Chapter 5

Stability of DNA passing through different geometrical pores

5.1 Introduction

The recent years witnessed significant advances in nanopore technology and the DNA sequencing. One of the efficient methods for DNA sequencing is denaturation mapping of the DNA molecule that is trapped inside a cylindrical geometry. In this chapter, we investigate the denaturation of homogeneous as well as heterogeneous DNA molecules which are confined in a conical as well as cylindrical geometric space. For the conically shaped confinement, we study the effect of the cone angle (θ) and the pore width (δ) on the melting profile of homogeneous DNA molecules. Similarly, for the cylindrically shaped confinement, we investigate the effect of cylinder diameter on the melting profile of homogeneous and heterogeneous DNA molecules. We consider the conditions when DNA is partially inside the confined space and partially outside the confined space. For the investigation, we vary the fraction of base pairs (φ) that remains inside the space and calculate the denaturation profile of DNA molecules in each condition.

The conventional *DNA mapping* scheme, which uses enzyme-based labelling, is quite expensive and exhaustive. A new concept of *denaturation mapping*, in comparision, is simple and affordable [125, 126]. The process, in which due to a thermal fluctuation the hydrogen bonds between the two bases in a pair on the opposite strands disintegrate, which leads to the separation of the two strands, is known as *thermal melting* of DNA. The method of *denaturation mapping* basically works on the fact that the GC rich region requires higher amount of energy than the AT rich region to melt. Reisner *et al* have predicted the DNA sequence through the *optical*

mapping of the denaturating DNA that is confined in a rectangular nanochannel [127]. By combining the experimental method with the computer simulation based on Poland-Scheraga model [128], which evaluates the sequence dependent melting probabilities, this method offers a new horizon to predict the genome sequence. In the current work, we extend these pioneering works and attempt to understand the melting of DNA that is fully or partially confined in a nanochannel. We consider melting of homogeneous DNA in two different geometries: (a) conical, (b) cylindrical. The motivation behind choosing two different geometries is the work related to translocation of DNA through the nanopores of different geometries. The nanopores are used widely to study the biopolymers like proteins and nucleic acids [79, 94–98, 189]. In general there are two kinds of nanopores which are used to decode the genome sequence: biotechnological or solid state.

The lipid-embedded α-hemolysin channels [99, 100, 104, 106] are used as model nanopores to explore the secondary structures of nucleic acids by using the electrical force to unzip duplex regions [107, 108]. The major shortcomings that hinders prospective biotechnological nanopore applications are the fragility of these lipid bilayers and lateral diffusion of these channels in the membrane [109, 110]. One of the recent development in this field is the use of solid-state nanopores [111–115] for sensing purpose as they are better than biological nanopores [116, 117] regarding high stability, easy controllability for size, adjustable surface properties and the potential for their integration into devices and arrays [118, 119]. DNA unzipping through these nanopores have applications in the field of clinical and laboratory use of nanopore sensors by integrating the solid-state nanopores into an all-in-one bioanalytical device. For solid-state nanopores, we still need more studies that show the current patterns corresponding to the complete unzipping process of DNA duplexes.

The current chapter investigate the melting of DNA that is confined in a nanochannel of two different geometries. The content of the findings are discussed under different sections, with a brief introduction to the statistical model. We discuss the melting profile of a DNA molecule that is partially or fully confined in a conical and cylindrical geometries in the third section. The chpater briefly presents the results and future prospects of the current work in the last section.

5.2 The mathematical model

In this chapter, we investigate the stability of dsDNA through these nanopores and explore the effect of these nanopores on the melting profile of dsDNA. We use the Peyrard Bishop and Dauxois model (PBD) model [64] and modify it to study the unzipping pattern of dsDNA through the nanopores. The essential components of the linear PBD models are the on-site potential between the bases in a pair, stacking potential along the DNA strand. The hamiltonian is

$$H = \sum_{i=1}^{N} \left[\frac{p_i^2}{2m} + V_M(y_i) \right] + \sum_{i=1}^{N-1} \left[V_s(y_i, y_{i+1}) \right], \tag{5.1}$$

The stacking interaction between two consecutive base pairs along the chain is $V_s(y_i, y_{i+1}) = \frac{k}{2}(y_i - y_{i+1})^2[1 + \rho \exp(-b(y_i + y_{i+1}))]$, where k represents the single strand elasticity. The anharmonicity in the strand elasticity is represented by ρ while b represents its range. For our studies, we tweaked the model parameters and found the values of $k = 0.02$ eVÅ^{-2} , $\rho = 50.0$, $b = 0.35$ Å^{-1} as suitable values for the current investigation. The hydrogen bonding between the two bases in the i^{th} pair is represented by the Morse potential, $V_M(y_i) = D_i[\exp(-a_i y_i) - 1]^2$ where D_i represents the potential depth, and a_i represents the inverse of the width of the potential well. These two parameters have a crucial role in DNA denaturation. The potential parameters are taken as $a = 4.2$ Å^{-1}, and $D_0 = 0.075$ eV. Thermodynamics of the transition can be investigated by evaluating the expression for the partition function. For a sequence of N base pairs, the canonical partition function can be written as, $Z = \int \prod_{i=1}^{N} \{dy_i dp_i \exp(-\beta H)\} = Z_p Z_c$, where Z_p corresponds to the momentum part of the partition function which is equal to $(2\pi m k_B T)^{N/2}$. The configurational part of the partition function, Z_c can be computed using transfer integral or by matrix multiplication method [174]. Thus, the configurational part of the partition function is, $Z_c = \int \prod_{i=1}^{N-1} dy_i K(y_i, y_{i+1})] dy_N K(y_N)$, where, $K(y_i, y_{i+1}) = \exp[-\beta H_c(y_i, y_{i+1})]$. For the homogeneous chain, one can evaluate the partition function by transfer integral (TI) method by applying the periodic boundary condition. In the case of a heterogeneous chain, with an open boundary, the configurational part of the partition function can be integrated numerically with the help of the matrix multiplication method [171, 186]. A point to note is that the nature of Morse potential in the PBD model leads to the divergence of the partition function. To overcome this problem, an upper cut-off for the integration needs to set up [140] or or some other way of confining the DNA molecule needs

to be applied. Approaches to how to do this has been discussed by several groups [140, 145, 146, 190]. In our previous study we use an integration limit of the integration from -5.0 Å to 200 Å. Once the limit of integration has been chosen, the task is reduced to discretize the space to evaluate the integral numerically. The space is discretized using the Gaussian quadrature formula. In our previous studies [174], we observed that in order to get a precise value for melting temperature (T_m) one has to choose large grid points. We found that 900 is a sufficient number for this purpose. As all matrices in the equation are identical in nature, the multiplication is done very efficiently.

5.3 DNA in the confined geometries

5.3.1 DNA confined in a conical geometry

How the confinement can be realized in the PBD model? We adopt the following scheme. Due to the surrounding cellular environment, the movements of base pairs are restricted which in turn affects the overall movements of the individual base pairs. We restrict the configuration space of the system as shown in Fig. 5.1a. Since the confining walls do not affect the movement below the equilibrium separation, we do not change the lower limit of integration. The same is not true for the upper limit of integration. The confining walls restrict the upper limit of the fluctuation, hence the upper limit of the integration needs some alteration. We modify the upper limit of integration for each base pair as x. The value of x changes with the angular separation and the the distance from the vertex as, $x = 3.4 \times i \times \tan(\theta)$. Here, i is the site index ($i = 1,2,3...,N$) and N is the total number of base pairs in the chain. The angular separation between the confining wall and the DNA strand is denoted by a symbol, θ. Since two base pairs are approximately at a distance of 3.4 Å, we take this number as the multiplication factor in our calculation. Thus, the configurational space for our calculation extends from -5 Å to x Å. The parameter δ represents the difference between the diameter of the pore and the diameter of DNA(we are defining that as pore width). In our calculations, we vary δ from 0 - 30 Å. With these modifications, we calculate the partition function and hence evaluate the thermodynamic properties of the system. First, we consider the DNA (of 100 & 200 base pairs) that is confined and translocated through a conical geometry and evaluate the thermodynamic properties of the system for the parameters, $\delta = 0$Å and the angular separation, $\theta = 20°$. When

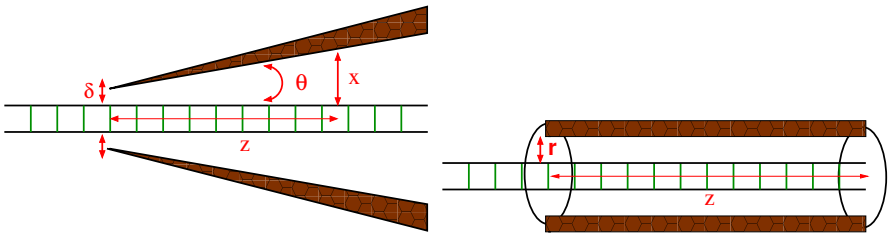


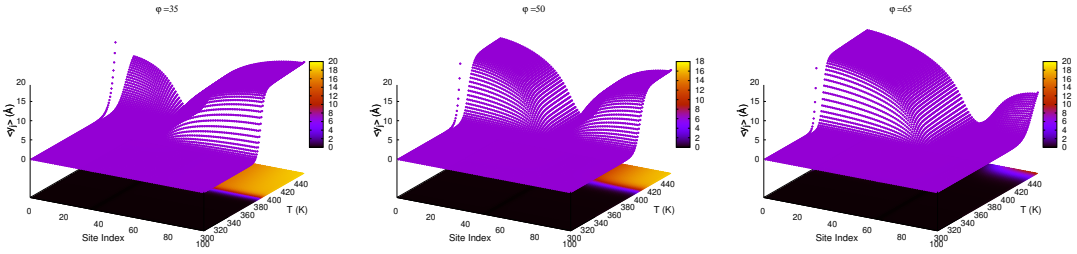
Figure 5.1: The schematic representation of the DNA chain confined in a cell. (a) Here, we schematically show a conically confined DNA molecule. The distance x represents the gap between a pair and the confining wall while θ is the angle at which the cellular wall is enclosing the DNA. Here $z = 3.4i$,the distance along x-axis. The parameter, δ, represents the pore width. (b) Here, we schematically show a cylindrically confined DNA molecule. The r is the distance of the confined wall from the DNA strand. So the radius of the cylinder becomes r+DNA radius.

DNA translocates through a nanopore, the fraction of the chain that stays inside the pore changes. We denote this fraction by φ. If the site index (or base pair) at the vertex of the cone is 30, the value of $\varphi = 30$. This means that 70 base pairs are inside the pore while 29 base pairs are outside the pore. We change the value of φ from 10 to 90 and study the melting profile of the chain in each case. We study the melting profie of the chain through the calculation of the average separation of the base pairs with temperature. The average separation of the j^{th} pair of the chain with temperature is,

$$\langle y_j \rangle = \frac{1}{Z_c} \int \prod_{i=1}^{N} y_j \exp(-\beta H_c) dy_i \tag{5.2}$$

The opening profie of the DNA chain for three different values of φ are shown in Fig.5.2. For this problem, there is a difficulty to define the standard *melting temperature* of the system. Hence, we define the temperature, above which more than 50% of the base pairs are open, as *the transition temperature*, T_c. By calculating the number of open pairs, we define a *transition temperature*, T_c of the system. We choose a thresold value of 1.0 Å, if $\langle y_j \rangle \geq 1.0$ Å, we call the base pair as *open* [173]. We count how many base pairs are in open state at a particular temperature as shown in Table 5.1. The opening of the chain of 100 base pairs with temperature is shown with the pm3d plot step by step for overall visualization in Fig.5.3 and the counting table is shown in Table 5.2

Upon further investigations, we find that as the chain moves from restricted to the unrestricted area, the stability of the chain increases. There is an increase



(a) The base pair that is at the vertex of thecone is 35 ($\varphi = 35$), and hence inside the cone there are 65 base pairs while on the other side (out of the cone) there are 34 base pairs.

(b) The base pair that is at the vertex of the cone is 50 ($\varphi = 50$), and hence inside the cone there are 50 base pairs while on the other side (out of the cone) there are 49 base pairs.

(c) The base pair that is at the vertex of the cone is 65 ($\varphi = 65$), and hence inside the cone there are 35 base pairs while on the other side (out of the cone) there are 64 base pairs.

Figure 5.2: The opening of the chain of 100 base pairs with temperature. The conical confined geometry makes the angle $\theta = 20°$ and $\delta = 0$Å.

Table 5.1: For $\varphi = 35, 50, 65$, we calculate the number of pairs that are open on the confined side (inside the cell) and the free side (outside the cell).

$\varphi = 35$						
Temperature (K)	377	378	409	424	427	450
Confined	0	36	60	61	61	61
Free	0	0	1	10	22	32
$\varphi = 50$						
Temperature (K)	382	383	393	394	403	450
Confined	0	0	0	18	43	46
Free	0	1	24	32	45	47
$\varphi = 65$						
Temperature (K)	368	369	377	389	427	450
Confined	0	0	0	0	12	30
Free	0	1	34	61	62	62

Figure 5.3: The opening of the chain of 100 base pairs with temperature. Here we consider the separation between the cone vortex and the chain i.e $\delta = 0$, and $\theta = 20°$.

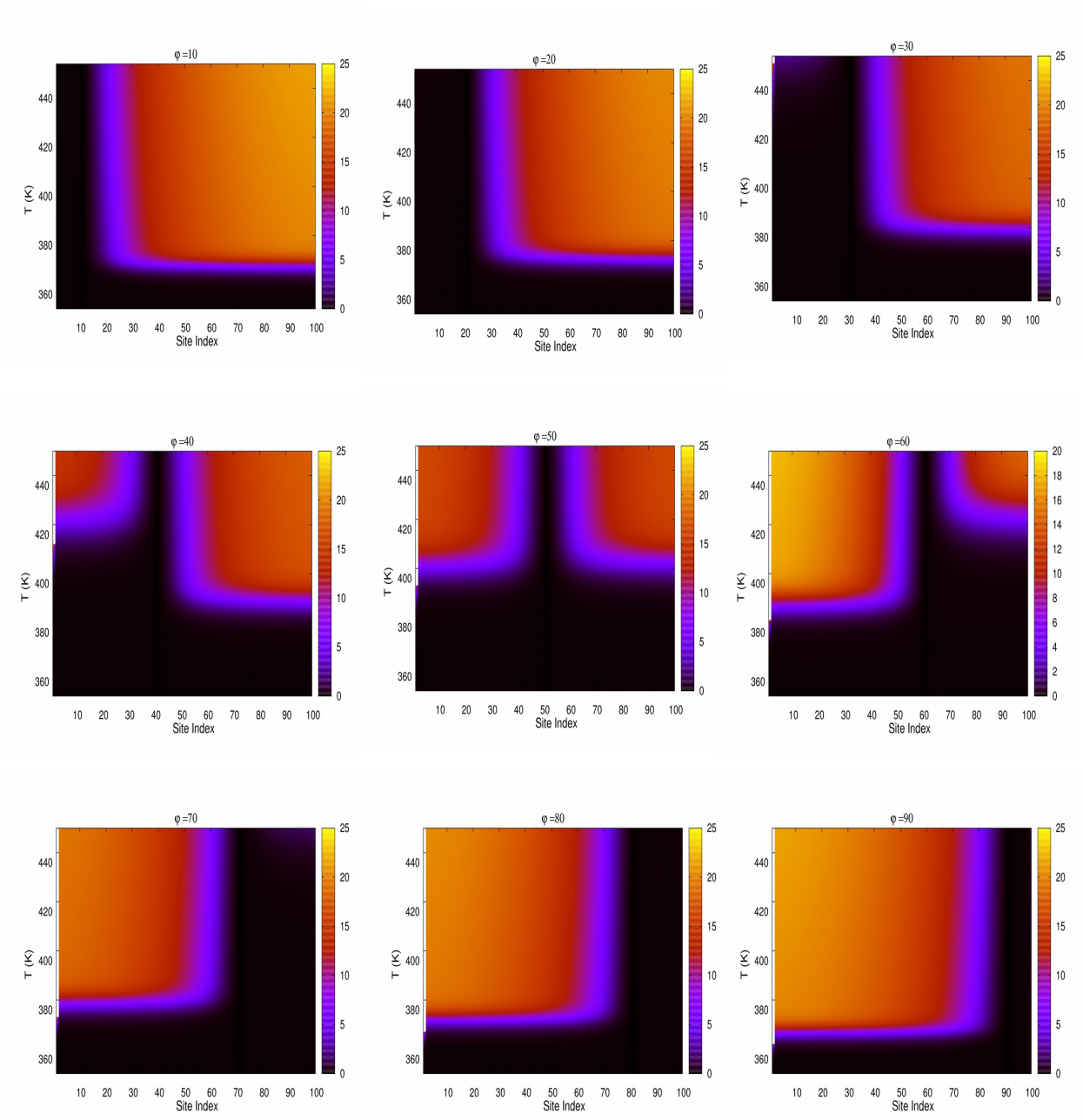


Table 5.2: Counting Table of Open base pairs: $y_i > 1\mathring{A}$, is taken as open pair : we calculate the number of pairs that are open on the confined side (inside the cell) and free side (outside the cell) for $\delta = 0$ and $\theta = 20°$.

	$\varphi = 20$		$\varphi = 35$		$\varphi = 50$		$\varphi = 65$		$\varphi = 80$	
Temp.	**Free**	**Conf.**	**Free**	**Conf.**	**Free**	**Conf.**	**Free**	**Conf.**	**Free**	**Conf.**
361K	0	0	0	0	0	0	0	0	1	0
367	0	0	0	0	0	0	0	0	24	0
368	0	44	0	0	0	0	0	0	57	0
369	0	60	0	0	0	0	1	0	65	0
377	0	74	0	0	0	0	34	0	75	0
378	0	74	0	36	0	0	46	0	75	0
383	0	75	0	56	1	0	58	0	76	0
394	0	75	0	60	32	18	61	0	76	0
409	0	75	1	60	46	45	61	0	76	0
424	0	76	10	61	47	46	62	0	77	0
427	0	76	22	61	47	46	62	12	77	0
450	0	76	32	61	47	46	62	30	77	0

in the T_c. However, this increase is not universal. There is a critical length above which the stability (measured through the T_c) decreases. This length is $\approx 50\%$ of the chain length. For 100 base pairs this is 50 while for 200 base pairs this is 100, see Fig. 5.4. An interesting point to note is that the curve is not symmetric about this critical length. The decreasing side of the curve has a slightly steeper slope than increasing side of the curve. In the Fig 5.6 it is distinguishable. An important question to be addressed. Why there is an increase and then a decrease in the stability of the DNA chain when it is translocating from the confined area to the unconfined area? Let us consider the three cases as shown in Fig.5.5. The segment that is outside the pore is having an unconstrained space while the portion that is inside the cone has constrained space. When a large portion of the DNA molecule is inside the cone (Fig.5.5a), there is a larger segment of the chain which is under the constrained space. In the constrained space, the effect of the conically shaped wall on the molecule varies (Fig.5.1 & 5.5). If one closely look into the geometry, one can find that the distance from the confining wall to the last few base pairs (on the right-hand side of the Fig.5.1a) is considerable ($\approx$ 100 Å). However, as the molecule comes out of the pore (Fig.5.5b), the length of the free segment increases while the length of the constrained segment decreases. Hence there is a decrease

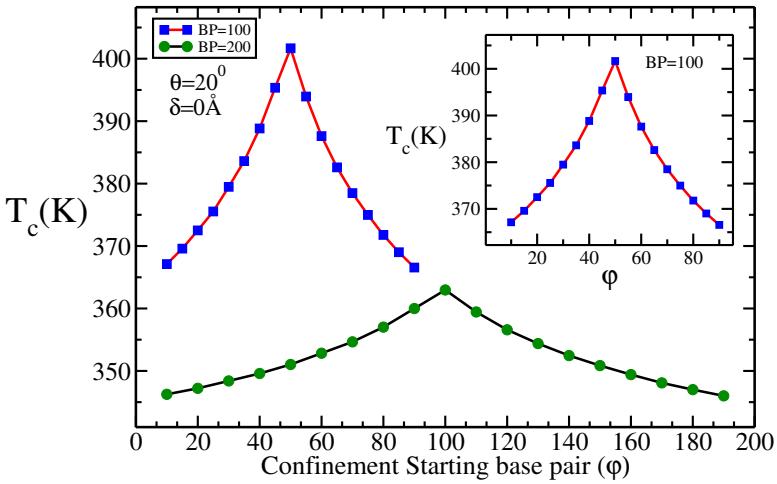


Figure 5.4: The change in the transition temperature of the chain that is confined in a conical geometry. $\theta = 20°$ for this graph. Here we consider two chains of different lengths, 100 and 200 base pairs.

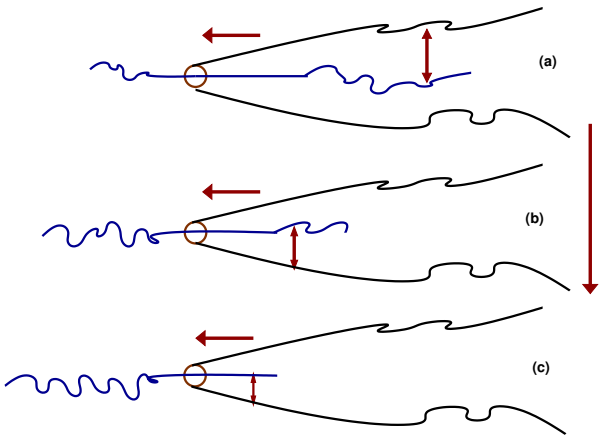


Figure 5.5: The schematic of translocation process. We show the translocation of DNA through the three figures (a), (b) and (c). The fraction of pairs which stay inside/outside the cone varies as DNA comes out of the cone.

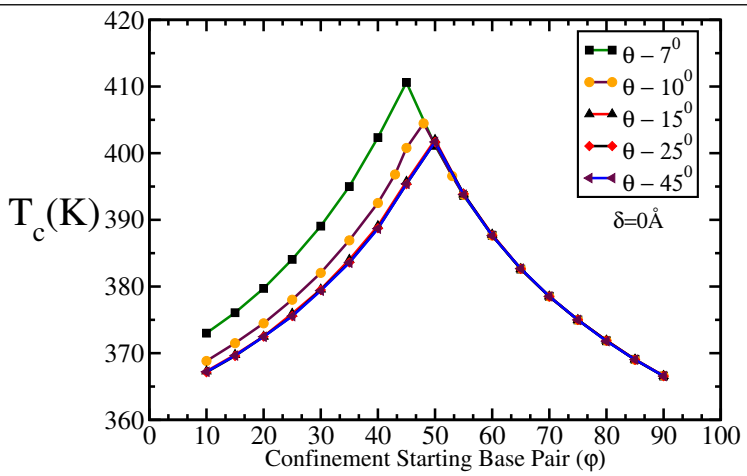


Figure 5.6: The plot shows the variation in the transition temperature, of DNA chain of 100 base pairs, with a fraction of chain outside the cell for different angular separations. Here the pore width, $\delta = 0$Å.

in the entropic contribution to the melting from the constrained portion. Thus, when the length of the constrained segment decreases, the entropic contribution to the melting decreases. There is a balance between the increase in the entropy due to unconstrained segment and decrease in the entropy due to the constrained segment. Up to the critical length, the decrease in the entropic contribution from the constrained space is more than the increase in the entropic contribution from the unconstrained segment. Hence the molecule is getting more stability at this point. The story is just reversed for the case in which a large segment of the chain is out of the cone (Fig.5.5c). In this case, the entropy of the molecule is higher; even there are few base pairs inside the pore. Hence the molecule is getting less stable as it is coming out of the cone. (decrease in the transition temperature, T_c). These results are in consistency with the simulation results by Nikoofard et al. [191]. They have shown that the force on the polymer that is translocating through a conical pore feels a driving force due to the confining walls. The origin of this force is purely entropic and is responsible for the translocation process.

We further investigate how the angular separation, θ, affects the stability of the DNA molecule while passing through the pore. We show our findings through the Fig. 5.6. In this plot, we calculate the transition temperature as the DNA translocate through the pore for different angular separation. Our findings show that with the increase in the angular separation, the stability of the translocating molecule decreases. For $\theta = 7°$, the maximum value of T_c is 410.5 K, it decreases to 404.5 K for $\theta = 10°$. The maximum value of T_c further reduces to 401.2 K for

$\theta = 15°$. We find that for $\theta > 15°$ there is no further decrease in the values of T_c. These results indicate that for higher values of θ, a larger segment of the chain, inside the cone, does not see or experience the wall restriction. Only a few base pairs that are closer to the pore experience the restriction due to the confining wall.

We investigate here the effect of the pore width (δ) on the stability of a DNA molecule confined in a conically shaped pore. We calculate the transition temperature (T_c) of the system for different values of δ ranging from 0 to 2 Å for a fixed value of $\theta = 10°$. Note that when we write $\varphi = 20$, this means that there are 19 base pairs outside the cone while 80 are inside the cone. We find that as the DNA molecule comes out of the pore, the stability of the molecule increases. After a certain length, there is a decrease in the stability of the chain. We have discussed this statement in the previous section in detail. When we change the value of pore width and study the stability of the DNA molecule during the translocation process, we find that with the increase in the pore width, the overall stability of the molecule decreases. This decrease is due to the more entropic contribution from the restricted base pairs. We show the results for $\delta = 0,\ 1,\ \&\ 2$Å only. There is a surprising feature we are observing for $\delta = 2$Å onwards. While for $\delta = 2$Å, there is a point around which the stability first increases and then decreases, for higher values of δ, there is no such unique point, but there is a region of ≈ 25 base pairs (and more) where the stability of DNA molecule does not change or change is not observable. For a chain of 100 base pairs, there is no substantial change in the stability of the molecule between the translocation of 30 to 60 base pairs, as T_c does not change substantially. After this point, the value of T_c decreases sharply.

5.3.2 DNA confined in a cylindrical geometry

Now consider, the DNA chain translocating through a pore of cylindrical geometry. To understand the melting profiles of the DNA chain in cylindrical confined geometry, we calculate the average separation $\langle y_j \rangle$ of the j^{th} pair of the chain with the temperature (see Fig 5.8). The melting nature of the DNA chain is different from the conical confined geometry because of the different geometry.

When the same chain translocates through the cylindrical geometry, there is a decrease in the transition temperature of the system as the chain comes out of the cone. The results are plotted in Fig.5.9. This is because of the difference in the geometries of the two cases. The geometry of a cone is such that, the radial distance between the two walls increases. Due to this, all the base pairs of

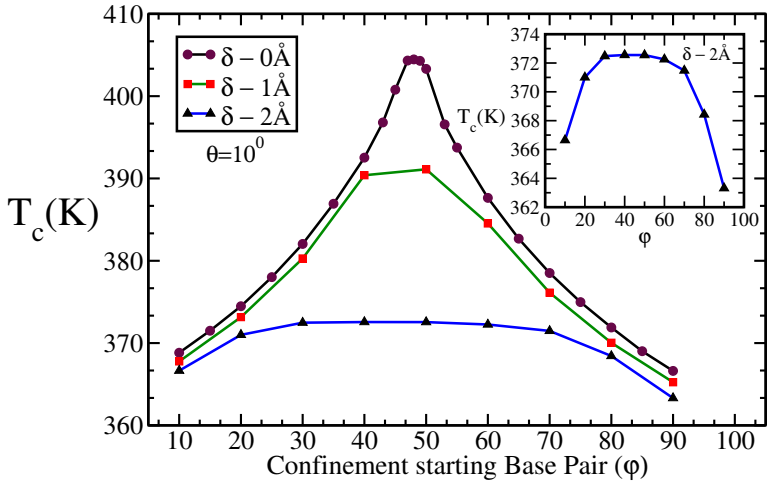


Figure 5.7: The plot shows the variation in the transition temperature, of DNA chain of 100 base pairs, with a fraction of chain outside (inside) the cell for different angular separations.

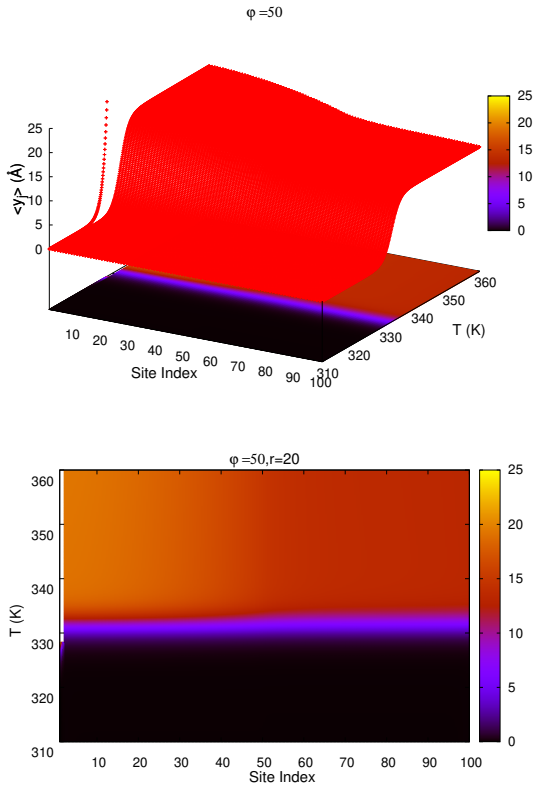


Figure 5.8: The opening of the chain of 100 base pairs with temperature. The distance (r) is 20Å and $\varphi = 50$. The confinement restriction starts from the base pair 50 ($\varphi = 50$), and hence inside the cone there are 50 base pairs while on the other side (out of the cylinder) there are 49 base pairs.

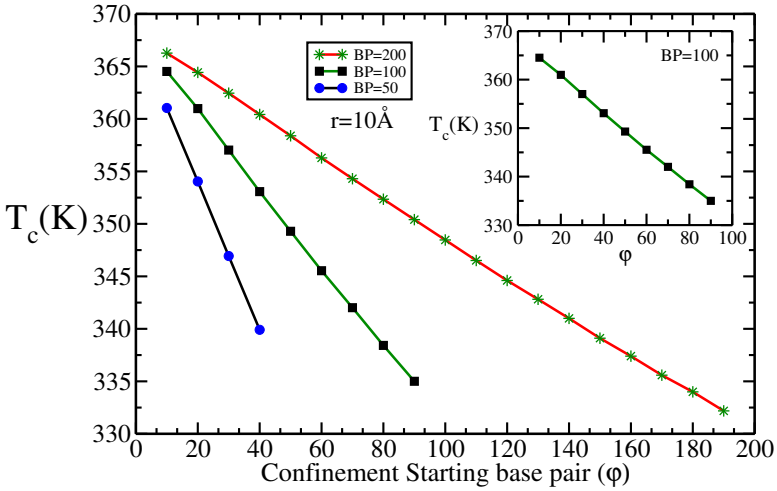


Figure 5.9: The change in the transition temperature of the chain that is confined in a cylindrical geometry. The distance, $r = 10$Å is considered. We consider three chains of different lengths, 50, 100 and 200 base pairs.

the DNA molecule (confined in a conical geometry) do not feel the same amount of restriction. The base pairs on the opposite side of the vertex are at a larger separation ($\approx$ 100 Å) in comparison to the base pairs near the vertex. In the case, in which the chain is confined in the cylindrical geometry, all base pairs (inside the pore) are having the same restriction. Hence there will be a more significant reduction in the transition temperature of the system as the chain translocates from the confined to the free state.

We also study the effect of the diameter of a cylinder on the stability of the DNA molecule that translocates through a cylindrically shaped pore. Our findings show that there is an impact of a geometry of the confinement on the stability of the DNA molecule that is translocating through the pore. For a DNA molecule translocating through the conical pore, the transition temperature increases up to a certain length and decreases after this length. The same argument is not valid for the DNA molecule translocating through a cylindrical pore. Here there is a decrease in the transition temperature as the molecule comes out of the pore. Please refer to Figs 5.7 and 5.10 for more details.

5.4 A random sequence

In the DNA mapping experiment, the researchers consider a real sequence of DNA molecules and observe the denaturation profile of the molecule. To obtain the

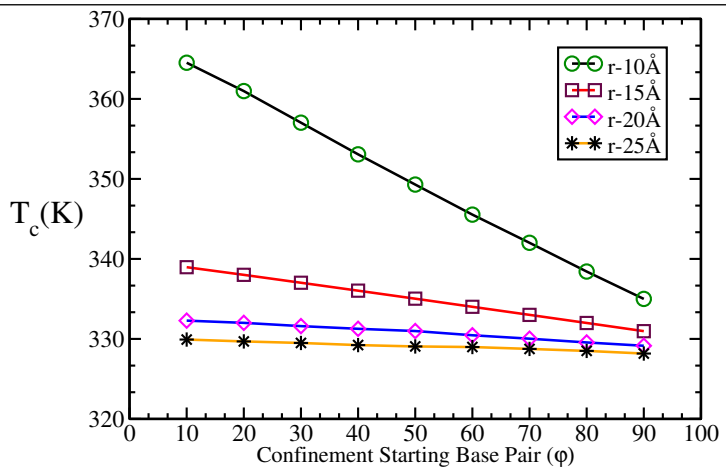


Figure 5.10: DNA translocates through a pore of cylindrical geometry. The plot shows the transition temperature through different radial distance, r. Here the chain is of 100 base pairs.

denaturation profile of a real DNA molecule that is having a random sequence of AT and GC pairs, we consider a 100 base pairs DNA molecule with a random sequence of AT and GC base pairs that are confined in cylindrical space. The sequence is,

$5' - GTGTGTGACCCGTTATGCGTAATTTAAATTAGAGCGTCATTGCGAGTA$
$GTCGTTTGCTTTCTCAGGCCCCGCGCGATTAAGCGTGACAGCCCCAGGGA$
$AC - 3'$.

The heterogeneity in the model is introduced via the Morse potential parameters, D_i and a_i. From the previous results, we know that the bond strengths of these two pairs are in an approximate ratio of $1.25 - 1.5$ as the GC pairs have three while AT pairs have two hydrogen bonds [11, 13, 14, 17, 128]. The value of D_{AT} = 0.075 eV while, $D_{GC} = 1.5 \times D_{AT}$ and $a_{AT} = 4.2$ Å^{-1}, $a_{GC} = 1.5 \times a_{AT}$. Here we monitor the average separation $\langle y_j \rangle$ of the molecule with the change in the temperature of the system. The obtained result is plotted in Fig. 5.11. The radius of the confined space is $r = 40$ Å. We are showing the results of the denaturation profile of a DNA molecule that is partially inside the core (50%) and partially outside the core. At a temperature 375 K, the $\langle y_j \rangle$ indicates the sequence effect on melting. As the temperature is increased from 375 K to 380 K, the section of the DNA molecule that is outside the pore opens, $\langle y_j \rangle \geq 1.2$ Å, while the section that is inside the core opens moderately. For a higher value of temperature the sequence effect disappears, only the effect of confinement is reflected. To avoid overflow of the data, we are not showing the results for the other conformations.

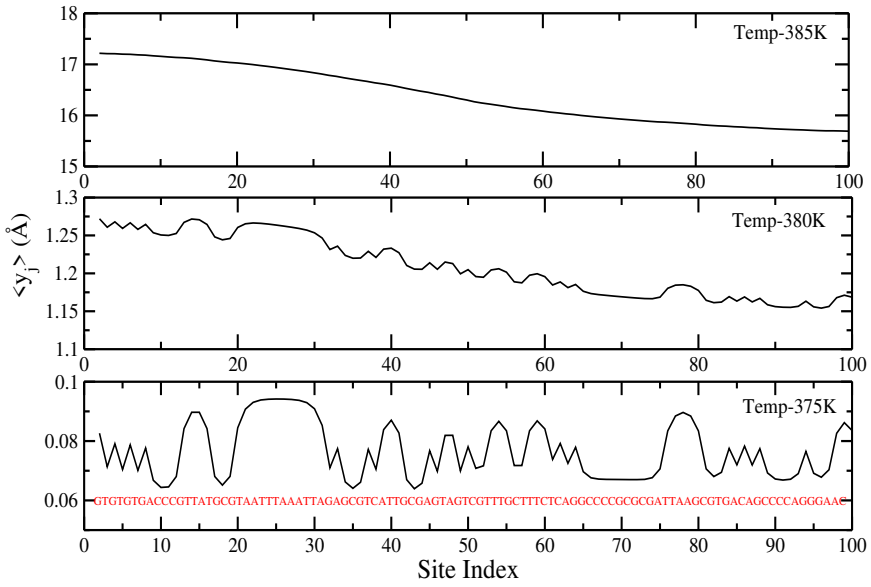


Figure 5.11: The opening of the heterogeneous chain of 100 base pairs that is confined in a cylindrical geometry and the value of $r = 20$Å and $\varphi = 50$.

5.5 Summary

In this chapter, we have studied the thermal denaturation of a DNA molecule passing through a pore. Through the numerical calculations in a thermal ensemble, we have calculated the average separation of the base pairs of a DNA molecule. We have considered a DNA molecule that is confined in conical and cylindrical shaped pores. Due to the geometry of each wall, the denaturation of the DNA molecule has been found different in each geometry. Not only, the denaturation but also the melting profile of the DNA molecule passing through these geometries differ significantly. The geometry of a cone is such that the distance between the molecule and the wall increases from vertex to end. Hence, the base pairs at the vertex of the cone are restricted while the base pairs at the end have a greater entropic contribution in the opening. The same argument is not valid when DNA is confined in a cylindrically shaped core. The parallel walls of a cylinder restrict the space uniformly, due to which there is suppression of the entropy to the opening of the DNA molecule. We have also considered different-length DNA molecules in these two geometries. We have found that for DNA confined in a conically shaped pore; the transition temperature is higher for the shorter molecule [180]. For DNA confined in a cylindrical pore, we have found that the transition temperature increases with the length of the molecule. We have also studied the effect of pore

width on the thermal stability of the molecule passing through these two pores. We have also considered a random sequence of DNA molecules that is confined in a cylinder and have studied the melting profile of it. Our findings show that the stability of the DNA molecule passing through a pore heavily depends on the geometry of the pore. We have also studied the sequence effect on the denaturation profile of a DNA molecule confined in a cylindrical geometry. We have shown how the denaturation profile of a heterogeneous DNA molecule differs from that of a homogeneous DNA molecule. In many interesting experiments, researchers have shown how the transport properties of the dsDNA encapsulated in a carbon nanotube change with the change in the diameter of the tube [92, 93]. To have a fair understanding of the stability of DNA inside the nanotube, experimentally one can explore the effect of pore geometry on the stability of a DNA molecule during translocation process. The role of entropy from the two regions (constrained and unconstrained) would be exciting to explore. The primary source of entropy is the open pairs and strand fluctuations. The theoretical model considered here underestimate the entropy due to the strand fluctuation. These two sources of entropy play crucial roles in the opening of dsDNA depending on the size of the pore. How the entropy of dsDNA affect the translocation process will be subject for future interest.