ and the temperature at which the assembly proceeds as $T^*$, the well known result for the nucleation time $\tau_{nucl}$ is

$$
\tau_{nucl} = A \exp \left( \frac{B}{(T_{m}^* - T^*)T^*} \right),
$$

(2.12)

where the prefactor $A$ mainly depends on the collision frequency of the subunits and $B$ is a constant depending on the thermodynamic quantities of the assembling system. Similarly, when the subunit concentration is the variable, the nucleation time depends on the super saturation which is the excess subunit concentration above the minimum concentration $c_{m}$ required for assembly. The result from the classical nucleation theory is

$$
\tau_{nucl} = A' \exp \left( \frac{B'}{\ln(c/c_{m})} \right),
$$

(2.13)

where $A'$ and $B'$ are constants. According to Eqs (2.12), (2.13), plots of $\ln \tau_{nucl}$ versus $1/(T_{m}^* - T^*)T^*$ and $1/\ln(c/c_{m})$ should be linear. These expectations based on the classical nucleation theory turn out to be valid as shown in figure 2.11a (at $c=0.5188$ M) and figure 2.11b (at $T^* = 0.476$). In fitting the data in figure 2.11a, the disassembly temperature $T_{m}^*$ is taken as the fitting parameter. The fit is indeed good. The value of the fitted disassembly temperature $T_{m}^* = 0.581$ (obtained by assuming that the critical size is a 3-mer) is consistent with simulation results where assembly did not proceed at temperatures higher than this in the absence of the polymer. If we choose the critical nucleus size as a 2-mer, then the fitted value for the disassembly temperature is $T_{m}^* = 0.583$ (the fit not shown). Thus the choice between 2-mer and 3-mer for the critical size does not make any significant difference in $T_{m}^*$. Similarly, the fit shown in figure 2.11b for the dependence of the nucleation time on
the supersaturation is good with the fitted minimum assembly concentration being 0.001mM. Again, the choice of either the 2-mer or the 3-mer as the critical nucleus size does not affect this value of the minimum subunit concentration for assembly. It must be emphasized that the actual values of $T_m^*$ and $c_m$ ought to be dependent on the specificities of the system. Our primary focus here is to find whether the assembly process of virus-like particles proceeds by the nucleation mechanism or not.

The linear growth rate in the G2 phase of crystallization phenomenon is broadly interpreted as an adsorption/desorption process for small molecular systems and as secondary nucleation process for polymeric systems. The theoretical expressions for the growth rate are different for these two mechanisms. For the adsorption/desorption mechanism, the growth rate in the G2 phase is

\[ G_2 = k_c \left[ 1 - \exp \left( \frac{-B_3(T_m^* - T_*)}{T_0 T_m^*} \right) \right], \]  

(2.14)

where $B_3$ is a thermodynamic factor and $k_c$ is a concentration dependent factor. For the secondary nucleation mechanism, the growth rate is given by

\[ G_2 = A'' \exp \left( \frac{B''}{(T_m^* - T_*)T_*} \right), \]  

(2.15)

where $A''$ and $B''$ are constants. We have fitted the data according to the two above equations, as shown in figure 2.12a and figure 2.12b. In figure 2.12a, $\ln(1-G_2/k_c)$ is plotted against $(T_m^*-T_*)/T_*$. If adsorption/desorption is the dominant mechanism in the G2 phase, then the plot should be a straight line. This is clearly not observed for any chosen value of $k_c$. On the other hand, the plot of $\ln G_2$ against $1/(T_m^*-T_*)T_*$ in figure 2.12b is linear in accordance with the secondary nucleation mechanism. It must also be emphasized that the value of the disassembly temperature $T_m^*$ obtained from the best fit in figure 2.12b is identical to that obtained in figure 2.11a for the primary nucleation. As a result, we conclude that the linear growth rate in the intermediate stage of assembly is analogous to that of the two-dimensional
Figure 2.11. Nucleation time follows classical nucleation theory. (a) Plot of $\ln(\tau_{\text{nucl}})$ against $1/(T_m^*-T^*)T^*$ is linear with the best fit value $T_m^*=0.581$ (and regression coefficient =0.9989). (b) Plot of $\ln(\tau_{\text{nucl}})$ against $1/\ln(c/c_m)$ is linear with the best fit value $c_m=0.001$ mM (and regression coefficient =0.9982).
lamellar growth in polymer crystallization. The concentration dependence of G2 is given in figure 2.12c demonstrating a linear proportionality between G2 and c. This linear dependence is mainly due to the stepwise addition of the subunit to the growing assembly. This is consistent with the approximate equal duration required for addition of one more subunit in the G2 phase as seen in figure 2.10.

Among the two mechanisms shown in figure 2.3, the self-assembly needs to follow the nucleation-growth pathway for proper assembled closed structure. The capsid needs to breathe while growing; if the subunits have associated wrongly then they need to rearrange and form the correctly assembled structure. With spontaneous assembly, correct assembled structure is possible, but the probability of formation of mis-assembled structures also exists. The mis-assembled structure do not have the chance to rearrange once formed. Since in most in-vitro experiments uniform distribution of virus like particles are observed, hence nucleation-growth is the most probable mechanism of formation of virus like particles.

2.4.2 Polymer-assisted assembly

As has already been alluded to in figures 2.8 and 2.9, the presence of the polymer chain significantly affects the assembly process. A typical trajectory is given in figure 2.13 for the chain length N = 130 at c = 0.5188mM and T* = 0.5266. The figures. 2.13a-2.13f are snapshots at times 0; 5 x 10^3; 1 x 10^4; 5 x 10^4; 1 x 10^5, and 5 x 10^5, respectively. It is evident from these snapshots that multiple capsid subunits bind with the polymer chain and the local concentration of the subunits is enhanced around the backbone of the chain. Since no specific sequence is endowed on the polymer, the binding occurs equally at all locations of the polymer. As the local concentration of the subunits is increased, the initial assembly of the subunits into the critical nucleus size is facilitated. Also, longer chains promote faster assembly, as long as they are not longer than the maximum length required for the fully assembled virus-like particle.
Figure 2.12. Dependence of linear growth rate on temperature and subunit concentration. (a) Adsorption/desorption model is not obeyed as indicated by the poor fit with Eq. (2.14); the line is with $k_c=0.1$ and regression coefficient $=0.8094$. (b) The linear fit (with $T_m^*=0.581$ and regression coefficient $=0.9997$) for $\ln(G2)$ vs. $1/(T_m^*-T^*)T^*$ supports the secondary nucleation model. (c) $G2$ is linear in subunit concentration.
As time progresses, the subunits then associate with the growing assembly further facilitated by the polymer. The polymer gets encapsulated by the subunits as seen in figures 2.13d and 2.13e. The positions of the united atoms belonging to the polymer are close to the capsid wall, as shown in figure 2.14a. The radially averaged density profile of the polymer from the capsid wall towards the center of the capsid is given in figure 2.14b. This density profile is analogous to the various experimental and theoretical density profiles for the genome in RNA-viruses. In the present coarse-grained model, there is an upper bound on the chain length to be fully packaged inside the assembled particle. This bound is roughly \( N = 130 \). If \( N \) is larger than the bound, a part of the chain hangs outside the particle as shown in figure 2.14c for \( N = 200 \). The quantitative aspects of the above mentioned features are seen in figure 2.8 and figure 2.9. The effect of chain length on the kinetics of assembly is given in figure 2.8c for \( T^* = 0.5266 \) and \( c = 0.5188 \text{mM} \). The curve corresponding to the polymer size being zero is the same as the blue curve in figure 2.9a. As the chain length increases, the growth kinetics becomes faster. For \( N = 130 \), the kinetics is an order of magnitude faster than in the absence of the polymer. The faster growth kinetics in the presence of the polymer can be attributed to the lowering of the free energy barrier for the assembly by the polymer. The free energy landscapes for \( N = 130 \) are given in figures figure 2.9a and 2.9b at \( T^* = 0.5266 \) and \( T^* = 0.631 \), respectively. Several important conclusions can be reached by comparing these landscapes with those in the absence of the polymer in figure 2.9. First, the nucleation barrier for assembly is reduced by the polymer. The barrier is about \( 4kBT \) in the presence of the polymer at both temperatures. On the other hand, as we have already noted, the barriers are about \( 5kBT \) and \( 11kBT \) at the lower and higher temperatures, respectively, in the absence of the polymer. Thus, the nucleation barrier is reduced by the polymer by about \( 1kBT \) and \( 7kBT \) at \( T^* = 0.5266 \) and \( T^* = 0.631 \), respectively. This reduction of the nucleation barrier is responsible for the faster kinetics of assembly in the presence...
of the polymer. The second conclusion is that the presence of the polymer makes the assembly process more thermodynamically favorable than without the polymer. As shown in figure 2.9b, the thermodynamically unfavorable assembly at $T^* = 0.631$ in the absence of the polymer is made favorable by the polymer. The free energy change is about $-5k_BT$ in the presence of the polymer, whereas it is $+4k_BT$ in the absence of the polymer at $T^* = 0.631$. At the lower temperature $T^* = 0.5266$, an already favorable assembly is made even more favorable by the polymer chain. Another feature of the free energy landscapes in the presence of the polymer is that there appears a metastable state with about ten subunits in the assembled structure. Given the smallness of the simulated system, we have not explored this feature in more detail in the present work. The mechanism of assembly thus turns out to be nucleation and growth, both in the presence and absence of the polymer, the nucleation barrier being reduced by the polymer. As already described for the assembly without the polymer, we take the critical nucleus size to be $k = 3$. Analogous to figure 2.10, we have carried out an analysis of $\tau_k$ in the presence of the polymer. The main result is similar to figure 2.10, except that the kinetics is faster in the presence of the polymer. Again, $k = 2$ is the critical nucleus size based on these kinetic data. As in the case of assembly without the polymer, the choice of $k = 2$ or $k = 3$ for the critical nucleus size does not affect the main conclusions regarding the applicability of the nucleation-growth mechanism for assembly of virus-like particles. Taking $k = 3$ for the critical nucleus size, we have collected the average nucleation time. The dependence of the nucleation time on the chain length is predicted by extending the classical nucleation theory for the present case. In view of the random binding of the subunits on the polymer chain without any cooperativity in the initial stages of assembly, the local concentration $c_{\text{local}}$ can be written as

$$c_{\text{local}} = c + \alpha N,$$  

(2.16)
where $c$ is the concentration of the subunits in the bulk and $\alpha$ is a coefficient for the assumed proportionality between the adsorbed subunits and the chain length. Substituting $c_{local}$ for $c$ in Eq. (2.13), we expect

$$\ln \tau_{nucl} \sim 1/\ln \left( \frac{c + \alpha N}{cm} \right),$$

(2.17)

**Figure 2.13.** Snapshots of assembly in the presence of the polymer ($T^*=0.5266$, $c=0.5188$ mM, $N=130$).

In figure 2.15a, $\ln \tau_{nucl}$ is plotted against $\ln[(c + \alpha N)/cm]$ at three reduced temperatures ($T^* = 0.5; 0.526$, and 0.556 for $c = 0.5188$ mM. The value of $cm$ is taken as the same as the value (0.001 mM) in the absence of the polymer as obtained from figures 2.11b. By using $\alpha$ as the single fitting parameter, Eq. (2.17) is found to be valid. The value of the coefficient $\alpha$ increases roughly linearly with an increase in the reduced temperature, with values 0.02, 0.07, and 0.15 for $T^*=0.5$, 0.526, and 0.556, respectively. This trend indicates that the enrichment of the local concentration of the subunits is facilitated by an increased conformational freedom of the polymer chain at higher temperatures. The linear growth rate in the G2 phase is also enhanced by the polymer. The dependence of G2 on $N$ is given in figure 2.15b at the same three
temperatures as in figure 2.15a \((c = 0.5188\text{mM})\). The data suggest that the linear growth rate \(G2\) is roughly proportional to \(N\). Since the local concentration of the subunit is proportional to \(N\), this result is consistent with the conclusions based on figure 2.12c.

According to our simulation, the self-assembly follows the first pathway in the schematic figure 2.4. This particular mechanism is expected in real cells, because as soon as the subunits are replicated inside a cell, they need to be captured by the genome otherwise they might interact with the other components of the host cell. The kinetics needs to be fast enough for the efficient production of viruses otherwise the subunits might start binding to other structures within the host cell.

The third pathway is the most probable mechanism of formation of virus like particle on pegylated surface of gold nanoparticles [22], where the template is already available and the subunits diffuse to the surface of the template forming virus like particles. The diffusion of the subunits to the template is expected to be slower than the collision of the long genome with the subunits (pathway I). During the diffusion process, the subunits might interact with some other components of the host cell on its pathway.

**Figure 2.14.** Snapshot of the polymer after co-assembly \((N=130)\). (b) Radially averaged monomer density distribution \((N=130)\). (c) The chain spills out of the icosahedron for large values of \(N\) \((N=200)\).
Figure 2.15. Plot of $\ln(\tau_{\text{nucl}})$ vs. $1/\ln((c+\alpha N)/c_m)$ is linear supporting the nucleation mechanism in the presence of polymer. (b) Linear growth rate is approximately linear with the polymer length. In both (a) and (b) the reduced temperatures are $T^*=0.556$ (circles), 0.526 (squares), and 0.5 (triangles).
leading to loss of subunits. The second pathway is very improbable, because sliding mechanism have not been reported anywhere else in the literature as far as we are concerned. Hence, we believe that the first pathway is the self-assembly pathway within a cell because of kinetic reasons.

2.5 Conclusions

We have constructed a coarse grained model for the MVM subunit based on the rich experimental details reported by Reguera et al [3]. The polymer is modeled as a flexible uniformly charged polyelectrolyte chain. Starting with a given number of the subunits in the simulation box, the time-evolution of the assembled structures and their populations were investigated using Langevin dynamics simulation. The free energy landscapes were constructed with the parallel tempering technique. It turns out that only within a narrow region of the parameter space, successful assembly into icosahedra occurs. After establishing this region, we have explored the mechanism of assembly and the dependence of the various measures of assembly on temperature, subunit concentration, and polymer length. The simulations show that the assembly of the subunits into the final icosahedron follows the nucleation and growth mechanism in the absence of the polymer. The features of the assembly kinetics are quite similar to many crystallization processes. There are three stages: nucleation, growth with linear growth rate, and the slowed down growth in the very late stage of assembly. We have shown that the nucleation time follows the expected laws from the classical nucleation theory regarding the dependencies on temperature and subunit concentration. In our model, the critical nucleus size is about three subunits. The second stage of assembly obeys the linear growth rate law. In the growth regime, we have shown that the growth rate obeys the laws expected from the model of secondary nucleation. This behavior is identical to the growth of two-dimensional lamella of polymer crystals where the well known mechanism is secondary nucleation. The third stage of slowed
down growth is due to the continuous depletion of the subunit concentration in the
growth medium. The same mechanism of nucleation and growth is observed in our
simulations of the assembly of subunits into icosahedra in the presence of the polymer
also. Now, the polymer reduces the free energy barrier that needs to be surmounted
for the assembly to occur. We have shown that the local concentration of the sub-
units is increased in the neighborhood of the polymer, due to random binding of the
subunits with the polymer segments, which in turn leads to a reduction in the free
energy barrier. We have shown that the nucleation time is smaller for longer chains in
accordance with the expectations from the classical nucleation theory, as long as the
chain is not too long to be spilling out of the finite sized icosahedron. Also, the linear
growth rate in the second stage of assembly has been found to be higher for longer
chains. It is remarkable that the expectations from the classical nucleation theory are
found to be valid even for such a small system involving only twenty subunits. The
present work is designed to explore the generic mechanism of assembly of icosahedra
and how this is affected by the presence of a flexible polymer bearing opposite charge
to the net positive charge on the inside surface of the icosahedra. The present model
is not suitable for addressing specific issues of virus assembly, although the capsid is
fashioned after a realistic virus. This is due to the crude nature of the coarse graining
used in the present simulations. Thus, comparisons with experimental data on any
particular virus cannot be readily made. However, the present work strongly suggests
that the assembly kinetics of viruses must generally follow the nucleation and growth
mechanism. It has been reported in the literature [38] that the self-assembly may
follow two mechanisms. One of the mechanism being nucleation-growth when the
association between subunit-polymer is relatively unfavorable. The other mechanism
proposed by them is the adsorption of the subunits onto the polymer in a disordered
fashion followed by cooperative rearrangement to form the ordered capsid. Assembly
occurs rapidly as multiple oligomers appear and coagulate to form an ordered capsid.
In our simulations we observed random binding of the subunits on the polyelectrolyte backbone resulting in a local increase in the subunit concentration leading to a faster nucleation-growth process. Experimentally the real mechanism of the self-assembly can be determined by doing single molecule fluorophore experiments, however it is a challenging experiment. In the experimental system, different colored fluorophores can be attached to the subunit-subunit binding sites and subunit-genome binding sites. Kinetics may be inferred from the fluorescence intensity as a function of time. If the mechanism is similar to what we proposed then we should observe subunit-genome fluorescence signal followed by fluorescence signal from the subunit-subunit interaction where the fluorescent intensity should correspond to a nucleation-growth process. In the future, more refined coarse graining can be employed to address more specific questions such as the charge balance between the capsid and the polymer and the sequence effects from the polymer on the assembly.
CHAPTER 3
EJECTION DYNAMICS OF DOUBLE STRANDED DNA FROM PHAGE

3.1 Abstract

Packing-ejection of chains of different persistence lengths (60 nm and 20 nm) in phi29 phage was investigated using Langevin dynamics simulation. The packing process happens in three different stages: motor driven, motor driven against resistance, saturation. Although the packing process is qualitatively similar for all the motor force, the amount of chain packed and the axial ordering of the packed chain depends on the persistence length of the chain as well as the magnitude of the motor force. The amount of chain packed decreased with the decrease in motor force but the axial ordering got better with reduced motor force. The persistence length of the chain plays an important role in packing: better ordering observed for more stiff chain for the corresponding motor force as well as in the absolute sense (ordering for a stiff polymer is always better than the less stiff polymer irrespective of the motor force). The amount of chain packed, stiffness and the axial order at the end of packing process plays a major role in the qualitative nature of the ejection. For chain of low stiffness, ejection happens smoothly with drift dominant regime at higher packing fraction followed by diffusion dominant regime at lower packing fraction. For chains of higher stiffness, the packing is qualitatively different than the chain of low stiffness. Since the persistence length of the chain is comparable to the diameter of the phage hence diffusion like behavior is observed only for the last fraction of the chain. Moreover, higher amount packed and lower ordering introduces jamming during the
ejection. The jamming of chain results in a slow intermediate ejection kinetics for most cases. With increase in ordering of the system (and decrease in the amount of chain packed), the jamming probability decreases resulting in a smoother exit. Hysteresis was observed for the packing-ejection cycle.

3.2 Introduction

Double stranded (ds) DNA viruses replicate by injecting their DNA inside the host cells. It is contended whether the DNA is injected into the cells due to the in-built pressure [39] or by enzymatic means [40, 41]. The process is at least pressure driven in the in-vitro condition. In the in-vitro equilibrium experiments the phages are dissolved in a solution containing poly ethylene glycol (PEG) [39, 42]. Depending on the applied osmotic pressure or the amount of PEG in the solution, the extent of ejection changes. It was observed that, with increase in applied osmotic pressure, the extent of ejection is decreased, ultimately ceasing at tens of atmospheres. Parameters such as salt-valency and salt-concentration were varied to determine the electrostatic nature of the interaction of intra-dsDNA strands and dsDNA-capsid. These equilibrium experiments give us an idea about the strength and nature of the forces acting but to actually understand the dynamics, kinetic measurements were done. The kinetic experiments can be broadly classified into bulk-phage and single-phage experiments. As the name suggests, the bulk phage experiments involve the measurement of ds-DNA ejected from the phages as a bulk (either total fluorescence intensity or light scattering intensity of the capsids) without distinguishing independent trajectories, while in single phage experiments the trajectory of individual dsDNA was followed by fluorescence imaging of the ejected dsDNA[43, 44, 45]. The total ejection time according to the bulk phage experiments is in minutes, while the single phage experiments show that the ejection happens in seconds[6, 4]. The difference in the time scale can be explained by the different stages in the ejection in these in-vitro
experiments. The first step in the ejection involves binding of the receptor protein on the phage mouth, followed by channel opening and ultimately the ejection of the dsDNA as identified by Chiaruttini et al [5]. The receptor proteins are believed to be only responsible for channel opening and not regulating the rate of ejection. The schematic of the steps involved in the in-vitro experiments are shown in figure 3.1. The channel opening does not happen spontaneously for all the phages, rather the time scale of phage opening is in minutes, that explains the bulk phage experiment dynamics. Hence bulk phage experiments are not suitable to determine the actual ejection dynamics. The single phage experiments have been conducted on two class of virus: T5 and lambda phage. For T5 phage, random pause were observed (figure 3.2a-b), whereas for lambda phage smooth kinetics without any pause was observed (figure 3.2c-d). The random pause was not completely random as it was observed only at high packing fractions (above 40 % DNA packed). Every ejection trajectory was different, with pause happening randomly anywhere at high packing fraction. There has been some speculation regarding the cause of random pause but the reason for this random pause remains unsolved as yet.

Existing studies on modeling the ejection dynamics[46, 47, 12, 48, 49, 50, 51] are limited. Most of the modeling studies assume that the DNA is arranged in a spool like order within the capsid, but recently it has been demonstrated that the conformation of the DNA inside the viral capsid is more concentric than coaxial [52]. Hence, the current modeling efforts has not yet solved the intermediate pause/slow kinetics. The reason cited for the intermediate pause in kinetics are a few. It was suggested by Ali et al [12], that the bad solvent quality causes the intermediate pause by jamming of the chain at about 50 % ejection. But pause has been observed even for good quality solvents as shown for T5 phage in figure 3.2a. For lambda phage, where the ejection is without any pause in good solvent condition, the change in quality of solvent has no effect on its quality of ejection except for increase in ejection time as shown in
figure 3.2d. Hence, the reason cited for pause can be ruled out. The other reason cited for pause has been single strand interruptions/nicks [4]. At high pressure highly bent local strands might interrupt smooth exit. This particular reason can also be ruled out because the pause happens within 60 % DNA ejected and the nicks are expected randomly over the whole DNA length [5]. Moreover, no random pause was observed with a simulated knot on DNA [53]. The other relevant reason speculated for intermediate pause, is the delays in local phase transition and defects underpinning at high packing fraction [5]. DNA arrangement within T5 phage during different stages of ejection was analyzed using cryo-EM [54], they showed that the DNA undergoes hexagonal-cholesteric-isotropic arrangement during the ejection but they did not find any correlation between the phase transition and pause. Hence local phase transition can also be ruled out.

One of the main goal of this simulation is to identify the cause of random intermediate pause during ejection. The other goal is to determine whether hysteresis exists in such system, since this is a non-equilibrium process.

**Figure 3.1.** Schematic of the steps involved in the viral ejection during in-vitro experiments.
3.3 Method

3.3.1 Modeling Capsid

Φ29 phage was chosen for investigation. The pseudo-atomic coordinates of the protein capsid is obtained from Protein Data Bank [55] and each residue is represented as a united atom bead of diameter 3 Å with the center at the C-α position. Since the coordinates represents a matured capsid without any opening for DNA packaging/ejection hence the topmost location on the capsid along the z-axis was chosen to create an artificial hole. The hole has a diameter (27.5 Å) which is slightly larger than the diameter of the coarse-grained DNA beads (25 Å) for packing and ejection. This diameter is comparable to the reported value of the pore in literature [13].
3.3.2 Modeling Motor Protein

A small hollow cylinder was constructed on top of the artificial hole created on capsid. The internal diameter of the cylinder was 27.5 Å and its length was 12 Å. The cylinder is constructed of beads of diameter 3 Å. The capsid along with the motor protein is shown in figure 3.3. The motor protein is quite complex in reality but for all practical purpose the role of the motor protein is to push the DNA beads with a constant downward force that venture into the cuboid (figure 3.3a). The cuboid has a length of 12.5 Å along the axis of the cylinder and a width of 27.5 Å with its axis coinciding with that of the hollow cylinder. A cuboid was used instead of a cylinder because of the software constraints.

3.3.3 Modeling Double-stranded DNA/Small persistence length polymer

A coarse grained bead-spring model with bending rigidity was used to create the polymer/DNA [52]. The diameter and the bond length of the DNA beads are 25 Å (hydrated diameter of DNA) and 12.5 Å [52] respectively to represent roughly cylindrical geometry for the DNA helix. One coarse-grained DNA bead has roughly eight base pairs. The dsDNA was modeled to have a persistence length of roughly 60 nm. The electrostatic interaction between the DNA beads was found to effect only the amount of DNA packed without any effect on the qualitative nature of dynamics[52]. Moreover the Debye length at physiological condition is smaller than the coarse grained diameter hence the electrostatic interactions can be neglected. For all practical purpose, the excluded volume interaction was sufficient for getting the qualitative trend. The goal of the project is to get a qualitative feel of the ejection dynamics hence only excluded-volume interaction between the DNA beads and DNA-capsid will be considered. The effect of electrostatics is being considered for future work where the effect of multivalent ions are expected to play a significant role in the qualitative nature of ejection dynamics. A polymer of small persistence length was
created by reducing the bending force constant. The polymer of smaller persistence length was considered for a complete understanding of the role of persistence length in the ejection process.

3.3.4 Simulation Technique

![Figure 3.3. Coarse grain model (a) Capsid (b) Capsid with dsDNA (c) close-up of initial configuration of DNA near the capsid mouth.](image)

The force fields on the individual beads were computed by Langevin dynamics simulation in LAMMPS [33, 34].

$$m_i \frac{d^2 r_{ij}}{dt^2} = -\zeta_i \frac{dr_{ij}}{dt} - \nabla_j U_i + f_{ij} + f_{i \text{motor}}^{\text{motor}}$$  \hspace{1cm} (3.1)

where $r_{ij}$ is the position vector of the jth component of the ith bead, $t$ is the time. $m_i$ and $\zeta_i$ are the mass and friction coefficient of the ith bead respectively. $U_i$ is the net potential acting on the i-th bead, as given below. $f_{ij}$ is the j-th component of
the random force acting on the i-th bead obeying the fluctuation-dissipation theorem with its magnitude given by $\sqrt{k_B T \zeta_i \Delta t}$ ($k_B T$ is the Boltzmann constant times the absolute temperature). $f^\text{motor}_{isj}$ represents the motor force on the i* bead/s with its center located in the motor protein region. The net potential acting on the i-th bead is the sum of all the non-bonded and bonded potentials.

$$U = U_{LJ} + U_{bond} + U_{angle}$$  \hspace{1cm} (3.2)

Excluded-volume interaction between the DNA-DNA beads and DNA-capsid beads are modeled as

$$U_{LJ} = 4\epsilon \left[ \left( \sigma / r \right)^{12} - \left( \sigma / r \right)^6 \right] + \epsilon_c \hspace{0.1cm} r < r_c, \hspace{0.1cm} 0 \hspace{0.1cm} otherwise$$  \hspace{1cm} (3.3)

where

$$\epsilon_c = -4\epsilon \left[ \left( \sigma / r_c \right)^{12} - \left( \sigma / r_c \right)^6 \right]$$  \hspace{1cm} (3.4)

$\sigma$ and $\epsilon$ are the parameters and $r_c$ is set to 1.12$\sigma$. Hookean bead spring model was used to represent the connectivity between adjacent beads in the DNA.

$$U_{bond} = K_{bond} (r' - r_0)^2$$  \hspace{1cm} (3.5)

where $r'$ is the bond length, $r_0$ is the equilibrium bond length and $K_{bond}$ is the force constant for the bond. A three body interaction potential was used to set the stiffness of the chain.

$$U_{angle} = K_{angle} (\cos \theta - \cos \theta_0)^2$$  \hspace{1cm} (3.6)

where $\theta$ is the angle between three consecutive beads along the chain, $\theta_0$ is the equilibrium value of bond angle, and $K_{angle}$ is the force constant for the angular stiffness.
The capsid and the motor protein are kept fixed throughout the simulation. The velocity and position of the DNA beads are updated at every timestep by velocity Verlet algorithm [56].

All variables are expressed in dimensionless Lennard-Jones (LJ) units, fully consistent with LAMMPS. To non-dimensionalize, the fundamental quantities of mass, length and energy are $m_0$ (taken to be $5.2\text{kg/mole}$ corresponding to the average molar mass of eight base pairs), $\sigma$ (taken to be $25\text{Å}$, the diameter of the coarse grained DNA bead), and $1\ k_B\ T$ (The excluded volume iteration is generally of this order), respectively. The other important quantities for non-dimensionalizing are expressed in terms of the fundamental quantities such as time ($\sqrt{\sigma^2 m_0/\epsilon}$), force ($\epsilon/\sigma$), temperature ($\epsilon/k_B$) and pressure ($\epsilon/\sigma^3$). The fundamental quantities that are based on the $\epsilon$ are estimates based on the approximation of $1\ k_B\ T$. In real they may be quite different, but they should be of the order of magnitude based on this approximation.

$m_0/\zeta$ was chosen to be 300 LJ time units. $\epsilon$ was chosen to be 1 unit, $\sigma$ was chosen to be 1 unit for the DNA beads and 0.12 units for capsid beads which are equivalent to $25\text{Å}$ and $3\text{Å}$ in real units. $K_{\text{bond}}$ value of 1500 units was used and it was sufficient to keep the bond length within 1% of its original $r_0$ value of 0.5 LJ units. $\theta_0$ was set to 180° and the $K_{\text{angle}}$ was tuned to set the stiffness of the chain as described below. Temperature was set to 1 unit and the timestep was chosen to be 0.0003 units for packing and 0.003 for ejection simulations, respectively.

### 3.3.5 Initial Configuration

Primarily three steps were executed. Stiffness (persistence length) of the chain was determined as a function of $K_{\text{angle}}$. In the second step polymer of certain stiffness (fixed $K_{\text{angle}}$ value) was packed into the capsid with the help of motor protein. In the last step, overhanging part of the DNA/polymer was removed and the ejection of the DNA was simulated without any interference from motor protein.
3.3.6 Persistence length determination

A polymer chain of 5000 beads (with a particular $K_{\text{angle}}$ value) was equilibrated in a cubical simulation box of length 200 LJ units with periodic boundary conditions. Following equilibration, the instantaneous average bond angle was monitored to calculate the persistence length [52]. The persistence length is given by:

\[ L_p = \frac{r_0}{1 + < \cos \theta >} \]  

(3.7)

where $r_0$ is the equilibrium bond length and $< \cos \theta >$ is the instantaneous average of the bond angle. Persistence length was obtained as a function of the $K_{\text{angle}}$ figure 3.4. The double stranded DNA is reported to have a persistence length close to 50-60 nm, hence $K_{\text{angle}}$ value of 750 was used to simulate a chain with stiffness similar to DNA [52]. The dynamics of packing a polymer chain of smaller persistence length was also investigated by choosing a chain with 20 nm persistence length ($K_{\text{angle}} = 100$). Five independent simulation were conducted and average value is shown in figure 3.4.

![Figure 3.4. Persistence length as a function of $K_{\text{angle}}$](image-url)
3.3.7 DNA Packing

Capsid with the motor protein were placed in the cubical simulation box of 200 units with the center of sphere coinciding with the center of simulation box. The outward normal to the hollow cylinder points in the positive z axis. A chain (with appropriate stiffness) of 5000 beads was placed in the simulation box, with three beads from one end of the chain inside the mouth of the pore as shown in figure 3.3c. The center of the third bead and the center of the outer face of the cylinder coincide, with the other two beads lying within the cylinder. The chain was equilibrated keeping the three polymer beads in the mouth immobile. After equilibration, the effect of motor protein was simulated by applying a downward normal force in the mouth of the pore. Any bead with its center lying within the motor protein region experiences a downward force of appropriate magnitude. Six different motor forces were used: 5pN, 10 pN, 20 pN, 30 pN, 40 pN, and 55 pN. The average magnitude of the applied motor force is known to be 55 pN [52] for phi29. The magnitude of motor force determines the amount and rate of packaging. The goal was to investigate the effect of packaging rate on the morphological ordering and pressure built in and to determine what is so special about the motor force of 55 pN. The simulation was executed for approximately 3000 time units. The position of the DNA beads, pressure and energy were collected at regular intervals for further analysis. Ten independent simulation runs were conducted and the average values are reported.

3.3.8 DNA Ejection

After the packaging, the DNA remaining outside of the capsid was removed followed by simulation without the action of motor protein (last term in equation (3.1) was removed). Again, the position of the DNA beads, pressure and energy were collected at regular intervals for further analysis. Twenty independent simulations are conducted.
3.3.9 Parameter evaluation

1. **Packed Length**: The packaged length and the ejected length at any point of time is evaluated by determining the bead index at the mouth. For the ejection case, when the minimum distance between the center of the mouth of the hole to the polymer bead exceeds 1 LJ unit, then the chain is considered as fully ejected. The total packed length is determined at the end of the 2750 LJ time units.

2. **Pressure**: The pressure of the whole system is calculated by LAMMPS using the given formula:

\[ P = \frac{Nk_B T}{V} + \frac{\sum \vec{r}_i \cdot \vec{f}_i}{dV} \]  

(3.8)

where \( N \) is the number of beads in the system, \( k_B T \) is the Boltzmann constant times temperature, \( d \) is the dimensionality of the system (3 in this case), \( V \) is the system volume and the second term is the virial computed for all pairwise (LJ in this case) as well as 2-body (bond stretching interaction), 3-body interactions (angular interaction potential).

\[ \sum \vec{r}_i \cdot \vec{f}_i = \sum \sum \vec{r}_{ij} \cdot \vec{f}_{ij} \]  

(3.9)

where \( \vec{r}_{ij} \) is the vector distance between ith and jth bead and \( \vec{f}_{ij} (= \nabla_{\vec{r}_{ij}} U(\vec{r}_{ij})) \) is the force between ith and jth bead and \( U(\vec{r}_{ij}) \) is the sum total of potential between the ith and jth bead. The reported value of pressure from LAMMPS is for the whole simulation box, but the contribution to pressure from the beads outside the capsid is negligible, hence the reported pressure value is multiplied by a factor \( V_{\text{Simulation box}}/V_{\text{capsid}} \) to estimate the pressure inside the capsid in LJ units, with \( V \) being the volume. The capsid is approximated as a sphere of inner diameter 41 Å, which is about 16.4 LJ units and the simulation box is a
cube of edge 200 LJ units. The obtained pressure is then multiplied by $\epsilon/\sigma^3$ to estimate the pressure in real units. As mentioned earlier $\epsilon$ is assumed to be 1 $k_BT$ and the value for $\sigma$ is 25 Å in real units. The average value of the pressure as a function of packed fraction is presented in the results section.

3. **Hysteresis:** The hysteresis is the energy loss per unit volume for a non-equilibrium process. The hysteresis of packing-ejection cycle was computed by measuring the difference in integrated area under pressure versus packed-fraction for the packing and ejection cycle. The average and standard deviation for the twenty ejection cycle and their corresponding packing cycle are presented in the next section.

4. **Ordering:** The morphology of the DNA inside the capsid was followed as a function of time. We were concerned with two different aspects of the DNA arrangement inside the capsid: radial order and axial order.

**Radial Order:** For radial order measurement the bead density at any instant was projected onto the plane perpendicular to the packaging axis (axis of the hollow cylinder). Projected densities are averaged azimuthally to obtain a radial distribution function [52].

**Axial Order:** The axial order parameter defines the degree to which the packaged DNA forms a toroid-like assemblage of stacked hoops aligned with the packaging axis. Fifteen consecutive beads were considered along the chain. A unit normal vector for each arc was defined by normalizing the cross-product of the two vectors connecting the middle bead to each of the end beads ($\vec{v}_1 = \vec{r}_1 - \vec{r}_8$ and $\vec{v}_2 = \vec{r}_{15} - \vec{r}_8$). For each arc along the chain contour a unit arc normal was computed and projected its length along the packing axis ($\vec{c} = \vec{v}_1 \times \vec{v}_2 / |\vec{v}_1 \times \vec{v}_2|$). The axial order parameter was determined by dividing the sum of the magnitude of projected arc normals by the number of arcs. The order parameter
approaches unity for a perfect stack of hoops, whereas a group of randomly oriented arc normals gives a value of 0.5 [52].

\[ AxialOrder = \frac{1}{N_{Arcs}} \sum_{i} |v_z| \]  
(3.10)

where \( N_{Arcs} \) is the number of arcs and \( v_z \) is the projection of the arc normal along z-axis.

5. **Trajectory and Displacement:** The trajectory and displacement of three beads were followed during the ejection process. The position of the beads were chosen randomly along the length of the polymer to be 375 nm, 1875 nm, and 2750 nm. Although they were randomly selected but the values were far apart. The goal was to investigate the path followed by beads at different packed fractions. The 3D trajectory followed by any bead is tracked by plotting a vector from the starting point to the destination point during a timestep. The trajectory gives an idea of the kind of path followed during the ejection: zig-zag or defined. The displacement (\( \Delta R \)) is the magnitude of the vector from the initial time (\( t=0 \)) to any time (\( t \)). The displacement is plotted against time to determine the nature of pathway followed by a stiff chain relative to less stiff chain. The trajectory and displacement both give an idea about the nature of pathway followed by a stiff chain and a less stiff chain at different packing fractions.

### 3.4 Results and Discussion

#### 3.4.1 Packing dynamics

Two parameters were used: (a) Motor force: determines the rate and amount of packaging (b) the persistence length of dsDNA, for a better understanding of the role of these parameters during packaging and ejection. The packaging was qualitatively
similar for all the cases (figure 3.5) of motor force (5 pN, 10 pN, 20 pN, 30 pN, 40 pN, 55 pN). The packaging proceeds at approximately constant rate until the motor protein reaches its limit. Beyond certain loaded length of the dsDNA, the packaging plateaus. Similar trends were observed by Forrey et al [57].

![Image: Figure 3.5. Packaging versus time (a) Effect of different motor force (b) Effect of persistence length of dsDNA](image)

The packing happens in three stages, motor driven, motor driven against crowding followed by saturation. In the first stage the motor protein is pushing the DNA against zero resistance as the packed pressure is not sufficient to exert an opposing force. In the second and third stage, the capsid is already crowded and the motor is still pushing the DNA leading to creation of inner layer of DNA and also addition of beads to the existing layers. This leads to a sharp increase in the bending energy as well as the LJ interaction energy towards the end (around 0.9 fraction packed) of the packing (figure 3.7) with a reduced rate of packing. The last stage, the system is trying to reach the mechanical equilibrium by packing as much DNA as possible to balance the applied pressure from the motor force. The total packed length as shown in inset figure 3.5a, saturates with increase in motor force rather than increasing linearly. Increase in packing rate introduces more randomness in the system hence the packing is not very
efficient leading to earlier saturation in packed length. A more ordered system would pack more length for a given motor force. The packing dynamics was followed for polymer with 20 nm persistence length as shown in figure 3.5b. The dynamics seems to be independent of the persistence length and is only dependent on the motor force although the total packed length is higher for less stiff polymer as expected (see inset figure 3.5b). The first stage lasts till at least 0.5 fraction packed (can go as high as 0.8 fraction for 55 pN motor force) and the saturation stage starts at about 0.9 fraction packed. The intermediate stage can start from somewhere between 0.5 fraction packed to 0.8 fraction packed. The interesting point to be noted about the packing trajectory is that the packing time is about the same for all the cases of motor force. The higher motor force is able to push DNA/polymer inside the capsid at a faster rate initially, but the rate of packing suddenly drops down at later stage. The sudden drop in packing is due to sudden increase in the resistance at high packing fraction to the entry due to sudden rise in the pressure inside the phage as shown in figure 3.6. The rate of packing with lower motor force decreases smoothly to saturation. As shown in figure 3.6, the pressure increases steadily with increase in packing fraction.

The pressure builds up to tens of atmosphere as expected (figure 3.6) due to increase in both bending energy as well as the repulsion within the system (figure 3.7). The morphology of the dsDNA inside the capsid was followed as a function of time. We were concerned with two different aspects of the dsDNA arrangement inside the capsid: (a) radial order (b) axial order. For radial order measurement the bead density from simulation (averaged for 50 equilibrated runs after packaging was ceased) was projected onto the plane perpendicular to the packing axis (figure 3.8). Projected densities are averaged azimuthally to obtain a radial distribution function (figure 3.9). The radial order matches qualitatively with the previous experimental study[58]. The radial order plot shows that the dsDNA is arranged radially in layers with a layer thickness of approximately 2.5 nm and the dsDNA is packaged more densely near the
Figure 3.6. Pressure buildup during packing for different motor force (Average value). The total length packed is shown in figure 3.5
Figure 3.7. Contribution of different component of energy to the packing pressure (Average value).
capsid walls because of its elasticity. The number of peaks decreases with decrease in motor force and the peaks are shifted more towards the center. To get a better understanding of these characteristics we investigated the axial ordering inside the phage (figure 3.10). The axial order parameter defines the degree to which the packaged dsDNA forms a toroid-like assemblage of stacked hoops aligned with the packing axis. The order parameter approaches unity for a perfect stack of hoops, whereas a group of randomly oriented arc normals gives a value of 0.5. As shown in figure 3.10, the axial order always decreases with increase in the loading length and the slow rate of packaging (low motor force) gives rise to more axial ordering because the dsDNA gets more time to order before another segment is pushed into the crowded environment. For the motor force corresponding to phi29 (Force of approximately 55 pN) there is some axial ordering in the beginning but it becomes more random as time progresses. Hence the dsDNA is more concentric than coaxial as shown in literature [57, 58] unlike many theoretical assumptions in the past. The axial order of dsDNA packed with 5 pN motor force and 55 pN are shown in figure 3.11 for demonstration. As shown in figure 3.11a, the DNA is ordered better but it is not perfectly stacked hoops, hence the value of axial order is about 0.75. Figure 3.11b shows that the packing is completely random with an order parameter approaching 0.5.

3.4.2 Ejection Dynamics

Figure 3.12a shows 20 independent traces of ejection of DNA from the phi29 phage. The plot shows the length of DNA remaining inside the phage as a function of time. There are two class of ejection, one that involves an intermediate slow ejection kinetics (85 % of total runs) and another that ejects without slow intermediate state(15 % of total runs). The ejection starts at a fast rate, but most of the time (85 % of total runs) the ejection is slowed down for a period of time, followed by a sudden increase in the rate of ejection and ultimately ending with a slow exit. The intermediate slow
Figure 3.8. Projected bead density on the xy plane

Figure 3.9. Radial probability density of dsDNA in capsid for different motor force
Figure 3.10. Axial order of the dsDNA in capsid for different motor force

Figure 3.11. DNA arrangement within the capsid after complete packing (after packing timestep of about 2750 LJ Units) (a) 5 pN: Better axial order (b) 55 pN: Random order.
rate starts somewhere between 2% to 60% and the last stage starts when the length of DNA remaining inside the phage is of the order of the phage diameter (50 nm). The intermediate slow rate is observed for a time period of 1000 LJ units to 10000 LJ time units.

**Figure 3.12.** Individual rate kinetics of ejection for the following packed conditions: (a) Force 55 pN, PL 60 nm (b) Force 55 pN, PL 20 nm (c) Force 20 pN, PL 60 nm (d) Force 5 pN, PL 60 nm.

Single phage experiments on T5 phage has yielded similar results (figure 3.2a-b, although a pause is observed instead of a kinetically trapped slow stage. The experimental setup consists of implanting the phages to the surface of a glass plate. The solvent consisting of the fluorophores and phage receptors, flows slowly over the plate. The receptors are required for phage channel opening and the fluorophores are required for detecting the exited DNA. The mild flow condition is needed for stretching the DNA for proper detection. It was observed that the channel opening
is a random event spread over minutes but the actual ejection event happens in a couple of seconds for pause free events. But majority of the ejection happens with intermediate pause lasting tens of seconds. The pause happens when the fraction of DNA remaining inside the phage is between 10\% – 60\%. The observation of a pause instead of slow ejection can be attributed to many reasons, the low resolution in the length and time scale, the flow condition, and the additional dynamics involved in the replacement of the counterions by fluorophores. Moreover, the last slow stage is not observed in their experiments most probably because of the flow condition.

We investigated the ejection kinetics for two different variables with phi29 capsid (a) persistence length of polymer chain (b) motor force, in order to determine the conditions when the intermediate slow kinetics doesn’t exist. The case of low motor force is considered for determining the kinetics of ejection from other phages in general and low persistence length polymer is considered for determining the role of chain stiffness in the ejection kinetics. Higher motor force was not considered because the packing kinetics saturates around 55 pN motor force (phi29) as mentioned in the packing section. The ejection kinetics of the system with lower persistence length (20 nm) didn’t show any pause. In fact the ejection happens in stages, pressure driven drift followed by diffusion. The pressure driven stage lasts till about 60 \% of the polymer is ejected. The crowding slowly vanishes hence the diffusion like behavior slowly creeps into the kinetics, replacing the pressure driven drift. During the pressure driven stage, the constant collision of the polymer beads with each other as well as the wall of the capsid leads to a strong driving force hence a drift like behavior. At later stage, there is less frequent collision of the polymer beads with each other as well as capsid. The less frequent collision allows for the polymer beads to explore within the available space inside the capsid before coming out of the capsid. Coming back to the effect of motor force on stiff chains, with decrease in the motor force, the number of events with intermediate slow kinetics decreased, almost completely
vanishing at 5 pN motor force. For 5 pN motor force, there are few traces that shows an intermediate slow kinetics, but the distinction between the traces are not very apparent, hence for all practical purpose the intermediate slow kinetics can be considered to be completely vanished.

In order to understand the reason for the slow intermediate ejection kinetics, the trajectory (figure 3.13) and the displacement versus time (figure 3.14) of three different beads along the DNA/polymer is tracked during the ejection. As it can be seen in the figure 3.13 3.14, the rigid DNA chain has to take circular trajectory, it can not take a sharp turn to come off the high pressure situation unlike the polymer of small persistence length. However, at low packing fraction the behavior of the less stiff chain and the stiff chain are similar. This can be explained by the availability of the space allowing both kind of chain to avoid sharp turns. Only at high packing fraction, the less stiff chains are able to take zig-zag (sharp turns) trajectory to get out of high pressure environment while the stiff chains doesn’t have the freedom to take zig-zag paths (sharp turns).

After spending considerable effort into the behavior of many different parts of the chain and the labeled beads, we have eventually found that the orientation of the beads in the neighborhood of the pore mouth is critical in dictating the local velocity of ejection. The angle between the beads near the pore mouth and the axis along the pore mouth indicates how bent the DNA is at the pore mouth. The angle was measured between the z-axis and the vector from the three adjacent beads (about to come out) just below the cylindrical pore mouth. For most of the trajectories, the DNA at high packed fraction remains bent, and waits for an opportunity to become less bent when the DNA packed fraction becomes low. This results in the intermediate slow kinetics. But sometimes the DNA gets an opportunity to become less bent at the pore mouth leading to smooth exit as shown in figure 3.15. The transition from highly bent to less bent near the pore mouth is purely stochastic. With decrease
in motor force, the packed DNA inside the capsid is less and as a result it has a higher probability to transition itself to a less bent form. Therefore, the percentage of trajectories with slow intermediate kinetics reduces with decrease in the motor force. Moreover, the system with low packing force is highly ordered, hence the stiff chain is mostly not bent near the mouth (not shown here) resulting in smoother exit. The lower persistence length polymer is able to come out smoothly, since the bending restriction is not so high.

In summary, we find that the jamming near the pore mouth to be the cause of pauses and intermediate slow rates in dsDNA ejection from phages. The higher the ordering and the lower the amount of chain packed, higher is the probability of transition of the conformation of the DNA near the mouth from bent to less bent state.

The averaging of kinetics does not gives an accurate representation of the actual kinetics because of very large standard deviation. Hence the average ejection dynamics was not pursued further.

### 3.4.3 Hysteresis

We observed a phenomena called hysteresis. The hysteresis shows the dissipative loss during a complete packing and ejection cycle (figure 3.16). The pressure inside the phage for a given packed fraction is higher for the packing cycle compared to the ejection cycle. The chain does not experiences any external force (due to motor) during the ejection phase, hence it has the freedom to arrange in such a manner that the chain is less bent and the inter strand repulsion is also less at a given packed fraction, relative to the packing process. Due to constant force from the motor during the packing process, the chain is unable to arrange to an energetically favorable state, more so for higher motor force. Hysteresis was estimated for a range of motor force varying from 55 pN to 20 pN (figure 3.17) and it was observed that with decrease in
Figure 3.13. Trajectory of three beads along the chain during ejection (lower value of bead represents the bead that comes out last) (a) Persistence length = 60 nm (b) Persistence length = 20 nm.
Figure 3.14. Displacement of three beads along the chain during ejection (lower value of bead represents the bead that comes out last) (a) Persistence length=60 nm (b) Persistence length =20 nm.
Figure 3.15. (a) The angle between the chain near the mouth with the pore axis (b) corresponding ejection kinetics (corresponding colors)
the motor force, the hysteresis also decreased, which is expected because the amount of chain packed decreases and the axial ordering increases with decrease in motor force. The hysteresis values for 10 pN and 5 pN motor force is not reported here because the average hysteresis was comparable to the standard deviation for these motor forces. The hysteresis was observed to be higher for the low persistence length at higher motor force because the amount of chain packed was higher for the low persistence length. With decrease in motor force, the hysteresis for both stiff and less stiff chains became comparable. Hysteresis has been observed in packing-ejection previously [47].

![Graph showing hysteresis in packaging-ejection cycle for dsDNA (55 pN)](image)

**Figure 3.16.** An example of hysteresis in the packaging-ejection cycle for dsDNA (55 pN)

In summary, hysteresis was observed during the packing-ejection cycle. The pressure at a given packed fraction was higher for the packing process compared to the ejection process because the polymer has more freedom to arrange itself during the ejection cycle due to the absence of a constant motor force. Hysteresis was observed to be decreasing with decrease in the motor force as expected.
3.5 Conclusions

The packing and ejection of dsDNA (persistence length of 60 nm) in phi29 (internal diameter of about 41 nm) virus was investigated using Langevin dynamics simulation. To understand the role of stiffness in the packing and ejection process, a chain with smaller stiffness (persistence length of 20 nm) was considered. Although the motor protein for phi29 exerts an average force of 55 pN, motor force values of 40, 30, 20, 10, and 5 pN were considered to understand the role of packing rate and the amount packed on the ejection process. The packing process is qualitatively the same for different motor force and chain stiffness and it happens in three stages: motor driven, motor driven against resistance, and saturation. If the packing is slow enough then the chain packed inside the phage is able to equilibrate until the next batch of chain gets inside. Hence the low motor force leads to slow and better ordered packing, whereas the packing with a high motor force leads to more beads packed but random
ordering. The pressure build up during the packing process and the ordering of the chain plays an important role during the ejection process. The ejection of the less stiff chain is qualitatively very different from the stiff chain. The less stiff chain ejects in two stages, mainly drift dominant regime at higher packing fraction and diffusion dominant regime at low packing fraction irrespective of the motor force. The less stiff chain is randomly ordered at the end of packing cycle, irrespective of motor force. However, the qualitative nature of the ejection of dsDNA is dependent on the motor force. With lower motor force, the DNA is highly ordered and the amount of DNA packed is low hence it ejects smoothly. With increase in motor force, the ordering of DNA at the end of packing cycle becomes more random. Hence, an intermediate slow kinetics is observed during the ejection. The intermediate slow kinetics is due to jamming (bent DNA at pore mouth). At intermediate pressure, the arrangement of the DNA near the mouth dictates the ejection rate. The more bent the DNA is, the slower is the exit. At lower packing fraction, the DNA has sufficient empty space to rearrange and at high packing fraction, the pressure is sufficient to push the DNA out, irrespective of the conformational state of DNA near the pore mouth. At intermediate packing fraction, the pressure inside the phage is not sufficient to push the DNA out (at a bent state near the pore mouth) and the empty space inside the phage is not sufficient to allow the arrangement of the DNA to allow smoother exit. Hence, jamming at the intermediate pressure can be attributed to the slow intermediate kinetics observed in phages.
CHAPTER 4
TRANSLOCATION OF SINGLE STRANDED DNA THROUGH MSPA PROTEIN PORE

4.1 Abstract

We investigated the translocation of DNA through Mycobacterium smegmatis (MspA) protein pore using Poisson-Nernst Planck computation and Langevin dynamics simulation. The chapter is divided into two sub sections. In the first part of the chapter we have resolved the recently reported experimental results on translocation of the DNA through MspA [1]. The time scale of translocation of the DNA through the pore was too small to be resolved from the blockade current experiments. It was not clear whether the DNA was successfully translocating or it was bouncing back to the cis chamber after probing the pore mouth. We resolved that the DNA translocates successfully most of the time. In the second part of the chapter we investigated the impact of two different methods: suitable mutations and application of alternating voltage signal on the translocation of DNA. The impact was measured in terms of the average translocation time and the width of the translocation time distribution. Both the methods lead to enhanced average translocation time but at the cost of wider translocation time distribution, which is not desirable. For proper sequencing using translocation, there are two main requirements: slow translocation for better signal to noise ratio and drift (not diffusion) of DNA (narrow translocation time distribution) so that each nucleotide can be detected only once. We conclude, therefore, that there is a need for another system before the MspA pore which has the inherent capability to slow down the DNA transport deterministically followed by translocation through M1MspA.
4.2 Background

Inspired by the translocation of biopolymers through biological channels, simpler systems have been designed to translocate isolated chains through pores [59, 60, 61, 62]. Since then protein pores as well as synthetic nanopores have been explored for possible application in genome sequencing. Recently MspA protein pore was explored for translocation of ssDNA [1]. MspA is obtained from mycobacteria. The protein pore is responsible for exchange of ions between mycobacteria and its surrounding. One MspA pore forms by the self-assembly of eight protein units on the membrane. In the experimental set-up eight units of the protein were assembled on a lipid bilayer to create one narrow channel (narrowest opening of about 1.2 nm). In the absence of any polymer the ionic current through the pore for an applied DC voltage is known as open pore ionic current. When the ssDNA translocates through the protein channel, the ionic current is reduced. The measured ionic current during the translocation of the polymer is called blockade current and it has been well characterized in the past for other protein channels as well for MspA [1]. With sufficiently narrow pore (pore diameter comparable to the size of the nucleotide), the individual bases can be distinguished based on their signature blockade current. The MspA pore is currently the best available channel because of its ability to distinguish all the bases in terms of their ionic current as shown in figure 4.1, because the narrowest end of the channel is about 1.2 nm (comparable to the projected diameter of a single stranded DNA) and the width of the narrow end of the channel is comparable to the size of the nucleotide (0.6 nm). The next best biological channel in terms of ionic current resolution of bases is alpha-hemolysin (AHL) [59], with a diameter at the narrowest end of channel of about 1.5 nm and the length of the narrow end of the channel is about eight nucleotides. Some terms that will be frequently used in this chapter are: blockade/dwell time, and translocation time will be defined. The blockade/dwell time is the time duration in which any part of the DNA is inside the pore. The translocation time is
the time duration in which the DNA blocks the narrow end of the channel leading to
lowest level of ionic current.

Figure 4.1. Hairpin experimental results [7] demonstrating the distinct blockade
current for different bases with MspA.

Although MspA is the best available pore in terms of ionic current resolution of
different bases, the translocation time is very small (about tens of $\mu$ seconds for a
ss DNA of 50 bases [1]). What it means, for example is that the translocation of
four different chains dT50, dA50, dC50, dG50 will show four distinct blockade cur-
rent, but a chain of dT25dC25 will not give two distinct blockade levels because of
the small time resolution. Reading one base at a time is a far fetched goal with the
current state of the pore. In order to resolve one base at a time, the current state of
art requirement is about millisecond per nucleotide. The difference in the blockade
current of the four nucleotides is very small (tens of picoAmps), hence this minimum
time requirement is for resolving the signal from the noise. The translocation time
for a DNA chain (dT50) through M2MspA is of the order of tens of microseconds.
Many strategies have been developed in the recent past to slow down the translo-
cation of polymer through these pores, one of them being engineering (site directed mutagenesis) the pore. In the context of MspA, the wild type protein channel is not practical for translocation experiments because at low applied voltage (60 mV) there is practically no translocation of the polymer and at high applied voltage conditions, there were spontaneous blockade of the pore even in the absence of polymers. The reason for this spontaneous blockade is unknown and we can not investigate the spontaneous blockade using our coarse grain modeling approach. However, they [1] resolved the situation by replacing the charged residues at the narrowest constriction of the pore to neutral residues (using site directed mutagenesis) resulting in a protein pore that doesn't get spontaneously blocked at higher applied voltage, hence making it more practical for translocation experiments. The problem with this engineered pore (M1MspA) is that the dwell time is too small (blockade time <30 µs) to even resolve whether the ssDNA is successfully translocating or bouncing back to the cis-side. To increase the translocation time the pore was further engineered to mutate some of the negatively charged residues near the pore mouth to positively charged residues. This led to a protein channel on which translocation experiments yielded sufficiently long dwell time (100 µs) but the translocation time (through the narrow constriction) was too small (<30 µs for 50 bases) to be resolved. To answer some of the experimentally unresolvable problems there is a need for modeling. The timescale of experiments was in tens of microseconds hence all atomic modeling of this system for such a long time is not feasible in a reasonable time frame with our current computational capabilities. Hence coarse-graining was used to answer some of the questions. The previous modeling work done on alpha hemolysin in our group [63] was a stepping stone for designing a more refined coarse-grained model for DNA translocation through MspA.

Another strategy for enhancing the translocation time is by application of time-dependent driving force [64] instead of a constant driving force. It was demonstrated
Figure 4.2. MspA structures and mutants: Original protein structure belongs to WTMspA, M1MspA is engineered from WTMspA by replacing the negatively charged residues (shown in red) to neutral residues (shown in grey) at the narrow end of the pore. M2MspA is created from M1MspA by replacing the negatively charged residues by positive residues (shown in blue). Replacement of residues are done by site directed mutagenesis. [1]
by Langevin dynamic simulation of a polymer chain translocating through a 2D nanopore, that the frequency of the alternating voltage has a significant effect on the average translocation time. If the average translocation time for an applied DC voltage value of $F$ is $\tau$, then by applying a time dependent dichotomic voltage $F + f(t)$ ($f(t)$ could be either $+A_d$ or $-A_d$ depending on the time), the average translocation time can be enhanced by orders of magnitude at certain conditions. For a frequency higher than $1/\tau$ and for an attractive pore average translocation time can be enhanced by multiple factors of $\tau$.

The goal of the project is two fold: understanding the dynamics of translocation of ssDNA for M1MspA and M2MspA, since the nature of translocation remains unresolved from the experimental studies [1] due to low time scales of blockade events. The second goal of the project is to investigate the performance of the pore with respect to the mutation and/or time dependent applied voltage. The performance of the pore will be measured in terms of the average translocation time and width of translocation time distribution of ssDNA through these protein pores.

### 4.3 Method

#### 4.3.1 Modeling Protein Pore

The all atomic information for the protein pore was obtained from the Protein Data Bank (PDB ID: 1UUN) [65]. A coarse grained model was built from the all atomic information. Each residue was represented by two coarse grain beads (except for glycine since it has no side group): backbone and side chain. The center of each bead is computed from the center of mass of the all atomic components constituting that coarse grained bead. The radius of the coarse grained bead is taken as the distance of the center of the mass of that group to the farthest atom constituting that group. The value of the radius computed from this method was averaged over all the backbones and all the different side groups to obtain a radius for backbone,
and all the side groups, respectively. A charge of $0e$ is assigned to all the beads except for the following side beads: D, E, R, and K. A charge of $-1e$ is assigned to D, and E and a charge of $+1e$ is assigned to R and K. The summary of the radius and charge information is tabulated below (table 4.1). The information for CYS is missing from the table because CYS was not found in the MspA. The same procedure was used to determine the radius of the backbone and side chains of alpha hemolysin [66] (PDB ID: 7AHL) for consistency test. The result were very similar to that obtained for MspA (not shown). No bond or angle information were recorded for the protein pore, because the protein will be kept static throughout the simulation. The protein pore’s orientation is set in such a way that the axis of the pore coincides with the x-axis, with the wider side of the pore on the positive side of the x-axis. The mid point along the axis of the pore was set at the origin (x=0). The y and z component of the center of mass of the pore were set at the origin (y=0, z=0).

4.3.2 Modeling Membrane

The membrane is modeled as two walls of monolayer neutral beads with bead radius 6.72 Å. Those walls are perpendicular to the axis of the pore and they are located at the narrow end of the pore and at the middle of the pore axis. Membrane beads within the pore were removed to allow passage of DNA through the pore. The length of the membrane along the y and z axis is equal to the length of the simulation box (box dimension will be specified later). The two monolayers serves the purpose of a bilayer lipid membrane barrier with minimal number of beads. The membrane will be kept immobile throughout the simulation.

4.3.3 Modeling ssDNA

The ssDNA is coarse-grained as three beads representing one nucleotide sequence: phosphate (charged), sugar and base (Thymine for this simulation) (figure 4.4). The strategy for obtaining the center and radius of the coarse grained beads are similar
Figure 4.3. (a) Coarse graining strategy for Histidine as an example (b) Coarse grain model for protein pore
Table 4.1. Coarse grain radius and charge of the backbone and side chains of MspA pore

<table>
<thead>
<tr>
<th>Component</th>
<th>Radius (Å)</th>
<th>Charge (e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone</td>
<td>4.2</td>
<td>0</td>
</tr>
<tr>
<td>ALA</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>SER</td>
<td>3.1</td>
<td>0</td>
</tr>
<tr>
<td>THR</td>
<td>4.3</td>
<td>0</td>
</tr>
<tr>
<td>CYS</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>VAL</td>
<td>4.4</td>
<td>0</td>
</tr>
<tr>
<td>LEU</td>
<td>4.5</td>
<td>0</td>
</tr>
<tr>
<td>ILE</td>
<td>5.4</td>
<td>0</td>
</tr>
<tr>
<td>MET</td>
<td>6.1</td>
<td>0</td>
</tr>
<tr>
<td>PRO</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td>PHE</td>
<td>7.4</td>
<td>0</td>
</tr>
<tr>
<td>TYR</td>
<td>4.2</td>
<td>0</td>
</tr>
<tr>
<td>TRP</td>
<td>5.1</td>
<td>0</td>
</tr>
<tr>
<td>ASP</td>
<td>4.3</td>
<td>-1</td>
</tr>
<tr>
<td>GLU</td>
<td>5.1</td>
<td>-1</td>
</tr>
<tr>
<td>ASN</td>
<td>4.3</td>
<td>0</td>
</tr>
<tr>
<td>GLN</td>
<td>5.1</td>
<td>0</td>
</tr>
<tr>
<td>HSD</td>
<td>5.6</td>
<td>0</td>
</tr>
<tr>
<td>LYS</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>ARG</td>
<td>6.9</td>
<td>1</td>
</tr>
</tbody>
</table>

to that of the Modeling Protein Pore section. The radius for phosphate, sugar and thymine are 3.0 Å, 4.0 Å, and 5.6 Å, respectively. Charge of 0e is assigned to the sugar and thymine and a charge of −1e is assigned to the phosphate group. Bond length and bond angle between the center of mass of different groups were calculated. The values are shown in figure 4.4. Potentials will be defined to keep the bond length and bond angles within reasonable limits during the simulation.

4.3.4 Simulation Summary

The first step in the simulation is obtaining voltage drop across the channel for a given applied voltage using Poisson-Nernst-Planck (PNP) calculation, followed by Langevin dynamics simulation of DNA translocation through the protein pore. The voltage calculated from the PNP solver will be used as an input parameter for the
Langevin dynamics simulation. The presence of a single DNA chain does not significantly influence the voltage profile across the channel (not shown), hence a decoupled Langevin-dynamics simulation and PNP calculation will be done. In this decoupled scheme, the voltage drop across the protein pore will be calculated across the channel without the DNA. The voltage drop thus obtained will be used for the Langevin dynamics simulation irrespective of the position of the DNA. Although the voltage drop across the protein pore is not significantly influenced by the presence of a single DNA chain but the current drop across the channel is significantly influenced by the presence of the DNA. In this decoupled scheme, the current drop across the pore will be calculated as a function of the DNA position (DNA position collected at regular intervals from the Langevin dynamics simulation).

4.3.4.1 Poisson-Nernst-Planck equation

The PNP solver is adapted from Roux group [67, 68, 69, 70]. The PNP equation is solved with and without the DNA chain. The case of no DNA is simple as everything is static except for the mobile ions. For the case involving DNA, the PNP equation is solved with an assumption that the small ions relax much faster than a large polyelectrolyte molecule and the concentration of electrolyte in the experiment is
very high in comparison with the monomer concentration. At regular intervals of the Langevin dynamics simulation, the electrolyte ions are assumed to be relaxed to the steady state and the polymer chain is taken only as a fixed charge distribution $\rho_p(r,t)$ at a time $t$. The maximum number of simulation runs over the frequency of data collection was 10000 and individual PNP calculations were finished in a couple of hours on a single processor on the cluster. The PNP equation is solved self-consistently for the local charge density $C_i(r,t)$ of the $i$th ionic species and $V(r,t)$ at any time $t$.

$$\nabla \left[ \nabla C_i(r) + \frac{Z_i C_i(r)}{k_BT} \nabla V(r) \right] = 0$$ (4.1)

$$\epsilon_0 \nabla [\epsilon(r) \nabla V(r)] = -\rho_f(r) - \rho_p(r,t) - \sum_i C_i(r)$$ (4.2)

$\rho_f(r)$ is the local charge density arising from the charge distribution on the fixed protein channel. The $\epsilon$ is set to 2 for the membrane, protein and the polymer whereas it is set to 80 everywhere else. The applied voltage is fixed at a box length of $-100\text{Å}$ and $100\text{Å}$ with the protein pore being fixed at the origin as mentioned in the Modeling Protein Pore section. The ion concentration is also fixed at 1M at $-100\text{Å}$ and $100\text{Å}$. The box length in the other two dimension is $125\text{Å}$, symmetric across the pore axis, periodic boundary condition is applied along these dimensions. The membrane boundary is set from $-46\text{Å}$ till $-1\text{Å}$ along the axis of the pore. Equations (4.1) and (4.2) are solved by successive over-relaxation method with a grid spacing of $1\text{Å}$. Using the successive over relaxation method, $V$, $C_+$ and $C_-$ are computed iteratively until they converge with a tolerance of $10^{-7}$ for voltage and $10^{-9}$ for concentration. The ionic current at any time can be obtained by

$$I(t) = \int d^2r (J_+ + J_-)$$ (4.3)
The ionic current is evaluated along the axis of the pore and the $D_+$ and $D_-$ are taken to be $1.96e^{-5}$ cm$^2$s$^{-1}$ and $2.03e^{-5}$ cm$^2$s$^{-1}$, respectively for $K^+$ and $Cl^-$ ions. Average value of the ionic current is reported.

### 4.3.4.2 Langevin dynamics simulation details

The force fields on the individual beads were computed by Langevin dynamics simulation in LAMMPS [33, 34].

\[
m_i \frac{d^2 r_{ij}}{dt^2} = -\zeta_i \frac{dr_{ij}}{dt} - \nabla_j U_i + f_{ij} + q_i E_j
\]  

(4.5)

where $r_{ij}$ is the position vector of the jth component of the ith bead, $t$ is the time. $m_i$ and $\zeta_i$ are the mass and friction coefficient of the ith bead respectively. $U_i$ is the net potential acting on the i-th bead, as given below. $f_{ij}$ is the j-th component of the random force acting on the i-th bead obeying the fluctuation-dissipation theorem with its magnitude given by $\sqrt{k_B T \zeta_i} dt$ ($k_B T$ is the Boltzmann constant times the absolute temperature). $q_i$ is the charge of ith DNA bead and $E_j$ is the space dependent electric field, calculated from the gradient of the voltage obtained from the PNP calculation.

The net potential acting on the i-th bead is the sum of all the non-bonded and bonded potentials.

\[
U = U_{LJ} + U_{DH} + U_{bond} + U_{angle}
\]  

(4.6)

Excluded-volume interaction between the DNA-DNA beads and DNA-capsid beads are modeled as

\[
U_{LJ} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] + \epsilon_c \quad r < r_c, \quad 0 \quad otherwise
\]  

(4.7)

where

\[
\epsilon_c = -4\epsilon \left[ \left( \frac{\sigma}{r_c} \right)^{12} - \left( \frac{\sigma}{r_c} \right)^6 \right]
\]  

(4.8)
and \( \epsilon \) are the parameters and \( r_c \) is set to 1.12\( \sigma \).

The pairwise electrostatic interaction potential between the \( i \)th bead of charge \( q_i \) and the \( j \)th bead of charge \( q_j \) separated by a distance \( r \) is assumed to be the Debye-Hückel potential. As explained in the simulation summary section, the effect of the DNA on the voltage across the pore was not significant hence the voltage drop across the pore was computed without the DNA. The voltage drop calculated from the PNP solver without the DNA is used as an input parameter for the Langevin dynamics simulation hence use of DH potential is justified.

\[
U_{DH} = \frac{q_i q_j}{4\pi \epsilon_0 \epsilon_r r} \exp((-\kappa r)) \quad r < r_c, \ 0 \ otherwise
\]  

(4.9)

where \( \epsilon_0 \) is the permittivity of the vacuum, \( \epsilon_r \) is the dielectric constant of the solution, and \( \kappa \) is the inverse Debye length.

Hookean bead spring model was used to represent the connectivity between adjacent beads in the DNA.

\[
U_{bond} = K_{bond} (r' - r_0)^2
\]

(4.10)

where \( r' \) is the bond length, \( r_0 \) is the equilibrium bond length and \( K_{bond} \) is the force constant for the bond. A three body interaction potential was used to set the stiffness of the chain.

\[
U_{angle} = K_{angle} (\cos \theta - \cos \theta_0)^2
\]

(4.11)

where \( \theta \) is the angle between any three beads, \( \theta_0 \) is the equilibrium value of bond angle, and \( K_{angle} \) is the force constant for the angular stiffness.

The position of the protein pore and the membrane is kept fixed throughout the simulation. The velocity and position of the DNA beads are updated at every timestep by velocity Verlet algorithm [56].

All variables are expressed in dimensionless Lennard-Jones (LJ) units, fully consistent with LAMMPS. To non-dimensionalize, the fundamental quantities of mass, length and energy are \( m_0 \) (taken to be 95g/mol) corresponding to the molar mass of
phosphate), $\sigma$ (taken to be 3 Å, the radius of the coarse grained phosphate bead), and $\epsilon$ (assumed to be $1 \, k_B T$), respectively. The other important quantities for non-dimensionalizing are expressed in terms of the fundamental quantities such as time ($\sqrt{\sigma^2 m_0 / \epsilon}$), force ($\epsilon / \sigma$), electric field ($\epsilon / [\sigma \sqrt{4 \pi \epsilon_0 \sigma \epsilon}]$), charge ($\sqrt{4 \pi \epsilon_0 \sigma \epsilon}$), and temperature ($\epsilon / k_B$).

$m_0 / \zeta$ was chosen to be 10 LJ time units. $\epsilon$ was chosen to be 1 unit. The $\sigma$ for all the beads are obtained from the values of the radius expressed in the modeling sections of protein, membrane, and DNA. The $\sigma$ is chosen to be $2 \times \text{radius}/1.12$, this choice of $\sigma$ ensures excluded volume interaction at the cutoff, $1.12\sigma$. The $\sigma$ thus obtained in real units, is converted to the non-dimensional LJ form by dividing the value by 3 Å. The Debye length $\kappa^{-1}$ is 1 LJ unit, corresponding roughly to 1M monovalent strong electrolyte solution at room temperature, and $1\epsilon = 14$ LJ units. $K_{\text{bond}}$ value of 1800 LJ units was used and it was sufficient to keep the bond length within 1 % of its original $r_0$ value. The $r_0$ value for phosphate-sugar, sugar-phosphate, and sugar-thymine are 1.3, 1.9, and 1.4 LJ Units, respectively. $\theta_0$ was set to 99 °, 91 °, 68 °, and 91 ° for phosphate-sugar-phosphate, phosphate-sugar-thymine, sugar-phosphate-sugar, and phosphate-sugar-thymine, respectively. $K_{\text{angle}}$ was chosen to be 100 units and it was sufficient to keep the angles within 5 % of the set $\theta_0$ values. Temperature was set to 1 unit and the timestep was chosen to be 0.005 units. Periodic boundary condition was used on a simulation box ranging from $-40$ to 66 units in the x-direction and $-22$ to 22 in the other two directions. Five hundred independent simulations were executed for every case. A ssDNA chain of dT50 at an applied voltage of 180 mV will be used for all the simulations, unless specified otherwise.

4.3.5 Obtaining Electric field from the Voltage

The voltage obtained from the PNP calculation is used as an input parameter for the Langevin dynamics simulation in the form of electric field. The electric field
is obtained by the following way: The voltage vs X was divided into two parts, one below 32.5 Å and the other above this limit of X. The calculated voltage versus X were fitted to sixth order polynomial functions in this limits, hence two sixth order polynomial equations are obtained for voltage as a function of X. The electric field was computed from the gradient of the voltage with respect to X (Electric field is assumed to be symmetric along the other two perpendicular directions). The gradient thus obtained has the units of $mV/Å$, which is converted to LJ units as described above. The non-linear X-dependent electric field thus obtained is used as an input parameter for the Langevin dynamics simulation.

4.3.6 Alternating voltage Application

In the later part of the simulation effect of Alternating voltage signal was investigated on the translocation of DNA. The alternating voltage is applied on top of a constant direct voltage signal of 112.5 mV. The amplitude of the AV (square signal was used) is 67.5 mV and the frequency was different for M1MspA and M2MspA. For M1MspA the voltage was switched to-fro 180 mV and 45 mV every 50 LJ timesteps and for M2MspA it was done every 125 LJ timesteps. For Langevin dynamics simulations, the corresponding electric field was used. The purpose of these simulations were to test the effects of alternating voltage on the average and standard deviation of translocation time distribution. The frequency was deliberately chosen to be higher than the inverse of the average translocation time at 180 mV. At low frequency, we do not expect any significant effect on translocation. Two possible scenarios will be considered for demonstrating our choice of high frequency case: in one scenario the chain is in the vicinity of the pore mouth when the voltage cycle is 180 mV, then the translocation of the chain will be very similar to the behavior at a direct voltage signal of 180 mV. In another scenario, the chain is in the vicinity of the pore mouth when the voltage cycle is 45 mV, then the chain will not enter the pore because of entropic
barrier, resulting in its diffusion to some other location. Ultimately, we will end up with a translocation time distribution very similar to the 180 mV direct voltage case. Hence high frequency alternating voltage signal was considered for the simulation.

### 4.3.7 Dielectric Mismatch Modeling

The choice of $\epsilon_r$ and $r_c$ in Eq. (5.11) is crucial for modeling the effect of mutation of protein on the translocation of the DNA. The initial choice of $\epsilon_r$ value of 80 everywhere yielded no significant difference between the translocation profile of DNA through M1MspA and M2MspA. The dielectric effect of water on the interaction between the DNA and the protein beads are negligible compared to the electric field driven drift of the DNA chain. In reality, the dielectric constant near the protein and DNA beads is much smaller (about 2). Hence $\epsilon_r$ value of 2 was chosen. The choice of $r_c$ was crucial for simulating the experimental results of Butler et al [1], where significant difference in the dwell time was observed for M2MspA relative to M1MspA. The dwelling time ratio of the dT50 in M2MspA to M1MspA is reported to be between 4.33 and 11. The choice of $r_c$ value of 1.20 yielded a dwell time ratio of M2MspA and M1MspA within the experimentally reported range. With a choice of lower value of $r_c$ we could not reproduce the experimental results and higher $r_c$ values resulted in DNA getting stuck to the walls of the protein pore.

### 4.3.8 Initial Configuration

The DNA was equilibrated separately and the equilibrated chain was placed near the pore mouth for the actual translocation simulation. The center of mass of the DNA chain was placed 5 units away from the pore mouth along the pore axis. In many simulations we were interested only in the histogram of translocation time distribution (not dwell time), in that scenario one end of the chain was put inside the mouth of the pore as a starting configuration.
4.3.9 Definitions

**Dwell time** Is the time the DNA spends inside the pore. The histogram for dwell time is created by computing the time gap between either end of the chain entering the cis side of the pore (13.5 LJ Units) and exiting from the trans side (-14.5 LJ Units) of the pore. Some chains might just probe the pore mouth and come back, hence the timer is started only when the chain goes little inside the pore. The timer is started only at X=13.5 LJ units and not 14.5 LJ units (the protein pore spans from 14.5 to -14.5 LJ Units along the X-direction).

**Translocation time** Is the time, the DNA spends in the neck of the pore. The histogram for translocation time is created by computing the time gap between either end of the chain entering the neck of the pore (-10 LJ units) and exiting to the trans side of the pore (-14 LJ Units).

4.3.10 Free Energy computation and Radial density measurement

The free energy of the polymer is calculated from the probability of finding monomers along the pore axis during translocation. The translocation time of the shortest strand of DNA (N=30) is two orders of magnitude higher than the equilibration time of the DNA (not shown). Hence the translocation process can be assumed to be in equilibrium. The free energy computation is based on the time spent by the monomers along the pore axis. The free energy as a function of the pore length($L$) is defined as the normalized time spent by the monomers along the pore axis $L$ ($n(L)/N$). Where $n(L)$ is the sum total of the time spent by all the monomers at a location $L$ along the pore axis and $N$ is the sum total of $n(L)$. Only successful translocation events are considered for the calculation. The free energy at the extreme end of the cis chamber is set as zero for reference.

\[
\frac{F(L)}{k_B T} = -ln \left( \frac{n(L)}{N} \right) \quad (4.12)
\]
The radial distribution profile of the phosphate group at $-9 \pm 1 \text{ Å}$ was investigated to determine the role of electrostatic attraction versus the entropy of the chain during translocation. The distribution of the phosphate groups in the mutation cross-section will determine the actual effects of mutation on the nature of translocation. The higher the probability of the beads near the mutated ring, higher is the enthalpic domination. For an ideal pore, introduction of mutation should introduce a delta distribution of the translocating monomers near the mutation ring. We will determine the impact of mutation on the translocation process by doing this analysis.

### 4.4 Results and Discussion

The Wild Type MspA (WTMspA) has been mutated at appropriate regions to create M1MspA and M2MspA as reported in the literature[1]. In addition to M1MspA and M2MspA, we will also consider M3MspA (Mutation of L88R of M2MspA) for our simulation. The protein pore is an octamer, hence the mutation L88R refers to mutation of eight residue along the pore diameter, creating a ring of positive charge.

#### 4.4.1 Voltage drop across the pore

The voltage drop across the pore is shown in figure 4.5. The voltage drop across the pore is not only sensitive to the geometry of the pore but also on the location of the charged ring. The voltage thus obtained is used as an input parameter of Langevin dynamics simulation after taking its gradient along the x-direction. The ionic current is generally overestimated by PNP calculation, even with all-atomic modeling [71]. Overestimation is attributed to the exclusion of the excluded volume effect of the ions and numerical error involved in overcounting of the ions if the ions are found in the grid border during calculation. With the coarse-grain model the ionic current estimates are even worse (figure 4.6). Hence the ionic current will not be used to estimate the translocation time, rather the position of the DNA chain along the pore
will be used for the dwell time and translocation time computation as mentioned in the Methods section.

Figure 4.5. Voltage drop across the pores. The voltage drop is sensitive to the location of the charged ring along the pore.

4.4.2 M1MspA vs. M2MspA

In order to observe the effect of the mutation, the dielectric mismatch was introduced as mentioned in the methods section. That leads to an average dwell time ratio of M2MspA to M1MspA value of 4.6, which is within the experimentally observed ratio of 4.33 to 11. The applied voltage of 180 mV is sufficient to translocate in most cases. Hence the experimental problem of Butler et al [1] is resolved that the translocation is indeed happening for both the M1MspA and M2MspA, more so for M2MspA because it is more attractive at the cis side. The choice of appropriate cut off value for the dielectric effect on the protein and DNA interaction was crucial for simulating the effect of mutation. By tuning, essentially one parameter we could
Figure 4.6. Ionic current across the pore as a function of applied voltage. The discrepancy between the experimental [1] and the model is very obvious.

simulate the mutation effect. This gave us the confidence to go further with our investigation on MspA. Effect of mutation and alternating voltage will be predicted for the translocation time distribution of dT50 through MspA. A typical translocation sequence is shown in figure 4.7, showing capture, threading and emptying stages.

4.4.3 Effect of the size of DNA and voltage on translocation through M1MspA

In order to test the generalized equilibrium translocation theories [72] on M1MspA, effect of the size of the DNA and the voltage on the average translocation time was investigated. The effect of the molecular weight (equivalently number of base sequence (N)) at constant applied voltage of 180 mV was investigated for five different N values: 30, 50, 70, 90, and 150. The effect of V on the average translocation time was investigated at constant N of 50 for four different voltages: 140 mV, 160 mV, 180
Figure 4.7. Snapshots during a translocation event.

mV, and 200 mV. The relationships, $\tau \sim N$ and $\tau \sim V^{-1}$ holds true for the M1MspA system as shown in figure 4.8a and 4.8b, respectively. Although the error bars are very high especially at low voltage. This agreement with the equilibrium translocation theory shows that the equilibration time of the DNA is smaller compared to the translocation time, hence the equilibrium translocation theories can be applied to predict the approximate average translocation behavior of DNA through this pore. Moreover we have computed the Rouse relaxation time for the DNA chain undergoing translocation (not shown) and it is found to be two orders of magnitude smaller than the translocation time. Hence the translocation can be assumed to be an equilibrium process within the explored parameter range.

4.4.4 M3MspA

With the mutation introduced on M2MspA to create a new pore M3MspA (figure 4.9), the average translocation time was observed to be enhance by a factor of 3.2 over M2MspA for dT50.
Figure 4.8. Dependence of translocation time through M1MspA on (a) Size of DNA ($V=180$ mV) (b) Applied Voltage ($N=50$). The red line represents the linear regression fit. Error bars are within the distribution.
The average value of the translocation time is increasing with the mutation but the width of translocation time distribution is also increasing, which is not desired. This wide distribution in translocation time indicates that the translocation time ratio of individual events of M3MspA to M2MspA can go up to a factor of 25. The histogram of the translocation time and dwell time for the three pores are presented below:

![Histograms of translocation times for M1MspA, M2MspA, and M3MspA.](image)

**Figure 4.10.** Translocation time histogram of (a) M1MspA (b) M2MspA (c) M3MspA.

The summary of the translocation time and dwell time is tabulated below:

In order to explain the increase in the average translocation time as well as the width of translocation time, we will consider the free energy and the radial distribution
Figure 4.11. Dwell time histogram of (a) M1MspA (b) M2MspA (c) M3MspA.

Table 4.2. Dwell time and Translocation time for dT50 at 180 mV through the three MspA pores

<table>
<thead>
<tr>
<th>Pore</th>
<th>$\tau_{Trans}$, LJ Units</th>
<th>$\tau_{Dwell}$, LJ Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1MspA</td>
<td>675 ± 576</td>
<td>1541 ± 755</td>
</tr>
<tr>
<td>M2MspA</td>
<td>4830 ± 3785</td>
<td>7093 ± 5482</td>
</tr>
<tr>
<td>M3MspA</td>
<td>15280 ± 11319</td>
<td>16808 ± 11708</td>
</tr>
</tbody>
</table>

function at the mutation cross section. Since the pore is becoming more and more attractive with increase in mutation from M1MspA to M2MspA and to M3MspA, hence the monomers are spending more time at mutation points. The cascading effect is felt at other regions of the pore since the monomers are connected. This leads to reduction in free energy as shown in figure 4.12. The free energy is getting deeper with the introduction of mutation, which is expected and desired.

However, the introduction of mutation does not ensure that all the phosphate groups will interact with the mutation site. If the pore mouth is wide then the entropy may still dominate during the translocation, resulting in a wider translocation time distribution. As shown in figure 4.13, the phosphate beads do not interact with the walls of M1MspA since the wall at that particular location ($x=-9 \text{ Å}$) is made up of excluded volume beads only. We expected to see a delta function like distribution with the introduction of mutation at $X=-9 \text{ Å}$ for M2MspA, but what we observe is that the probability of the phosphate groups near the vicinity of the protein wall has increased but not very significantly. Many of the phosphate groups are still not interacting with
Figure 4.12. The free energy of the DNA monomers translocating through the protein pores. Mutation site for M2MspA is shown in purple, and the blue color indicates the additional mutation introduced in M3MspA.
the mutation site during translocation. There seems to be a remnant of the density distribution of M1MspA in M2MspA and an additional peak near the protein wall. With the introduction of additional mutation (at $X=-28.5\ \text{Å}$) for M3MspA, there seems to be hardly any effect on the density distribution of the phosphate groups at the mutation cross-section ($X=-9\ \text{Å}$). Hence, making the pore more attractive by introducing mutation at other location ($X=-28.5\ \text{Å}$) does not ensure more intimate interaction of the phosphate group with the previous mutation site ($X=-9\ \text{Å}$). The geometry of the pore seems to play a major role in the case of MspA. The pore is very wide throughout, except for the neck region, hence entropy seems to be dominating all along the translocation through the pore mouth.

![Figure 4.13. Radial distribution of phosphate group in the mutation cross section ($x=-9\ \text{Å}$).](image)

Though we achieved our goal of enhancing the average translocation time with appropriate mutation, we ended up in situation (B) rather than (A) in figure 4.14.
The goal of increasing translocation time for genome sequencing is slowing down the process in a more deterministic way, but the standard deviation of the translocation time distribution are 576, 3785, and 11319 for M1MspA, M2MspA and M3MspA respectively. The diameter of the pore is very wide at the entry and the neck of the pore is limited to only a very small region. This leads to a situation, where all the DNA beads are not in intimate contact with the mutated regions. We fear, that the introduction of mutation near the neck of the pore might lead to a situation like WTMspA. The WTMspA had a ring of charged residues near the neck region and spontaneous blockade events were observed even without the presence of DNA rendering it useless for the translocation experiments. We do not have the tools for investigating the spontaneous blockade of the WTMspA, hence we did not explore any mutation near the pore neck. Some of the possible scenarios for the translocation events are shown in figure 4.15. The ideal scenario is case (a) of schematic figure 4.15, where the polymer chain interacts with the mutation site all along its transit. But there are other cases, where the chain does not intimately interact with the mutation site all along its transit. Some times, part of the chain interacts with the mutation site (figure 4.15b-c), and sometime the whole chain can go without interacting with the mutation site (figure 4.15d), although the probability for this situation is very low.

### 4.4.5 Effect of Alternating Voltage

We presumed that the introduction of alternating voltage will have a ratcheting effect on the translocation of DNA. During the high voltage cycle, the chain will move forward and during the low voltage cycle the chain will stick to the mutation site. We do observe increase in the average translocation time with the introduction of alternating voltage but the translocation time distribution becomes wider with this kind of alternating signal. We have previously investigated the effect of direct voltage
Figure 4.14. Schematic of the expected and real situation of mutation on MspA.
Figure 4.15. Schematic of the possible scenarios of interaction of the DNA with the protein pore (a) complete interaction of the DNA chain with the mutation site (b)-(c) Partial interaction (d) no interaction.
on the translocation behavior of dT50 through M1MspA (figure 4.8a). We observed that the average translocation time increases but the distribution becomes wider with reduced voltage. This feature may be particular to MspA pore, which has such a wide mouth that the entropy seems to dominant in most situations. The summary of the effect of alternating voltage is depicted in figure 4.16. The introduction of alternating voltage resulted in a change in the average translocation time from $675 \pm 576$ LJ time units to $1284 \pm 1250$ LJ time units for M1MspA. The average translocation time for M2MspA changed from $4833 \pm 3785$ to $6652 \pm 6208$. The ratcheting effect is not observed because the volume available inside the pore mouth allows conformational entropy of the captured DNA chain, resulting in more situations like schematic figure 4.15b-c depicted above.

Hence both the strategies: introduction of mutation on the pore, and application of alternating voltage on an attractive pore resulted in the increase of the average translocation time but increase in the width of translocation time distribution was also observed, which is undesirable. We have been careful in introducing mutation far away from the pore neck because the WTMspA consisted of charged residues near the neck of the pore and it was not very useful for translocation experiments due to spontaneous blockade events.

## 4.5 Conclusions

The translocation of ssDNA through MspA was investigated using PNP calculations and Langevin dynamics simulations. Previous experimental dilemma about the position of the DNA during the translocation experiment [1] is resolved now. It was not clear whether the blockade current signal from the translocation experiments was due to the successful DNA translocation or due to the probing of the DNA at the cis chamber of the pore followed by bouncing back to the cis side of the chamber. The ssDNA translocates successfully most of the time for both the M1MspA and
Figure 4.16. Summary of the effect of the alternating voltage on the translocation time distribution of (a) M1MspA: Direct voltage (b) M1MspA: Alternating Voltage (c) M2MspA: direct voltage (d) M2MspA: Alternating Voltage.
M2MspA. The introduction of mutation and application of alternative voltage as a strategy to enhance the translocation time worked well as the average translocation time increased but the width of the translocation time distribution also increased, which is not desirable. Hence there is a need for another system which has the capability to release DNA in a deterministic manner into the M1MspA, followed by translocation through M1MspA. Recent effort by Gundlach group [73] has been quite successful in detecting one base at a time by using polymerase to control the DNA release rate into the MspA pore.
CHAPTER 5

SEGMENTAL DYNAMICS OF A POLYMER TRANSLOCATING THROUGH A SYNTHETIC NANOPORE

5.1 Abstract

We investigated the validity of polyelectrolyte theory in the simulation of polyelectrolyte translocation through nanopores. The validity of the Debye Hückel (DH) approximation in the confinements of a nanopore is investigated by comparing the analytical solution of Debye Hückel equation with the numerical solution of non linear Poisson Boltzmann equation for a point charge across a nanopore. The DH approximation was found to be acceptable for most coarse grain model systems as long as the coarse grained bead size is greater than the Debye length. The next part of the project investigates the validity of the Rouse model to the segmental dynamics of a polyelectrolyte translocating through a nanopore, using Langevin dynamics simulation. Three different cases were considered: (a) polyelectrolyte length $<<$ nanopore length (b) polyelectrolyte length $\sim$ nanopore length (c) polyelectrolyte length $>>$ nanopore length. For all the cases an initial symmetric configuration of the polyelectrolyte was considered (the central bead of the polyelectrolyte coincides with the center of the nanopore). For unbiased translocation, Rouse model fails to predict the segmental dynamics of the polyelectrolyte translocating through a nanopore, except for the case where the polyelectrolyte length $<<$ nanopore length. The Rouse models makes a good qualitative prediction of the segmental dynamics of a polyelectrolyte through a nanopore in a biased field.
5.2 Background

In cellular transport, the property of the channel is often unknown. Determining the length of the channel and the voltage across the channel is often challenging. With the advent of fluorescence microscopy techniques, it is becoming possible to trace the motion of fluorescently tagged molecules inside the cells. Fluorescent tagging of polymer chains are routinely done and it is a general practice to tag only a part of the polymer chain and not the entire chain. In principle, it is possible to determine the length and voltage across the channel by measuring the mean square displacement of a polymer segment as a function of time. It is interesting to note that the behavior of a polymer segment changes dramatically when it moves from an open space to confinement during diffusion. The Rouse model predicts the dependence of mean square displacement \((msd)\) of a polymer segment on time \((t)\):

\[
< [R_i(t) - R_i(0)]^2 > \sim \begin{cases} 
t^{2\nu/(2\nu+1)}, & \text{for } t < \tau_R \\
t, & \text{for } t > \tau_R \end{cases}
\]  

(5.1)

where \(R_i(t)\) is the position vector of a polymer segment index \(i\) at time \(t\) and \(R_i(0)\) is the position vector of the segment at initial condition, \(\nu\) is the Flory exponent, and \(\tau_R\) is the Rouse relaxation time \((\tau_R \sim N^{2\nu+1})\). \(N\) is the number of Kuhn segment in the polymer chain.

Although the polymer segment has interesting dynamics, the dynamics of the center of mass of the polymer is not so interesting.

\[
< [R_{com}(t) - R_{com}(0)]^2 > \sim Dt
\]

(5.2)

where \(R_{com}\) is the position vector of the center of mass and \(D \sim 1/N\) is the diffusion coefficient of the polymer chain.

The Flory exponent \(\nu\) changes from 0.6 to 1 when the chain moves from a 3D space to a 1D channel. Moreover the \(msd \sim t^2\) for a voltage driven process. In
principle, this dramatic dependence of the $msd$ on time for a polymer segment can be used to identify the pore dimension and voltage across the pore, but there are various issues with the measurement. First of all the system inside the cell is too complicated and the polymer segment may interact with the various components of the cell. The Rouse model is valid only for an entropically driven system. Secondly, Rouse model is a simplified model for polymer dynamics where hydrodynamics is not taken into account and in the real systems hydrodynamics plays a major role. It is rightly pointed out by Muthukumar [72] that the Rouse model is generally used by the computational and theoretical community. But the Rouse model is used to analyze the translocation simulation without checking its validity. Switching gears, the Rouse model is a very important tool to identify whether a system has reached equilibrium or not. The dependence of mean square displacement of a monomer on time is very sensitive to time. The exponent of time is different for a equilibrated system from a non-equilibrium system, as discussed above. Equilibrium translocation theory is generally applied to interpret the simulation results without checking the equilibrium state of the translocation process. One way to identify the equilibrium state of the system is by determination of the Rouse relaxation time. If the translocation time is comparable to the Rouse relaxation time then the system can not be assumed to be in equilibrium. We will determine the validity of the Rouse model in some typical translocation simulations.

Secondly the Debye H"uckel (DH) theory is used by the translocation community for modeling the mean field electrostatic interaction. The mean-field approximation is good for a low potential and low ionic strength solution (The DH equation and Poisson-Boltzmann (PB) equation is described in detail in the methods section). Hence, in principle the DH equation can be used in many situations. But the analytical expression of DH equation does not have the capability to handle the geometrical constraints of a polymer translocating through a narrow channel. The mean-field
distribution of the salt ions within the cylindrical confinement may not be applicable. So how good is a DH approximation in the confinement of a cylinder? We will determine the validity of DH approximation by comparing the analytical solution of the DH theory with the numerical solution of the non-linear Poisson-Boltzmann equation (NPBE). In DH theory and Linearized Poisson-Boltzmann equation (LPBE), the coions as well as the counterions are assumed to be uniformly distributed in the medium. In NPBE no such assumption is made hence it is the most accurate description of the electrostatic potential.

5.3 Method

The PBE solver will be used to determine the validity of DH approximation in some specific scenario and Langevin dynamics simulation will be used to determine the validity of Rouse model in the segmental dynamics of a polymer translocating through a nanopore.

5.3.1 Poisson-Boltzmann Equation (PBE)

The PBE solver is adapted from CHARMM \[74, 75\]

\[
\nabla \epsilon \nabla \phi = -4\pi \rho_0 - 4\pi \sum_i e z_i c_i \exp \left( \frac{-ez_i \phi}{k_B T} \right)
\]

(5.3)

where \( \epsilon \) is the dielectric constant, \( \phi \) is the electrostatic potential, \( \rho_0 \) is the fixed charge density, \( e \) is the unit charge, \( z_i \) is the valence of ion type \( i \), \( c_i \) is the number density of the ion type \( i \), \( k_B \) is the Boltzmann constant, and \( T \) is the absolute temperature.

For a solution with symmetric 1:1 salt, Eq. (5.3) can be simplified to

\[
\nabla \epsilon \nabla \phi = -4\pi \rho_0 + \frac{\epsilon_{out} k^2}{C} \sinh(C\phi)
\]

(5.4)
where $\kappa^2 = \frac{8\pi z^2 I}{\epsilon_{\text{out}} k_B T}$ and $C = \frac{\epsilon z}{k_B T}$. Here “out” denotes the outside solvent, $I$ represents the ionic strength of the solution and $I = z^2 c$. If the electrostatic potential is weak and the ionic strength is low, the nonlinear PBE can be simplified to the linear form

$$\nabla \epsilon \nabla \phi = -4\pi \rho_0 + \epsilon_{\text{out}} \kappa^2 \phi$$

(5.5)

The linearized PBE is easier to solve but it is not very accurate in modeling highly charged systems, while the nonlinear PBE predictors have been shown to yield good agreement with experiments and explicit ion simulations [76, 77].

Generally, the analytical form of the Linearized LPBE equation is used for modeling and theory, which is known as Debye Hückel (DH) potential equation. The DH equation is described below and the solution of the DH equation should be the same as the LPBE in open space but they may be different in confinement due to boundary conditions. The LPBE is solved numerically hence the confinement is taken into account while solving it, but DH equation is generally solved analytically without taking into account the geometrical constraint factor.

$$\phi = \frac{q}{4\pi \epsilon_0 \epsilon r} \exp(-\kappa r)$$

(5.6)

where $q$ is the fixed charge, and $\kappa$ is the inverse debye length.

Nonlinear PBE and linear PBE will be solved numerically using CHARMM, and DH equation will be solved analytically for the following system, and the validity of the DH approximation inside the nanopores will be tested by comparing the LPBE solution and the analytical solution of DH equation with the solution of nonlinear PBE.

A solid wall of 30 Å X 30 Å X 20 Å (x y z) is created at the center of a cuboidal box of 30 Å X 30 Å X 80 Å (x y z). A cylindrical hole of 20 Å length and a diameter of 10 Å is created in the center of the solid wall with its axis along the z-direction.
A fixed boundary condition of 1M salt solution is used at z=-40 Å and z= 40 Å. Periodic boundary condition is applied at the x and y boundaries of the cuboidal box. The $\epsilon$ for the solid wall was set at 2.0 and a value of 80.0 was set for all other location. A point charge ($1e$) was fixed anywhere (x y z) in the solvent accessible domain. The system was discretized into cubical box of 2 Å$^3$ and the Linearized PBE was solved using Successive Over Relaxation (SOR) method [74, 75] until a tolerance limit of $2 \times 10^{-7}$ is reached. Initially, we started the computation with larger grid size and solved for LPBE. Gradually the grid size was reduced to 0.1 Å$^3$ using the previous solution as the initial guess. After reaching the desired grid dimension of 0.1 Å$^3$, the solution of the LPBE was used as an initial guess for the NPBE. The NPBE was solved using under relaxation method until a tolerance limit of $2 \times 10^{-6}$ is reached. The potential with respect to the fixed charge location is reported as a function of z. The analytical solution of the DH equation is reported for comparison. The schematic of the system is shown below:

![Figure 5.1. 2D representation of the schematic of the system for PB calculation.](image)

Figure 5.1. 2D representation of the schematic of the system for PB calculation.
5.3.2 Langevin dynamics simulation details

Different simulations will be conducted, with or without pore, with or without external voltage. Same sized coarse grained bead will be used to create both the polymer beads as well as the nanopore.

1. **Rouse model verification for an uncharged chain in free solution** A polymer chain of a suitable length (N=30, 50, 70, 100) will be created. The parameters for the chain is described in the following section. The chain will be equilibrated and the center of mass of the polymer chain will be placed at the center of the simulation box at initial condition. Since shrink wrapped non-periodic boundary condition will be used hence the choice of the simulation box dimension is not a concern, but the box dimension should be big enough to contain the polymer chain. The DH potential and the electric field part of the langevin equation will not be used. Five hundred independent simulation will be conducted and the mean square displacement of the segments of polymer chain as well as the center of mass of the polymer chain will be computed at regular time intervals. The radius of gyration ($R_g$) of the polymer chain will also be obtained for determining the Flory coefficient.

2. **Rouse model verification for a charged chain in free solution** A polymer chain of a suitable length (N=30, 50, 70, 100) will be created. The chain will be equilibrated and the center of mass of the polymer chain will be placed at the center of mass of the simulation box at initial condition. The electric field part of the langevin equation will not be used. Five hundred independent simulation will be conducted and the mean square displacement of the segments of polymer chain as well as the center of mass of the polymer chain will be determined at regular time intervals. The radius of gyration ($R_g$) of the polymer chain will also be obtained for determining the Flory coefficient.
3. **Rouse model verification for a charged chain in a long cylindrical confinement**

A cylindrical pore of length of 500 Lenard Jones (LJ) units and a radius of 2 LJ units will be created by placing uniformly placed monolayers of bead along the axis of the cylinder at a radial distance of 2 units from the axis. The diameter of the pore that is available for polymer access is 9 Å. A polymer chain of length (N=30, 50, 70, 100) will be created with its center coinciding with the center of the nanopore. The chain will be equilibrated before start of the simulation (the central bead of the polyelectrolyte will be fixed during the simulation) and the simulation box will encapsulate the cylindrical pore. The electric field part of the Langevin equation will not be used. Fifty independent simulation will be conducted and the mean square displacement of the segments of polymer chain as well as the center of mass of the polymer chain will be determined at regular time intervals. The radius of gyration ($R_g$) of the polymer chain will also be obtained for determining the Flory coefficient.

4. **Unbiased translocation of polymer chain**

A cylindrical pore of length of 20 LJ units and a radius of 2 LJ units will be created by placing uniformly placed monolayers of bead along the axis of the cylinder at a radial distance of 2 units from the axis. A polymer chain of suitable length (N=20, 40, 60, 100, 220) will be created at the center of the cylinder. The chain will be equilibrated before start of the simulation and the simulation box should be big enough to encapsulate both the chain and the cylinder. In real situations the cylindrical pore is embedded on a solid substrate, but we do not take that into account in our simulation for computational efficiency. The electric field part of the Langevin equation will not be used. One fifty independent simulation will be conducted and the mean square displacement of the segments of polymer chain as well as the center of mass of the polymer chain will be determined at regular
time intervals. The radius of gyration \((R_g)\) of the polymer chain will also be obtained for determining the Flory coefficient.

5. **Biased translocation of polymer chain** Only two chain lengths will be considered (\(N=20\), and 220) and the electric field part of the langevin dynamics equation will be used. All other details are similar to the description given in the previous section (Unbiased translocation of polymer chain).

The force fields on the individual beads were computed by Langevin dynamics simulation in LAMMPS [33, 34].

\[
m_i \frac{d^2 r_{ij}}{dt^2} = -\zeta_i \frac{dr_{ij}}{dt} - \nabla_j U_i + f_{ij} + q_i E_j
\]  

(5.7)

where \(r_{ij}\) is the position vector of the jth component of the ith bead, \(t\) is the time. \(m_i\) and \(\zeta_i\) are the mass and friction coefficient of the ith bead respectively. \(U_i\) is the net potential acting on the i-th bead, as given below. \(f_{ij}\) is the j-th component of the random force acting on the i-th bead obeying the fluctuation-dissipation theorem with its magnitude given by \(\sqrt{k_B T \zeta_i dt}\) \((k_B T\) is the Boltzmann constant times the absolute temperature). \(q_i\) is the charge of ith DNA bead and \(E_j\) is the electric field across the pore, assumed to be \(V/L_{\text{pore}}\), where \(V\) represents the applied voltage and \(L_{\text{pore}}\) represents the length of the nanopore channel.

The net potential acting on the i-th bead is the sum of all the non-bonded and bonded potentials.

\[
U = U_{LJ} + U_{DH} + U_{\text{bond}}
\]  

(5.8)

Excluded-volume interaction between the beads are:

\[
U_{LJ} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] + \epsilon_c \quad r < r_c, \quad 0 \text{ otherwise}
\]  

(5.9)

where

\[
\epsilon_c = -4\epsilon \left[ \left( \frac{\sigma}{r_c} \right)^{12} - \left( \frac{\sigma}{r_c} \right)^6 \right]
\]  

(5.10)

113
\( \sigma \) and \( \epsilon \) are the parameters and \( r_c \) is set to \( 1.12\sigma \).

The pairwise electrostatic interaction potential between the \( i \)th bead of charge \( q_i \) and the \( j \)th bead of charge \( q_j \) separated by a distance \( r \) is assumed to be the Debye-Hückel potential,

\[
U_{DH} = \frac{q_i q_j}{4\pi \epsilon_0 \epsilon_r r} \exp(-\kappa r) \quad r < r_c, \ 0 \text{ otherwise} \quad (5.11)
\]

where \( \epsilon_0 \) is the permittivity of the vacuum, \( \epsilon_r \) is the dielectric constant of the solution, and \( \kappa \) is the inverse Debye length.

FENE potential was used to represent the connectivity between adjacent beads in the polymer chain.

\[
U_{bond} = -0.5 K_{bond} R_0^2 \ln \left[ 1 - \left( \frac{r'}{R_0} \right)^2 \right] + 4\epsilon \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right) + \epsilon \quad (5.12)
\]

where \( r' \) is the bond length, \( R_0 \) is the maximum extent of the bond and \( K_{bond} \) is the force constant for the bond. The other terms are defined in the excluded-volume section.

The position of the synthetic nanopore and the membrane is kept fixed throughout the simulation. The velocity and position of the polymer beads are updated at every timestep by velocity Verlet algorithm [56].

All variables are expressed in dimensionless Lennard-Jones (LJ) units, fully consistent with LAMMPS. To non-dimensionalize, the fundamental quantities of mass, length and energy are \( m_0 \) (taken to be \( 206g/mole \) corresponding to the molar mass of a styrene sulfonate), \( \sigma \) (taken to be \( 3 \, \text{Å} \), the approximate coarse grained radius of a styrene sulfonate molecule), and \( \epsilon \) is assumed to be \( 1 \, k_B T \), respectively. The other important quantities for non-dimensionalizing are expressed in terms of the fundamental quantities such as time (\( \sqrt{\sigma^2 m_0/\epsilon} \)), force (\( \epsilon/\sigma \)), electric field (\( \epsilon/|\sigma \sqrt{4\pi \epsilon_0 \epsilon \sigma^3}| \)), charge (\( \sqrt{4\pi \epsilon_0 \epsilon \sigma} \)), and temperature (\( \epsilon/k_B \)).
\( m_0/\zeta \) was chosen to be 1 LJ time units, \( \epsilon \) was chosen to be 1 unit and \( \sigma \) was chosen to be 1 unit. The Debye length \( \kappa^{-1} \) is 1 LJ unit, corresponding roughly to 1M monovalent strong electrolyte solution at room temperature, and \( \epsilon_r \) value of 80 was used. \( K_{\text{bond}} \) value of 10 LJ units was used and \( R_0 \) value of 1.5 units was used. Temperature was set to 1 unit and the timestep was chosen to be 0.001 units. Shrink-wrapped non-periodic boundary condition was used with an initial box size big enough to contain all the components within.

5.4 Results and Discussion

5.4.1 Poisson-Boltzmann equation

Only one salt concentration case (1 M NaCl) will be considered throughout the chapter, since typical translocation experiments are conducted at this salt concentration. Two cases will be considered: a system with fixed point charge outside the confinement and another system with fixed point charge within the confinement. As shown in figure 5.2a, the solution of the LPBE and DH match as expected but they do not match exactly in the confinement case (figure 5.2b), which is expected because the LPBE although assumes equal distribution of the coions and the counterions in the medium but it does not assumes the system geometry. The system geometry is taken into account during the numerical solution of the LPBE. But DH equation is just an analytical equation and it should be used with caution if applied within confinement. When we compare the NPBE solution with the LPBE solution, the potential calculated near the point charge is much smaller compared to the LPBE. In NPBE, equal distribution of the coions and the counterions around the point charge is not assumed unlike the LPBE. The counterion is concentrated near the fixed point charge compared to the co-ions. But the distribution of the counterions and co-ions become more uniform as we move away from this fixed point charge. The distribution of the ions are more or less uniform up and above the debye length of the system (3
Å at 1 M NaCl concentration). Most of the translocation experiments and simulations are done at this salt concentration and the coarse grain models generally used in the simulation are normally above 3 Å, hence the DH approximation is quite acceptable for most of the simulations.

![Figure 5.2](image)

**Figure 5.2.** Solution of the NPBE, PBE, and DH equation for different position of the fixed point charge (1e) along the pore, blue color symbol represents NPBE solution, red color symbol represents linear PBE solution, and green color symbol represents the solution of the DH equation (a) Outside the confinement (x=4, y=0, z=16), (b) Inside the confinement (x=4, y=0, z=0).

### 5.4.2 Rouse Model Verification

Rouse model will be verified for the following system: (a) Neutral chain in free solution (b) Charged chain in free solution (c) Charged chain in a very long narrow pore. After verifying the Rouse model, a system will be considered where the polymer chain is partly confined and partly in the free solution, just like a situation during the polymer translocation. Only symmetric case will be considered where the center of the polymer chain will coincide with the center of the pore at the start of the simulation. Lastly the effect of the externally applied voltage on the MSD of the polymer segment will be investigated.
5.4.2.1 Neutral chain in free solution

The Flory exponent for the polymeric system was computed by determining the slope of \( \log(R_g) \) versus \( \log(N) \) as shown in figure 5.3a, where \( R_g \) is the radius of gyration of the polymer chain and \( N \) is the number of beads in the polymer segment. The average value of the Flory exponent \( \nu \) was found to be 0.622, which represents a good solvent condition for the polymeric system.

![Figure 5.3](image)

**Figure 5.3.** (a) Flory exponent determination for the neutral chain (dashed green line represents the linear fit of average value of \( \log(R_g) \) versus \( \log(N) \), a slope of 0.622 is obtained, (b) MSD of center of mass of chain versus time: the slope determines the diffusivity of the polymer chain, four different values of \( N \) were used (30, 50, 70, and 100) (c) Plotting \( \log(D) \) versus \( \log(N) \) to verify Rouse model: A slope of -0.98 is obtained which is very close to the expected slope of -1. Averaged over 500 independent runs.

The Rouse model was verified for the neutral chain by plotting diffusivity \( D \) versus \( N \) as shown in figure 5.3c. According to the Rouse model, the diffusivity of the center of mass of a polymer chain is inversely proportional to the size of the polymer chain. Larger the size of the polymer chain, slower is the diffusivity. \( D \) is obtained from the slope of \( MSD \) of the center of mass of the chain versus time \( t \) (figure 5.3b). \( D \sim \frac{1}{N} \) holds true for the polymer chain hence the Rouse model is verified for the center of mass of the chain.

The validity of the Rouse model on the individual polymer segments was tested by plotting \( MSD \) versus \( t \) for individual beads. According to the Rouse model, the
polymer segment moves slowly \((msd \sim t^{\frac{2\nu}{2\nu+1}}})\) if the time of observation is below the Rouse relaxation time \(\tau_R\), and above \(\tau_R\) the dependence of \(msd\) on \(t\) is linear just like the center of mass of the chain. Hence above \(\tau_R\), it will be difficult to distinguish the motion of the polymer segment and the center of mass of the polymer chain based on the \(msd\) versus \(t\).

With a \(\nu\) value of 0.622, the expected slope for \(\log(msd)\) versus \(\log(t)\) for a polymer segment is 0.554 for \(t\) below \(\tau_R\), and a slope of 1 is expected above \(\tau_R\). We could not verify Rouse model for all type of chains and all the beads along the segment. A slope of higher than the expected value (0.55) is observed for the beads at either chain ends for \(t < \tau_R\). Higher values of the slopes at the chain ends have been previously observed \([78]\). This anomaly was attributed to chain end effect. Similarly the slope of the \(msd\) versus \(t\) for a central monomer for a small polymer segment \((N=30)\) was found to be higher than the expected value of 0.55 as shown in figure 5.4a. This anomaly was attributed to the finite size effect. An expected slope of 0.55 is obtained for the central part of the chain for a large polymer segment as shown in figure 5.4b. This anomaly is most likely due to the following reason: in the Rouse model the segment is expected to slow down due to random collision with its neighboring segments. When we consider a polymer segment at the chain ends or a middle segment of a small polymer than the segments are not experiencing the collision from all the sides, hence they are not as slow as expected. But the central part of a long polymer chain experiences random collision from all the directions as it is surrounded by a lot of segments from all the directions. Hence Rouse model should be applied with caution to a polymer chain. Above \(\tau_R\), all the polymer segments have a \(msd\) versus \(t\) slope closer to 1. Hence all the polymer segments are following Rouse model for \(t\) above \(\tau_R\).
5.4.2.2 Charged Chain in free solution

The Rouse model was verified for a charged chain in the free solution and the result is presented in figure 5.5.

5.4.2.3 Charged Chain in a long cylindrical confinement

The Rouse model was verified for a charged chain (N=100) in a long cylindrical confinement (L=500 LJ Units) and the result is presented in figure 5.6. The polymer chain with its

5.4.3 Charged Chain in a small cylindrical confinement: Unbiased

Polymer chain is placed in the pore with its center coinciding with the center of the pore (Pore length=20) as a starting configuration. Five different polymer chain lengths were considered (N=20, 40, 60, 100, and 220) to investigate the dynamics of the segments without any bias. The msd of the polymer chain shows some interesting behavior as a function of time. Below $\tau_R$ ($\sim$ 100 LJ time units), all the chains behave like Rouse chain with log(msd) versus log(time) slope of about 0.67, as expected.

Figure 5.4. MSD versus time for the central monomer for (a)N=30: a slope of 0.66 is observed below $\tau_R$ and a slope of 0.99 is observed above $\tau_R$, $\tau_R$ is about 100 LJ time units. (b)N=100: a slope of 0.55 (expected slope=0.55) is observed below $\tau_R$ and a slope of 0.97 (expected slope=1) is observed above $\tau_R$, $\tau_R$ is somewhere between 100 to 1000 LJ time units. Averaged over 500 independent runs.
Figure 5.5. (a) Flory exponent determination for the charged chain (dashed green line represents the linear fit of average value of $\log(R_g)$ versus $\log(N)$, a slope of 0.622 is obtained, (b) MSD of center of mass of chain versus time: the slope determines the diffusivity of the polymer chain, four different values of $N$ were used (30, 50, 70, and 100), (c) Plotting log(D) versus log(N) to verify Rouse model: A slope of -0.96 is obtained which is very close to the expected slope of -1 (d) MSD versus time for the central monomer for chain length of 100: a slope of 0.55 (expected slope=0.56) is observed below $\tau_R$ and a slope of 0.97 (expected slope=1) is observed above $\tau_R$, $\tau_R$ is somewhere between 100 to 1000 LJ time units. Averaged over 500 independent runs.
Figure 5.6. (a) Flory exponent determination for the charged chain confined in a long nanopore (dashed green line represents the linear fit of average value of \( \log(R_g) \) versus \( \log(N) \), a slope of 1.0 is obtained, (b) MSD of center of mass of chain versus time: the slope determines the diffusivity of the polymer chain, four different values of N were used (30, 50, 70, and 100), (c) Plotting log(D) versus log(N) to verify Rouse model: A slope of -1.06 is obtained which is very close to the expected slope of -1, (d) MSD versus time for the central monomer for chain length of 100: a slope of 0.70 (expected slope=0.67) is observed below \( \tau_R \) and a slope of 1.0 (expected slope=1) is observed above \( \tau_R \); \( \tau_R \) is about 500 LJ time units. Averaged over 50 independent runs.
Above the $\tau_R$, the short chains and long chains behave completely differently. The short chains (N=20 and N=40) drift even without any external bias. If either end of the chain diffuses little farther away from the pore due to thermal fluctuations then the opposite end of the chain is pulled into the pore. The chain end that has been pulled into the pore can not go back to the free solution because that will lead to entropic penalty for the free chain at the other end. Moreover, there is no pulling force from its own side, since almost everything at its end is inside the pore. The only way out is to translocate through the pore. Since there is nothing to hold it back, therefore the segments drift through the pore. For the large polymer, there seems to be a tug of war between both ends. The system does not gains a lot of free energy by pulling in beads into the nanopore from either end because of the large size of the polymer hanging from each end of the nanopore. The system never reaches equilibrium, since the slope of log(msd) versus log(t) yields a slope of about 0.8 after the Rouse relaxation time. Hence equilibrium theories can not be applied to either short or long chains diffusing through a pore because the short chain drifts through the pore without any external bias and the long polymer never reaches the equilibrium coefficient.

### 5.4.4 Charged chain in a small cylindrical confinement: External bias

Two different polymer chains (N=20, and 220) will be considered to investigate the effect of externally applied voltage signal on the segmental dynamics. For N=20, the dynamics of the middle bead is shown in figure 5.8. The chain drifts, no mater what the applied voltage is, although the drifting starts earlier with increase in the applied voltage. The relaxation time of the polymer chain is reduced with increase in applied voltage. For unbiased case, the polymer segment is already equilibrated when it comes out of the nanopore (above the dotted line in figure 5.8), since the slope is 1 at the exit. The segment is expected to be relaxed at the exit as there is no external
Figure 5.7. MSD of the central bead as a function of time for different N values (20, 40, 60, 100, 220) (averaged over 150 independent runs).
force to induce non-equilibrium. But for biased case, the segment relaxes only after spending some time in the bulk (the exponent at the exit \( \sim 0.55 \)). After relaxing for 100s of LJ time step, the exponent reaches 1 indicating a relaxed segment. Larger the voltage, larger is the relaxation time at the exit.

![Figure 5.8](image)

**Figure 5.8.** MSD of the central bead of a polymer chain (N=20) as a function of time for different applied voltage (averaged over 150 independent runs).

With a large polymer chain (N=220), the segment does not drifts without the external applied voltage unlike the short chains (N=20). The chain shows drift like behavior (slope \( \sim 2 \)) under biasing condition. The segment starts drifting sooner with increase in the applied voltage. For the same applied voltage, the chain starts drifting earlier for a shorter chain compared to a longer chain. When we compare
Figure 5.9. MSD of the central bead of a polymer chain (N=220) as a function of time for different applied voltage (averaged over 150 independent runs).
the green colored plot (V=50 mV) from figure 5.8 and figure 5.9, the drifting started much before 10 time units for N=20, but for N=220, the drifting started only after 20 time units. This behavior is expected, since a longer chain offers more resistance to motion. Lastly, the polymer segment is non-equilibrated when it comes out of the nanopore. Above the dotted line (nanopore end) in the figure 5.9, the slope is about 0.55 and the slope is expected to reach 1 (equilibrium slope) only after some time (beyond our simulation time). The larger the voltage, larger is the time for equilibration. The black and red colored lines in figure 5.9 corresponding to 0, and 10 mV show a slope of 1 at the exit (above the dotted line) indicating an equilibrated system.

For both the cases (N=20 and N=220), the Rouse model is valid for a segment undergoing translocation under biased condition. The snapshots of a long polymer chain undergoing translocation is shown in figure 5.10 below:

In summary, caution must be taken while applying the Rouse model to the translocation of a polymer chain through a nanopore under unbiased condition. The Rouse model is valid only for a small polyelectrolyte chain translocating through a very long nanopore. The Rouse model fails to predict the segmental dynamics of a polyelectrolye chain which is either comparable or larger than the nanopore in terms of length. The Rouse model is applicable to the segmental dynamics of a polyelectrolyte undergoing translocation through a nanopore under biased condition.

5.5 Conclusions

The validity of the Debye Hückel approximation in the confinements of a nanochannel was investigated by solving non linear Poisson Boltzmann equation for a point charge across various location of the nanopore. By comparing the analytical solution of DH equation with the NPBE, the DH approximation was found to be valid as long as the coarse grained bead size was comparable or greater than the Debye length. The
Figure 5.10. Snapshots at different stages of a long polymer chain (N=220) undergoing translocation through a synthetic nanopore (L=20).
investigation was done only at 1 M KCl condition which is typical condition for most translocation experiments. The validity of Rouse model was tested for the dynamics of a polymer segment undergoing translocation through a nanopore, using Langevin dynamics simulation. The polymer chains that are equivalent to the size of the pore are found to have unusual translocation behavior. They are found to drift even without any external field. Very large chains never reach equilibrium under unbiased condition, because of the tug of war happening between the overhanging polymer at both ends of the pore. Under unbiasing condition, the Rouse model is valid only for a small polyelectrolyte chain translocating through a very long nanopore. The Rouse model fails to predict the segmental dynamics of a polyelectrolyte chain undergoing translocation under unbiased condition, which is either comparable or larger than the nanopore in terms of length. The Rouse model seems valid for segmental dynamics of a polyelectrolyte undergoing translocation through a nanopore under biased condition.
CHAPTER 6
CONCLUSIONS

We investigated the interaction of polyelectrolyte with proteins and nanopores in the context of four different problems: self-assembly of virus, ejection of DNA from phages, translocation of DNA through MspA protein pore, and translocation of polyelectrolyte through synthetic nanopores.

In the first investigation, we determined the mechanism of the self-assembly of virus (with and without polyelectrolyte) using Langevin dynamics simulation and Replica exchange computation. We constructed suitable coarse grained model of viral subunit and polyelectrolyte. The subunits self-assemble to form a closely assembled structure in a narrow temperature range. The effect of parameters (temperature, concentration of subunits, and presence of polyelectrolyte) were investigated in this narrow temperature range. By investigating the effect of temperature and concentration of subunits on the assembly kinetics of virus (without polyelectrolyte), we determined that the self-assembly of virus follows nucleation-growth kind of mechanism. The growth mechanism can be subdivided into a linear growth regime followed by a slow growth regime. In the presence of a polyelectrolyte, the same mechanism is followed but the nucleation time is reduced as well as the linear growth rate is increased because of the increase in the local concentration of the subunits due to the adsorption of the subunits on the backbone of the polyelectrolyte.

In the second investigation, we determined the dynamics of DNA ejection from phi29 phage using Langevin dynamics simulation on a suitable coarse grain model of DNA and phi29 phage. Our simulation results show significant variations in the local...
ejection speed, consistent with experimental observations reported in the literature. In efforts to understand the origin of such variations in the local speed of ejection, we have investigated the correlations between the local ejection kinetics and the packaged structures created at various motor forces and chain flexibility. For a higher motor force, typical of a realistic phage the amount of DNA packed is high and the system is more disordered relative to a low motor force case. At high packing fraction the pressure inside the phage is sufficient to eject the DNA smoothly. The ejection of the DNA is smooth at low packing fraction because of the availability of more empty space inside the phage for DNA rearrangement. However, at the intermediate packing fraction the pressure inside the phage is not high enough to eject the DNA out smoothly once the DNA is stuck at a bent conformation at the pore mouth. Moreover, the DNA does not have enough empty space for its rearrangement inside the phage so that it can exit smoothly. In this scenario we observe a slow intermediate ejection until sufficient empty space is created within the phage for DNA rearrangement or favorable DNA conformation is reached during this slow exit process. Hence slow intermediate DNA ejection kinetics is attributed to the DNA jamming at the pore mouth.

In the third investigation, the translocation of single stranded DNA through MspA protein pore is investigated using PNP computation and Langevin dynamics simulation. Several methods have been suggested in the literature to slow down the translocation process which is desired for genome sequencing. We investigate two such methods: mutation of the protein pore and the use of alternating voltage. Use of both the techniques, mutation as well as the alternating voltage approach resulted in an enhanced average translocation time, but the translocation time distribution also became wider. For genome sequencing a deterministic process is required so that each nucleotide along the DNA strand can be detected only once, diffusion may lead to erroneous sequencing. There is a need for a deterministic process before the
translocation of DNA through MspA. Recent effort in the literature has been directed towards using polymerase [13] to slow down the translocation process as well as to make the translocation process deterministic.

The fourth investigation, determines the validity of the polyelectrolyte theory in the translocation simulation of polyelectrolyte through nanopores using Poisson Boltzmann solver and Langevin dynamics simulation. We investigated the validity of Debye Hückel approximation in the confinement of nanopores by comparing the analytical solution of DH potential equation with the numerical solution of non-linear Poisson Boltzmann equation (NPBE) potential for a point charge across the nanopore. We demonstrated that the Debye Hückel approximation is valid as long as the coarse grain bead is larger than the Debye length. We also investigated the validity of the Rouse model on the segmental dynamics of a polyelectrolyte chain undergoing translocation through a synthetic nanopore. We demonstrated that the qualitative behavior of the segmental dynamics agrees well with the Rouse model when the polyelectrolyte is translocating under a biased field. The Rouse model fails to predict the segmental dynamics of a polyelectrolyte chain undergoing translocation in unbiased condition, except for the case when the polyelectrolyte is much smaller than the nanopore in terms of length.


