Experimental study and molecular dynamics (MD) modeling of a nucleic acid in nano-confinement

Çağla Okyay

Thèse de doctorat

NNT: 2024UPASF063

université PARIS-SACLAY

Experimental study and Molecular Dynamics (MD) modeling of a nucleic acid in nano-confinement

Etude expérimentale et en modélisation de dynamique moléculaire d'un acide nucléique en nano-confinement

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°571 : sciences chimiques : molécules, matériaux, instrumentation et biosystèmes (2MIB)
Spécialité de doctorat : Physique
Graduate School : Chimie. Référent : Université d'Évry Val d'Essonne

Thèse préparée dans l'unité de recherche **LAMBE** (Université Paris-Saclay, Univ Evry, CNRS), sous la direction de **Jérôme MATHÉ**, Professeur des universités, le co-encadrement de **Nathalie BASDEVANT**, Maîtresse de conférences

Thèse soutenue à Paris-Saclay, le 06 décembre 2024, par

Cagla OKYAY

Composition du Jury

Membres du jury avec voix délibérative

Luca MONTICELLI Directeur de Recherche, Université de Lyon	Président du Jury
Elisa FREZZA Maîtresse de conférences, HDR, Université Paris Cité	Rapporteur
Christophe YBERT Directeur de Recherche, Université Claude-Bernard-Lyon-1	Rapporteur
Tâp HA DUONG Professeur des universités, Université Paris-Saclay	Examinateur
Fabien MONTEL Chargé de Recherche, ENS de Lyon	Examinateur
Marco PASI Maître de conférences, Université Paris-Saclay	Examinateur

Contents

1 State of the Art **9**
- 1.1 Overview of Nanopore Technology . . . 9
  - 1.1.1 αHL for Nanopore Technology Applications . . . 11
  - 1.1.2 Importance of DNA Translocation and Unzipping Studies . . . 13
- 1.2 Experimental Studies of DNA transport . . . 14
  - 1.2.1 DNA Translocation Studies . . . 14
  - 1.2.2 DNA Unzipping Studies . . . 23
- 1.3 Theoretical studies . . . 25
  - 1.3.1 Continuous models . . . 25
  - 1.3.2 Molecular Dynamics Simulations . . . 29

2 Methods **37**
- 2.1 Experimental Methods . . . 37
  - 2.1.1 Single-molecule experiments, experimental technique . . . 37
  - 2.1.2 Experimental Setup . . . 39
  - 2.1.3 *IV* Curves and Their Experimental Measurement . . . 40
  - 2.1.4 Analysis of Translocation Events in Nanopore Experiments . . . 42
  - 2.1.5 DNA Unzipping Experiments . . . 44
- 2.2 Computational Methods . . . 50
  - 2.2.1 Molecular Dynamics Simulations . . . 50
  - 2.2.2 Ensembles in SMD simulations . . . 56
  - 2.2.3 Steered Molecular Dynamics (SMD) . . . 58
  - 2.2.4 All-atom and Coarse-Grained Molecular Force Fields . . . 58

3 Experimental study of DNA translocation and unzipping using aHL **73**
- 3.1 Experimental model and parameters . . . 74
- 3.2 Effect of orientation on translocation time . . . 76
- 3.3 Effect of the duplex structure on unzipping time . . . 78
- 3.4 Effect of the duplex length on unzipping time . . . 85
- 3.5 Conclusion . . . 94

4 Poly(dA) translocation using CG-SMD **97**
- 4.1 Cf-SMD system preparations and simulation parameters . . . 99
  - 4.1.1 Simulation Setup for ssDNA Molecules . . . 99
  - 4.1.2 Preparation of the ssDNA-Pore System . . . 100
  - 4.1.3 Steered Molecular Dynamics Simulations . . . 103

4.2 Calculating ssDNA Translocation Time per Base . . . 103
4.2.1 Translocation time distribution for 3' oriented ssDNA. . . . 105
4.2.2 Influence of ssDNA orientation on translocation time. . . . 107
4.2.3 Influence of ssDNA charges on translocation time . . . 109
4.3 DNA-nanopore interactions during translocation . . . 110
4.4 DNA tilting angles during translocation: 3' vs 5' . . . 113
4.5 Conformations of ssDNA during translocation . . . 117
4.6 Conclusion . . . 123

5 Study of the Effect of Sequence and Pulling Force on Translocation Dynamics Using CG-SMD 125
5.1 CG-SMD System Preparations and Simulation Parameters . . . 126
5.1.1 Simulation Setup for ssDNA Molecules . . . 126
5.1.2 Steered Molecular Dynamics Simulations . . . 127
5.2 Influence of sequence on translocation time . . . 127
5.2.1 Translocation time distributions of poly(dA) and poly(dC) molecules 128
5.2.2 DNA-pore interactions depending on the sequence . . . 130
5.2.3 DNA Tilt Angle Differences Depending on the Sequence . . . 132
5.2.4 Conformational differences depending on the sequence . . . 135
5.3 Influence of Pulling Force on Translocation Time . . . 136
5.3.1 Distributions of Translocation Time Per Base Under Varying Pulling Forces . . . 137
5.3.2 The relationship between pulling force and translocation velocity . . 139
5.3.3 Influence of pulling force on ssDNA conformations . . . 141
5.4 Conclusion . . . 148

List of Figures

1.1 Figure taken from Bonthuis *et al.*. Experimental setup with αHL inserted into a membrane and surrounded by ionic solution. Applying a voltage across the system induces ionic flux through the pore. The presence of DNA molecules in one compartment causes them to translocate through the pore, resulting in characteristic current blockages. 10

1.2 The crystal structure of αHL from the side (on the left) and top view (on the right), each chain is indicated in a different color. The diameter, D, of the inner wall of the β-barrel channel is given in the right figure. 12

1.3 Proposed insertion mechanism of αHL in the lipid bilayer. αHL monomer binds to the target cell membrane. αHL monomers begin to oligomerize on the membrane surface, assembling into a heptameric structure composed of seven monomers, called the pre-pore complex. The pre-pore undergoes a conformational change, driving the insertion of the heptameric structure into the membrane. 13

1.4 Figure taken from Kasianowicz *et al.*. The poly(U) (0.1 mg/ml; mean length, 210 nucleotides) produced channel blockades with three distinct characteristic lifetimes. These lifetimes were analyzed by fitting three Gaussian distributions to the data, corresponding to the three peaks observed. Peak 1 represents non-translocation attempts, where the polymers collide with the pore but fail to pass through. Peaks 2 and 3 correspond to successful translocation events occurring with different orientations (3' or 5'). 15

1.5 Figure taken from Kasianowicz *et al.*. Poly(U) channel blockade lifetimes were observed to be (a) proportional to the mean polymer length and (b) inversely proportional to the applied voltage. The plots illustrate the lifetimes for (a) peaks 1 (+), 2 (□), and 3 (•) in experiments conducted with a -120 mV applied voltage, using 13 different size-selected poly(U) polymers, and (b) for peaks 2 (□) and 3 (•) for poly(U) with a mean length of 215 nucleotides at the indicated voltages. 16

1.6 Figure taken from Akeson *et al.*. Representative current traces recorded from the α-hemolysin pore in the presence of poly(A), poly(C), and poly(U) homopolymers. The average lengths of the homopolymers were approximately 130 for poly(C), 175 for poly(A), and 150 nucleotides for poly(U). . 17

1.7 Key findings of Meller *et al.* on the dynamics of DNA translocation. On the left, the quadratic relationship between polymer velocity and applied voltage is depicted. Data points are shown for poly(dA) with lengths of 12 (•) and 30 (□) nucleotides at 2°C. The lines represent quadratic fits to the data, and error bars are provided except for data points below 90 mV, where insufficient events occurred to estimate errors. On the right, the plot shows the dependence of the most probable translocation time, t_P, on polymer length, N. Measurements were performed for poly(dA) (•) and poly(dCdT) (■) at 2°C. For polymers longer than $N = 12$, the translocation time scales proportionally with polymer length, while a sharp transition is observed for shorter polymers ($N < 12$). The inset shows the average polymer velocity, calculated from t_P and translocation trajectories. 18

1.8 Figure taken from Stoddart *et al.*. Schematic representation of a homopolymeric DNA oligonucleotide (blue circles) immobilized inside the αHL nanopore (grey, cross-section) via a 3' biotin-TEG (yellow) streptavidin (red) complex. The nucleobase recognition sites (R1, R2, and R3) within the β-barrel of the nanopore are shown. 19

1.9 Figure taken from Cockroft *et al.*. Monitoring DNA polymerase-catalyzed single-nucleotide primer extensions using a single-molecule nanopore device. (a) Initially, the DNA primer is interlocked within the αHL–DNA–PEG complex, where its length is measured via ion current readings in the monitoring mode (+40 mV). (b) In the elongation mode (−30 mV), the 3' end of the primer becomes accessible to DNA polymerase. (c) The DNA polymerase binds to the primer-template complex with the correct incoming deoxynucleotide triphosphate (dNTP). (d) The catalytic activity of the DNA polymerase incorporates a single dNTP into the primer against the DNA template threaded through the pore. (e) Successive base incorporation steps result in the templated extension of the primer sequence. 21

1.10 Figure adapted from Mayer *et al.*. Some of the biological nanopores used in single-molecule nanopore experiments. The top view and the side view inserted in a lipid membrane are given for each biological nanopore. 23

1.11 Figure taken from Mathé *et al.*. The figure shows characteristic timescales as a function of applied voltage for different hairpin structures: the perfect match hairpin (HP1, solid circles), the single mismatch hairpin (HP2, triangles), and the 7-bp hairpin (HP3, squares). Exponential fits (lines) approximate the data well, with the same slope for all hairpins. The intercept values are approximately $\tau_0 \sim 2.1 \pm 0.2$ s for HP1, 1.2 ± 0.1 s for HP2, and 0.34 ± 0.05 s for HP3. 24

1.12 The current-voltage (IV) relationship of αHL, taken from Aksimentiev *et al.*. Open circles represent currents computed under pH 8.0, while solid squares correspond to pH 4.5. Semi-solid squares indicate simulations where one of the loops at the trans end of the channel detached, momentarily obstructing the pore entrance. The dashed line represents a linear fit through the data point at 120 mV and the origin. Consistent with experimental findings, the IV curve shows sublinear behavior at $V < 0$ in 1 M KCl. 31

1.13 Figure taken from Mathé *et al.*. The tilt of the DNA bases inside the αHL pore depends on the global orientation of the strand: molecules oriented with their 3' end forward benefit from an upward tilt of their bases (b), whereas those with their 5' end forward experience a downward tilt (c). . . . 32

1.14 Figure taken from Dessaux *et al.*. *IV* curves for the WT αHL. The asymmetric current response is shown in green, with lines representing linear regression fits for both positive and negative potentials. 36

2.1 The resistance of the channel is the sum of the resistance from the nanopore and the surroundings, which is called the access. Insulating membrane is shown as gray rectangles. 39

2.2 Illustration of the experimental setup. The cis side is magnified to show details. An α-hemolysin is inserted in a thin lipid bilayer, and DNA molecules are translocated through the pore under an applied voltage. 41

2.3 Current-Voltage relationship of αHL nanopore at 20°C at pH 7.5 in 1M KCl solution. The signature values of αHL nanopore at +120 mV, 110 pA and at -120 mV, -90 pA are shown on the plot. 42

2.4 Current Trace during the recording of experiments with HP39sf molecules. Open pore current (I_0), Blocked pore current (I_b), Time between events (t_i), and Translocation duration (t_u) are shown. 43

2.5 Current Trace and the distribution of the current during the recording of experiments with HP39sf molecules. On the plot of the distribution of the current trace, three peaks are visible: I_0, $I_{undesired}$, and I_b. 44

2.6 Design of the experimental dsDNAs: HP9s and HP9sf sequences fold into hairpin structures. The difference between the s and sf structure is coming from the presence of a short unpaired sequence of a poly(dA) structure present in the sf structure which is called a flap. 45

2.7 Design of the experimental dsDNAs.The O60 sequence serves as a template for pairing with longer sequence strands than 9 bases. The same color is used to indicate the identical sequences used to pair longer sequences with the O60 sequence. Each time a new sequence of 10 nucleobase is added to make the hybridized portion longer. 47

2.8 Illustration of the opening mechanisms in the presence of a flap (left) and blunt-end (right). In the presence of a flap structure, the mechanism of opening is likely to occur in a zipper-like fashion. In the absence of this flap, when a blunt-end is present, the opening is likely to occur as a result of shear forces. 48

2.9 Histograms of (a) normalized blocked current (b) The time between the start of each event, defined as the inter-event time (t_i), which is fitted by an exponential decay function $Ae^{-\lambda t}$, λ being the frequency of events, A being a constant (c) Probability, represented as $1 - e^{-t/\tau}$ of unzipping corresponding to a HP19s under 120 mV bias voltage 49

2.10 Bonded and non-bonded interactions considered in MD simulations. 51

2.11 Representation of the Lennard-Jones potential as a function of the distance r_{ij} between two atoms i and j. The parameters σ_{ij} and ε_{ij} are indicated on the graph according to equation 2.12. $r_{\min}$, the distance between two particles at which the LJ potential energy reaches its minimum, is also shown in red. 53

2.12 Periodic boundary conditions in two dimensions, illustrating the replication of the system in all directions. How $r_{cut-off}$ is chosen by respecting PBC conditions is shown. 54

2.13 Molecular Dynamics iterative cycle, as explained in Verlet algorithm. Each step is illustrated and represented. The Potential Energy U is calculated and the Force **F** exerted on each particle is calculated as performed in step 2. The acceleration **a** is calculated to update the velocities and positions of each atom. 56

2.14 (a) All-atom and MARTINI (b) coarse-grained representations of α-hemolysin, pdb code: 7ahL. VDW representation in VMD is used in VMD to show the decreased number of beads present in the CG representation 60

2.15 Chemical structure of (a) DPPC and (b) DPhPC lipids, methyl groups on the two chains of DPhPC are highlighted in red 63

2.16 MARTINI representation of a DPPC molecule with twelve interaction sites of four types. 64

2.17 Representation of martini amino acids. 65

2.18 W and PW water model of MARTINI 2 force field. 67

2.19 ssDNA chemical structure. DNA bases are shown as *. 69

2.20 DNA base pairing. Deoxyribose sugars are shown as *. 70

2.21 Representation of a DNA nucleotide in the MARTINI model. 71

3.1 2 state theory for the unzipping process. E_b is the energy required to pass from state 1 (closed) to state 2 (open). Applied voltage V reduces the energy barrier by $-q_eV$ where q_e stands for the effective charge. 75

3.2 Results of the ssDNA translocation experiments. The upper figure shows the experimental translocation time distribution for poly$(dA)_{50}$ for both orientations with the αHL nanopore at 120 mV and room temperature. The scatter plot of events, event cloud, is represented in the lower figure. Each dot represents an event (about 12,500 events on this plot) with its time of translocation per molecule and the residual current (I_b/I_0, with I_0 being the open pore current and I_b the blocked pore current when the molecule is translocating through the pore). Both current and time are distinguishable between 3' oriented and 5' oriented ssDNA molecules. The 3' and 5' events were separated using an MC algorithm based on the double Gaussian distributions in the current-time plot of the cloud of events. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events. 77

3.3 Scatter plots of HP9s at different voltages at room Temperature 22°C at pH 7.5 in 1M KCl solution. From 80 mV to 180 mV, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events. 79

3.4 Current trace of poly$(dA)_{50}$ at 22°C at pH 7.5 in 1M KCl solution. Translocation time was calculated as 281 μs. The start and the end of the translocation are given in red and blue dashed lines, respectively. 80

3.5 Comparison of unzipping times, t_u, extracted from the Gaussian fits of the time value histograms for HP9s and HP9sf experiments under 120 mV applied voltage. The t_u of the HP9sf molecule is shorter than that of the HP9s molecule under the same experimental conditions. 81

3.6 Scatter plot of HP9sf at different voltages at room Temperature 22°C at pH 7.5 in 1M KCl solution. From 80 mV to 140 mV, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events. 82

3.7 The unzipping times t_u obtained from HP9s and HP9sf are plotted against the applied voltage on a logarithmic scale. Both datasets were fit linearly on the logarithmic time scale to calculate q_e, where the slope provides $\frac{-q_e}{k_B T}$. The results from the linear fits are as follows: For HP9sf (blue line), the fit is $\log(t_u) = -0.03 \times V + 9.30$ with an R^2 value of 0.98. For HP9s (green line), the fit is $\log(t_u) = -0.02 \times V + 10.90$ with an R^2 value of 0.99. . . . 84

3.8 The probability function introduced in Equation 3.5 for different duplex lengths. $\sigma(\log(t_u))$ kept constant and $\mu(\log(t_u))$ assumed to be proportional to the duplex length L for the illustration, with a linear function: $t_u = k \times L$, k being a random constant. As the duplex length gets longer, the distribution shifts to the longer t_u values and gets wider on linear time scale. 86

3.9 Scatter plot of HP9s, 19s, 29s at 120 mV applied voltage at room Temperature 22°C at pH 7.5 in 1M KCl solution. From HP9s to 29s, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events. 87

3.10 Histogram of the unzipping times t_u on logarithmic binning from the 39sf experiments. The t_u values spread along the time axis, varying between 10^2 to $10^8 \mu s$. The t_u was calculated as 2600 $\pm$ 820 μs. 88

3.11 Scatter plot of HP9sf, 19sf, 29sf, 39sf at 120 mV applied voltage at room Temperature 22°C at pH 7.5 in 1M KCl solution. From HP9sf to 39sf, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events. 89

3.12 Relationship between E_b and the duplex length was investigated using the results from `mfold` for blunt-end and flap structures. The slopes for blunt-end structures and for flap structures were calculated to be 2.0 and 2.2, respectively. 91

3.13 Unzipping times t_u obtained from HP9s, 19s, 29s and HP9sf, 19sf, 29sf, 39sf are plotted against the duplex length in structure on a semi-logarithmic scale. Both datasets were fit on the semi-logarithmic time scale according to equation 3.7, except for 39sf, to calculate n_0. The results from the fits are as follows: for the flap structures of different duplex lengths (blue line), the fit is $\log(t_u) = 0.12 \times n + 5.20$ with an R^2 value of 0.92. For the blunt-end structures of different duplex lengths (green line), the fit is $\log(t_u) = 0.40 \times n + 5.10$ with an R^2 value of 0.99. 92

3.14 Unzipping times t_u obtained from HP9sf, 19sf, 29sf, 39sf are plotted against the duplex length in structure in linear scale. Fit parameters for the first three points were found as $t_u = 230 \times l - 1640$ with a R^2 value of 0.99 . . 93

3.15 New design of the experimental dsDNAs.The template sequence (O60) serving for pairing with longer sequence strands than 9 bases has been changed and adapted for each length to prevent having a free single-stranded sequence from the 5' end. The same color is used to indicate the identical sequences used to pair longer sequences with the adapted sequence. Each time, a new sequence of 10 nucleobases is added to make the hybridized portion longer, with the adapted base sequence called O19 or O29, depending on the sequence studied. 95

4.1 ssDNA simulation for a poly(dA) DNA molecule of 10 nucleotides in the presence of an electric potentiel difference of approximately 800 mV. The applied electric field is not sufficient to drive the ssDNA molecule through the nanopore, and the ssDNA molecule interacts with the cap of the nanopore. Snapshot is taken after 500 ns. 98

4.2 The illustration of the initial system requirements. The distance between the COM of the ssDNA extremity and the COM of the pore constriction should be smaller than 6 nm, and the entry angle, which was defined as the angle between two vectors: Vector 1 one from the COM of the ssDNA 3' or 5' end nucleotide to the COM of the pore constriction ring (E111/K147), and Vector 2 from the COM of the pore constriction ring to the COM of the pore bottom (D127/K131), is limited to a maximum of 10°. 101

4.3 CG representation of the system at its initial conformation with 10-nucleotide poly(dA) molecule. The DPPC lipid bilayer is depicted with brown points, ions with cyan and purple points, poly(dA) in yellow beads, and the α-hemolysin nanopore in surface representation. Water is not shown. 102

4.4 Surface representation of coarse-grained αHL nanopore. Residues of the central constriction (E111/K147), and bottom (D127/K131) are shown in cyan and purple respectively. 105

4.5 Distribution of translocation time per base of the ensemble of all CG cf-SMD simulations of charged ssDNA pulled from 3' end, on logarithmic time scale. The average of the translocation time per base is shown in red dashed lines. 106

4.6 Translocation time per molecule as a function of the ssDNA length in cf-SMD simulations. The average translocation time per molecule is given for each system, and the standard error is shown as error bars. No data is available for 18-nucleotide poly(dA), and only one total translocation event was observed for 19-nucleotide poly(dA). 107
4.7 Distribution of translocation time per base of ssDNA molecules of 16 nucleotide length for both orientations with logarithmic binning representation. The average translocation times per base for 3' and 5' strand orientations are indicated by red and purple dashed lines, respectively. 109
4.8 Distribution of translocation time per base of ssDNA molecules of 16 nucleotide length for both neutral and charged systems with logarithmic binning representation. The average translocation times per base for the neutral and charged ssDNA molecules are indicated by blue and red dashed lines, respectively. 110
4.9 Contact maps of (a) 3'-A_{16}-1 (b) 3'-A_{16}-1 (c) 3'-A_{16}N-1 and (d) 3'-A_{16}N-2 with each ring in the stem during 1 μs MD simulation as a function of time. Dashed lines show the end of each translocation event. The cumulative contacts every 20 ns are given with a color code for visualisation. 111
4.10 Representation of the tilt angle (α) defined between the center of mass of BB1, BB3, and SC3 beads of the ssDNA molecule, represented in cyan, orange, and purple beads, respectively. 5' and 3' extremities are indicated in the illustration. 114
4.11 Evolution of the tilt angles inside the nanopore during ssDNA translocation for both orientations. Average tilt angles of DNA bases relative to the backbone for 3' end and 5' end oriented 16-nucleotide poly(dA) molecules are given as red and blue circles as a function of positions in the z-direction inside the simulation box. The displayed values represent the mean tilt angle for the 14 bases (excluding bases at both extremities) across all translocation events in the two simulation replicas, accompanied by their standard error. The black dashed lines denote the constriction and bottom of the stem. 115
4.12 Evolution of the tilt angles inside the nanopore during ssDNA translocation for both orientations. Average tilt angles of DNA bases relative to the backbone for 3' end and 5' end oriented 19-nucleotide poly(dA) molecules are given as red and blue circles as a function of positions in the z-direction inside the simulation box. The displayed values represent the mean tilt angle for the 17 bases (excluding bases at both extremities) across all translocation events in the two simulation replicas, accompanied by their standard error. The black dashed lines denote the constriction and bottom of the stem. 5'-A_{19}-1 simulation was extended for another 200 ns to achieve one complete translocation for this analysis. 117
4.13 Positions in the z-coordinate of the bases of 3'-A_{16}-1 over time. The stem is depicted as a blue-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147). The inset focuses on the positions of bases in the z-direction at the entrance of the nanopore before the constriction ring during the third translocation event. 118

4.14 Evolution of the base positions in z-direction of 3'-A_{16} systems. 119

4.15 Inter-base distance observed in two control simulations: ssDNA free in solution and ssDNA dragged in solution by the pulling force from 3' end. . 120

4.16 Inter-base distances for 3'-A_{16} simulations. The distance between two successive bases (inter-base distance) for all pairs except the first and last ones is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 representing the pore constriction ring (E111/K147). Left panel: Probability density of inter-base distance before the entry of the DNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: Probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm). 121

4.17 Inter-base distance of the first pair of bases for 3'-A_{16} simulations. As the SMD force is applied to the first base, this inter-base is overstretched. . . 122

4.18 Inter-base distances for 5'-A_{16} simulations. The distance between two consecutive bases (inter-base distance) for all pairs except for the first and last one is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 on the z-axis corresponding to the pore constriction ring (E111/K147). Left panel: probability density of inter-base distance before the entry of the ssDNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm). 122

5.1 Distribution of translocation time per base of 10, 13 and 16 nucleotide length of poly(dA) (blue) and poly(dC) (orange) molecules pulled from 3' end, on logarithmic time scale with a force of 400 kJ mol^{-1} nm^{-1}. The average of the translocation time per base is shown in yellow and blue dashed lines for poly(dC) and poly(dA) molecules, respectively. 130

5.2 Contact maps for 3'-C_{16}-400 simulations showing the interactions of poly(dC) ssDNA with each ring in the stem of the αHL nanopore during a 1 μs MD simulation. Panels (a) and (b) represent two replicas of the simulation with the color scale used for poly(dA) for visualizing cumulative contacts every 20 ns, while panels (c) and (d) display the same data with a color scale adapted to the data. The dashed lines indicate the end of each translocation event. Both scales demonstrate that the constriction (E111/K147) forms the most significant interactions, while differences in overall interaction intensity can be better appreciated with the alternative color scale in panels (c) and (d). 131

5.3 Representation of the tilt angle (α) defined between the center of mass of BB1, BB3, and SC3 beads of the ssDNA molecule, represented in cyan, orange, and purple beads, respectively. 5' and 3' extremities are indicated in the illustration. 133

5.4 Evolution of tilt angles inside the nanopore during ssDNA translocation for poly(dA) and poly(dC) molecules. Average tilt angles of DNA bases relative to the backbone for both 16-nucleotide poly(dA) and poly(dC) molecules are shown as red and blue lines, respectively, as a function of z-position in the simulation box. The values represent the mean tilt angle for 14 bases (excluding terminal bases) across all translocation events in two simulation replicas, with error bars representing standard error. The black dashed lines denote the constriction and bottom of the stem. 134

5.5 Inter-base distances for 3'-C_{16}-400 simulations. The distance between successive bases (inter-base distance) for all pairs, except the first and last ones, is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 representing the pore constriction ring (E111/K147). Left panel: Probability density of inter-base distance before the entry of the DNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: Probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm). 135

5.6 Distributions of translocation times per base of (a) poly(dA) (b) poly(dC), on logarithmic time scale under varying pulling forces. The average of the translocation times per base is shown in dashed lines for varying pulling forces, with the same color code: green, blue, orange, yellow, and red for 450, 400, 350, 300, 250 and 200 kJ mol^{-1} nm^{-1}, respectively. 138

5.7 Comparison of translocation velocities of poly(dA) (red squares) and poly(dC) (blue circles) under different pulling forces on a logarithmic velocity scale. The standard error can be consulted in Table 5.5. The semi-logarithmic plot shows the exponential fit for poly(dA) and poly(dC) based on $v = A \times e^{bF}$, where F is the pulling force (in kJ mol^{-1} nm^{-1}), and v is the translocation velocity (in Å ns^{-1}). For poly(dA), the values of A and b were calculated as $A = 1.41 \times 10^{-4}$ Å ns^{-1} and $b = 0.0213$ nm mol kJ^{-1}, while for poly(dC), $A = 7.48 \times 10^{-4}$ Å ns^{-1} and $b = 0.0186$ nm mol kJ^{-1}. 140

5.8 Inter-base distances for 3'-C_{10}-F simulations under various pulling forces (F-values). For simplicity, the axis labels are shown on the first probability density plot. The distance between two successive bases (inter-base distance) for all pairs except the first and last ones is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 representing the pore constriction ring (E111/K147). Left panel: Probability density of inter-base distance before the entry of the DNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: Probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm). 142
5.9 Evolution of the probabilities of each inter-base state before entering the pore as a function of the pulling force. 144
5.10 Positions in the z-coordinate of the bases of 3'-C_{10}-F simulations for F = 200 and 250 kJ mol^{-1} nm^{-1} over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147). 145
5.11 Positions in the z-coordinate of the bases of 3'-C_{10}-F simulations for F = 300 and 350 kJ mol^{-1} nm^{-1} over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147). 146
5.12 Positions in the z-coordinate of the bases of 3'-C_{10}-F simulations for F = 400 and 450 kJ mol^{-1} nm^{-1} over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147). 147

Introduction

The single-molecule (SM) technique is a powerful tool for providing valuable information about the properties, molecular mechanisms, and interactions of individual biomolecules that are obscured by the bulk. SM studies enable investigating the equilibrium or non-equilibrium processes, rare events, short-lived or transition states, and mechanical and thermodynamic information can be deduced. Techniques such as fluorescence resonance energy transfer (FRET) for studying conformational changes or biochemical processes, atomic force microscopy (AFM) for probing the energies involved in molecular interactions, optical tweezers to tether the forces involved in various processes, and nanopore technology for extracting kinetic parameters and study the intrinsic properties of the molecules have been developed for SM studies. Among these, nanopore technology stands out due to its distinct advantages, including label-free detection, real-time analysis, and the ability to analyze a wide variety of macro- and bio-molecules under diverse conditions. These features make it especially useful for applications like DNA sequencing, one of its most popular uses. In nanopore experiments, an electric potential is applied across an insulating membrane where a nanopore is inserted, inciting ions to flow through and generate an ionic current. When a molecule passes through the nanopore, it partially or totally blocks the nanopore, creating a characteristic reduced ionic current that depends on the molecule's properties as well as the experimental conditions. By monitoring these alterations in the current signal, it is possible to study the size, structure, or conformation of molecules in real-time at small costs.

The α-hemolysin (αHL) nanopore, a toxin which is produced by *Staphylococcus aureus*, is widely used in nanopore experiments due to several reasons such as its well-characterized structure, its inner pore dimensions, and its hydrophilicity. It forms a mushroom-shaped heptamer consisting of a cap and a stem region. The use of nanopores, particularly αHL, has led to significant advances in molecular sensing and sequencing, with early studies demonstrating the ability to detect nucleic acids, DNA or RNA, and other biomolecules by evaluating translocation events. For instance, proteins and peptides can be detected and characterized through this nanopore. By analyzing current blockages, proteins' conformational changes, folding dynamics, and ligand binding interactions can be identified as they pass through the nanopore. The use of αHL has also extended to the detection of small molecules and polymers. Studies have shown that polyethylene glycol (PEG) and other synthetic polymers can also be differentiated based on their size and shape from the current blockages during their passage through the nanopore. This ability to detect synthetic molecules has expanded the potential applications of nanopores beyond biological systems into areas like material science and drug delivery.

The study of DNA translocation through αHL remains central in nanopore technology, particularly for sequencing applications. Early experiments demonstrated that nucleotides

can be distinguished by their current blockage levels or translocation times, which are influenced by diverse factors. These factors include intrinsic properties of the molecules, such as length, sequence, structure, and orientation within the nanopore, as well as characteristics of the nanopore itself, including its charge distribution, diameter, and chemical composition. Additionally, experimental conditions like temperature, ionic concentration, and applied voltage also influence current blockage levels and translocation times.

While these experimental techniques have yielded valuable information, they cannot unveil the complete molecular mechanisms of molecule transport at the microscopic level. Therefore, different theoretical and computational approaches have been developed to complement experimental findings, including continuum models, such as Poisson-Nernst-Planck (PNP) theory, Brownian dynamics, and molecular dynamics (MD) simulations. Early MD studies successfully provided a valuable understanding of ionic transport and DNA translocation through nanopores, confirming the influence of DNA molecule length, orientation, sequence composition, and pore charge distribution. However, a large system is needed to model a DNA molecule passing through a pore inserted into a membrane. Modeling such systems using all-atom models limits the simulation time. Knowing that the transport of biomolecules through nanopores occurs over long timescales, using all-atom models is challenging when studying these systems.

Utilizing simplified models, such as coarse-grained (CG) models, for simulating these systems could be an alternative to the all-atom models. CG models group several atoms into a single site, enabling longer simulations. These models make the study of ion and biomolecule transport closer to the experimental time scales possible. These CG simulations with αHL nanopore, which have been performed for the first time with the MARTINI force field by our group, were able to identify key residues within the nanopore that affect current asymmetry and anion selectivity. Additionally, the first simulations of single-stranded DNA (ssDNA) translocation using this CG system paved the way for this thesis.

This thesis investigates the DNA transport through αHL nanopore using experimental approaches and CG-steered molecular dynamics (CG-SMD) simulations. The study focuses on two key aspects: the experimental work mainly focuses on the unzipping dynamics of double-stranded DNA (dsDNA) as a function of duplex length, duplex structure, and the applied voltage, and the MD simulations work focuses on ssDNA translocation dynamics as a function of various parameters such as the ssDNA orientation, ssDNA charges, sequence composition, and the steered molecular dynamics(SMD) forces. The MD studies of this work are essential for studying experimentally inaccessible details and validating this CG MARTINI system for further simulations of the unzipping process, the molecular details of which are yet to be elucidated.

In the first chapter, we will introduce the principles of nanopore technology and the characteristics of αHL. We will then introduce the experimental studies involving the transport of biomolecules and ions through this nanopore, focusing on DNA translocation and unzipping dynamics. This section will also cover the theoretical approaches, including continuum models and molecular dynamics simulations, which have been employed in previous studies to understand these transport processes better.

The second chapter will describe the experimental and MD simulation methods employed in this thesis. The experimental methods offering insights into the setup, procedures, and analytical techniques used in investigating ssDNA translocation and unzipping

through the αHL nanopore, along with the DNA molecules utilized, will be explained in detail. This chapter will also include the principles of MD simulations, force fields, and coarse-grained models. Specifically, the MARTINI force field parameters and the characteristics assigned to each type of molecule used in the simulations will be presented.

The third chapter will present the results of our experimental studies on ssDNA translocation and dsDNA unzipping through the αHL nanopore. It will discuss the behavior of ssDNA translocation as a function of its orientation, comparing the 3' and 5' ends, which serves as a complement to our MD simulation results. Additionally, the chapter will discuss the results of dsDNA unzipping experiments, examining various parameters such as the duplex length, duplex structure, and the applied voltage. We seek to explore the mechanisms that govern the unzipping processes and how different structural and experimental conditions influence these dynamics, with the goal of understanding the factors affecting DNA unzipping through nanopores.

The fourth chapter presents the results of our molecular dynamics simulations on ssDNA translocation through the αHL nanopore. This section will primarily focus on the distribution of translocation times, examining the influence of ssDNA orientation (3' vs. 5') and the electrostatic interactions between the nanopore and the DNA. We will analyze the tilting angles of DNA bases to understand how orientation affects translocation times, and we will investigate the conformational changes of ssDNA by examining inter-base distances during the translocation process. These results will be compared with prior experimental and theoretical studies to demonstrate how the MARTINI CG model effectively captures critical aspects of DNA translocation dynamics and provides a reliable framework for future studies.

The fifth chapter shifts focus to the effects of sequence composition and pulling force on ssDNA translocation through the αHL nanopore. We will present our findings on the sequence-dependent differences in translocation times, comparing the translocation times of poly(dA) and poly(dC) molecules of various nucleotide lengths. Additionally, we will explore how different pulling forces affect translocation dynamics. Our analysis will highlight the conformational changes ssDNA undergoes during translocation, offering insights into how the magnitude of the pulling force affects DNA inter-base distances. By comparing our simulation results with existing experimental and theoretical data, we aim to confirm the MARTINI CG model as an effective tool for studying further transport processes.

Finally, we will summarize the key findings from both our experimental and computational investigations on DNA translocation and unzipping through the αHL nanopore. After reflecting on these results, we will explore potential avenues for future research.

Introduction

La technique de la molécule unique est un outil puissant qui permet d'obtenir des informations précieuses sur les propriétés, les mécanismes moléculaires et les interactions des biomolécules individuelles qui sont masquées par la masse (bulk). Les études de la molécule unique permettent d'étudier des processus d'équilibre ou de hors équilibre, des événements rares, des états de transition ou de courte durée, et des informations mécaniques et thermodynamiques peuvent être déduites de ces études. Des techniques telles que le transfert d'énergie par résonance de fluorescence (FRET) pour étudier les changements de conformation ou les processus biochimiques, la microscopie à force atomique (AFM) pour sonder les énergies impliquées dans les interactions moléculaires, les pinces optiques pour déduire les forces impliquées dans divers processus, et la technologie des nanopores pour extraire les paramètres cinétiques et étudier les propriétés intrinsèques des molécules ont été mises au point pour les études de la molécule unique. Parmi celles-ci, la technologie des nanopores se distingue par ses avantages, notamment la détection sans label, l'analyse en temps réel et la capacité d'analyser une grande variété de macromolécules et de biomolécules dans diverses conditions. Ces caractéristiques la rendent particulièrement utile pour des applications telles que le séquençage de l'ADN, l'une de ses utilisations les plus populaires. Dans les expériences sur les nanopores, un potentiel électrique est appliqué au système contenant un nanopore inséré dans une membrane isolante, ce qui induit le passage des ions à travers le nanopore et génère par conséquent un courant ionique. Lorsqu'une molécule passe à travers le nanopore, elle bloque partiellement ou totalement le nanopore, engendrant un courant ionique réduit caractéristique qui dépend des propriétés de la molécule ainsi que des conditions expérimentales. En analysent ces variations du signal de courant, il est possible d'étudier la taille, la structure ou la conformation des molécules en temps réel et à faible coût.

Le nanopore d'α-hémolysine (αHL), une toxine produite par *Staphylococcus aureus*, est largement utilisé dans les expériences de nanopores pour plusieurs raisons telles que sa structure bien caractérisée, ses dimensions et son hydrophilie. Il forme un heptamère en forme de champignon composé d'un chapeau et d'une tige. L'utilisation des nanopores, en particulier de l'αHL, a permis des avancées significatives en matière de détection et de séquençage moléculaires, les premières études ayant démontré la capacité de détecter les acides nucléiques, l'ADN ou l'ARN, et d'autres biomolécules en évaluant les événements de translocation. Par exemple, les protéines et les peptides peuvent être détectés et caractérisés par ce nanopore. En analysant les blocages de courant, il est possible d'identifier les changements de conformation, la dynamique de repliement et les interactions de liaison des ligands des protéines lors de leur passage dans le nanopore. L'utilisation de l'αHL s'est également étendue à la détection de petites molécules et de polymères. Des études ont montré que le polyéthylène glycol (PEG) et d'autres polymères synthétiques

peuvent également être différenciés sur la base de leur taille et de leur forme à partir des blocages de courant lors de leur passage à travers le nanopore. Cette capacité à détecter les molécules synthétiques a élargi les applications potentielles des nanopores au-delà des systèmes biologiques, dans des domaines tels que la science des matériaux et l'administration de médicaments.

L'étude de la translocation de l'ADN à travers l'αHL reste centrale dans la technologie des nanopores, en particulier pour les applications de séquençage. Les premières expériences ont démontré que les nucléotides peuvent être distingués par leur niveau de blocage de courant ou leur temps de translocation, qui sont influencés par divers facteurs. Ces facteurs comprennent les propriétés intrinsèques des molécules, telles que la longueur, la séquence, la structure et l'orientation de l'ADN dans le nanopore, ainsi que les caractéristiques du nanopore lui-même, notamment sa distribution de charge, son diamètre et sa composition chimique. En outre, les conditions expérimentales telles que la température, la concentration ionique et la tension appliquée influencent également les niveaux de blocage du courant et les temps de translocation.

Bien que ces techniques expérimentales aient fourni des informations précieuses, elles ne peuvent pas dévoiler tous les mécanismes moléculaires du transport de molécules au niveau microscopique. C'est pourquoi différentes approches théoriques et informatiques ont été développées pour compléter les résultats expérimentaux. Il s'agit notamment de modèles de continuum, tels que la théorie de Poisson-Nernst-Planck, la dynamique brownienne et les simulations de dynamique moléculaire. Les premières études de dynamique moléculaire ont permis de mieux comprendre le transport ionique et la translocation de l'ADN à travers les nanopores, confirmant l'influence de facteurs tels que la longueur des oligo-nucléotides, leur orientation, la composition des séquences et la distribution de la charge des pores. Cependant, il faut créer un large système pour modéliser une molécule d'ADN passant à travers un pore inséré dans une membrane. La modélisation de tels systèmes à l'aide de modèles tout-atome limite le temps de simulation. Sachant que le transport de biomolécules à travers les nanopores se produit sur de longues échelles de temps, l'utilisation de modèles tout-atome semble être un défi pour l'étude de ces systèmes.

L'utilisation de modèles simplifiés, tels que les modèles gros-grains, pour simuler ces systèmes peut être une alternative aux modèles tout-atome. Les modèles gros-grains regroupent plusieurs atomes en un seul site, ce qui permet des simulations plus longues. Ces modèles permettent d'étudier le transport des ions et des biomolécules en se rapprochant des échelles de temps expérimentales. Ces simulations gros-grains avec le nanopore αHL, qui ont été réalisées pour la première fois avec le champ de force MARTINI par notre groupe, ont permis d'identifier des résidus clés dans le nanopore qui affectent l'asymétrie du courant et la sélectivité anionique. En outre, les premières simulations de translocation d'ADN simple brin (ADNsb) à l'aide de ce système gros-grains ont également ouvert la voie à cette thèse.

Cette thèse étudie le transport de l'ADN à travers le nanopore αHL en utilisant des approches expérimentales et des simulations de dynamique moléculaire dirigées par gros-grains (CG-SMD). L'étude se concentre sur deux aspects clés : la partie expérimental se concentre principalement sur la dynamique de dézippage de l'ADN double brin (ADNdb) en fonction de la longueur du duplex, de la structure du duplex et de la tension appliquée, et la partie de simulation dynamique moléculaire se concentre sur la dynamique de translo-

cation de l'ADNsb en fonction de divers paramètres tels que l'orientation de l'ADNsb, les charges de l'ADNsb, la composition de la séquence et les forces SMD. Les études de dynamique moléculaire de ce travail sont essentielles pour étudier des détails inaccessibles expérimentalement et valider ce système gros-grains MARTINI pour d'autres simulations comme l'étude du processus de dézippage, dont les détails moléculaires doivent encore être élucidés.

Dans le premier chapitre, nous présenterons les principes de la technologie des nanopores et les caractéristiques de l'αHL. Nous présenterons ensuite les études expérimentales impliquant le transport de biomolécules et d'ions à travers ce nanopore, en nous concentrant sur la translocation de l'ADN et la dynamique de dézippage. Cette section couvrira également les approches théoriques, y compris les modèles continus et les simulations de dynamique moléculaire, qui ont été utilisées dans des études précédentes pour mieux comprendre ces processus de transport.

Le deuxième chapitre décrira les méthodes expérimentales et de simulation de dynamique moléculaire employées dans cette thèse. Les méthodes expérimentales offrant un aperçu de la configuration, des procédures et des techniques analytiques utilisées pour étudier la translocation et la dézippage de l'ADNdb à travers le nanopore αHL, ainsi que les molécules d'ADN utilisées, seront expliquées en détail. Ce chapitre comprendra également les principes des simulations dynamique moléculaire, des champs de force et des modèles gros-grains. Plus précisément, les paramètres du champ de force MARTINI et les caractéristiques attribuées à chaque type de molécule utilisée dans les simulations seront présentés

Le troisième chapitre présentera les résultats de nos études expérimentales sur la translocation de l'ADNsb et le dézippage de l'ADNdb à travers le nanopore αHL. Il discutera du comportement de la translocation de l'ADNsb en fonction de son orientation, en comparant les extrémités 3' et 5', ce qui sert de point de comparaison aux résultats de nos simulations dynamique moléculaire. En outre, ce chapitre abordera les résultats des expériences de dézippage de l'ADNdb, en examinant divers paramètres tels que la longueur du duplex, la structure du duplex et la tension appliquée. Nous cherchons à explorer les mécanismes qui régissent les processus de dézippage et comment différentes conditions structurelles et expérimentales influencent cette dynamique, dans le but de fournir une compréhension des facteurs qui affectent le dézippage de l'ADN à travers les nanopores.

Le quatrième chapitre présente les résultats de nos simulations de dynamique moléculaire sur la translocation de l'ADNsb à travers le nanopore αHL. Cette section se concentre principalement sur la distribution des temps de translocation, en examinant l'influence de l'orientation de l'ADNsb (3' *vs.* 5') et les interactions électrostatiques entre le nanopore et l'ADN. Nous analyserons les angles d'inclinaison des bases de l'ADN pour comprendre comment l'orientation affecte les temps de translocation, et nous étudierons les changements de conformation de l'ADNsb en examinant les distances inter-bases pendant la translocation. Ces résultats seront comparés à des études expérimentales et théoriques antérieures afin de démontrer comment le modèle MARTINI capture efficacement les aspects critiques de la dynamique de translocation de l'ADN et fournit une base fiable pour les études futures.

Le cinquième chapitre se concentre sur les effets de la composition de la séquence et de la force appliquée sur la translocation de l'ADNsb à travers le nanopore αHL. Nous

présenterons nos conclusions sur les différences de temps de translocation en fonction de la séquence, en comparant les temps de translocation de molécules de poly(dA) et de poly(dC) de différentes longueurs de molécules d'ADN. En outre, nous étudierons comment différentes valeurs de forces appliquées affectent la dynamique de translocation. Notre analyse mettra en évidence les changements de conformation que subit l'ADNsb au cours de la translocation, ce qui permettra de comprendre comment la valeurs de la force appliquée influe sur les distances inter-bases de l'ADN. En comparant nos résultats de simulation avec les données expérimentales et théoriques existantes, nous visons à confirmer que le modèle MARTINI gros-grains est un outil efficace pour étudier d'autres processus de transport.

Enfin, nous résumerons les principaux résultats de nos études expérimentales et des simulations dynamique moléculaire sur la translocation et le dézippage de l'ADNdb à travers le nanopore αHL. Après avoir synthetisé ces résultats, nous explorerons des pistes potentielles pour de futures recherches.

Chapter 1

State of the Art

Contents

1.1 Overview of Nanopore Technology 9
1.1.1 αHL for Nanopore Technology Applications 11
1.1.2 Importance of DNA Translocation and Unzipping Studies . . . 13
1.2 Experimental Studies of DNA transport 14
1.2.1 DNA Translocation Studies 14
1.2.2 DNA Unzipping Studies 23
1.3 Theoretical studies 25
1.3.1 Continuous models 25
1.3.2 Molecular Dynamics Simulations 29

1.1 Overview of Nanopore Technology

Nanopore technology, which involves single-molecule studies, has become essential for analyzing a wide range of molecules. The principle of molecule translocation through nanopores is akin to the Coulter counter, a device used for counting and sizing particles suspended in a fluid. In the Coulter counter, particles are drawn through a small aperture between two compartments, causing temporary changes in electrical resistance of this aperture as each particle passes through. Similarly, in nanopore experiments, two compartments are separated by a membrane with a nanometer-scale pore. Molecules are added to one compartment, and their passage through the pore is monitored by measuring changes in ionic current under a constant applied voltage. When a molecule traverses the pore, it partially or completely blocks the channel, leading to a decrease in current. This current reduction, known as the blocking current (I_b), has an amplitude and duration that are characteristic of the translocated molecule. (See Figure 1.1 for the illustration of the nanopore experiments.) Further details to this nanopore method will be detailed in Chapter 2, Methods.

The studies incorporating this nanopore technique date back to the beginning of the 1990s. For example, Krasilnikov *et al.* and Sabirov *et al.* used polyethylene glycol (PEG)

polymers of varying sizes to estimate the pore's radius by analyzing the decrease in conductance with different PEGs [1, 2, 3]. Henrickson *et al.* used polynucleotides, whose ends could bind to large molecules that could not pass through the pore, to measure its length. They revealed that a molecule of 45 to 50 nucleotides of thymine or cytosine could span the entire length of an α-hemolysin (αHL), one of the most employed protein nanopore in these studies [4]. It has also been shown that, by this nanopore method, molecules interacting with a nanopore can cause current blockages without being transported through the pore [5, 6].

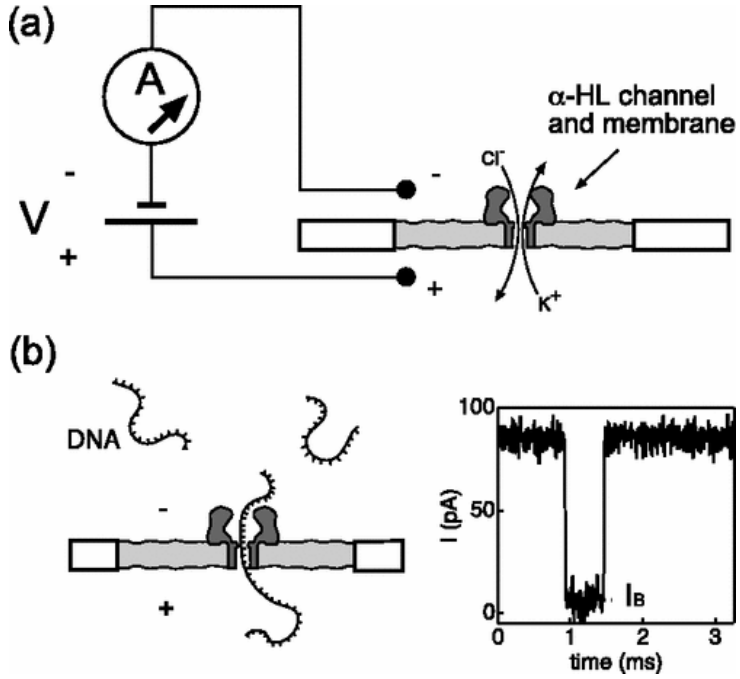


Figure 1.1: Figure taken from Bonthuis *et al.*[7]. Experimental setup with αHL inserted into a membrane and surrounded by ionic solution. Applying a voltage across the system induces ionic flux through the pore. The presence of DNA molecules in one compartment causes them to translocate through the pore, resulting in characteristic current blockages.

Nanopores can be used to analyze various types of biomolecules, peptides [8, 9, 10] and proteins [11, 12], nucleic acids [13, 14, 15, 16, 17, 18, 19], TNT [20] and recently polysaccharides [21]. αHL has also been used to determine the size of PEG polymers, demonstrating its ability to discriminate a mixture of PEG molecules of different sizes by analyzing the blocking currents and residence times of the polymers inside the nanopore channel [22].

The application of nanopores for analyzing nucleic acids has gained significant interest over time. One of the most promising applications of nanopore technology is in the field of DNA sequencing. For instance, pioneering research by David Deamer, Daniel Branton, George Church, and John Kasianowicz in 1996 laid the groundwork for using nanopores in DNA transport and sequencing. They explored the feasibility of employing small pores, particularly αHL, for efficient single-stranded DNA (ssDNA) sequencing. The concept involves detecting and analyzing different bases of a ssDNA strand by measuring real-time changes in electrical current as molecules pass through the pore under an applied electric voltage. It offers real-time analysis, long-read, and label-free sequencing, and

circumvents the limitations of traditional methods by eliminating the need for extensive sample preparation and PCR amplification [13].

Solid-state nanopores, introduced in the early 2000s by Li et al.[23], expanded the capabilities of nanopore technology by offering tunable pores and enhanced stability under elevated applied electric voltages. Fabricated from materials such as silicon nitride or graphene, these solid-state pores are integral to sophisticated microfluidic systems, enabling precise control such as temperature and fluidic mixing, broadening the scope of nanopore applications [23, 24, 25, 26, 27, 28]. The development of hybrid nanopores, which combine biological and solid-state components, is another area of active research, aiming to leverage the advantages of both types of nanopores for improved performance in molecular sensing and analysis [29, 30].

Nanopore technology finds application in a multitude of domains, such as mutation detection in the human genome [31], pathogen detection in remote areas [32], or rapid diagnosis in clinical settings [32], while also promising to facilitate access to genomic information globally due to its cost-efficiency and scalability. Additionally, nanopores are useful in studying folding dynamics [33], ligand interactions [34], and biomolecule modifications, which have implications for drug discovery and biophysics.

Beyond biology, nanopore technology is also adressed in materials science, allowing precise characterization of nanoparticles, polymers, and other nanoscale materials. For example, nanopores can be used to analyze the size distribution and aggregation of nanoparticles [35, 36], study the properties and dynamics of polymers, and investigate the mechanical properties of nanomaterials.

They are also utilized for monitoring and detecting pollutants, toxins, and biomarkers in air, water, and soil samples, thereby contributing to environmental protection and public health. They can detect heavy metals, pesticides, and other hazardous substances with high sensitivity and specificity, making them valuable tools for environmental monitoring [37].

In this thesis, we are interested in DNA translocation and its unzipping dynamics through wild-type (WT) αHL nanopore, combining experimental and theoretical approaches. Therefore, in the following sections, we will focus only on DNA transport through protein nanopores, specifically through αHL nanopore.

1.1.1 αHL for Nanopore Technology Applications

αHL is a pore-forming cytotoxin produced by the bacterium *Staphylococcus aureus*. Its initial recognition dates back to investigations in the late 1800s, where studies linked *S. aureus* secretions to severe effects such as hemolysis and lethality in animals. An important moment occurred in 1928 when contaminated diphtheria toxin-antitoxin preparations led to fatalities in Bundaberg, Australia, prompting further investigation by F. McFarlane Burnet. He identified a heat-labile antigenic substance in *S. aureus* responsible for these toxic effects, although it wasn't until 1964 that αHL was identified as a key virulence factor [38, 39, 40, 41].

In the late 20th century, the Human Genome Project catalyzed a paradigm shift for αHL. David Deamer proposed its potential for single-molecule DNA analysis. His group hypothesized that ssDNA could translocate through the αHL pore, with each nucleotide producing distinct changes in ionic current. Subsequent studies by Kasianowicz *et al.*

observed that the translocation of ssDNA and RNA through the αHL pore caused distinct, measurable blockades in ionic current. These blockades were characteristic of the nucleotide composition and sequence of the polynucleotides. They confirmed that while ssDNA and RNA could pass through the pore, double-stranded DNA (dsDNA) could not, highlighting the size limitations of the αHL nanopore for different forms of biomolecules. Moreover, the duration of the current blockades was found to be proportional to polymer length, allowing channel blockades to be used for measuring polynucleotide length. The study underscored the potential of αHL for DNA transport and sequencing by showing that different nucleic acids produced unique signatures in the ionic current, paving the way for developing nanopore-based sequencing technologies. These findings underscored the utility of αHL in studying DNA transport, not only for sequencing purposes but also for understanding fundamental aspects of DNA behavior and interactions. Examples include examining DNA-protein interactions, studying the effects of DNA damage and repair, investigating DNA folding and unfolding dynamics, and analyzing the mechanical properties of DNA as it moves through confined spaces [13].

The publication of αHL's crystal structure in 1996 provided critical insights into its dimensions and the amino acid residues lining its channel: a mushroom-shaped heptamer of 7 identical chains with a 5 nm stem and an inner channel diameter varying from 1.4 nm to 2.4 nm, influenced by amino acid side chains. The nanopore's β-barrel consists of two charged rings at its two extremities: the narrowest part, the constriction zone of the nanopore is formed by amino acid residues Glu111, Lys147, and Met113, the trans exit of the pore is composed of both positively and negatively charged residues Asp127, Asp128, and Lys131 on each of the 7 chains [42].

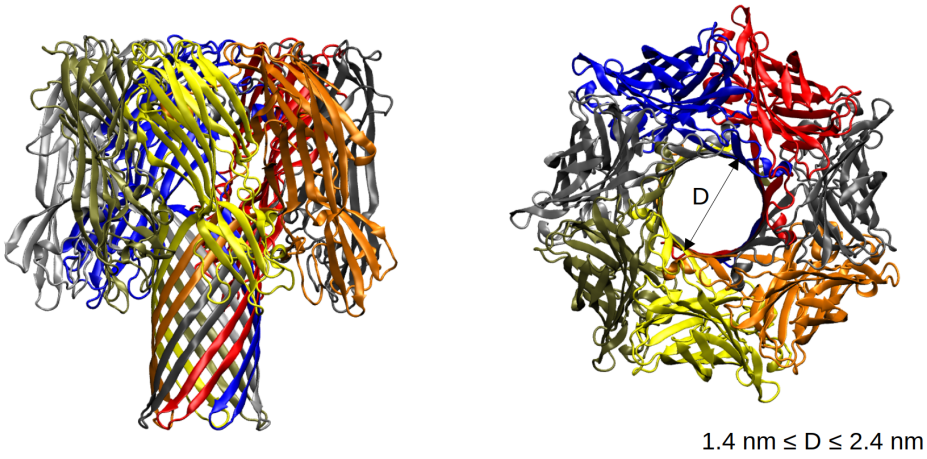


Figure 1.2: The crystal structure of αHL from the side (on the left) and top view (on the right), each chain is indicated in a different color. The diameter, D, of the inner wall of the β-barrel channel is given in the right figure.

Meanwhile, the insertion mechanism of αHL into lipid bilayers has been extensively studied [42, 43, 44]. The formation of the αHL pore involves several sequential steps, as illustrated in Figure 1.3. Initially, the αHL monomer binds to the target cell membrane. Upon binding, αHL monomers begin to oligomerize on the membrane surface, assembling into a heptameric structure composed of seven identical monomers. This oligomerized complex initially forms a pre-pore, a non-inserted, membrane-bound intermediate that is

stable and serves as a precursor to the transmembrane pore. The pre-pore undergoes a conformational change, driving the insertion of the heptameric structure into the membrane. This insertion involves the extension of beta-barrel domains from each monomer into the lipid bilayer, forming a β-barrel pore that spans the membrane. Once inserted, the heptamer forms a stable transmembrane pore, featuring a stem that traverses the membrane and a cap that remains on the surface. Fully inserted, the αHL pore permits the passage of ions and small molecules through the membrane, disrupting membrane integrity and leading to cell lysis and death, which constitutes the cytotoxic effect of αHL [45, 46].

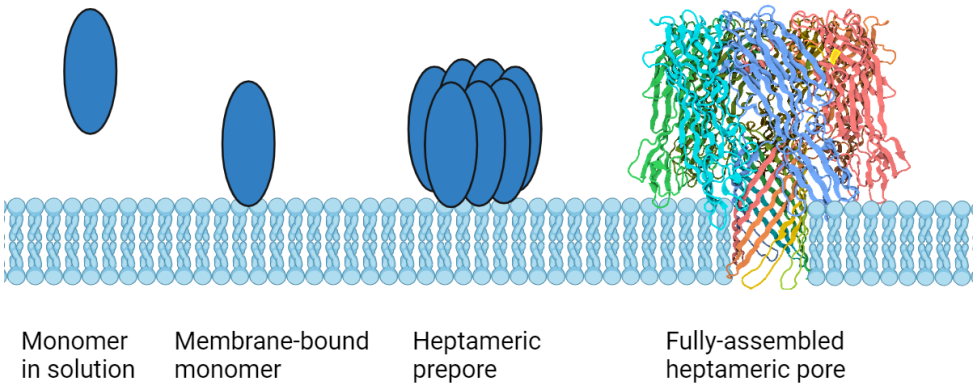


Figure 1.3: Proposed insertion mechanism of αHL in the lipid bilayer [42, 43, 44] .αHL monomer binds to the target cell membrane. αHL monomers begin to oligomerize on the membrane surface, assembling into a heptameric structure composed of seven monomers, called the pre-pore complex. The pre-pore undergoes a conformational change, driving the insertion of the heptameric structure into the membrane. (Image created using Biorender.)

These advancements marked a significant step forward in the development of methods for studying DNA transport, as will be detailed in the Experimental and Theoretical Studies section.

1.1.2 Importance of DNA Translocation and Unzipping Studies

Since Watson and Crick's discovery of DNA in 1953,[47] the study of DNA has opened new horizons in biological research. DNA contains the necessary genetic information that determines the characteristics of all living organisms. By studying DNA, scientists can decipher how genes are organized and regulated, how they express traits, and how genetic variation contributes to species diversity. Additionally, studying DNA has unveiled essential processes such as replication, transcription, and repair, which are crucial for cell division, gene expression, and maintaining genetic integrity.

The study of DNA translocation through nanopores is a subject for both fundamental and applied sciences, providing insights and applications beyond sequencing. From a physical point of view, understanding DNA translocation helps elucidate the molecular forces and interactions involved when DNA moves through confined spaces, shedding light on electrostatic interactions, hydrodynamic effects, and molecular conformations. This knowledge extends to DNA-protein interactions, critical for replication, transcription, and

repair, thus enhancing our understanding of these fundamental biological mechanisms [13]. Additionally, studying DNA transport mechanisms, including diffusion and electrophoresis, offers insights into how charged molecules behave under electric fields and the role of sequence, length, and secondary structure in molecular translocation [14, 16]. Furthermore, insights from DNA translocation inform the design and optimization of nanopores for sequencing applications, such as adjusting pore size, shape, and surface chemistry to improve sensitivity and selectivity [48].

In material science, studying DNA translocation aids in understanding the properties of materials used in nanopore fabrication, such as silicon nitride and graphene. It guides the development of new nanopores with enhanced features [23].

Overall, studying DNA translocation through nanopores bridges the gap between biological phenomena and physical principles. Theoretical and computational modeling benefits greatly from DNA translocation studies, as experimental data validate theoretical models and simulations of molecular transport. Accurate models, in turn, predict the behavior of DNA and other biopolymers in nanopores, facilitating the design of new experiments and technologies [49].

Moreover, beyond elucidating transport phenomena, investigations into DNA unzipping studies offer insights into the mechanical dynamics of nucleic acids. These studies, originating from nucleosome structure investigations, involve the mechanical unzipping of base pairs under piconewton-scale forces exerted by DNA-binding proteins during processes like transcription, replication, and repair. Understanding these unzipping mechanisms and dynamics is crucial for studying secondary and tertiary structures in complicated polynucleotides, such as ribozymes [50], the affinity of DNA-binding proteins and enzymes [51, 52], mismatch detection [53], interactions of base pairing [54] and the structure of nucleosomes [55]. Techniques such as optical tweezers [56, 57], atomic force microscopies [58], and nanopores [50, 59, 60], inspired by sequencing studies, have been employed since the late 20th century to explore these dynamics.

In the next two sections, the advances in nanopore techonology by experimental and theoretical studies will be explained in detail for studying DNA translocation and unzipping.

1.2 Experimental Studies of DNA transport

The experimental principle was explained in previous sections and detailed in Section 2.1, in Methods. Over decades, experimental work with nanopores and αHL has evolved significantly.

1.2.1 DNA Translocation Studies

The first experiments on polynucleotide translocation were conducted in 1996 by Deamer and collaborators [13]. These experiments confirmed that the αHL channel could serve as a nano-scale sensor for the translocation of individual nucleic acid molecules. It was demonstrated by Kasianowicz *et al.* that ssRNA and ssDNA passage through the nanopore could be detected as discrete changes in ionic current. The study provided insights into the translocation mechanism of nucleic acids through αHL. The characteristic lifetimes of channel blockades caused by poly(U) RNA molecules revealed three distinct peaks in the

blockade lifetimes, as illustrated in Figure 1.4, each corresponding to different translocation events. Peak 1 represented instances where RNA molecules approached the pore but did not translocate. Peaks 2 and 3 corresponded to translocation events occurring via different orientations of the nucleic acid molecules (3' or 5'). This observation highlighted the influence of the orientation of the nucleic acids on the translocation process. Further investigation by Kasianowicz *et al.* into the factors influencing the blockade lifetimes showed that the lifetimes of poly(U) RNA channel blockades were directly proportional to the length of the polymer, as shown in panel (a) of Figure 1.5. Additionally, an inverse relationship between the blockade lifetimes and the applied voltage across the nanopore was also observed as illustrated in panel (b) of Figure 1.5.

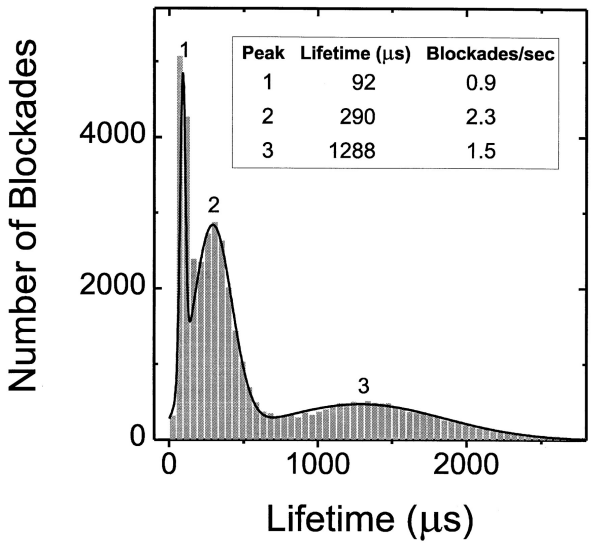


Figure 1.4: Figure taken from Kasianowicz *et al.*.[13] The poly(U) (0.1 mg/ml; mean length, 210 nucleotides) produced channel blockades with three distinct characteristic lifetimes. These lifetimes were analyzed by fitting three Gaussian distributions to the data, corresponding to the three peaks observed. Peak 1 represents non-translocation attempts, where the polymers collide with the pore but fail to pass through. Peaks 2 and 3 correspond to successful translocation events occurring with different orientations (3' or 5').

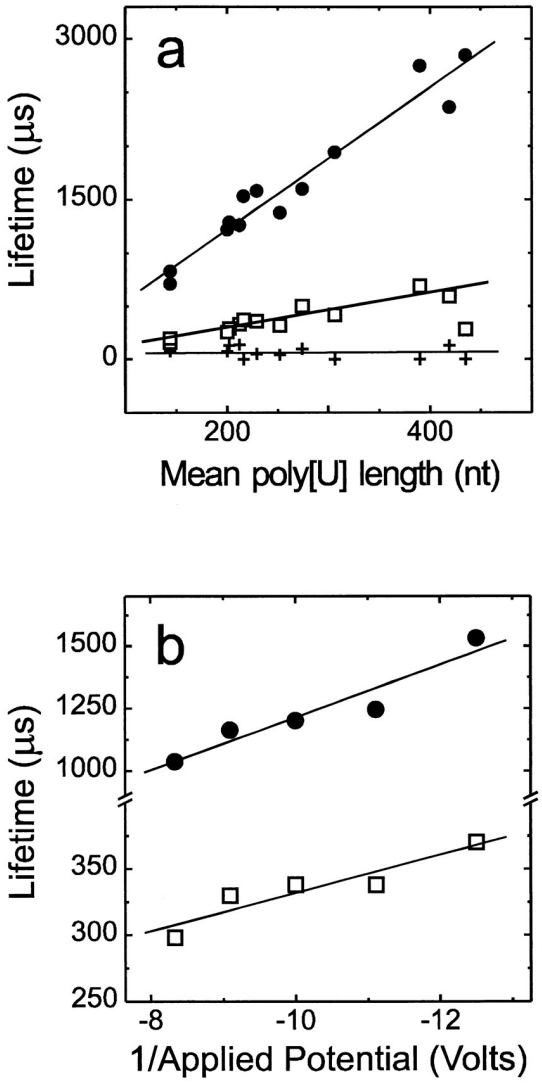


Figure 1.5: Figure taken from Kasianowicz *et al.* [13]. Poly(U) channel blockade lifetimes were observed to be (a) proportional to the mean polymer length and (b) inversely proportional to the applied voltage. The plots illustrate the lifetimes for (a) peaks 1 (+), 2 (□), and 3 (•) in experiments conducted with a -120 mV applied voltage, using 13 different size-selected poly[U] polymers, and (b) for peaks 2 (□) and 3 (•) for poly(U) with a mean length of 215 nucleotides at the indicated voltages.

In 1999, Akeson *et al.* expanded upon previous research by demonstrating that αHL nanopores can distinguish between different homopolymeric polynucleotides. Their study revealed that individual single-stranded molecules of poly(C), poly(A), poly(U), and poly(dC) could be differentiated based on blockade amplitude, blockade duration, or both, as illustrated in Figure 1.6 for poly(C) and poly(A) molecules. Poly(A) (RNA) exhibited notably slower translocation compared to the other polynucleotides, likely due to its helical structure, which has a diameter of 2.1 nm—wider than the pore constriction. In contrast, poly(dC), the deoxyribonucleotide version of cytosine (with "d" indicating DNA rather than RNA), translocated the fastest among the studied polynucleotides. This

behavior was attributed to RNA molecules maintaining a helical structure during translocation, whereas DNA molecules adopted a disordered structure and did not translocate as helices. The study also examined the blockade amplitudes of RNA copolymers $A_{30}C_{70}$, finding a higher frequency of entries from the 3' end compared to the 5' end, suggesting potential directional preferences in translocation [61].

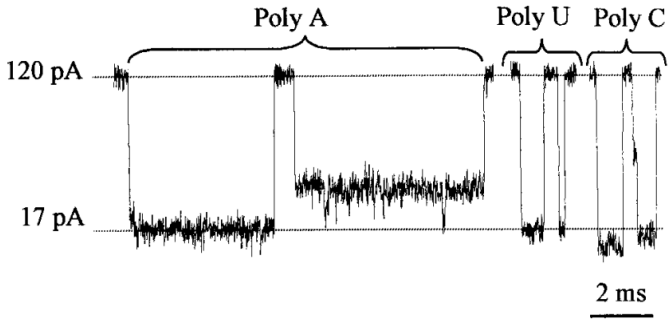


Figure 1.6: Figure taken from Akeson *et al.* [61]. Representative current traces recorded from the α-hemolysin pore in the presence of poly(A), poly(C), and poly(U) homopolymers. The average lengths of the homopolymers were approximately 130 for poly(C), 175 for poly(A), and 150 nucleotides for poly(U).

In the early 2000s, Meller *et al.* conducted significant studies on the translocation times of ssDNA molecules through αHL nanopores, further advancing the field [14, 15, 62]. It was demonstrated that different homopolymeric DNA sequences, such as poly(dC) and poly(dA) molecules, showed significant differences in passage times, allowing discrimination between them; specifically, purines translocated approximately three times slower than pyrimidines [14]. The slower translocation of poly(dA) was hypothesized to be due to its strong tendency for base stacking, which increases rigidity. Subsequent experiments with polymers having varying numbers of adenines spaced within cytosine polymers revealed that a few evenly spaced adenines in a predominantly cytosine polymer markedly slowed translocation, while increasing adenine content beyond 50% caused only a mild decrease in the translocation rate. These findings suggested specific interactions between adenines and the nanopore wall. Additionally, it was observed that at lower temperatures, the difference in translocation times between purines and pyrimidines increased, indicating stronger interactions between adenines and the pore at lower temperatures. The effect of applied voltage on translocation speed was also investigated, noting that higher voltages accelerated DNA translocation but compromised nucleotide discrimination accuracy. The relationship between applied voltage and DNA translocation speed through nanopores was described as quadratic. At low to moderate voltages, the relationship may appear somewhat linear, where increasing voltage results in a proportional increase in translocation speed due to the enhanced electrophoretic force. However, as the voltage increases further, the effect of this increased voltage seems to be stronger on the translocation dynamics [15, 62]. It was also noted that translocation time is linearly dependent on the length of DNA molecules longer than 12 nucleotides [62]. For shorter polymers ($N \leq 12$) that do not extend the full length of the channel, a different balance between the driving electrical force and the polymer-pore wall interactions is attained for each polymer length, resulting in a steep dependence of the translocation rate on polymer

length. These main findings from Meller *et al.* are illustrated in Figure 1.7.

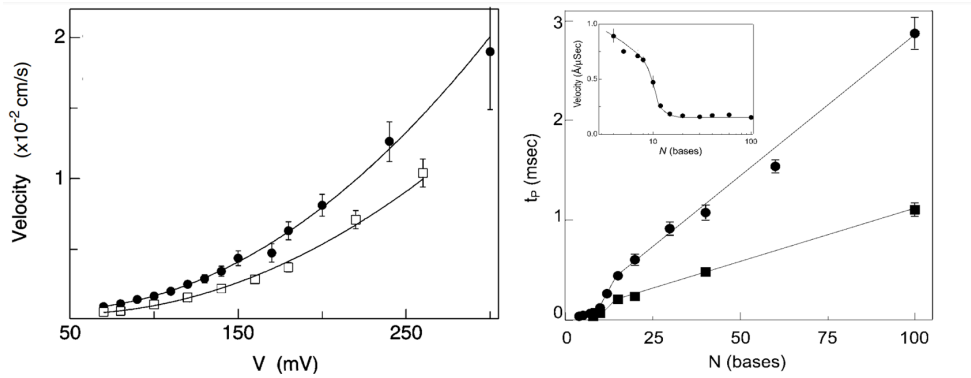


Figure 1.7: Key findings of Meller *et al.* on the dynamics of DNA translocation. On the left, the quadratic relationship between polymer velocity and applied voltage is depicted. Data points are shown for poly(dA) with lengths of 12 (•) and 30 (□) nucleotides at 2°C. The lines represent quadratic fits to the data, and error bars are provided except for data points below 90 mV, where insufficient events occurred to estimate errors [14]. On the right, the plot shows the dependence of the most probable translocation time, t_P, on polymer length, N. Measurements were performed for poly(dA) (•) and poly(dCdT) (■) at 2°C. For polymers longer than $N = 12$, the translocation time scales proportionally with polymer length, while a sharp transition is observed for shorter polymers ($N < 12$). The inset shows the average polymer velocity, calculated from t_P and translocation trajectories [62].

In 2005, Mathé *et al.* investigated the effect of orientation on ssDNA translocation dynamics, revealing that ssDNA molecules exhibit distinct behaviors depending on their orientation inside the αHL nanopore [63]. They first performed ssDNA translocation experiments with poly(dA_{60}) and measured the mean translocation current distribution. This distribution displayed two distinct peaks, indicating the impact of orientation on the measured ionic current. To identify the sources of each peak, the ion current flowing through the pore and the escape time for DNA hairpin molecules threaded from the *cis* chamber with either their 3' end (HP3') or their 5' end (HP5') were measured. The histogram of the mean blockade current for both hairpin molecules yielded the same peaks observed for ssDNA translocation, thus identifying the peaks associated with each orientation. Additionally, the dynamics of entry and exit from the *cis* side of the pore were examined. The experimental procedure involved setting the voltage to 120 mV to draw the DNA molecules into the pore, followed by setting the voltage to zero for a period of waiting time, and finally applying a small positive probing voltage, V_p, to determine whether the molecules remained in the pore. By repeating each measurement over 1,500 times, it was observed that the voltage-free diffusion of the 3'-threaded DNA (in the trans to *cis* direction) was twice as slow as that of the corresponding 5'-threaded DNA with the same poly(dA) sequence. This study was complemented by molecular dynamics (MD) simulations aimed at elucidating the underlying mechanisms of this behavior, which will be further detailed in Section 1.3.2.

However, the challenge of achieving single-base resolution during DNA transport through nanopores remained unresolved until 2005. Bayley's and Ghadiri's groups demon-

strated that a single adenine nucleotide substitution in a poly(dC) strand could be distinguished from cytosines based on its distinctive impact on αHL's ion conductance [64]. By the late 2000s, research began to focus on single-base discrimination with the goal of optimizing nanopore sequencing techniques. In 2008, Maglia *et al.* investigated the effect of internal charges in αHL on translocation dynamics. Their study showed that manipulating internal charges via mutagenesis allowed for fine-tuning the nanopore's properties. Introducing positive charges within the αHL nanopore increased the frequency of translocation events and reduced the threshold voltage required for translocation. This is due to positively charged residues attracting the negatively charged phosphate backbone of DNA, lowering the energy barrier for DNA entry into the nanopore, and enhancing the frequency of observed translocation events. Furthermore, the reduced threshold voltage indicates positive charges modify the nanopore's electrical potential profile.

Stoddart *et al.* expanded on this work by exploring the sensing capabilities of αHL nanopores. Their study in 2009 examined the nanopore's ability to discriminate single nucleotides in DNA strands immobilized within the pore [18]. They demonstrated that individual nucleotides could be distinguished based on current blockades, in accordance with previous experimental studies. The blocked current in αHL nanopores was found to be intricately linked to the transport dynamics since as a DNA molecule enters the nanopore, it displaces ions from the ion current passing through the pore, resulting in a reduction in current, detected as a blockade. The magnitude and duration of this blockade were discovered to depend on factors such as the speed of translocation, the sequence and structure of the DNA, and the electrostatic interactions between the DNA and the nanopore. The authors emphasized that single nucleotide discrimination depends on the subtle differences in current blockades caused by various nucleotides passing through the nanopore, with neighboring nucleotides affecting these blockades due to differences in sequence composition and arrangement.

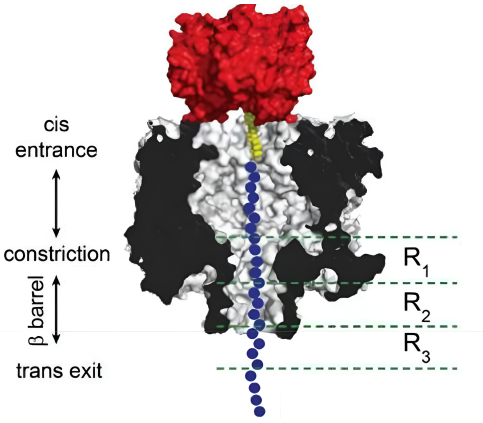


Figure 1.8: Figure taken from Stoddart *et al.* [18]. Schematic representation of a homopolymeric DNA oligonucleotide (blue circles) immobilized inside the αHL nanopore (grey, cross-section) via a 3' biotin-TEG (yellow) streptavidin (red) complex. The nucleobase recognition sites (R1, R2, and R3) within the β-barrel of the nanopore are shown [18, 19, 65].

In 2010, Stoddart *et al.* further investigated the sensing spots of αHL nanopores, confirming the presence of three distinct sensing spots: R_1 (Glu111/Lys147), R_2 (Gly119/Asn121),

and R_3 (Asn123/Thr125). (See Figure 1.8) With the central constriction as the primary site, interactions between the phosphate groups and the Lys147 ring were explained as a barrier to the translocation process. These interactions were considered dominant and important for the conformational dynamics of the ssDNA molecule in confined geometry, as these interactions are key to future nucleobase discrimination in sensing devices [19]. They further emphasized the importance of considering these three sensing spots when interpreting current signatures for accurate base identification, noting the challenges posed by using WT αHL. Continuing their research, they explored mutated αHL nanopores with enhanced base recognition capabilities [65]. Their study compared the base discrimination abilities of WT αHL nanopores with mutants, showing improved discrimination at recognition site number 2 (R2). One mutant featured strong recognition at sites R1 and R2, functioning like a "two-head sensor" that reads each nucleotide twice (at R1 and R2), thus enhancing base discrimination accuracy with a built-in proofreading mechanism. In 2015, the same group investigated further mutations in αHL nanopores and the effects of DNA stretching under applied potential on nucleobase recognition [66]. They found that mutations, such as Glu111Asn/Lys147Asn/Met113Tyr, enhanced discrimination of all four nucleobases at sites R2 and R3. This improvement was crucial for achieving high-fidelity discrimination between bases. Additionally, under applied potential, ssDNA stretched as it translocated through the nanopore. This stretching was discovered to reduce conformational variability, resulting in more consistent and distinguishable current signatures for each nucleobase. The study demonstrated how altering recognition sites through mutations could significantly enhance αHL nanopore base discrimination capabilities, providing a pathway for improving single base discrimination accuracy in applications like DNA sequencing.

A pivotal breakthrough, initially conceptualized in 1998 [67], involved introducing processing enzymes to regulate DNA translocation through the nanopores. Early investigations demonstrated that tethering enzymes such as the Klenow fragment or *Escherichia coli exonuclease-I* significantly slowed down DNA translocation through the αHL pore. Despite their effectiveness in slowing down movement, these methods did not achieve precise base-by-base control. Further research and experiments by Ghadiri and collaborators [68] made notable advancements in this direction. They employed a DNA polymerase to monitor primer strand elongation with single-nucleotide precision, alternating the voltage across the membrane to facilitate observation of the polymerase's interactions with the DNA template. This approach, given in Figure 1.9, allowed for real-time monitoring of polymerase dissociation from the DNA, positioning the extended primer/template junction precisely at the nanopore's sensing aperture. However, challenges persisted due to the tendency of enzymes to dissociate under the voltage load necessary to drive DNA through the nanopore, limiting consecutive nucleotide additions [69].

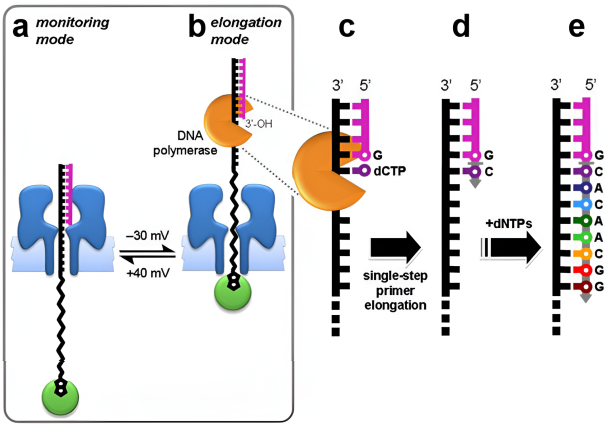


Figure 1.9: Figure taken from Cockroft *et al.* [68]. Monitoring DNA polymerase-catalyzed single-nucleotide primer extensions using a single-molecule nanopore device. (a) Initially, the DNA primer is interlocked within the αHL–DNA–PEG complex, where its length is measured via ion current readings in the monitoring mode (+40 mV). (b) In the elongation mode (−30 mV), the 3' end of the primer becomes accessible to DNA polymerase. (c) The DNA polymerase binds to the primer-template complex with the correct incoming deoxynucleotide triphosphate (dNTP). (d) The catalytic activity of the DNA polymerase incorporates a single dNTP into the primer against the DNA template threaded through the pore. (e) Successive base incorporation steps result in the templated extension of the primer sequence.

More successful experiments were performed with phi29 DNA polymerase - a replicative polymerase from the *Bacillus subtilis phage phi29*- which exhibited remarkable stability in forming binary complexes with DNA on the αHL pore compared to other enzymes [70]. Techniques incorporating blocking oligomers further refined the control over phi29-DNA interactions within the nanopore, enabling unprecedented examination of up to 500 synthetic DNA strands. In these studies, approximately 25 nucleotides of each captured template strand were read in the 3' to 5' and 5' to 3' directions. These studies were subject to several publications that validated the feasibility of practical nanopore DNA strand sequencing with other biological nanopores [71, 72].

Through extensive experimental work, it has been shown that the translocation process is a complex phenomenon, resulting in a wide distribution of translocation times per base. This complexity arises from various factors, such as the chemical composition, where variations in nucleotide sequence or modifications alter DNA's interaction with the nanopore walls. For instance, adenine-rich regions exhibit slower translocation due to stronger base-pore interactions than cytosine-rich regions [13, 14, 16, 17, 15, 62]. Additionally, the conformation of the translocating DNA molecule significantly impacts its passage through the nanopore. Structured or folded DNA, such as hairpins or loops, encounters increased friction within the pore, resulting in slower translocation rates [73, 50, 59]. Changes in salt concentration affect the electrostatic environment, such as increasing the nanopore conductance or enhancing the screening of electrostatic interactions. Higher salt concentrations reinforce the electric capture field, increasing the molecule capture rate [74], but also lead to slower DNA translocation rates [75, 74]. Furthermore, the orientation of the DNA molecule inside the nanopore dictates how different DNA behaves in the pore's in-

terior, influencing dwell times and translocation speeds [63, 59]. pH variations modify the ionization states of DNA bases and nanopore residues, impacting hydrogen bonding and electrostatic interactions during translocation. Higher pH conditions generally promote faster translocation rates due to reduced electrostatic interactions and increased mobility of DNA molecules [76, 77]. Mutations in nanopore proteins can change the distribution and charge density of amino acids lining the pore, altering the electrostatic potential and affecting DNA capture and threading dynamics [78, 18, 19, 79]. Temperature also plays an important role, where higher temperatures enhance molecular motion and flexibility, thereby increasing translocation speeds [16]. Finally, the applied voltage influences the electrophoretic force driving DNA through the pore; higher voltages accelerate translocation rates but can also induce non-specific effects or alter interactions between DNA and the nanopore [15, 62].

In light of these experimental findings, a groundbreaking innovation was introduced in 2014 by Oxford Nanopore Technologies with the release of the first portable single-molecule sequencing device based on nanopores [80, 81, 82]. These compact devices significantly advanced single-molecule DNA sequencing, enabling extremely long reads in a short amount of time in a user-friendly manner.

However, the WT αHL pore was considered sub-optimal for base-by-base sequencing of ssDNA molecules for several reasons. The translocation time per base was very short, a few microseconds per base, making it difficult to recognize single bases during translocation. Additionally, the β-barrel, with its three primary sensing regions where nucleobases interact during DNA translocation, influences the current levels observed during blockade events [18, 19, 66]. The stem's capacity to accommodate approximately 12 nucleobases further complicates the precise identification of individual bases. These limitations led to the development of engineered versions of this pore-forming toxin to achieve the required resolution.

Research has extended beyond αHL to explore other biological nanopores and their mutants, such as MspA [83, 84], aerolysin [85, 10], and recently CsgG:CsgF [86]. (See Figure 1.10 MspA, found in *Mycobacterium smegmatis*, is notably smaller than αHL, with a narrower constriction of approximately 1.2 nm. This pore has a funnel geometry providing a better sensing resolution for biopolymers like DNA or proteins but may exhibit reduced signal strength compared to other nanopores [87, 88]. Aerolysin, derived from *Aeromonas hydrophila*, exhibits a narrower diameter than αHL of around 1 nm and a longer pore lumen of around 10 nm. These unique features of aerolysin provided an opportunity to study the effect of pore electrostatics on ion selectivity, ionic conductance, and molecular sensing capabilities [10, 88, 21]. The Bacterial curli transport lipoprotein CsgG:CsgF complex, a nonameric peptide transporter, from *Escherichia coli* also features a narrower pore constriction compared to αHL of 1 nm that shows a high sensitivity for DNA sequencing. Oxford Nanopore Technologies currently use this biological nanopore in their sequencing devices [89]. However, the utilization of αHL has been crucial in addressing numerous questions and advancing the field of nanopore sequencing. Pioneering studies in this domain were made possible thanks to the use of this protein nanopore. The advancements in nanopore sequencing technology have been significantly driven by the insights from αHL, demonstrating its pivotal role in developing this transformative technology. Without the foundational work using αHL, these remarkable achievements in nanopore sequencing would not have been possible.

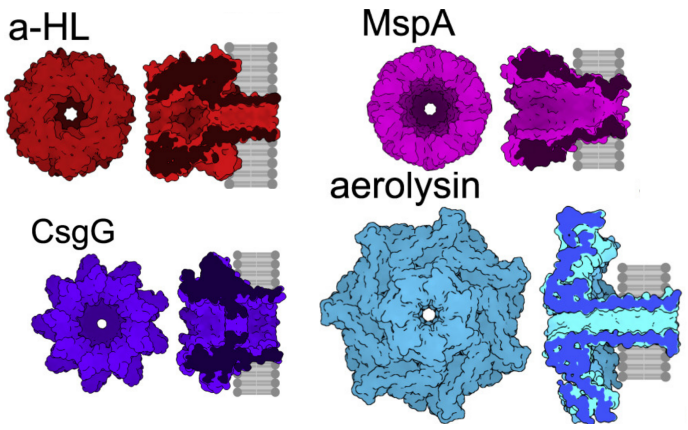


Figure 1.10: Figure adapted from Mayer *et al.* [88]. Some of the biological nanopores used in single-molecule nanopore experiments. The top view and the side view inserted in a lipid membrane are given for each biological nanopore.

1.2.2 DNA Unzipping Studies

When, in 1996, Kasianowicz *et al.* [13] discovered that αHL does not allow the passage of dsDNA molecules, the idea of studying unzipping using nanopores emerged. Unzipping via aforementioned methods such as optical tweezers [56, 57, 90] and atomic force microscopy (AFM) [58, 91] is highly precise but technically demanding due to the requirement for complex instrumentation and experimental conditions. Optical tweezers measure force by tracking the displacement of beads attached to the DNA ends; therefore, force is measured via a spring, the constant of which can be estimated from the thermal motion of the probe. AFM uses a cantilever to apply force directly to the molecule. Nanopore unzipping, in contrast, simplifies the setup by utilizing an applied voltage to induce unzipping, allowing for high-throughput measurements and eliminating the need for sophisticated mechanical manipulation or any covalent modification of the molecules studied. dsDNA domains can lodge in the 2.5 nm vestibule part but cannot enter the 1.5 nm β-barrel channel of the αHL nanopore. Therefore, the only way for it to translocate is via unzipping under applied voltage inside or at the cap region. Building on these advances, studies on unzipping through αHL were conducted in the early 2000s.

Sauer-Budge *et al.* [54] explored the unzipping mechanism of dsDNA through nanopores, considering various factors such as ionic strength, temperature, voltage, and DNA sequence. Their study demonstrated that the time distribution of strand openings can reveal detailed information about the unzipping process. They proposed a two-step unzipping mechanism: the first step involves the reversible unzipping of the DNA up to a 4 base-pair mismatch, while the second step encompasses the unzipping of the remaining strand. By varying experimental conditions, they measured several biophysical properties, including the unzipping reaction rate constants, the effective charge on nucleotides as they traverse the nanopore, and the enthalpy barrier height for the reaction. Their results showed that increasing ionic strength reduced the rate constants for both steps of the process. This reduction was attributed to the screening effect, where ions in the solution shield the negative charge on the DNA, reducing the electrostatic interactions that drive the process. As a result of this screening, the DNA appears to carry a lower effective charge, which they determined to be approximately $q \approx 0.1e$. This concept of effective charge will be

discussed in more detail in Chapter 3. Higher temperatures generally provided additional thermal energy, facilitating the overcoming of energy barriers and accelerating the unzipping process. Applied voltage increased the force exerted on the DNA, reducing the energy barrier and leading to shorter unzipping times; the characteristic unzipping time exhibited an exponential dependence on the applied voltage. The nucleotide sequence also influenced the local energy landscape, with regions of higher Guanine-Cytosine content presenting higher energy barriers due to the greater number of hydrogen bonds. These results underscored the nanopore technique's capability for efficient, label-free exploration of unzipping mechanisms and measuring biophysical properties under varying conditions.

Mathé *et al.* investigated DNA hairpin unzipping using nanopores with active voltage control. They performed unzipping experiments across a broad voltage range with different hairpin structures to measure the effect of mismatch and length of the hairpin. They observed that the characteristic unzipping time decays exponentially with increasing unzipping voltage. Additionally, they observed that the introduced mismatches facilitate the unzipping process, as seen in Figure 1.11. A simple two-state model, open and closed states according to Kramer's theory, effectively described the data for voltages above approximately 30 mV, yielding a consistent estimate of the effective charge on ssDNA within the αHL channel. Dynamic force measurements at high ramp rates (>4 V/s) supported the two-state model, demonstrating that the unzipping time correlates inversely with the logarithm of the ramp rate. Their approach allowed investigation into unzipping forces below the critical dsDNA-ssDNA transition, thus expanding the capabilities of traditional techniques.

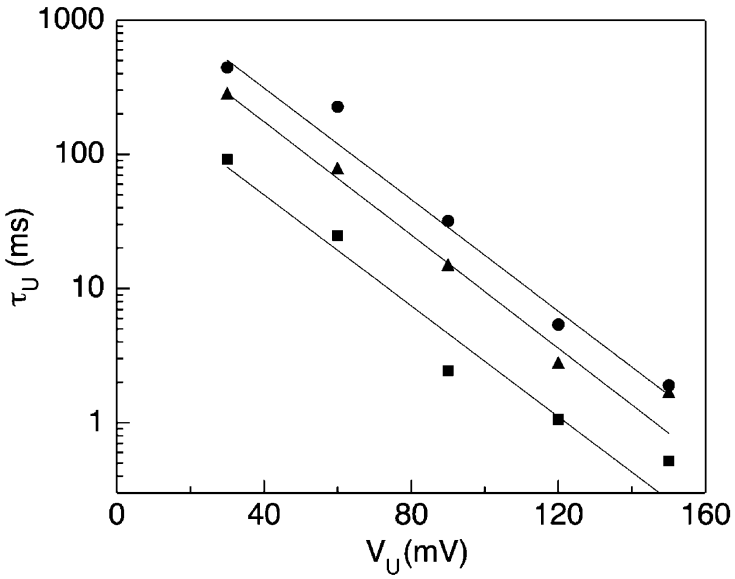


Figure 1.11: Figure taken from Mathé *et al.* [50]. The figure shows characteristic timescales as a function of applied voltage for different hairpin structures: the perfect match hairpin (HP1, solid circles), the single mismatch hairpin (HP2, triangles), and the 7-bp hairpin (HP3, squares). Exponential fits (lines) approximate the data well, with the same slope for all hairpins. The intercept values are approximately $\tau_0 \sim 2.1 \pm 0.2$ s for HP1, 1.2 ± 0.1 s for HP2, and 0.34 ± 0.05 s for HP3.

Dudko *et al.* explored the force-dependent kinetics of DNA unzipping by examining

DNA hairpins in nanopores [92]. They employed two experimental conditions: constant voltage and voltage ramp, measuring the forces necessary for hairpin unzipping. The study compared simple models, such as Bell's formula and Kramers' theory, with more sophisticated microscopic models. In the simplest models, force-induced rupture is described as the escape over an energy barrier, with the logarithm of the rupture rate (the forces required to break the bonds) changing linearly with force. In the voltage ramp experiments, the mean rupture force was found to vary linearly with the logarithm of the force loading rate. However, more complex microscopic models accounted for the nonlinear dependence of the logarithm of rupture rate on the force, which arises because the transition state distance decreases as force increases. Consequently, at high forces, rupture behavior diverges from Bell's prediction. Their findings indicated that the nonlinear relationship between force and the logarithm of the voltage ramp speed influenced the distribution of rupture forces.

Muzard *et al.* explored the geometric effects of αHL nanopores on both ssDNA translocation and DNA hairpin unzipping [59]. In their study, "forward translocation" refers to the movement of DNA from the *cis* side to the *trans* side of the nanopore, while "backward translocation" denotes the reverse direction, from the *trans* side to the *cis* side. For ssDNA translocation, they found that the translocation times of single-stranded poly(dA) were primarily dependent on DNA orientation upon entering the pore, irrespective of whether the translocation occurred in the forward or backward direction. The study revealed distinct blocked current levels enable differentiation between single-stranded and double-stranded regions within the same molecule. They confirmed that during DNA hairpin unzipping, introducing a single-strand overhang facilitated entry into the pore and enhanced mechanical force transduction from electrophoretic bias. Stable current levels in the forward direction indicated consistent DNA orientation within the vestibule cavity, while fluctuations in current levels in the backward direction suggested positional noise due to varying DNA-pore interactions. Discrimination criteria based on these fluctuations enabled the identification of unsuccessful translocation events originating from DNA-nanopore collisions. The unzipping process exhibited shorter times in the forward direction compared to the backward direction, influenced by positional fluctuations that affected mechanical force application.

1.3 Theoretical studies

Since the molecular details of ion and DNA transport through nanopores are inaccessible in experiments, theoretical studies are necessary to unveil the physical mechanisms and dynamics governing this transport process. With the increase in computational resources and methods, theoretical studies have complemented experimental findings over the decades. This section discusses continuous models and molecular dynamics (MD) studies focused on ion and DNA transport through nanopores.

1.3.1 Continuous models

Continuum theory approaches have been pivotal in understanding ionic transport within nanopores. The Poisson-Nernst-Planck (PNP) model is a widely used framework that

combines the Poisson equation for the electrostatic potential of the pore and the Nernst-Planck equation for ion fluxes. In the PNP model, ions are treated explicitly as charged species, and their transport is governed by diffusion and electromigration under the influence of both concentration gradients and electric fields. The nanopore is considered static. The Poisson equation is given by:

$$\nabla \cdot (\epsilon \nabla \phi) = -\rho, \tag{1.1}$$

where ϕ is the electrostatic potential, ϵ is the dielectric constant, and ρ is the charge density. The Nernst-Planck equation describes the flux $\mathbf{J}_i$ of ion species i as:

$$\mathbf{J}_i = -D_i \left(\boldsymbol{\nabla} c_i + \frac{z_i e c_i}{k_B T} \boldsymbol{\nabla} \phi \right), \tag{1.2}$$

where c_i is the concentration of ion species i, D_i is the diffusion coefficient, z_i is the ion valence, e is the elementary charge, k_B is the Boltzmann constant, and T is the temperature. The PNP model is particularly effective in describing ionic transport in channels and nanopores [93], providing insights into phenomena such as ion selectivity and conductance profiles in accordance with the experiments.

The Poisson-Boltzmann and Nernst-Planck (PBNP) model [94] extends the PNP model by incorporating the Poisson-Boltzmann equation to account for the equilibrium distribution of ions in the presence of an electrostatic potential ϕ:

$$\nabla^2 \phi = -\frac{e}{\epsilon} \sum_i z_i c_i^0 e^{-z_i e \phi / k_B T}, \tag{1.3}$$

where c_i^0 is the bulk concentration of ion species i. In the PBNP model, the electrostatic potential ϕ determines the ion distribution, which then influences the Nernst-Planck equation (1.2) to describe the dynamic behavior of ion transport. This model is particularly useful for predicting ion concentration profiles and current-voltage characteristics through narrow nanopores and channels. By solving these differential equations, one can obtain the electrostatic potential, ion concentration, and ionic flux. Parameters for these calculations are defined based on experiments or molecular dynamics simulations.

While continuum models offer computational efficiency and insights into ionic transport dynamics, they have notable limitations. The Poisson-Nernst-Planck (PNP) model encounters significant limitations when applied to narrow nanopores [95]. One critical issue arises from its continuum approach, which assumes a uniform charge distribution and smooth electrostatic potential variation over large spatial scales. In narrow nanopores, where dimensions approach or fall below the Debye screening length, these assumptions break down. The model fails to accurately capture high electric field gradients and ion crowding effects near the pore walls. Moreover, PNP neglects steric effects and ion-ion interactions that become pronounced in confined spaces, further limiting its accuracy [95]. Additionally, these models may oversimplify nanopore geometry, treating it as a static entity without considering conformational changes or dynamic interactions with surrounding molecules. Despite these challenges, PNP remains useful for qualitative predictions and exploratory studies in nanopore research, providing insights into ion selectivity and permeation [96, 97, 98], conductance characteristics [99, 100], and current-voltage relationships. The Poisson-Boltzmann and Nernst-Planck (PBNP) models address limitations

of simpler models by incorporating additional complexities such as ion-ion correlations and steric effects. The Poisson-Boltzmann (PB) component calculates the electrostatic potential by adjusting for ion interactions through modifications to the dielectric constant or using more complex potential models. The Nernst-Planck (NP) component describes ionic fluxes based on this potential and concentration gradients, with adjustments to diffusion coefficients and flux expressions to account for ion-ion correlations and steric effects. These combined effects help capture the flexible and dynamic nature of biological ion channels more accurately [101]. PBNP models provide accurate predictions of ionic transport parameters such as electrostatic potentials, ion concentration profiles, and current-voltage (I-V) curves through ion channels and nanopores [93, 101, 100]. Studies on αHL with these models revealed that the neutralization of the charged amino acids at the trans part of αHL reduces the current rectification [93] and the two charged residues at the constriction, Lys147 and Glu111, were identified to influence the selectivity [93, 100]. Additionally, PBNP allows for the simulation of large systems over long timescales at a relatively low cost and has been extensively used in studies concerning ion transport in αHL [93, 100, 102, 103].

The Brownian Dynamics (BD) method provides an intermediate approach between simplified continuum models and detailed Molecular Dynamics (MD) simulations. In this method, ions and the nanopore are treated explicitly, while solvent molecules are described implicitly. The interactions between the solvent and other molecules are characterized by frictional and stochastic forces, which are linked through the fluctuation-dissipation theorem. The equation of motion for a particle of mass m in Brownian Dynamics is governed by Langevin Dynamics, which includes both friction and random forces to account for the interaction of particles with the surrounding medium. The Langevin equation is given by:

$$m\frac{d^2\mathbf{r}}{dt^2} = \boldsymbol{\nabla} U(\mathbf{r}) - \gamma\frac{d\mathbf{r}}{dt} + \mathbf{F}_{\text{random}}(t), \tag{1.4}$$

where $\mathbf{r}$ is the position of the particle, $U(\mathbf{r})$ is the potential energy, γ is the friction coefficient, and $\mathbf{F}_{\text{random}}(t)$ is the random force due to thermal fluctuations.

In contrast to Langevin Dynamics, which explicitly includes both inertial and non-inertial forces, Brownian Dynamics simplifies the simulation by neglecting inertial effects. Brownian Dynamics focuses solely on the stochastic aspect of the motion, describing particle diffusion through the random forces and friction, without considering the inertia of the particles.

Typically, a single, uniform dielectric constant is used to represent the solvent in simulations. However, in confined environments like nanopores, the dielectric constant can vary significantly. To accurately capture electrostatic screening effects, it is crucial to use environment-dependent dielectric constants, which can be derived from molecular dynamics simulations. Studies employing Brownian Dynamics methods have shown their efficacy in modeling various aspects of ion channels, such as ion selectivity, conductance, gating mechanisms, and ion permeation pathways [100, 98, 104, 105]. Additionally, these methods are useful for understanding current fluctuations within channels and the stochastic movement of ions [100].

As an alternative to Brownian and Langevin dynamics, Monte Carlo simulations based on the Metropolis acceptance algorithm can be utilized to study ion transport [93, 106, 107]. In this method no differential equations are solved. Instead, an ion selected randomly

is moved stochastically within a defined maximum distance. Moves of the selected ions are proposed and accepted or rejected based on a probability criterion. This approach is advantageous for studying rare events because particle movement involves jumps between different states, allowing traversal of energy barriers within the system. This modeling technique has been notably employed to gather insights into ion transport through αHL, including obtaining current-voltage (IV) curves as a function of applied potential [93, 107].

In addition to ionic transport, various simulation methods have been employed to study polymer translocation through solid or biological nanopores at a reasonable computational cost. These methods include Brownian dynamics [108, 109, 104], Langevin dynamics [110, 111, 112, 113, 114, 104], and Monte Carlo simulations [106].

The theoretical research on polymer translocation dynamics often concentrated on the role of the Flory exponent (ν), which describes how a polymer's size scales with its length (N) in different solvent environments. This exponent is critical in understanding the behavior of polymers during translocation. For instance, Sung and Park [115] and Muthukumar [108] analyzed the equilibrium entropy of polymers and predicted a translocation time (τ) scaling as $\tau \sim N^2$ for field-free translocation, reflecting the polymer's entropic barrier. They suggested a linear dependence of τ on N for forced translocation, aligning with some experimental results for the α-hemolysin channel [13].

However, Chuang *et al.* [109] argued that for self-avoiding polymers, the translocation time scales as $\tau \sim N^{1+2\nu}$, considering the polymer's equilibration time ($\tau_{\text{equil}} \sim N^{1+2\nu}$). Kantor and Kardar [112] provided a lower bound for forced translocation time scaling as $\tau \sim N^{1+\nu}$. More recent studies using Monte Carlo simulations[106] and Langevin dynamics [114] verified the scaling of $\tau \sim N^{1+2\nu}$ for free translocation and observed a crossover from $\tau \sim N^{2\nu}$ for short polymers to $\tau \sim N^{1+\nu}$ for longer chains under force. This crossover remains unaffected for heteropolymer translocation.

Among these studies, Lou *et al.* explored the translocation time distributions of purines and pyrimidines using a hetero-DNA molecule model by employing both 2D Langevin dynamics simulations [110, 111] and Monte Carlo simulations [116]. Despite the simplifications of their DNA and pore model (a chain and bead model using only Lennard-Jones (LJ) interactions) and the 2D representation of their system, they demonstrated a faster translocation velocity for poly(dC) molecules. Additionally, their investigation of hetero-DNA translocation yielded that the translocation time distributions for hetero-DNA copolymers varied significantly depending on the sequence and composition of the copolymers. They found that certain copolymer sequences, especially those with larger block lengths of repeating units of poly(dC) and poly(dA), showed distinct translocation behaviors compared to poly(dA) and poly(dC) homo-polymer translocations, consistent with experimental observations [14, 15]. Lou *et al.* also noted that simulations at different forces reveal distinct translocation dynamics between different polymer structures. At lower forces, the translocation times were longer and more varied compared to higher forces. However, they did not account for the effect of orientation inside the pore, an important factor influencing translocation time as identified by various experimental studies.

Continuum theory studies further elucidated the effect of the type and intensity of external forces on translocation dynamics of polymer chains. Both voltage-driven [112, 106, 117] and pulling [112, 118, 114, 113, 119, 120] translocation models have been developed and studied. These studies consistently yielded results similar to those observed with voltage-driven experiments [13, 14, 15] including a wide distribution of translocation

times, an increase in translocation time and the variety of the translocation time with lower pulling forces, and a diminishing difference in translocation times between various polymers at higher pulling forces, regardless of the type of the external force applied (voltage or mechanical pulling).

In addition to constant driving forces, the translocation process has been examined theoretically under time-dependent external forces [121, 122, 123]. For example, Langevin dynamics simulations have explored polymer translocation under time-dependent alternating forces, revealing that resonant activation occurs at an optimal frequency. This means that the translocation rate is significantly enhanced when the frequency of the alternating force matches the system's natural dynamics, facilitating the movement of the polymer through the pore, particularly when the polymer-pore interaction is attractive [121]. This concept has practical applications in biology, such as the translocation of α-helical and linear peptides through an αHL nanopore under an AC field [124], and the use of alternating current signals to monitor DNA escape from αHL nanopores [125].

In summary, theoretical studies are useful to study the electrostatic properties of nanopores and ionic and polymer transport through nanopores at a low cost. The choice of a particular theoretical method depends on the aim of the study, the system under study, the desired resolution, and the computational cost.

1.3.2 Molecular Dynamics Simulations

The popularity of MD simulations for studying complex biological processes has grown significantly over the past few decades. Starting in the late 20th century, the exponential increase in computational power, coupled with the development of efficient algorithms and accurate force fields, enabled more realistic and detailed modeling of biological systems. The introduction of accessible software packages like GROMACS, CHARMM, AMBER, NAMD, and LAMMPS in the 1990s further motivated the use of MD simulations. As simulation results were validated against experimental data, confidence in the predictive power of MD grew, leading to its widespread adoption. The study of DNA translocation across nanopores using MD simulations began in the early 2000s after the discovery of the crystal structure of the αHL by Song *et al.*, [42] building on the foundation of experimental research in nanopore technology.

In this section, advancements in MD simulations utilizing all-atom (AA) or coarse-grained (CG) force-fields to study DNA translocation through nanopores, particularly focusing on αHL, will be elucidated.

All-atom simulations

The methodology for AA MD simulations is detailed in Section 2.2.2 of the Methods. Aksimentiev *et al.* initiated the study of DNA transport through artificial nanopores in 2004, providing invaluable insights and motivating further simulations involving protein nanopores [126]. In their work, they employed AA molecular dynamics to explore the microscopic kinetics of DNA translocation through synthetic nanopores. Key findings of their study included the observation of distinct ionic current blockades corresponding to different nucleotide forms and the effect of applied voltage on translocation speed. Their conclusions highlighted that translocation of dsDNA inside the nanopore can proceed at 10–100 base pairs per microsecond. They noted that interactions between the pore wall

and DNA base pairs can slow translocation through the pore. A strong hydrophobic interaction of the DNA bases with the pore's surface favors an unzipped conformation of dsDNA inside the pore. They observed that a significant reduction of the ionic current can be detected even when DNA is not transiting the pore, indicating that only part of the ionic current blockade measured experimentally reflects actual DNA translocation events. They observed that increased voltage speeds up the translocation process. The chosen voltages up to 22.1 V were used to collect data by accelerating the translocation process. These insights provided a quantitative understanding of DNA translocation through nanopores and laid the groundwork for future simulations to optimize nanopore shapes, coatings, and application of mechanical and electrical forces for more controlled DNA translocation, as well as to improve the resolution and accuracy of nanopore-based sequencing technologies.

In 2005, Aksimentiev *et al.* published another computational study, this time on ionic transport through αHL nanopores, demonstrating the potential of MD simulations to investigate biological confined systems and paving the way for computational studies of translocation phenomena through protein nanopores [127]. This study highlighted the behavior of ions within the nanopore, showcasing how molecular dynamics simulations can reveal the detailed mechanisms of ionic transport and interaction with the pore structure. Their repeating simulations at several voltage biases yielded αHL's current-voltage relationship as shown in Figure 1.12. The selectivity of αHL to Cl^- was found to depend on the direction and magnitude of the applied voltage bias. By analyzing the trajectories of all water molecules, they computed the αHL's osmotic permeability for water and its electroosmotic effect, and characterized the permeability of seven side channels. These side channels were found to connect seven His144 residues surrounding the stem of the protein to the bulk solution. The side channels facilitate the movement of water and ions, contributing to the osmotic permeability and electroosmotic effect of αHL. The protonation of these His144 residues was observed to affect the ionic conductance, suggesting the seven His144 residues gate conductance of the αHL channel. Subsequently, Bhattacharya *et al.* [118] expanded their investigation into the ionic characteristics of αHL by focusing on current rectification, which refers to the variation in ionic current through the nanopore based on the direction of the applied voltage, leading to asymmetric conductance properties. They studied how different alkali metal cations influence this rectification phenomenon in αHL pores. They discovered that both the direction and magnitude of current rectification are highly dependent on the cation type. This variation was attributed to the distinct interactions between the cations and the pore, which affect the distribution and movement of ions. Simulations were performed at transmembrane biases of $\pm$180 mV, $\pm$600 mV, and $\pm$1.2 V, with a duration of the MD trajectories varying between 20 and 30 ns for 600 mV and 1.2 V bias simulations and was at least 100 ns at 180 mV.

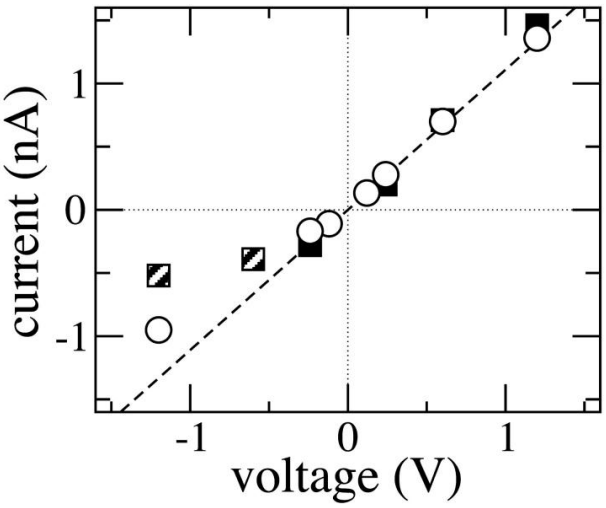


Figure 1.12: The current-voltage (IV) relationship of αHL, taken from Aksimentiev *et al.* [127]. Open circles represent currents computed under pH 8.0, while solid squares correspond to pH 4.5. Semi-solid squares indicate simulations where one of the loops at the trans end of the channel detached, momentarily obstructing the pore entrance. The dashed line represents a linear fit through the data point at 120 mV and the origin. Consistent with experimental findings [128, 1], the IV curve shows sublinear behavior at $V < 0$ in 1 M KCl [128, 1].

Zhao *et al.* [129] performed AA molecular dynamics simulations to study the translocation of ssDNA and dsDNA through solid-state nanopores, examining the effects of DNA molecular weight and structure on translocation velocity. Their results showed that ssDNA translocation velocities followed the sequence poly(dC) > poly(dT) > poly(dA) > poly(dG) of 20 nucleotides long, under a 300 mV transmembrane voltage potential. The translocation velocity for ssDNA under a 300 mV voltage potential was calculated as 27 bases/ns. For its dsDNA counterpart, the translocation velocity was calculated as 18 base pairs/ns. The study noted that much higher translocation velocities were observed in the simulations compared to experimental targets, indicating the need for simulations on much longer time scales (ms) and with smaller driving forces for practical device development. Given current computational limitations, they suggested using a CG method to model the systems closer to experimentally relevant time scales.

However, these first MD studies were limited by simulation times not exceeding a few hundred ns due to computational constraints at the time, including insufficient resources and the absence of GPUs. This necessitated increased external fields and accelerated transport processes to observe dynamic properties. Despite these limitations, these pioneering studies demonstrated that MD simulations offer both qualitative and quantitative understanding of translocation phenomena, reproducing numerous experimental properties such as the current rectification, selectivity for ions, and dependence of ionic current on the DNA structures, and encouraged extensive research on DNA transport through this αHL or other nanopores. While AA MD simulations are limited in their ability to simulate microscale timescale events, like DNA translocation, alternative techniques have been explored to investigate the dynamic properties of DNA translocation while conserving computational resources. Mathé *et al.* [63] pioneered the study by examining the static behavior of DNA molecules depending on the orientation of the translocation inside nanopores

without inducing translocation. Subsequently, techniques, such as Steered Molecular Dynamics, which will be explained in detail in Chapter 2, accelerated the permeation process through nanopores for translocation simulations [130, 131, 132, 133, 134, 135, 85]. Bond *et al.* [136] and Guy *et al.* [137] introduced truncated pore models to enhance computational efficiency by reducing the number of atoms simulated in MD simulations.

In a seminal investigation on DNA molecule orientation, Mathé and colleagues were the first ones to explore how the orientation of DNA molecules influences their behavior during the translocation process. Their experimental and theoretical analyses highlighted that the orientation of ssDNA within a confined and narrow cylinder affects its dynamics. Utilizing MD simulation techniques, they demonstrated that in a confined space, the bases of the ssDNA naturally tilt towards the 5' end. (See Figure 1.13) This observation suggested that during translocation, molecules oriented with their 3' end forward benefit from an upward tilt of their bases within the narrower sections of the pore, whereas those with their 5' end forward experience a downward tilt. However, it's important to note that Mathé *et al.* primarily focused on the structural conformation of ssDNA inside the pore through calculating the tilting angle of the bases, rather than the dynamics of the translocation process itself.

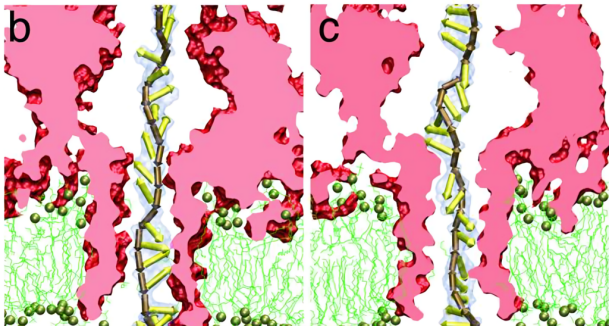


Figure 1.13: Figure taken from Mathé *et al.* [63]. The tilt of the DNA bases inside the αHL pore depends on the global orientation of the strand: molecules oriented with their 3' end forward benefit from an upward tilt of their bases (b), whereas those with their 5' end forward experience a downward tilt (c).

Some computational studies have employed non-equilibrium MD simulation techniques, such as Steered Molecular Dynamics (SMD) [130, 131, 132, 133, 134, 135, 85]. SMD involves applying an external force to one or more atoms of a molecule, mimicking the electric field-driven force used experimentally to guide the translocation molecule, DNA, through a nanopore. This significantly accelerates the permeation process, allowing multiple MD simulations to be executed concurrently. SMD studies have proven efficient in capturing the dynamics of DNA translocation through confined geometries, including various conformations that DNA may adopt within a nanopore.

Wells *et al.* [134] demonstrated the preferential orientation of ssDNA bases tilting towards the 5' end near the constriction site during DNA translocation using AA grid-SMD simulations. In these simulations, pre-computed interaction potentials between the DNA and the nanopore are stored in a grid format. During the simulation, these potentials are applied together with the steering forces to drive the ssDNA through the nanopore. Their work confirmed a slower translocation velocity for 5' oriented DNA yet without analyz-

ing any tilting angle. Additionally, they found that the ratio of permeation rates, entry rates to the nanopore, for poly(dA) is inferior to those of poly(dC) ssDNA molecules, and heteropolymers of poly(dCdA) molecules have a permeation rate between poly(dA) and poly(dC), validating the suitability of SMD for studying DNA translocation. However, they also noted that the translocation rate of a single molecule is 1000 times faster than in the experiment, and the corresponding applied voltages exceeded the experimental voltages by approximately 100 times.

Martin *et al.* compared poly(A) RNA and poly(dC) DNA using AA constant-velocity SMD (cv-SMD) [133] and Adaptive Bias Force (ABF) simulations [135], compared the results obtained by these two methods with previous experimental and computational studies. Their studies demonstrated the suitability of SMD simulations for studying the translocation of different nucleic acid molecules, and their observations closely matched experimental results [13]. Specifically, they provided insights into the free energy profiles of poly(A) and poly(dC) molecules during their translocations. Martin *et al.* revealed distinct energy barriers and interaction patterns for poly(A) and poly(dC) as they translocate through nanopores via two different SMD methods: cv-SMD and ABF. They observed that poly(A) RNA experiences higher energy barriers compared to poly(dC) DNA, which correlates with slower translocation speeds for poly(A). They also noted that this energy barrier comes primarily from the constriction of the αHL via electrostatic interactions caused by lysine at position 147 on each of the seven chains. These findings highlighted the impact of molecular structure and sequence on the translocation process using SMD simulations.

Studies exploring DNA translocation through a truncated αHL nanopore, β-barrel only, including the works of Bond *et al.* [136] and Guy *et al.* [137] have provided insights into the role of specific amino acid residues in the translocation process. These studies systematically showed that lysine residues at position 147 on each chain of the αHL nanopore interact with the phosphate groups of the DNA backbone through electrostatic interactions. Bond *et al.* demonstrated that these electrostatic interactions significantly influenced the translocation process by stabilizing specific conformations of ssDNA as it moved through the nanopore. Guy *et al.* highlighted that mutations introduced in αHL could alter the translocation dynamics of DNA: introducing a positive charge could hinder, and neutralizing a positive charge could facilitate the movement of ssDNA through the nanopore. These computational findings were consistent with experimental observations by Maglia *et al.* [78] and Stoddart *et al.* [18, 19, 65], which showed that positively charged residues, specifically lysine at position 147, significantly affecting the translocation velocity. In all these computational studies, electric voltage biases 3 to 6 times higher than those used in experimental setups were applied to induce DNA translocation events. This approach resulted in translocation times per base that exceeded experimental values by approximately 100 times.

Coarse-grained simulations

CG models simplify complex systems by regrouping atoms which smooths the energy landscape. They significantly increase the computational efficiency and allow the simulation time scales of large biological systems to be closer to the experimental time scales [138]. Consequently, CG MD simulations are an interesting option to study biomolecule translocation through nanopores. The details of the CG modelling methods are given in detail

in section 2.2.2, Methods.

In 2011, Ramachandran *et al.* [139] investigated how chemical modifications to solid-state nanopores affect DNA translocation using a CG model under an applied electric field. This model simplifies DNA by representing each nucleotide with two interaction sites: a backbone site (Bk) for the sugar and phosphate group and a base site (Tb) for the DNA bases. The interaction potentials between DNA sites were derived from matching thermodynamic properties obtained from AA MD simulations and experimental data. The CG model used an LJ potential, also derived from AA MD simulations, to describe interactions with the nanopore surface. The study revealed that the DNA translocation process is significantly influenced by the density and type of nanopore surface functionalization, the interaction potential between DNA and the pore surface, and the applied bias voltage. Functional groups on the nanopore surface alter electrostatic interactions, steric hindrance, and DNA affinity, affecting both the rate and smoothness of DNA movement through the pore. Additionally, an increased electric field accelerates DNA translocation by enhancing the force on the negatively charged DNA.

Among other studies utilizing a CG approach for DNA translocation process, the work of Stachiewicz *et al.* is particularly noteworthy. They utilized a modified MARTINI-like CG model; the MARTINI CG force field will be explained in detail in the following chapter. Adjustments were made to nonbonded interaction parameters, including increasing the dielectric constant and modifying ion-DNA interactions. The DNA model was adapted from Dans *et al.*,[140], six superatoms representing each nucleotide: PX for the phosphate group (P), KX and KN for the sugar (C5' and C1', respectively), and three pseudo atoms for each base, incorporating CHARMM27 force field parameters for bonds, angles, and dihedrals, while adjusting LJ parameters for DNA beads. They tested two solid nanopore configurations with diameters of 1.6 nm and 3 nm to simulate DNA unzipping and ionic current flow. They investigated the DNA unzipping process through nanopores subjected to an electric voltage, utilizing CG simulations [141, 142, 143]. Their unzipping approach was based on a protocol initially proposed by Comer *et al.* [144]: Hydrogen bonds were considered as broken when the distance between superatoms KN was larger than 4 Å. They used an inorganic, electrically neutral nanopore made of silicon nitride and the electrolyte according to the models explained above [141]. Their CG model enabled the examination of various translocation and unzipping parameters, considering factors such as pore size [141, 142], hairpin length, and DNA sequence [143]. Larger nanopore sizes facilitated faster DNA movement and more efficient unzipping than smaller pores, which restricted DNA passage. Longer DNA hairpins required more time and higher voltage to unzip due to their greater mechanical stability. DNA sequence variations also impacted the process, with different sequences affecting the interaction between DNA and the nanopore surface and influencing unzipping efficiency. Higher applied electric fields, the corresponding applied voltages for DNA translocation and unzipping simulations typically ranged between 0.1 and 1.0 V, increased the rate of DNA translocation by generating stronger forces to drive the negatively charged DNA through the pore. Their CG simulations thus revealed how nanopore dimensions, DNA structural features, and electric fields collectively affect DNA unzipping and translocation. However, they noted that their CG DNA model was not suitable for studying unzipping in nanopores with diameters smaller than 1.6 nm, rendering it inadequate for unzipping through the αHL nanopore, as the αHL nanopore has a constriction of just 1.4 nm.

In our laboratory, our group has been interested in computational simulations and experiments involving the αHL nanopore for over 10 years. With the arrival of the first PhD student, Delphine Dessaux, in 2016, research on this subject accelerated, combining computational studies with experiments. Before simulating DNA translocation with the MARTINI CG model, the group investigated the ionic and electrostatic properties of this protein nanopore to ensure the model's capabilities [145, 146, 147].

In 2019, our group published a study [145] on the ionic transport through αHL and found that the model accurately describes the experimental conditions under a potential bias. The ionic conductivity values with a polarizable water model, discussed in detail in the following chapter, at 320 K, were reported to agree with experimental and AA theoretical data for concentrations of KCl below 1.15 M. They also showed that the MARTINI model is accurate in reproducing electrostatic properties around a lipid membrane in the presence of an external electric field. They further performed CG molecular dynamics simulations of the ionic transport through αHL inserted in a lipid bilayer under different electric fields for 2–3 microseconds. The resulting IV curve was qualitatively consistent with experiments and showed asymmetry, although the computed current was one order of magnitude smaller due to the CG size of beads. The authors also noted that channel flexibility plays an important role in ion transport; with a fixed nanopore, no current was observed.

In 2022, our group investigated the current rectification and ionic selectivity of αHL [146, 147]. The role of charged amino acids in current asymmetry and anion selectivity was identified by approximately 100 CG MD simulations with numerous mutations on the αHL nanopore. The study demonstrated that current rectification originates from the cation current, which is significantly reduced at negative voltages through the pore (see Figure 1.14), due to the charges on the *trans* side, particularly Asp128. Anion selectivity was attributed to Lys147 residues located at the constriction. Extensive CG MD simulations revealed that current asymmetry and anion selectivity do not stem from the global charge of the pore or channel extremities but from the local electrostatic potential influenced by Asp128 and Lys147. This work highlighted the effectiveness of the MARTINI CG model in simulating the nanopore complex, enabling longer simulations and a broader range of voltages than AA simulations.

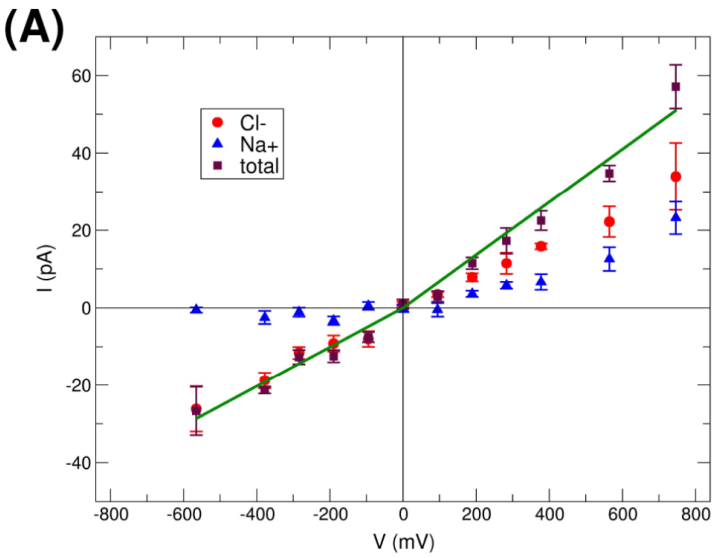


Figure 1.14: Figure taken from Dessaux *et al.* [147]. *IV* curves for the WT αHL. The asymmetric current response is shown in green, with lines representing linear regression fits for both positive and negative potentials.

This validation of the model opened the way to simulate nanopore molecular transport, such as DNA translocation. At the end of Delphine Dessaux's PhD, the group began studying DNA translocation through WT αHL under an applied electric field [146]. However, these CG MD simulations did not result in DNA capture by the nanopore or any translocation event. Subsequently, the use of SMD was implied, and the first encouraging results were obtained, yet without any complete translocation events of the DNA molecule used. These CG MD simulations were performed in a limited time; however, they encouraged the research to continue and became the subject of this thesis.

In the following Chapter, Methods, the details of single molecule analysis experimental and computational methods, specifically CG modeling with MARTINI, will be explained.

Chapter 2

Methods

Contents

2.1 Experimental Methods 37
2.1.1 Single-molecule experiments, experimental technique 37
2.1.2 Experimental Setup 39
2.1.3 *IV* Curves and Their Experimental Measurement 40
2.1.4 Analysis of Translocation Events in Nanopore Experiments . . 42
2.1.5 DNA Unzipping Experiments 44
2.2 Computational Methods 50
2.2.1 Molecular Dynamics Simulations 50
2.2.2 Ensembles in SMD simulations 56
2.2.3 Steered Molecular Dynamics (SMD) 58
2.2.4 All-atom and Coarse-Grained Molecular Force Fields 58

This chapter provides an overview of both the experimental and theoretical methods employed in this thesis. It delves into the experimental nanopore studies techniques and computational methodologies that were utilized to conduct the research.

2.1 Experimental Methods

2.1.1 Single-molecule experiments, experimental technique

Nanopore technology involves measuring the ionic current through either a biological or solid-state nanopore, which is embedded in an insulating membrane, thereby creating a separation between two chambers filled with an ionic solution [13]. When an electric potential bias is applied across this membrane, the ionic current is recorded. Biological nanopores are typically inserted into a lipid bilayer, while solid-state nanopores do not rely on such a bilayer. Instead, they are embedded in a thin sheet of solid insulating material, with the nanopores being created through methods such as ion beam sculpting, electron beam drilling, or chemical etching [23, 25].

In our experiments, we used a biological heptameric bacterial toxin named α-hemolysin, (αHL) as mentioned in the previous chapter. Hence, for the rest of this experimental methods section, only our experimental setup and protocol will be explained.

In the study of nanopore systems, the electrical behavior of the setup can be simplified to a resistance (R) for the pore and a capacitance (C) for the membrane. When we apply a voltage (V) across the system, the resulting current (I) is a combination of a resistive component (proportional to V) and a capacitive component (proportional to the rate of change of V). The current is expressed as:

$$I = \frac{V}{R} + C\frac{\mathrm{d}V}{\mathrm{d}t} \tag{2.1}$$

When a constant voltage is applied, the capacitive part becomes null, and the resistance of the pore stands out. As molecules, such as DNA or proteins, move through the nanopore, they cause temporary changes in the resistance of the pore. This resistance spikes sharply whenever a molecule passes through the pore. This phenomenon forms the basis of the "resistive pulse technique," also known as Coulter counting.

According to the Coulter Principle, as particles (like DNA molecules in our experiments) move through an aperture under an electric current, they cause changes in impedance. In nanopore experiments, the DNA molecules temporarily block the pore, leading to a drop in the ionic current. These drops, or "pulses," provide valuable data about the molecules' size, shape, and concentration.

This change in resistance (ΔR) when a particle is present in the nanopore can be expressed in terms of conductance. The initial conductance G_0 of the nanopore, before any particles enter, is related to the channel resistance R_{channel} by the following expression:

$$R_{\text{channel}} = \frac{1}{G_0} = \frac{4l + \pi d}{\sigma \pi d^2} \tag{2.2}$$

where σ is the ionic conductivity of the solution, d is the diameter of the nanopore and l is the length of the nanopore.

$R_{channel}$ can be expressed as the resistance of the nanopore plus any resistance, includeing the access to the nanopore from both sides, as illustrated in Figure 2.1, according to the formula:

$$R_{channel} = R_{pore} + R_{acces} \tag{2.3}$$

When a particle enters the nanopore, it causes a change in the pore's resistance (ΔR), resulting in a new resistance value R:

$$R = R_{channel} + \Delta R \tag{2.4}$$

The corresponding conductance G with the particle present can be calculated using equation 2.2

By monitoring these changes in conductance, the presence and properties of molecules translocating through the nanopore can be deduced.

Additionally, the capacitance C of the membrane in equation 2.2 is measured as an average over the thickness of the membrane according to:

$$C = \frac{\varepsilon A}{L} \tag{2.5}$$

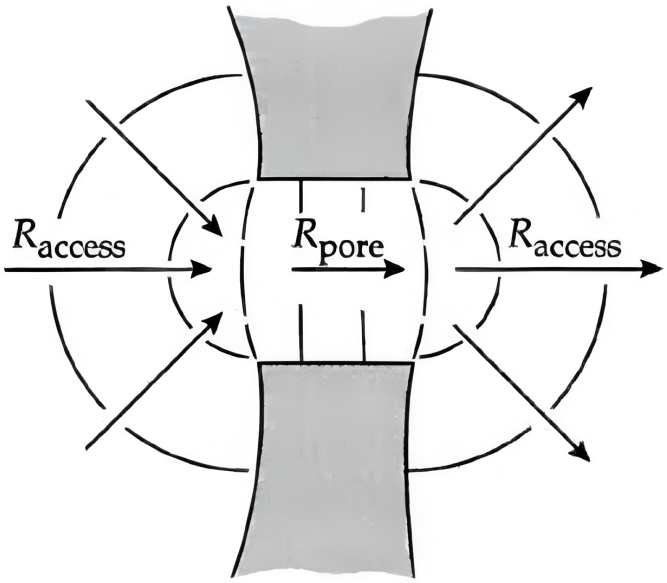


Figure 2.1: The resistance of the channel is the sum of the resistance from the nanopore and the surroundings, which is called the access. Insulating membrane is shown as gray rectangles [148].

where ε is the permittivity of the bulk, A is the surface areas of the membrane, L is the thickness of the membrane.

By combining the resistive pulse technique with the conductance measurements, a comprehensive understanding of the molecular transport dynamics and characteristics of the molecules interacting with the nanopore can be obtained by linking the observed changes in current and resistance to the physical processes occurring within the nanopore.

2.1.2 Experimental Setup

We conducted our nanopore experiments using a custom-made Teflon device inspired by Akeson *et al.* [61]. An illustration of this setup can be found in Figure 2.2. The setup consists of two chambers connected by a tube with a hole approximately 20–30 μm in diameter at its *cis* end. These two chambers, *cis* and *trans*, are filled with an ionic solution of the desired concentration. In our experiments, we used a 1 M KCl solution at pH 7.5 with Ag/AgCl electrodes installed in both chambers to apply a voltage and measure the ionic current. These electrodes are commonly used in electrochemical measurements because of their stable and well-defined reference potential. The Ag/AgCl electrode consists of a silver (Ag) wire coated with silver chloride (AgCl), which is immersed in a chloride-containing solution. When a voltage is applied, the potential difference across the chambers causes ions in the solution to move, thus generating an ionic current that can be monitored.

Initially, the hole was pre-treated with a 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC)-hexane lipid solution (at a concentration of 4 $\mu g.\mu L^{-1}$). For this purpose, 7 μL of the solution was deposited onto the hole, which was then dried. This step was repeated two times to ensure a good attachment of the lipid bilayer during the experiments, which

is the most important step for these single-molecule experiments.

Once the pre-treatment is performed, the chambers are filled with 100 μL of the buffer solution. A lipid bilayer is then formed across the conical aperture, forming bubbles using DPhPC-Xlipid solution, separating the *cis* and *trans* chambers. The capacitance of the membrane is measured to ensure it is thin enough for αHL to be inserted in a way that the β-barrel of the nanopore pierces the membrane from both sides. This is done by applying a small, known voltage and measuring the resulting current. The capacitance C is calculated using the equation 2.1, introduced above:

$$C = \frac{I}{\dot{V}} \tag{2.6}$$

where I is the current and $\dot{V}$ is the rate of change of the applied voltage. A typical setup involves a triangle wave voltage signal, and the resulting charging and discharging currents are analyzed.

To further verify the membrane's suitability, a high tension is applied: if the membrane bursts, it indicates it is appropriately thin for protein insertion.

αHL proteins are diluted in buffer (0.04 μg of αHL in 100 μL). A 1 μL aliquot of this solution is added to the *cis* side, and a single αHL channel insertion is monitored using a 120 mV potential. In the absence of the nanopore, the current is negligible due to the absence of an aperture between the chambers. Upon nanopore insertion, a significant rise in current, known as the open pore current, is observed. To prevent further protein insertions, the chambers are rinsed with the same ionic solution using a circulation pump. After confirming that only one αHL channel is inserted by the open pore current measurement, oligonucleotides of interest are added to the *cis* side at a concentration of 1 μM.

The current signals generated by the oligonucleotides are amplified and measured with an Axon 700B patch clamp amplifier. This amplifier converts the current into voltage, and a fourth-order Bessel low-pass filter is applied by the external Krohn-Hite filter with a cutoff frequency generally set to $f_c = 30\,\text{kHz}$. This filter configuration effectively reduces signal noise while preserving the translocation events.

The filtered signal is digitized using a National Instruments 16-bit acquisition card at a 1 MHz sampling rate (one sample per microsecond). A custom LabVIEW program (National Instruments) stores the digitized data on a computer hard drive and controls the patch-clamp amplifier. Depending on experimental requirements, either the entire signal or just the translocation events are recorded. For the latter, the LabVIEW program detects translocation events and triggers the recording of the relevant signal segments. This includes capturing 500 samples before (pre-trigger) and 500 samples after (post-trigger) each detected event to provide a comprehensive view of the translocation process.

2.1.3 *IV* Curves and Their Experimental Measurement

IV curves, also known as current-voltage curves, represent the relationship between the electrical current (I) flowing through a conductor or device and the voltage (V) applied across it. By analyzing the *IV* curve of a nanopore, valuable insights into its ion transport properties can be deduced. The slope of the *IV* curve stands for the pore's ionic conductance; the formula can be addressed in equation 2.2. Additionally, some nanopores

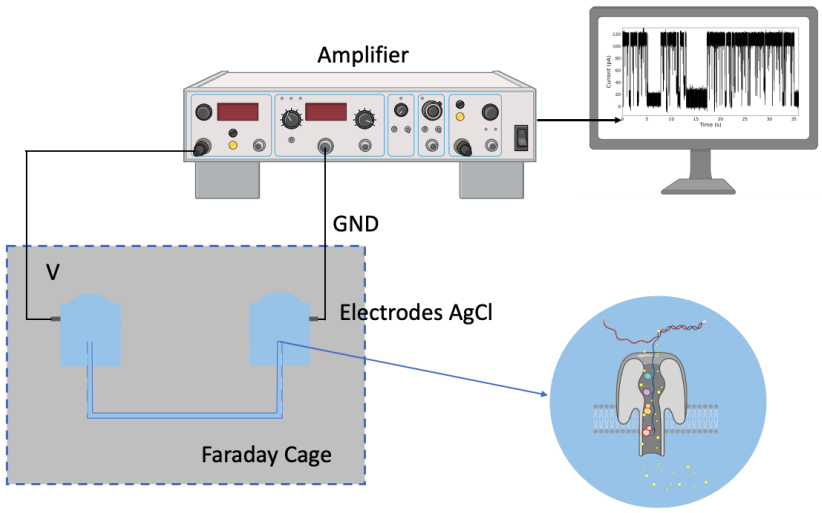


Figure 2.2: Illustration of the experimental setup. The cis side is magnified to show details. An α-hemolysin is inserted in a thin lipid bilayer, and DNA molecules are translocated through the pore under an applied voltage.

exhibit rectification, where their conductance depends on the sign of the applied voltage. This behavior is manifested as an asymmetry in the IV curve for positive and negative voltages. Finally, the shape and specific features of the IV serves as a fingerprint of a nanopore's ion transport characteristics.

To experimentally determine the IV curves of the αHL nanopore used in our experiments, we applied a triangular voltage waveform ranging from $-150\,\mathrm{mV}$ to $+150\,\mathrm{mV}$ at a frequency of a few Hz. Because the voltage isn't continuous, the capacitive component in Equation 2.1 becomes significant. Typically, voltage ramps of about $1\,\mathrm{V}\cdot\mathrm{s}^{-1}$ across a membrane with a capacitance of $15\,\mathrm{pF}$ generate a capacitive current of $15\,\mathrm{pA}$. As a result, IV curves obtained using ascending voltage ramps are shifted by $+15\,\mathrm{pA}$, while those obtained with descending ramps are shifted by $-15\,\mathrm{pA}$ to account for this effect.

The voltage signal generated by the amplifier comes from a signal emitted by the computer through a LabVIEW program. This creates a ramp function that isn't a true straight line but rather a staircase function due to the digital nature of the signal generation. Each step in this staircase function induces a capacitive response, increasing signal noise. To address this and eliminate the capacitive component, 50 consecutive ascending and descending ramps are recorded. By averaging the ascending and descending slopes, the signal noises are effectively reduced, and the capacity component is cancelled, resulting in an IV curve that passes through the origin.

The IV curve of the αHL nanopore has several key characteristics. Firstly, the αHL nanopore displays rectification, meaning the current flow is significantly higher in one direction compared to the other for positive and negative voltages. This rectification can be attributed to the asymmetric structure of the nanopore and the distribution of the charged amino acids wthin the nanopore, which affects ion flow differently depending on the voltage polarity. Secondly, the IV curve is also dependent on the surrounding environment. The acidity/alkalinity (pH) and the concentration of dissolved salts (ionic strength) can alter the flow of ions through the channel, further influencing the shape of

the IV curve.

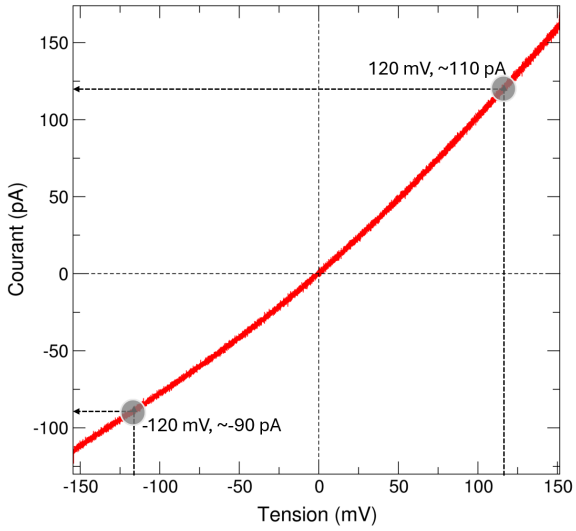


Figure 2.3: Current-Voltage relationship of αHL nanopore at 20°C at pH 7.5 in 1M KCl solution. The signature values of αHL nanopore at +120 mV, 110 pA and at -120 mV, -90 pA are shown on the plot.

The typical IV curve of αHL nanopore in 1M KCl solution, at room temperature is given in Figure 2.3. At +120 mV, we observe 110 pA while at -120 mV, the current is around -90 pA.

2.1.4 Analysis of Translocation Events in Nanopore Experiments

Principle of Analysis

The analysis of translocation events in nanopore experiments involves identifying and characterizing the dynamics of molecules passing through the nanopore based on recorded electrical signals. One of the fundamental parameters analyzed is the Open Pore Current (I_0), which represents the current when no molecule is obstructing the pore. This measurement serves as a reference to normalize data across different experiments and to establish a baseline for comparison. On the other hand, the Blocked Pore Current (I_b) refers to the current observed when a molecule translocates through the pore, either partially or completely blocking it. By analyzing I_b, characteristics such as the size and nature of the translocating molecule can be discerned. Another critical parameter is the Time Between Events (t_i), which denotes the interval between the onset of successive translocation events. This metric provides insights into the frequency and temporal distribution of molecule passage through the nanopore. Additionally, the Translocation Duration (t_u) measures the time taken for a molecule to traverse the nanopore entirely, the decision for calling this parameter t_u is due to the unzipping experiments, which will be explained shortly. This duration offers valuable information about the dynamics of the processes governing the translocation and unzipping events. Figure 2.4 illustrates a typical current trace during nanopore experiments, highlighting these key parameters: Open Pore Current (I_0), Blocked Pore Current (I_b), Time Between Events (t_i), and Translocation Duration (t_u).

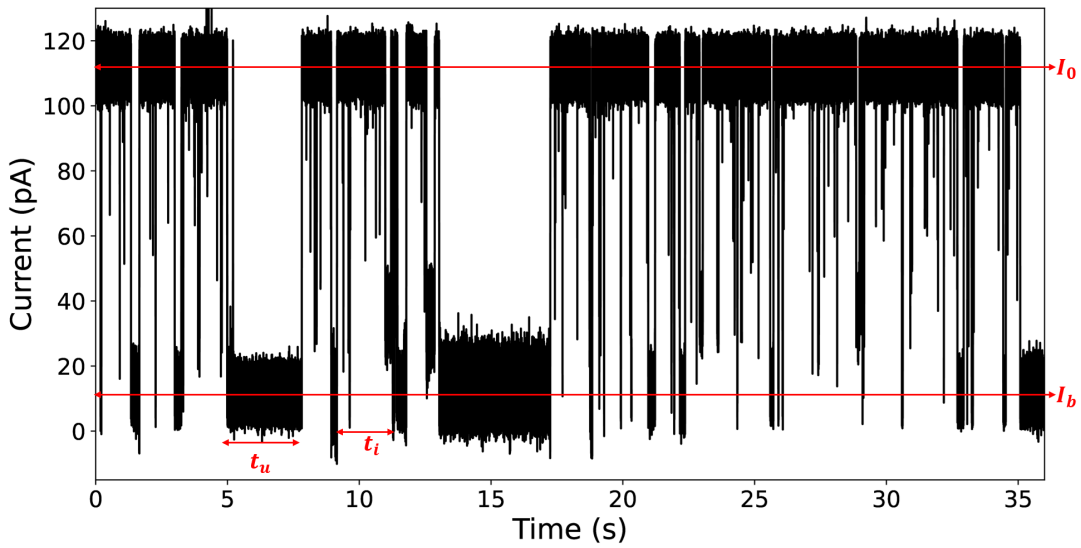


Figure 2.4: Current Trace during the recording of experiments with HP39sf molecules. Open pore current (I_0), Blocked pore current (I_b), Time between events (t_i), and Translocation duration (t_u) are shown.

Histogram Analysis

The analysis of a specific DNA molecule of interest in nanopore experiments begins with constructing a histogram of the monitored current based on all recorded current samples. Figure 2.5 illustrates a typical distribution of current levels observed during the recording process. In the histogram, two primary peaks are typically discernible: The I_0 represents the baseline current level when the nanopore is unobstructed by any molecule. This peak serves as a reference for the normal operating state of the nanopore and the I_b corresponds to the current level when a DNA molecule, either partially or completely, blocks the nanopore during translocation. This peak is typically of lower magnitude compared to the open pore peak because the majority of the recording time is spent in the open state of the nanopore.

Additionally, the histogram may show extra peaks, $I_{undesired}$ indicating undesirable events recorded during data acquisition. These peaks could arise from unsuccessful translocation attempts, rapid translocations of impurities, non-paired single-stranded DNA (ssDNA), or mispaired double-stranded DNA (dsDNA) rather than the intended dsDNA unzipping events in our experiments.

Thresholding, Event Identification, and Parameter Extraction

Once the histogram of the recorded current is analyzed, the next step is to distinguish between valid translocation events and unwanted noise across the collected data. To achieve this, two thresholds are defined: the detection threshold, which is used to identify the beginning and end of a translocation event and is typically set slightly above the level of the blocked pore current peak (I_b commonly noted to be around 10% of I_0), and the rejection threshold, which is used to eliminate poorly recorded events and is set below the open pore current noise level. An event is considered valid only if it meets specific

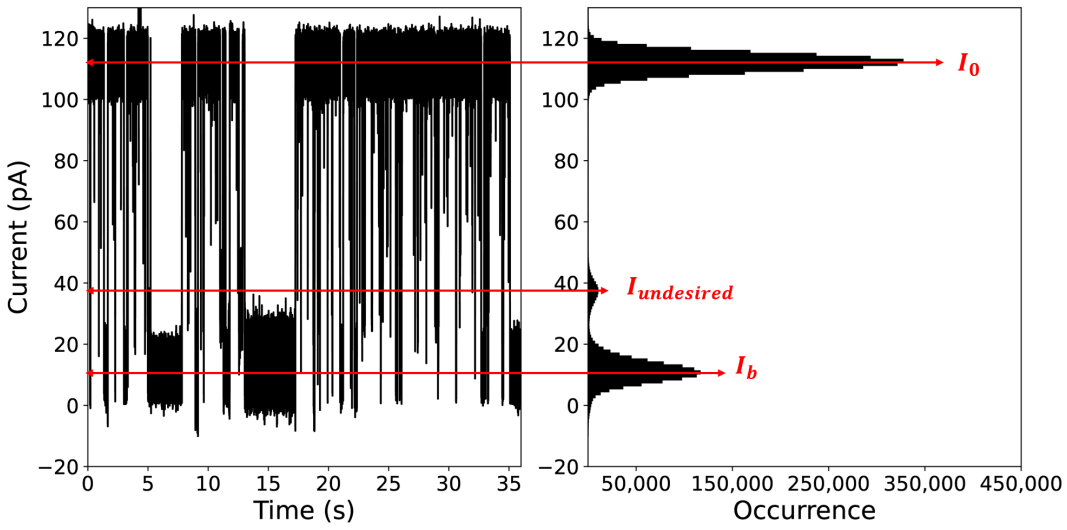


Figure 2.5: Current Trace and the distribution of the current during the recording of experiments with HP39sf molecules. On the plot of the distribution of the current trace, three peaks are visible: I_0, $I_{undesired}$, and I_b.

criteria: the average open pore current before and after the event must be above the rejection threshold, and all blocked pore current (I_b) samples within the event window must fall below the detection threshold. This analysis is typically performed automatically using dedicated software, such as IgorPro (Wavemetrics), which can be programmed to extract the translocation times (t_u), inter-event times (t_i), open pore currents (I_0), and blocked pore currents (I_b) for each identified event.

Building on these analyses and protocols, the next section will detail how the unzipping experiments are analyzed and the methods employed for their investigation.

2.1.5 DNA Unzipping Experiments

Our study aims to investigate how the length and structure of hybridized sequences influence the opening time of DNA molecules using the αHL nanopore. In these experiments, we utilize the experimental setup and protocol described in Section 2.1.2. The oligonucleotides used in this study include HP9s, HP9sf, 19s, 19sf, 29s, 29sf, 39s, 39sf, and O60. The detailed sequences of these oligonucleotides will be provided shortly.

The DNA molecules of interest are introduced from the *cis* side under an applied electrical voltage bias, and the typical recording procedure begins. The applied voltage serves as the force required to unzip the molecules through mechanisms that differ based on the structure of the molecules, whether flap or blunt-end. These differences will be detailed shortly. The translocation duration (t_u) is defined as the total time from the insertion of the DNA into the nanopore, through its unzipping, to its complete translocation and exit from the nanopore. In our experiments, the unzipping phase is the most time-consuming part of this process. Therefore, the overall translocation time (t_u) is predominantly influenced by the unzipping duration. This is why we use t_u to refer to the total translocation time.

The sequences of the oligonucleotides introduced above are listed below:

- HP9s: 5'-CGC-TAT-CCA-CCC-CTG-GAT-AGC-G-$[A]_{37}$-3'
- HP9sf: 5'-AAA-AAA-AAA-ACG-CTA-TCC-ACC-CCT-GGA-TAG-CG-$[A]_{40}$-3'
- 19s: 5'-CGC-TAT-CCA-CTC-TTC-ATT-A-3'
- 19sf: 5'-AAA-AAA-AAA-ACG-CTA-TCC-ACT-CTT-CAT-TA-3'
- 29s: 5'-CGC-TAT-CCA-CTC-TTC-ATT-ACC-AAC-CTG-TC-3'
- 29sf: 5'-AAA-AAA-AAA-ACG-CTA-TCC-ACT-CTT-CAT-TAC-CAA-CCT-GTC-3'
- 39s: 5'-CGC-TAT-CCA-CTC-TTC-ATT-ACC-AAC-CTG-TCC-GCT-CCG-CTA-3'
- 39sf: 5'-AAA-AAA-AAA-ACG-CTA-TCC-ACT-CTT-CAT-TAC-CAA-CCT-GTC-CGC-TCC-GCT-A-3'
- O60: 5'-GCG-AAG-ATG-ACG-ATG-GAT-AGT-AGC-GGA-GCG-GAC-AGG-TTG-GTA-ATG-AAG-AGT-GGA-TAG-CGC-$[A]_{39}$-3'

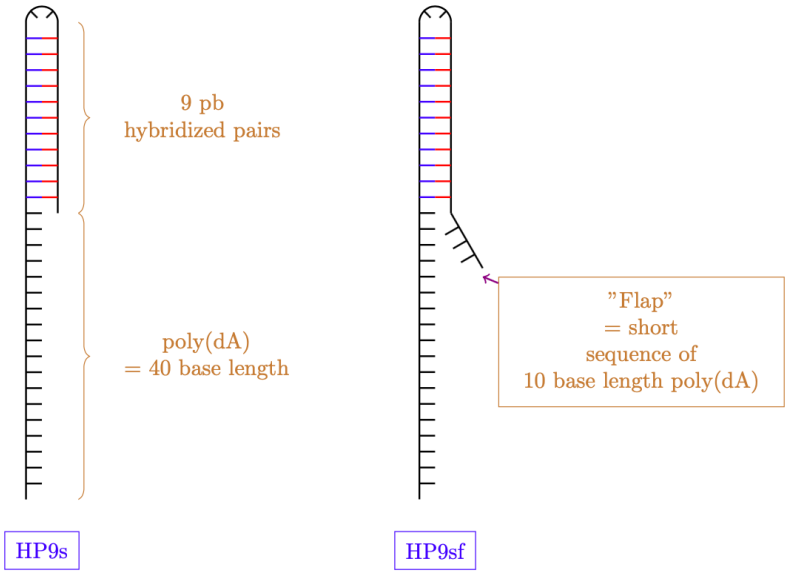


Figure 2.6: Design of the experimental dsDNAs: HP9s and HP9sf sequences fold into hairpin structures. The difference between the s and sf structure is coming from the presence of a short unpaired sequence of a poly(dA) structure present in the sf structure which is called a flap.

The HP9s and HP9sf sequences fold into hairpin structures (see Fig 2.6). The O60 sequence serves as a template for pairing sequence strands longer than 9 bases (see Fig 2.7). All structures with the same number of hybridized bases (s or sf) are expected to have

the same pairing energy. The key difference between blunt-end and flap structures of all lengths lies in the presence of a short non-hybridized sequence at the 5' end of the flap structure strand, as shown in Figure 2.8. This free, non-hybridized DNA sequence, known as a "flap," may hinder the entry of the double-stranded portion into the αHL vestibule. Therefore, molecule opening is likely to occur at the cap entrance in a zipper-like fashion, facilitating the unzipping. When the DNA has a blunt end, the unzipping is likely to occur within the vestibule as a result of the shear forces, which can be seen in Figure 2.8). The study aims to examine the influence of a flap's presence on the molecule's unzipping time (t_u) as a function of hybridized portion length.

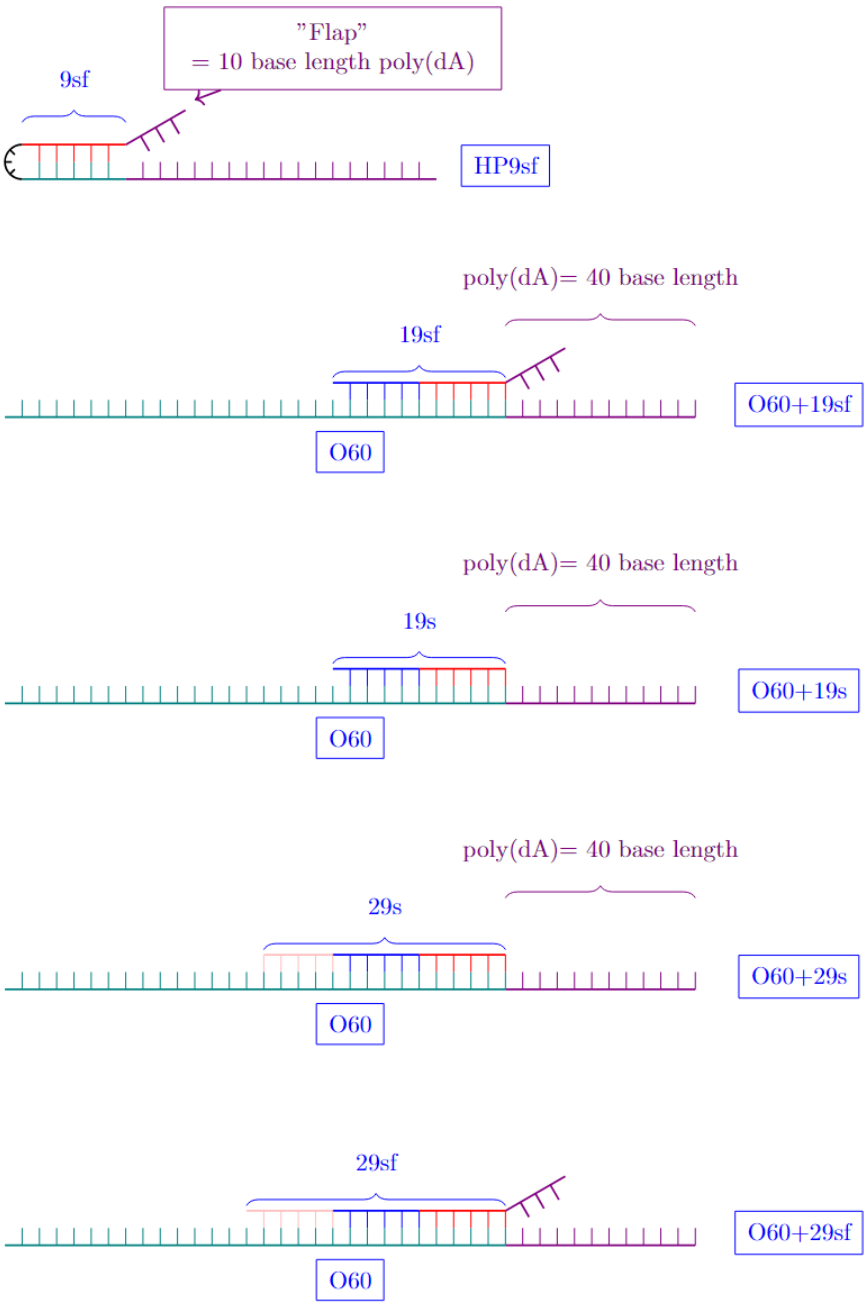


Figure 2.7: Design of the experimental dsDNAs.The O60 sequence serves as a template for pairing with longer sequence strands than 9 bases. The same color is used to indicate the identical sequences used to pair longer sequences with the O60 sequence. Each time a new sequence of 10 nucleobase is added to make the hybridized portion longer.

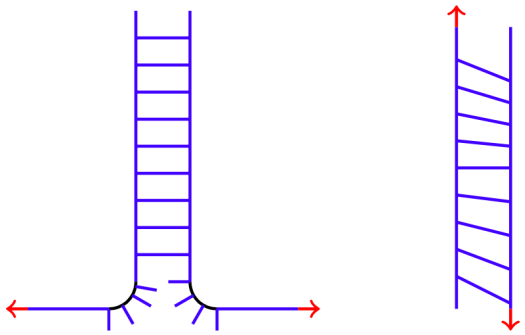
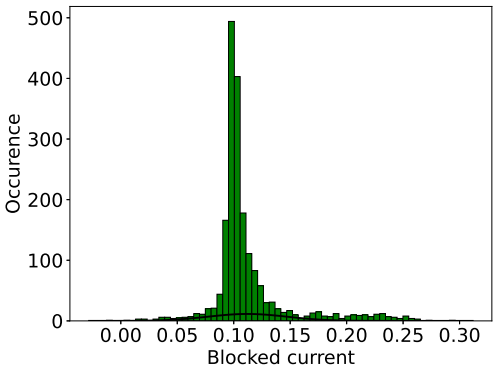
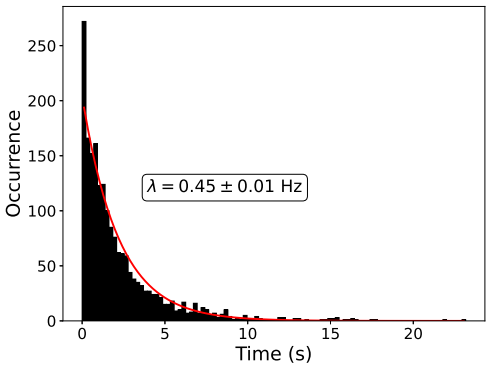


Figure 2.8: Illustration of the opening mechanisms in the presence of a flap (left) and blunt-end (right). In the presence of a flap structure, the mechanism of opening is likely to occur in a zipper-like fashion. In the absence of this flap, when a blunt-end is present, the opening is likely to occur as a result of shear forces.

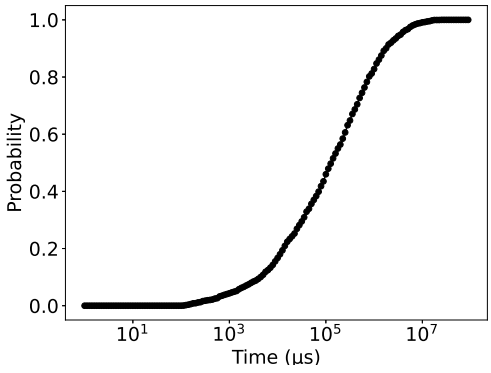
By plotting the normalized blocking current (I_b/I_0) against the unzipping time (t_u), where each event corresponds to a point, which will be called event clouds in the forthcoming chapters, we can identify distinct clusters of points that represent different categories of events. These clusters enable us to isolate and analyze specific events of interest. To assess the relative density of each event, we use the Kernel Density Estimator (KDE). This method involves calculating the Gaussian function for each point in terms of both the normalized blocking current and the logarithm of unzipping time, thereby determining the probability density function (PDF) for each point based on its location and the surrounding data points. We then assign density values to each point according to the Gaussian distributions calculated in the previous step, summing the contributions of all Gaussian functions at each point to determine the overall density. To visualize these densities, we use a color code, where each point is colored according to its density value, with different colors representing different density levels. This KDE color-coded representation allows us to observe the distribution and concentration of unzipping events more clearly and to analyze the characteristic unzipping times. The histograms are also generated to analyze normalized blocking currents (I_b/I_0) in DNA unzipping experiments. This histogram typically exhibits a Gaussian distribution centered around approximately 0.1 ($I_b = 0.1 \times I_0$), facilitating standardized comparisons across different experimental conditions. Inter-event times (t_i) following an exponential distribution indicate the statistical independence of events. While analyzing the inter-event time, performing an exponential decay fit with $Ae^{-\lambda t}$ on this histogram on a normal scale, as shown in Figure 2.9b, allows for the extraction of the frequency of the events, λ, usually on the order of Hz, providing insights into the stochastic dynamics of DNA unzipping within the nanopore environment.



(a)



(b)



(c)

Figure 2.9: Histograms of (a) normalized blocked current (b) The time between the start of each event, defined as the inter-event time (t_i), which is fitted by an exponential decay function $Ae^{-\lambda t}$, λ being the frequency of events, A being a constant (c) Probability, represented as $1-e^{-t/\tau}$ of unzipping corresponding to a HP19s under 120 mV bias voltage

The characteristic unzipping time (t_u) can be determined from the integrated unzipping probability, which describes the fraction of unzipping events that have occurred by a given time. The integrated probability is defined as:

$$P_u(t) = \frac{N(t)}{N_t} \tag{2.7}$$

where $N(t)$ is the number of events with unzipping times less than or equal to t, and N_t is the total number of events. This probability function captures the accumulation of unzipping events over time, providing insight into the dynamics of the process.

To analyze the unzipping behavior, the integrated probabilities are plotted as a function of time, producing sigmoidal curves on a semi-logarithmic scale, as illustrated in Figure 2.9c. These curves reflect the gradual increase in unzipping events over time, with the majority of events occurring after an initial delay. The curves are fitted using the sigmoidal function:

$$P_u(t) = 1 - e^{-t/\tau} \tag{2.8}$$

where τ represents the characteristic time constant of the unzipping process. The parameter τ, extracted from the fit, provides a measure of how quickly the unzipping events occur on average.

The characteristic unzipping time, t_u, is defined as the time at which the integrated probability reaches 0.5, indicating that half of the unzipping events have taken place. This point corresponds to the inflection point of the sigmoidal curve and is calculated directly from the experimental data. The fitted time constant τ, while related to the time scale of the process, does not exactly coincide with the characteristic unzipping time.

By plotting the integrated probability curves for different molecules and conditions, we can compare their unzipping times and extract both the characteristic unzipping time, t_u, and the time constant τ. This analysis allows for a detailed comparison of the unzipping dynamics under varying experimental conditions, with the inflection point of the curves shifting to longer times for molecules that take longer to unzip.

With these explanations, the experimental methods are concluded. In the next section, we will explain the computational methods, more specifically, the ones we utilized during this thesis.

2.2 Computational Methods

In this thesis, we used classical Molecular Dynamics (MD) simulations to study the translocation of ssDNA. In this section, we will review the methods and frameworks of MD simulations, with a particular focus on the coarse-grained approach using the MARTINI force field.

2.2.1 Molecular Dynamics Simulations

MD simulations function as computational microscopes, offering insights into the behavior of atomic and molecular systems. At its core, MD involves numerically integrating Newton's equations of motion to track the position and velocity of each atom or molecule over time, collectively referred to as a molecular trajectory.

MD simulations operate on principles of classical mechanics and statistical physics. Starting with well defined atomic coordinates, velocities, and other parameters like temperature or pressure, the MD algorithm iteratively advances the system's state through successive time steps. This iterative process generates a detailed trajectory that captures the dynamic evolution of the molecular system. By analyzing this trajectory, information about the system's structural fluctuations, thermodynamic properties, and kinetic processes can be extracted. At each time step, forces acting on each atom are computed based on the system's current configuration and force field parameters, which will be detailed in the following section. These forces govern the subsequent movement of particles, leading to incremental changes in their positions and velocities.

The accuracy of MD simulations depends on several factors, including the precision of force field parameters, appropriateness of simulation conditions, and adequacy of the sampling strategy used. Additionally, simplifications are often made to increase computational efficiency without compromising the essential physics. Atoms are typically treated as point particles with well-defined masses devoid of internal degrees of freedom, such as electronic excitation. Moreover, the assumption of perfectly elastic collisions between

atoms ensures the conservation of total energy throughout the simulations. This simplifies the computational framework while maintaining the overall integrity of the system dynamics. However, it's important to acknowledge that these simplifications inherently limit the scope of conventional MD simulations, particularly in scenarios involving bond formation, bond breaking, and charge transfer events. This limitation underscores the necessity of integrating MD simulations with complementary computational methods, when needed, such as quantum mechanics-based approaches or hybrid methods like QM/MM (Quantum Mechanics/Molecular Mechanics), to address phenomena involving electronic rearrangements or chemical transformations.

Force Fields

The most important step of MD simulations is the accurate calculation of potential energy, which is essential for predicting the position and velocity of each atom within the system over time. This potential energy inherently depends on the force field employed, which defines the interactions between atoms. The force field, comprised of empirical parameters, assigns specific energy values to different types of atoms present in the system. These parameters are typically derived from a combination of experimental data and quantum mechanical calculations, ensuring that the force field captures the essential features of molecular interactions. It's important to note that there are several types of force fields used in molecular dynamics simulations: all-atom, united-atom, and coarse-grained (CG). All-atom force fields explicitly represent each atom in the system, considering its individual interactions with other atoms. These force fields provide detailed representations of molecular structures and interactions, making them suitable for studying systems with high resolution. United-atom force fields further simplify the representation by combining certain atoms into single interaction sites, typically hydrogen atoms with their adjacent heavy atoms. This approach reduces computational costs compared to all-atom models. CG force fields simplify the representation of molecules by grouping multiple heavy atoms into single interaction sites. While they sacrifice some level of detail compared to all-atom force fields, CG models are valuable for studying larger biomolecular systems and processes that occur over longer timescales.

The potential energy in MD simulations is commonly divided into two distinct components, independent of the type of force field used: bonded interactions and non-bonded interactions.

Bonded interactions encompass atoms connected by covalent bonds, while non-bonded interactions consider the forces exerted through electrostatic and Van der Waals interactions. Figure 2.10, extracted from *Waidyasooriya et al.* [149], visually illustrates the bonded and non-bonded interactions considered in MD simulations.

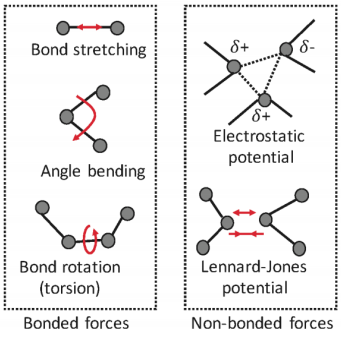


Figure 2.10: Bonded and non-bonded interactions considered in MD simulations [149].

The total potential energy U is expressed as the sum of bonded (U_{Bonded}) and non-

bonded ($U_{\text{Non-Bonded}}$) components:

$$U = U_{\text{Bonded}} + U_{\text{Non-Bonded}} \tag{2.9}$$

Bonded potential Energy The bonded potential energy (U_{Bonded}), see equation 2.10, in molecular dynamics simulations encompasses several components: covalent bonds (U_{Bonds}), bending angles (U_{Angles}), and torsion dihedral angles ($U_{\text{Dihedrals}}$).

$$U_{\text{Bonded}} = U_{\text{Bonds}} + U_{\text{Angles}} + U_{\text{Dihedrals}} \tag{2.10}$$

Harmonic potentials are typically associated with these bonds and angles. The constants involved correspond to the stiffness of these springs, indicating how strongly the system resists deviations from equilibrium values. The dihedral angle potential, $U_{Dihedrals}$, applies between two atoms separated by three covalent bonds, with parameters representing the height of the energy barrier and the periodicity of the rotation. The equilibrium values for bond length, bond angle, and dihedral angle are the values at which the respective potential energies reach their minimum.

An "improper" dihedral angle potential can also be added in structures that exhibit greater rigidity, such as groups with double bonds. In this case, a harmonic potential complements $U_{\text{Dihedrals}}$ between atoms 1-4 separated by three covalent bonds.

However, it is important to note that depending on the force field used, the equations used for treating bonded and non-bonded interactions can differ. In the following sections, we will explain these bonded interactions specifically for the MARTINI force field, detailing the particular formulations and constants used in this context.

Non-bonded potential energy The non-bonded potential energy $U_{\text{Non-Bonded}}$ is given by:

$$U_{\text{Non-Bonded}} = U_{\text{Elec}} + U_{\text{VdW}} \tag{2.11}$$

$$U_{\text{Non-Bonded}} = \sum_{\text{i}} \sum_{\text{j}} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} + \sum_{\text{i}} \sum_{\text{j}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.12}$$

The electrostatic interactions are defined between two charged particles according to the Coulomb's Rule which can be consulted in equation 2.12 where ε_0 is the permittivity of free space, ε is the relative permittivity (dielectric constant), q_i and q_j are the charges of atoms i and j, r_{ij} is the vector distance between atoms i and j. Van der Wals interactions are often modeled using the Lennard-Jones (LJ) potential, also known as the "12-6" potential. This potential captures both the attractive and repulsive forces and is given by the equation 2.12 where ϵ_{ij} is the depth of the LJ potential well between particles i and j, which varies depending on the specific pair of interacting particles, resulting in different interaction levels. This parameter ϵ_{ij} is calculated using the Berthelot rule, taking the geometric mean of the ϵ values of the individual particles depending on the force field: $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$. σ_{ij} is the distance at which the potential energy between atoms i and j is zero, effectively defining the size or diameter of the particles. The parameter σ_{ij} is calculated using the Lorentz rule, which averages the σ, LJ radius, values of the individual particles: $\sigma_{ij} = (\sigma_i + \sigma_j)/2$. The attractive forces, known collectively as van

der Waals forces, include Keesom, Debye, and London interactions, and are represented by a term proportional to $\frac{1}{r_{ij}^6}$. Conversely, the repulsive component stems from the Pauli exclusion principle, which prevents the electron clouds of two atoms from overlapping. This repulsion is described by a term proportional to $\frac{1}{r_{ij}^n}$, with n typically in the range of 10 to 12, usually dominating at very short distances.

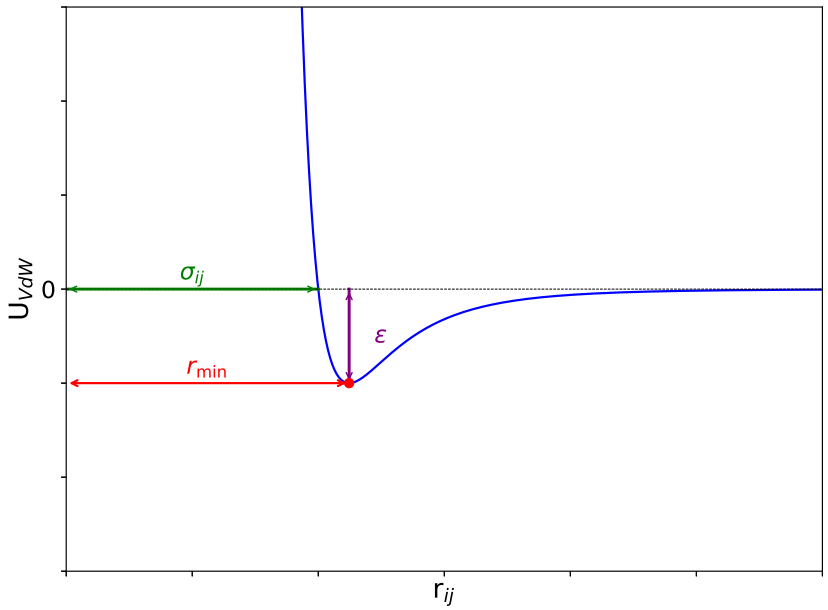


Figure 2.11: Representation of the Lennard-Jones potential as a function of the distance r_{ij} between two atoms i and j. The parameters σ_{ij} and ε_{ij} are indicated on the graph according to equation 2.12. $r_{\min}$, the distance between two particles at which the LJ potential energy reaches its minimum, is also shown in red.

Efficient computation of non-bonded interactions poses challenges due to their long-range nature and the necessity to consider interactions between all atom pairs. Various strategies, such as cutoff distances and Ewald summation techniques [150, 151], are employed to mitigate these computational demands while maintaining accuracy. In MD simulations, the introduction of a cut-off radius impacts van der Waals (VdW) and electrostatic interactions differently. VdW interactions decay rapidly with distance (r_{ij}^{-6}), making them less sensitive to cut-off effects, often mitigated with switching functions to ensure a smoother energy curve near the cut-off radius. In contrast, electrostatic interactions decay more slowly (r_{ij}^{-1}), remaining significant at larger distances and thus more affected by cut-off errors. Particle Mesh Ewald (PME) is a method used in MD simulations to compute long-range electrostatic interactions accurately. Unlike simple cut-off techniques, PME combines direct space and reciprocal space calculations using Fourier Transforms to ensure accuracy. Short-range interactions within a specified cutoff radius are computed directly in real space, while PME maps atomic charges onto a grid and calculates the electrostatic potential via convolution in reciprocal space. Key parameters include the grid size for Fourier Transforms and the Ewald parameter (α), which control computational efficiency and accuracy.

Periodic boundary conditions (PBC) allow the simulation box to represent an infinite

system. This approach ensures that particles interact as if they were in an unbounded space, mitigating edge effects inherent in finite boxes as shown in Figure 2.12. PBC enable efficient exploration of large-scale systems while conserving computational resources, allow obtaining statistically meaningful results without the limitations of finite boundaries. PME method explained above is particularly essential for accurately handling electrostatic interactions across periodic boundaries, ensuring interactions between periodic images of the system are appropriately accounted for, meaning that a particle does not see its own replica. (see Figure 2.12) [150, 151, 152]

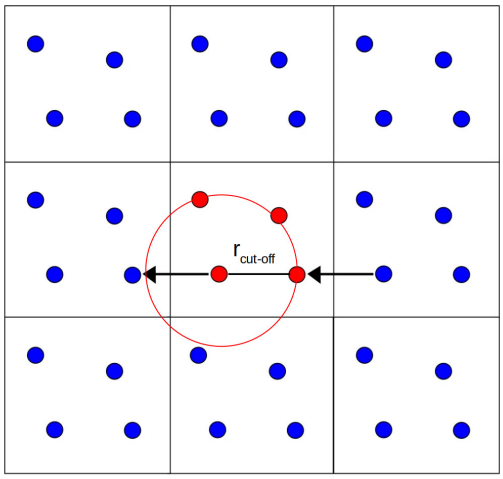


Figure 2.12: Periodic boundary conditions in two dimensions, illustrating the replication of the system in all directions. How $r_{cut-off}$ is chosen by respecting PBC conditions is shown.

In summary, the force field serves as the foundation of MD simulations, providing the framework to compute the potential energy landscape that governs atomic interactions. The details for the calculation of each bonded and non-bonded potential, with the specific parameters, will be given in section 2.2, which will explain the MARTINI force field used in this thesis.

Equations of Motion

To calculate the macroscopic properties of atoms or molecules in a system over time, the initial positions of all atoms or molecules must be known. This information can be obtained from experimental studies such as X-ray crystallography, NMR, or cryo-EM. Systems for MD simulations are created based on these known structures. Initial velocities, on the other hand, are assigned using the Maxwell-Boltzmann distribution for a given temperature.

$$P(\mathbf{v_i}) = \sqrt{\frac{m_i}{2\pi k_B T}} e^{-\frac{m_i \mathbf{v_i}^2}{2k_B T}} \tag{2.13}$$

Here, $P(\mathbf{v_i})$ represents the probability that atom i has a velocity $\mathbf{v_i}$, m_i is the mass of atom i, k_B is the Boltzmann constant, and T is the temperature.

Before starting MD simulations, it is important to ensure that the system is in a stable conformation. To achieve this, the energy of the system is minimized to reach the most stable configuration. This minimization can be performed using various methods. In this thesis, the *Steepest Descent* algorithm, also known as gradient descent, is used.

The Steepest Descent algorithm iteratively updates the positions of atoms to minimize the system's energy. It follows the negative gradient of the potential energy surface with respect to the atomic coordinates. Mathematically, the update rule for the position of atom i at iteration k is expressed as:

$$\mathbf{r}_i^{(k+1)} = \mathbf{r}_i^{(k)} - \alpha_{SD} \boldsymbol{\nabla} U(\mathbf{r}_i^{(k)}) \tag{2.14}$$

In this equation, $\mathbf{r}_i^{(k)}$ represents the position of atom i at iteration k, α_{SD} is the step size, and $\boldsymbol{\nabla} U(\mathbf{r}_i^{(k)})$ is the gradient of the potential energy with respect to the position of atom i at iteration k. This gradient indicates the direction of steepest ascent in the energy landscape, and the negative sign ensures that the algorithm moves in the direction of decreasing energy.

The step size α_{SD} is a crucial parameter in the Steepest Descent algorithm. It controls the size of each step taken along the gradient direction. Choosing an appropriate step size is essential to ensure convergence to the minimum energy configuration without overshooting or oscillating around it. One common approach is to use a fixed step size, while more sophisticated methods, such as line search or backtracking, dynamically adjust the step size at each iteration based on the change in energy and the gradient magnitude.

After the minimization steps, which define the starting point of simulations, integrating Newton's equations of motion determines atomistic trajectories. Various molecular dynamics algorithms are used for this integration step. Integration of equations of motion determines all the velocities and the positions of atoms at any time. A key parameter for all algorithms is the definition of the time step Δt, which is related to the frequency of integration. Therefore, the computational cost highly depends on Δt. Ideally, Δt should be large enough to reduce the cost but small enough to provide a reasonable approximation of the macroscopic properties [153]. Verlet's algorithm, one of the most popular algorithms used for molecular dynamics purposes, is utilized [154] and is as follows [154, 152]:

1. Calculate the potential energy for all pairs U_{ij}, according to the chosen force field parameters.

2. Calculate the force exerted on each particle $\mathbf{F}_i = -\boldsymbol{\nabla} U(\mathbf{r}_i) = -\frac{\partial U}{\partial \mathbf{r}_i}$, where $\mathbf{r}_i = (x_i, y_i, z_i)$ is the position vector of particle i.

3. Calculate the acceleration $\mathbf{a}_i(t)$ for each particle using $\mathbf{a}_i(t) = \frac{\mathbf{F}_i}{m_i}$, where m_i is the mass of particle i.

4. Update the position $\mathbf{r_i}(t + \Delta t)$ using the previous position $\mathbf{r_i}(t)$ and the position $\mathbf{r_i}(t - \Delta t)$:
$$\mathbf{r_i}(t + \Delta t) = 2\mathbf{r_i}(t) - \mathbf{r_i}(t - \Delta t) + \mathbf{a_i}(t)\Delta t^2, \tag{2.15}$$

5. The updated positions are then used in the next iteration of the cycle.

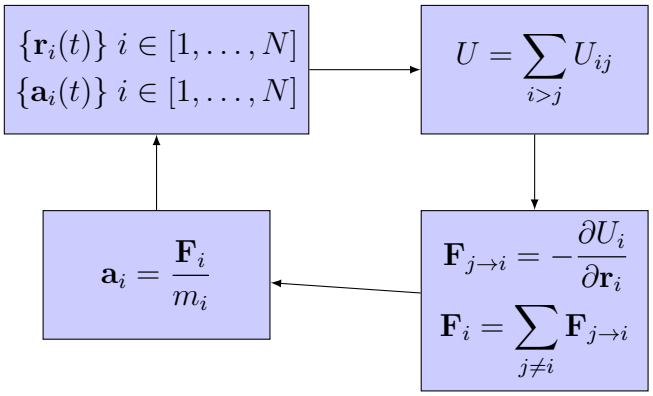


Figure 2.13: Molecular Dynamics iterative cycle, as explained in Verlet algorithm. Each step is illustrated and represented. The Potential Energy U is calculated and the Force $\mathbf{F}$ exerted on each particle is calculated as performed in step 2. The acceleration $\mathbf{a}$ is calculated to update the velocities and positions of each atom.

In the Verlet algorithm, velocities are not explicitly calculated. However, velocities can be derived using the positions from the Verlet algorithm. The velocity at time t can be estimated as follows:

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t)}{2\Delta t} \tag{2.16}$$

In addition to the Verlet algorithm, other algorithms are commonly used in Molecular Dynamics (MD) simulations. The Velocity-Verlet algorithm improves upon Verlet by explicitly updating velocities. The Leapfrog algorithm is a modified version of the Verlet algorithm, which integrates positions and velocities in a staggered manner. Choosing the appropriate algorithm for MD simulations depends on factors such as the required accuracy, computational efficiency, and specific simulation goals.

2.2.2 Ensembles in SMD simulations

In MD simulations, an "ensemble" refers to a specific set of conditions that dictate how fundamental thermodynamic quantities are treated during the simulation, influencing the statistical behavior and properties of the simulated system. The default ensemble in MD simulations is the NVE ensemble, also known as the microcanonical ensemble, where N (number of particles), V (volume), and E (total energy) remain constant throughout the simulation. This isolation ensures that the system does not exchange particles or energy with its surroundings. In an NVE ensemble, the probability of every accessible state is equally likely. Assuming Ω possibilities, every state in an NVE simulation has a probability of $1/\Omega$.

Beyond the NVE ensemble, other commonly used ensembles include the NVT ensemble (canonical ensemble), where N, V, and T (temperature) are held constant, allowing the system to exchange energy with a heat bath to maintain a specified temperature. In the NVT ensemble, the total energy of the system and the heat bath is kept constant, and the probability of every accessible state is given by the Boltzmann distribution as follows:

$$P_i = \frac{e^{-E_i/k_B T}}{Z} \tag{2.17}$$

where P_i is the probability of the system being in state i with energy E_i, k_B is the Boltzmann constant, T is the temperature, and Z is the partition function, defined as:

$$Z = \sum_i e^{-E_i/k_B T} \tag{2.18}$$

Another important ensemble is the NPT ensemble (isothermal-isobaric ensemble), which maintains the number of particles (N), pressure (P), and temperature (T) constant. In this ensemble, pressure is controlled through algorithms such as the Berendsen or Parrinello-Rahman barostats, which adjust the simulation box size to maintain the desired pressure. This ensemble is particularly useful during equilibrations to obtain the optimal simulation box size needed for NVT simulations during the MD production runs. Additionally, the NPT ensemble is suitable for studying systems under mechanical stress or in contact with a surrounding medium, where both volume and energy can fluctuate.

Each ensemble offers a unique perspective on system behavior under different thermodynamic conditions, providing insights into equilibrium states, phase transitions, and dynamic processes. The choice of ensemble in MD simulations depends on the specific physical properties and phenomena being investigated, ensuring that computational studies accurately capture the relevant aspects of experimental systems.

In this thesis, we used the NVT ensemble with Steered Molecular Dynamics (SMD), which will be explained in the following section. The choice of this ensemble is based on the fact that translocation simulations with SMD are not feasible using the NPT ensemble. In the NVT ensemble, the number of particles (N), volume (V), and temperature (T) are kept constant:

To maintain a constant temperature T, the v-rescale thermostat is used. This thermostat scales the velocities of particles based on their kinetic energy to achieve the desired temperature T_{des}.

The temperature T of the system is related to the average kinetic energy $\langle K \rangle$ by:

$$\langle K \rangle = \frac{3}{2} N k_B T \tag{2.19}$$

where N is the number of particles, k_B is the Boltzmann constant, and T is the temperature of the system.

To conserve the temperature of the system the average kinetic energy $\langle K \rangle$ is fixed, while the instantaneous K is allowed to fluctuate. The instantaneous kinetic energy K of the system is expressed as:

$$K = \frac{1}{2} \sum_i m_i \mathbf{v}_i^2 \tag{2.20}$$

where m_i is the mass of particle i and $\mathbf{v}_i$ is its velocity.

The v-rescale thermostat adjusts the velocities $\mathbf{v}_i$ of particles i by a scaling factor λ:

$$\mathbf{v}_i^{\text{new}} = \mathbf{v}_i^{\text{old}} \cdot \lambda \tag{2.21}$$

where λ is calculated as:

$$\lambda = \sqrt{\frac{T_{\text{des}}}{T(t)}} \tag{2.22}$$

Here, T_{des} is the desired temperature, $T(t)$ is the current temperature of the system at time t, and k_B is the Boltzmann constant. The thermostat ensures that the average kinetic energy $\langle K \rangle$ corresponds to $\frac{3}{2}k_B T_{\text{des}}$, maintaining the system at the desired temperature throughout the simulation.

2.2.3 Steered Molecular Dynamics (SMD)

Steered Molecular Dynamics (SMD) is an advanced non-equilibrum MD technique that allows for the manipulation of a molecular system along a specified reaction coordinate or pathway. Unlike conventional MD simulations, SMD actively applies an external force to drive the system towards a desired conformational change or transition state. This approach finds widespread applications in studying biomolecular processes such as protein folding, ligand binding, conformational changes, and molecular transport.

The external force, also called a steering force, is typically applied to specific atoms or groups within the molecular system. By monitoring the system's response to the applied force, insights into the underlying energy landscape and kinetics of the process can be obtained.

These simulations include Constant Velocity SMD (cv-SMD), which applies a constant pulling velocity to investigate mechanical unfolding or transitions in biomolecules. Constant Force SMD (cf-SMD) maintains a fixed force to explore force-induced conformational changes and mechanical properties. Targeted Molecular Dynamics (TMD) [155] guides the system from an initial state to a desired target configuration using biasing potentials or forces. Adaptive Biasing Force (ABF) SMD [156] employs adaptive bias potentials to enhance sampling efficiency and explore free energy landscapes. Jarzynski's Equality SMD [157] estimates free energy differences using non-equilibrium work measurements.

In this thesis, cf-SMD simulations are employed to investigate ssDNA translocation through through αHL nanopore, providing insights into the molecular mechanisms and energy barriers involved in this process. Therefore we will detail this type of the SMD simulations.

Constant Force SMD

In cf-SMD, an external force with a defined magnitude is applied to specific atoms or groups of atoms along a given direction and geometry. This force is maintained as long as it is desired throughout the simulation to explore mechanical properties and transition pathways of molecular systems under non-equilibrium conditions. cf-SMD modifies the equations of motion to incorporate the applied external force, enabling detailed investigation into the dynamics of molecular processes.

2.2.4 All-atom and Coarse-Grained Molecular Force Fields

This section provides an exploration of the interactions and parameters pertinent to the force fields, especially for the MARTINI force field, in light of the foundation for molecular dynamics simulations discussed previously. Initially, force field interactions necessary for the calculation of the potential energy U in the MARTINI force field, including their specific parameters, will be comprehensively elucidated.

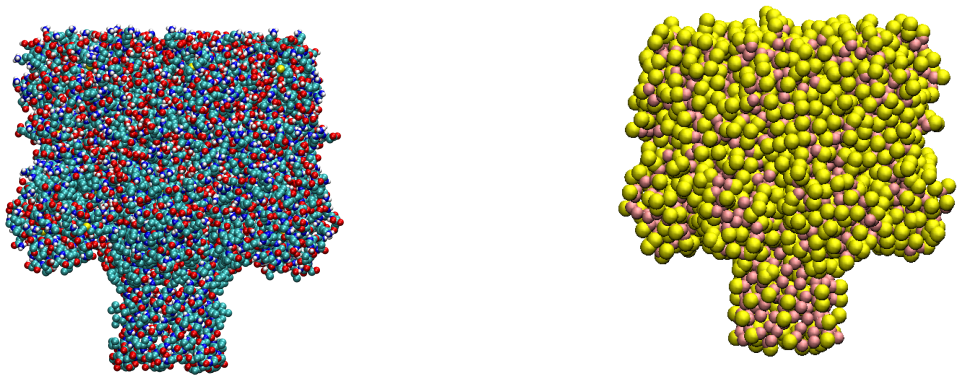
Subsequently, each biological molecule featured in this thesis will be introduced from a biological perspective. Following this biological overview, the molecules will be described within the context of the MARTINI force field, emphasizing their representation and characterization in this CG simulation framework.

All-atom Force Fields

All-atom force fields represent each atom in a system explicitly, providing detailed descriptions of molecular structures and interactions. Among the most commonly used all-atom force fields are AMBER (Assisted Model Building with Energy Refinement), which is highly regarded for simulating biomolecular systems like proteins and nucleic acids; CHARMM (Chemistry at HARvard Macromolecular Mechanics), widely used for proteins, lipids, and nucleic acids; GROMOS (GROningen MOlecular Simulation), optimized for biomolecular systems in aqueous solutions; OPLS-AA (Optimized Potentials for Liquid Simulations - All Atom), known for its versatility in proteins, organic molecules, and polymers; COMPASS (Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies), suitable for polymers and materials; and MMFF (Merck Molecular Force Field), designed for small organic molecules and pharmaceuticals. Choosing the right force field depends on the specific needs of the simulation, the type of molecules involved, and the level of accuracy required. All-atom force fields include parameters for bond lengths, bond angles, dihedral angles, and non-bonded interactions such as van der Waals and electrostatic forces, as discussed previously. The specific parameters and interactions may differ from force field to force field, impacting their suitability for different types of simulations.

Coarse-Grained Force Fields

CG force fields, dating back to the 1970s, aim to simulate complex and large systems to understand their global behavior using a simplified representation, which reduces the computational cost. The reduction in computational cost is achieved through several factors: a reduced number of degrees of freedom in the system and a simplified system with fewer particles (See Figure 2.14), which results in a smoother energy landscape [138]. CG models are well-suited for understanding the overall behavior of certain biological events, especially those occurring over timescales of several microseconds [158].



(a) All-atom structure of aHL, VDW representation in VMD

(b) Coarse-grained model of aHL, VDW representation in VMD

Figure 2.14: (a) All-atom and MARTINI (b) coarse-grained representations of α-hemolysin, pdb code: 7ahL [159] VDW representation in VMD is used in VMD [160] to show the decreased number of beads present in the CG representation

While CG models offer significant advantages, such as reduced computational cost and the ability to simulate larger systems for longer timescales, they also have limitations. The primary drawback is the loss of atomic detail, which can result in inaccuracies for certain physical properties. Force-field parameters for CG models are typically derived from all-atom models through approximations, which means some specific interactions and detailed structural features might not be accurately represented.

Despite these limitations, several successful CG models provide a sufficiently accurate representation of inter-residue properties, including the SIRAH, SDK, and PRIMO force fields. One of the most prominent CG models is the MARTINI force field, developed in the early 2000s, initially to model lipid bilayer assemblies [161]. The MARTINI force field has since been extended to include proteins [162], carbohydrates [163, 164], and nucleic acids [165], making it a widely used tool for CG-MD studies. In this thesis, the focus is on modeling large protein channels embedded within lipid bilayers. Modeling such extensive systems typically involves including the lipid bilayer, the embedded protein channel, and the surrounding environment. Conducting all-atom molecular dynamics simulations for these molecules over long time scales is challenging and often requires simplifications using implicit solvent or membrane models. The MARTINI force field offers a solution for modeling these proteins by reducing the number of interaction sites while maintaining an explicit representation of both the solvent and the membrane. Although the MARTINI 3 force field has been released recently, it is not used in this thesis because the polarizable water and DNA models are not yet incorporated, making it unsuitable for our specific application.

The MARTINI Coarse-Grained Force Field The MARTINI CG force-field was initially introduced in 2004 for modeling lipids [166]. In 2007, it was expanded to encompass other biomolecules and was named MARTINI [161]. This force field is tailored to study molecular interactions with polar and nonpolar solvents, facilitating comparisons with experimental data and all-atom simulations. Its parameters are fine-tuned to ensure that the simulations closely match experimental findings.

The MARTINI 2 force field, which we use in this thesis, simplifies the representation of molecules by averaging approximately four heavy atoms and their associated hydrogen atoms into a single interaction center, called a particle or bead. This design makes the model user-friendly, with a limited number of bead types. For the sake of simplicity, each time we use the term MARTINI will refer to the MARTINI 2 force field.

There are four main types of particles, categorized based on the polarity of the atom groups they represent: P for polar, N for nonpolar, C for apolar, and Q for charged particles. Hydrophobic groups are modeled with C-type particles, while N-type particles represent groups that are both partially polar and partially hydrophobic.

Each particle type within the MARTINI force field is further classified into subtypes that provide additional information about the chemical characteristics of the represented group. For instance, particles like N and Q have four sub types denoting their hydrogen-bonding capabilities: **0** indicates groups that cannot form hydrogen bonds, **d** signifies hydrogen bond donors, **a** denotes hydrogen bond acceptors, and **da** indicates groups that can both donate and accept hydrogen bonds. On the other hand, particles P and C have five subtypes labeled from 1 to 5, offering a more detailed description of their polarity, where 1 represents weak polarity and 5 indicates high polarity.

In terms of mass, all particles except those representing cyclic groups have a mass of 72 g/mol. Cyclic particles, which typically model 2 or 3 heavy atoms along with their associated hydrogen atoms to maintain the geometric properties of the all-atom cycle accurately, have a reduced mass of 45 g/mol.

The MARTINI CG force field relies on the same physical principles of force fields discussed earlier. The system's potential energy is divided based on the nature of interactions (bonded or non-bonded) between particles.

Non-bonded interactions in the MARTINI force field are described using a Van der Waals or LJ 12-6 potential, represented by equation 2.12:

$$U_{\text{vdW}} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.23}$$

Here, ε_{ij} denotes the interaction strength between particles i and j, and r_{ij} stands for the vector distance between these particles i and j, as explained in equation 2.12. Table 2.1 provides a comprehensive overview of these levels of LJ interactions, categorized with Roman numbers. The minimum distance between particles, defining their effective size, is $\sigma = 0.47\,\text{nm}$ for most particles, except for cyclic particles where $\sigma = 0.43\,\text{nm}$. Additionally, interactions between cyclic particles involve an ε reduced to 75% of its standard value.

Charged particles (Q-type) carry a full charge and interact via a Coulomb potential as described earlier by eq 2.12:

$$U_{\text{Elec}} = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_i q_j}{r_{ij}} \tag{2.24}$$

Table 2.1: Different Lennard-Jones interaction levels between particles in the MARTINI force field [161].

Interaction Level	ε **(kJ/mol)**
0 (supra attractive)	5.6
I (attractive)	5.0
II (almost attractive)	4.5
III (semi attractive)	4.0
IV (intermediate)	3.5
V (almost intermediate)	3.1
VI (semi repulsive)	2.7
VII (almost repulsive)	2.3
VIII (repulsive)	2.0

Here, ε_0 is the permittivity of free space, and $\varepsilon_r = 15$ is the relative dielectric constant for explicit screening, which balances the increased hydration strength of many coarse-grained (CG) particles. The dielectric constant ε_r depends on the solvent used, either the classic MARTINI water model (W) or the polarizable water model (PW). The force field empirically sets this parameter to reproduce water's dielectric properties, solvation effects, and electrostatic interactions, ensuring accurate simulation of these interactions in aqueous environments.

Bonded interactions in the MARTINI force field utilize harmonic potentials:

$$U_{\text{Bonds}} = \frac{1}{2} k_d (d - d_0)^2 \tag{2.25}$$

where $d_0 = 0.47\,\text{nm}$ and $k_d = 1250\,\text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$. These parameters, such as the equilibrium bond distance d_0 and force constant k_d, are adjusted based on the specific atom groups represented by the beads. For instance, cyclic particles may have different d_0 values to maintain the original cycle geometry.

To model bending-angle potential, MARTINI employs a cosine harmonic potential for angles:

$$U_{\text{Angles}} = \frac{1}{2} k_\theta (\cos\theta - \cos\theta_0)^2 \tag{2.26}$$

where $\theta_0 = 180^\circ$ and $k_\theta = 25\,\text{kJ} \cdot \text{mol}^{-1}$. Similar to bond potentials, k_θ and θ_0 can vary depending on the specific chemical structures represented by the beads. For example, lipid molecules containing *cis* or *trans* double bonds have distinct k_θ and θ_0 values.

Additionally, MARTINI treats cyclic particles with an improper dihedral angle potential in complex geometries:

$$U_{\text{IDihedrals}} = k_{\text{IDihedrals}} (\psi - \psi_0)^2 \tag{2.27}$$

where ψ represents the angle between planes formed by atoms i, j, k and $j, k, l,$, ψ_0 is the equilibrium angle, and $k_{\text{IDihedrals}}$ is the force constant, depending on the chemical group represented.

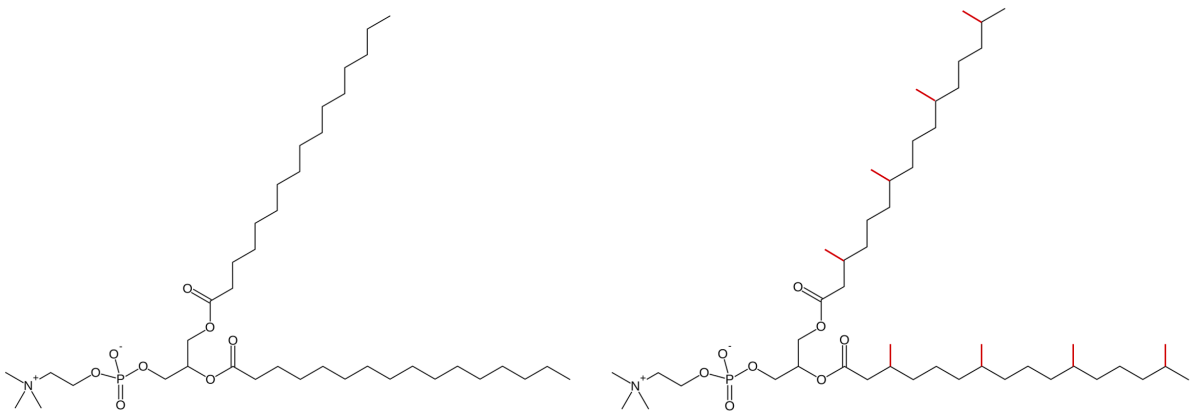
The MARTINI CG force field significantly reduces computational complexity by reducing the number of interaction sites in the system. This reduction allows integration of

equations of motion with larger time steps, up to 20 fs, thereby accelerating computation times.

With the MARTINI force field, it is possible to model lipids [161], simple sugars [163, 164], polymers [167], peptides and proteins [162], as well as nucleic acids [165]. Different solvent models are available, such as the MARTINI water models (W), including the polarizable water model PW [168]. This thesis utilizes the MARTINI force field for modeling proteins, lipids, and DNA, utilizing the polarizable water model PW. In the following sections, we will detail each molecule used, first within the biological and then from the MARTINI perspective.

MARTINI Lipids - DPPC Lipids are a diverse group of hydrophobic molecules that form the structural foundation of biological membranes. They play an important role in maintaining the integrity and functionality of cells by creating a barrier and facilitating various cellular processes. Among the different types of lipids, glycerophospholipids are a major class, characterized by a glycerol backbone linked to fatty acid chains and a phosphate group. In this thesis, two glycerophospholipids containing a choline group, namely diphytanoylphosphatidylcholines (DPhPC), were utilized for experimental purposes to form lipid bilayers into which αHL was inserted.

However, parameters specific to DPhPC are not available in the MARTINI force field. Instead, parameters for dipalmitoylphosphatidylcholine (DPPC) are available and structurally similar to DPhPC, as depicted in Figure 2.15. Both DPPC and DPhPC have saturated carbon chains of identical length. The sole difference lies in the presence of methyl groups on the two chains of DPhPC, highlighted in red in Figure 2.15b.



(a) (b)

Figure 2.15: Chemical structure of (a) DPPC and (b) DPhPC lipids, methyl groups on the two chains of DPhPC are highlighted in red

As explained earlier, the MARTINI force field combines on average four heavy atoms and associated hydrogen atoms into a single CG particle, called a bead. The addition of a methyl group does not require an additional coarse-grained bead on the fatty acids. Therefore, the differences between DPhPC and DPPC molecules are considered negligible within the MARTINI force field for the purposes of this thesis, where DPPC is used in MD simulations to represent experimentally used DPhPC .

DPPC molecules are extensively used for modeling lipid bilayers and were among the first lipids introduced and parameterized in the MARTINI force field [161]. In the MARTINI force field, DPPC lipids are represented by twelve interaction sites of four different types (Figure 2.16) [161]. The polar head consists of a Q0 site (positively charged) and a Qa site (negatively charged), representing the choline part and phosphate group, respectively. The glycerol backbone is represented by two Na sites of intermediate polarity. Four highly hydrophobic C1 sites form each of the two lipid tails. Additionally, the bond length, σ, between the two Na particles of the glycerol group is reduced to 0.37 nm.

The transition parameters for DPPC are essential for justifying the simulation temperature. With the MARTINI force field, at temperatures above 310 K, DPPC lipids correspond to the liquid phase. Below 310 K, down to 290 K, they indicate the gel phase. These values are crucial for selecting the simulation temperature to accurately represent the lipid behavior [169].

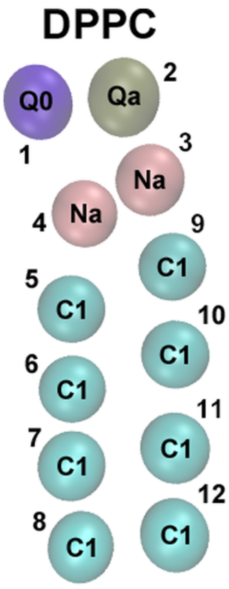


Figure 2.16: MARTINI representation of a DPPC molecule with twelve interaction sites of four types [161].

MARTINI Proteins Proteins are essential biomolecules with a wide range of functions. They can be involved in cellular recognition, maintain tissue structure, or facilitate the transport of chemical compounds.

Proteins are biological polymers made up of 20 different units called amino acids, which are linked by amide bonds, also known as peptide bonds. These bonds are typically rigid and planar. When incorporated into a protein, amino acids are referred to as residues. Each residue consists of an amine group, a carboxyl group, and a variable side chain that determines its properties. The sequence of amino acids forms the primary structure, which can fold into specific shapes known as secondary structures, including alpha helices, beta sheets, and beta turns. Portions of the protein chain that do not form structured regions are called random coils. Additionally, proteins exhibit tertiary structures, which defines the function of a protein, which are the overall folding of a single protein molecule, and quaternary structures, which are the arrangement of multiple protein sub units within

a larger complex. Modeling proteins is crucial for understanding their functions and eventually enables their modification when necessary for various applications in different domains.

In 2008, Monticelli et al. [162] developed parameters for amino acids within the MARTINI force field. Each amino acid is represented by 1 to 5 beads instead of approximately twenty atoms in all-atom force fields, with one bead representing the protein backbone (BB type), i.e., the parts of amino acids involved in peptide bonds. Figure 2.17 shows the mapping of the 20 amino acids in the MARTINI coarse-grained model, with the degree of polarity of the different beads.

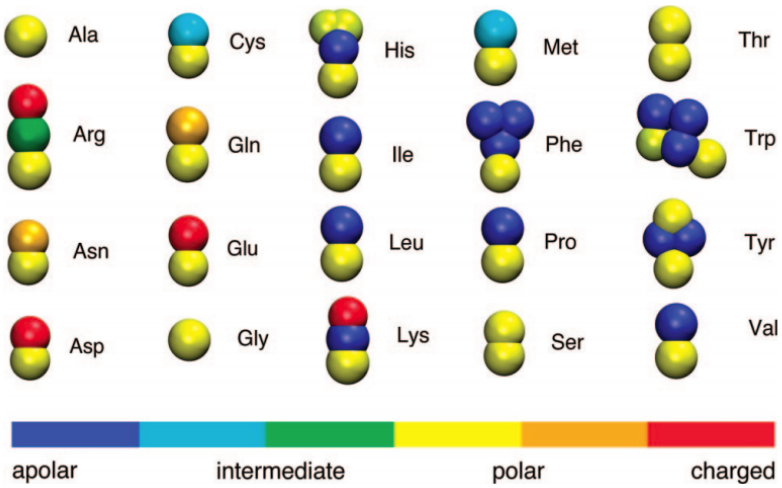


Figure 2.17: Representation of martini amino acids [162].

Glycine (Gly) and alanine (Ala) have minimal side chains, consisting of a single hydrogen atom and a methyl group, respectively. In the MARTINI model, both are represented by a single backbone bead labeled as BB, reflecting their status as the smallest amino acid residues.

Proline (Pro), distinguished by its unique ring structure involving the backbone amine, is represented by two MARTINI beads. One of these beads is a BB bead with specific properties tailored for Proline's distinct backbone structure compared to other amino acids.

Aromatic amino acids such as histidine (His), tyrosine (Tyr), phenylalanine (Phe), and tryptophan (Trp) contain ring structures in their side chains. These are modeled using special beads prefixed by S (small), which have a restricted LJ radius of $\sigma = 0.43\,\text{nm}$ for interactions with other S beads. Additionally, their LJ interactions are reduced to 75% of the standard value to accurately simulate their behavior in protein structures.

Four amino acids are charged: aspartic acid (Asp) and glutamic acid (Glu) carry negative charges, while lysine (Lys) and arginine (Arg) carry positive charges. Each of these amino acids has two forms for their side chains: a charged form represented by a Q bead with a charge of $\pm 1e$, and a neutral form corresponding to a protonated state for Asp and Glu, and a deprotonated state for Lys and Arg.

Maintaining the correct geometry of peptide bonds, crucial for protein folding and secondary structures, is ensured through a dihedral angle potential:

$$U_{Dihedrals} = k_d[1 + \cos(n\psi - \psi_0)] \tag{2.28}$$

This potential plays an important role in accurately modeling protein folding dynamics and structure. The force constant k_d and equilibrium angle ψ_0 depend on the secondary structures involving the particles. For helices, k_d is set at 400 kJ/mol with $\psi_0 = 60°$, while for beta sheets, k_d is reduced to 10 kJ/mol and $\psi_0 = 180°$.

An improper dihedral angle potential, as described by equation 2.27, maintains the planar configuration of cyclic amino acids like His, Phe, Tyr, and Trp. Unlike all-atom force fields, $U_{Dihedrals}$ in MARTINI is specific to certain structures and excludes lipids.

The polarity of peptide bonds varies with their secondary structure. In random coils, the backbone is polar and modeled with a P-type particle, while in alpha helices and beta sheets, it is less polar and represented by N-type particles. The backbone particle types for glycine, alanine, and proline differ from others.

The bond length between two BB beads is 0.35 nm, regardless of secondary structure. The bond length and force constant between a BB bead and a side chain bead, or between side chain beads, depend on the amino acid. Equilibrium angle parameters involving the 2 backbone beads and 1 side chain bead are consistent across amino acids, while those between 1 BB bead and 2 side chain beads are structure-dependent. The angle force constant also depends on the amino-acid stucture. Dihedral angles are imposed only when all four interacting beads have the same secondary structure [162].

The `martinize.py` tool, introduced in 2013, converts all-atom protein structures into MARTINI coarse-grains [170]. Updates included a charged form for His and adjusted bond parameters for BB beads in alpha helices, constrained to 0.31 nm, resulting in the MARTINI 2.2 force field used in this thesis.

The polarizable version, 2.2P, developed by de Jong *et al.* [170], introduced new charged side chain particles and polar particles for Ser, Thr, Gln, and Asn side chains. Detailed parameters for versions 2.2 and 2.2P are available in their supplementary materials [170].

One significant drawback of the MARTINI force field, particularly when applied to proteins, is its inability to maintain protein secondary structures. To address this issue, an elastic network model consisting of a set of harmonic potentials is added on top of the MARTINI force field to stabilize the secondary structures of the protein [171]. The elastic network model used in this thesis is called ElNeDyn (Elastic Network in Dynamics). This network applies a spring-like restraint between two BB beads, i and j, if their distance r_{ij} is less than 0.9 nm and they belong to amino acids separated by at least two residues ($j \geq i + 3$) [171]. The potential of a bond in this elastic network can be written as:

$$U_{\mathrm{EN}} = \frac{k_{\mathrm{elastic}}}{2}(d - d_0)^2 \tag{2.29}$$

In our study, the spring constant for this restraint is set to $k_{\mathrm{elastic}} = 500\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. The equilibrium length d_0 corresponds to the distance between two Cα atoms in the initial structure. Consequently, the elastic network springs act as pseudo-covalent bonds between atoms, ensuring that beads linked by this network do not interact through non-bonded interactions.

The introduction of the ElNeDyn network necessitated adjustments to certain parameters in the conventional MARTINI force field. Specifically, the positions of BB beads

representing the protein backbone correspond to the Cα atom positions, rather than the center of mass of the atom group they represent (N-Cα-C-O). This change prompted modifications to several amino acid interaction parameters. Additionally, the representation of aromatic residues (Phe, Tyr, Trp, and His) was adjusted to account for the asymmetry of the histidine and tryptophan rings. For phenylalanine and tyrosine, an internal bond was added to stabilize the ring structure. Further details on these parameter adjustments can be found in the supplementary material of the paper by Periole *et al.* [171].

In this thesis, the elastic network is applied on the entire secondary structue of the protein, accounting for inter and intra chains. This ensures maintaining the structure of the αHL throughout the simulations, without observing deformation of the channel.

MARTINI Polarizable Water model and Ions The MARTINI force field's polarizable water model introduces a refined representation of water molecules in molecular simulations compared to the simple W model [168]. In this model, each water molecule is typically represented by 3 MARTINI beads, effectively capturing the dipole moment of water. These charges are balanced to ensure overall neutrality. Polarizable interactions are incorporated between neighboring water molecules by allowing the dipole moment to adapt based on the local environment. This adaptation is achieved through the use of virtual sites, which are additional interaction points used to model the polarization effects. These virtual sites enable the beads to adjust their positions and interactions in response to the surrounding electrostatic field, thereby providing a more accurate representation of water's polarizable nature.

The constituent particles in the MARTINI polarizable water model, (See Figure 2.18), include the negatively charged particles (**WM**), positively charged particles(**WP**), and neutral particle representing the oxygen (**W**). The molar mass of these constituent particles is 24 g/mol, totaling 72 g/mol.

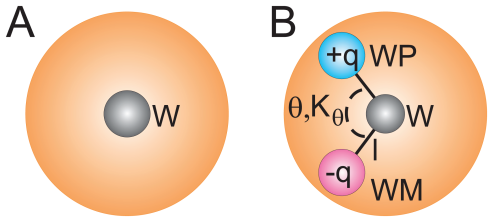


Figure 2.18: W and PW water model of MARTINI 2 force field [168].

WM and WP particles belong to the same grain and do not interact directly with each other. However, they can rotate around the central particle W, influencing the dipole moment of the PW model. When their charges align, the dipole moment μ is zero; when they are opposite, μ equals $2lq$, where l is the bond length (0.14 nm) between W and the charged particles (WP and WM), and q denotes their charge (± 0.46). The LJ radius is approximately 0.23 nm, ensuring that charged particles from different PW grains cannot overlap.

Additionally, there is an angular harmonic potential governing the rotation of WP and WM around W, with equilibrium angle parameters $\theta = 0$ and a force constant $k_\theta = 4.2$ kJ/mol.rad^{-2}.

The central particle W in the MARTINI polarizable water model is of type P4, similar to a non-polarizable water model. The depth of the LJ potential well (ε) for interactions between W particles is 4.0 kJ/mol, with a LJ radius (σ) of 0.47 nm. For interactions with particles of other types, ε is reduced to 95% of its value, except for type Q particles, which have differently adjusted interaction forces. The global dielectric constant is reduced to $\varepsilon = 2.5$. This adjustment is made to more accurately represent the interactions between charged and polar groups in a low-dielectric medium, such as a membrane interior, and the reduced screening effects in the CG model, which compensates for the fewer degrees of freedom compared to an all-atom model. This ensures that the interactions between charged particles are appropriately scaled in the CG simulation [168].

Previous research from our laboratory has also demonstrated the significance of using the polarizable water model in MARTINI with the system we use in this thesis. Specifically, it was found that with an external field, the potential differences across the membrane align with those observed in all-atom (AA) simulations [145]. This highlights the necessity of employing a polarizable water model to accurately capture electrostatic interactions.

Apart from the PW model, in the MARTINI force field, ions are represented as single-type charged beads (**Q**) with a charge of $\pm 1e$. These beads interact with other particles in the system according to Coulombic interactions. The van der Waals interactions between ions and other particles are modeled using LJ potentials, with parameters specific to each ion-particle interaction pair. The inclusion of ions allows for the simulation of electrolyte solutions and the study of ion-protein interactions in biological systems.

MARTINI Nucleic Acids - DNA Deoxyribonucleic acid (DNA) is the molecule that holds the genetic blueprint for life. Found in every cell, DNA is a biopolymer made up of four distinct nucleotides arranged in a sequence that varies from organism to organism. Each nucleotide consists of a nitrogenous base attached to a sugar molecule, deoxyribose, which is further linked to a negatively charged phosphate group. These nucleotides are joined by phosphodiester bonds between the third carbon of one sugar and the fifth carbon of the next, creating a sugar-phosphate backbone with distinct 3' and 5' ends (see Figure 2.19). DNA sequences are read from the 5' to the 3' end.

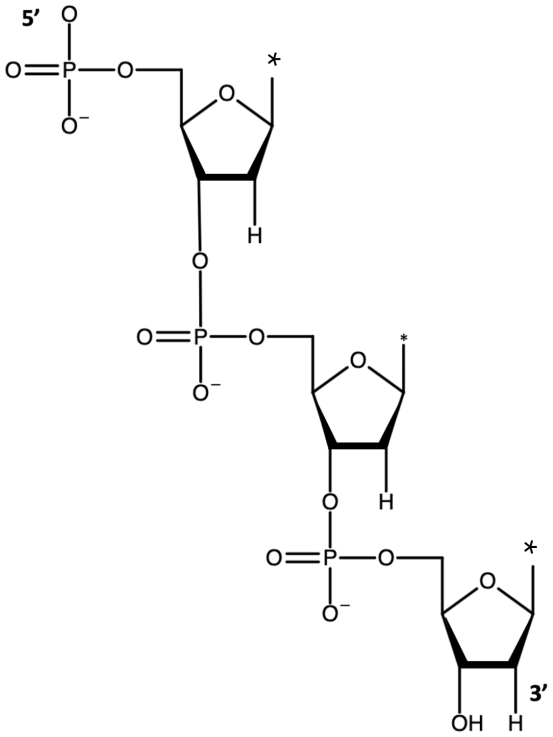


Figure 2.19: ssDNA chemical structure. DNA bases are shown as *.

The nitrogenous bases are divided into two categories: pyrimidines, which have a single ring structure (cytosine (C) and thymine (T)), and purines, which have a double ring structure (adenine (A) and guanine (G)). Adenine pairs with thymine via two hydrogen bonds, while cytosine pairs with guanine via three hydrogen bonds, as shown in Figure 2.20. Within the cell, DNA exists as a double helix, where two complementary anti parallel strands are held together by hydrogen bonds between paired bases. In a single-strand form, DNA tends to fold back on itself, forming structures like hairpins when complementary sequences are nearby. The number of GC pairings in these structures affects the stability and strength of the DNA duplex.

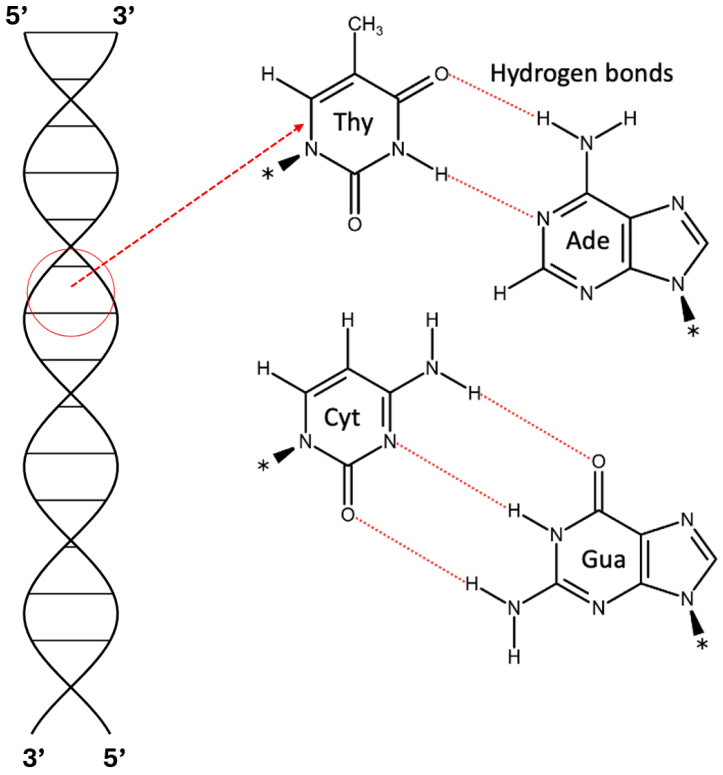


Figure 2.20: DNA base pairing. Deoxyribose sugars are shown as *.

In 2015, Uusitalo *et al.*[165] developed a MARTINI force field specifically for modeling DNA. This CG model represents each nucleotide with 6 or 7 beads instead of the thirty or so atoms found in all-atom models. The DNA backbone (BB) is modeled with three beads: a phosphate group (type Q) and two sugar rings (types SN0 and SC2). The nitrogenous bases are represented with three beads for pyrimidines (C and T) and four beads for purines (A and G) (see Figure 2.21). The backbone beads are designated BB1, BB2, and BB3, with BB1 representing the phosphate group, BB2 and BB3 representing sugar in 3' direction. The nitrogenous base beads are labeled SC1, SC2, SC3, and, if needed, SC4, with SC1 linked to the BB2 bead. In a double-stranded DNA helix, the SC2 and SC3 beads pair with complementary strands.

The helical structure of double-stranded DNA necessitates smaller beads than standard MARTINI particles. A new class of beads, labeled "T" for tiny, was introduced for the MARTINI DNA model to model hydrogen bonding interactions between nitrogenous bases. These beads have a LJ radius of $\sigma = 0.32$ nm when interacting among themselves. When interacting with cyclic particles (S), the LJ radius σ is increased to 0.43 nm, and the interaction energy ε is reduced to 75% of its standard value. The tiny particles interact with other particles using standard MARTINI parameters, for which $\sigma = 0.47$ nm and ε varies depending on the particle type.

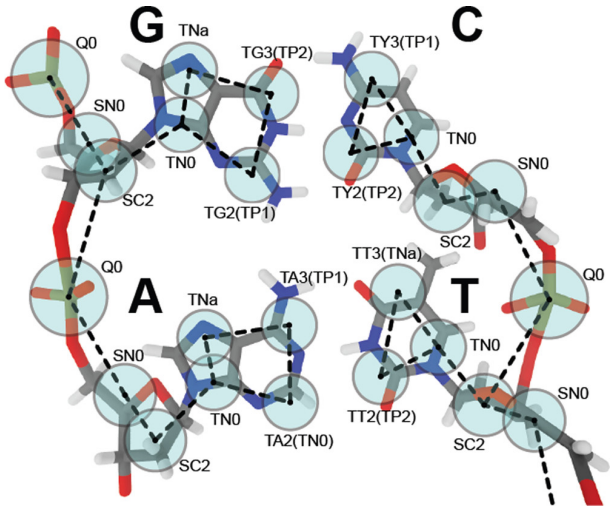


Figure 2.21: Representation of a DNA nucleotide in the MARTINI model [165].

Despite these advancements, maintaining the structural integrity of DNA during simulations remains a challenge. To address this, an elastic network model, inspired by the ElNeDyn model used for proteins, has been developed for DNA simulations. This model applies spring-like forces between specific beads in the DNA structure, preventing excessive deformation and maintaining the overall structural stability. The `martinize-dna.py` script was developed to convert all-atom DNA structures to CG MARTINI models and add the elastic network with chosen stiffness parameters.

In the double-stranded DNA structure, a rigid network connects all types of beads, including backbone and nitrogenous bases, within a 1.0 nm distance. This network applies a strong restoring force, preserving the molecule's overall structure with a force constant of $k_{\text{elastic}} = 500\,\text{kJ}\,\text{mol}^{-1}\,\text{nm}^{-2}$.

Alternatively, a flexible network selectively links backbone beads (BB) and nitrogenous base beads (SC1) up to 1.2 nm apart. This arrangement allows greater freedom for base movements while maintaining structural integrity. The force constant for this flexible network is lower at $k_{\text{elastic}} = 13\,\text{kJ}\,\text{mol}^{-1}\,\text{nm}^{-2}$, striking a balance between stability and flexibility to accommodate dynamic conformational changes without compromising DNA's overall structure.

For ssDNA structures that are the subject of this thesis, the option exists to utilize either a rigid or flexible elastic network. The rigid network, similar to its double-stranded counterpart, acts as a robust scaffold, preventing deformation of the DNA strand. Conversely, simulations with a flexible elastic network offer greater flexibility but may sacrifice some degree of structural stability. In this thesis, we used the flexible elastic network for the ssDNA molecule, since the rigid elastic network model is adapted for dsDNA molecules.

These parameters conclude the description of the methods employed in this thesis.

Chapter 3

Experimental study of DNA translocation and unzipping using aHL

Contents

3.1 Experimental model and parameters 74

3.2 Effect of orientation on translocation time 76

3.3 Effect of the duplex structure on unzipping time 78

3.4 Effect of the duplex length on unzipping time 85

3.5 Conclusion 94

This chapter will detail our experimental results on DNA translocation and unzipping. The experimental method, protocol, and the design of the DNA molecules are explained in Chapter 2, Methods. As mentioned in previous chapters, we are interested in combining experimental and computational studies to investigate the translocation of DNA molecules and gain insights into the dynamic aspects of the process. Despite the main experimental work focusing on the dynamics of unzipping, to support our findings as a result of our CG-SMD simulations, which will be explained in the next chapter, we also studied the translocation of a single-stranded DNA (ssDNA) poly$(dA)_{50}$ molecule. Our unzipping studies aim to build on the previous experimental work that demonstrated the efficacy of DNA unzipping in the αHL vestibule compared to the trans site [59]. Leveraging these findings, we studied the DNA unzipping from the cis side, focusing on the effects of applied voltage, duplex length, and structure (flap or blunt-end) within the DNA molecule.

We will begin by quickly summarizing the DNA molecules studied for both translocation and unzipping, along with the experimental parameters such as applied voltage, recording settings, and analysis methods. Next, we will present our findings on the effect of orientation on the translocation of a ssDNA molecule, which complements the computational results discussed in Chapter 4, Section 4.2. Then we will present our results on the DNA unzipping as a function of various parameters mentioned above. Finally, we will discuss our results on DNA translocation and unzipping.

3.1 Experimental model and parameters

The ssDNA molecule used for the translocation experiments is a single-stranded poly(dA)$_{50}$ without any modifications and labels. This means that the ssDNA molecule can enter the αHL nanopore from both extremities, 3' and 5', for its translocation.

The process of ssDNA transport through a nanopore involves several key physical principles. When an electric field is applied across the nanopore, the negatively charged DNA is driven through the pore by electric forces. The force acting on the DNA molecule is proportional to the applied voltage; thus, higher voltages result in faster translocation rates. As the ssDNA translocates through the nanopore, the molecule's passage temporarily blocks the ionic current flowing through the nanopore, resulting in a characteristic drop in current (illustrated in Figure 1.1).

The magnitude and duration of these current blockades can provide information about the properties of the translocating molecule, such as its length, sequence, and orientation. Specifically, longer ssDNA molecules produce longer current blockades [13, 62], and the sequence composition can affect the translocation times, such as purines translocating approximately three times slower than pyrimidines due to differences in the molecular interactions of nucleic acids within the pore [13, 61, 14]. Additionally, the orientation of the ssDNA (whether it enters the pore 3' end first or 5' end first) can influence the translocation dynamics, with 5' end ssDNA translocating slower than its 3' end counterpart [63, 59].

For the unzipping experiments, the DNA molecules used are detailed in Chapter 2, Section 2.1.5, with illustrations in Figures 2.6 and 2.7. From a physical point of view, the unzipping process of the DNA molecule can be seen as a two-state process, a transition between closed and open states separated by an energy barrier E_b, as illustrated in Figure 3.1, which can be explained by Kramer's theory of diffusive energy barrier crossing [172]. Kramer's theory describes the rate at which a system transitions over an energy barrier due to thermal fluctuations. An Arrhenius-type expression gives the rate of such thermally activated processes and is proportional to $\exp\left(-\frac{E_b}{k_BT}\right)$. Since the average unzipping time t_u is inversely proportional to this rate, we have:

$$t_u \propto \exp\left(\frac{E_b}{k_BT}\right) \tag{3.1}$$

The term $\frac{E_b}{k_BT}$ is a dimensionless quantity for the ratio of the energy barrier E_b to the thermal energy k_BT. A higher $\frac{E_b}{k_BT}$ implies that the barrier is much more significant than the thermal energy, making the unzipping process less probable and, by consequence, slower. The magnitude of this E_b is accepted to be proportional to the duplex length present in the DNA structures when the sequence is random; therefore, t_u is expected to be exponentially longer as the duplex length gets longer. Additionally, the presence of the flap should facilitate DNA unzipping. This behavior arises because the presence of a flap in the dsDNA structure reduces the initial resistance to unzipping, effectively pre-opening the duplex and making it easier for the unzipping to proceed, as illustrated in Figure 2.8. Consequently, the unzipping time t_u for dsDNA structures with a flap is expected to be shorter than for those with blunt ends.

During the unzipping or translocation process under a voltage bias, the applied voltage

V modifies the energy landscape by reducing the energy barrier E_b. (See Figure 3.1) Therefore, the unzipping time t_u depends on this voltage and is inversely proportional to it. The concept of effective charge q_e is used to understand the influence of voltage on the unzipping process. This term is homogeneous to a charge that measures how effectively the voltage is applied to facilitate the translocation or unzipping. DNA molecules have negatively charged phosphate groups, and the ionic solution contains both positively and negatively charged ions. When an electric voltage is applied, the electric field acts on the charged particles—phosphates of the DNA molecules, ions, and charged residues of the protein nanopore. This electric field exerts a force that guides the DNA molecule through the nanopore and initiates the unzipping of the duplex region. By doing work on the system, the electric force effectively lowers the energy barrier E_b, providing the additional energy required to overcome it and thereby facilitating the unzipping process. Importantly, the electric field is strongest inside the nanopore and much weaker outside. This is because the applied voltage is concentrated over a very short distance inside the nanopore, resulting in a much higher electric field compared to the surrounding solution. Thus, the effective charge q_e is accepted to be related to the charges inside the nanopore, where the electric force is most significant.

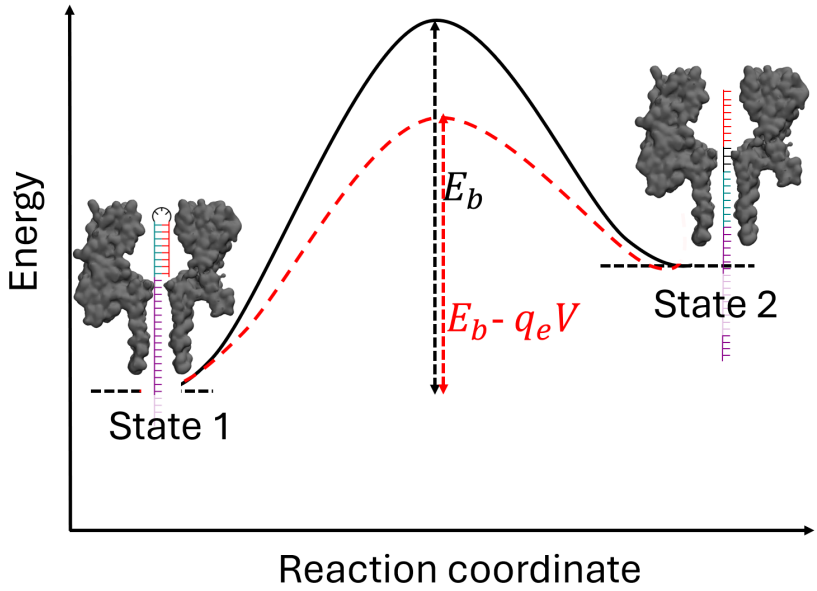


Figure 3.1: 2 state theory for the unzipping process. E_b is the energy required to pass from state 1 (closed) to state 2 (open). Applied voltage V reduces the energy barrier by $-q_eV$ where q_e stands for the effective charge.

The average unzipping time t_u is therefore proportional to a function that includes the energy barrier E_b, the applied voltage V, and the effective charge q_e:

$$t_u \propto A \exp\left(\frac{E_b - q_e V}{k_B T}\right) \tag{3.2}$$

The unzipping experiments were therefore performed first to see the effect of applied voltage on the unzipping with flap and blunt-end structures of the same duplex length.

For this step of the experiments, we worked with hairpin (HP) structures (See Figure2.6): HP9s and HP9sf. We varied the voltage from 80 mV to 180 mV, increasing by 20 mV each time under identical experimental conditions, as explained in Chapter 2.

Later, we investigated how the length of the duplex region affects the unzipping process in both flap and blunt-end DNA structures. We decided to work at a single voltage bias of 120 mV and used DNA structures presented in Chapter 2, Section 2.1.5: HP9s, HP9sf, 19s, 19sf, 29s, 29sf, and 39sf.

For both simple translocation and the unzipping experiments, the experimental setup was prepared according to the protocol explained in Chapter 2. The current is recorded using the same amplifier, Axon Multiclamp 700B (Molecular Devices, USA), low-pass filtered by a four-pole Bessel filter with a cutoff frequency of 50 kHz for simple translocation and 30 kHz for unzipping experiments (Khron-Hite, USA). It is then digitized at a sampling frequency of 1 MHz or 100kHz in case of longer events using a 16-bit acquisition card (National Instruments, Austin, TX) and saved directly to a computer's hard drive. The acquisition card is controlled via a homemade program written with LabView (National Instruments). The events were detected, and their characteristics were measured using the protocol described in Chapter 2, Section 2.1.4, based on the threshold and rejection parameters. In-house written Python code was used to illustrate the results.

3.2 Effect of orientation on translocation time

As explained in the outline of this Chapter, we will start by presenting the results of the simple translocation of ssDNA molecules. Figure 3.2 shows the translocation results of poly$(dA)_{50}$ in the form of colored event clouds, which have been explained in Chapter 2. Since in our translocation experiments, ssDNA could enter from either its 3' or 5' orientations; we observed two peaks at different normalized I_b levels with different characteristic translocation times in the event clouds plot. We computed the density of events for both orientations for the normalized blocked current, I_b/I_0, and the logarithm of time,$\log(T_t)$, and fit both as a double 2D-Gaussian curve, T_t standing for the translocation time per molecule. The fitting procedure identified two peaks with different translocation times T_t and residual currents I_b/I_0: the 3' orientation at $T_t = 160 \pm 10$ μs and $I_b/I_0 = 9.4 \pm 0.4$ %, and the 5' orientation at $T_t = 255 \pm 20$ μs and $I_b/I_0 = 13.1 \pm 0.9$ %. These assignments are made based on the observation that the 5' end-oriented DNA translocates slower than its 3' counterpart. This observation has been noted in previous experimental studies [61, 63, 59] and theoretical works [63, 134], as explained in Chapter 1.

To distinguish the events corresponding to the 3' end and the 5' end translocations on the event clouds plot, we used a Monte Carlo (MC) algorithm. This algorithm simulates the distribution of events between the two types based on their probability densities. The MC algorithm is a statistical procedure that works as follows: at each step, a random point from the plot is selected, and the probability of the point being either 3' end or 5' end-oriented is calculated based on the double Gaussian distributions: the normalized blocked current and the logarithm of time. These probabilities are then normalized, and a random selection process determines the assignment of each event to either the 3' or 5' category. The result of the Monte Carlo algorithm is displayed in Figure 3.2, where different colors are used to differentiate between the 3' end and 5' end orientations. According to these results, the translocation time ratio of 5' compared to 3' translocation was calculated as

2.22, based on more than 10,000 events recorded. This ratio is in accordance with the previous experimental results [13, 59, 63].

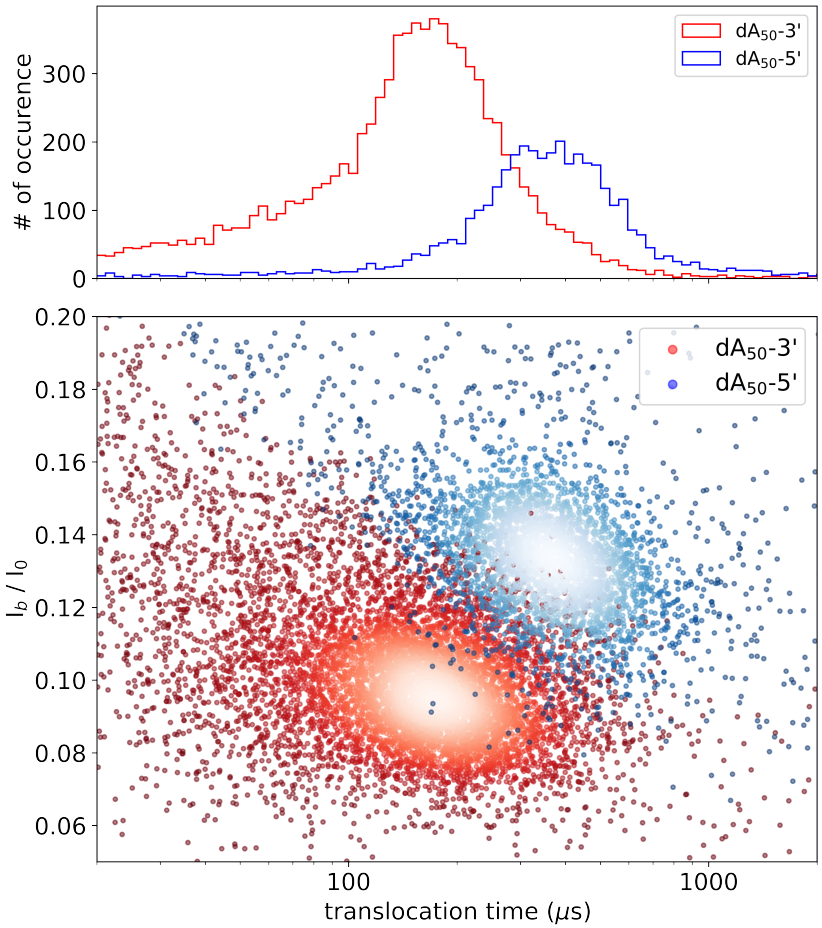


Figure 3.2: Results of the ssDNA translocation experiments. The upper figure shows the experimental translocation time distribution for poly(dA)$_{50}$ for both orientations with the αHL nanopore at 120 mV and room temperature. The scatter plot of events, event cloud, is represented in the lower figure. Each dot represents an event (about 12,500 events on this plot) with its time of translocation per molecule and the residual current (I_b/I_0, with I_0 being the open pore current and I_b the blocked pore current when the molecule is translocating through the pore). Both current and time are distinguishable between 3' oriented and 5' oriented ssDNA molecules. The 3' and 5' events were separated using an MC algorithm based on the double Gaussian distributions in the current-time plot of the cloud of events. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events.

In the next sections, we will present the results of the DNA unzipping experiments, which are the main subject of the experimental part of this thesis.

3.3 Effect of the duplex structure on unzipping time

Figure 3.3 presents scatter plots of unzipping events, colored according to the Kernel Density Estimator (KDE) explained in Chapter 2, for HP9s at various applied voltages. The data exhibit a broad distribution of unzipping times with two distinct regions: a shorter time scale (100-330 μs) and a longer time scale extending beyond 1 ms. The shorter time scale disappears with increasing voltage, while the longer time scales shift to shorter durations as voltage increases. We aim to investigate the origins of these time regions and determine which one is primarily related to the unzipping process.

To identify the characteristic unzipping times, we applied a Gaussian fit directly to the logarithmic distribution of the unzipping events, as shown in Figure 3.3. The peak positions of the Gaussian fit provided the characteristic times for both the shorter and longer time scales. The associated errors for both time scales were calculated based on the bin sizes corresponding to the time scale at each peak. These characteristic times, along with the number of events recorded at each applied voltage, are summarized in Table 3.1 At 80, 100, 120, 140, and 160 mV, this shorter time range remains between 100-330 μs. At 180 mV, insufficient data prevented us from determining this shorter-time scale correctly. Additionally, these shorter-time scales are on the same order of magnitude as the 50 nucleotide-length ssDNA translocation, leading us to conclude that these events could not be unzipping-related, as supported by the experimental results presented in Figures 3.2 and 3.4. Furthermore, the fact that these shorter-scale events dissapear as the voltage increases suggests that these events could originate from the dsDNA molecules which are not correctly inserted for their unzipping. Given these observations, the characteristic unzipping times, t_u, were extracted by focusing on the times corresponding to the longer-time scale values.

It is important to note that conducting experiments at higher voltages introduced challenges, such as increased event frequency, which results in frequent pore blockages. Despite efforts to reduce the concentration of injected DNA molecules, the number of recorded events at higher voltages, particularly at 180 mV, was lower compared to lower voltages, as shown in the top plot of Figure 3.3 and in Table 3.1.

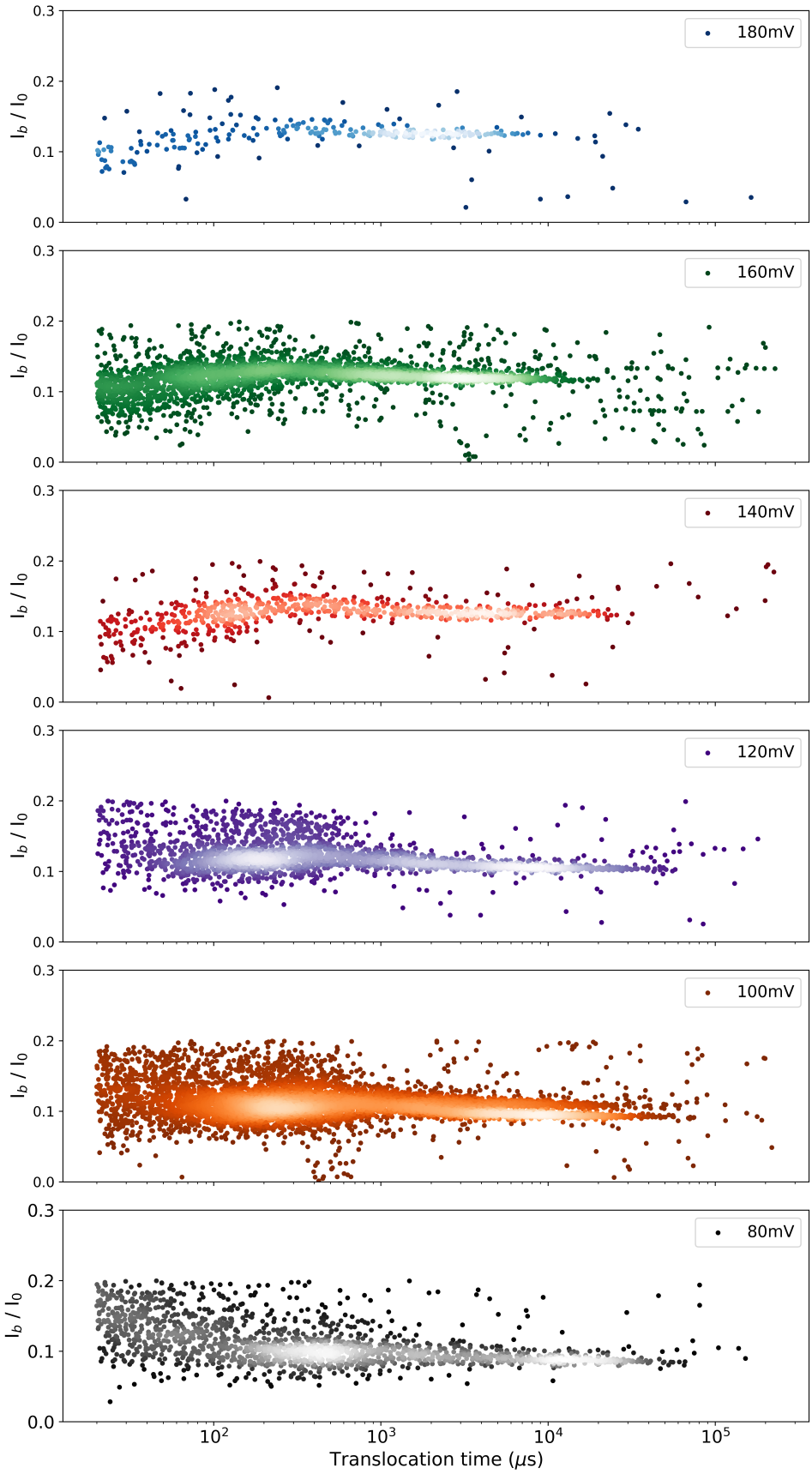


Figure 3.3: Scatter plots of HP9s at different voltages at room Temperature 22°C at pH 7.5 in 1M KCl solution. From 80 mV to 180 mV, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events.

Table 3.1: Characteristic unzipping times of HP9s at different voltages at room temperature (22°C) in 1 M KCl solution.

Voltage (mV)	Shorter-range Time (μs)	Longer-range Time (μs)	Number of events
80	330 ± 150	12040 ± 2000	2458
100	320 ± 90	7920 ± 1500	6135
120	200 ± 70	5560 ± 2500	2545
140	120 ± 50	3530 ± 1300	849
160	100 ± 60	2900 ± 1200	3946
180	-	1790 ± 600	426

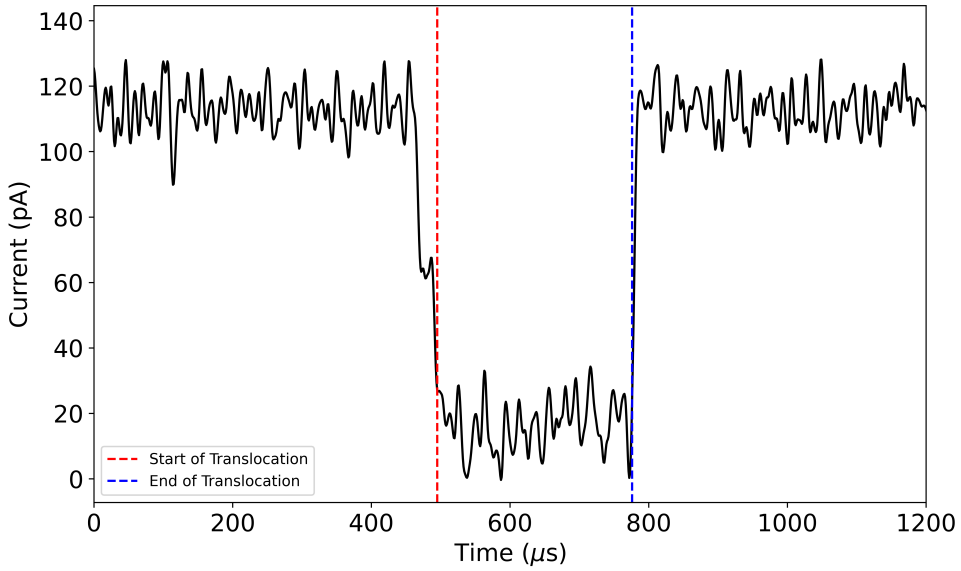
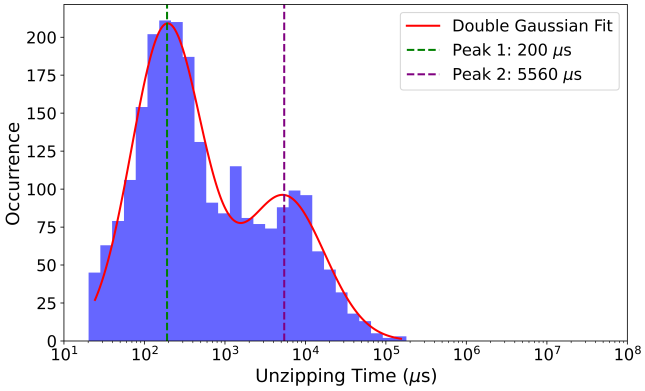
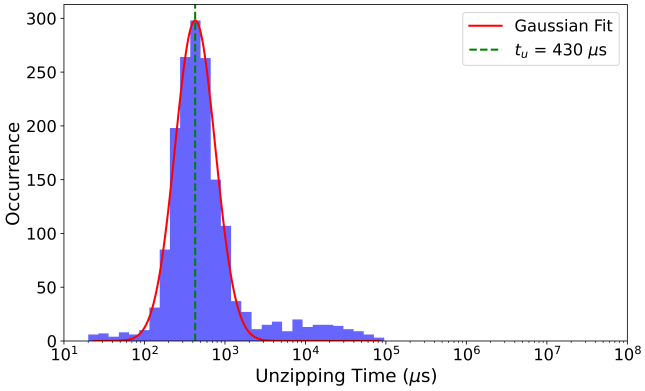


Figure 3.4: Current trace of poly(dA)$_{50}$ at 22°C at pH 7.5 in 1M KCl solution. Translocation time was calculated as 281 μs. The start and the end of the translocation are given in red and blue dashed lines, respectively.

Figure 3.6 represents the scatter plots of unzipping events, colored according to KDE, obtained for HP9sf at different applied voltages. Unlike the HP9s molecules displayed in Figure 3.3, the unzipping events for HP9sf exhibit single characteristic unzipping times, significantly shorter at each applied voltage. This difference highlights the potential distinct unzipping mechanisms between the flap and blunt-end structures, as introduced earlier in Figure 2.8. Unfortunately, the experiments do not allow obtaining this kind of information at the microscopic level. Additionally, the high-probability regions identified by KDE in Figure 3.6 shift towards shorter unzipping times as the applied voltage increases. This trend aligns with findings from HP9s studies. The characteristic unzipping times, extracted from these event clouds using the same method as for HP9s, along with the number of events for each voltage, are presented in Table 3.1. Notably, the unzipping times for HP9sf structures are consistently shorter than those for HP9s structures, as shown in Figures 3.3 and 3.6, as well as in Tables 3.1 and 3.2. To further illustrate this observation, Figure 3.5 provides histograms of the unzipping times for both structures at 120 mV, plotted on the same time scale.



(a) The double Gaussian fit for extracting the shorter and longer-time scales of HP9s experiments under 120 mV applied voltage. Times for shorter and longer, t_u, scales were calculated as 200 $\pm$ 70 and 5560 $\pm$ 2500 μs.



(b) The Gaussian fit for extracting the unzipping times of HP9sf experiments under 120 mV applied voltage. t_u was calculated to be 430 $\pm$ 120 μs.

Figure 3.5: Comparison of unzipping times, t_u, extracted from the Gaussian fits of the time value histograms for HP9s and HP9sf experiments under 120 mV applied voltage. The t_u of the HP9sf molecule is shorter than that of the HP9s molecule under the same experimental conditions.

For the HP9sf structure, experiments at higher voltages (160 and 180 mV) proved to be challenging due to frequent pore blockages and the occasional escapes of the αHL nanopore from the membrane after multiple unsuccessful attempts. Consequently, experiments were halted at 140 mV.

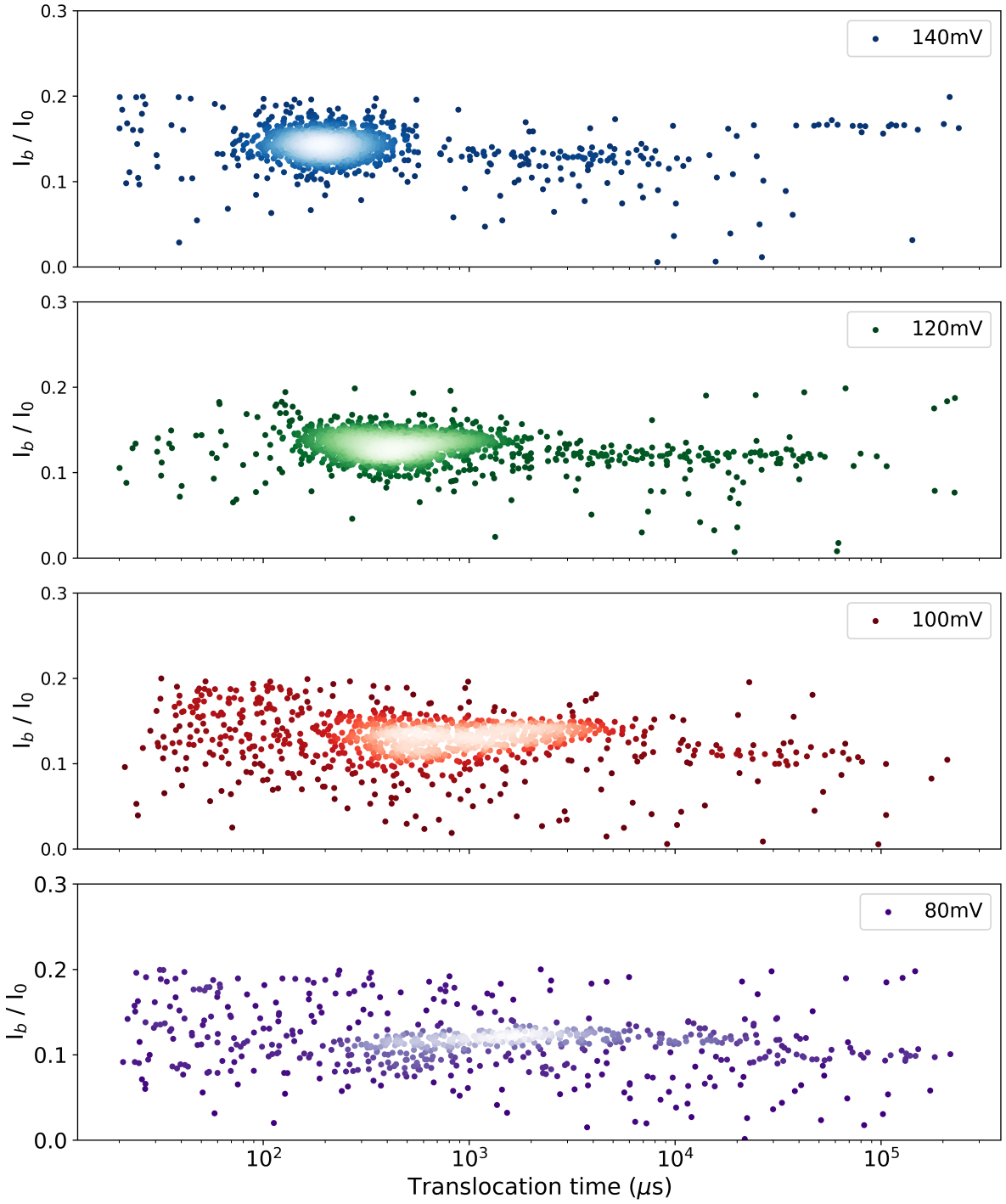


Figure 3.6: Scatter plot of HP9sf at different voltages at room Temperature 22°C at pH 7.5 in 1M KCl solution. From 80 mV to 140 mV, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events.

Table 3.2: Characteristic unzipping times of HP9sf at different voltages at room temperature (22°C) in 1 M KCl solution.

Voltage (mV)	Time (μs)	Number of events
80	1070 ± 360	1139
100	700 ± 200	1222
120	430 ± 120	1851
140	200 ± 60	1499

In this initial phase of the experiments, we kept the duplex length constant to evaluate how the applied voltage bias influences the unzipping times of flap HP9sf and blunt-end HP9s DNA structures. According to Equation 3.2, the unzipping time t_u exponentially depends on $E_b - q_e V$, where E_b mainly depends on the duplex length and can be calculated [173]. However the flap structure introduces a new variable in the calculation of E_b. To estimate E_b for both structures, HP9s and HP9sf, we utilized the `mfold` [173] server. `mfold` server predicts nucleic acid secondary structures by minimizing the free energy (ΔG), involves thermodynamic parameters that account for base pairing, base stacking, and loop penalties, and incorporates experimentally determined nearest-neighbour thermodynamic parameters and accounts for salt and temperature effects. The calculated E_b values were 9.4 kcal/mol (15.8 k_BT at 22C) and 10.3 kcal/mol (17.4 k_BT at 22C) for HP9s and HP9sf structures, respectively. This higher E_b predicted for the flap structure was unexpected, which proves the complexity of calculating the E_b when a flap structure is introduced. We cannot directly calculate the corresponding E_b values from our results. However, We can calculate the differences between the E_b values of both structures according to:

$$t_u = A \exp\left(\frac{E_b}{k_B T}\right) \times \exp\left(\frac{-q_e V}{k_B T}\right) \tag{3.3}$$

Where A can be considered as the inverse of an attempt rate for the hairpin opening [174] according to Kramer's theory and the same for both structures. When we take the logarithm of both sides, we have:

$$\log(t_u) = \left(log(A) + \frac{E_b}{k_B T}\right) - \left(\frac{q_e}{k_B T} V\right) \tag{3.4}$$

If we plot $\log(t_u)$ against the voltage for both structures, these plots should reveal a straight line with the slope giving us the $\frac{q_e}{k_B T}$, and the intercept giving the $log(A) + \frac{E_b}{k_B T}$. As mentioned, if we consider A the same for both structures, the differences in the intercept values should give the difference between the E_b values of both structures and can be compared with the `mfold` results. Figure 3.7 presents a semi-logarithmic plot of t_u versus applied voltage, along with fits to the experimental data for HP9s and HP9sf. The data generally follow a linear trend on the semi-logarithmic scale for both cases, indicating an exponential dependence of unzipping time on the applied voltage. This exponential behavior suggests the presence of two states, open and closed, supporting the application of Kramers' theory. We can consider the unzipping as a jump over a single energy barrier. This implies that the hairpin structures are unzipped before translocation through the

nanopore rather than through a series of partial unzipping and step-like translocation processes [60].

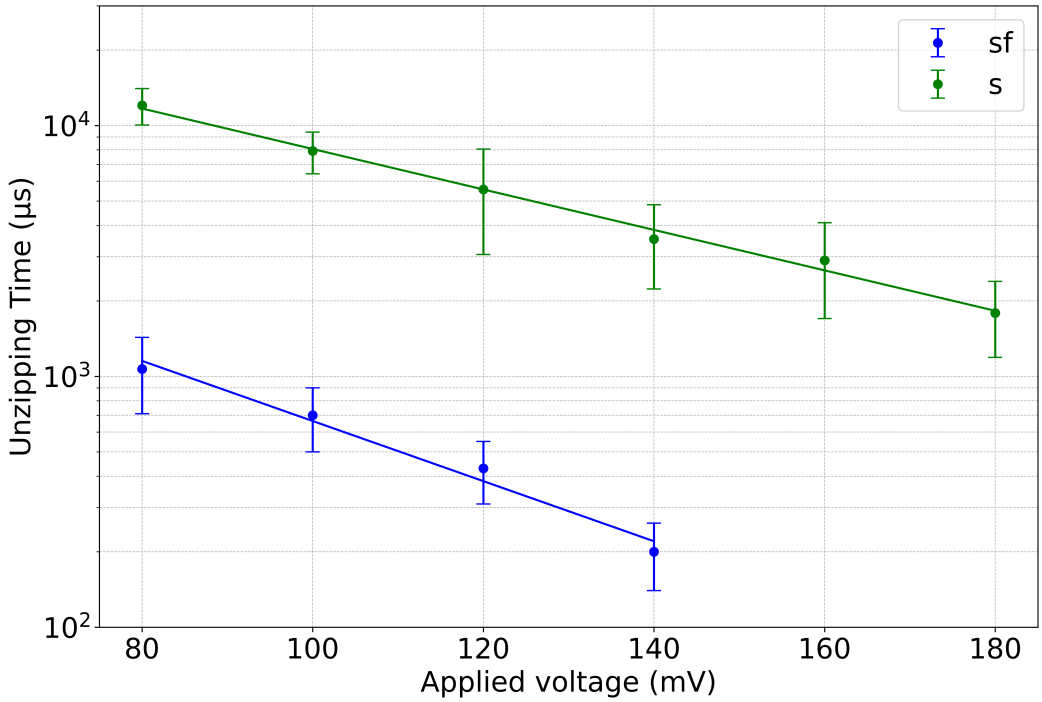


Figure 3.7: The unzipping times t_u obtained from HP9s and HP9sf are plotted against the applied voltage on a logarithmic scale. Both datasets were fit linearly on the logarithmic time scale to calculate q_e, where the slope provides $\frac{-q_e}{k_B T}$. The results from the linear fits are as follows: For HP9sf (blue line), the fit is $\log(t_u) = -0.03 \times V + 9.30$ with an R^2 value of 0.98. For HP9s (green line), the fit is $\log(t_u) = -0.02 \times V + 10.90$ with an R^2 value of 0.99.

The q_e values for HP9sf and HP9s were found to be 0.71 e and 0.47 e, respectively. These values align well with literature values and fall within an acceptable range for q_e, which is between 0 and 1 e [175, 54]. The larger q_e observed for HP9sf is unexpected, as q_e should be identical for both structures, given that q_e depends on the portion of DNA within the pore. Ideally, the fits in Figure 3.7 should be parallel if q_e were identical. This deviation suggests that the unzipping process for blunt-end and flap structures predicted by the theoretical model described in equation 3.2 is very simple and does not take into account that along the reaction coordinate (inside the nanopore), the energy barrier location changes as the energy $q_e V$ depends on this coordinate. As explained by Merstorf *et al.* [60], the applied voltage dynamically alters the energy landscape of the unzipping process. As the voltage increases, the energy barrier for unzipping not only lowers but also shifts along the reaction coordinate within the nanopore. This shift implies that the effective charge q_e and the energy barrier are not constant but vary depending on the position of the DNA hairpin within the nanopore. Consequently, the unzipping process for HP9sf and HP9s might occur under different energy landscapes, leading to the observed differences in q_e.

Dudko *et al.*[92] also studied the voltage dependence of hairpin unzipping and found that at high voltage regimes, the exponential dependence of t_u on voltage diminishes as the energy states (closed and open) come closer along the reaction coordinate. The different slopes observed in fits in Figure 3.7 for HP9sf and HP9s suggest that these structures might be unzipping under different voltage regimes, with HP9sf possibly experiencing a transition into a regime where the standard exponential voltage dependence no longer holds.

The energy barrier differences E_b between the two structures, HP9sf and HP9s, were determined from the experimental data using equation 3.4. Our analysis revealed a difference of $1.6\,k_BT$, indicating that HP9s has a slightly higher energy barrier compared to HP9sf. However, predictions from *mfold* suggested the opposite trend, with a calculated difference of $-1.6\,k_BT$, implying that HP9sf should possess a larger energy barrier compared to HP9s. The discrepancy between our results and the *mfold* predictions highlights the challenges in modeling the dynamic unzipping process for flap and blunt-end structures. While *mfold* predicts a higher E_b for HP9sf, this may be due to factors not fully captured by the software, such as dynamic structural changes during unzipping or interactions between the DNA and the nanopore environment.

Our experimental results show faster unzipping times for HP9sf structures, which would suggest a smaller energy barrier compared to HP9s. This observation supports the calculated E_b differences from Figure 3.7. However, it should be noted that this E_b difference was calculated by making two assumptions: first, the assumption that the process can be modeled as a two-state system using equation 3.2, which may be overly simplified. In reality, the unzipping process could be more complex than this model can account for, and unfortunately, these details cannot be fully visualized by our experiments. This underscores the importance of complementing experimental results with unzipping simulations. Alternatively, the assumption that the attempt rate A is identical for both structures might not hold. If these assumptions are inaccurate, they could lead to incorrect estimations of the energy barrier difference. These findings emphasize the difficulties in accurately modeling and interpreting the energy landscapes of DNA unzipping. The divergence between experimental data and theoretical predictions underscores the need for careful consideration of the underlying assumptions and the limitations of the models used.

3.4 Effect of the duplex length on unzipping time

Following the effect of the flap structure in the unzipping time, we continued the effect of the duplex length in both structures. Figure 3.9 illustrates the scatter plots of unzipping events for HP9s, 19s, and 29s structures at 120 mV. The same color code method is used to represent the relative density of events. From Figure 3.9, it can be seen that as the duplex portion gets longer, drastic shifts on the time axis occur. Specifically, there is nearly a two-order of magnitude difference every time the duplex length is increased by ten additional base pairs, showing an exponential dependence between t_u and E_b on the logarithmic scale. The unzipping times show two different peaks, as for the HP9s molecules; however, the probability of having the shorter-time scale times is reduced as the duplex length gets longer. Additionally, the unzipping times t_u spread more along the time axis for longer sequences, supporting the theory that the distribution widens

as the characteristic unzipping times increase. This behavior can be mathematically approximated by considering the probability distribution of unzipping times $P(t_u; L)$ for a DNA sequence of length L a log-normal distribution:

$$P(t_u; L) = \frac{1}{\sigma(\log(t_u))\sqrt{2\pi}} \exp\left(-\frac{(\log(t_u) - \mu(\log(t_u)))^2}{2\sigma^2(\log(t_u))}\right) \tag{3.5}$$

where $\mu(\log(t_u))$ is the mean of the logarithm of the unzipping time, which tends to increase with the duplex length, and $\sigma(\log(t_u))$ is the standard deviation, representing the spread of the distribution.

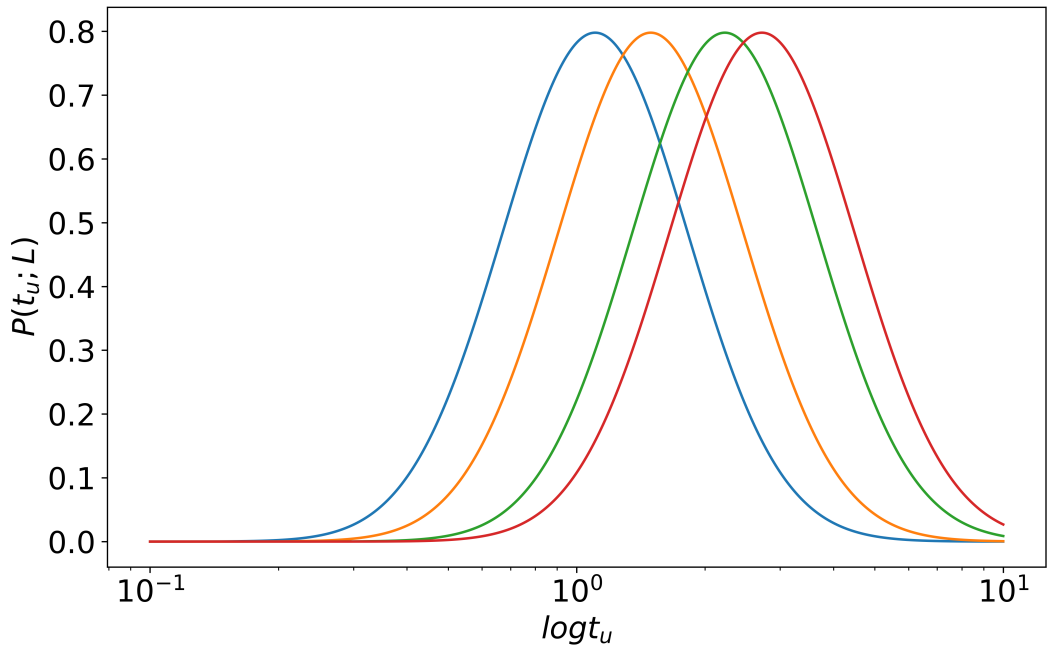


Figure 3.8: The probability function introduced in Equation 3.5 for different duplex lengths. $\sigma(\log(t_u))$ kept constant and $\mu(\log(t_u))$ assumed to be proportional to the duplex length L for the illustration, with a linear function: $t_u = k \times L$, k being a random constant. As the duplex length gets longer, the distribution shifts to the longer t_u values and gets wider on linear time scale.

Figure 3.8 shows a hypothetical case where as the sequence length L increases, the mean unzipping time $\mu(t_u)$ proportionally increases and the standard deviation $\sigma(t_u)$ stays the same, leading to a broader distribution $P(t_u; L)$ on linear time scale.

The characteristic times for shorter and longer-time scales extracted as explained above for HP9s and the number of events are given in Table 3.3. For the same reasons as explained above, we will be interested in the longer-time scales for the rest of this analysis.

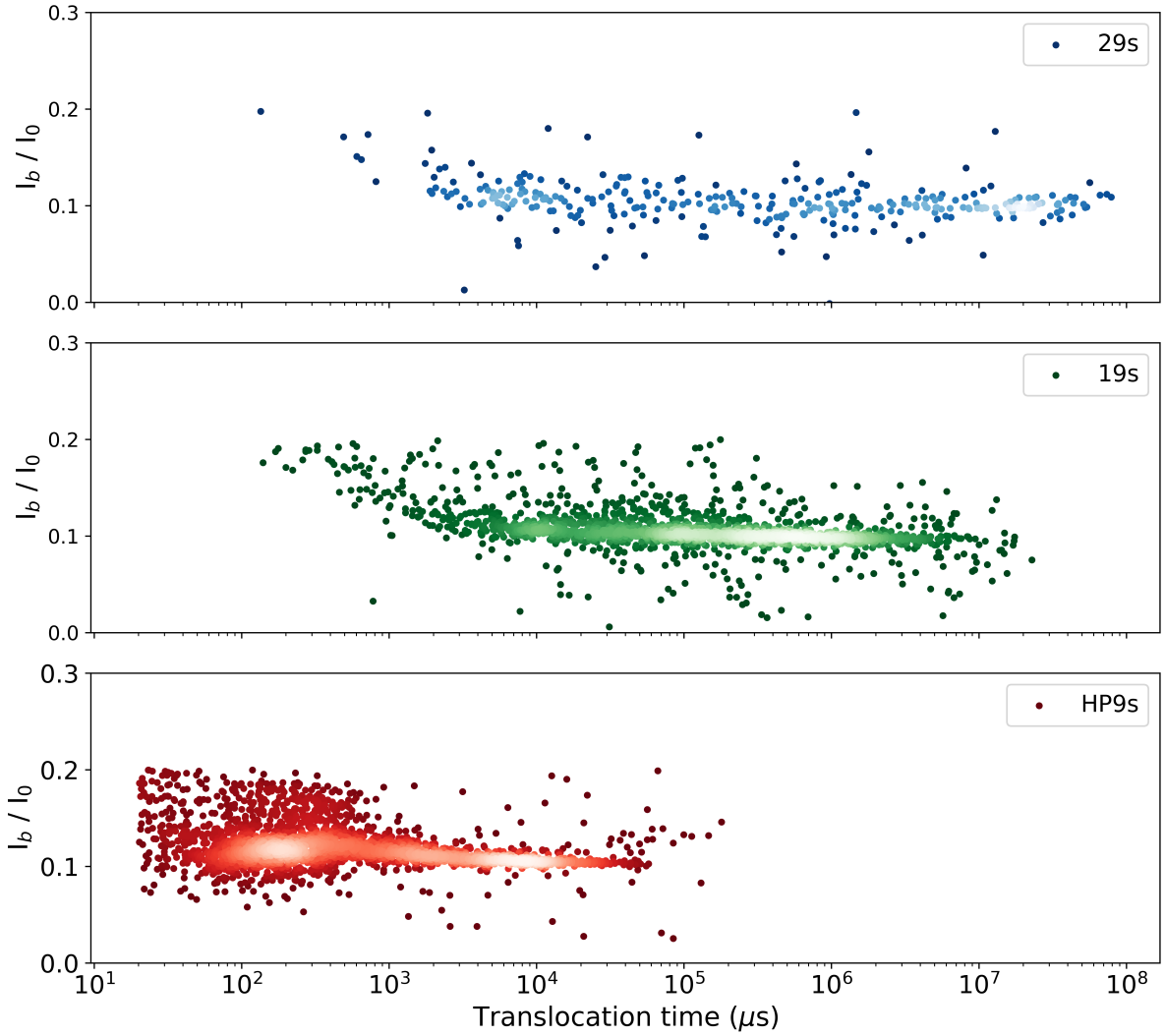


Figure 3.9: Scatter plot of HP9s, 19s, 29s at 120 mV applied voltage at room Temperature 22°C at pH 7.5 in 1M KCl solution. From HP9s to 29s, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events.

Table 3.3: Characteristic unzipping times of HP9s, 19s, 29s at 120 mV at pH 7.5 at room temperature (22°C) in 1 M KCl solution.

Length of duplex	Shorter-scale Time (μs)	Longer-scale Time (μs)	Number of events
9	200 ± 70	5560 ± 2500	2487
19	12370 ± 4780	$3.3x10^5$ ± 71400	1939
29	7980 ± 220	$1.6 \times 10^7 \pm 1.4 \times 10^6$	347

Figure 3.11 illustrates the scatter plots of unzipping events for HP9sf, 19sf, 29sf, and 39sf structures at 120 mV. The same color code method is used to represent the relative density of events. As for HP9sf, we observed a single characteristic unzipping time for each

duplex length. From Figure 3.11, it is not evident that the unzipping times (t_u) shift to longer durations on a logarithmic time scale as the length of the duplex portion increases, especially between 19sf, 29sf, and 39sf structures, no significant shifts are visible.

Although the scatter plot for the 39sf experiments does not contain enough unzipping events to deduce a characteristic unzipping time (t_u), our conclusion regarding the most probable events according to the KDE in Figure 3.11 is that these points likely represent the unzipping events. However, the characteristic t_u of these events falls around the same characteristic t_u as observed in the 29sf experiments, which seems misleading. We observed that the longest unzipping times observed during the 39sf experiments exceeded 10 s, as can be seen on the histogram of the t_u on Figure 3.10, making it difficult to collect a sufficient number of unzipping events. Therefore the events from 39sf experiments were not considered in the following analyses. Table 3.4 summarizes the characteristic unzipping times obtained from HP9sf, 19sf, 29sf, and 39sf structures, along with the number of events.

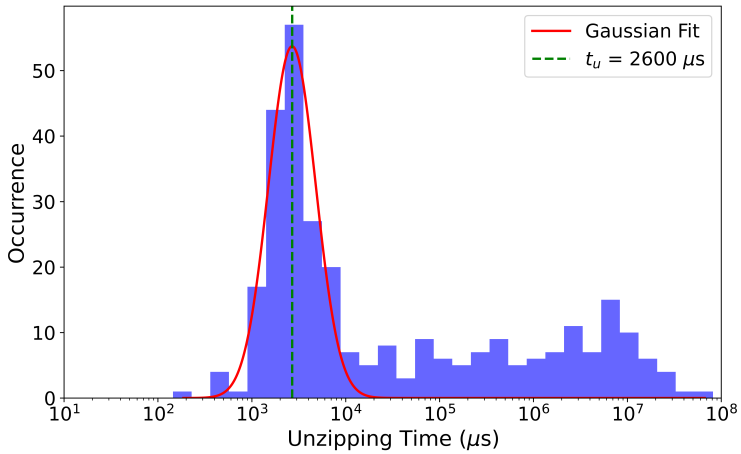


Figure 3.10: Histogram of the unzipping times t_u on logarithmic binning from the 39sf experiments. The t_u values spread along the time axis, varying between 10^2 to $10^8 \mu s$. The t_u was calculated as 2600 ± 820 μs.

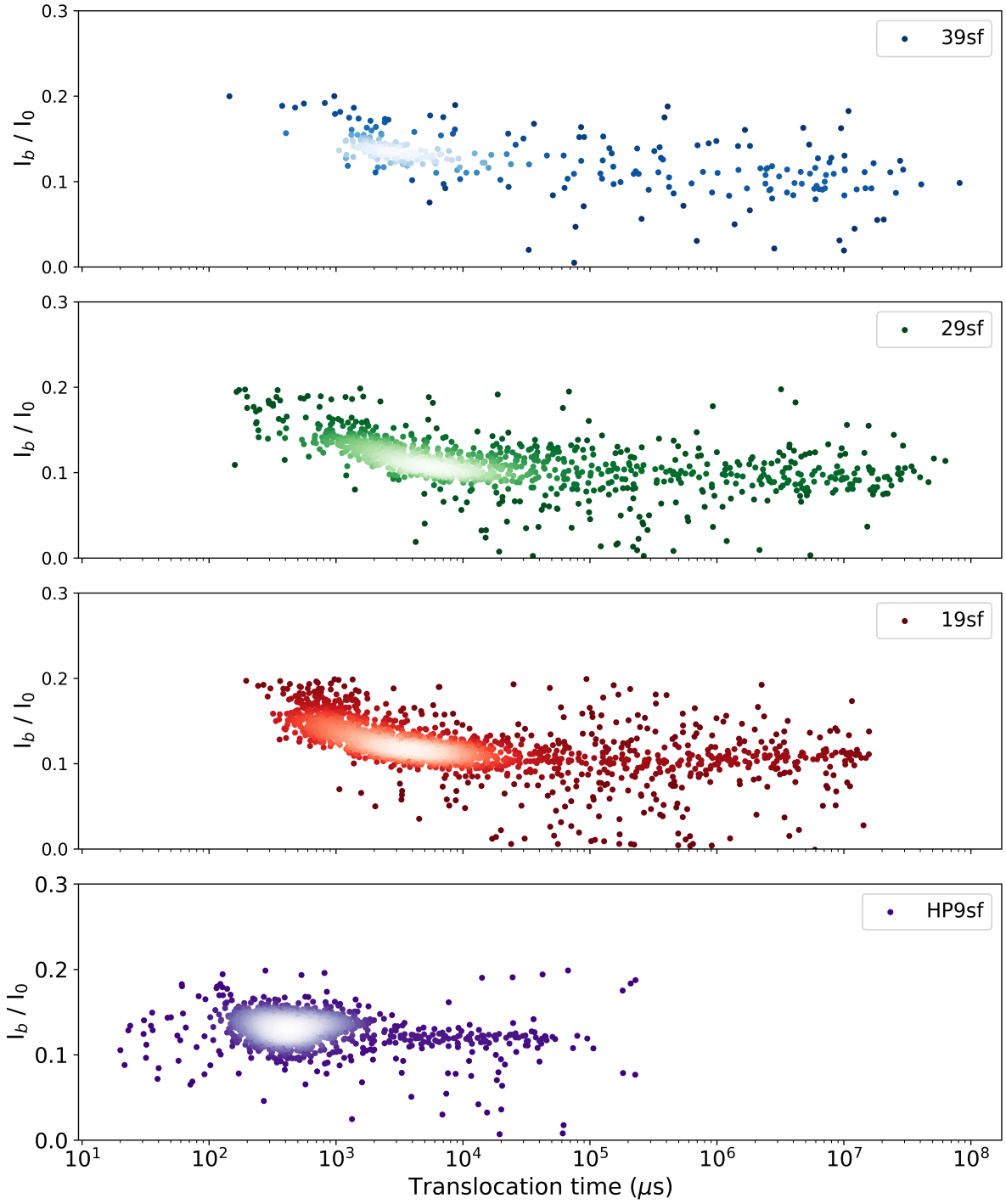


Figure 3.11: Scatter plot of HP9sf, 19sf, 29sf, 39sf at 120 mV applied voltage at room Temperature 22°C at pH 7.5 in 1M KCl solution. From HP9sf to 39sf, bottom to top. Each dot represents an unzipping event under the indicated applied voltage bias. A color code highlights the relative density of events on the plots: the brighter the dot, the denser the events.

Table 3.4: Characteristic unzipping times of HP9sf, 19sf, 29sf at 120 mV at pH 7.5 at room temperature (22°C) in 1 M KCl solution.

Length of double-stranded portion	Time (μs)	number of events
9	430 ± 120	1851
19	2770 ± 800	1913
29	5050 ± 1260	1214
39	2600 ± 820	345

For blunt-end structures, we observed that the unzipping time t_u increases exponentially with the duplex length, as demonstrated in Figure 3.9. If we plot t_u against the duplex length on a semi-logarithmic scale, the exponential fit should appear as a straight line. In contrast, a clear exponential behavior is not readily apparent for flap structures, as illustrated in Figure 3.11, suggesting that the unzipping mechanism may deviate from the two-state Kramer's theory. To investigate the relationship between t_u and the duplex length for flap structures, we considered both a two-state exponential model and a linear model. In the linear model, the unzipping time is assumed to increase proportionally with the duplex length, where each base pair contributes to the unzipping process via consecutive unzipping events. By comparing these models, we aimed to gain deeper insights into the underlying mechanism.

When plotting t_u against duplex length on a semi-logarithmic scale, assuming 2-state Kramer's theory, we aimed to investigate the effect of flap and blunt-end structures by calculating the duplex length that corresponds to the thermal energy k_BT during unzipping. Under constant voltage, the dependence of t_u on different duplex lengths can be described by rearranging Equation 3.2 as follows:

$$t_u \propto \exp\left(\frac{E_b}{k_BT}\right) \propto \exp\left(\frac{n\epsilon}{k_BT}\right), \tag{3.6}$$

where n represents the number of base pairs in the duplex, and ϵ is the energy associated with each base pair. The right-hand side of Equation 3.6 can be further simplified to:

$$t_u \propto \exp\left(\frac{n}{n_0}\right), \tag{3.7}$$

where n_0 represents the number of base pairs in a duplex that corresponds to the thermal energy k_BT. The value of n_0 can be determined from the inverse of the slope of t_u against the duplex length on a semi-logarithmic plot that we will calculate.

For the random sequences used in our experiments, where regions rich in G-C or A-T content are intentionally avoided during synthesis, the slope of the plot of E_b versus duplex length represents the energy required to unzip a single base pair in the DNA duplex. Therefore we plotted the E_b values predicted from `mfold` against duplex length, although `mfold` systematically predicts higher E_b values for flap structures compared to blunt-end structures. This slope is expected to fall between 2 and 3 because G-C pairs have 3 hydrogen bonds, while A-T pairs have 2 hydrogen bonds.

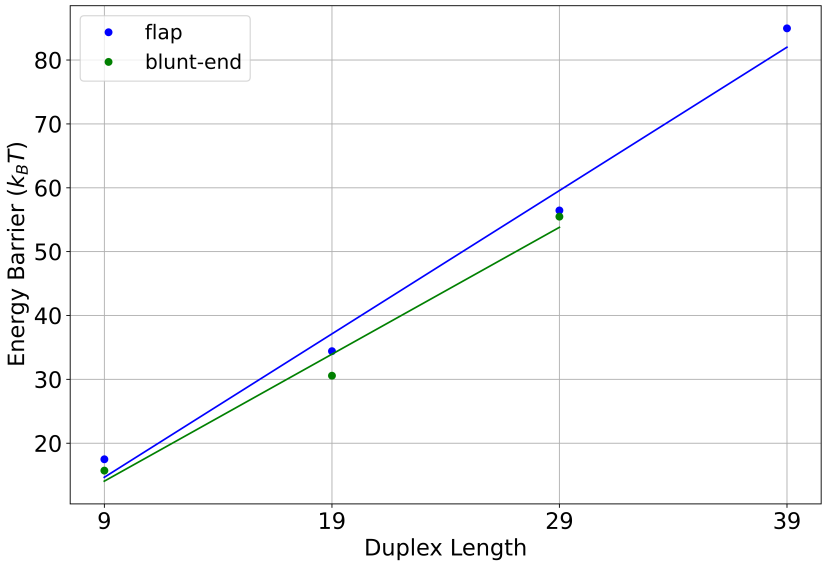


Figure 3.12: Relationship between E_b and the duplex length was investigated using the results from `mfold` for blunt-end and flap structures. The slopes for blunt-end structures and for flap structures were calculated to be 2.0 and 2.2, respectively.

The slope obtained from `mfold`, as shown in Figure 3.12, provides an estimate of the energy per base pair, which we will compare to the experimental n_0 value to validate the theoretical predictions. We found that the slope for blunt-end structures was 2.0, while for flap structures, it was 2.2.

Figure 3.13 represents the semi-logarithmic plots of t_u against duplex length for both structures. The differences in the scale of the time axis between flap and blunt-end structures are visible, with much longer unzipping times for the blunt-end structures, confirming the difficulty of unzipping these types of molecules compared to flap structures despite the same number of base pairs being unzipped. According to equation 3.7, n_0 was calculated as 8.33 for flap structures and 2.50 for blunt-end structures.

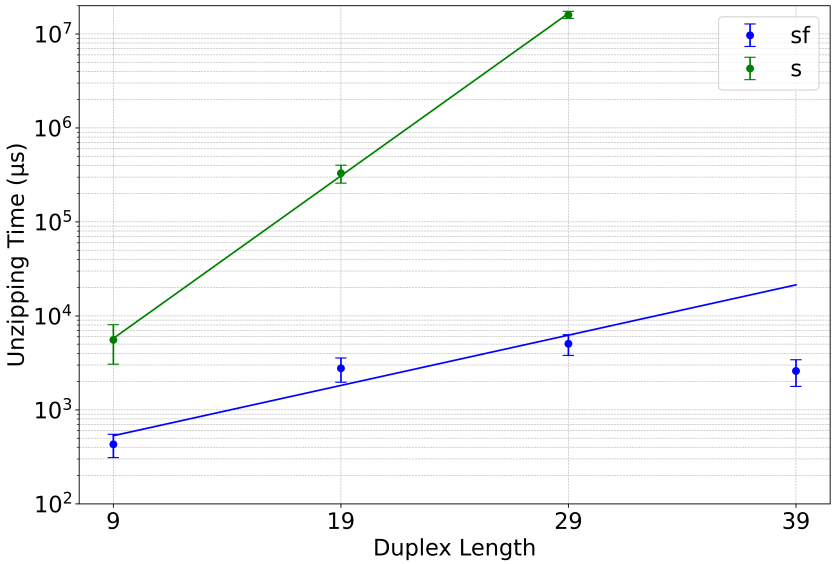


Figure 3.13: Unzipping times t_u obtained from HP9s, 19s, 29s and HP9sf, 19sf, 29sf, 39sf are plotted against the duplex length in structure on a semi-logarithmic scale. Both datasets were fit on the semi-logarithmic time scale according to equation 3.7, except for 39sf, to calculate n_0. The results from the fits are as follows: for the flap structures of different duplex lengths (blue line), the fit is $\log(t_u) = 0.12 \times n + 5.20$ with an R^2 value of 0.92. For the blunt-end structures of different duplex lengths (green line), the fit is $\log(t_u) = 0.40 \times n + 5.10$ with an R^2 value of 0.99.

The n_0 calculated for the flap structures, 8.33, which corresponds to 0.12 k_BT per base pair, is much higher than the expected values. Furthermore, the exponential fit in Figure 3.13 shows some deviations from the expected behavior, as mentioned above. Specifically, the unzipping data does not follow an exponential dependence on duplex length, as equation 3.7 suggested. This is supported by a poorer R^2 value accompanying the fit compared to the R^2 value found for blunt-end structures. The poorer R^2 shows that the variability in unzipping times for flap structures is less well-explained by the model. To further investigate the unzipping dynamics of flap structures, we plotted t_u against the duplex length on a linear scale, as shown in Figure 3.14.

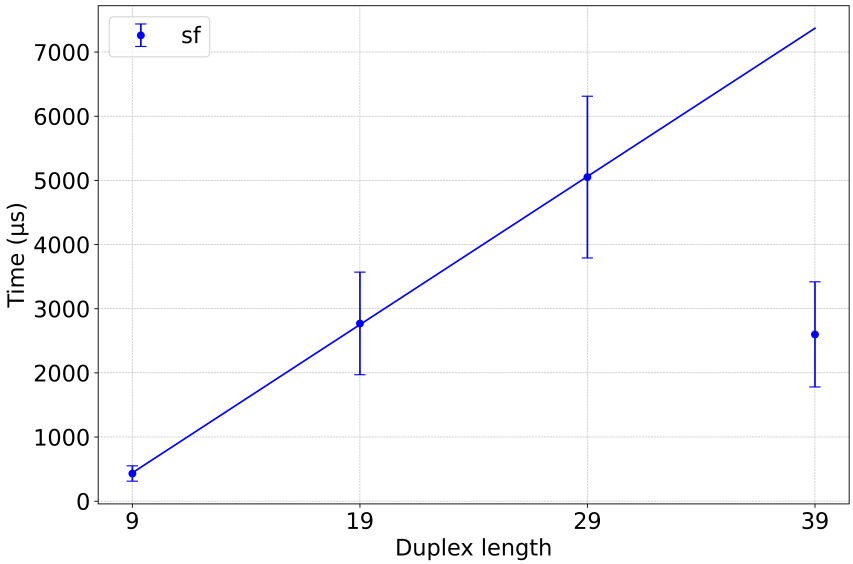


Figure 3.14: Unzipping times t_u obtained from HP9sf, 19sf, 29sf, 39sf are plotted against the duplex length in structure in linear scale. Fit parameters for the first three points were found as $t_u = 230 \times l - 1640$ with a R^2 value of 0.99

The linear plot of t_u against the duplex length was performed using results from HP9sf, 19sf, and 29sf, as the data set for 39sf was deemed insufficient. The observed linear dependence of t_u on the duplex length in flap structures with a high R^2 value suggests that the unzipping process deviates from a simple two-state transition, as described by Kramers' theory where the rate of unzipping is exponentially related to the energy barrier. The linear relationship observed here implies that the unzipping process involves a series of sequential events rather than a single transition. Specifically, this linear dependence indicates that each base pair contributes an amount of time to the overall unzipping process. This sequential unzipping and translocation lead to a cumulative time effect that scales linearly with the number of base pairs rather than following an exponential dependence, contrasting with the two-state model's predictions. Such a mechanism is likely confined to the entrance of the nanopore, as the vestibule does not provide space for repeated unzipping and translocation of DNA duplexes. A previous experimental study [60] also observed that the unzipping time distribution for long duplexes is slower than exponential, suggesting that a two-state Kramer's-like approach is insufficient for these cases. It is more probable that unzipping occurs progressively toward the duplex end. This same experimental study also mentioned that for longer duplexes, the dynamics might depend on the energy landscape of the system, particularly if G-C-rich regions induce pauses during the unzipping process [90, 54]. In our design, we avoided this by ensuring a more homogeneous distribution of base pairs.

3.5 Conclusion

We performed ssDNA translocation and dsDNA unzipping experiments. As a result of the translocation experiments, we observed that the differences in translocation times between the 3' end and the 5' end, as well as the blocked current levels originating from the translocation of different orientations, were consistent with the literature. Specifically, the 3' end was observed to translocate faster than the 5' end. Additionally, the blocked current levels varied depending on the orientation of the DNA, with the 3' end causing more blockage of the current during its translocation.

During the unzipping experiments, the existence of two distinct time scales—short and long—along with systematically longer unzipping times for the blunt-end structures, confirmed the influence of an additional flap structure on unzipping mechanisms. These observations reveal the existence of different unzipping mechanisms between duplex structures. The shorter-time scale values observed during the unzipping of blunt-end structures might correspond to the simple translocation of the 50 nucleotide-length poly(dA) or might be due to the dsDNA, which are not correctly inserted in the nanopore and, therefore, were not taken into account as unzipping events. The t_u depends exponentially on the applied voltage, indicating a two-state unzipping process: closed and open states. This two-state process implies that the duplex is first completely unzipped and then translocated. For short flap structures (HP9sf), we also observed an exponential dependence of t_u on the applied voltage with a single unzipping time at each voltage. The differences between the unzipping processes of the blunt-end and flap structures can be attributed to the positions at which the unzipping occurs. The unzipping might occur either at the entrance of the nanopore or in the vestibule just above the pore constriction for blunt-end structures. However, for flap structures, the unzipping likely occurs at the entrance of the nanopore in a zipper-like fashion, as illustrated in Figure 2.8. Additionally, the different effective charge, q_e, values calculated for the HP9s and HP9sf structures indicate the complexity of the unzipping process. These observations and results highlight the need for a more advanced model of unzipping mechanisms, especially when considering the effect of the additional sequence in the flap structure on E_b, which may enhance electrical forces during DNA unzipping due to the additional negative charges present on this structure.

The effect of duplex length on unzipping times was also evident in our experiments, with longer unzipping times measured as the duplex length increased. The exponential dependence of unzipping time on duplex length observed for blunt-end structures supports the two-state unzipping process. In contrast, a linear dependence of t_u on duplex length was observed for longer duplex lengths in flap structures. This linear dependence of t_u on duplex length, shown in Figure 3.11, suggests a series of unzipping and translocation processes for these flap structures, with the presence of intermediate states during unzipping contributing to the overall unzipping time.

To explain the observation of two different time scales for blunt-end structures, we also considered the possible entry orientations of the DNA duplexes. As depicted in Figure 2.7, the design of the blunt-end DNA allows entry from either the 5' end or 3' end, which may lead to different unzipping times (t_u). The variation in translocation times between the 5' end and the 3' end entry could account for observing different unzipping time regions observed in blunt-end structures. However, as discussed in Section 3.2, entrance from the

5' end and 3' end results in different blocked current levels, which was not observed as a result of our experiments, apart from the experiments with HP9s at 100 mV which can be seen in Figure 3.2. In contrast, for flap structures, entry from the 5' end is less likely due to their design, which is further supported by the observation of a single t_u.

However, our experimental setup does not allow us to conclusively determine the causes of the differences between flap and blunt-end structures. To address potential orientation differences and eliminate uncertainties, we decided to redesign our DNA molecules to prevent entry from the 5' end and test our hypothesis. The new DNA molecule design, shown in Figure 3.15, adapts the length of the base O60 sequence to prevent a free single-stranded part at the 5' end after hybridization. Testing these redesigned DNA molecules is of great importance to provide more explanations on the dynamics of unzipping, as these modifications will eliminate some of the sources of doubt mentioned above.

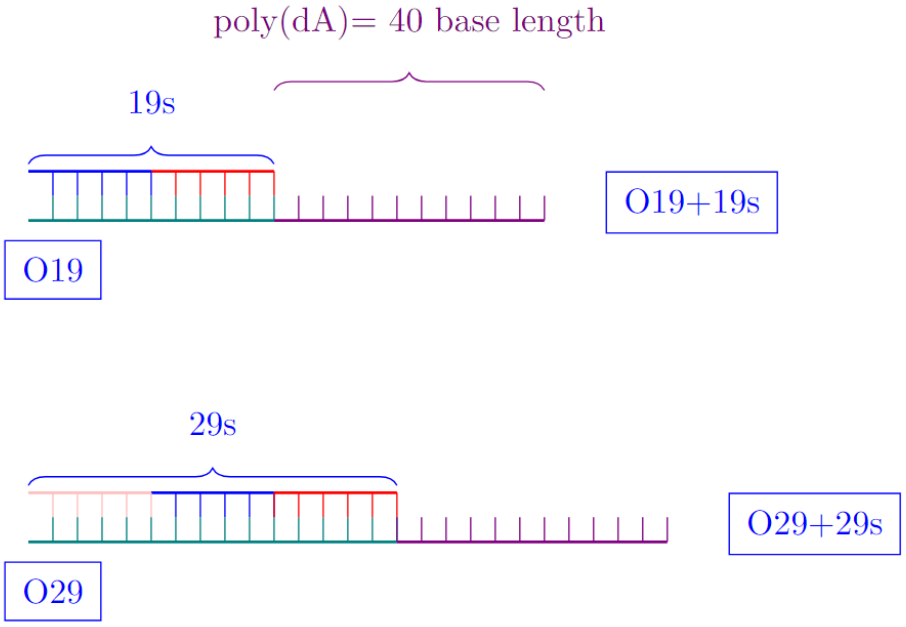


Figure 3.15: New design of the experimental dsDNAs.The template sequence (O60) serving for pairing with longer sequence strands than 9 bases has been changed and adapted for each length to prevent having a free single-stranded sequence from the 5' end. The same color is used to indicate the identical sequences used to pair longer sequences with the adapted sequence. Each time, a new sequence of 10 nucleobases is added to make the hybridized portion longer, with the adapted base sequence called O19 or O29, depending on the sequence studied.

In conclusion, our experiments demonstrated that the unzipping mechanisms for DNA molecules are influenced by their structural features and duplex length. The complexity of the unzipping process exceeds the predictions of Kramers' theory, highlighting the need for further investigation with improved theoretical models and computational simulations since the experiments are not allowing us to conclude the mechanisms underlying the unzipping of the molecules studied in this thesis. Our group has previously modeled the ionic current through this nanopore, and in Chapters 4 and 5, we will present our results

on the modeling of ssDNA translocation, which are essential for further modeling the dsDNA unzipping process for discovering the microscopic details of this process.

Chapter 4

Poly(dA) translocation using CG-SMD

Contents

4.1 Cf-SMD system preparations and simulation parameters . . . 99

4.1.1 Simulation Setup for ssDNA Molecules 99

4.1.2 Preparation of the ssDNA-Pore System 100

4.1.3 Steered Molecular Dynamics Simulations 103

4.2 Calculating ssDNA Translocation Time per Base 103

4.2.1 Translocation time distribution for 3' oriented ssDNA. 105

4.2.2 Influence of ssDNA orientation on translocation time. 107

4.2.3 Influence of ssDNA charges on translocation time 109

4.3 DNA-nanopore interactions during translocation 110

4.4 DNA tilting angles during translocation: 3' vs 5' 113

4.5 Conformations of ssDNA during translocation 117

4.6 Conclusion 123

This chapter presents the findings from coarse-grained (CG) simulations of poly(dA) molecules translocating through the α-hemolysin (αHL) nanopore exploiting constant-force Steered Molecular Dynamics (cf-SMD) simulations. We examined how DNA length, orientation, and charge affect the translocation time, which is the main measurable parameter in DNA translocation experiments and one of the most analyzed aspects of theoretical studies.

While experimental data primarily offer translocation times and current measurements, computational simulations can provide more detailed information on the DNA translocation process. By comparing experimental data with results from simulations, we can validate our models and gain deeper insights into the physical mechanisms involved. In this study, we used the wild-type (WT) αHL nanopore system, as previously investigated [145, 147], and employed the MARTINI CG force field (described in detail in Section 2.2.2). Our initial simulations used an external electric field to mimic the experimental conditions, with a field strength of 0.03 V/nm, corresponding to an electric potential difference of approximately 800 mV according to the relation $V = E \times d$ where V is the voltage difference, E is the electric field strength and d is the distance across which the

field is applied. However, this approach led to the ssDNA collapsing at the pore entrance and did not result in a translocation (see Figure 4.1). This behavior can be attributed to the simplified electrostatics of the MARTINI force field, as detailed in Chapter 2. The MARTINI model, applied to both the protein embedded in a DPPC bilayer and the bilayer itself, cannot accommodate excessively strong electric fields (above 0.04 V/nm), as discussed in our group's previous studies [146]. At such high field strengths, the system becomes unstable, and the lipid membrane breaks down. To address this limitation, we utilized constant-force Steered Molecular Dynamics (cf-SMD), which allows us to apply an external force to guide the translocation of ssDNA through the nanopore.

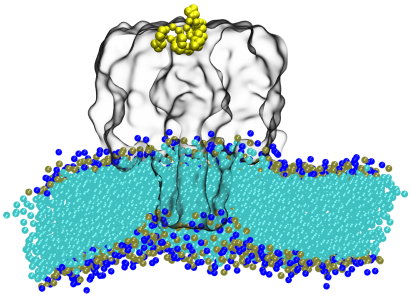


Figure 4.1: ssDNA simulation for a poly(dA) DNA molecule of 10 nucleotides in the presence of an electric potentiel difference of approximately 800 mV. The applied electric field is not sufficient to drive the ssDNA molecule through the nanopore, and the ssDNA molecule interacts with the cap of the nanopore. Snapshot is taken after 500 ns.

In Section 4.1, we detail the preparation of our systems before delving into the results. The findings presented in Sections 4.2 to 4.6 are based on our previously published work [176]. Section 4.2 presents our findings on translocation time per base calculations for both orientations and compares our results with experimental data. Additionally, this section discusses MD simulations involving a neutral ssDNA, comparing translocation times between charged and neutral ssDNA molecules. Section 4.3 examines the interactions between the ssDNA molecule and the inner surface of the nanopore using contact maps. In Section 4.4, we investigate the tilting angle of the ssDNA molecule bases during their translocation to understand the impact of ssDNA orientation on the translocation process. Section 4.5 focuses on the conformational dynamics of the ssDNA molecule within the αHL nanopore, particularly on inter-base distances during translocation. Finally, Section 4.6 concludes this first part of our computational study.

4.1 Cf-SMD system preparations and simulation parameters

In this section, we detail the preparation of the ssDNA molecules and the final ssDNA–αHL nanopore–DPPC system. Our simulations utilize the wild-type (WT) αHL nanopore and the polarizable water (PW) model of the MARTINI force field. The PW water model was chosen due to its ability to accurately represent electrostatic interactions between water molecules and charged particles, as well as the behavior of charged particles under an applied electric field [168]. The suitability of the PW model for this coarse-grained (CG) system has been demonstrated in previous studies by our group [145, 146, 147]. Specifically, it was found that when an external electric field is applied, the potential differences across the membrane in the CG simulations align closely with those observed in all-atom (AA) simulations [145]. As mentioned earlier, our investigations began with the aim of mimicking the electric voltage bias applied in single-molecule nanopore experiments. Therefore, using the PW water model was essential to ensure that the behavior of charged particles, such as ions and ssDNA, under an electric field was accurately represented.

All molecular dynamics (MD) simulations were conducted using GROMACS software version 2020.1, employing the MARTINI 2.2p (polarizable) force field [161]. The system temperature was maintained at 320 K using the v-rescale thermostat, while pressure, during the equilibration, was controlled at 1 bar with the Berendsen barostat. Electrostatic interactions were managed using the particle-mesh Ewald (PME) method [150, 177], with a 2 Å Fourier grid spacing and a 13 Å direct space cutoff radius. Periodic boundary conditions were applied throughout.

4.1.1 Simulation Setup for ssDNA Molecules

ssDNA molecules, ranging from 10 to 19 nucleotides poly(dA), were constructed by modifying dsDNA using the builder function in PyMol [178], followed by the deletion of one strand. The `martinize_dna.py` tool [179] (version 2.2) was then used to convert the ssDNA into a CG model, using the `-dnatype ss`option. To account for the phosphate group missing at the 5' end, a $-1e$ charged Q0 type MARTINI bead was manually added. The ssDNA molecule was positioned within a 25 $\times$ 25 $\times$ 25 nm^3 simulation box and was oriented along the z-axis. Initial energy minimization was performed for 100 steps using the steepest descent algorithm, followed by a pulling process on both ends of the ssDNA with a 100 kJ mol^{-1} nm^{-1} force over 10 ps using a 2 fs time step. The system was then solvated with approximately 130,000 PW water molecules and subjected to further minimization (50,000 steps) and equilibration over 50 ns with an incrementally increasing time step from 2 fs to 10 fs. Subsequently, 1M NaCl MARTINI ions were added using the `genion` tool in GROMACS, and the system was minimized again before a final 20 ns equilibration in the NPT ensemble with positional restraints on the DNA backbone, applying a force constant of 1,000 kJ mol^{-1} nm^{-2}.

For the neutral poly(dA) ssDNA molecule comprising 16 nucleotides, modifications to the MARTINI force field were made by substituting the charged $Q0$ type BB1 beads (representing phosphate groups) with neutral $N0$ type BB1 beads. This modification was performed to avoid bias from Lennard-Jones (LJ) interactions. In the MARTINI force field, each bead type has specific interaction parameters governed by LJ potentials.

The $Q0$ type beads, being charged by definition, have a high interaction ε value with other bead types, which can introduce undesired interactions of neutral molecules with the charged residues of the αHL (See Chapter 2.2.2 and Table2.1 for further details). By using the neutral $N0$ type beads, we ensured that the interactions between beads more accurately represent those of a neutral DNA molecule, thereby providing a more proper depiction of the neutral poly(dA) ssDNA.

4.1.2 Preparation of the ssDNA-Pore System

The CG system for the αHL nanopore embedded in a DPPC lipid membrane had been validated in previous studies [145, 146, 147], as explained in Chapter 1. The ssDNA was positioned just above the αHL nanopore, with either the 3' or 5' end facing the pore entrance in a $15 \times 15 \times 27$ nm^3 simulation box. Figure 4.3 provides an example of the initial system configuration. The initial setup ensured that the distance between the center of mass (COM) of the ssDNA 3' or 5' end and the COM of the pore constriction was no more than 6 nm, and the entry angle was limited to a maximum of 10°, as detailed in Table 4.1. The entry angle was defined as the angle between two vectors: one from the COM of the ssDNA 3' or 5' end nucleotide to the COM of the pore constriction ring (E111/K147), and the other from the COM of the pore constriction ring to the COM of the pore bottom (D127/K131), see Figure 4.2 for illustration. The system was solvated with approximately 40,000 PW-type water molecules. Excess water within the lipid membrane was removed using a script adapted from `water_deletor.pl`, as provided in the GROMACS tutorial [180]. The system underwent a maximum of 100,000 steps of energy minimization using the steepest descent algorithm, followed by equilibration first in the NVT ensemble for 20 ns and then in the NPT ensemble for 30 ns, with constraints on the DNA and protein backbones using a 10 fs integration step. After adding 1 M NaCl MARTINI ions using the `genion` tool to replicate the ionic strength used in our experiments, the system was minimized for another 100,000 steps and equilibrated for an additional 50 ns using the same protocol.

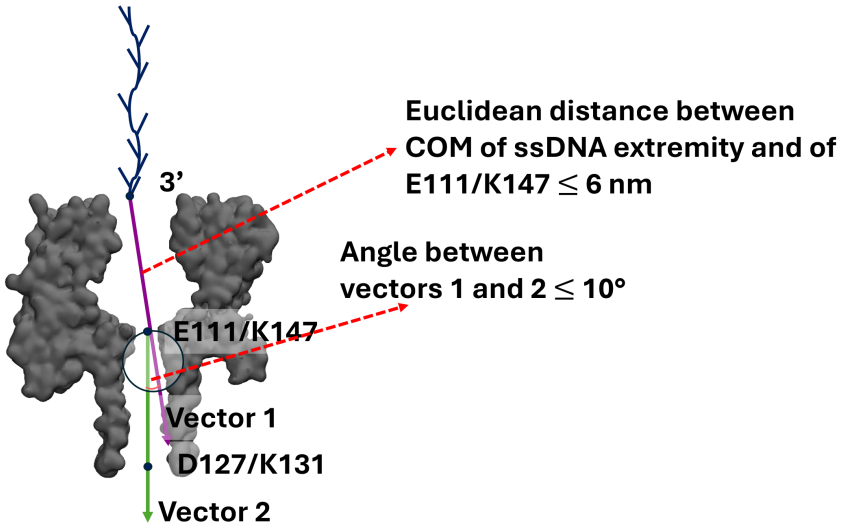


Figure 4.2: The illustration of the initial system requirements. The distance between the COM of the ssDNA extremity and the COM of the pore constriction should be smaller than 6 nm, and the entry angle, which was defined as the angle between two vectors: Vector 1 one from the COM of the ssDNA 3' or 5' end nucleotide to the COM of the pore constriction ring (E111/K147), and Vector 2 from the COM of the pore constriction ring to the COM of the pore bottom (D127/K131), is limited to a maximum of 10°.

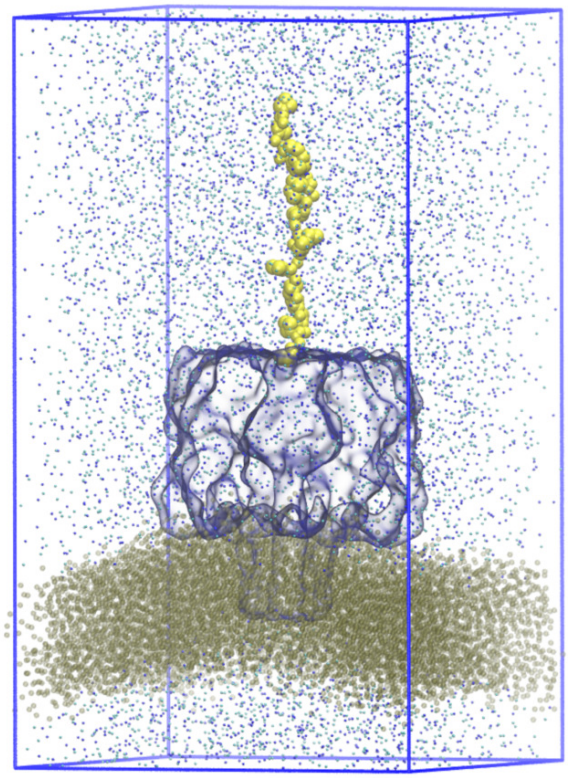


Figure 4.3: CG representation of the system at its initial conformation with 10-nucleotide poly(dA) molecule. The DPPC lipid bilayer is depicted with brown points, ions with cyan and purple points, poly(dA) in yellow beads, and the α-hemolysin nanopore in surface representation. Water is not shown.

Table 4.1: Entry angle and distance of ssDNA molecules in initial structures.

System name	Entry angle (°)	Distance (nm)
3'-A_{10}	4.56	5.41
3'-A_{11}	9.57	5.46
3'-A_{12}	8.47	5.06
3'-A_{13}	4.30	4.47
3'-A_{14}	2.35	4.73
3'-A_{15}	7.45	5.32
3'-A_{16}	6.51	4.70
3'-A_{16}N	9.69	3.86
5'-A_{16}	8.92	4.94
3'-A_{17}	0.83	5.23
3'-A_{18}	2.28	5.11
3'-A_{19}	5.04	5.02
5'-A_{19}	6.64	5.33

System names are given: Orientation- # of nucleotides, replicates are mentioned as system name-simulation # (*e.g.* 3'-A_{10}-1 and 3'-A_{10}-2)

4.1.3 Steered Molecular Dynamics Simulations

Cf-SMD simulations were performed to investigate ssDNA translocation. Depending on the orientation, either the 3' or 5' end nucleotide of the ssDNA was chosen as the pull group, with the central constriction of the αHL nanopore serving as the reference group. In cf-SMD simulations, the pulling force was applied in a specified direction relative to the reference group. With the reference group fixed, the pulling force moves the pull group along the defined direction. A constant force of 400 kJ mol^{-1} nm^{-1} was applied in the z-direction to the COM of the pull group, inducing ssDNA translocation through the nanopore. A time step of 10 fs was used to accurately capture the dynamics of the ssDNA movement under the applied force as the DNA moves rapidly. A total of twelve MD systems were prepared, as summarized in Table 4.2. The length of the ssDNA was varied from 10 to 19 nucleotides for 3' end pulling. Additionally, one system was prepared for 5' end pulling and another for neutral ssDNA pulling using a 16-nucleotide poly(dA) molecule. Each system was simulated in two replicas for at least 1 μs in the NVT ensemble. Finally, two cf-SMD simulations with 16-nucleotide ssDNA were conducted for 500 ns without the αHL nanopore, applying the same pulling force to either the 3' or 5' end. A reference MD simulation with a 16-nucleotide ssDNA free in solution was also performed for 500 ns.

4.2 Calculating ssDNA Translocation Time per Base

In our CG-SMD simulations, we aimed to study the translocation time per base since this parameter is the main measurable from the nanopore experiments and the most studied one in the previous theoretical studies. The translocation time calculated via CG-SMD simulations can be directly compared with the experimental and previous theoretical

Table 4.2: Summary of ssDNA translocation simulations

System name	# of nucleotides	Orientation	Total # of complete translocations per μs		Total # of bases translocated per μs	
			sim. #1	sim. #2	sim. #1	sim. #2
3'-A_{10}	10	3'	10	8	102	85
3'-A_{11}	11	3'	8	9	92	99
3'-A_{12}	12	3'	5	5	61	62
3'-A_{13}	13	3'	1	2	13	26
3'-A_{14}	14	3'	3	2	44	29
3'-A_{15}	15	3'	0[b]	4	3	60
3'-A_{16}	16	3'	3	2	52	35
3'-A_{16}N	16[a]	3'	3	5	51	87
5'-A_{16}	16	5'	0[b]	1	6	16
3'-A_{17}	17	3'	2	1	39	19
3'-A_{18}	18	3'	0[b]	0[b]	4	4
3'-A_{19}	19	3'	1	0[b]	23	12
5'-A_{19}[c]	19	5'	0[b]	0[b]	7	5

[a] with neutralized phosphates groups
[b] partial translocation
[c] Additional simulations for Section 4.4, which are not taken into the calculations of the histograms
Replica simulations are mentioned as system name-simulation # (*e.g.* 3'-A_{10}-1 and 3'-A_{10}-2)

studies. We varied the length of the ssDNA molecule, its orientation during translocation, and its charge and observed the differences in the translocation times of these molecules. For this purpose, a complete base translocation event is defined when a single base enters the constriction (E111/K147) and passes through the COM of the bottom (D127/K131) of the αHL nanopore without returning (see Figure 4.4). Therefore, the translocation time corresponds to the time elapsed between these two moments. To perform this calculation, we tracked the COM position of the bases, the constriction (E111/K147), and the bottom (D127/K131) rings of the αHL nanopore in the z-direction during each simulation and calculated the average translocation time per base and the standard error of the mean (SEM) for different systems by taking into account all translocation events in both replicas of each system. As a result of our CG simulations, we observed a total of 1024 bases translocated over 24 μs. In the next section, we will detail the translocation time per base for each case studied.

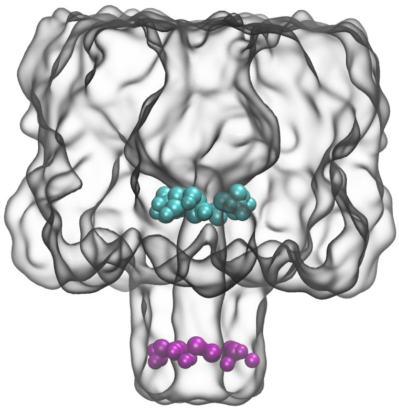


Figure 4.4: Surface representation of coarse-grained αHL nanopore. Residues of the central constriction (E111/K147), and bottom (D127/K131) are shown in cyan and purple respectively.

4.2.1 Translocation time distribution for 3' oriented ssDNA.

First, we measured the translocation time per base for only 3' oriented ssDNA molecules and constructed a histogram on a log binning to facilitate comparison with the experimental data. Figure 4.5 illustrates the distribution of translocation times per base for the ensemble of 3' end CG cf-SMD simulations detailed in Table 4.2. Our set of 20 simulations covers 864 base translocations, showcasing a wide range in translocation times per base, spanning from 2 to 1200 ns. The mean translocation time per base is 124 ns with a standard error of 6 ns. This average translocation time per base using CG cf-SMD is approximately one order of magnitude smaller than experimental results for poly(dA) with the αHL nanopore at 120 mV, as shown in Figure3.2. This discrepancy is attributed to the higher pulling forces applied to the ssDNA molecule in our simulations, equivalent to 660 pN in contrast to the experimental forces, which are estimated to be around 20 pN [15, 16, 181, 27]. However, our forces align with other all-atom (AA) SMD simulations [134, 133, 135], which we will detail in the coming paragraph.

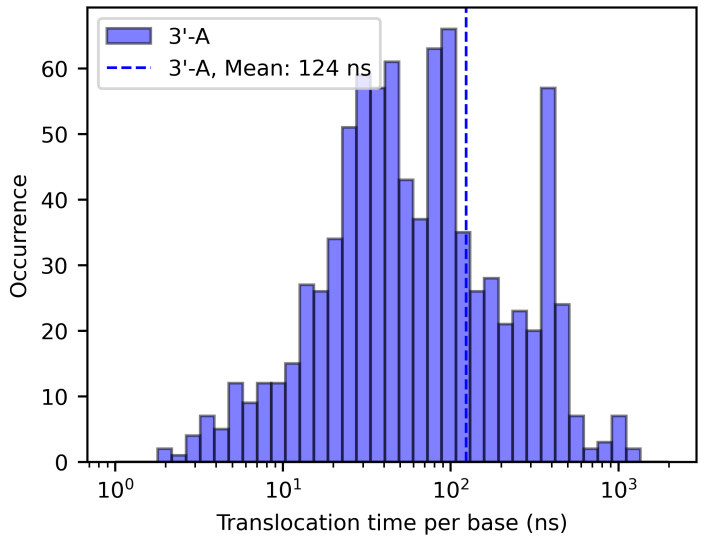


Figure 4.5: Distribution of translocation time per base of the ensemble of all CG cf-SMD simulations of charged ssDNA pulled from 3' end, on logarithmic time scale. The average of the translocation time per base is shown in red dashed lines.

Previous all-atom (AA) MD simulations by Aksimentiev *et al.* noted that in most of their MD simulations, the 20 base-length ssDNA translocated within approximately 50 ns under a field strength of 1.4 V, albeit through a synthetic pore [126]. Similarly, Wells *et al.* reported extremely fast translocation times, 1000 times faster than experimental results, under steering forces corresponding to 4.8 V of applied potential through αHL nanopore [134]. Guy *et al.* observed average times per base ranging from 0.7 to 4.0 ns under a 300 mV electric field through αHL nanopore [137]. Haynes *et al.* reported an average translocation time per base of 0.62 ns under a field strength of 0.2 V/nm through hydrophobic nanopores [182]. Rattu *et al.* observed translocation of 12-nucleotide poly(dA) within approximately 70 ns, yielding an average translocation time per base of 5.83 ns through CsgF-CsgG, a protein nanopore, using constant-velocity SMD (cv-SMD) simulations [132]. These studies, however, did not analyze the distribution of translocation times, likely due to computational constraints associated with longer simulations using all-atom force fields. In contrast, our CG-SMD simulations reveal a wide distribution of translocation times per base, consistent with prior CG Langevin dynamics simulations [183] and experimental studies [13, 15, 59].

Figure 4.6 presents the translocation time per molecule plotted against the number of nucleotides of the translocating ssDNA molecule. Analysis across different ssDNA lengths indicates that the translocation time tends to increase with the length of the ssDNA molecule. This observation aligns with previous experimental studies [13, 16, 62] and theoretical work [184]. However, it is important to note that not all studies have observed a linear relationship for varying ssDNA lengths. For example, Meller *et al.* [62] reported that when the DNA length is shorter than the pore's stem length, the dependence of translocation time on DNA length becomes non-linear, as explained in Chapter 1 in Figure 1.7.

A common result across these studies is that the length of the DNA directly influences the translocation time. Longer DNA chains tend to take more time to translocate and

exhibit greater variability in translocation times. This observation can be attributed to the increased likelihood of interactions between the DNA and the nanopore, as well as the possibility of secondary structure formation during translocation [15, 185]. In our cf-SMD simulations, this variability is also observed, as shown in Figure 4.6. Although the trend is subtle, as the ssDNA length increases, the SEM generally shows a tendency to increase, indicating greater variability in translocation times for longer sequences. This is particularly noticeable when comparing shorter lengths (10, 11, and 12 nucleotides) with longer lengths, where the standard errors become larger.

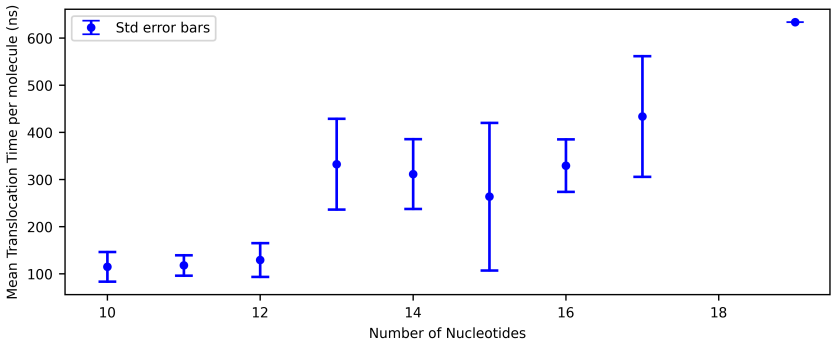


Figure 4.6: Translocation time per molecule as a function of the ssDNA length in cf-SMD simulations. The average translocation time per molecule is given for each system, and the standard error is shown as error bars. No data is available for 18-nucleotide poly(dA), and only one total translocation event was observed for 19-nucleotide poly(dA).

The absence of any ssDNA translocation for 18-nucleotide poly(dA) and differences in the ssDNA translocations between various events across replicates, such as the lack of translocation events for 15-nucleotide and 19-nucleotide poly(dA) in one of the replicates and the differences observed in total number of the translocation events (see Table 4.2), underscores the random nature of translocation events and the wide distribution of the translocation time observed. Halted ssDNA molecules, when the ssDNA translocation is stopped, contribute to the tail of the translocation time distribution, a phenomenon also observed experimentally. Longer simulations may capture the tail of this distribution of translocation times.

Moreover, these differences in translocation time per base/molecule between the replicates of the same systems (see Table 4.2), akin to findings from previous experimental studies [13, 61, 14, 15], also emphasize the necessity of conducting multiple MD simulations for robust statistical analysis. This need is effectively addressed by CG simulations compared to previous AA simulations, as these simulations require more resources.

4.2.2 Influence of ssDNA orientation on translocation time.

We continued with the analysis of the influence of molecular orientation on the translocation time by comparing CG cf-SMD simulations for 5'-A_{16} and 3'-A_{16} ssDNA molecules. Previous experimental studies [13, 63, 59] have demonstrated that the orientation during translocation directly influences the translocation time. However, experimental work does not reveal the mechanisms behind this behavior and therefore cannot explain why the 5' end-oriented ssDNA translocates slower. This is where the advantage of MD studies

becomes apparent. Previous all-atom MD studies [63, 134] confirmed this slower translocation for the 5' end and proposed a tilting mechanism which we will detail in Section 4.4. To the best of our knowledge, no CG study has ever examined this phenomenon.

We chose to study this phenomenon with 16 base-length ssDNA, which fits the αHL nanopore from the cap to the exit. This length is neither too short to translocate too quickly nor too long to interact with the cap and halt the translocation. The results showed a significant difference in translocation events, with 87 bases translocating from the 3' end compared to 22 bases from the 5' end during the same simulation time, 2 μs (see Table 4.2). Figure 4.7 presents the distribution of translocation times per base for both orientations. The average translocation time per base was 146$\pm$12 ns for the 3'-A_{16} system and 334$\pm$46 ns for the 5'-A_{16} system, resulting in a ratio between the translocation times of the 5' end and its 3' counterpart of 2.28, indicating a faster translocation for the 3' oriented ssDNA. Despite the fourfold increase in the number of translocated bases in the 3'-A_{16} system, this does not directly translate to a proportional decrease in translocation time per base due to differences in the time distributions of both orientations (see Figure 4.7). Our CG SMD results align closely with experimental findings from studies on αHL nanopores with poly$(dA)_{50}$ molecules [13, 59, 63]. Our experimental data, based on over 12,000 recorded events, showed that the ratio of translocation times between the 5' end and the 3' end of ssDNA was 2.22 (see Chapter 3 and Figure 4.5 for details). The findings of Muzard *et al.* reported a similar ratio of 2.07 [59]. Both values are in excellent agreement with our simulation results. In contrast, Wells *et al.*, using all-atom MD simulations, observed a much lower ratio of 1.16 for the translocation times of molecules oriented at the 5' and 3' ends [134], which deviates more significantly from both experimental data and our CG results.

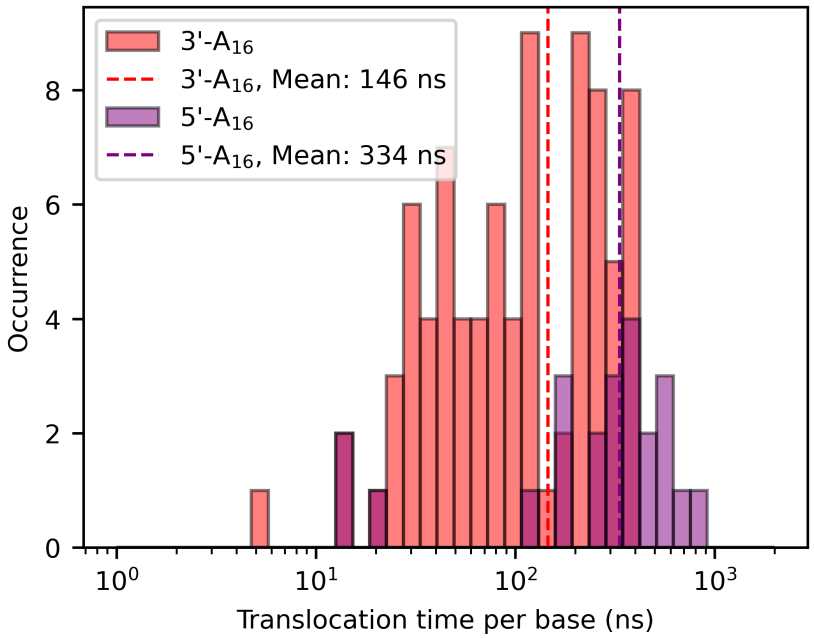


Figure 4.7: Distribution of translocation time per base of ssDNA molecules of 16 nucleotide length for both orientations with logarithmic binning representation. The average translocation times per base for 3' and 5' strand orientations are indicated by red and purple dashed lines, respectively.

These findings highlight the MARTINI CG model's ability to accurately replicate the effect of ssDNA orientation on translocation time. The initial orientation's influence on translocation dynamics will be further examined in Section 4.4.

4.2.3 Influence of ssDNA charges on translocation time

As a final step in the translocation time studies, we aimed to investigate the effect of electrostatic interactions between the nanopore and the translocating ssDNA molecule from a different perspective. Instead of using a neutral αHL nanopore, we created a neutral ssDNA molecule comprising 16 nucleotides (3'-A_{16}N) by modifying the MARTINI force field parameters for DNA. This modification neutralized the negatively charged phosphate groups present in ssDNA. We then compared the translocation time of this neutral ssDNA molecule with that of a charged ssDNA molecule (referred to as "charged" in this section for comparison) through the αHL nanopore. This approach allowed us to directly observe the effect of the negatively charged phosphate groups on the translocation velocity through αHL.

Our CG cf-SMD simulations comparing the translocation of 3'-A_{16} with its neutral counterpart, 3'-A_{16}N, revealed notable differences in the number of translocated bases along with the translocation times per base. We observed a total of 138 base translocations for the neutral 3'-A_{16}N across both replicas, compared to 87 for the charged 3'-A_{16} (see Table 4.2). The distribution of translocation times per base for both systems is shown in Figure 4.8. The average translocation time per base is significantly shorter for the neutral ssDNA molecule, at 84 ± 8 ns, compared to 146 ± 12 ns for the charged ssDNA.

The faster translocation time for the neutral ssDNA molecules indicates that the electrostatic interactions between the negatively charged phosphate groups of the ssDNA and the nanopore are responsible for slowing down the translocation speed, a topic that will be explored further in Section 4.3.

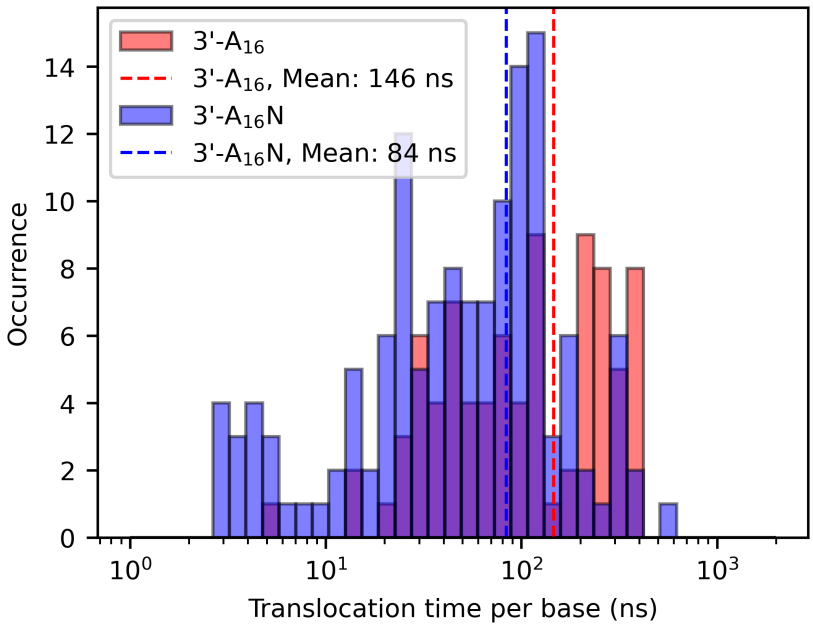


Figure 4.8: Distribution of translocation time per base of ssDNA molecules of 16 nucleotide length for both neutral and charged systems with logarithmic binning representation. The average translocation times per base for the neutral and charged ssDNA molecules are indicated by blue and red dashed lines, respectively.

4.3 DNA-nanopore interactions during translocation

In order to understand the reasons behind the observed differences in translocation times between charged and neutral ssDNA, this section analyzes the interactions between the DNA strand and the inner wall of the αHL stem. The αHL structure features two lysine residues on each of its seven chains, positioned at both ends of the stem: the K147 ring at the pore constriction and the K131 ring at the bottom of the pore. Existing evidence from both experimental studies [186, 18, 19, 105] and computational simulations [136, 137, 130, 133, 135] suggests that these positively charged lysine residues are responsible for attractive electrostatic interactions with the negatively charged ssDNA molecule.

In contrast, the rest of the inner stem is composed of neutral residues, which are hypothesized to have weaker attractive interactions with the DNA, and negatively charged residues, which have repulsive interactions with the negatively charged DNA residues. In the MARTINI force field, these weaker and repulsive interactions can be quantified based on the types and energy levels of the MARTINI beads used to represent these residues, as detailed in Table 2.1. For instance, neutral beads have minimal interactions with other types of beads, while negatively charged beads can repel the negatively charged phosphate group beads.

To quantify the ssDNA-pore interactions, we generated contact maps throughout the MD simulations by analyzing the number of contacts between the beads of the ssDNA molecule and those on the inner surface of the pore. The `mindist` command in GROMACS, with the `-group` option, was used for this analysis. The `-group` option allows multiple beads to be treated as a single group, ensuring that if any bead within the group is in contact with the ssDNA, the entire group is counted as having made contact. A contact is considered to exist if the center of mass (COM) of two beads are within 6 Å of each other.

We focused on the amino acid residues on each of the seven chains pointing towards the lumen in the stem of the αHL nanopore: E111/K147, M113/T145, T115/G143, T117/S141, G119/N129, N121/G137, N123/L135, T125/G133, and D127/K131. An in-house written Python code was used to calculate the cumulative contacts between the ssDNA molecule and each ring of these amino acids every 20 ns and to visualize them on a contact map.

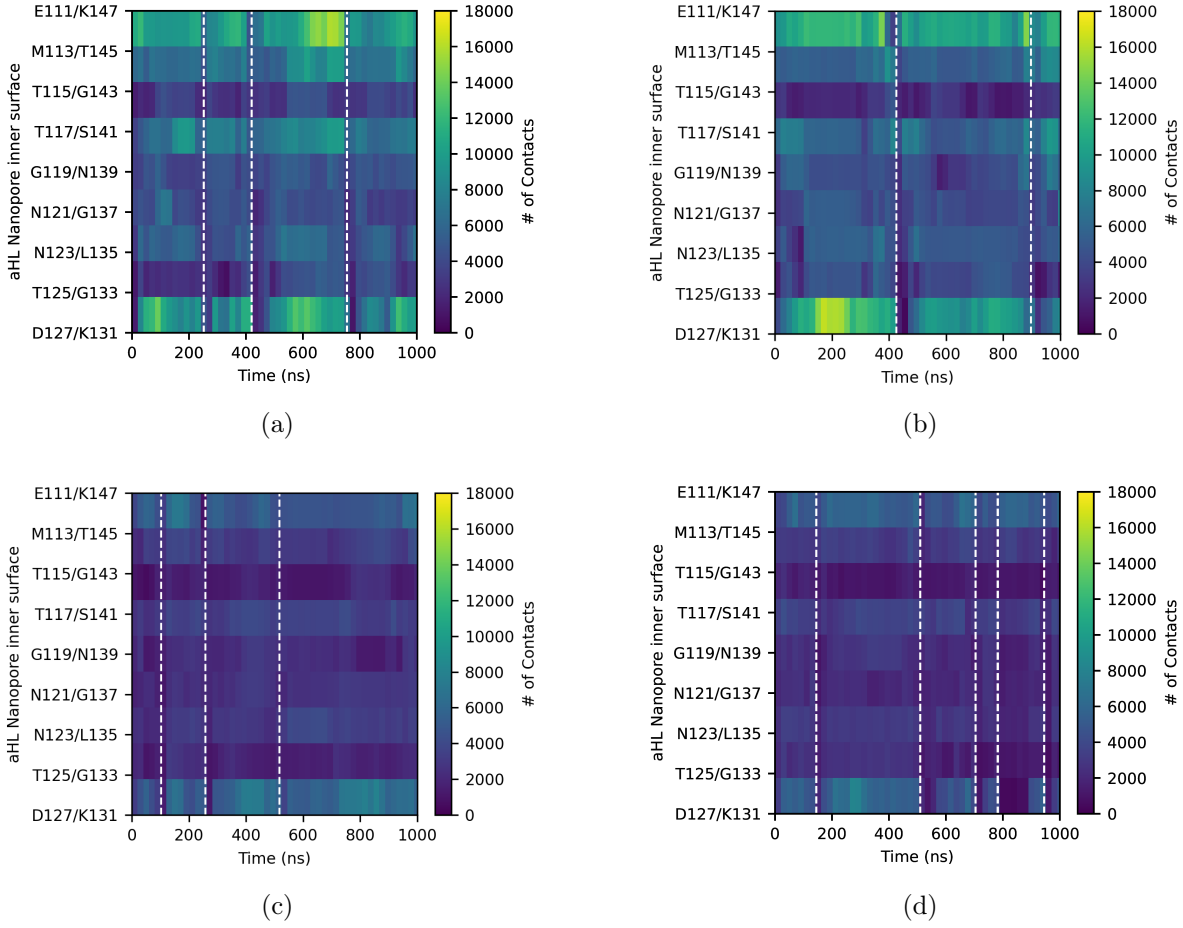


Figure 4.9: Contact maps of (a) 3'-A_{16}-1 (b) 3'-A_{16}-1 (c) 3'-A_{16}N-1 and (d) 3'-A_{16}N-2 with each ring in the stem during 1 μs MD simulation as a function of time. Dashed lines show the end of each translocation event. The cumulative contacts every 20 ns are given with a color code for visualisation.

Figure 4.9a and 4.9b display the contact maps for both charged ssDNA molecules of

16 nucleotides length during SMD simulations over time. Each row corresponds to an inner wall ring within the stem of the nanopore. The dashed lines mark the completion of the ssDNA molecule's translocation process. Notably, during a complete translocation event, the constriction, E111/K147, exhibits more contact with each base of the ssDNA than any other ring in the stem. This increased number of contacts can be attributed to the constriction being the narrowest part of the αHL nanopore. The bottom part of the nanopore, although having the widest radius (mean radius provided in Table 4.3), also shows numerous contacts, especially when the DNA is charged. The explication for that observation would be the attractive interactions between the phosphate charges and the positively charged lysine residues at position 131.

Table 4.3: Mean radius along with the standard error of constriction (E111/K147) and bottom (D127/K131) rings during ssDNA translocation simulations of 3'-A_{16} and 3'-A_{16}N systems.

System	Ring	Mean radius in nm	
		sim. #1	sim. #2
3'-A_{16}	E111/K147	1.04 ± 0.02	1.03 ± 0.02
	D127/K131	1.22 ± 0.03	1.22 ± 0.04
3'-A_{16}N	E111/K147	1.05 ± 0.02	1.06 ± 0.02
	D127/K131	1.22 ± 0.03	1.26 ± 0.03

Figures 4.9c and 4.9d display the contact maps for the 3'-A_{16}N simulations. These maps indicate that, although the translocation of the neutralized ssDNA molecule is occasionally halted, it forms fewer contacts with the inner wall of the stem compared to the charged ssDNA molecule. Similar to the charged ssDNA simulations, the highest number of contacts forms at the constriction and bottom rings of the stem, though the magnitude is much lower, as detailed in Table 4.4. Since we eliminated the effect of electrostatics by neutralizing the ssDNA molecule, the contacts with the constriction can be explained by size limitations, which leave little room for the ssDNA and result in contacts. The reason behind the contacts with the bottom can be attributed to LJ interactions. The charged particles, represented by Q0-type beads in the MARTINI force field, are present at the constriction and the bottom of the stem. These Q0-type beads have stronger interaction potentials compared to other types of MARTINI beads. This means that the Q0-type beads interact more strongly with any other beads in the system. (See Table 2.1 for further details on the interaction levels.)

Table 4.4: Number of contacts between ssDNA of 3'-A_{16} and of 3'-A_{16}N systems with constriction (E111/K147) and bottom (D127/K131) rings in the stem during complete translocation events for each replica.

System	sim. #	Ring	T#1	T#2	T#3	T#4	T#5
3'-A_{16}	1	E111/K147	126,933	75,569	192,116		
		D127/K131	115,882	67,343	155,697		
	2	E111/K147	226,842	247,202			
		D127/K131	246,282	197,655			
3'-A_{16}N	1	E111/K147	24,701	40,546	64,415		
		D127/K131	20,954	39,765	76,030		
	2	E111/K147	34,006	100,186	46,795	17,658	43,995
		D127/K131	34,803	105,316	35,107	13,023	17,111

The higher frequency of translocation events observed with neutral ssDNA (see Table 4.2) suggests that electrostatic interactions between the ssDNA and the inner wall of the stem contribute significantly to slowing down DNA passage. This observation is consistent with previous experimental studies showing that mutations converting neutral residues to positively charged ones within the pore stem reduce ssDNA translocation speed [78, 79].

As a result of this section, we concluded that although the CG model simplifies electrostatic interactions by using only $1e$ and $-1e$ charged residues, it still effectively captures the electrostatic contributions to the ssDNA translocation process. This is observed through the lower number of contacts with the pore stem for neutral ssDNA molecules, particularly at the constriction (E111/K147), leading to a smaller translocation time per base. However, it is important to note that the physical constriction region of the αHL nanopore also presents a barrier to translocation through steric interactions. This is demonstrated by instances where the translocation of neutral ssDNA molecules is halted, specifically in Fig. 4.9c after the third translocation.

4.4 DNA tilting angles during translocation: 3' vs 5'

In Section 4.2, our observations indicate that 3'-oriented ssDNA molecules translocate faster through the nanopore compared to 5'-oriented molecules. This finding is consistent with our experimental data (see Figure 3.2) and aligns with previous experimental results [63, 59] as well as MD simulation studies [134].

Mathé *et al.* [63] provided an explanation for this difference by examining the tilt angle of ssDNA molecules confined within a narrow cylinder using all-atom (AA) simulations. They discovered that as the cylinder's diameter decreases around the DNA molecule, the DNA bases tend to tilt towards the 5' end. This upward tilt of the bases for 3'-oriented molecules in the narrowest part of the pore facilitates their translocation, while the downward tilt of bases for 5'-oriented molecules hinders it. However, Mathé *et al.* focused primarily on the conformational aspects of the ssDNA inside the pore rather than the dynamic translocation process, as detailed in Chapter 1.

Building on this work, a subsequent AA study by Wells *et al.* using grid-SMD [134] confirmed the impact of base tilting on the differences in translocation velocities between

5' and 3' orientations. By observing DNA conformations during translocation in both orientations, they reinforced the conclusion that base tilting toward the 5'end slows down the translocation for 5' end oriented ssDNA molecules, even though they did not quantitatively analyze the tilt angle values throughout their simulations.

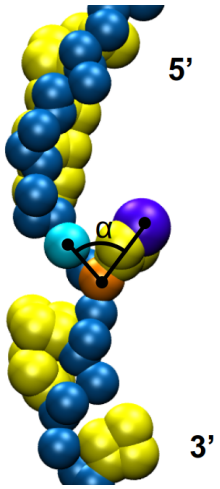


Figure 4.10: Representation of the tilt angle (α) defined between the center of mass of BB1, BB3, and SC3 beads of the ssDNA molecule, represented in cyan, orange, and purple beads, respectively. 5' and 3' extremities are indicated in the illustration.

Leveraging these studies, we investigated the orientation-dependent tilting of the bases and its effect during translocation on 16-nucleotide length ssDNA molecules. We calculated the tilt angles between the ssDNA bases and the backbone as a function of the base position inside the pore during their translocation throughout the simulations. This angle was determined by measuring the angles between the BB1, BB3, and SC3 beads of the ssDNA molecule for all bases, as illustrated in Figure 4.10, excluding those at both extremities. We excluded nucleotides subjected to pulling forces and those at the opposite ends, as they lack a consecutive base. The GROMACS `angle` command was used to compute the relevant angles, and the `traj` command was used to track the z-direction positions of the bases. To facilitate the visualization of the results, the coordinates were translated so that the constriction point became the new origin position inside the simulation box. Running averages over a 2 ns time window were computed, and the data were further averaged over defined intervals of 0.2 nm in the z-direction. The average and SEM of the translocation angles of the bases within each 0.2 nm interval were calculated for both replica simulations and are represented on the same plot. Figure 4.11 presents the evolution of these tilt angles for the 3'-A_{16} and 5'-A_{16} systems.

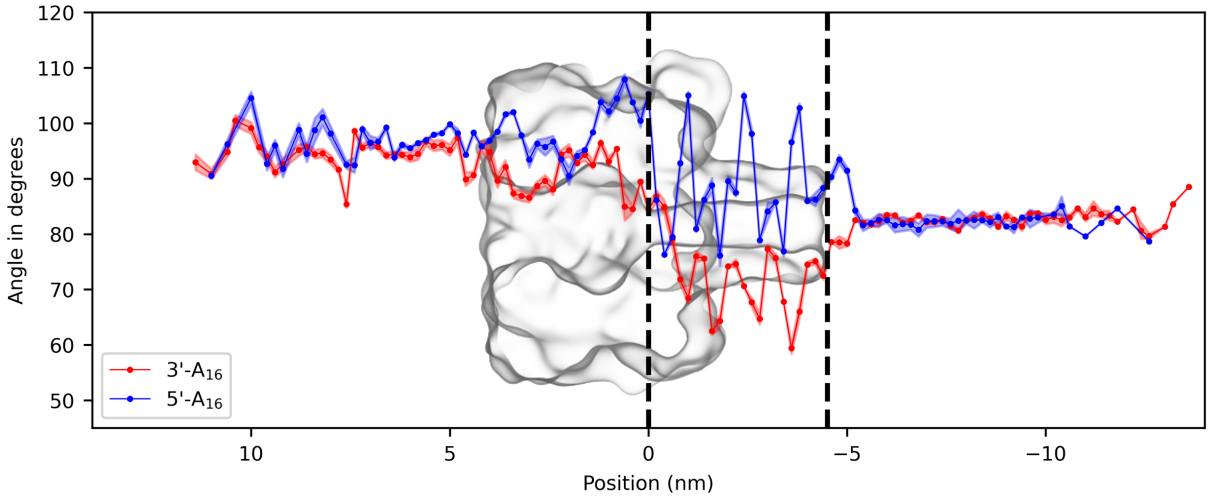


Figure 4.11: Evolution of the tilt angles inside the nanopore during ssDNA translocation for both orientations. Average tilt angles of DNA bases relative to the backbone for 3' end and 5' end oriented 16-nucleotide poly(dA) molecules are given as red and blue circles as a function of positions in the z-direction inside the simulation box. The displayed values represent the mean tilt angle for the 14 bases (excluding bases at both extremities) across all translocation events in the two simulation replicas, accompanied by their standard error. The black dashed lines denote the constriction and bottom of the stem.

The tilt angles for the 3'-A_{16} and 5'-A_{16} molecules exhibit similar values before they enter the pore, averaging 94.6° for the 3' end and 96.6° for the 5' end and after exiting the pore, the angles average 82.7° for the 3' end and 82.5° for the 5' end. Notably, the tilt angle before pore entry closely resembles the average value calculated for a ssDNA molecule in a solvent without constraints (96.8°, see Table 4.5), indicating a weaker force on the DNA bases prior to pore entry.

Upon exiting the nanopore, the tilt angles align closely with an average of 83.3° for the ssDNA pulled by the 3'-end without a nanopore (see Table 4.5). In contrast, for ssDNA pulled by the 5'-end in the absence of the nanopore, the average tilt angle is about 93.3° (Table 4.5). However, it is essential to differentiate between pulling ssDNA in the absence of the nanopore and pulling it during translocation through the nanopore. When ssDNA is pulled in the absence of the nanopore, it experiences no confinement, and its behavior is governed by the applied force and the interactions with the solvent, primarily friction. In contrast, when ssDNA is pulled during translocation through the nanopore, the nanopore's geometry introduces additional constraints. The presence of the nanopore slows down the ssDNA molecule due to the spatial restrictions imposed by the pore, more specifically by the stem, which affects the conformational changes of the ssDNA, as we observed in Figure 4.11. Specifically, the nanopore's constriction region can impede the movement of the ssDNA, causing it to adopt different conformations and altering the tilt angle compared to when it is pulled freely without the nanopore. After the nanopore exit, the ssDNA bases are predominantly influenced by the pulling force, as their interactions with the nanopore are no longer present. Post-exit oscillations are smaller because the pulling force is stronger on the exiting bases while other bases remain within the stem. As will be discussed in Section 4.5, similar conformational behaviours were observed during ssDNA stretching at the nanopore exit.

Table 4.5: Average tilt angles along with the standard errors of DNA bases relative to the backbone for a 16-nucleotide poly(dA) molecule under different conditions.

System	Average Tilt Angle (°)
F'-A_{16}	96.8 ± 0.9
3'-A_{16} (no pore)	83.3 ± 1.4
5'-A_{16} (no pore)	93.0 ± 1.7

System names are mentioned as before, F stands for free, meaning ssDNA in solution without any steering forces applied. Simulations in the absence of the nanopore are mentioned in parenthesis.

Conversely, as shown in Figure 4.11, the tilt angles inside the nanopore, particularly in the stem, differ significantly depending on the ssDNA molecule's orientation. In the vestibule, where the molecule starts experiencing interactions with the pore wall, the tilt angle starts to change, decreasing slightly for the 3'-A_{16} and increasing for the 5'-A_{16}. Inside the stem, the tilt angle for the 3'-A_{16} molecule decreases sharply and oscillates, reaching a minimum of about 60° to pass through the narrowest part of the pore. For the 5'-A_{16} molecule, the tilt angle shows more pronounced oscillations, increasing to nearly 110° but never reaching the 120° complementary minimum tilt angle observed for the 3'-A_{16} molecule. Additionally, the tilt angle frequently reverts to around 85°, which is less favorable for translocation as the bases are nearly orthogonal to the backbone. To ensure this observation is independent of DNA sequence length, we also analyzed the tilt angles in 3'-A_{19} and 5'-A_{19} simulations. As shown in Figure 4.12, the same phenomenon was observed.

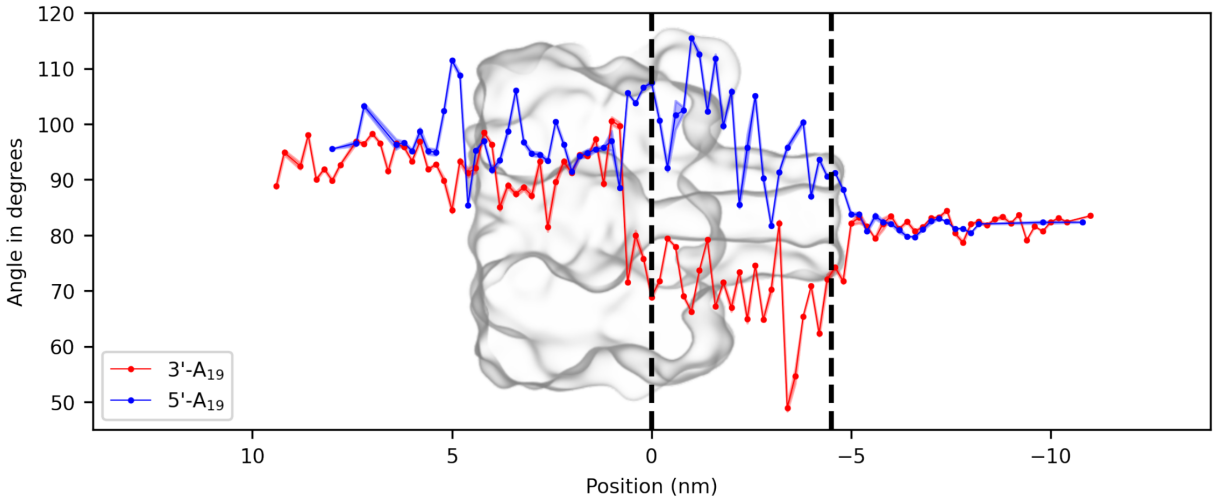


Figure 4.12: Evolution of the tilt angles inside the nanopore during ssDNA translocation for both orientations. Average tilt angles of DNA bases relative to the backbone for 3' end and 5' end oriented 19-nucleotide poly(dA) molecules are given as red and blue circles as a function of positions in the z-direction inside the simulation box. The displayed values represent the mean tilt angle for the 17 bases (excluding bases at both extremities) across all translocation events in the two simulation replicas, accompanied by their standard error. The black dashed lines denote the constriction and bottom of the stem. 5'-A_{19}-1 simulation was extended for another 200 ns to achieve one complete translocation for this analysis.

In conclusion, consistent with previous experimental [63, 59] and all-atom MD studies [63, 134], our results indicate that differences in tilt angles are responsible for the slower translocation of 5' end-oriented ssDNA. In addition to these aforementioned MD studies, we investigated the tilt angle phenomenon during the translocation process through extensive CG-SMD simulations for over 140 base translocations across two systems of different lengths of ssDNA molecules. We observed that the base tilting required to navigate the ssDNA through the narrowest part of the pore is less favorable in the 5' end orientation. These insights provide a more comprehensive understanding of how molecular orientation affects ssDNA translocation dynamics through the αHL nanopore.

4.5 Conformations of ssDNA during translocation

Pulling on the ssDNA by one end unfolds the molecule, facilitating its entry into the nanopore. This unfolding eventually stretches the inter-base links if the translocation is halted and the ssDNA is held within the pore. Consequently, the degree of stretching serves as a useful measure to understand if the chain is interacting with the pore and hindering its translocation.

To investigate the conformational changes of ssDNA during its translocation, we monitored and analyzed the positions of the bases along the z-axis and the inter-base distances throughout the MD simulations. Figure 4.13 illustrates the positions of the ssDNA bases

along the z-axis for the 3'-A_{16}-1 simulation, with an inset showing the base positions in the vestibule. Position graphs from other simulations, which we will discuss in this section, are available in Figure 4.14. As the ssDNA is pulled through the nanopore via cf-SMD, the molecule is initially dragged through the solvent, taking an elongated conformation. The sharp transition at the stem entrance distinctly shows that the first 3 to 4 bases of the 3' end enter the pore stem rapidly. The rest of the molecule adopts a less extended conformation while waiting to pass through the pore constriction for its translocation, as indicated by the crossover of the base position curves in the inset of Figure 4.13.

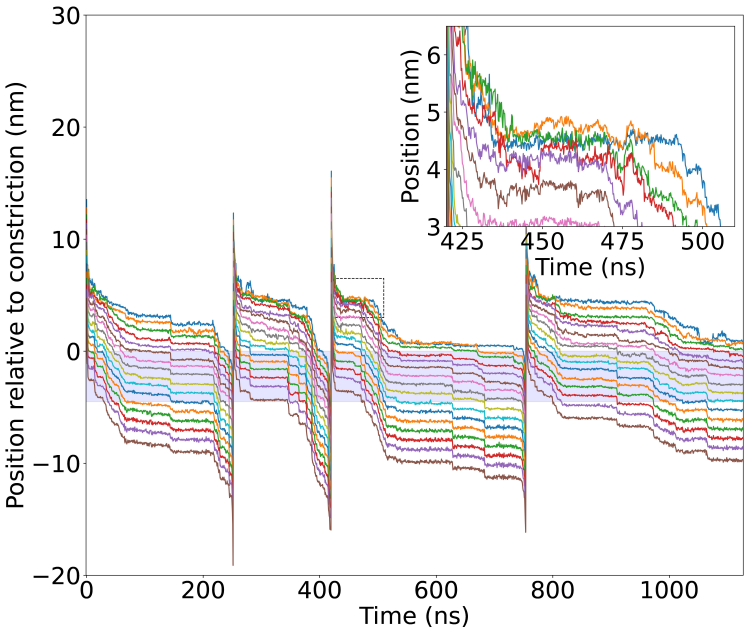
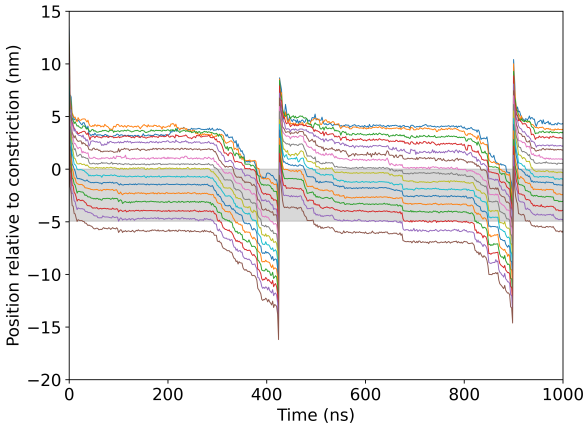
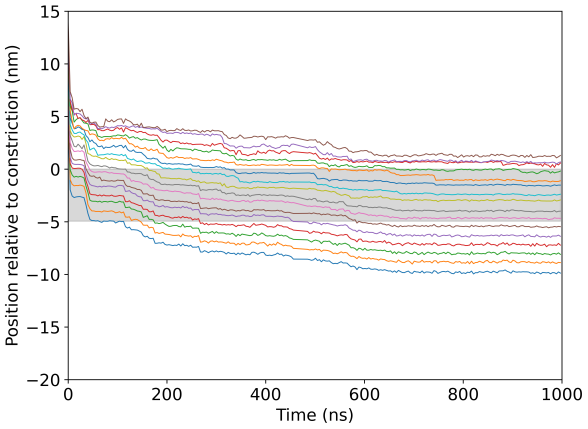


Figure 4.13: Positions in the z-coordinate of the bases of 3'-A_{16}-1 over time. The stem is depicted as a blue-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147). The inset focuses on the positions of bases in the z-direction at the entrance of the nanopore before the constriction ring during the third translocation event.

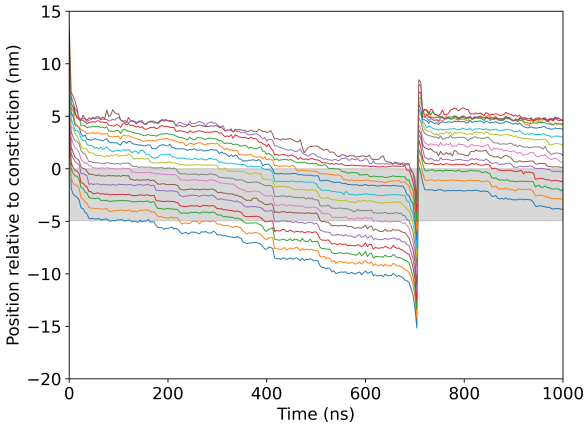
In contrast, the position curves evolve in a parallel manner once the nucleotides enter the stem, indicating that the nucleotides align with the z-axis of the pore. This behavior persists even after exiting the pore, while some bases remain inside the stem. Once the ssDNA has completely exited the pore, it is drawn through the solvent again until the next translocation event occurs.



(a) Positions in the z-coordinate of the bases of 3'-A_{16}-2 over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147).



(b) Positions in the z-coordinate of the bases of 5'-A_{16}-1 over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147).

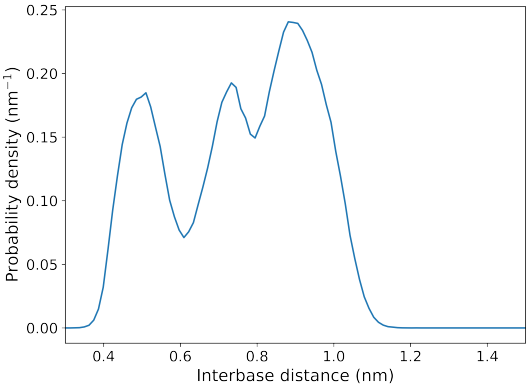


(c) Positions in the z-coordinate of the bases of 5'-A_{16}-2 over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147).

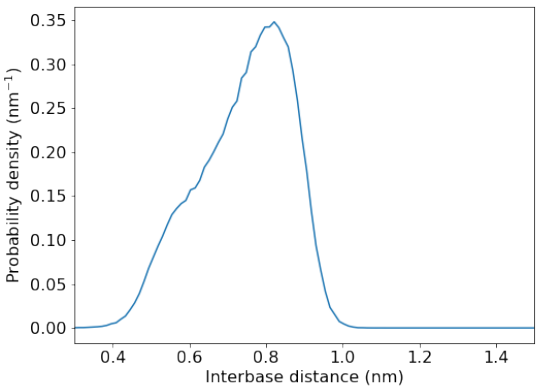
(d) Legend for the plots, each color indicating a base number

Figure 4.14: Evolution of the base positions in z-direction of 3'-A_{16} systems.

As a second step of this conformational study, we measured the inter-base distances along the molecule as Euclidean distances between the COM of each pair of consecutive bases. These inter-base distance measurements were also performed on a free DNA molecule in solution (in the absence of the αHL nanopore and without any applied force) to serve as a reference (see Figure 4.15a). The inter-base distances of the free ssDNA showed three primary values with nearly equal probabilities of occurring at: 0.50 ± 0.08 nm, 0.71 ± 0.08 nm, and 0.91 ± 0.12 nm.



(a) Probability density of inter-base distance for a A_{16} ssDNA free in solution. We observe 3 main conformations at 0.50 ± 0.08 nm, 0.71 ± 0.08 nm and 0.91 ± 0.12, which are coexisting along the chain.



(b) Probability density of inter-base distance for a 3'-A_{16} ssDNA pulled by the SMD force in solution (in the absence of the pore). We observe one main conformation at 0.88 ± 0.11 nm and a minor one at 0.55 ± 0.15 nm.

Figure 4.15: Inter-base distance observed in two control simulations: ssDNA free in solution and ssDNA dragged in solution by the pulling force from 3' end.

During the translocation process for 3'-A_{16} simulations, the inter-base distances were analyzed and represented in Figure 4.16 as 2D histograms against the position along the pore axis, highlighting the stem region in red, as explained above. The position of an inter-base was defined as the mean position of the two constituting bases in the z-direction relative to the constriction. To calculate the 2D histograms, the z-axis was divided into intervals of 0.2 nm. Within each 0.2 nm interval, the distribution of inter-base distances was computed with a bin size of 0.02 nm along the inter-base distance axis. All these distributions were then concatenated and normalized to obtain the probability density. The final 2D histogram, as seen in Figure 4.16, provides the probability density of inter-base distances at various positions along the z-axis. On either side of the 2D histogram, the probability density of inter-base distances is shown before entry (left panel) and after exit (right panel) from the stem, obtained by integrating the 2D probability densities for $z < 0$ and $z > z_{\text{bottom}}$, respectively. The first inter-base distance, which includes the base to which the SMD pulling force was applied, was excluded from the calculation since it is always more stretched than the rest of the molecule, as shown in Figure 4.17, with a mean inter-base distance of 1.16 ± 0.05 nm. The last inter-base distance was also excluded for consistency between both orientations.

Before the ssDNA enters the pore, two main populations of inter-base distances are observed at 0.55 ± 0.10 nm and 0.89 ± 0.12 nm. These values are very close to the first and third states observed in the free ssDNA molecule and the ssDNA pulled in solution (see Figure4.15a and 4.15b). This observation shows that one of the free ssDNA inter-base states is not accessible when the molecule is pulled during translocation. After exiting the pore, the inter-base distances converge to a single population at 0.93 ± 0.06 nm, corresponding to the most stretched conformation seen in both the free ssDNA and the ssDNA pulled in solution from the 3' end. The insertion and subsequent pulling of the ssDNA through the pore influence the equilibrium among the accessible states,

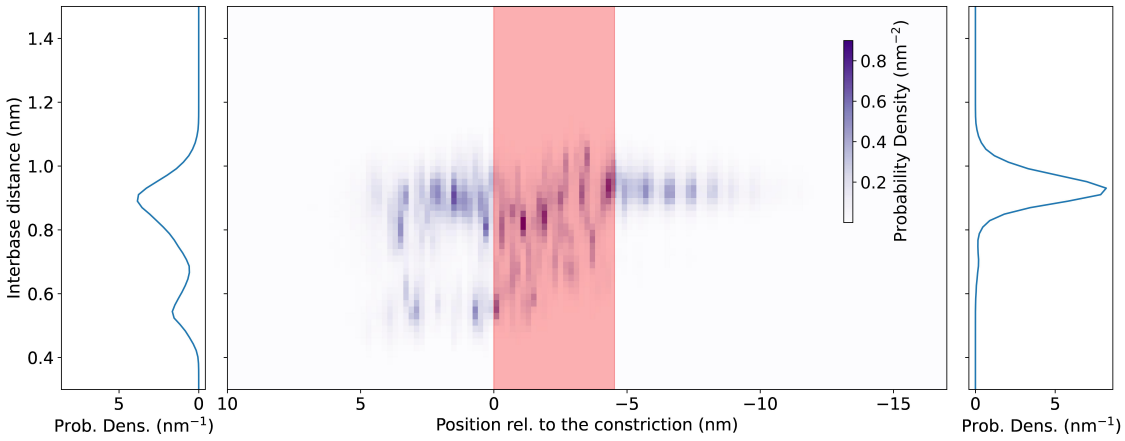


Figure 4.16: Inter-base distances for 3'-A_{16} simulations. The distance between two successive bases (inter-base distance) for all pairs except the first and last ones is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 representing the pore constriction ring (E111/K147). Left panel: Probability density of inter-base distance before the entry of the DNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: Probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm).

favoring a more stretched conformation once the molecule is extended out of the pore, as anticipated. Similar findings were obtained from the 5'-A_{16} simulations and are presented in Figure 4.18, proving that the phenomenon observed is independent of the orientation during the translocation.

Experimental measurements of the mean inter-base distance of a loaded ssDNA have been previously reported. Using a 10 pN force applied with optical tweezers [91] or a 160 mV voltage through the nanopore technique [66], the mean distance was found to be 0.42 nm. This value corresponds to the lowest inter-base distance observed in our simulations despite the much larger force applied. Previous AA-MD simulations indicated a mean inter-base distance of about 0.6 nm under a 300 mV applied voltage [137], which is closer to the one of the populations observed before entering the pore at 0.55 nm.

Our observations on inter-base distances demonstrate that the ssDNA exhibits greater freedom to fluctuate before entering the stem than after exiting the pore, consistent with findings regarding the tilt angles described in Section 4.3. Additionally, two potential configurations for the portion waiting to enter the stem were observed. In one configuration, the applied SMD force directs the ssDNA into the stem, stretching it into an elongated conformation with bases aligned along the pore's axis. In the alternative configuration, the ssDNA does not experience the steering force and can move freely within or outside the vestibule.

This phenomenon occurs due to the bases already trapped in the pore, resulting in less stretched pairs of bases in the vestibule and a more folded conformation for that part of the molecule. Upon exiting the pore, the ssDNA is subjected to the applied force,

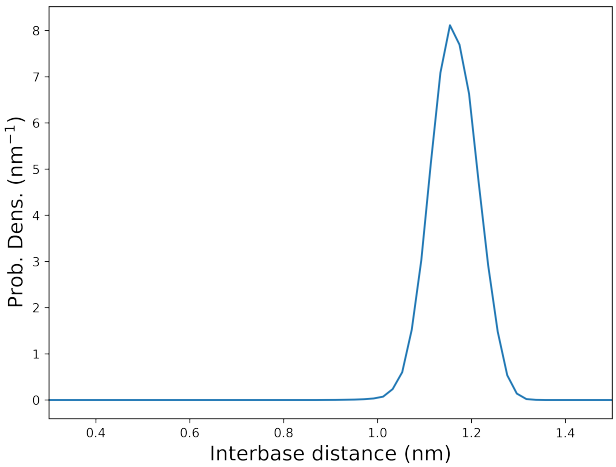


Figure 4.17: Inter-base distance of the first pair of bases for 3'-A_{16} simulations. As the SMD force is applied to the first base, this inter-base is overstretched.

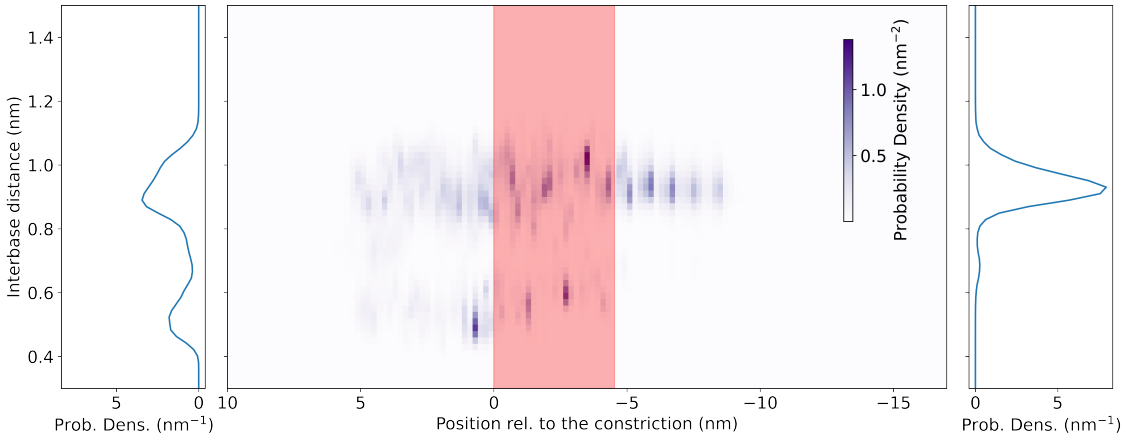


Figure 4.18: Inter-base distances for 5'-A_{16} simulations. The distance between two consecutive bases (inter-base distance) for all pairs except for the first and last one is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 on the z-axis corresponding to the pore constriction ring (E111/K147). Left panel: probability density of inter-base distance before the entry of the ssDNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm).

thus adopting a straight conformation with a distinct stretched inter-base distance. This behavior is difficult to observe in experimental studies due to the challenges of probing such fluctuations. However, conformational fluctuations have been noted in experimental studies under different conditions. For instance, an experimental study under applied voltage observed a DNA hairpin fluctuating at the pore entrance while a ssDNA over-

hang was threaded in the pore, experiencing a less effective pulling force [59], confirming our observations about the fluctuations depending on the position with respect to the nanopore.

4.6 Conclusion

This study demonstrates the robustness of the MARTINI CG model in exploring DNA translocation through αHL. Our simulations using cf-SMD effectively replicate key experimental findings, including the wide distribution of translocation times per base, the influence of DNA orientation on translocation time, and the role of electrostatic interactions between the nanopore and ssDNA in slowing down the translocation process.

The analysis of contact maps highlighted the impact of these electrostatic interactions, particularly at the constriction, between ssDNA and αHL. Additionally, our simulations provided insights into the effect of 3' vs. 5' orientation on translocation times, which are linked to the observed tilting angles of ssDNA nucleotides. The conformational changes of ssDNA during translocation were further illustrated through inter-base distance calculations. These findings are consistent with prior experimental and simulation studies, offering a comprehensive understanding of ssDNA translocation dynamics via αHL.

Our CG simulations revealed diverse characteristics in each translocation event, underscoring the necessity of analyzing multiple events to avoid misleading conclusions. This study emphasizes the importance of robust statistical analysis in both experimental and theoretical investigations. Consequently, our work validates the effectiveness of CG simulations as a complementary approach to the experimental work compared to more computationally expensive all-atom simulations.

Moving forward, we will extend our MARTINI CG modeling approach to investigate additional aspects of ssDNA translocation, such as the impact of ssDNA sequence and the effects of the steering force. These topics will be explored in the next chapter.

Chapter 5

Study of the Effect of Sequence and Pulling Force on Translocation Dynamics Using CG-SMD

Contents

5.1 CG-SMD System Preparations and Simulation Parameters . 126

5.1.1 Simulation Setup for ssDNA Molecules 126

5.1.2 Steered Molecular Dynamics Simulations 127

5.2 Influence of sequence on translocation time 127

5.2.1 Translocation time distributions of poly(dA) and poly(dC) molecules 128

5.2.2 DNA-pore interactions depending on the sequence 130

5.2.3 DNA Tilt Angle Differences Depending on the Sequence 132

5.2.4 Conformational differences depending on the sequence 135

5.3 Influence of Pulling Force on Translocation Time 136

5.3.1 Distributions of Translocation Time Per Base Under Varying Pulling Forces 137

5.3.2 The relationship between pulling force and translocation velocity 139

5.3.3 Influence of pulling force on ssDNA conformations 141

5.4 Conclusion 148

This chapter investigates the effect of nucleotide sequence and the magnitude of the pulling force on the translocation dynamics of single-stranded DNA (ssDNA) molecules through an α-hemolysin (αHL) nanopore embedded in a DPPC bilayer. We utilized coarse-grained (CG) constant-force Steered Molecular Dynamics (cf-SMD) simulations, similar to the approach used in Chapter 4. Our study first focuses on the translocation time differences between poly(dC) and poly(dA) ssDNA sequences under the force conditions used in Chapter 4 to elucidate how the intrinsic properties of DNA, such as sequence composition, influence translocation behavior.

In addition to exploring the sequence-dependent effects, we aimed to thoroughly understand how varying the magnitude of the pulling force impacts the translocation dynamics.

Therefore, we applied a range of pulling forces to both ssDNA molecules, poly(dA) and poly(dC), of the same length. Specifically, we investigated whether the relationship between the pulling force and the translocation rate aligns with theoretical predictions. Furthermore, by comparing our simulation results with existing experimental and theoretical studies, we seek to provide deeper insights into ssDNA translocation mechanisms.

In this Chapter 5, we will follow the same outline as in the previous Chapter 4. In section 5.1, we detail the preparation of our systems, when different from Chapter 4, before delving into the results. Section 5.2 presents the results of the influence of the sequence on translocation time and dynamics by performing analyses similar to those presented in Chapter 4 when exploring differences in translocation times, such as contact maps, tilt angle, and inter-base distance. Section 5.3 examines the effect of the applied pulling force on the translocation time and dynamics, as well as the relationship between the translocation velocity and the pulling force. We investigate the effect of a range of pulling forces on the conformational differences during the ssDNA translocation. Finally, in Section 5.4, we will conclude this second computational part of this thesis.

5.1 CG-SMD System Preparations and Simulation Parameters

This section describes the simulation setup and parameters used for the poly(dC) ssDNA-αHL nanopore-DPPC and poly(dA) ssDNA-αHL nanopore-DPPC systems. The methodology follows the procedures outlined in Section 4.1, with modifications specific to the use of poly(dC) ssDNA. Unlike the poly(dA) simulations, which involved the application of an electric field, we directly employed CG-SMD simulations for these studies.

5.1.1 Simulation Setup for ssDNA Molecules

ssDNA molecules consisting of 10, 13, and 16 nucleotides of poly(dC) were constructed by modifying double-stranded DNA (dsDNA) using the builder function in PyMol [178], followed by the deletion of one strand. The `martinize_dna.py` tool [179] (version 2.2) was then used to convert the ssDNA into a coarse-grained (CG) model, using the `-dnatype ss` option. To account for the phosphate group missing at the 5' end, a $-1e$ charged Q0 type MARTINI bead was manually added. The ssDNA molecule was positioned within a $25 \times 25 \times 25$ nm^3 simulation box and was oriented along the z-axis. Initial energy minimization was performed for 100 steps using the steepest descent algorithm, followed by a pulling process on both ends of the ssDNA with a 100 kJ mol^{-1} nm^{-1} force over 10 ps, using a 2 fs time step. The system was then solvated with approximately 130,000 PW water molecules and subjected to further minimization (max 50,000 steps) and equilibration over 50 ns with an incrementally increasing time step from 2 fs to 10 fs. Subsequently, 1M NaCl MARTINI ions were added using the `genion` tool in GROMACS. The system was minimized again before a final 20 ns equilibration in the NPT ensemble with positional restraints on the DNA backbone, applying a force constant of 1,000 kJ mol^{-1} nm^{-2}.

Table 5.1: Summary of ssDNA translocation simulations

System name	# of nucleotides	F[a]	Simulation time (μs)	# of total strand translocations		Total # of bases translocated	
				sim.#2	sim.#1	sim.#2	sim.#1
3'-C_{10}-400	10	400	1	18	16	183	160
3'-C_{13}-400	13	400	1	11	10	143	136
3'-C_{16}-400	16	400	1	6	3	104	50
3'-C_{10}-450	10	450	1	24	40	240	400
3'-C_{10}-350	10	350	2	10	12	100	121
3'-C_{10}-300	10	300	5	-[b]	9	-[b]	92
3'-C_{10}-250	10	250	5	2	4	21	41
3'-C_{10}-200	10	200	5	-[b]	-[b]	-[b]	-[a]
3'-A_{10}-450	10	450	1	10	8	310	221
3'-A_{10}-400[c]	10	400	1	10	8	102	85
3'-A_{10}-350	10	350	2	4	4	41	40
3'-A_{10}-300	10	300	5	7	-[b]	70	-[b]
3'-A_{10}-200	10	200	5	-[b]	-[b]	-[b]	-[a]

[a] Pulling Force (kJ mol^{-1} nm^{-1})
[b] None of the bases are translocated
[c] Data taken from Chapter 4

5.1.2 Steered Molecular Dynamics Simulations

Cf-SMD simulations were performed to investigate the translocation of ssDNA through the αHL nanopore. In these simulations, the 3' end of the ssDNA was designated as the pull group, while the central constriction of the αHL nanopore served as the reference group. The pulling force was applied along the z-axis, directing the ssDNA through the nanopore channel. For examining the effect of nucleotide sequence on translocation dynamics, a constant pulling force of 400 kJ mol^{-1} nm^{-1}, the same as the pulling force used in Chapter 4, was applied to the center of mass (COM) of the 3' end of poly(dC) molecules. To investigate the influence of pulling force on the translocation dynamics, we applied forces of 200, 250, 300, 350, 400, and 450 kJmol^{-1} nm^{-1} to the COM of the 3' end of 10-nucleotide-long poly(dC) and poly(dA) ssDNA strands. All simulations employed a time step of 10 fs to accurately capture the rapid dynamics of ssDNA movement under applied forces. In total, 8 MD systems were prepared for the poly(dC) and 4 MD systems for poly(dA), as summarized in Table 5.1. Each system was simulated in two replicas for a duration of at least 5 μs for the pulling forces of 200, 250 and 300 kJ mol^{-1} nm^{-1}, at least 2 μs for the pulling forces of 350 kJ mol^{-1} nm^{-1} and at least 1 μs for the pulling forces of 400 and 450 kJ mol^{-1} nm^{-1} to ensure adequate sampling and reproducibility of results under the NVT ensemble conditions.

5.2 Influence of sequence on translocation time

Similar to Section 4.2, we first investigated the translocation time per base by varying the length of the poly(dC) molecules—10, 13, and 16 nucleotides—and compared the

results with those obtained for poly(dA) molecules of the same lengths, as presented in the previous chapter. We chose these 3 lengths to ensure sufficient statistical data for comparison with poly(dA).

The translocation time was calculated in the same manner as described in Section 4.2. Over the course of our CG simulations, we observed a total of 776 bases translocated over 6 μs. In the following section, we will provide a detailed analysis of the translocation time per base for poly(dC) structures and compare these results with those obtained for poly(dA) molecules. We will then compare our results with the previous experimental and theoretical studies.

5.2.1 Translocation time distributions of poly(dA) and poly(dC) molecules

First, we measured the translocation times per base for poly(dC) molecules. To facilitate comparative analysis, we extracted the translocation times of poly(dA) molecules of the same length and constructed a histogram with both sequences, employing logarithmic binning. Figure 5.1 presents the distribution of translocation times for poly(dA) and poly(dC) molecules. The data were obtained from 6 independent SMD simulations for each ssDNA type, resulting in a total of 776 base translocations for poly(dC) and 313 base translocations for poly(dA).

Both ssDNA sequences exhibited a wide range of translocation times: 1 to 400 ns for poly(dC) and 3 to 1100 ns for poly(dA). Notably, the spread of translocation times was narrower for poly(dC) molecules. The broader distribution of translocation times observed for poly(dA) is consistent with prior experimental studies [14, 62] and 2D Langevin dynamics simulations [110]. Additionally, this observation aligns with the phenomenon that longer translocation times typically lead to wider distributions. Specifically, the average translocation time for poly(dA) was found to be 108 ns with a standard error of 8 ns, while poly(dC) exhibited a faster average translocation time of 48 ns with a standard error of 2 ns.

These findings are consistent with the well-documented experimental work that pyrimidines, such as cytosine, translocate faster through nanopores than purines, such as adenine, due to several factors: stronger electrostatic attractions between adenine-rich sequences and the positively charged residues within the nanopore [14, 62, 187], a greater potential for hydrogen bonding with the pore's amino acid side chains, and the inherent conformational rigidity of poly(dA) [62, 187]. Meller *et al.* [62] reported that poly(dC)$_{100}$ translocated approximately 2.75 times faster than poly(dA)$_{100}$, with average translocation times of 120 and 330 μs, respectively, using the αHL nanopore at 120 mV applied voltage.

Some Langevin dynamics simulations have also explored the sequence dependence of ssDNA translocation, investigating both homopolymers and heteropolymers [110, 111, 116]. Although these studies did not explicitly report translocation time ratios between poly(dC) and poly(dA), their simulation parameters were derived from experimental translocation time distributions that accurately reflected the observed experimental ratios. Specifically, Luo *et al.* [110] demonstrated that stronger Lenard-Jones (LJ) attractive interactions between the nanopore and adenine bases lead to significantly longer translocation times for poly(dA)$_{100}$ compared to poly(dC)$_{100}$. Furthermore, their studies on hetero-DNAs with repeating units of A_mC_n showed that translocation times decrease ex-

ponentially with increasing cytosine content, underscoring the sensitivity of translocation dynamics to sequence composition [110, 111]. Sun *et al.* [116] extended these findings using Monte Carlo simulations by showing that different copolymer structures of monomer dA and monomer dC with the same overall composition can result in markedly different translocation behaviors due to variations in monomer-pore LJ attractive interactions. Complementary to these studies, Payne *et al.* [129] employed all-atom MD simulations and found that 20 nucleotide length poly(dC) translocated 1.34 times faster than its poly(dA) counterpart through a nanopore, further highlighting the influence of base size, mass, and structural differences on translocation times. In particular, poly(dC) is smaller and has a single-ring structure compared to the larger, double-ringed poly(dA). The all-atom simulations allowed for detailed observation of how these structural differences contribute to the faster translocation of poly(dC). However, the results were based on a single simulation without extensive sampling of translocation events.

Our coarse-grained cf-SMD simulations showed that, on average, poly(dC) translocated 2.45 times faster than poly(dA). However, the average translocation times are approximately two orders of magnitude smaller than experimental values reported by Meller *et al.* [14, 62], a remark which has been pointed out in Chapter 4 due to the high pulling forces in our CG-SMD simulations. Our results agree with the aforementioned experimental and theoretical findings, validating our computational model for its ability to capture sequence-dependent translocation dynamics. Furthermore, the CG approach allows for gathering data through multiple translocation events, which is essential for capturing a broader statistical representation of the translocation process, as we emphasized in Chapter 4. In the following sections, we will explore the underlying mechanisms contributing to the differences in translocation times between poly(dA) and poly(dC) molecules.

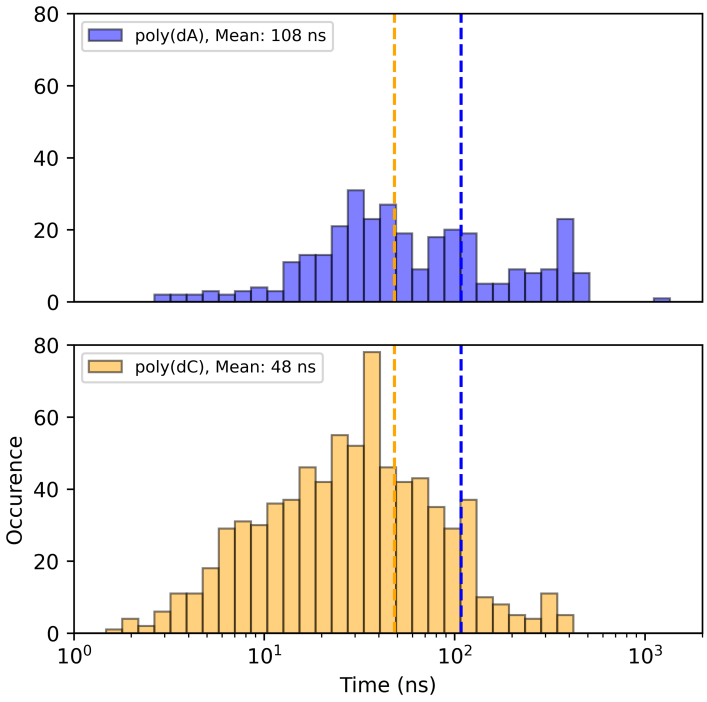


Figure 5.1: Distribution of translocation time per base of 10, 13 and 16 nucleotide length of poly(dA) (blue) and poly(dC) (orange) molecules pulled from 3' end, on logarithmic time scale with a force of 400 kJ mol^{-1} nm^{-1}. The average of the translocation time per base is shown in yellow and blue dashed lines for poly(dC) and poly(dA) molecules, respectively.

5.2.2 DNA-pore interactions depending on the sequence

In order to explore the differences observed in translocation times of poly(dC) and poly(dA) molecules, this section analyzes the interactions of poly(dC) molecules of 16 nucleotide length with the inner wall of the αHL stem in the same way as Section 4.3. These interactions were then compared with those of the poly(dA) molecules presented in Section 4.3.

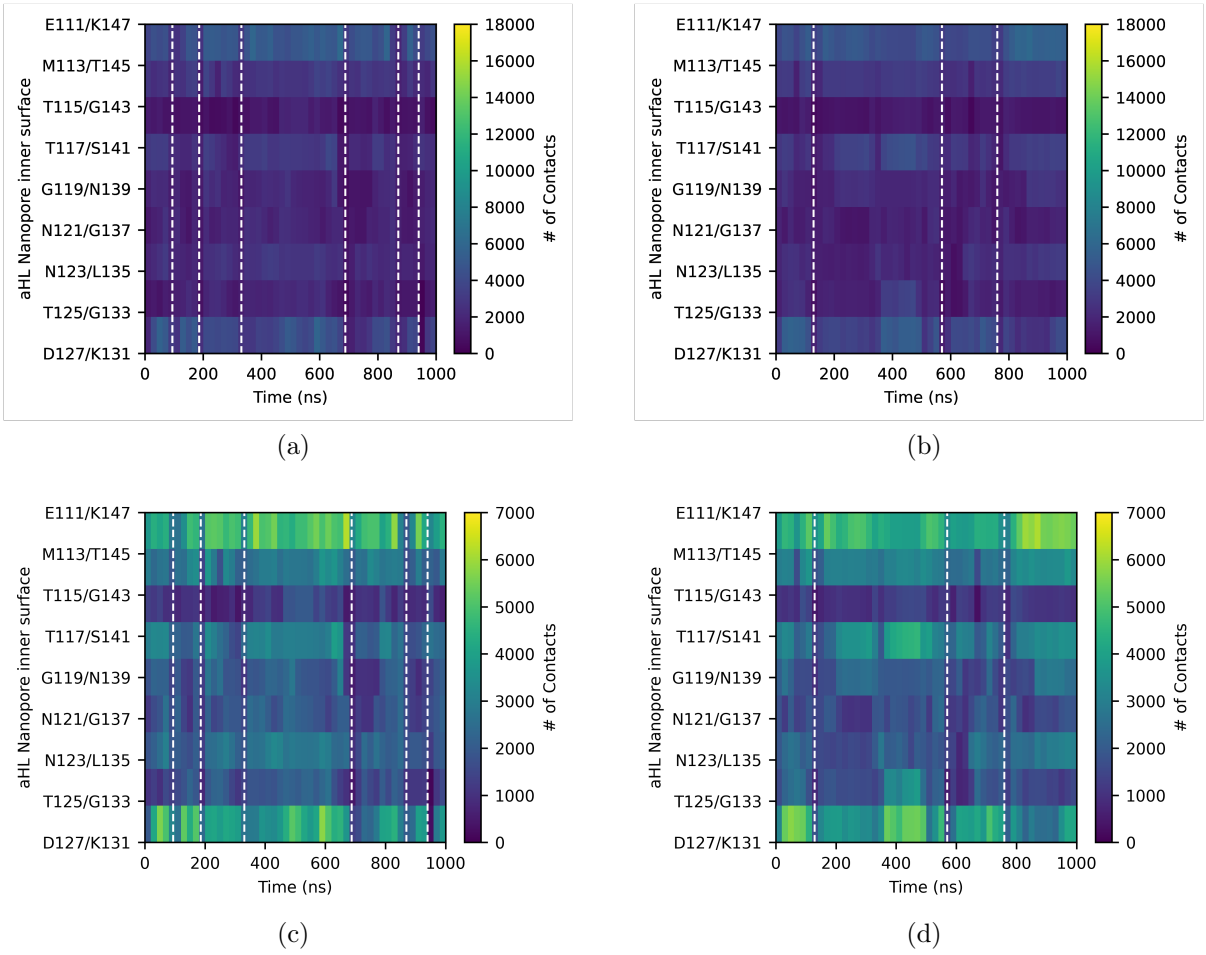


Figure 5.2: Contact maps for 3'-C_{16}-400 simulations showing the interactions of poly(dC) ssDNA with each ring in the stem of the αHL nanopore during a 1 μs MD simulation. Panels (a) and (b) represent two replicas of the simulation with the color scale used for poly(dA) for visualizing cumulative contacts every 20 ns, while panels (c) and (d) display the same data with a color scale adapted to the data. The dashed lines indicate the end of each translocation event. Both scales demonstrate that the constriction (E111/K147) forms the most significant interactions, while differences in overall interaction intensity can be better appreciated with the alternative color scale in panels (c) and (d).

Figure 5.2a and 5.2b display the contact maps for both poly(dC) ssDNA molecules of 16 nucleotides length during SMD simulations over time, with the same color scale as poly(dA) molecules presented in Figures 4.9a and 4.9b. Each row corresponds to an inner wall ring within the stem of the nanopore. The dashed lines mark the completion of the ssDNA molecule's translocation process. However, the interaction intensity compared to poly(dA) ssDNA molecules is significantly lower and more comparable to that of the neutral poly(dA) molecule presented in Figures 4.9c and 4.9d. Table 5.2 compares these number of contacts between ssDNA of 3'-A_{16} and 3'-C_{16}-400 systems with the constriction (E111/K147) and bottom (D127/K131) rings in the stem during complete translocation events for each replica. It is evident that poly(dC) interacts less with the constriction and bottom rings, which could originate from the smaller size of poly(dC) and, therefore,

translocates faster. To explore poly(dC)'s interaction with the pore inner wall further, we reconstructed the contact maps using a different color scale, as shown in Figures 5.2c and 5.2d. These new contact maps suggest that the constriction at E111/K147 is the primary site of interaction between the poly(dC) ssDNA molecule and the pore inner wall, similar to its poly(dA) counterpart, as the constriction is the narrowest part of the αHL nanopore. Although the bottom part of the nanopore has the widest radius, as shown in poly(dA) MD simulations, poly(dC) still exhibits numerous contacts with this region, despite being smaller by one MARTINI bead compared to poly(dA). This behavior may be explained by the electrostatic interactions between the negatively charged phosphate backbone and the positively charged lysine residues at position 131.

Table 5.2: Number of contacts between ssDNA of 3'-A_{16} and of 3'-C_{16}-400 systems with constriction (E111/K147) and bottom (D127/K131) rings in the stem during complete translocation events for each replica.

System	sim. #	Ring	T#1	T#2	T#3	T#4	T#5	T#6
3'-A_{16}	1	E111/K147	126,933	75,569	192,116			
		D127/K131	115,882	67,343	155,697			
	2	E111/K147	226,842	247,202				
		D127/K131	246,282	197,655				
3'-C_{16}-400	1	E111/K147	26,187	97,604	39,648			
		D127/K131	29,674	87,615	37,108			
	2	E111/K147	19,375	15,536	35,665	86,462	41,579	12,983
		D127/K131	19,954	17,303	29,275	69,572	28,172	10,152

The higher frequency of translocation events observed with poly(dC) ssDNA (see Table 5.1) and its reduced interaction with the nanopore's inner wall compared to poly(dA) (see Table 5.2) suggest that these interactions play a significant role in slowing down DNA translocation. This finding aligns with previous experimental studies showing that poly(dA) translocates more slowly than poly(dC) due to its stronger and more prolonged interactions with the pore walls. The MARTINI CG model successfully captures the faster translocation times of poly(dC) compared to poly(dA), demonstrating that it is well-suited for studying sequence-dependent differences in ssDNA translocation dynamics. .

5.2.3 DNA Tilt Angle Differences Depending on the Sequence

Building on the approach detailed in Chapter 4, Section 4.3, where we demonstrated that the CG MARTINI model successfully captured orientation-dependent differences in tilt angles, we extended this analysis to examine sequence-dependent variations. In Chapter 4, we showed that increased tilt angles in 5'-oriented ssDNA led to slower translocation. Here, we investigate whether similar tilt angle differences exist between poly(dA) and poly(dC) sequences and how they contribute to the faster translocation of poly(dC). To our knowledge, sequence-dependent tilt angle variations during translocation have not been previously studied.

We calculated the tilt angles between the ssDNA bases and the backbone as a function of the base's position within the nanopore during translocation, using the same method-

ology described in Section 4.3. For poly(dC), we measured angles between the BB1, BB3, and SC3 beads, excluding the terminal bases. (see Figure 5.3) These angles were computed using the GROMACS `angle` command, while the `traj` command was used to track the z-direction positions of the bases. The coordinates were adjusted to set the constriction point as the origin in the simulation box. Running averages were computed over a 2 ns time window, and data were averaged over 0.2 nm intervals along the z-axis. The mean and standard error of the tilt angles across both replicas were calculated, with results presented in Figure 5.4.

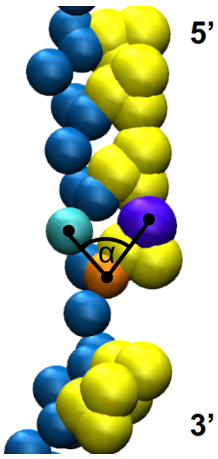


Figure 5.3: Representation of the tilt angle (α) defined between the center of mass of BB1, BB3, and SC3 beads of the ssDNA molecule, represented in cyan, orange, and purple beads, respectively. 5' and 3' extremities are indicated in the illustration.

Table 5.3: Average tilt angles along with the standard errors of DNA bases relative to the backbone for 16-nucleotide poly(dA) and poly(dC) molecules before entering and after exiting the pore.

System	Average tilt angle before entering pore (°)	Average tilt angle after exiting pore (°)
3'-A_{16}	94.6 ± 2.6	82.7 ± 1.6
3'-C_{16}-400	87.7 ± 1.6	77.8 ± 0.8

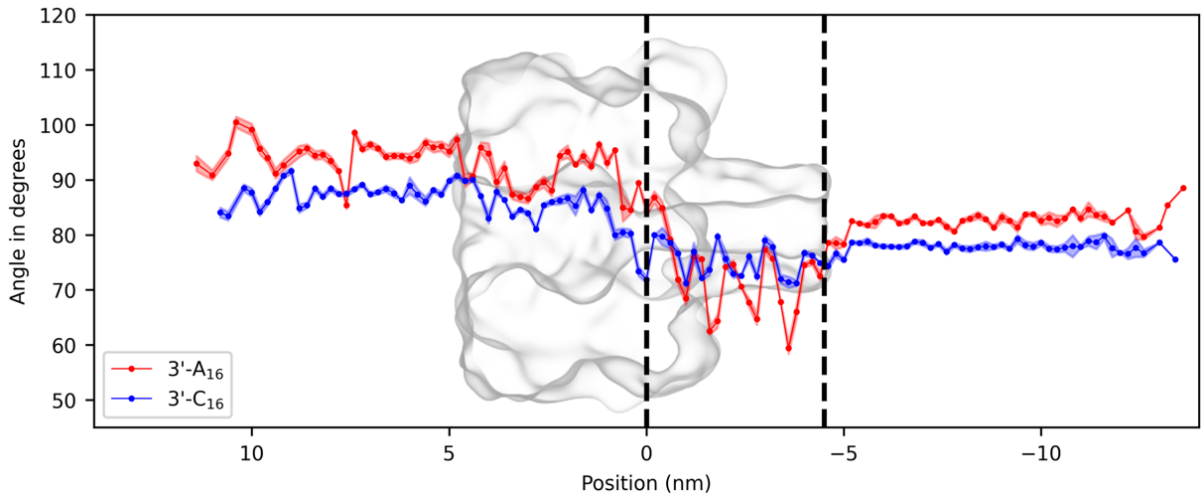


Figure 5.4: Evolution of tilt angles inside the nanopore during ssDNA translocation for poly(dA) and poly(dC) molecules. Average tilt angles of DNA bases relative to the backbone for both 16-nucleotide poly(dA) and poly(dC) molecules are shown as red and blue lines, respectively, as a function of z-position in the simulation box. The values represent the mean tilt angle for 14 bases (excluding terminal bases) across all translocation events in two simulation replicas, with error bars representing standard error. The black dashed lines denote the constriction and bottom of the stem.

The tilt angle values, summarized in Table 5.3, show that poly(dC) consistently exhibits smaller and less oscillatory tilt angles compared to poly(dA), both before entering and after exiting the pore. This difference likely stems from the structural variations between the two monomers, as discussed in Chapter 2 (see Figure 2.21). The tilt angles for poly(dA) and poly(dC) are calculated using different MARTINI beads due to their structures. Specifically, poly(dC) has one fewer bead per nucleotide than poly(dA), meaning the angles do not represent the same physical measurement. Despite this, the tilt angles remain useful for comparing the evolution of the strands during translocation.

As seen in Figure 5.4, the evolution of tilt angles within the stem differs depending on the sequence. In the vestibule, where the molecule starts interacting with the pore walls, both molecules exhibit slight decreases in tilt angles. However, within the stem, the tilt angle of the 3'-A_{16} molecule drops sharply and oscillates, reaching a minimum of about 60° as it passes through the narrowest part of the pore. In contrast, the tilt angle of the 3'-C_{16}-400 molecule shows fewer oscillations, decreasing to approximately 70° but never reaching the same minimum tilt angles seen for poly(dA). This shows that poly(dC)

requires less tilting and deformation of its bases to pass through the constriction.

From our extensive CG-SMD simulations covering over 230 base translocations across two different ssDNA sequences, the faster translocation of poly(dC) compared to poly(dA) appears to be linked to its less tilting behaviour during translocation, as explained by the structural differences mentioned above. These findings highlight the important role that tilt angle, influenced by molecular sequence, plays in determining the translocation speed through the αHL nanopore.

5.2.4 Conformational differences depending on the sequence

To investigate the sequence-dependent conformational changes of ssDNA during its translocation, we analyzed the inter-base distances throughout the MD simulations, as described in Section 4.5. These inter-base distances were then compared with those of the poly(dA) molecules presented in Section 4.5.

During the translocation process for 3'-C_{16}-400 simulations, the inter-base distances were analyzed and represented in Figure 5.5 as 2D histograms against the position along the pore axis, with the stem region highlighted in red, as explained in Section 4.5.

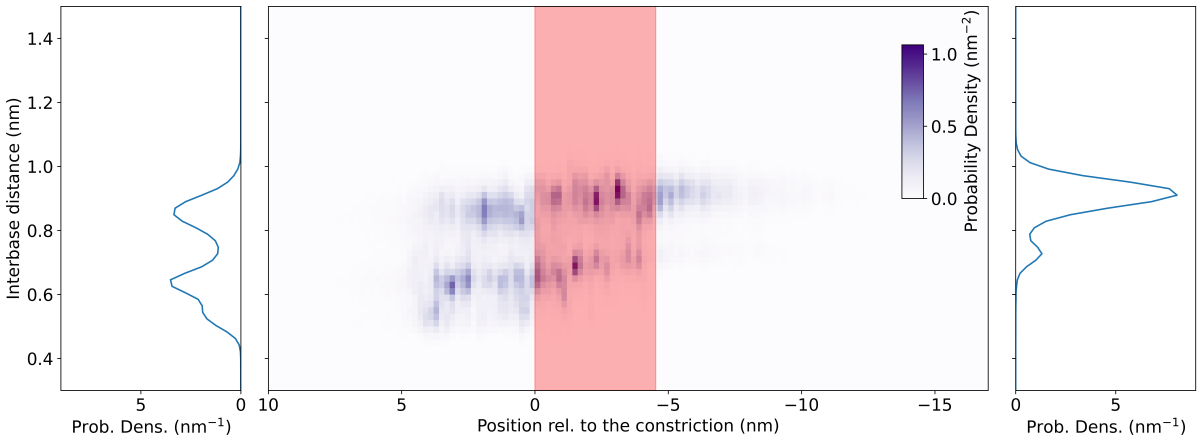


Figure 5.5: Inter-base distances for 3'-C_{16}-400 simulations. The distance between successive bases (inter-base distance) for all pairs, except the first and last ones, is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 representing the pore constriction ring (E111/K147). Left panel: Probability density of inter-base distance before the entry of the DNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: Probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm).

Before the ssDNA enters the pore, three populations of inter-base distances are observed at 0.52 ± 0.04, 0.65 ± 0.05, and 0.85 ± 0.06 nm. The first and third populations closely match those seen in 3'-A_{16} simulations, 0.52 ± 0.10 and 0.89 ± 0.12 nm. However, the second population, absent in poly(dA) translocation simulations (see Figure 4.16), indicates that poly(dC) fluctuates between more conformational states before entering

the pore, with its inter-base distance populations resembling those of poly(dA) free in solution (see Figure 4.15a).

After exiting the pore, poly(dC) exhibits a dominant inter-base distance population at 0.91 ± 0.10 nm, similar to poly(dA), but also a smaller population at 0.73 ± 0.04 nm, though with low probability density. This additional population, negligible in poly(dA) translocations, supports that poly(dC) fluctuates between multiple conformations more readily. The dominant stretched conformation is expected since a more elongated conformation is favored once the molecule exits the pore, as similarly observed in 3'-A_{16} and 5'-A_{16} simulations.

These observations on inter-base distances reveal that poly(dC) exhibits the same primary populations as poly(dA) both before entering the stem and after exiting the pore. However, poly(dC) displays more conformational states before and after exiting the pore. This suggests that poly(dC) experiences more frequent conformational changes during its translocation, consistent with prior experimental findings that poly(dC) is less structured than poly(dA) [62].

Additionally, the increased probability of the most stretched configuration observed both before entering the stem and after exiting the pore can be attributed to the applied SMD force directing the ssDNA through the pore, stretching the molecule into an elongated conformation with its bases aligned along the pore's axis. In the alternative conformations, the ssDNA likely experiences less influence from the applied force, allowing for more fluctuation in its structure, particularly within the vestibule region.

In summary, the conformational differences between poly(dC) and poly(dA) may contribute to the translocation velocity differences observed. This study provides insights into the role of sequence-dependent conformational changes during ssDNA translocation and their potential impact on translocation dynamics.

5.3 Influence of Pulling Force on Translocation Time

In Sections 4.2 and 5.2, we measured the translocation time per base under a pulling force of 400 kJ mol^{-1} nm^{-1}. Our CG-SMD simulations revealed that the translocation times for both poly(dA) and poly(dC) were two orders of magnitude smaller than experimental results for both molecules with the αHL nanopore at 120 mV. This difference, as mentioned before, arises from the higher pulling forces applied in our simulations. Therefore, this section will explore the effect of the pulling force on translocation time and dynamics by varying the applied force: we applied four forces smaller than 400 kJ mol^{-1} nm^{-1} and one larger.

For this section, we performed CG-SMD simulations using 10 nucleotide-long ssDNA sequences of both poly(dA) and poly(dC) to ensure enough statistics since this length translocated the fastest for both sets of ssDNA molecules (see Table 5.1). Given that decreased translocation velocity was expected with decreased pulling forces, we chose to keep the ssDNA length minimal for both types of molecules. We conducted SMD simulations with forces of 200, 250, 300, 350, 400, and 450 kJ mol^{-1} nm^{-1} for poly(dC) and 200, 300, 350, 400, and 450 kJ mol^{-1} nm^{-1} for poly(dA), as explained in Section 1.1.2 and illustrated in Table 5.1. For poly(dA) molecules, we prioritized the 300, 350, and 450 kJ mol^{-1} nm^{-1}.

5.3.1 Distributions of Translocation Time Per Base Under Varying Pulling Forces

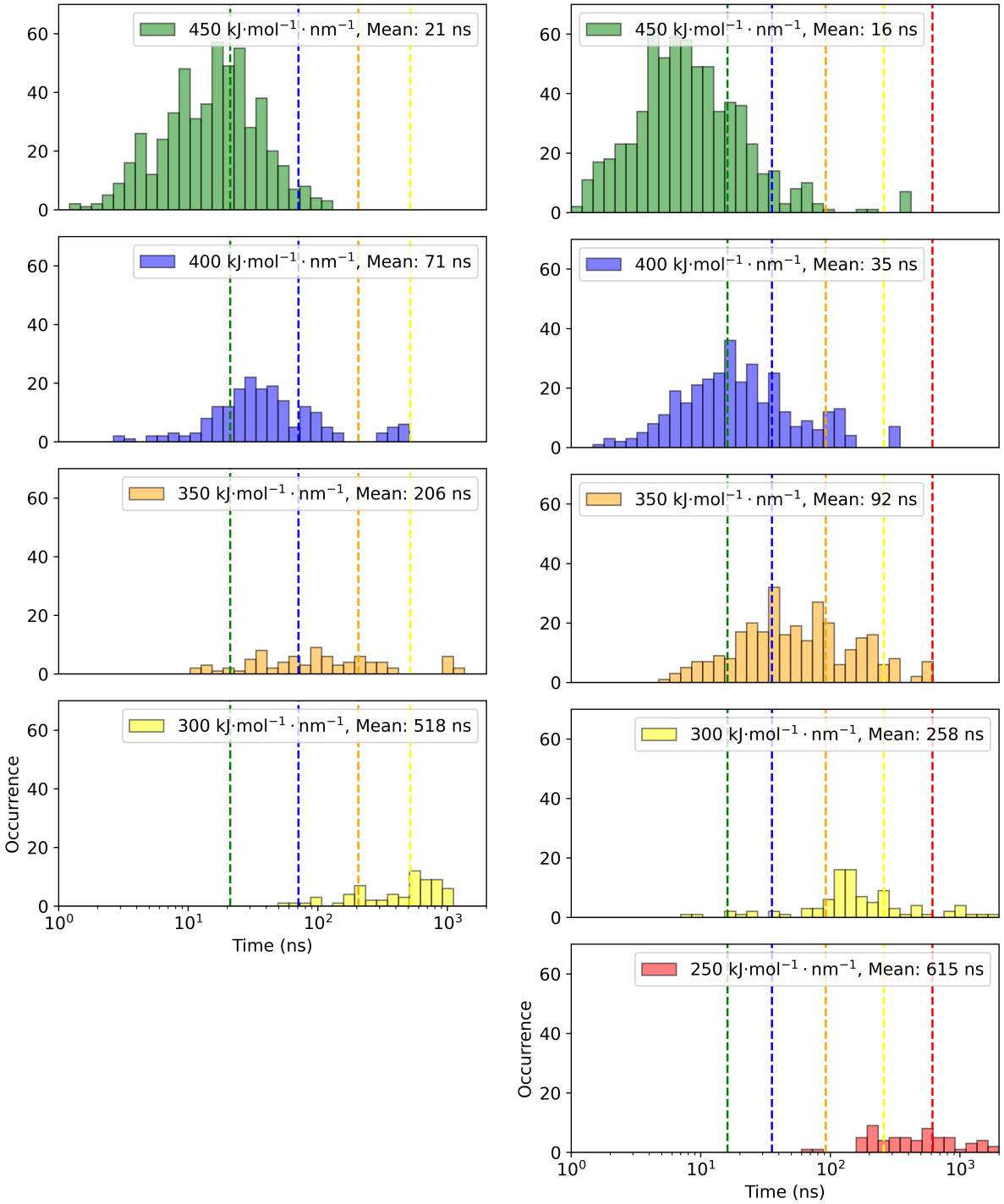
The translocation times for each pulling force were calculated as described earlier in Section 5.2 and 4.2, and their distributions are presented with a log binning, consistent with our previous analyses. Figures 5.6a and 5.6b illustrate the distribution of translocation times per base for poly(dA) and poly(dC) molecules under different pulling forces. The data used to generate these distributions are detailed in Table 5.1.

Both poly(dA) and poly(dC) exhibit a broad range of translocation times, which is consistent with prior experimental [13] and theoretical [114, 119] observations. As the pulling force decreases, the range of translocation times generally expands for both sequences, as can be seen from Figures 5.6a and 5.6b. Table 5.4 presents mean translocation times and standard errors. The results clearly show that the translocation time per base increases as the pulling force decreases. It is also important to note that poly(dC) consistently translocates faster than poly(dA) across all tested forces, as evidenced by the mean translocation times presented in Table 5.4. Furthermore, standard errors associated with poly(dC) are systematically smaller than those associated with poly(dA). These findings are consistent with the sequence-dependent dynamics discussed and analyzed earlier in Section 5.2. However, the relationship between pulling force and translocation dynamics still needs to be addressed and will be further explored in the subsequent sections.

Table 5.4: Comparison of average translocation times per base and standard errors for 10-nucleotide length Poly(dA) and Poly(dC) at different applied forces.

Force ($kJ{\cdot}mol^{-1}{\cdot}nm^{-1}$)	Translocation time per base (ns) (Poly(dA))	Translocation time per base (ns) (Poly(dC))
450	21 ± 1	16 ± 2
400	71 ± 8	35 ± 3
350	206 ± 33	92 ± 6
300	518 ± 36	258 ± 34
250	—[a]	615 ± 65
200	—[b]	—[b]

[a] No data available.
[b] None of the bases are translocated.



(a) (b)

Figure 5.6: Distributions of translocation times per base of (a) poly(dA) (b) poly(dC), on logarithmic time scale under varying pulling forces. The average of the translocation times per base is shown in dashed lines for varying pulling forces, with the same color code: green, blue, orange, yellow, and red for 450, 400, 350, 300, 250 and 200 kJ mol^{-1} nm^{-1}, respectively.

5.3.2 The relationship between pulling force and translocation velocity

Translocation velocity is a parameter that can be easily calculated from experimental measurements, typically by dividing the length of the αHL stem (48 Å) by the mean translocation time, which can be expressed as:

$$v = \frac{L}{t} \tag{5.1}$$

where v is the average translocation velocity, L is the length of the αHL stem (48 Å), and t is the mean translocation time.

To facilitate comparison with experimental data, particularly the work of Meller *et al.* [15], we calculated translocation velocities using the mean translocation times per base, which were derived in the previous section. The velocities and associated standard errors for both poly(dC) and poly(dA) molecules under varying pulling forces are presented in Table 5.5.

Table 5.5: Comparison of average velocities and standard errors for 10-nucleotide length Poly(dA) and Poly(dC) at different applied forces, along with the average velocity ratios of Poly(dC) to Poly(dA).

Force ($kJ{\cdot}mol^{-1}{\cdot}nm^{-1}$)	Velocity (Poly(dA)) (Å/ns)	Velocity (Poly(dC)) (Å/ns)
450	2.28 ± 0.20	2.98 ± 0.32
400	0.68 ± 0.16	1.36 ± 0.23
350	0.23 ± 0.10	0.52 ± 0.086
300	0.093 ± 0.022	0.19 ± 0.095
250	—[a]	0.078 ± 0.015
200	—[b]	—[b]

[a] No data available.
[b] None of the bases are translocated.

Figure 5.7 illustrates the relationship between pulling force and the average translocation velocity for poly(dA) (red squares) and poly(dC) (blue circles), plotted on a semi-logarithmic scale. This is essentially the inverse of plotting mean translocation time against force. As the pulling force increases, both poly(dA) and poly(dC) exhibit enhanced translocation velocities, as expected. This observation aligns well with prior experimental and computational studies, which have repeatedly demonstrated that the translocation velocities increase with the increased electrical [13, 15, 62] or pulling force [114, 119].

Several models have been proposed to explain the relationship between pulling force and translocation velocity in DNA translocation studies. Payet *et al.* [188] employed an Arrhenius-type model, demonstrating that the dynamics of ssDNA transport through nanopores are characterized by an initial energy barrier for pore entry, followed by a translocation process that accelerates exponentially with the applied voltage. Alternatively, Meller *et al.* [15] used a quadratic fit to describe the relationship between voltage and translocation velocity, although they did not provide a detailed justification for this choice [187]. Additionally, some theoretical models, such as those proposed by Kasianowicz *et al.* [13], predict a linear response in translocation velocity over a restricted range

of applied potentials (70-120 mV). Beyond this range, non-linear behaviors have been observed, including the quadratic response identified by Meller *et al.*

Given these varying models, we opted to use an exponential dependence, $v = A \times e^{bF}$, to describe the relationship between pulling force and translocation velocity in our study, following the approach by Payet *et al.*, who demonstrated that the translocation velocity increases exponentially with applied voltage in both αHL and aerolysin nanopores. As shown in Figure 5.7, the exponential model fits our data well, with R^2 values of 0.98 for poly(dA) and 0.94 for poly(dC). For poly(dA), the fitting parameters were calculated as $A = 1.41 \times 10^{-4}$ and $b = 0.0213$, and for poly(dC), $A = 7.48 \times 10^{-4}$ and $b = 0.0186$. On the contrary, the quadratic model suggested by Meller *et al.* and the linear model from Kasianowicz *et al.* were not well-suited for our data.

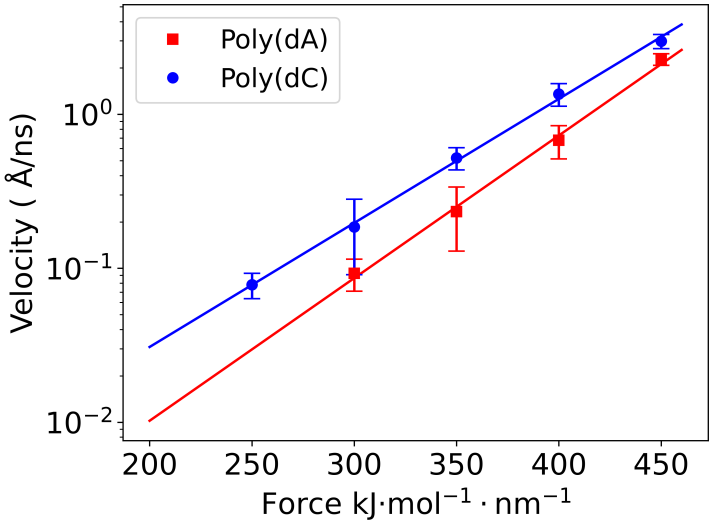


Figure 5.7: Comparison of translocation velocities of poly(dA) (red squares) and poly(dC) (blue circles) under different pulling forces on a logarithmic velocity scale. The standard error can be consulted in Table 5.5. The semi-logarithmic plot shows the exponential fit for poly(dA) and poly(dC) based on $v = A \times e^{bF}$, where F is the pulling force (in kJ mol^{-1} nm^{-1}), and v is the translocation velocity (in Å ns^{-1}). For poly(dA), the values of A and b were calculated as $A = 1.41 \times 10^{-4}$ Å ns^{-1} and $b = 0.0213$ nm mol kJ^{-1}, while for poly(dC), $A = 7.48 \times 10^{-4}$ Å ns^{-1} and $b = 0.0186$ nm mol kJ^{-1}.

However, directly comparing the experimentally applied voltage regimes with our SMD pulling forces is not straightforward. As discussed in Chapter 4, the difference arises from how the force is applied in experiments versus simulations. In the experiments, an electric force is exerted on each charged group of the DNA, while in our simulations, a mechanical force is applied only to the 3' end of the DNA strand.

From Figure 5.7, we observed that the exponential fits for the velocities of both poly(dA) and poly(dC) converge as the pulling force increases, indicating that the sequence-dependent differences in translocation dynamics diminish under stronger forces. This trend is consistent with previous experimental findings, which demonstrated that higher voltage accelerates DNA translocation but reduces nucleotide discrimination accuracy [62]. The velocity ratios between poly(dC) and poly(dA) range from 2.26 to 1.31 across different pulling forces as a result of our CG-SMD simulations, with these ratios decreasing as

the pulling force increases, as can be seen in Table 5.6.

These results underscore the impact of varying pulling forces on translocation times and dynamics. The ability of the MARTINI model to accurately capture these differences highlights its effectiveness in simulating the translocation of different ssDNA sequences under diverse conditions.

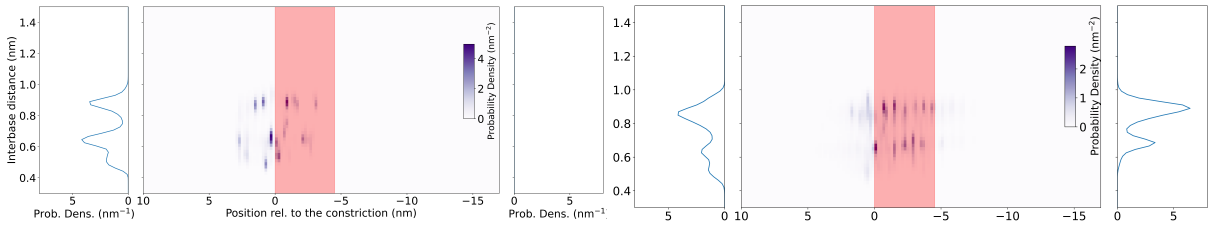
Table 5.6: Ratios of translocation velocities for 10-nucleotide length Poly(dC) to Poly(dA) at different applied forces.

Force ($kJ \cdot mol^{-1} \cdot nm^{-1}$)	Ratio (Poly(dC)/Poly(dA))
450	1.31 ± 0.3
400	2.00 ± 0.8
350	2.26 ± 1.3
300	2.11 ± 1.6
250	—[a]
200	—[b]

[a] No data available.
[b] None of the bases are translocated.

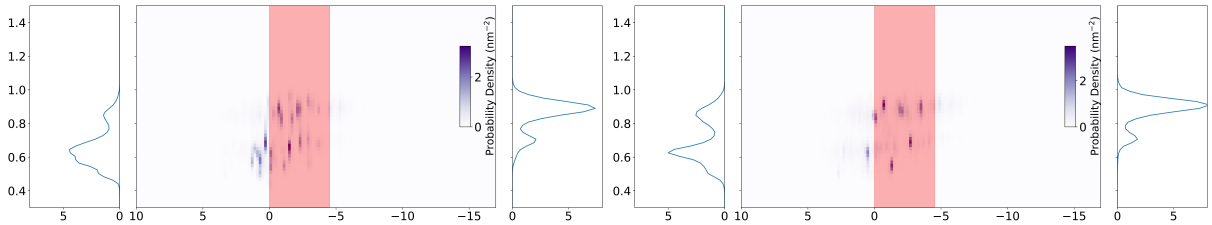
5.3.3 Influence of pulling force on ssDNA conformations

To investigate the effect of pulling force on the conformation of ssDNA during translocation through the αHL nanopore, we analyzed the inter-base distances for poly(dC) molecules under varying pulling forces. The inter-base distances were examined relative to the position of the ssDNA within the nanopore and are represented in Figure 5.8 as 2D histograms. Table 5.7 and 5.8 summarize the inter-base distance states before and after the nanopore stem across different pulling forces. These states were found by fitting Gaussian distributions to the probability densities presented in Figure 5.8 before and after the stem.



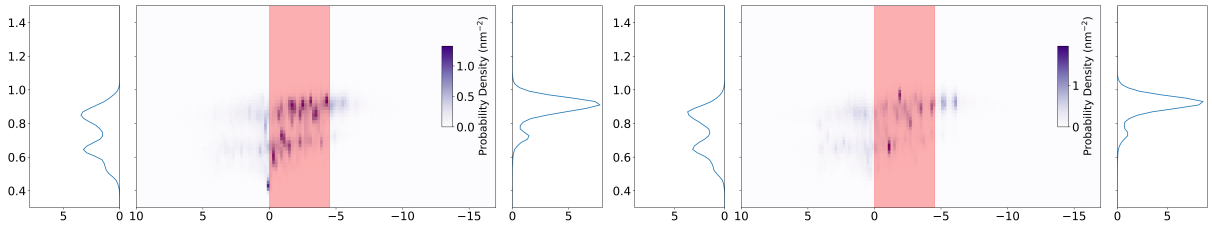
F = 200 kJ mol^{-1} nm^{-1}

F = 250 kJ mol^{-1} nm^{-1}



F = 300 kJ mol^{-1} nm^{-1}

F = 350 kJ mol^{-1} nm^{-1}



F = 400 kJ mol^{-1} nm^{-1}

F = 450 kJ mol^{-1} nm^{-1}

Figure 5.8: Inter-base distances for 3'-C_{10}-F simulations under various pulling forces (F-values). For simplicity, the axis labels are shown on the first probability density plot. The distance between two successive bases (inter-base distance) for all pairs except the first and last ones is represented as a probability density. Central panel: probability density of inter-base distance as a function of the position along the pore axis. The stem is depicted as a red-filled region, with 0 representing the pore constriction ring (E111/K147). Left panel: Probability density of inter-base distance before the entry of the DNA in the stem. It is the integration of the central panel over the position above the stem ($z > z_{\text{constriction}} = 0$ nm). Right panel: Probability density of inter-base distance after the exit of the DNA from the stem. It is the integration of the central panel over the position below the stem ($z < z_{\text{bottom}} \simeq -4.5$ nm).

Table 5.7: Inter-base distance states before the stem for poly(dC) under varying pulling forces.

Pulling Force ($kJ \cdot mol^{-1} \cdot nm^{-1}$)	State 1 (nm)	State 2 (nm)	State 3 (nm)
200	0.51 ± 0.04	0.64 ± 0.04	0.88 ± 0.04
250	0.51 ± 0.04	0.63 ± 0.06	0.86 ± 0.06
300	0.52 ± 0.04	0.63 ± 0.06	0.85 ± 0.05
350	0.52 ± 0.03	0.63 ± 0.04	0.85 ± 0.06
400	0.52 ± 0.04	0.65 ± 0.05	0.85 ± 0.06
450	0.52 ± 0.04	0.65 ± 0.05	0.85 ± 0.06

Table 5.8: Inter-base distance states after the stem for poly(dC) under varying pulling forces.

Pulling Force ($kJ \cdot mol^{-1} \cdot nm^{-1}$)	State 1 (nm)	State 2 (nm)
200	-	-
250	0.69 ± 0.04	0.89 ± 0.05
300	0.70 ± 0.04	0.89 ± 0.04
350	0.72 ± 0.04	0.91 ± 0.04
400	0.73 ± 0.04	0.91 ± 0.05
450	0.76 ± 0.05	0.92 ± 0.04

Before entering the nanopore stem, the inter-base distances show three populations ranging from approximately 0.50 to 0.87 nm under all applied forces, as seen in Figure 5.8 and Table 5.7. At the pulling force of 200 kJ mol^{-1} nm^{-1}, the inter-base distance states for poly(dC) appear at 0.51 ± 0.04 nm, 0.64 ± 0.04 nm, and 0.88 ± 0.04 nm. In this case, no complete translocation of any of the bases was observed, indicating that the DNA remains in a more relaxed state and adopts all three states, similar to when it is free in solution. This observation aligns with our findings for free poly(dA) in Section 4.5, where we calculated inter-base distances in solution and identified three distinct populations at 0.50 ± 0.08 nm, 0.71 ± 0.08 nm, and 0.91 ± 0.12 nm (see Figure 4.15a). Despite increasing the pulling force, the positions of these states remain largely unchanged, and the probabilities of each state (Figure 5.9) show no clear force dependence. These results suggest that the applied force does not significantly affect the conformations of the DNA before entering the nanopore stem.

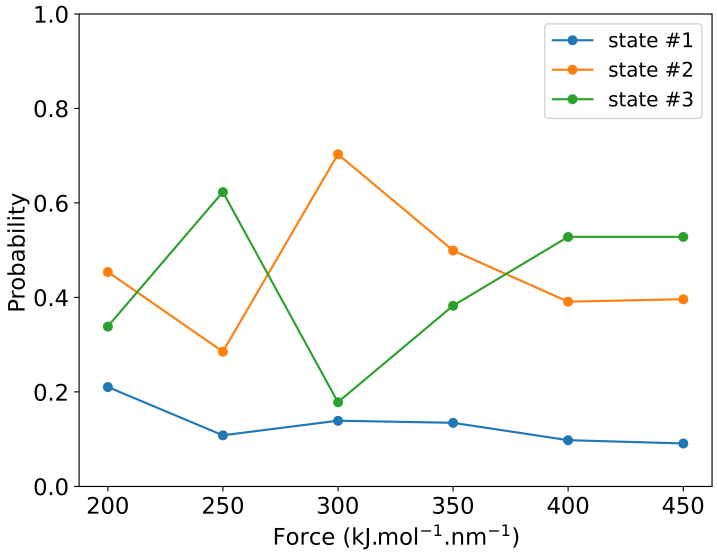
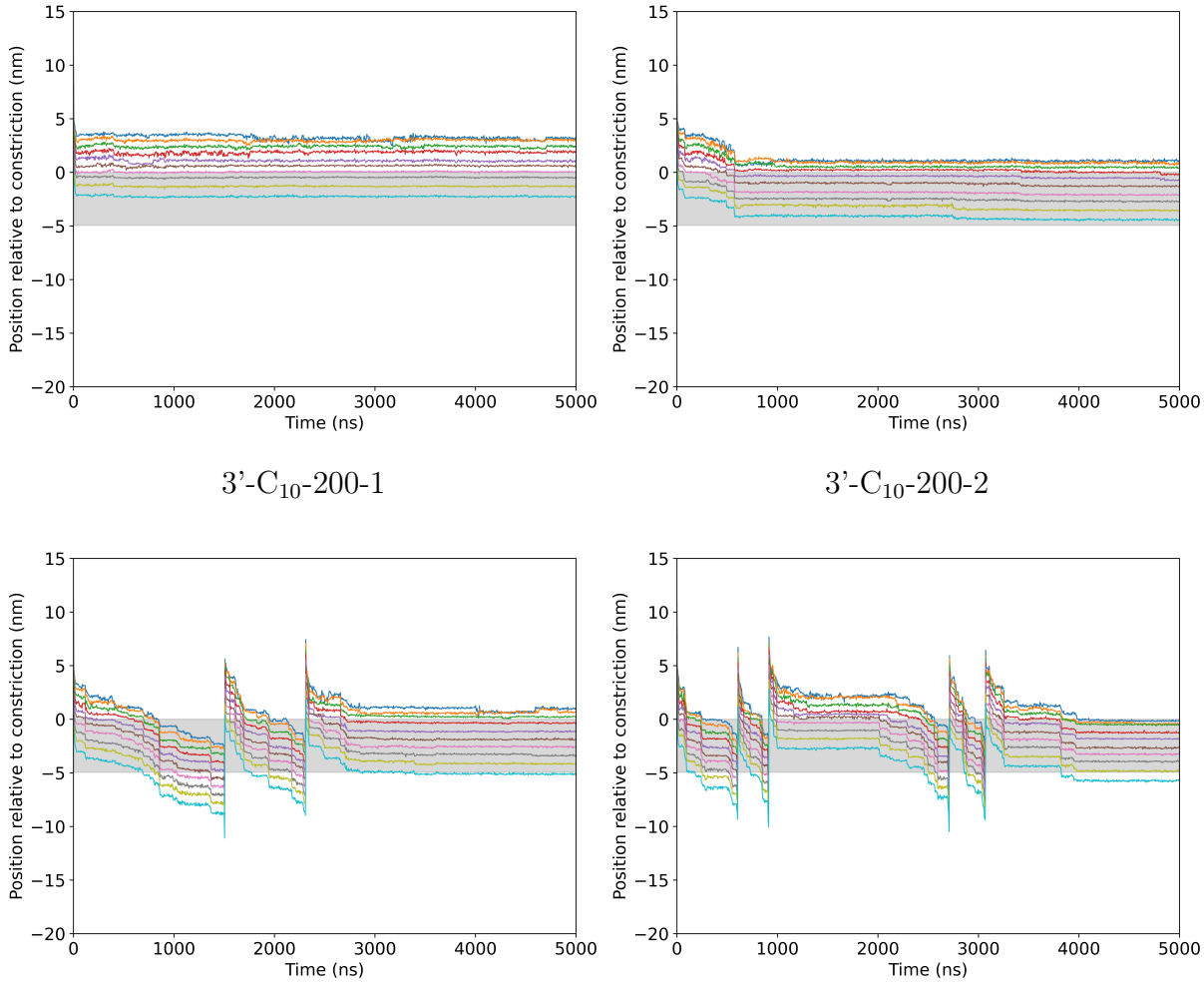


Figure 5.9: Evolution of the probabilities of each inter-base state before entering the pore as a function of the pulling force.

After exiting the stem, the inter-base distances exhibit two populations across all pulling forces, at around 0.7 and 0.9 nm, but the smaller population tends to diminish as the pulling force increases. This trend highlights how an increased pulling force stretches the DNA more uniformly as it exits the stem, reducing the occurrence of shorter inter-base distances. These observations closely align with the results obtained in Section 4.5, where we observed a single stretched population of poly(dA) at 0.93 nm at the exit of the pore, indicating that the stretching at the exit of the nanopore is sequence-independent and force-related.

In order to understand the probability distributions of inter-base distances prior to the stem, we extended our analysis by tracing the z-positions of the COM of the bases, similar to the approach discussed in Section 4.5. This analysis provided insight into how the strand behaves while waiting for translocation and allowed us to correlate the z-positions of the bases with the inter-base populations observed.

At low force constants, 200 and 250 kJ mol^{-1} nm^{-1}, we observed that the ssDNA spends a significant amount of time outside the stem, as shown in Figure 5.10. During this waiting phase, the z-positions of the bases exhibit frequent crossovers, indicating a less elongated and linear conformation. These crossovers suggest that at these lower forces, the DNA does not adopt a fully stretched conformation where the bases are aligned on top of each other, resembling more closely the behavior of DNA free in solution. It is important to highlight that these longer waiting times at lower forces dominate the histogram distributions because the probability densities are integrated over the entire simulation. In contrast, at higher pulling forces, 300, 350, 400, and 450 kJ mol^{-1} nm^{-1}, the ssDNA passes through the pre-stem region more quickly, as can be seen in Figures 5.11 and 5.12. The shorter waiting times at higher forces reduce the contribution of inter-base distances at this pre-stem region to the histograms, resulting in more similar inter-base populations at these forces, especially at 400 and 450 kJ mol^{-1} nm^{-1}.



3'-C_{10}-200-1 3'-C_{10}-200-2

3'-C_{10}-250-1 3'-C_{10}-250-2

Figure 5.10: Positions in the z-coordinate of the bases of 3'-C_{10}-F simulations for F = 200 and 250 kJ mol^{-1} nm^{-1} over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147).

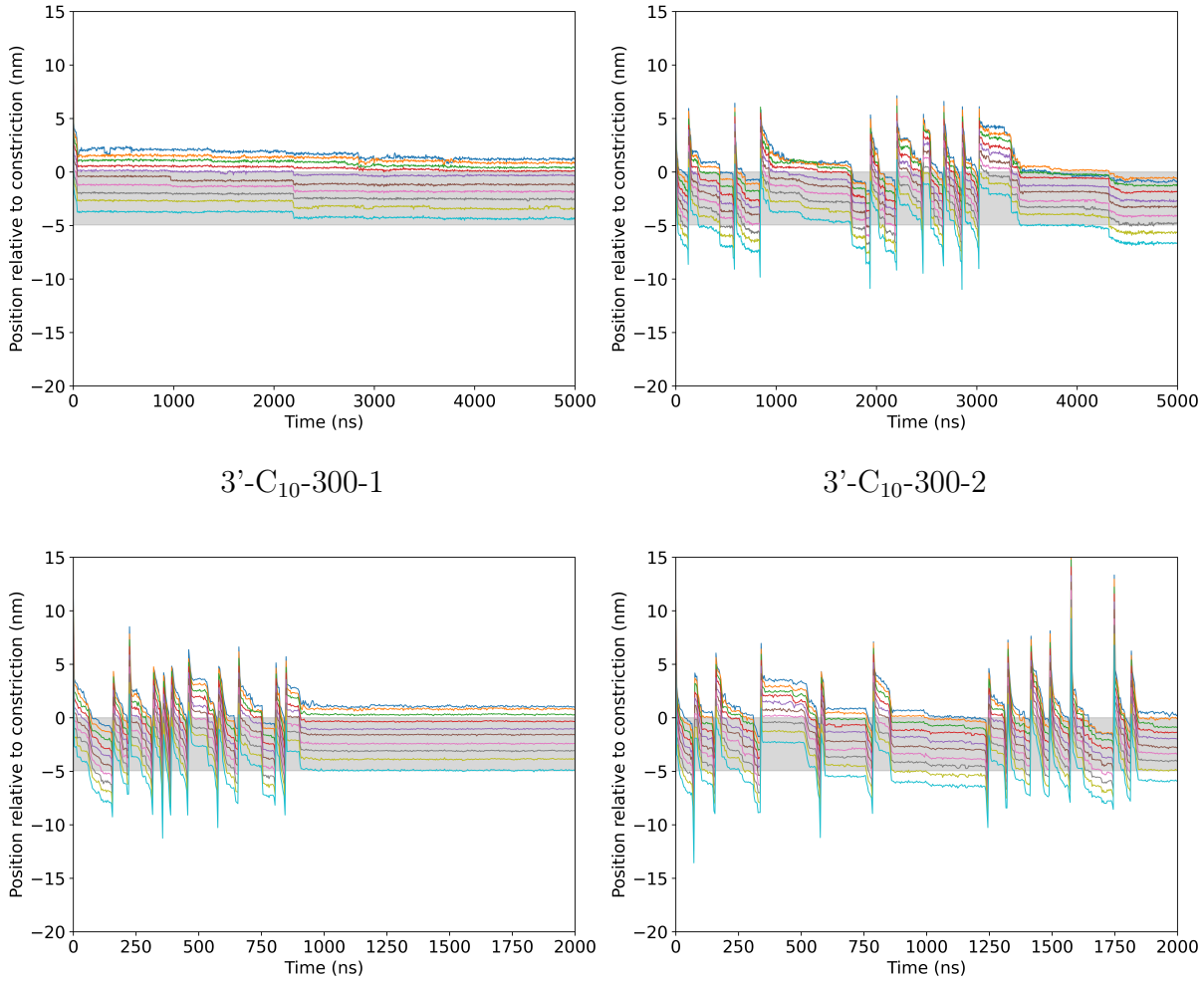


Figure 5.11: Positions in the z-coordinate of the bases of 3'-C_{10}-F simulations for F = 300 and 350 kJ mol^{-1} nm^{-1} over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147).

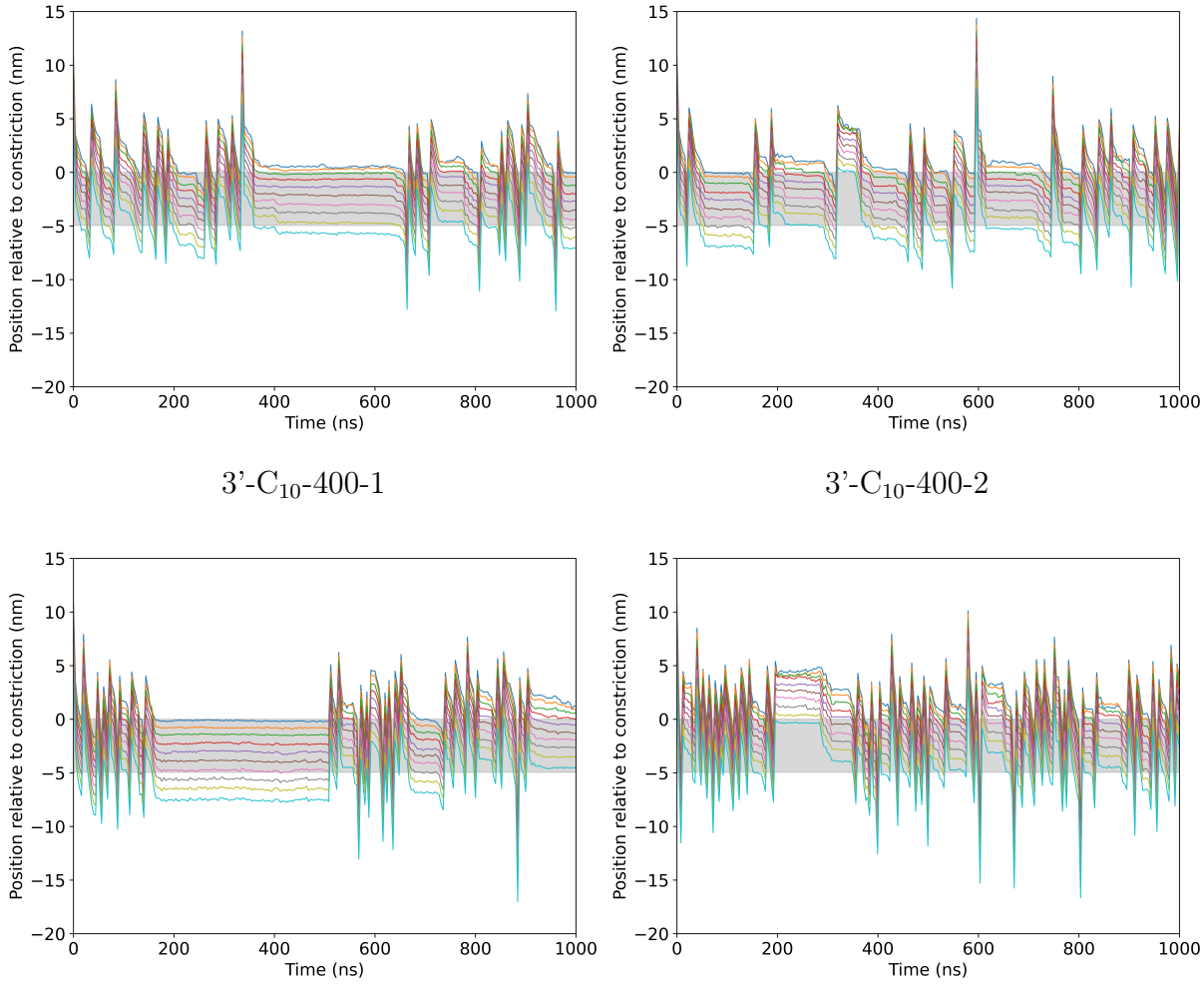


Figure 5.12: Positions in the z-coordinate of the bases of 3'-C_{10}-F simulations for F = 400 and 450 kJ mol^{-1} nm^{-1} over time. The stem is depicted as a gray-filled region, with 0 on the position axis corresponding to the pore constriction (E111/K147).

Our analysis of inter-base distances under varying pulling forces reveals changes in DNA conformation as the applied force increases. Before entering the stem, at lower forces, the inter-base probability densities are very similar to the ssDNA behavior free in solution. However, the inter-base distance populations before stem entry show no significant dependence on the applied force, suggesting that the pulling force does not directly affect these populations. After exiting the stem, the influence of the applied force becomes more pronounced. The inter-base distances gradually converge to a stretched conformation as the pulling force increases, demonstrating a clear force-dependent modulation of DNA conformation at the exit of the stem. These findings highlight the differential impact of the applied force on ssDNA conformation before and after passing through the αHL nanopore.

5.4 Conclusion

In this chapter, we extended our investigation into the influence of sequence composition and pulling force on ssDNA translocation through the αHL nanopore using the MARTINI CG model. Using the CG model enabled us to gather a large amount of statistical data, making it possible to effectively compare sequence-dependent behaviors and the impact of pulling forces on translocation dynamics. Our cf-SMD simulations successfully replicated key experimental findings, such as the faster translocation times of poly(dC) compared to poly(dA) and the non-linear relationship between pulling force and translocation velocity.

Our study first highlighted the sequence-dependent differences in translocation times between poly(dA) and poly(dC). We found that poly(dC) translocated significantly faster than poly(dA) under the same pulling force, with an average translocation time approximately 2.45 times shorter than that of poly(dA). This observation is consistent with experimental studies through αHL, where the ratio of poly(dC) translocation over that of poly(dA) was found to be 2.75, reflecting a good match. Experimental and theoretical studies attributed the slower translocation of poly(dA) to its stronger attractive interactions with the nanopore and its greater structural rigidity, which were confirmed by our contact map analysis and inter-base distance measurements. Furthermore, our analysis of tilt angles revealed that poly(dC) requires less base tilting to pass through the pore compared to poly(dA), which might contribute to its faster translocation.

Next, we examined the influence of varying pulling forces on the translocation dynamics of both poly(dA) and poly(dC). Across all forces, poly(dC) consistently translocated faster than poly(dA), with the velocity ratio between the two sequences decreasing as the pulling force increased, indicating that stronger forces reduce the sequence-dependent differences in translocation times. Our results demonstrated an exponential relationship between pulling force and translocation velocity for both sequences.

Finally, our analysis of inter-base distances under varying pulling forces revealed that the applied force does not influence the inter-base distance populations before stem entry. However, after exiting the nanopore, the DNA adopted a more stretched conformation as the pulling force increased. These observations highlight the differences in conformations of ssDNA and their dependence on the pulling force before entering and after exiting the stem.

These findings validate the ability of the MARTINI CG model to capture the role of both sequence composition and pulling force in ssDNA translocation. This study reinforces the use of CG simulations as an efficient alternative to all-atom simulations, providing valuable complementary information to experimental work.

Conclusion and Perspectives

In this thesis, we investigated the dynamics of DNA translocation and unzipping through the α-hemolysin (αHL) nanopore using both experimental techniques and computational simulations. The experimental work explored the single-stranded DNA (ssDNA) translocation to complement the computational studies but mainly focused on the unzipping dynamics of the double-stranded DNA (dsDNA) molecule as a function of the duplex length, duplex structure (flap or blunt-end), and the applied voltage. The computational work investigated the ssDNA translocation using coarse-grained (CG) steered molecular dynamics (SMD) simulations. The influence of nucleotide length, ssDNA orientation, phosphate charges present on the DNA, sequence composition, and finally, SMD forces is explored at the molecular level via analyzing molecular interactions and conformational changes during translocation, which is not accessible through the nanopore experiments.

In Chapter 3, we presented our results on ssDNA translocation and dsDNA unzipping. The ssDNA translocation experiments demonstrated distinct orientation-dependent translocation times, with the 3' end consistently translocating faster than its 5' counterpart. This observation, in agreement with previous experimental studies, highlights the influence of strand orientation on translocation behavior and enables comparing experimental data with our computational simulations.

In the case of dsDNA unzipping, our experiments demonstrated that unzipping dynamics strongly depend on duplex structure, duplex length, and the applied voltage.

The experiments to study the effect of the duplex structure demonstrated that this factor influences the unzipping mechanism. Our results revealed two distinct unzipping timescales for blunt-end structures, with longer timescales corresponding to the unzipping events. In contrast, flap structures exhibited a single and always shorter unzipping times. Furthermore, the exponential dependence of unzipping time on the applied voltage supports a two-state unzipping process, yet the molecular details remain inaccessible through experimental methods.

We also observed that longer duplexes resulted in longer unzipping times, with blunt-end structures exhibiting an exponential dependence on duplex length, while flap structures showed a linear relationship. This linear dependence in flap structures shows the deviation from the two-state unzipping process and suggests intermediate states during unzipping, contributing to the overall unzipping time.

Our experimental results show that the unzipping process exceeds the predictions of the two-state theory. However, the molecular details of the process cannot be studied in nanopore experiments. Our results suggest that further investigation with improved theoretical models and simulations is necessary to fully understand the mechanisms underlying the unzipping of DNA molecules.

In Chapters 4 and 5, we demonstrated that the MARTINI coarse-grained (CG) model

is well-adapted for studying the molecular mechanisms of biomolecule transport through nanopores. By complementing experimental study with computational simulations, our study provided detailed insights into ssDNA translocation dynamics. CG approximation enabled us to perform multiple MD simulations over long timescales, which was important for gathering sufficient statistical data for exploring these translocation processes.

In Chapter 4, our CG simulations qualitatively reproduced key experimental findings, including the broad distribution of translocation times and the influence of ssDNA orientation on translocation dynamics. Specifically, the 3' end molecule translocated faster than the 5' end. Additionally, the CG MARTINI model was successful in predicting the electrostatic interactions between the phosphate groups of ssDNA and the nanopore's inner surface, particularly at the constriction.

In Chapter 5, we showed that sequence- and force-dependent translocation dynamics could be studied with the CG MARTINI model. Our results demonstrated that poly(dC) translocates faster than poly(dA), with velocity ratios aligning with experimental findings. Inter-base distance measurements further demonstrated that poly(dC) exhibited more conformational states during translocation than poly(dA). We also investigated the dependence of translocation velocity on the applied pulling force. Our CG results also demonstrated that higher pulling forces diminished the sequence-dependent differences in translocation velocity between poly(dA) and poly(dC).

These findings underscore the MARTINI CG model's potential to study biomolecular transport through nanopores, paving the way for future research involving other biomolecules, such as peptides, proteins, or polysaccharides, through various protein or solid-state nanopores.

In future research, it is important to address the limitations posed by experimental studies of unzipping, which only provide blocked current levels and unzipping times without offering molecular-level insights, such as the precise location where unzipping occurs or how the mechanisms differ depending on the duplex structure. Since the timescales in experiments have already reached several seconds for the longest duplexes, extending these studies beyond the current duplex lengths becomes impractical. Therefore, detailed molecular-level investigations, which could elucidate the unzipping mechanisms, must be pursued through MD simulations.

The CG-SMD simulations, as discussed in Chapters 4 and 5, have successfully reproduced key experimental findings for ssDNA translocation. This validation of the MARTINI CG model encourages us to explore the next step: unzipping dsDNA. However, the current MARTINI dsDNA model represents hydrogen bonds between complementary strands as covalent-like bonds that cannot be broken, which presents a limitation for simulating the unzipping process.

To simulate the unzipping between complementary DNA bases, more specific interactions representing the hydrogen bonds between adenine (A)-thymine (T) and guanine (G)-cytosine (C) must be introduced. A promising approach could be to employ polarizable coarse-grained (CG) beads to capture the polar nature of hydrogen bonding, as proposed by de Jong *et al.* for polarizable amino acids [170]. The reparametrization could differentiate between the stronger G-C pairs (three hydrogen bonds) and weaker A-T pairs (two hydrogen bonds). The MARTINI force field would need to incorporate new potentials to represent these hydrogen bonds while adjusting the elastic network to allow strand separation during the unzipping process.

Once the new parameters are developed, they need to be validated. The re-parameterized model should be tested on short sequences of dsDNA to check whether it accurately forms and breaks hydrogen bonds under mechanical pulling forces. Those simulations could be helpful in complementing the unzipping experiments by demonstrating the differences between the mechanisms governing the unzipping of the DNA duplexes used in this thesis. Additionally, this newly parameterized CG model for hydrogen bonding could be useful beyond DNA unzipping, extending its applicability to simulations of DNA origami, an important field in drug delivery.

In addition to dsDNA unzipping, our laboratory has recently expanded its focus on the translocation of polysaccharides, which is an emerging application of nanopore technology. Experiments conducted in our laboratory have demonstrated that the wild-type aerolysin nanopore can successfully detect glycosaminoglycan oligosaccharides (GAGs), distinguishing structural features such as sulfate patterns, osidic bonds, and epimers of uronic acid residues. Despite these advancements, there is still a need to decipher the fine molecular details of the discrimination of each constituent unit of GAGs at the single-molecule level. CG simulations of polysaccharide translocation could help address this challenge by shedding light on understanding the mechanisms that enable nanopores to sense these structural variations and encouraging the development of more precise mutant nanopores. The MARTINI force field has been extended to sugar molecules, and we have demonstrated that the ssDNA transport through protein nanopores can be successfully studied with this force field. Similarly, the application of this approach to polysaccharides would enable molecular-level insights that are not achievable through experiments.

This research not only enhances our understanding of DNA translocation and unzipping through nanopores but also opens up the possibility of extending CG MARTINI simulations to other biomolecules and nanopores.

Bibliography

[1] O V Krasilnikov, R Z Sabirov, V I Ternovsky, P G Merzliak, and B A Tashmukhamedov. The structure of staphylococcus aureus alpha-toxin-induced ionic channel. *Gen. Physiol. Biophys.*, 7(5):467–473, October 1988.

[2] O V Krasilnikov, R Z Sabirov, V I Ternovsky, P G Merzliak, and J N Muratkhodjaev. A simple method for the determination of the pore radius of ion channels in planar lipid bilayer membranes. *FEMS Microbiol. Lett.*, 105(1-3):93–100, September 1992.

[3] R Z Sabirov, O V Krasilnikov, V I Ternovsky, and P G Merzliak. Relation between ionic channel conductance and conductivity of media containing different nonelectrolytes. a novel method of pore size determination. *Gen. Physiol. Biophys.*, 12(2):95–111, April 1993.

[4] Sarah E Henrickson, Edmund A DiMarzio, Qian Wang, Vincent M Stanford, and John J Kasianowicz. Probing single nanometer-scale pores with polymeric molecular rulers. *J. Chem. Phys.*, 132(13):135101, April 2010.

[5] L Q Gu, M Dalla Serra, J B Vincent, G Vigh, S Cheley, O Braha, and H Bayley. Reversal of charge selectivity in transmembrane protein pores by using noncovalent molecular adapters. *Proc. Natl. Acad. Sci. U. S. A.*, 97(8):3959–3964, April 2000.

[6] Dvir Rotem, Lakmal Jayasinghe, Maria Salichou, and Hagan Bayley. Protein detection by nanopores equipped with aptamers. *J. Am. Chem. Soc.*, 134(5):2781–2787, February 2012.

[7] Douwe Jan Bonthuis, Jingshan Zhang, Breton Hornblower, Jérôme Mathé, Boris I. Shklovskii, and Amit Meller. Self-energy-limited ion transport in subnanometer channels. *Phys. Rev. Lett.*, 97:128104, September 2006.

[8] Amy E. Chavis, Kyle T. Brady, Grace A. Hatmaker, Christopher E. Angevine, Nuwan Kothalawala, Amala Dass, Joseph W. F. Robertson, and Joseph E. Reiner. Single molecule nanopore spectrometry for peptide detection. *ACS Sensors*, 2(9):1319–1328, September 2017.

[9] Fabien Piguet, Hadjer Ouldali, Manuela Pastoriza-Gallego, Philippe Manivet, Juan Pelta, and Abdelghani Oukhaled. Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nature Communications*, 9(1):966, March 2018.

[10] Chan Cao, Nuria Cirauqui, Maria Jose Marcaida, Elena Buglakova, Alice Duperrex, Aleksandra Radenovic, and Matteo Dal Peraro. Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores. *Nature Communications*, 10(1):4918, October 2019.

[11] Laura Restrepo-Pérez, Chirlmin Joo, and Cees Dekker. Paving the way to single-molecule protein sequencing. *Nature Nanotechnology*, 13(9):786–796, September 2018.

[12] Veerle Van Meervelt, Misha Soskine, Shubham Singh, Gea K. Schuurman-Wolters, Hein J. Wijma, Bert Poolman, and Giovanni Maglia. Real-time conformational changes and controlled orientation of native proteins inside a protein nanoreactor. *Journal of the American Chemical Society*, 139(51):18640–18646, December 2017.

[13] J J Kasianowicz, E Brandin, D Branton, and D W Deamer. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences*, 93(24):13770–13773, November 1996.

[14] Amit Meller, Lucas Nivon, Eric Brandin, Jene Golovchenko, and Daniel Branton. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 97(3):1079–1084, 2000.

[15] A Meller, L Nivon, and D Branton. Voltage-driven DNA translocations through a nanopore. *Phys Rev Lett*, 86(15):3435–3438, April 2001.

[16] David W Deamer and Daniel Branton. Characterization of nucleic acids by nanopore analysis. *Acc Chem Res*, 35(10):817–825, October 2002.

[17] James Clarke, Hai-Chen Wu, Lakmal Jayasinghe, Alpesh Patel, Stuart Reid, and Hagan Bayley. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.*, 4(4):265–270, 2009.

[18] David Stoddart, Andrew J Heron, Ellina Mikhailova, Giovanni Maglia, and Hagan Bayley. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *PNAS*, 106(19):7702–7707, April 2009.

[19] David Stoddart, Andrew J. Heron, Jochen Klingelhoefer, Ellina Mikhailova, Giovanni Maglia, and Hagan Bayley. Nucleobase recognition in ssdna at the central constriction of the α-hemolysin pore. *Nano Lett.*, 10(9):3633–3637, September 2010.

[20] Xiyun Guan, Li-Qun Gu, Stephen Cheley, Orit Braha, and Hagan Bayley. Stochastic sensing of TNT with a genetically engineered pore. *Chembiochem*, 6(10):1875–1881, October 2005.

[21] Parisa Bayat, Charlotte Rambaud, Bernard Priem, Matthieu Bourderioux, Mélanie Bilong, Salomé Poyer, Manuela Pastoriza-Gallego, Abdelghani Oukhaled, Jérôme Mathé, and Régis Daniel. Comprehensive structural assignment of glycosaminoglycan oligo- and polysaccharides by protein nanopore. *Nature Communications*, 13(1):5113, August 2022.

[22] Joseph W F Robertson, Claudio G Rodrigues, Vincent M Stanford, Kenneth A Rubinson, Oleg V Krasilnikov, and John J Kasianowicz. Single-molecule mass spectrometry in solution using a solitary nanopore. *Proc. Natl. Acad. Sci. U. S. A.*, 104(20):8207–8211, May 2007.

[23] Jiali Li, Derek Stein, Ciaran McMullan, Daniel Branton, Michael J. Aziz, and Jene A. Golovchenko. Ion-beam sculpting at nanometre length scales. *Nature*, 412(6843):166–169, July 2001.

[24] Peng Chen, Jiajun Gu, Eric Brandin, Young-Rok Kim, Qiao Wang, and Daniel Branton. Probing single dna molecule transport using fabricated nanopores. *Nano Letters*, 4(11):2293–2298, November 2004.

[25] Cees Dekker. Solid-state nanopores. *Nature Nanotechnology*, 2(4):209–215, April 2007.

[26] Bala Murali Venkatesan, Brian Dorvel, Sukru Yemenicioglu, Nicholas Watkins, Ivan Petrov, and Rashid Bashir. Highly sensitive, mechanically stable nanopore sensors for DNA analysis. *Adv Mater*, 21(27):2771, July 2009.

[27] Jiali Li and David S Talaga. The distribution of DNA translocation times in solid-state nanopores. *J Phys Condens Matter*, 22(45):454129, October 2010.

[28] S. Garaj, W. Hubbard, A. Reina, J. Kong, D. Branton, and J. A. Golovchenko. Graphene as a subnanometre trans-electrode membrane. *Nature*, 467(7312):190–193, September 2010.

[29] Meni Wanunu and Amit Meller. Chemically modified Solid-State nanopores. *Nano Lett.*, 7(6):1580–1585, June 2007.

[30] Adam R Hall, Andrew Scott, Dvir Rotem, Kunal K Mehta, Hagan Bayley, and Cees Dekker. Hybrid pore formation by directed insertion of α-haemolysin into solid-state nanopores. *Nat Nanotechnol*, 5(12):874–877, November 2010.

[31] Peijie Zheng, Chuntao Zhou, Yuemin Ding, Bin Liu, Liuyi Lu, Feng Zhu, and Shiwei Duan. Nanopore sequencing technology and its applications. *MedComm*, 4(4), July 2023.

[32] Yi-Lun Ying, Zheng-Li Hu, Shengli Zhang, Yujia Qing, Alessio Fragasso, Giovanni Maglia, Amit Meller, Hagan Bayley, Cees Dekker, and Yi-Tao Long. Nanopore-based technologies beyond dna sequencing. *Nature Nanotechnology*, 17(11):1136–1146, September 2022.

[33] Wei Si and Aleksei Aksimentiev. Nanopore sensing of protein folding. *ACS Nano*, 11(7):7091–7100, July 2017.

[34] Vera Arnaut, Martin Langecker, and Friedrich C. Simmel. Nanopore force spectroscopy of aptamer–ligand complexes. *Biophys J*, 105(5):1199–1207, September 2013.

[35] Eva Weatherall and Geoff R. Willmott. Applications of tunable resistive pulse sensing. *The Analyst*, 140(10):3318–3334, 2015.

[36] F. Caputo, J. Clogston, L. Calzolai, M. Rösslein, and A. Prina-Mello. Measuring particle size distribution of nanoparticle enabled medicinal products, the joint view of euncl and nci-ncl. a step by step approach combining orthogonal measurements with increasing complexity. *Journal of Controlled Release*, 299:31–43, 2019.

[37] Vineet Kumar and Praveen Guleria. Application of DNA-Nanosensor for environmental monitoring: Recent advances and perspectives. *Curr Pollut Rep*, pages 1–21, December 2020.

[38] F.M. Burnet. The exotoxins of staphylococcus pyogenes aureus. *The Journal of pathology and bacteriology*, 32(4):717–734, January 1930.

[39] F. M. Burnet. The production of staphylococcal toxin. *The Journal of Pathology and Bacteriology*, 33(1):1–16, 1930.

[40] Royal Commission of Inquiry into Fatalities at Bundaberg. *Report of the Royal Commission of Inquiry into Fatalities at Bundaberg, together with appendices*. Printed and published for the Government of the Commonwealth of Australia by H.J. Green, Government Printer, Canberra, 1928.

[41] Bryan J Berube and Juliane Bubeck Wardenburg. Staphylococcus aureus α-toxin: nearly a century of intrigue. *Toxins (Basel)*, 5(6):1140–1166, June 2013.

[42] L Song, M R Hobaugh, C Shustak, S Cheley, H Bayley, and J E Gouaux. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science*, 274(5294):1859–1866, 1996.

[43] B Walker, M Krishnasastry, L Zorn, and H Bayley. Assembly of the oligomeric membrane pore formed by staphylococcal alpha-hemolysin examined by truncation mutagenesis. *Journal of Biological Chemistry*, 267(30):21782–21786, 1992.

[44] Barbara Walker, Orit Braha, Stephen Cheley, and Hagan Bayley. An intermediate in the assembly of a pore-forming protein trapped with a genetically-engineered switch. *Chemistry & Biology*, 2(2):99–105, 1995.

[45] G Menestrina, M D Serra, and G Prévost. Mode of action of beta-barrel pore-forming toxins of the staphylococcal alpha-hemolysin family. *Toxicon*, 39(11):1661–1672, November 2001.

[46] Mauricio Montoya and Eric Gouaux. β-barrel membrane protein folding and structure viewed through the lens of α-hemolysin. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1609(1):19–27, 2003.

[47] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[48] Daniel Branton, David W. Deamer, Andre Marziali, Hagan Bayley, Steven A. Benner, Thomas Butler, Massimiliano Di Ventra, Slaven Garaj, Andrew Hibbs, Xiaohua Huang, Stevan B. Jovanovich, Predrag S. Krstic, Stuart Lindsay, Xinsheng Sean Ling, Carlos H. Mastrangelo, Amit Meller, John S. Oliver, Yuriy V. Pershin, J. Michael Ramsey, Robert Riehn, Gautam V. Soni, Vincent Tabard-Cossa, Meni Wanunu, Matthew Wiggin, and Jeffery A. Schloss. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.*, 26(10):1146–1153, October 2008.

[49] Murugappan Muthukumar. Mechanism of DNA transport through pores. *Annu. Rev. Biophys. Biomol. Struct.*, 36(1):435–450, 2007.

[50] Jérôme Mathé, Hasina Visram, Virgile Viasnoff, Yitzhak Rabin, and Amit Meller. Nanopore unzipping of individual dna hairpin molecules. *Biophys J*, 87(5):3205–3212, 2004.

[51] Steven J. Koch, Alla Shundrovsky, Benjamin C. Jantzen, and Michelle D. Wang. Probing protein-DNA interactions by unzipping a single DNA double helix. *Biophys J*, 83:1098–1105, 2002.

[52] Steven J. Koch and Michelle D. Wang. Dynamic force spectroscopy of protein-DNA interactions by unzipping DNA. *Phys Rev Lett*, 91:028103, 2003.

[53] Jie Jiang, Lin Bai, Jennifer A. Surtees, Zekeriya Gemici, Michelle D. Wang, and Eric Alani. Detection of high-affinity and sliding clamp modes for MSH2-MSH6 by single-molecule unzipping force analysis. *Mol Cell*, 20:771–781, 2005.

[54] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton. Unzipping kinetics of double-stranded dna in a nanopore. *Physical Review Letters*, 90(23):238101, 2003.

[55] Mekonnen Lemma Dechassa, Katharina Wyns, Ming Li, Michael A. Hall, Michelle D. Wang, and Karolin Luger. Structure and scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat Commun*, 2:313, 2011.

[56] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot. Molecular stick-slip motion revealed by opening DNA with piconewton forces. *Phys. Rev. Lett.*, 79:4489–4492, 1997.

[57] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot. DNA strand separation studied by single molecule force measurement. *Phys. Rev. E*, 58:2386–2394, 1998.

[58] Simona Cocco, Rémi Monasson, and John F. Marko. Force and kinetic barriers to unzipping of the dna double helix. *Proceedings of the National Academy of Sciences*, 98(15):8608–8613, 2001.

[59] J Muzard, M Martinho, J Mathé, U Bockelmann, and V Viasnoff. DNA translocation and unzipping through a nanopore: some geometrical effects. *Biophys J*, 98(10):2170–2178, May 2010.

[60] Céline Merstorf, Benjamin Cressiot, Manuela Pastoriza-Gallego, Abdel Ghani Oukhaled, Laurent Bacri, Jacques Gierak, Juan Pelta, Loïc Auvray, and Jérôme Mathé. DNA unzipping and protein unfolding using nanopores. *Methods Mol Biol*, 870:55–75, 2012.

[61] Mark Akeson, Daniel Branton, John J. Kasianowicz, Eric Brandin, and David W. Deamer. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single rna molecules. *Biophys J*, 77(6):3227–3233, 1999.

[62] Amit Meller and Daniel Branton. Single molecule measurements of DNA transport through a nanopore. *Electrophoresis*, 23(16):2583–2591, August 2002.

[63] Jérôme Mathé, Aleksei Aksimentiev, David R. Nelson, Klaus Schulten, and Amit Meller. Orientation discrimination of single-stranded dna inside the α-hemolysin membrane channel. *Proc. Natl. Acad. Sci. U. S. A.*, 102(35):12377–12382, August 2005.

[64] Nurit Ashkenasy, Jorge Sánchez-Quesada, Hagan Bayley, and M Reza Ghadiri. Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopores. *Angew Chem Int Ed Engl*, 44(9):1401–1404, February 2005.

[65] David Stoddart, Giovanni Maglia, Ellina Mikhailova, Andrew J Heron, and Hagan Bayley. Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew Chem Int Ed Engl*, 49(3):556–559, 2010.

[66] David Stoddart, Lorenzo Franceschini, Andrew Heron, Hagan Bayley, and Giovanni Maglia. DNA stretching and optimization of nucleobase recognition in enzymatic nanopore sequencing. *Nanotechnology*, 26(8):084002, February 2015.

[67] George Church, David Deamer, Daniel Branton, Robert Baldarelli, and John Kasianowicz. Characterization of individual polymer molecules based on monomer-interface interactions, 1998.

[68] Scott L Cockroft, John Chu, Manuel Amorin, and M Reza Ghadiri. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J Am Chem Soc*, 130(3):818–820, January 2008.

[69] John Chu, Marcos González-López, Scott L Cockroft, Manuel Amorin, and M Reza Ghadiri. Real-time monitoring of DNA polymerase function and stepwise single-nucleotide DNA strand translocation through a protein nanopore. *Angew Chem Int Ed Engl*, 49(52):10106–10109, December 2010.

[70] Kate R Lieberman, Gerald M Cherf, Michael J Doody, Felix Olasagasti, Yvette Kolodji, and Mark Akeson. Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *J Am Chem Soc*, 132(50):17961–17972, December 2010.

[71] Elizabeth A Manrao, Ian M Derrington, Andrew H Laszlo, Kyle W Langford, Matthew K Hopper, Nathaniel Gillgren, Mikhail Pavlenok, Michael Niederweis,

and Jens H Gundlach. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol*, 30(4):349–353, March 2012.

[72] Andrew H Laszlo, Ian M Derrington, Henry Brinkerhoff, Kyle W Langford, Ian C Nova, Jenny Mae Samson, Joshua J Bartlett, Mikhail Pavlenok, and Jens H Gundlach. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A*, 110(47):18904–18909, October 2013.

[73] Wenonah Vercoutere, Stephen Winters-Hilt, Hugh Olsen, David Deamer, David Haussler, and Mark Akeson. Rapid discrimination among individual dna hairpin molecules at single-nucleotide resolution using an ion channel. *Nature Biotechnology*, 19(3):248–252, March 2001.

[74] Yuhui He, Makusu Tsutsui, Ralph H. Scheicher, Chun Fan, Masateru Taniguchi, and Tomoji Kawai. Mechanism of how salt-gradient-induced charges affect the translocation of dna molecules through a nanopore. *Biophys J*, 105(3):776–782, 2013.

[75] Ralph M. M. Smeets, Ulrich F. Keyser, Diego Krapf, Meng-Yue Wu, Nynke H. Dekker, and Cees Dekker. Salt dependence of ion transport and dna translocation through solid-state nanopores. *Nano Letters*, 6(1):89–95, December 2005.

[76] Ranulu S S de Zoysa, D M Milan Krishantha, Qitao Zhao, Jyoti Gupta, and Xiyun Guan. Translocation of single-stranded DNA through the α-hemolysin protein nanopore in acidic solutions. *Electrophoresis*, 32(21):3034–3041, October 2011.

[77] Loredana Mereuta, Alina Asandei, Ioan Andricioaei, Jonggwan Park, Yoonkyung Park, and Tudor Luchian. Considerable slowdown of short DNA fragment translocation across a protein nanopore using ph-induced generation of enthalpic traps inside the permeation pathway. *Nanoscale*, 15(36):14754–14763, September 2023.

[78] Giovanni Maglia, Marcela Rincon Restrepo, Ellina Mikhailova, and Hagan Bayley. Enhanced translocation of single dna molecules through alpha-hemolysin nanopores by manipulation of internal charge. *Proceedings of the National Academy of Sciences*, 105(50):19720–19725, 2008.

[79] Marcela Rincon-Restrepo, Ellina Mikhailova, Hagan Bayley, and Giovanni Maglia. Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes. *Nano Lett*, 11(2):746–750, January 2011.

[80] H. Lu, F. Giordano, and Z. Ning. Oxford nanopore minion sequencing and genome assembly. *Genomics Proteomics Bioinformatics*, 14:265–279, 2016.

[81] M. Jain and et al. Improved data analysis for the minion nanopore sequencer. *Nature Methods*, 12:351–356, 2015.

[82] M. Jain, H. E. Olsen, B. Paten, and M. Akeson. The oxford nanopore minion: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17:239–249, 2016.

[83] Michael Faller, Michael Niederweis, and Georg E Schulz. The structure of a mycobacterial outer-membrane channel. *Science*, 303(5661):1189–1192, February 2004.

[84] Ian M. Derrington, Tom Z. Butler, Marcus D. Collins, Elizabeth Manrao, Mikhail Pavlenok, Michael Niederweis, and Jens H. Gundlach. Nanopore dna sequencing with mspa. *Proceedings of the National Academy of Sciences*, 107(37):16060–16065, 2010.

[85] Chan Cao, Meng-Yin Li, Nuria Cirauqui, Ya-Qian Wang, Matteo Dal Peraro, He Tian, and Yi-Tao Long. Mapping the sensing spots of aerolysin for single oligonucleotides analysis. *Nature Communications*, 9(1):2823, July 2018.

[86] Sander E Van der Verren, Nani Van Gerven, Wim Jonckheere, Richard Hambley, Pratik Singh, John Kilgour, Michael Jordan, E Jayne Wallace, Lakmal Jayasinghe, and Han Remaut. A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nat Biotechnol*, 38(12):1415–1420, July 2020.

[87] Tom Z. Butler, Mikhail Pavlenok, Ian M. Derrington, Michael Niederweis, and Jens H. Gundlach. Single-molecule dna detection with an engineered mspa protein nanopore. *Proceedings of the National Academy of Sciences*, 105(52):20647–20652, December 2008.

[88] Simon Finn Mayer, Chan Cao, and Matteo Dal Peraro. Biological nanopores for single-molecule sensing. *iScience*, 25(4):104145, 2022.

[89] David Deamer, Mark Akeson, and Daniel Branton. Three decades of nanopore sequencing. *Nat Biotechnol*, 34(5):518–524, May 2016.

[90] U. Bockelmann, P. Thomen, B. Essevaz-Roulet, V. Viasnoff, and F. Heslot. Unzipping dna with optical tweezers: high sequence sensitivity and force flips. *Biophys J*, 82:1537–1553, 2002.

[91] C Bustamante, S B Smith, J Liphardt, and D Smith. Single-molecule studies of DNA mechanics. *Curr Opin Struct Biol*, 10(3):279–285, June 2000.

[92] Olga K. Dudko, Jérôme Mathé, Attila Szabo, Amit Meller, and Gerhard Hummer. Extracting kinetics from single-molecule force spectroscopy: Nanopore unzipping of dna hairpins. *Biophys J*, 92(12):4188–4195, 2007.

[93] Sergei Yu Noskov, Wonpil Im, and Benoît Roux. Ion permeation through the alpha-hemolysin channel: theoretical studies based on brownian dynamics and Poisson-Nernst-Plank electrodiffusion theory. *Biophys J*, 87(4):2299–2309, October 2004.

[94] Qiong Zheng and Guo-Wei Wei. Poisson-Boltzmann-Nernst-Planck model. *J Chem Phys*, 134(19):194101, May 2011.

[95] D G Levitt. Interpretation of biological ion channel flux data–reaction-rate versus continuum theory. *Annu Rev Biophys Biophys Chem*, 15:29–57, 1986.

[96] D.G. Levitt. Modeling of ion channels. *The Journal of General Physiology*, 113(6):789–794, June 1999.

[97] N. Modi, M. Winterhalter, and U. Kleinekathöfer. Computational modeling of ion transport through nanopores. *Nanoscale*, 4(20):6166–6180, September 2012.

[98] Christopher Maffeo, Swati Bhattacharya, Jejoong Yoo, David Wells, and Aleksei Aksimentiev. Modeling and simulation of ion channels. *Chemical Reviews*, 112(12):6250–6284, December 2012.

[99] I. Cozmuta, J.T. O'Keeffe, and V. Stolc. Hybrid md-pnp simulations of the α-hemolysin open channel ionic current. *TechConnect Briefs*, 1:143–146, March 2004.

[100] Benoît Roux, Toby Allen, Simon Bernèche, and Wonpil Im. Theoretical and computational models of biological ion channels. *Q Rev Biophys*, 37(1):15–103, February 2004.

[101] Nikolai A. Simakov and Maria G. Kurnikova. Soft wall ion channel in continuum representation with application to modeling ion currents in α-hemolysin. *The Journal of Physical Chemistry B*, 114(46):15180–15190, November 2010.

[102] Ioana Cozmuta, James T. O'Keeffe, Deepak Bose, and Viktor Stolc. Hybrid md-nernst planck model of α-hemolysin conductance properties. *Molecular Simulation*, 31(2–3):79–93, February 2005.

[103] Witold Dyrka, Maciej M. Bartuzel, and Malgorzata Kotulska. Optimization of 3d poisson-nernst-planck model for fast evaluation of diverse protein channels: 3d pnp for diverse protein channels. *Proteins: Structure, Function, and Bioinformatics*, 81(10):1802–1822, August 2013.

[104] Pablo M De Biase, Suren Markosyan, and Sergei Noskov. Microsecond simulations of DNA and ion transport in nanopores with novel ion-ion and ion-nucleotides effective potentials. *J Comput Chem*, 35(9):711–721, April 2014.

[105] Pablo M. De Biase, Eric N. Ervin, Prithwish Pal, Olga Samoylova, Suren Markosyan, Michael G. Keehan, Geoffrey A. Barrall, and Sergei Yu. Noskov. What controls open-pore and residual currents in the first sensing zone of alpha-hemolysin nanopore? combined experimental and theoretical study. *Nanoscale*, 8:11571–11579, 2016.

[106] Kaifu Luo, Ilkka Huopaniemi, Tapio Ala-Nissila, and See-Chen Ying. Polymer translocation through a nanopore under an applied external field. *The Journal of Chemical Physics*, 124(11):114704, 03 2006.

[107] C. Millar, R. Madathil, O. Beckstein, et al. Brownian simulation of charge transport in α-haemolysin. *Journal of Computational Electronics*, 7:28–33, 2008.

[108] M. Muthukumar. Polymer translocation through a hole. *The Journal of Chemical Physics*, 111(22):10371–10374, December 1999.

[109] Jeffrey Chuang, Yacov Kantor, and Mehran Kardar. Anomalous dynamics of translocation. *Physical Review E*, 65(1), December 2001.

[110] Kaifu Luo, Tapio Ala-Nissila, See-Chen Ying, and Aniket Bhattacharya. Sequence dependence of DNA translocation through a nanopore. *Phys Rev Lett*, 100(5):058101, February 2008.

[111] Kaifu Luo, Tapio Ala-Nissila, See-Chen Ying, and Aniket Bhattacharya. Dynamics of DNA translocation through an attractive nanopore. *Phys Rev E Stat Nonlin Soft Matter Phys*, 78(6 Pt 1):061911, December 2008.

[112] Yacov Kantor and Mehran Kardar. Anomalous dynamics of forced translocation. *Phys. Rev. E*, 69:021806, February 2004.

[113] Mohammadreza Niknam Hamidabad and Rouhollah Haji Abdolvahab. Translocation through a narrow pore under a pulling force. *Scientific Reports*, 9(1):17885, November 2019.

[114] Ilkka Huopaniemi, Kaifu Luo, Tapio Ala-Nissila, and See-Chen Ying. Polymer translocation through a nanopore under a pulling force. *Phys. Rev. E*, 75:061912, June 2007.

[115] W. Sung and P. J. Park. Polymer translocation through a pore in a membrane. *Phys. Rev. Lett.*, 77:783–786, July 1996.

[116] Li-Zhen Sun and Meng-Bo Luo. Simulation study on the translocation and separation of copolymers. *Polymer*, 54(4):1448–1454, 2013.

[117] P. M. De Biase, C. J. Solano, S. Markosyan, L. Czapla, and S. Y. Noskov. Bromocd: Brownian dynamics/monte-carlo program suite to study ion and dna permeation in nanopores. *Journal of Chemical Theory and Computation*, 8(7):2540–2551, July 2012.

[118] A Bhattacharya, W H Morrison, K Luo, T Ala-Nissila, S-C Ying, A Milchev, and K Binder. Scaling exponents of forced polymer translocation through a nanopore. *Eur Phys J E Soft Matter*, 29(4):423–429, August 2009.

[119] Timothée Menais. Polymer translocation under a pulling force: Scaling arguments and threshold forces. *Phys. Rev. E*, 97:022501, February 2018.

[120] Jalal Sarabadani and Tapio Ala-Nissila. Theory of pore-driven and end-pulled polymer translocation dynamics through a nanopore: an overview. *Journal of Physics: Condensed Matter*, 30(27):274002, June 2018.

[121] Timo Ikonen, Jaeoh Shin, Wokyung Sung, and Tapio Ala-Nissila. Polymer translocation under time-dependent driving forces: Resonant activation induced by attractive polymer-pore interactions. *The Journal of Chemical Physics*, 136(20), May 2012.

[122] A. Fiasconaro, J. J. Mazo, and F. Falo. Active polymer translocation in the three-dimensional domain. *Phys. Rev. E*, 91:022113, February 2015.

[123] Ashutosh Dubey and Malay Bandyopadhyay. Polymer translocation across an oscillating nanopore: study of several distribution functions of relevant brownian functionals: Brownian functionals of polymer translocation across oscillating nanopore. *The European Physical Journal B*, 92(11), November 2019.

[124] Radu I. Stefureac, Anton Kachayev, and Jeremy S. Lee. Modulation of the translocation of peptides through nanopores by the application of an ac electric field. *Chem. Commun.*, 48:1928–1930, 2012.

[125] Daniel K. Lathrop, Eric N. Ervin, Geoffrey A. Barrall, Michael G. Keehan, Ryuji Kawano, Michael A. Krupka, Henry S. White, and Andrew H. Hibbs. Monitoring the escape of dna from a nanopore using an alternating current signal. *Journal of the American Chemical Society*, 132(6):1878–1885, January 2010.

[126] Aleksij Aksimentiev, Jiunn B Heng, Gregory Timp, and Klaus Schulten. Microscopic kinetics of DNA translocation through synthetic nanopores. *Biophys J*, 87(3):2086–2097, September 2004.

[127] Aleksij Aksimentiev and Klaus Schulten. Imaging α-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.*, 88(6):3745–3761, June 2005.

[128] Gianfranco Menestrina. Ionic channels formed by staphylococcus aureus alpha-toxin: Voltage-dependent inhibition by divalent and trivalent cations. *The Journal of Membrane Biology*, 90(2):177–190, June 1986.

[129] Zhao, Christina M. Payne, and Peter T. Cummings. Controlled translocation of dna segments through nanoelectrode gaps from molecular dynamics. *The Journal of Physical Chemistry C*, 112(1):8–12, January 2008.

[130] Richard M A Manara, Andrew T Guy, E Jayne Wallace, and Syma Khalid. Free-Energy calculations reveal the subtle differences in the interactions of DNA bases with α-Hemolysin. *J. Chem. Theory Comput.*, 11(2):810–816, 2015.

[131] Richard M.A. Manara, E. Jayne Wallace, and Syma Khalid. Dna sequencing with mspa: Molecular dynamics simulations reveal free-energy differences between sequencing and non-sequencing mutants. *Scientific Reports*, 5(1):12783, August 2015.

[132] Punam Rattu, Flo Glencross, Sophie L. Mader, Chris-Kriton Skylaris, Stephen J. Matthews, Sarah L. Rouse, and Syma Khalid. Atomistic level characterisation of ssdna translocation through the e. coli proteins csgg and csgf for nanopore sequencing. *Computational and Structural Biotechnology Journal*, 19:6417–6430, 2021.

[133] Hugh S. C. Martin, Shantenu Jha, Stefan Howorka, and Peter V. Coveney. Determination of free energy profiles for the translocation of polynucleotides through α-hemolysin nanopores using non-equilibrium molecular dynamics simulations. *J. Chem. Theory Comput.*, 5(8):2135–2148, 2009. PMID: 26613153.

[134] David B Wells, Volha Abramkina, and Aleksei Aksimentiev. Exploring transmembrane transport through alpha-hemolysin with grid-steered molecular dynamics. *J Chem Phys*, 127(12):125101, 2007.

[135] Hugh S C Martin, Shantenu Jha, and Peter V Coveney. Comparative analysis of nucleotide translocation through protein nanopores using steered molecular dynamics and an adaptive biasing force. *J Comput Chem*, 35(9):692–702, January 2014.

[136] Peter J. Bond, Andrew T. Guy, Andrew J. Heron, Hagan Bayley, and Syma Khalid. Molecular dynamics simulations of dna within a nanopore: Arginine-phosphate tethering and a binding/sliding mechanism for translocation. *Biochemistry*, 50(18):3777–3783, 2011. PMID: 21428458.

[137] Andrew T Guy, Thomas J Piggot, and Syma Khalid. Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophys J*, 103(5):1028–1036, 2012.

[138] Helgi I Ingólfsson, Cesar A Lopez, Jaakko J Uusitalo, Djurre H de Jong, Srinivasa M Gopal, Xavier Periole, and Siewert J Marrink. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip Rev Comput Mol Sci*, 4(3):225–248, May 2014.

[139] A. Ramachandran, Q. Guo, S. M. Iqbal, and Y. Liu. Coarse-grained molecular dynamics simulation of dna translocation in chemically modified nanopores. *The Journal of Physical Chemistry B*, 115(19):6138–6148, May 2011.

[140] Pablo D. Dans, Ari Zeida, Matías R. Machado, and Sergio Pantano. A coarse grained model for atomic-detailed dna simulations with explicit electrostatics. *Journal of Chemical Theory and Computation*, 6(5):1711–1725, May 2010.

[141] Anna Stachiewicz and Andrzej Molski. A coarse-grained MARTINI-like force field for DNA unzipping in nanopores. *J Comput Chem*, 36(13):947–956, February 2015.

[142] Anna Stachiewicz and Andrzej Molski. Diffusive dynamics of dna unzipping in a nanopore. *Journal of Computational Chemistry*, 37(5):467–476, 2016.

[143] Anna Stachiewicz and Andrzej Molski. Sequence-dependent unzipping dynamics of DNA hairpins in a nanopore. *The Journal of Physical Chemistry B*, 123(15):3199–3209, March 2019.

[144] Jeffrey Comer, Valentin Dimitrov, Qian Zhao, Gregory Timp, and Aleksei Aksimentiev. Microscopic mechanics of hairpin DNA translocation through synthetic nanopores. *Biophys J*, 96(2):593–608, January 2009.

[145] Nathalie Basdevant, Delphine Dessaux, and Rosa Ramirez. Ionic transport through a protein nanopore: a coarse-grained molecular dynamics study. *Sci. Rep.*, 9(1):15740, 2019.

[146] Delphine Dessaux. *Expériences et modélisation par dynamique moléculaire gros-grain du transport de l'ADN à travers l'alpha-hémolysine.* PhD thesis, 2020. 2020UPASE002.

[147] Delphine Dessaux, Jérôme Mathé, Rosa Ramirez, and Nathalie Basdevant. Current rectification and ionic selectivity of α-hemolysin: Coarse-grained molecular dynamics simulations. *J. Phys. Chem. B*, 126(23):4189–4199, June 2022.

[148] Bertil Hille. *Ion Channels of Excitable Membranes*, volume 18. 01 2001.

[149] Hasitha Muthumala Waidyasooriya, Masanori Hariyama, and Kota Kasahara. An fpga accelerator for molecular dynamics simulation using opencl. *International Journal of Networked and Distributed Computing*, 5:52–61, 2017.

[150] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[151] Ulrich Essmann, Lalith Perera, Max Berkowitz, Tom Darden, Hsing Lee, and Lee Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577, November 1995.

[152] Daan Frenkel and Berend Smit. Understanding molecular simulation. In Daan Frenkel and Berend Smit, editors, *Understanding Molecular Simulation, From Algorithms to Applications (Second Edition)*. Academic Press, San Diego, second edition edition, 2002.

[153] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, volume 1 of *Computational Science Series*. Academic Press, San Diego, second edition, 2002.

[154] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, July 1967.

[155] J. Schlitter, M. Engels, and P. Krüger. Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. *Journal of Molecular Graphics*, 12(2):84–89, June 1994.

[156] Jeffrey Comer, James C. Gumbart, Jérôme Hénin, Tony Lelièvre, Andrew Pohorille, and Christophe Chipot. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *The Journal of Physical Chemistry B*, 119(3):1129–1151, October 2014.

[157] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, April 1997.

[158] Siewert J. Marrink and D. Peter Tieleman. Perspective on the martini model. *Chem. Soc. Rev.*, 42:6801–6822, 2013.

[159] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 01 2000.

[160] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd - visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[161] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The martini force field: Coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007. PMID: 17569554.

[162] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008. PMID: 26621095.

[163] Cesar A. López, Andrzej J. Rzepiela, Alex H. de Vries, Lubbert Dijkhuizen, Philippe H. Hünenberger, and Siewert J. Marrink. Martini coarse-grained force field: Extension to carbohydrates. *Journal of Chemical Theory and Computation*, 5(12):3195–3210, 2009. PMID: 26602504.

[164] Antti-Pekka Hynninen, James F. Matthews, Gregg T. Beckham, Michael F. Crowley, and Mark R. Nimlos. Coarse-grain model for glucose, cellobiose, and cellotetraose in water. *Journal of Chemical Theory and Computation*, 7(7):2137–2150, 2011. PMID: 26606485.

[165] Jaakko J. Uusitalo, Helgi I. Ingólfsson, Parisa Akhshi, D. Peter Tieleman, and Siewert J. Marrink. Martini coarse-grained force field: Extension to dna. *Journal of Chemical Theory and Computation*, 11(8):3932–3945, 2015. PMID: 26574472.

[166] Siewert J Marrink, Alex H de Vries, and Alan E Mark. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B*, 108(2):750–760, 2004.

[167] Hwankyu Lee, Alex H de Vries, Siewert-Jan Marrink, and Richard W Pastor. A coarse-grained model for polyethylene oxide and polyethylene glycol: conformation and hydrodynamics. *J Phys Chem B*, 113(40):13186–13194, October 2009.

[168] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable water model for the coarse-grained martini force field. *PLOS Computational Biology*, 6:1–17, 06 2010.

[169] Yin Wang, Paraskevi Gkeka, Julian E Fuchs, Klaus R Liedl, and Zoe Cournia. DPPC-cholesterol phase diagram using coarse-grained molecular dynamics simulations. *Biochim Biophys Acta*, 1858(11):2846–2857, August 2016.

[170] Djurre H. de Jong, Gurpreet Singh, W. F. Drew Bennett, Clement Arnarez, Tsjerk A. Wassenaar, Lars V. Schäfer, Xavier Periole, D. Peter Tieleman, and Siewert J. Marrink. Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation*, 9(1):687–697, 2013. PMID: 26589065.

[171] Xavier Periole, Marco Cavalli, Siewert-Jan Marrink, and Marco A. Ceruso. Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition. *Journal of Chemical Theory and Computation*, 5(9):2531–2543, 2009. PMID: 26616630.

[172] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[173] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, July 2003.

[174] J. Mathé, A. Arinstein, Y. Rabin, and A. Meller. Equilibrium and irreversible unzipping of dna in a nanopore. *Europhysics Letters*, 73(1):128, November 2005.

[175] Murugappan Muthukumar, Calin Plesa, and Cees Dekker. Single-molecule sensing with nanopores. *Physics Today*, 68(8):40–46, 08 2015.

[176] Cagla Okyay, Delphine Dessaux, Rosa Ramirez, Jérôme Mathé, and Nathalie Basdevant. Exploring ssdna translocation through α-hemolysin using coarse-grained steered molecular dynamics. *Nanoscale*, 16:15677–15689, July 2024.

[177] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[178] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[179] Jaakko J. Uusitalo, Helgi I. Ingólfsson, Parisa Akhshi, D. Peter Tieleman, and Siewert J. Marrink. Martini coarse-grained force field: Extension to dna. *J. Chem. Theory Comput.*, 11(8):3932–3945, August 2015.

[180] water_deletor.pl. http://www.mdtutorials.com/gmx/membrane_protein/Files/water_deletor.pl.

[181] Ulrich F Keyser, Bernard N Koeleman, Stijn van Dorp, Diego Krapf, Ralph M M Smeets, Serge G Lemay, Nynke H Dekker, and Cees Dekker. Direct force measurements on DNA in a solid-state nanopore. *Nature Physics*, 2(7):473–477, July 2006.

[182] Taylor Haynes, Iain P. S. Smith, E. Jayne Wallace, Jemma L. Trick, Mark S. P. Sansom, and Syma Khalid. Electric-field-driven translocation of ssdna through hydrophobic nanopores. *ACS Nano*, 12(8):8208–8213, August 2018.

[183] Ining Jou and Murugappan Muthukumar. Effects of nanopore charge decorations on the translocation dynamics of DNA. *Biophys J*, 113(8):1664–1672, October 2017.

[184] D K Lubensky and D R Nelson. Driven polymer translocation through a narrow pore. *Biophys J*, 77(4):1824–1838, October 1999.

[185] Arnold J. Storm, Cornelis Storm, Jianghua Chen, Henny Zandbergen, Jean-François Joanny, and Cees Dekker. Fast dna translocation through a solid-state nanopore. *Nano Letters*, 5(7):1193–1197, July 2005.

[186] Robert F. Purnell and Jacob J. Schmidt. Discrimination of single base substitutions in a dna strand immobilized in a biological nanopore. *ACS Nano*, 3(9):2533–2538, September 2009.

[187] Amit Meller. Dynamics of polynucleotide transport through nanometre-scale pores. *Journal of Physics: Condensed Matter*, 15(17):R581, April 2003.

[188] Linda Payet, Marlène Martinho, Céline Merstorf, Manuela Pastoriza-Gallego, Juan Pelta, Virgile Viasnoff, Loïc Auvray, Murugappan Muthukumar, and Jérôme Mathé. Temperature effect on ionic current and ssdna transport through nanopores. *Biophys J*, 109(8):1600–1607, 2015.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences Chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)

Titre : Etude expérimentale et en modélisation de dynamique moléculaire d'un acide nucléique en nano-confinement

Mots clés : MD, gros-grain, α-hémolysine, translocation d'ADN, dézippage d'ADN

Résumé : La technologie des nanopores s'est imposée comme un outil puissant pour étudier le transport biomoléculaire, en particulier pour la translocation et le dézippage des molécules d'ADN. Les études expérimentales ont montré la capacité des nanopores de l'α-hémolysine (αHL) à distinguer différentes séquences et orientations d'ADN. Cependant, les résultats expérimentaux fournissent principalement des informations sur le courant bloqué et le temps de translocation, laissant les détails au niveau moléculaire du processus de dézippage inexplorés. Bien que les simulations de dynamique moléculaire tout-atome soient informatives, elles sont limitées par des échelles de temps réduites. En revanche, les simulations de dynamique moléculaire à gros-grains utilisant le champ de force MARTINI permettent l'étude du transport de l'ADN sur des échelles de temps plus longues, se rapprochant ainsi de celles observées expérimentalement.

Cette thèse explore les dynamiques de translocation de l'ADN simple-brin et de dézippage de l'ADN double-brin à travers le nanopore αHL à l'aide d'approches expérimentales et de simulations de dynamique moléculaire dirigées (SMD) à gros-grains. Les différences de temps de translocation entre les extrémités 3' et 5' de l'ADN simple-brin et les temps de dézippage de l'ADN double-brin dans différentes conditions, telles que la structure du duplex d'ADN et la tension appliquée, ont été observées dans les études expérimentales. En particulier, nous avons mesuré des temps de dézippage distincts pour les molécules d'ADN double-brin utilisées, et il a été observé que la dépendance du temps de dézippage à la tension appliquée suivait une loi exponentielle. À mesure que la longueur du duplex augmente, les mécanismes semblent changer en fonction de la structure du duplex. Cependant, les raisons derrière les comportements de translocation et de dézippage restent inaccessibles expérimentalement.

En utilisant des simulations de dynamique moléculaire gros-grains, l'influence de l'orientation de l'ADN simple-brin, de la composition en séquences et de la force appliquée sur les dynamiques de translocation a été examinée de manière computationnelle. Nos résultats de simulation ont reproduit les principales observations expérimentales, telles que la large distribution des temps de translocation, les comportements de translocation dépendants de l'orientation, le rôle crucial des interactions électrostatiques entre l'ADN et le nanopore, soulignant l'impact des charges des phosphates de l'ADN sur les taux de translocation, et les dynamiques de translocation dépendantes des séquences sous des forces appliquées variables. En particulier, le rapport entre les temps de translocation des bases puriques et pyrimidiques a également été en bon accord avec les résultats expérimentaux. À la suite des simulations gros-grains, une relation non linéaire entre la vitesse de translocation et la force appliquée a été observée. De plus, les différences de conformations de l'ADN à l'intérieur du nanopore ont apporté des explications supplémentaires aux comportements de translocation dépendants de la séquence et de l'orientation.

Cette étude valide le modèle MARTINI à gros-grains comme un outil efficace pour l'étude du transport de l'ADN, montrant sa capacité à compléter les travaux expérimentaux. Nos résultats suggèrent que les simulations de MD gros-grains sont bien adaptées pour dévoiler les mécanismes moléculaires du dézippage de l'ADN, offrant des perspectives inaccessibles par les techniques expérimentales actuelles.

Title: Experimental study and Molecular Dynamics (MD) modeling of a nucleic acid in nano-confinement
Keywords: MD, coarse-grained, α-hemolysin, DNA translocation, DNA unzipping

Abstract: Nanopore technology has emerged as a powerful tool for studying biomolecular transport, particularly for the translocation and unzipping of DNA molecules. Experimental studies have shown the ability of α-hemolysin (αHL) nanopores to distinguish between different DNA sequences and orientations. However, experimental results primarily provide blocked current and translocation time information, leaving molecular-level details of the unzipping process unexplored. All-atom molecular dynamics (MD) simulations, though informative, are limited by short time scales. Coarse-grained (CG) MD simulations using the MARTINI force field, on the other hand, enable the study of DNA transport over extended time scales, approaching those observed experimentally.

This thesis investigates the dynamics of both ssDNA translocation and dsDNA unzipping through the αHL nanopore using a combination of experimental techniques and CG-steered MD (SMD) simulations. Experimental studies explored the translocation times of ssDNA at the 3' and 5' ends, as well as the unzipping times of dsDNA under various conditions, including different duplex structures and applied voltages. Our findings on ssDNA translocation aligned with previous experimental and theoretical results, confirming faster translocation of 3' oriented ssDNA. Additionally, distinct unzipping times were observed for the different duplex structures under identical experimental conditions, with an exponential relationship noted between unzipping time and applied voltage. As the duplex length increased, the unzipping mechanisms appeared to vary depending on the duplex structure. However, the underlying molecular mechanisms behind these translocation and unzipping behaviors remain experimentally inaccessible, highlighting the need for further theoretical studies.

By employing CG MD simulations, the influence of ssDNA orientation, sequence composition, and pulling force on translocation dynamics were computationally examined. Our simulation results reproduced the key experimental findings, such as the wide distribution of the translocation times, the orientation-dependent translocation behaviors, the crucial role of electrostatic interactions between DNA and the nanopore, highlighting the impact of DNA phosphate charges on translocation rates, and the sequence-dependent translocation dynamics under varying applied forces. Specifically, the ratio of translocation times of purine and pyrimidine bases was also found to be in good agreement with the experimental findings. As a result of the CG simulations, a non-linear relationship between translocation velocity and the applied force was observed. Additionally, differences in DNA conformations inside the nanopore provided additional explanation for the sequence- and orientation-dependent translocation behaviors.

This study validates the MARTINI CG model as an effective tool for investigating DNA transport, demonstrating its ability to complement experimental data. Our findings suggest that CG MD simulations are well suited for uncovering the molecular mechanisms underlying DNA unzipping, offering insights that are otherwise inaccessible through current experimental techniques.

Maison du doctorat de Université Paris-Saclay
2^{e} étage, aile ouest, École normale supérieure Paris-Saclay
4 avenue des Sciencs
91190 Gif-sur-Yvette, France