Statistical Mechanics of DNA in Confined Geometry

Arghya Maity

Department of Physics

BITS Pilani, Pilani-Campus

A thesis submitted for the degree of

Doctor of Philosophy

2020

Future Scope: Following Research Work

Introduction

In this chapter, we study the change in the structural properties of duplex DNA in confined geometry. This is an interesting issue in the field of DNA research. We study 1-BNA in cylindrical confined geometry and calculate the RMSD (root mean square deviation) with a change in the temperature of the system. We study the structural parameters for the different manifestations of this confined geometry. How the rate of increase of temperature influence the opening and other structural parameters of the DNA molecule is the subject of current interest. The temperature dependence of DNA dynamics has been studied by various research groups [192–198]. Still, less attention has been paid for the temperature-dependent DNA dynamics in confined geometry. This motivates us to study how the different structural parameters of DNA changes with a different changing rate of temperature in confined geometry.

Methods

We choose 1-BNA (PDB code) as our system. It holds 12 base pairs. For our problem, we choose AMBER14sb_parmbsc1 force field. We restrict every base pair according to geometry. For cylindrical geometry, the radius of the cylinder is our fundamental parameter. The parameter r is the radius of the confined cylinder (see Fig. 5.12). We use spc216, which is a generic equilibrated 3-point solvent model. We add $0.1M$ salt concentration of NaCl also. The structure is relaxed through energy minimization (EM), and it stops the minimization. EM ensured that we have a reasonable starting structure, in terms of geometry and solvent orientation. To begin real dynamics, we must equilibrate the solvent and

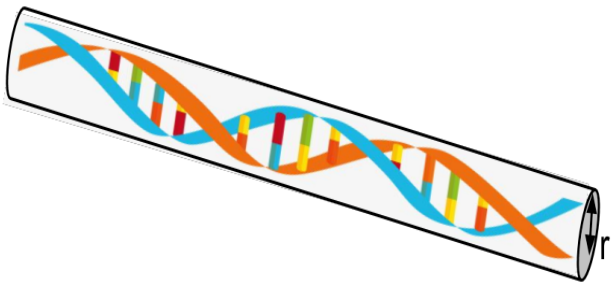

Figure 5.12: The schematic representation of the confined geometry. For cylindrical confined geometry r represents the radius of the cylinder.

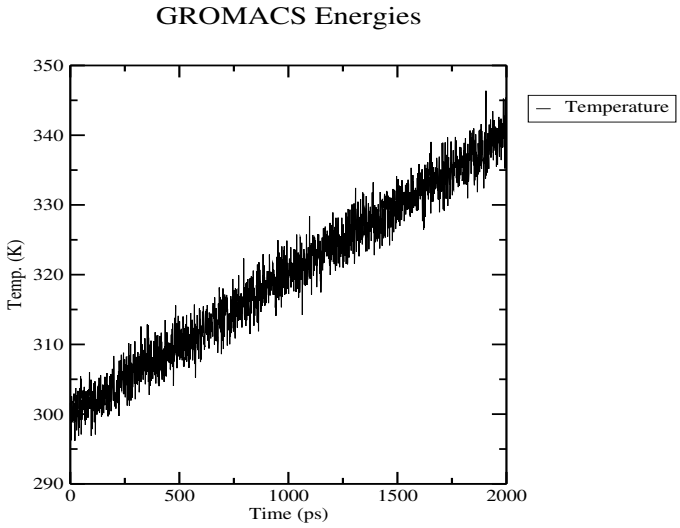

Figure 5.13: The rate of increasing of temperature with time.

ions around the DNA. If we have tried to attempt unrestrained dynamics at this point, the system may collapse. The reason is that the solvent is mostly optimized within itself, and not necessarily with the solute. It needs to be brought to the temperature we wish to simulate and establish the proper orientation about the solute (the DNA). After we arrive at the correct temperature (based on kinetic energies), we apply pressure to the system until it reaches the proper density. The well known simulated annealing (SA) protocol is supported in GROMACS with an arbitrary number of reference temperatures that change during the simulation. Since we are changing the reference temperature, it is important to remember that the system will not instantaneously reach this value. We need to allow for the inherent relaxation time in the coupling algorithm too. If we are changing the annealing reference temperature faster than the temperature relaxation, we will

probably end up with a crash when the difference becomes too large. The previous step, NVT equilibration, stabilized the temperature of the system. Prior to data collection, we must also stabilize the pressure (and thus also the density) of the system. Equilibration of pressure is conducted under an NPT ensemble. Upon completion of the two equilibration phases, the system is now well-equilibrated at the desired temperature and pressure. We are now ready to release the position restraints and run production MD for data collection. Here we prefer to run a 10-ns MD simulation for our system. We use trjconv as a tool to strip out coordinates, correct for periodicity, or manually alter the trajectory (time units, frame frequency, etc.). We look at structural stability first, and GROMACS has a built-in utility for RMSD calculations called rms. We study the RMSD for different geometry. All simulations are conducted with the GROMACS package, version 4.0.4.

To analysis different structural features of DNA, we take the help of curves+. Curves+ is the revised version of the Curves approach [199]. It can analyze molecular dynamics trajectories and generate time series. It follows the international conventions for nucleic acid analysis. The necessity of the detail crystal structure of DNA brought together many crystallographers, and hence an EMBO workshop was conducted in Cambridge in 1988 to settle this issue. There are many inconsistencies about the parameters which are useful to describe the helical DNA figure. The Cambridge meeting finally proposed a geometrical set of parameters that can describe the base pair location with respect to the helical axis [200]. The proposed parameters were not accepted by many groups [201], but there was not a general consensus. The final agreement was achieved in 1999 at Tsukuba meeting [202]. Curves+ provides the standard geometrical figure of DNA bases in a data file by analyzing molecular dynamics trajectories. It provides all the helical and backbone parameters.

We know that to describe the parameters in a group, there are two choices, and one is based on local helix axes between two base pairs or global parameters that are calculated relative to an overall helix axis. The values differ considerably indeed. Curves+ avoids this issue of local and global helical parameters. Curves+ set its reference system using chosen base atoms. There are C1', N1(Y)/N9(R), and C2(Y)/C4(R) in standard bases where Y represents a pyrimidine base, and R represents a purine base. To understand the underlying mathematics see refs [199]. It can produce a complete set of helical parameters. In our present study, we are interested in the full set of helical parameters that include translational and

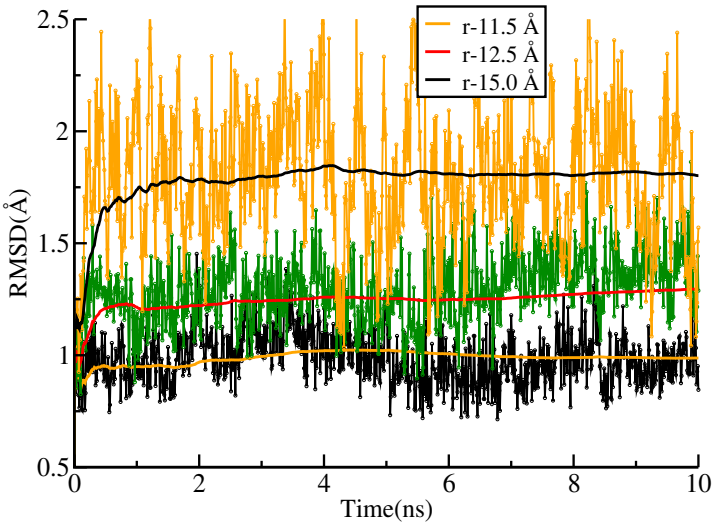

Figure 5.14: RMSD calculation results are plotted for three cylindrical radius (r). The running average is shown for each RMSD fluctuations.

rotational parameters.

Results

Entire Confinement Geometry

We consider a 1-BNA (12bp) DNA chain sequence and restrict it in the cylindrical confinement. We increase the temperature of the system as our external force from 300 K to 340 K linearly. The radius of the cylinder(r) is changed, and we calculate RMSD for each radius. The results are plotted in Fig.5.14.The first radius that we consider is 11.5 Å, and the running average of RMSD for this radius is about $\sim$ 0.98 Å $\pm$ 0.054 Å standard error of the mean (SEM). Then the radius is increased to 12.5 Å, and the running average of RMSD for this radius is about $\sim$ 1.2 Å $\pm$ 0.052 Å (SEM). Further increasing of radius to 15.0 Å leads to RMSD value 1.8 Å $\pm$ 0.171 Å (SEM). So confinement will increase the melting temperature of DNA. Protein and DNA are treated as macromolecules, and it is expected that both are supposed to behalf similarly in confinement effects [203]. Some simulation studied are done in confinement, and it shows that folding temperature of protein is raised with confinement strength [204–212]. To study the detail changes of structural features after 10ns simulation of DNA depending on confinement geometry, we analyze geometrical structure for two confined radius 12.5 Å and 15.0 Å through curves+. The intra-base pairs parameters hold three translational parameters

Table 5.3: The average value of Intra-base pair parameters, Inter-base pair parameters and Base pair-axis parameters.

Intra-BP Parameters						
Parameters	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(^{0})	Propeller(^{0})	Opening(^{0})
$r = 12.5$ Å	0.07	0.02	0.09	3.37	-8.78	1.26
$r = 15.0$ Å	0.07	0.02	-0.17	1.55	-13.21	1.68

Inter-BP Parameters						
Parameters	Shift(Å)	Slide(Å)	Rise(Å)	Tilt(^{0})	Roll(^{0})	Twist(^{0})
$r = 12.5$ Å	0.15	-0.36	3.34	0.51	0.73	35.83
$r = 15.0$ Å	0.12	-0.30	3.24	-1.07	3.11	35.13

Base pair-axis Parameters				
Parameters	X-disp(Å)	Y-disp(Å)	Inclin(^{0})	Tip(^{0})
$r = 12.5$ Å	-0.91	-0.07	0.94	-0.29
$r = 15.0$ Å	-1.11	-0.33	4.05	1.50

- Shear, Stretch, Stagger, and three rotational parameters - Buckle, Propeller, Opening. The zero values of these parameters represent canonical Watson-Crick base pairs, and the non-zero values show distortion with respect to the short and long axis of the base pairs and their normal. Through rigid-body transformation, the parameters are calculated, and that maps one base reference system onto the other. Keeping in mind the end effect, we execute 10 ns MD simulation to ensure equilibration that puts more flexibility of the terminal base pairs. There are many ways to reduce the end effect in simulations [213–219]. One of the crude way among them is to focus on the structural changes of the central base pairs assuming that they are unaffected by the fluctuations at the end. Curves+ provides the average value of base pair parameters (Intra, Inter, and BP-axis parameters) from 1bp to 12bp but to avoid the end effect distortion, we prefer to have an average value of 4bp to 9bp for Intra-base pair and Base pair-axis parameters. Inter-base pair parameters are pair junction parameters hence the average values are taken by neglecting the fisrt two and last two base pair effects.

5.6 Three dimensional geometrical structure of DNA

The large stagger values represent a highly perturbed hydrogen bond and opening of the base pairs. However, all kinds of Watson-Crick base pairs hold a negative value of Propeller. The large standard deviations of the rotational parameters further show the different Hydrogen bonding interactions between the bases. The inter-base pair parameters also hold three translational parameters- Shift, Slide, Rise, and three rotational parameters - Tilt, Roll, Twist. These translational and rotational parameters report the relative positions of two successive bps with respect to their short axes, long axes, and their normals, so these are also called pair junction parameters. Reference frames are defined for the two base pairs, and each base pair frame is referred to as the mid-frame that is based on Rodrigues' formula(for details, please see refs [199]). Generally, it is found that the angle of twist belongs between 20^0 and 40^0 in B-DNA structures. Then the angle of Roll generally belongs between 0^0 and 15^0 in B-DNA structures. The values of Roll and Twist are associated with the changes in the translational parameter slide [220]. A large negative difference in the Buckle of consecutive base pairs sometimes shows extreme changes in Rise. Similarly, a large positive value of ΔOpening enhances the Shift. On the other hand, large negative values of ΔStagger increase the Tilt and a large positive value of ΔShear increase Twist. The value of Slide is not significantly influenced by base-pair deformations like Roll. The base pair-axis parameters are also generated, and among four, there are two translational parameters- Xdisp, Ydisp, and two rotational parameters - Inclination, Tip. The Xdisp reports the movement of the bases towards the grooves, and Ydisp reports the movement perpendicular to the grooves. Inclination shows rotation around the short axis of the base pairs, and Tip shows the rotation around the long axis of the base pairs. Pseudodyad symmetry of the base pairs maintains the magnitude of these parameters unchanged with changing the direction of analysis, but the Ydisp and Tip change their sign due to that.

Partially Confined Geometry

Still, now we deal with the DNA system, which is confined entirely, but if the restriction of the confinement partially acts on the system, then RMSD also changes according to the location of the confinement. We take the same DNA chain (1-

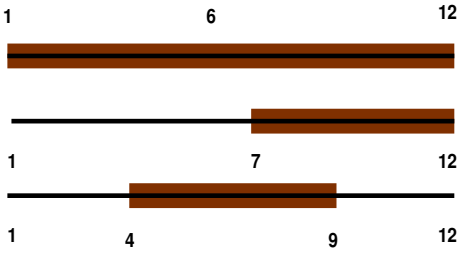

Figure 5.15: The location of the confinement. The radius of the confined cylinder for three systems are $r = 12.5$ Å.

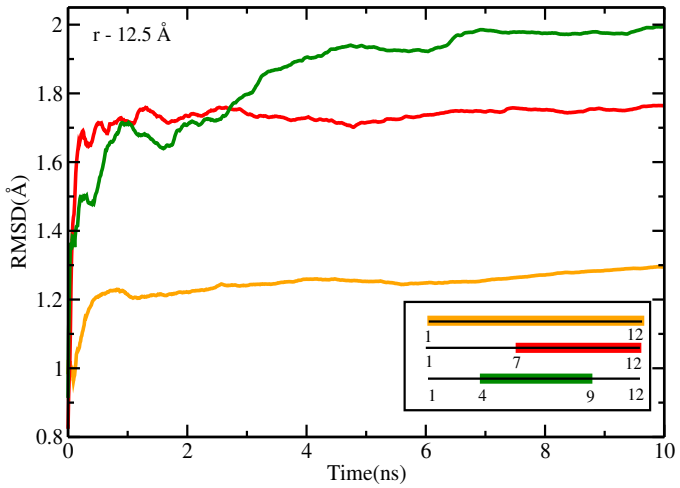

Figure 5.16: The RMSD calculation results are plotted for three different confined systems. The running average is shown for each RMSD fluctuations.

BNA) and increase the temperature of the system from 300 K to 340 K linearly too. We put the restriction of confinement of $r = 12.5$ Å into three ways. The diagram is shown in Fig. 5.15. The first diagram shows the full confinement system and the next one represents half confinement that starts from 7^{th} bp and lasts up to 12^{th} bp (6bp) and the last one indicates the same quantity of base pairs (6bp) confinement but the location of the confinement begins from 4^{th} bp and lasts up to 9^{th}. Since the opening manner of the chain for these three confined systems should be different [221] hence the RMSD value would be different. The RMSD value of these three systems is shown in Fig. 5.16. In the case of full confinement of the DNA system, the RMSD value is shown as 1.2Å, and for the second system that holds confinement in the last six-bps shows the saturated RMSD value as 1.7Å. The last system that holds confinement in the middle six-bps shows the saturated RMSD value as 2.0Å. We can observe one point in this graph that the