CHAPTER 5

Effect of solvent gradient on DNA confined in a strip

Life started 4.5 billion years ago when most of the Earth was covered under the sea-water. At that period, formation of life is believed to be happened in deep sea hydrothermal vents[92, 133–136]. Hydrothermal vents are the faults in deep ocean floor which eject hot chemicals [137]. These chemical compounds reacted with hydrogen and carbon-di-oxide to produce organic compounds. Those organic compounds under suitable conditions (warm deep sea water of high pH value) form complex structures like cell membranes to nucleotides (DNA / RNA) in the cell. These are the first evidences of life which involves a cascade of biological / chemical processes.

One of the major challenge related to the inception of life at deep sea is to answer the question 'how could prebiotic molecules concentrated in otherwise highly diluted or low density salty water? ' [138] This is now well established that non-equilibrium boundary conditions can only result in such reduction of entropy to accumulate the life molecules [139–142]. Apart from this, living organisms must have a way to transfer the genetic information to the next generation. In deep sea, during DNA replication the molecules replicate itself and at the same time those replicated segments are forbidden to diffuse into ocean. Those replicated molecules accumulate complex form of life against diffusion.

Thermal convection along with DNA thermophoresis [139–141] can delineate the replication and accumulation of DNA in hydrothermal vent (see Fig. 5.1). Thermal convection is transfer of heat energy from a region to another mediated by the movement of fluid. The flow of fluid (gas or liquid) from heated source region to cold region (thermal gra-

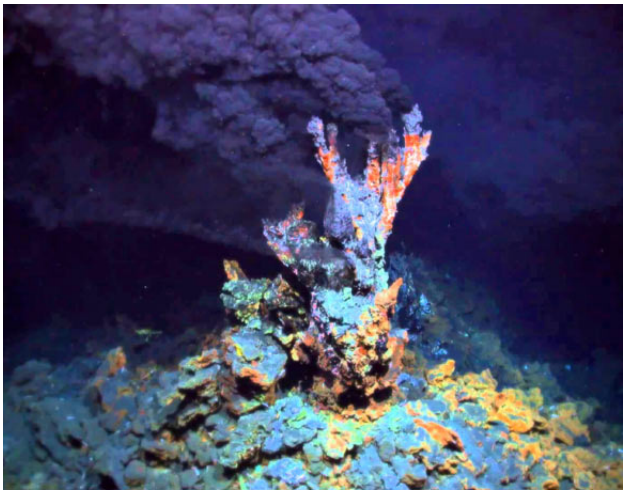

Figure 5.1. Deep sea hydrothermal vent ejecting mineral-rich chimneys. Image credit: Oregon State University / CC BY-SA 2.0.

dient) is known as convection current. When the hot fluid rises, it gets cooled due to the atmospheric temperature. This cold fluid again descends due to its high density. This cycle is well understood in terms of basic physics. Whereas thermophoresis is the process where the particles deplete from heated region. Particles move along thermal gradient (usually from hot region to cold) due to thermophoresis. While thermal convection shuttles the molecules up and down within a region (due to gravity), thermophoresis drives the molecule perpendicular to the motion of convection. The hydrothermal vents has large porous rocks which provides thermal gradient across it. DNA under convection denatures near the hot region. These separated strands shuffle back to the cold region by the convection current cycle. In cold region DNA replicates itself by DNA polymerase protein. Now, due to thermophoresis these replicated DNAs accumulate near the bottom of the cold region. Kreysing et al [143] noted that a heat flux across an open pore shows the exponential replication under convective thermal cycling. Experimental results showed that laminar convection can trigger and perform this DNA replicating reaction cycle over and over again. This process involves heat flux, convection and thermophoresis which is a dynamic and non-equilibrium

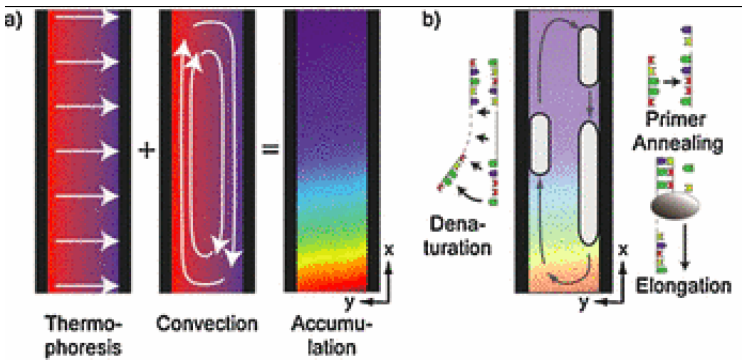

Figure 5.2. a) Trapping and accumulation of DNA by convection and thermophoresis. Thermophoresis drive the molecule to the right. DNA gets accumulated in the bottom of the right b) DNA convection cycle is shown. In each cycle DNA denatures by short primer and replicates by DNA polymerase. This is taken from ref [139]

process (see Fig. 5.2) [144–147].

Efforts have been made to understand behavior of biopolymers when subjected to pH or chemical potential gradient within confinement due to its potential applications in replication, transcription, segregation, isolation of nano particle bookmarker etc. [149–153]. In this context, solvent gradient arising due to the change in chemical potential plays a significant role in the understanding of colloidal accumulation, various separation processes, crowding of nucleotides, control of flow etc. at the nano-metric scale [149, 154, 155]. In the polymeric systems, experiments have shown that the wall of a capillary tube grafted with polymers can act as a valve. This valve selectively allows the flow of fluid through it depending on the solvent quality or temperature of the system. This is known as polymer gating [148]. When a good solvent is introduced inside the capillary pore there will be solvent gradient. Due to the gradient, polymers extend with in the pore which block the flow. Whereas, in low temperature or poor solvent inside the pore, polymers collapse to make the flow continue within the pore. This phenomenon is also an example of motion of molecules due to the thermal gradient (see Fig. 5.3). The thermal gradient drives the system to the steady state and there exists a

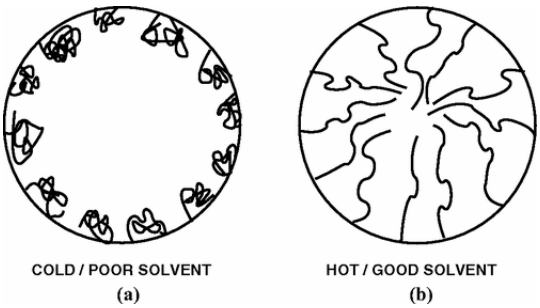

Figure 5.3. Schematic representations of end grafted polymers in capillary valve: (a) Low temperature (or poor solvent condition); (b) high temperature (or good solvent condition).This is taken from ref [148]

local thermal equilibrium where concept of statistical mechanics may be utilized [148] to understand such process.

Motivated by the above studies [143, 156–158], in this chapter we have studied a dsDNA confined in a strip of width L in Fig. 5.4. To model the linear thermal gradient, we keep one wall of the strip is at high temperature (T_H), whereas other wall is at the low temperature (T_L). This temperature gradient system allows the DNA to denature near hot surface (bottom surface in the model) and accumulate near cold surface (upper surface of the strip) in the zipped form.

Now equilibrium statistical mechanics usually is derived from partition function which is the sum of Boltzmann factors ($exp(\beta\epsilon)$). Now, when the confining surfaces are put at different temperatures then usual concept of equilibrium thermodynamics (uniform temperature) cannot be applied. In order to overcome the issue we used interaction gradient (ϵ) to model the gradient across the strip. This mimics the high temperature region containing a good solvent quality and low temperature region as poor solvent quality. One can have solvent gradient $\epsilon_H - \epsilon_L)/L$ across the channel analogous to the thermal gradient. In this framework, it is possible to apply the concept of statistical mechanics to understand the dynamics under the non-equilibrium condition by varying the base-pairing interactions across the strip.

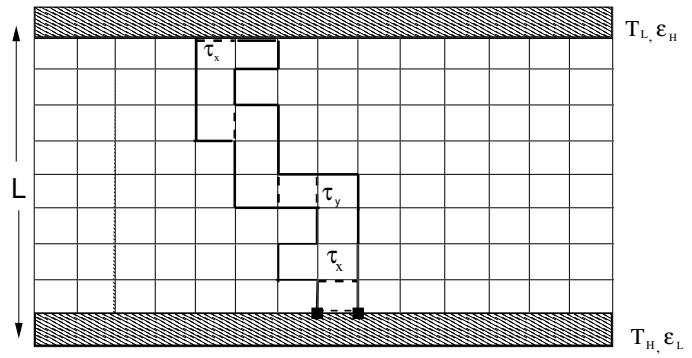

Figure 5.4. Schematic representation of MASAW (two mutually self-attracting self avoiding walk) dsDNA chain on a square lattice in a confined system (strip of length L) kept at different temperatures (T_H and T_L). We mimic the thermal gradient in terms of the solvent gradient by assigning different base-pairing interactions (ϵ_L and ϵ_H) corresponding to lower and upper walls respectively. The Boltzmann weight for the base-pairing interactions (τ_x and τ_y) are depicted by dotted lines. For the homogeneous medium τ_x and τ_y are same, but for a system having solvent gradient, τ is a function of y ($\tau_x \neq \tau_y$).

Electrostatic nature of H-bonds changes due to dielectric constant and thus directly affects the base-pairing of the DNA. The lattice model introduced in chapter 1 may be used to study effect of varying dielectric constant (solvent quality) to study the effect of solvent gradient on the melting transitions of DNA. We would like to extend our studies also to explore the effect of heterogeneous sequence of dsDNA on the phase diagram. The small chain length of dsDNA is studied using the exact enumeration technique [4, 18],whose results have been substantiated by Monte Carlo simulations [46] for the longer chain.

5.1 Model and method

We model the dsDNA as two mutually self-attracting self avoiding walks (MASAW) on a square lattice confined in a strip of length L. The channel is made up of 9 equidistant layers ($L = 0$ to 8) along y-axis (distance between two layers is one lattice spacing). We fix one end of the dsDNA chain at any one of the layers and the other end of each strand is free to

be anywhere Fig. 5.4. Now restrictions are imposed in such a way that the dsDNA chain is not allowed to escape the channel. Monomers are sites occupied by the polymers, and interactions are among non-bonded nearest neighbour monomers. This interacting pair, we referred as base-pair. Here, base-pair is restricted to native contacts only, where i-th monomer of one strand interact with i-th monomer of the other strand only [18, 85, 104, 122].

The thermodynamic properties of the system under study can be obtained by the average of physical quantities calculated in the canonical ensemble. So, the estimation of partition function is essential in the framework of equilibrium statistical mechanics and the same is given by,

$$Z = \sum_{all\ walks} \tau^{N_p} \tag{5.1}$$

Here, the sum is over all the individual chain configurations. $\tau = \exp(\beta\epsilon)$ is the Boltzmann weight for the base-pairing interaction. $\beta = \frac{1}{k_B T}$, where k_B is the Boltzmann constant and T is temperature. From here onward we set $k_B = 1$ and work on the reduced unit. N_p is the number of base-pairs. We are trying to mimic a system having temperature gradient along y-axis. So, the solvent quality is allowed to vary along y-axis which can be thought of as a varying (gradient) dielectric along y-axis. This allows us to vary the strength of base-pairing interaction as a function of y. The base-pairing interaction at zero-th layer ($L = 0$) is 0 and at the upper layer ($L = 8$) is ϵ. This permits us to define solvent gradient interaction as $\Delta\epsilon = \frac{\epsilon}{L}$ and the base-pairing interaction as function of y will be $\epsilon(y) = y\Delta\epsilon$. Now, to incorporate the inhomogeneity in the base-pairing interaction, we assign the base-pairing interaction in the y direction different than x direction. For x-direction, solvent quality is uniform and hence the Boltzmann weight for the base-pairing interaction is given by $\tau_x(y) = \exp(\beta\epsilon(y))$. However, when a base-pair is involved between two layers, we take the average interaction and the corresponding Boltzmann weight for base-pairing interaction formed between layers y

and $y+1$ can be written as $\tau_y(y) = \exp\left(\frac{\beta}{2}(\epsilon(y)+\epsilon(y+1))\right)$.

All the results are averaged over all possible choices of layers where one end of the polymer is can be anchored and other end is free to be anywhere. So, the DNA has a tendency to be at any layer with equal probability for a uniform channel ($\beta\Delta\epsilon = 0$). Our interest is to see which configurations are preferred when $\beta\Delta\epsilon \neq 0$. We consider here a short chain for which all possible configurations of the chain can be enumerated exactly using exact enumeration technique, The canonical partition function of such system can be written as,

$$Z_N^L = \sum_{(N_p^x, N_p^y, L)} C(N_p^x, N_p^y, n_m, L)\tau_x^{N_p^x}\tau_y^{N_p^y}, \tag{5.2}$$

where N_p^x, N_p^y, and n_m are number of base-pairing formed along x-direction (along the layer), number of base-pairing formed along y-direction (between two layers), number of monomers (nucleotides) respectively.
$C(N_p^x, N_p^y, n_m, L)$ is the number of distinct configurations having $N_p(= N_p^x + N_p^y)$ bound base-pairs with n_m monomers (nucleotides) corresponding to L-th layer.

The average number of monomers ($< n_m >$) and average number of base-pairs ($< N_p >$) in each layer can be calculated from the following equations:

$$< n_m >= \frac{1}{Z_N^L} \sum_{(N_p^x, N_p^y, L)} n_m C(N_p^x, N_p^y, n_m, L)\tau_x^{N_p^x}\tau_y^{N_p^y} \tag{5.3}$$

and

$$< N_p >= \frac{1}{Z_N^L} \sum_{(N_p^x, N_p^y, L)} N_p C(N_p^x, N_p^y, n_m, L)\tau_x^{N_p^x}\tau_y^{N_p^y} \tag{5.4}$$

Since, for large N the number of conformations increases as $C_N \sim \mu^N N^{\gamma-1}$, where μ is the connectivity constant of the lattice and γ is the critical exponent [4], it is not possible to go for a larger value of N using exact enumeration method. Therefore, in this study, we have to restrict

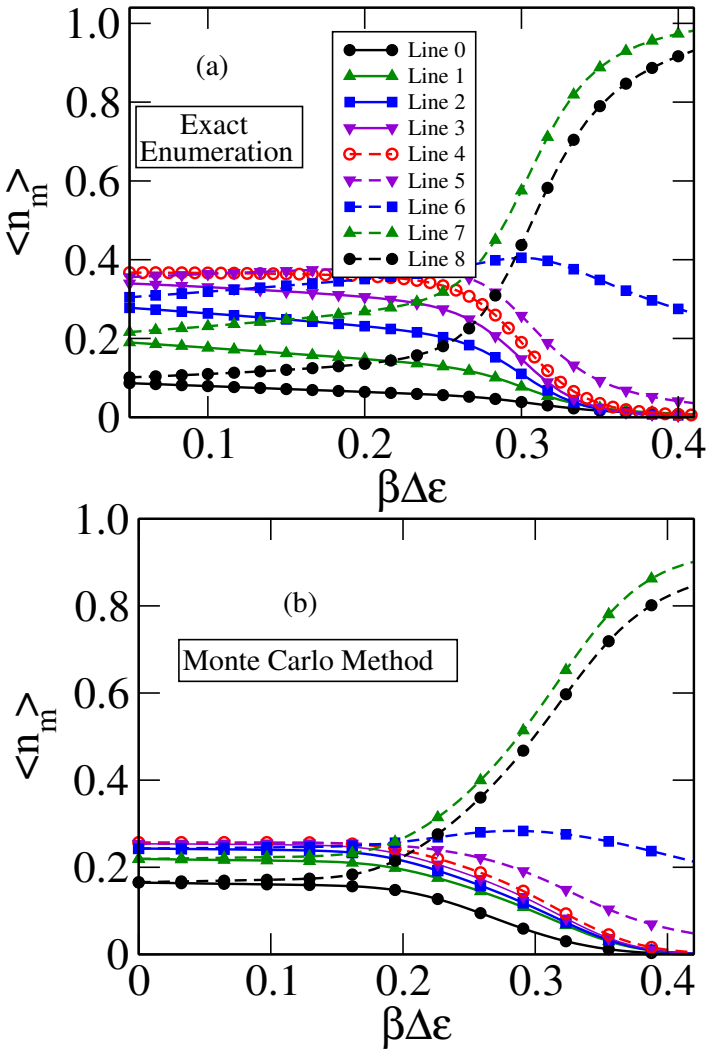

Figure 5.5. shows the variation of average number of monomers (nucleotides) ($< n_m >$) scaled by its length as a function of solvent gradient ($\beta\Delta\epsilon$) using (a) exact enumeration method (short chain); (b) Using Monte Carlo method (long chain). For both the cases, nucleotides move along the interaction gradient (thermal gradient) and prefer to stay near the upper layers (7th and 8th layers) at higher interaction gradient (cold temperature) to minimize the free energy.

ourselves for $N \leq 28$ (14 base pairs).

For a longer chain ($N = 200$ (100 base pairs)), we have used Monte Carlo method (Wang-Landau sampling) [46, 159] to simulate the dsDNA confined in a strip of width $L = 8$ (9 equidistant layers). It is pertinent to mention here that the present sampling gives quite accurate picture of the density of states (DOS) ($g(E)$) of the system with energy, E and thus allow us to compare it with the exact enumeration method where DOS can be calculated exactly. The details of Wang-Landau method is discussed in the introduction chapter.

Now, to calculate the physical quantities of the system (average nucleotide density, average number of base pairs, end-to-end distance etc.) such as the observable quantity Q can be calculated in the following way:

$$< Q(\beta) >= \frac{1}{Z(\beta)} \sum_{E} g(E) \exp(-\beta E) \overline{Q(E)}, \tag{5.5}$$

where $\overline{Q(E)}$ is expressed as,

$$\overline{Q(E)} = \frac{\sum_{Q} h(E,Q)Q}{\sum_{Q} h(E,Q)}, \tag{5.6}$$

where, $h(E,Q)$ is the two dimensional histogram in E and Q.

For the system under consideration, we have calculated the average number of nucleotides in each layer as,

$$< n_m >= \frac{1}{Z(\beta)} \sum_{E} g(E) \overline{n_m(E)} \exp(-\beta \Delta\epsilon E)) \tag{5.7}$$

where,

$$\overline{n_m(E)} = \frac{\sum_{n(y)} h(y,E,n(y))n(y)}{\sum_{n(y)} h(y,E,n(y))} \tag{5.8}$$

where $n(y)$ is the number of nucleotides in the layer y, $h(y,E,n(y))$ is the histogram in nucleotide number $n(y)$ in y-th layer corresponding to energy E, and $Z(\beta) = \sum_E g(E) \exp(-\beta \Delta\epsilon E)$. Similarly, we have also

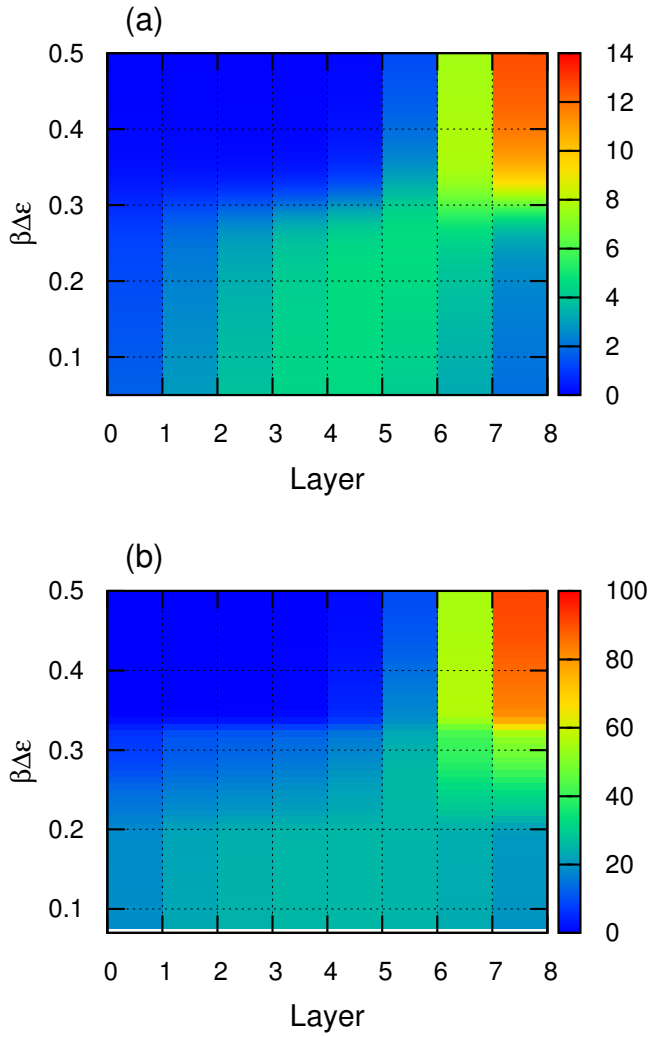

Figure 5.6. Distribution of an average number of monomers (nucleotides) as a function of layer number and solvent gradient ($\beta\Delta\epsilon$) using (a) exact enumeration method, and (b) Monte Carlo simulation. The colour corresponds to the density.

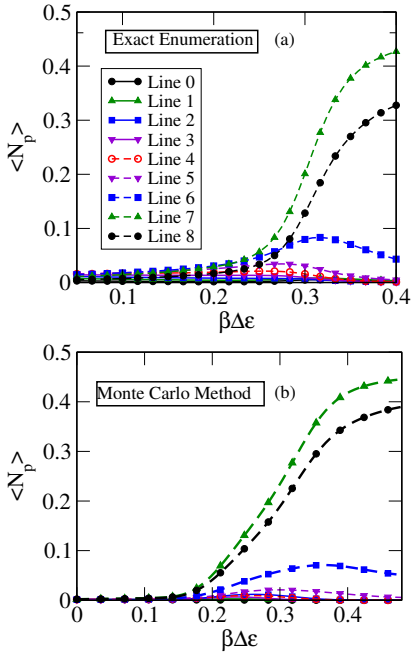

Figure 5.7. show the average number of base-pairing ($\langle N_p \rangle$) scaled by length of the chain (N) as a function of solvent gradient ($\beta\Delta\epsilon$) using (a) exact enumeration method (short chain); (b) Monte Carlo method (long chain). It is evident from the plots that as we increase the solvent gradient, zipping (increase in number of base-pairing) takes place towards the upper surface.

calculated the average number of base-paring ($< N_p >$) in each layer using the same procedure.

The efficiency of the method depends on how accurately the allowed configuration space to the polymer is sampled. For that we have employed several local and global moves, like, kink, Krank-shaft, pivot and pull moves during the simulation.

5.2 Distribution of nucleotide and base-pair across the strip

We first calculated the distribution of average number of nucleotides in each layer obtained from the exact enumeration method for a short chain as a function of $\beta\Delta\epsilon$ Fig. 5.5 (a). It is evident from the plots that at

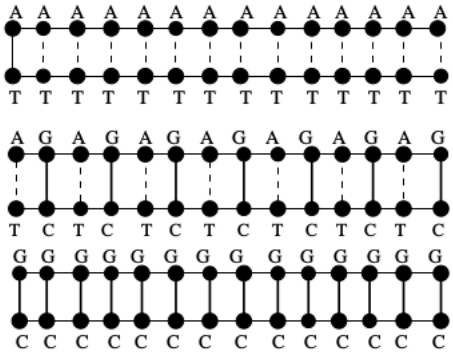

Figure 5.8. Schematic representations of three different sequences of DNA: (a) a homo-sequence of AT, (b) a di-block of DNA which contains 50 % AT and 50 % GC, and (c) a homo-sequence of GC.

low interaction gradient, the dsDNA prefers to stay at the entropically favourable layers (near the middle) and decreases when one moves away from the middle layer. The increase in interaction gradient leads to the increase in average nucleotide density. It can be easily seen from Fig. 5.5 (a) that when $\beta\Delta\epsilon$ increases, the nucleotide density shows a sharp jump for layer 7 and 8. This happens at a certain value of interaction gradient (~ 0.30), where the zipping takes place and dsDNA moved towards the upper layers. This is analogous to the thermophoresis where dsDNA moves under thermal gradient [139–141]. In order to see whether such behaviour is generic in nature, we performed Monte Carlo simulations to study the average nucleotide density across the strip which is depicted in Fig. 5.5 (b). Here, like exact enumeration study, we also observe that nucleotides move along the interaction gradient and preferred to stay near the upper layers to minimize the free energy. For the longer chain we observe the increase in nucleotide density around the interaction gradient ~ 0.33. In Figure Fig. 5.6 we showed the distribution of nucleotide as a function of the layer and $\beta\Delta\epsilon$. The colour corresponds to the number of nucleotides. The transition point is consistent with Fig. 5.5. The qualitative features for short and long chain remain almost the same. One may notice from Fig. 5.5 (a) and (b) that the value of $\beta\Delta\epsilon$ is bit higher for a

long chain in compassion to short chain. This may be attributed to the confinement arising due to net increase in the density in a strip of width L.

We expect that the dsDNA is in zipped form near the top surface (high interaction energy). To verify that, we have shown the variation of number of base-pairs as a function of $\beta\Delta\epsilon$ in Fig. 5.7 . As observed in Fig. 5.5, below a certain value of $\beta\Delta\epsilon$ there is no base-pairing which is clearly reflected here. When $\beta\Delta\epsilon$ increases, the zipping takes place towards the upper layer *i.e.*, at higher interaction gradient. Similar behaviour has been observed for a longer chain (Fig. 5.7 (b)) using Monte Carlo simulation study. Therefore, the manipulation of interaction gradient within a strip may drive the dsDNA from the lower surface to the upper surface which will be helpful for fast molecular transport.

5.3 Effect of sequence on the melting profile

Wienken et al [160] performed set of experiments on DNA thermophoresis with different sequences to study the stability of DNA. They found that any mismatch in sequence of DNA leads to lower melting temperature (T_m). Motivated by this study, we also study the effect of sequence and solvent interaction gradient on the melting profile of DNA. We have considered following three sequences as shown in Fig. 5.8: (i) a homo-sequence of AT, (ii) a di-block of DNA which contains 50% AT and 50% GC, (iii) a homo-sequence of GC. For this we choose the MASAW model of polymer [18, 122] described in section II which captures essential physics of DNA denaturation. Since, A-T and G-C base pairs contain two and three hydrogen bonds respectively, we assign the base-pairing interaction for A-T as $\epsilon = -1$ and for G-C as $\epsilon = -2$ in our model.

In Fig. 5.9, we plot the average end-to-end distance of dsDNA as a function of interaction gradient for the above mentioned sequences. One can see from these plots that the solvent gradient effect on base-pairing interactions for DNA chain reveals that the GC rich DNA will melt at higher temperature compare to that of AT rich DNA. However, both small

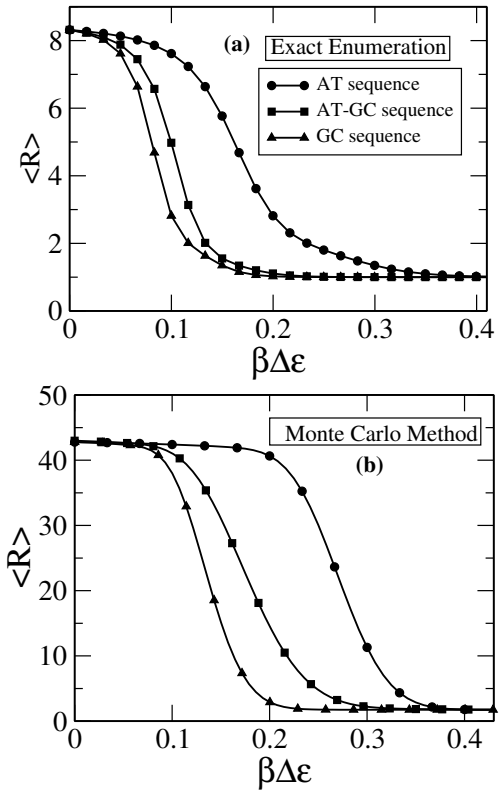

Figure 5.9. Variation of average end-to-end distance ($\langle R \rangle$) as a function of solvent gradient ($\beta\Delta\epsilon$) for three sequences using (a) exact enumeration (short chain) and (b) Monte Carlo method (long chain). Both short and long chain results show that GC rich DNA chain denatures at higher temperature than that of AT rich DNA.

and long chain results ensure that while it migrates from hot to cold region, it transforms from coil structure to zipped rod like structure.

To understand the melting behaviour of DNA under solvent gradient in confined geometry along with the accumulation of biomolecule in the specific region, we vary both temperature and solvent interaction gradient. By adjusting the interaction and temperature, we studied the phase diagram which represents the zipping-unzipping transition of DNA. The phase diagram can be drawn by identifying the peak of the density fluctuations as a function of the temperature. The phase diagrams for different sequences of DNA obtained from exact enumeration and Monte Carlo

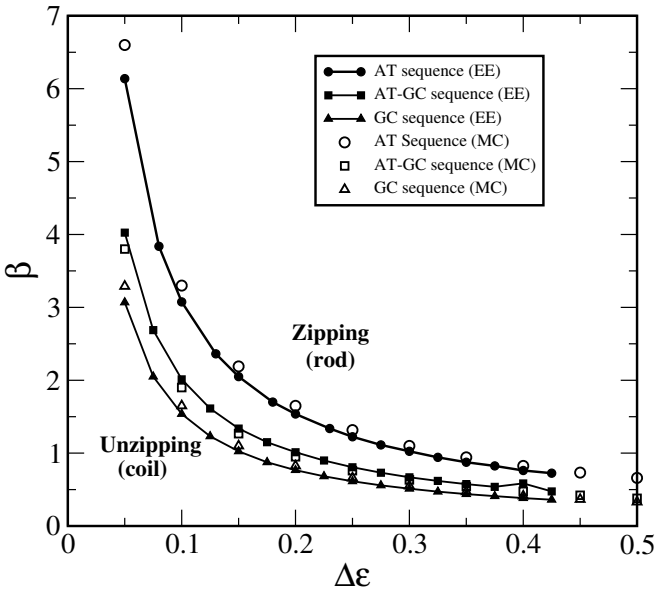

Figure 5.10. Phase diagram of DNA melting under solvent gradient for different sequences: (a) exact enumeration (EE), and (b) Monte Carlo (MC) method. There is an excellent agreement between exact enumeration and Monte Carlo results. It shows a transition from DNA unzipping to zipping.

method are shown in Fig. 5.10. From the plot, it is evident that the transition temperature is higher for GC rich sequence of dsDNA.

5.4 Summary

We have studied the transfer of dsDNA from one side of the wall to the other inside a strip of finite width under the solvent gradient. A simple but realistic lattice model of polymer presented here, is able to show that solvent gradient drives the dsDNA from the lower layer to the upper layer. We observe that for short chain, low interaction gradient is required for successful transfer from one side (hot) of the wall to another (cold) compare to long chain. As we are increasing the interaction energy between base pairs by varying the solvent quality (dielectric constant), the system always prefers to have zipped like structure to minimize its free energy. On the other hand, we also varied the GC content of the dsDNA to see the effect of sequential preferences on thermophoresis. We have observed

that the GC rich DNA will transfer to higher solvent interaction gradient side earlier than both AT-GC sequence and AT rich DNA. All results obtained for short chain using exact enumeration is in nice agreement with the results obtained from Monte Carlo simulation for longer chain.

By using a simple lattice model in a varying solvent setting we successfully able to show that nucleotides depletes from the hot surface and accumulates near cold surface. This mimics DNA thermophoresis like effect. But, the convection cycle (hot-cold-hot) of DNA is not possible to mimic in the scope of our study. In experimental studies oligonucleotides are exponentially replicated under temperature oscillations following a convection path. This may lead to formation of patterns in the confined system .