Effects of temperature on Nucleic Acids unzipping

By

Naman Kumar Bharti

A dissertation submitted for the partial fulfilment of a BS-MS dual degree in Science

Indian Institute of Science Education and Research Mohali

April 2019

Certificate of Examination

This is to certify that the dissertation titled "**Effects of temperature on Nucleic Acids unzipping**" submitted by Naman Kumar Bharti (MS14120) for the partial fulfilment of BS-MS dual degree program of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Shashi Bhushan Pandit	Dr. Ujjal K Gautam	Dr. Monika Sharma
		(Supervisor)

Dated: 26 April 2019

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Monika Sharma at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Naman Kumar Bharti

(MS14120)

Dated: 26 April 2019

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Monika Sharma

(Supervisor)

Acknowledgement

First I would like to thanks my family for all their support and to my thesis supervisor Dr. Monika Sharma for her full support and guidance and giving me this opportunity to work with her. I thank to committee members Dr. Shashi Bhushan Pandit and Dr. Ujjal K Gautam and IISERM computing facility. I thank to my lab members Prateek meena and my friends for moral support.

List of Figures

Notation

DNA – Deoxyribonucleic Acid

RNA – Ribonucleic Acid

NMR – Nuclear Magnetic Resonance

SMD – Steered Molecular Dynamics

MD – Molecular Dynamics

MC – Monte Carlo

NAMD – NAno scale Molecular Dynamics

VMD – Visual Molecular Dynamics

CHARMM – Chemistry at HARvard Macromolecular Mechanics

OPLS – Optimized Potential for Liquid Simulations

PDB – Protein Data Bank

PSF – Protein Structure Bank

PBC – Periodic Boundary Conditions

SBC – Stochastic Boundary Conditions

GBIS – Generalized Born Implicit Solvation

vs – versus

ps – picoseconds

ns – nanoseconds

Contents

Abstract

This thesis deals with the molecular aspects of unzipping of Nucleic Acids using steered molecular dynamics computationally. In a human body or any living organism, this conversion of double strand to single strand process is very common in many biological processes. The unzipping of nucleic acids can also be studied experimentally using optical tweezers and AFM techniques but to get the idea of how these processes work without any experimental setup we use molecular dynamics technique. With this molecular dynamics technique we can model any kind of biological system to study the behavior of those microbiological system. In this work, we will be unzipping RNA and DNA with a hairpin using Steered Molecular Dynamics (SMD) in different environmental conditions. With different forces and pulling velocity the unzipping of nucleic acids was studied. And a comparative study between DNA & RNA and a DNA hairpin & without hairpin and DNA hairpin at room temperature & high temperature was also done.

Chapter 1

Introduction

1.1. Computer Simulations

Computer simulations is a tool to understand the behavior/properties of atoms, molecules assemblies of molecules with respect to their structures or interactions between them computationally. Simulation of a system act as the running of a model based on that system. The atoms and molecules are allowed to interact over a certain period of time which gives a dynamic evolution of the system. Why do we need simulations …? In some cases, experiments is impossible like inside of stars, weather forecast; too dangerous like explosion; expensive like high pressure simulations; some cannot be observed on very short time-scales and very small-spaces. So in short, this is a way to understand something that cannot be found in other ways (difficult or expensive in laboratories). Computer simulations is like a connection between microscopic time scale and the macroscopic world of laboratory providing predictions of the interactions of molecules in bulk or in another words, a connection between theory and experiment and the accuracy of the results from these simulations are governed by our computational power. These simulations are useful because they can provoke, explain and helps in establishing intellectual properly.

This can be done with either Molecular Dynamics (MD) or Monte Carlo (MC) or there are other methods available which have combined features of both MC and MD.

1.1.1. Molecular Dynamics

In this technique, basically a simple numerical integrations of newton's equations of motion. The particles of the system are allowed to move according to the interactions happen in between them. It is basically a time evolution of the system we are working on. There is a disadvantage of this technique is that, because of the large number of particle interactions to be calculated, small time step will required, so our simulations are restricted to small time scales like several picoseconds, nanoseconds.

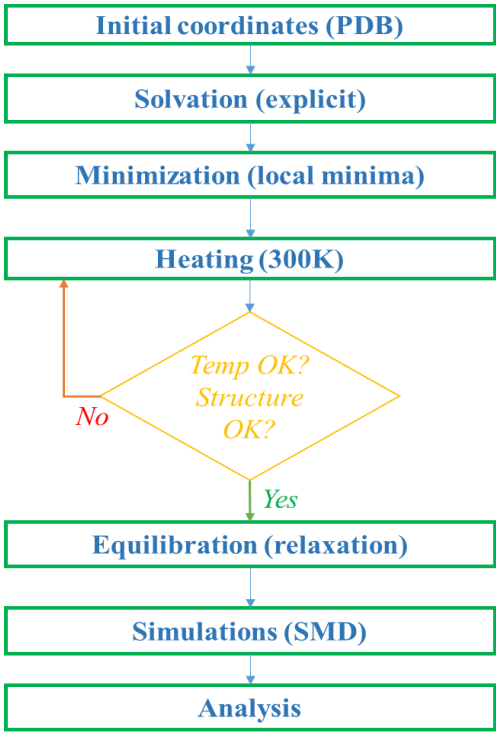

Figure 1 Overview of Molecular Dynamics

1.1.2. Monte Carlo

Any simulation model that can be simulated using dynamics can also be done by MC. MC algorithm are designs to sample the equilibrium unlike MD which is focused on the dynamics. Systems like lattice based models, are difficult to study with MD but with MC they can be readily studied.

The choice between MC and MD is selected by the phenomena or process we are going to study. Like for a simulation of gas or low density molecule, MC is preferred. As there can be large energy barriers which can be easily crossed by the random moves of MC while in MD, it can stuck in few low energy conformations, leading to poor conformational sampling. In case of MC, there is large probability of selecting large random moves, leading to large rejected moves decrease the efficiency of sampling. In determination of transport and other time dependent phenomena, MD is a suitable choice whereas

simulations with varying number of particles MC is better by adding moves with creation and destruction of particles.

1.2. History

The method was first developed in 1950's. In 1957, Alder and wainwright used an IBM 704 computer simulate collisions between hard spheres[13]. The next major simulation is done in 1964, using a realistic potential for liquid argon. The first molecular dynamics simulations of a realistic system was done in 1974 of liquid water. Then protein comes in picture, in 1977.

Presently, these simulations are applying on various systems like proteins, DNA, lipids etcetera. And also, there exist now many different effective technique for particular problems, including quantum mechanics and molecular mechanics that can be applied to enzymatic reactions and also in experimental techniques such as X-Ray crystallography, NMR etcetera.

1.3.Nucleic Acids

Nucleic acids are biopolymers which can be found in all well-known life form. They have a monomer unit called nucleotide which consists of three components: a 5-carbon ring sugar, a phosphate chain and a nitrogenous base. They are of two types DNA (Deoxyribonucelic acid) and RNA (ribonucelic acid). In DNA, there is ribose compound in sugar and in RNA, a derived form of ribose as deoxyribose in present. They are assembled form of base-pairs (monomers). DNA has Adenine, Thymine, Guanine, and Cytosine whereas RNA has Uracil instead of Thymine.

1.4. Software used

In this work, we have used two softwares: NAMD for simulations and VMD for modelling and visualization of simulations.

1.4.1. NAno scale Molecular Dynamics (NAMD v2.12)

NAMD is a software for parallel molecular dynamics which utilizes the maximum computational power for high performance simulation of particularly large biomolecular systems. NAMD can use up to hundreds of processors on high-end computer systems and also it can work on less number of processors on low-end clusters and can be run also individual computer or laptops[3].

NAMD software works with Amber and CHARMM force field which are basically mathematical expression used to calculate potential energies using different parameters and files. NAMD is a free Software which is distributed free with source code at www.ks.uiuc.edu. Software NAMD can do only simulation but for the visualization, we need another software called VMD, visual molecular dynamics.

1.4.2. Visual Molecular Dynamics (VMD v1.9.3)

VMD is molecular graphics software which is designed to modelling, visualize, and analysis of different biological molecules like nucleic acids, proteins etcetera and whatever simulation we have done in NAMD we can look and examine the whole simulations in this software. So VMD act as a graphical output of the MD simulations by displaying and animating that molecule. VMD can control a multiple number of structures at a time using a large variety of rendering styles, different coloring method and other different exciting features[4].

In this software we have control on different structures and even in a single structure we can control single atom or molecule and also we can find out coordinates of a system bond distances, angles, energies and many other different parameters of particular structure. So, basically, VMD provides us a complete graphical user interface for program control and also a text interface using TCL scripting for complex scripts and control of biomolecules.

1.5. Force Fields

A force field is just a set of equations which explains the dependence of the energy of a system on the coordinates of its particles. It consists of an analytical form of the interatomic potential energy (U) and a set of parameters entering into this form[10]. The parameters are typically obtained either from *ab initio* or semi-empirical quantum mechanical calculations or by fitting to experimental data such as neutron, X-ray and electron diffraction, NMR, IR, Raman and neutron spectroscopy, etc. There are number of force fields available having different degrees of complexity and oriented to treat different kind of systems.

A basic expression for a force field looks like:

$$U = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)]$$
$$+ \sum_{impropers} V_{imp} + \sum_{LJ} 4 \in_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{electrostatics} \frac{q_i q_j}{r_{ij}}$$

In this equation first four terms are for bonded interactions (bond stretching, angle bending, and, dihedral and improper) and last two terms are for non-bonded interactions (12-6 Lennard-Jones potential and the coulombic interactions). There are several force fields present over the years for simulations. It was all started roughly in 1980, when MD and MC simulations of proteins was just begins. Some of these are Amber, CHARMM, OPLS and other. CHARMM is the default force field used by software VMD.

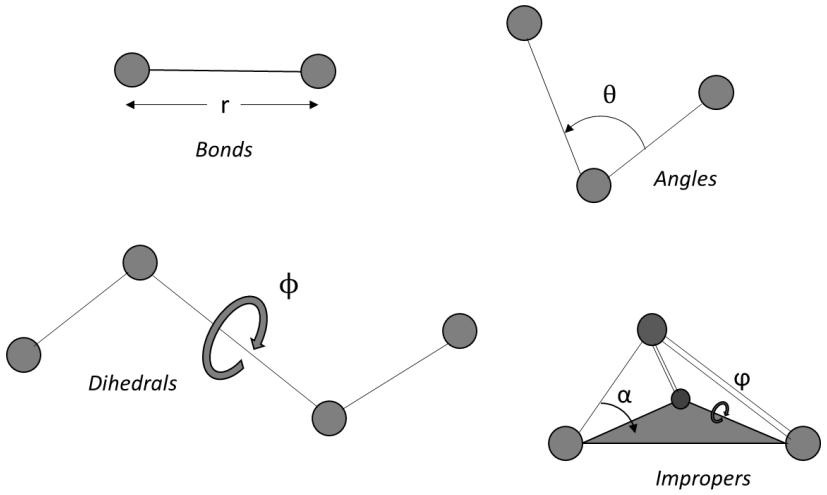

Figure 2 Molecular Parameters

1.6. PDB file

PDB is a file consisting of atomic coordinates of a particular molecule and other details which describes the structure of that biological macromolecule. Structures (atomic coordinates) is well defined by structural biology using experimental techniques such as x-ray crystallography, NMR spectroscopy and cryo electron microscopy. With these techniques they determine the location of each atom relative to another in the molecule and they have a collection these structures on an online server: https://www.rcsb.org

This is a constantly growing server as research is going on in labs all over the world on different biological molecules. On this server every molecule have a unique identity given by the rcsb pdb (Protein Data Bank) server. For simulation we can pick any molecule, we want to work on, to get the basic structure to start with.

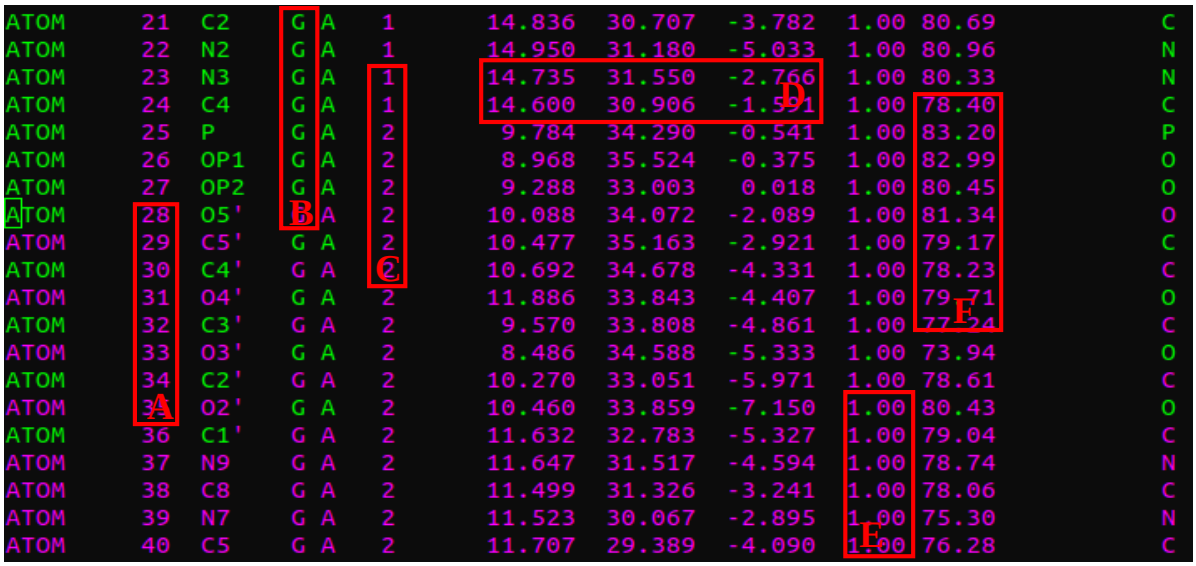

Figure 3 PDB file

- A – *indexes* – every atom has a unique index
- B – *residue* – G represents Guanine here
- C – *residue number* – represent the n^{th} residue (base-pair)
- D – *coordinates* – XYZ coordinate of every atom
- E – *occupancy* – ratio of two different conformation (1 means 50:50)
- F – *beta* – average distance of an atom from its mean position

Occupancy and beta column can also be used to fix or selecting atom for simulation like in SMD, one atom can stay fixed and other will be steered by constant force.

1.7. Topology file

Topology file consist of basic structures like in case of nucleic acids they contain structure and information of bonds and angles about base pairs like guanine, cytosine. Topology file also contain the mass and charge on every atom in every base pair of nucleic acids. It also contains the internal coordinates that allow the automatic assignments of hydrogen atoms and other missing atoms to a PDB file. As coordinates in PDB file is produced by x-ray crystallography and different experimental techniques which doesn't consider any hydrogen atoms, topology file will complete that missing information in the PDB file.

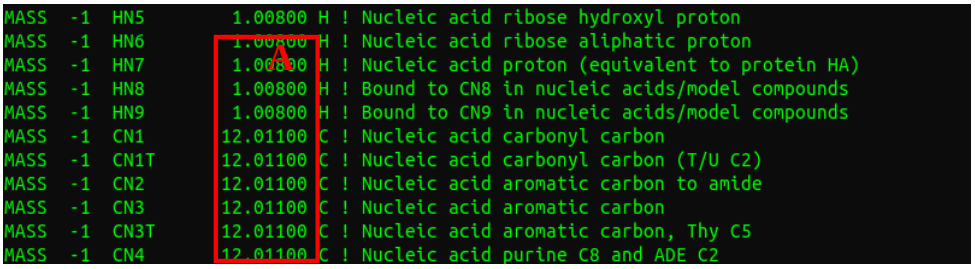

Figure 4 Topology file (1)

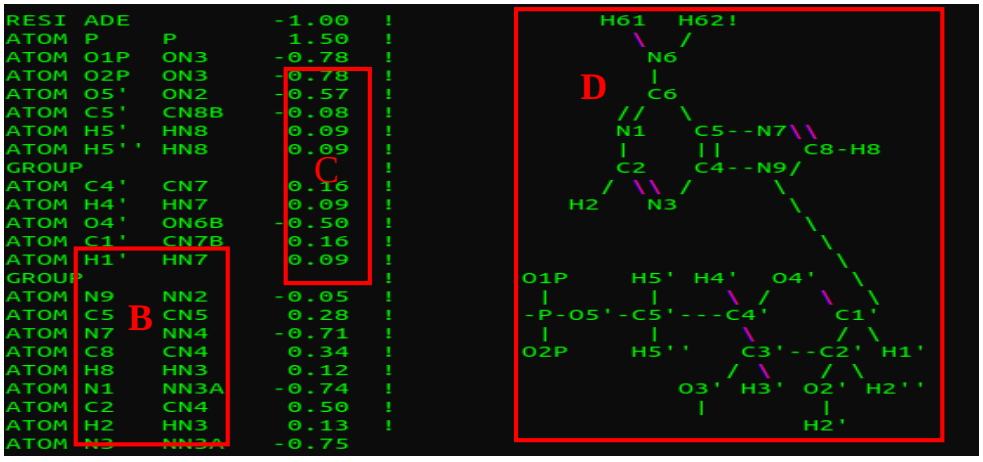

Figure 5 Topology file (2)

- A – atom masses
- B – atom name and atom type

- C – charge on that atom
- D – basic structue of a residue

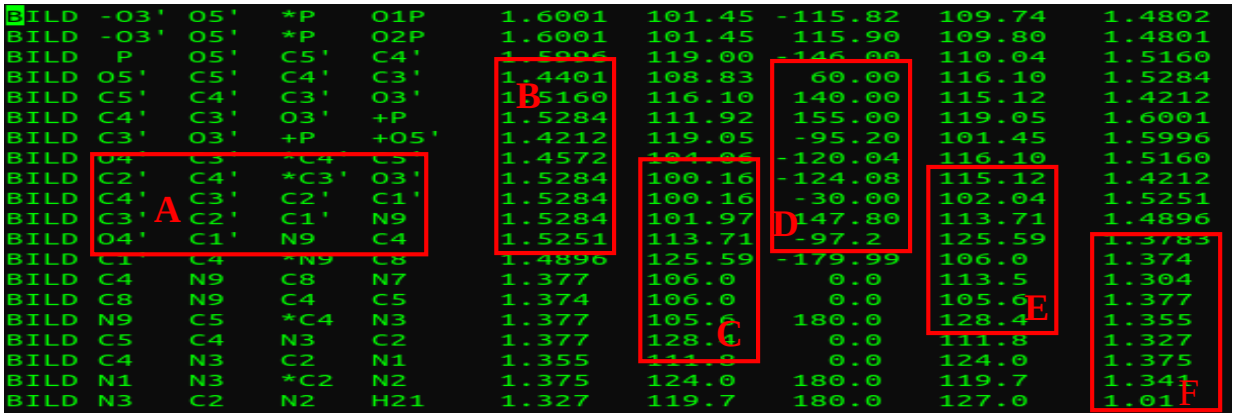

Figure 6 Topology file (3)

- A – *4 different atoms* – identity of four different atoms
- B and F – *bond distance* – bond distance between 1st & 2nd atom and 3rd & 4th atom respectively
- C and E – *angle* – angle between 1st, 2nd & 3rd and 2nd, 3rd & 4th atom respectively
- D – *dihedral* – dihedral angle between 1st, 2nd, 3rd & 4th atoms

1.8. Parameter file

This file contains all the numerical constants that is required by force fields to calculate energies and forces to generate a full PSF file for a particular molecule.

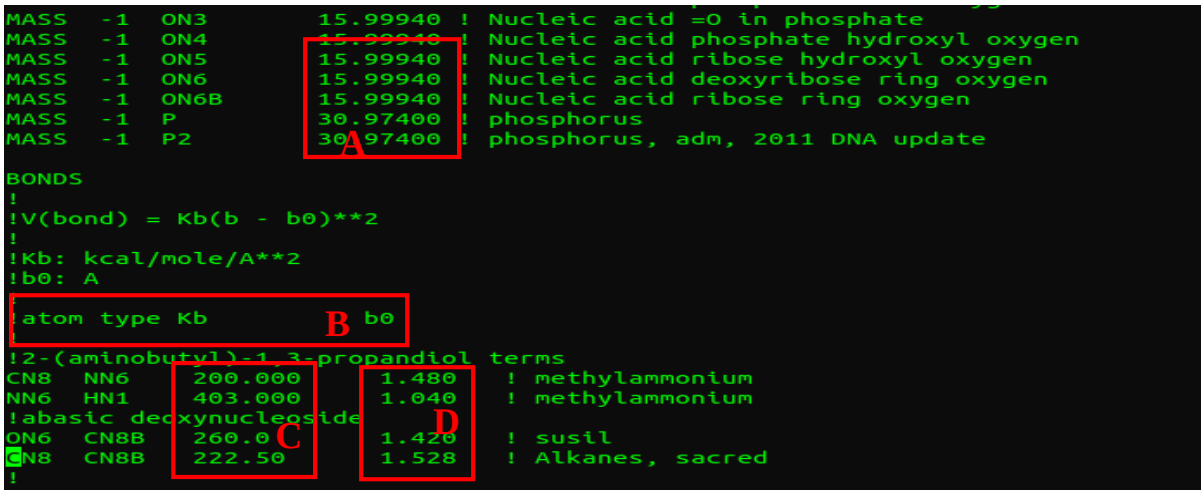

Figure 7 Parameter file

- A – *mass* – masses of different atom
- B – *equation* – general equation for potential energy
- C – *constant* – spring constant acting as bond between two atoms
- D – *equilibrium distance* – bond equilibrium distance between two atoms

1.9. PSF file

This is the file which contains all the information about your molecule. This file is generated by applying force fields on the coordinate file PDB, topology and parameter file. This file has the information about bonds, angles, dihedrals, impropers and non-bonded interactions, all the atoms which have these interactions, their indexes values are mentioned in this file.

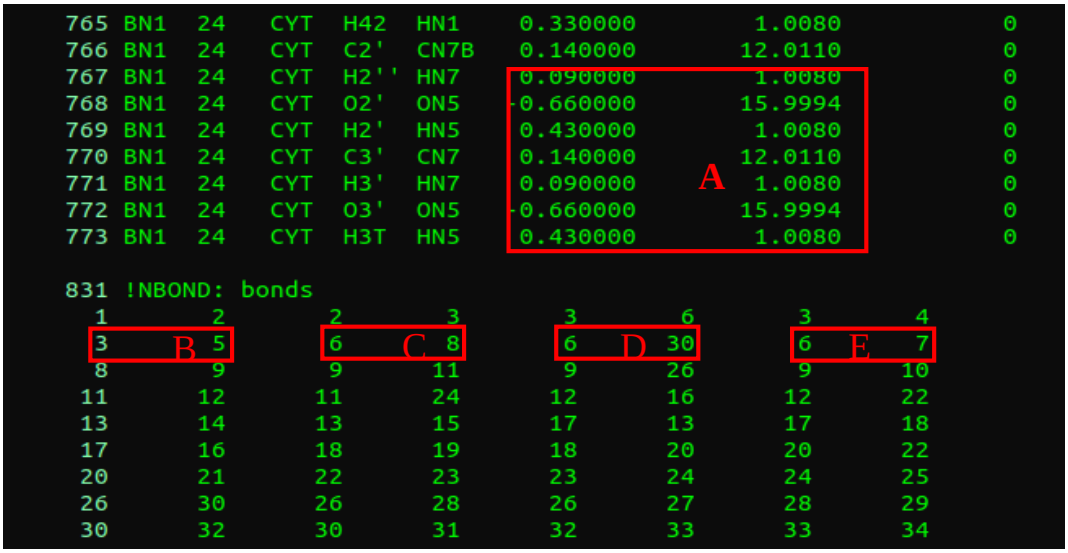

Figure 8 PSF file (1)

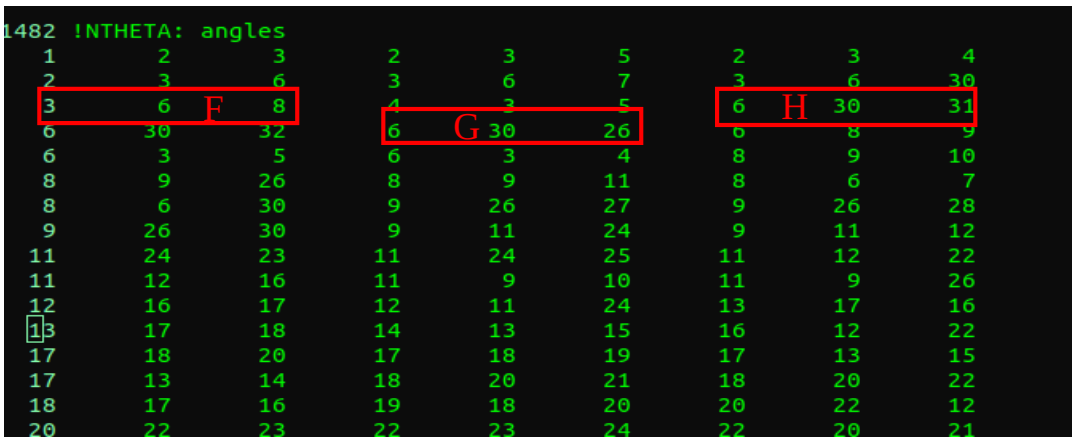

Figure 9 PSF file (2)

- A – charge and masses on atoms
- B, C, D. E – set of atoms having bond in between
- F, G, H – set of atoms having angle in between

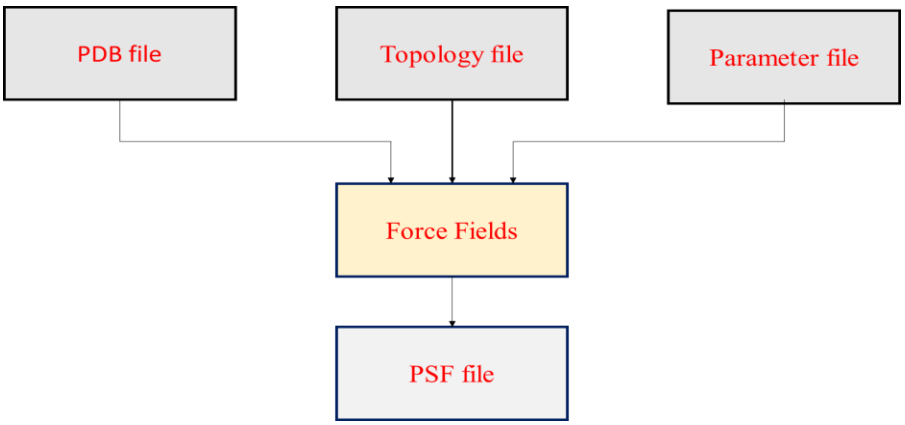

Figure 10 Schematic diagram of PSF file generation

1.10. Solvation

To try to mimic the cellular environment for simulation we will be solvating our system then adding ions and dielectrics and other different parameters create this cellular picture. In order to this, we create a solvent box around a molecule shaped cavity in which our molecule is present and the interaction between our molecule and solvent makes it more real such as short range interactions like H-bonds near the molecule and long range interactions like electrostatics polarization (charge screening). There are two types of solvation exists, implicit and explicit.

1.10.1. Explicit solvation

This type of solvation includes all solvent (generally water) molecules explicitly and all other details like ions in solvent to provide the cellular environment. Because of this much of details, this method is much more accurate but at the same time it is computationally expensive. There are boundary conditions used in this solvation type:

- **Periodic Boundary Conditions (PBC)**

In this scheme, the molecule will be solvated in a solvent cubical box and it has infinite replicas of that solvent box all around it, so with one simulation cell it will be surrounded by 26 solvent cell replicas. If we work only with one simulation box, the molecules on the surface and the molecules which are in bulk experience different forces, so to counteract this surface effect we need to add infinite replicas all around. So during simulation, if a molecule leaves from the simulation box from one side same

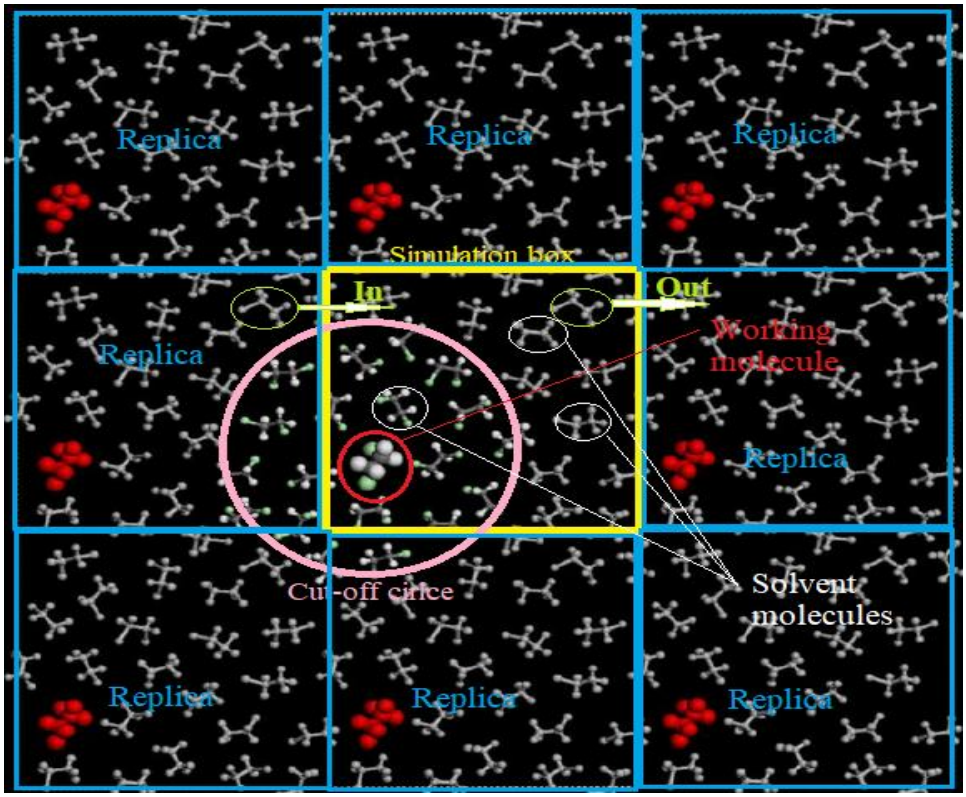

Figure 11 Periodic Boundary Conditions

molecule can enter from the replica of other side.

This PBC condition are usually used with the minimum image convention for short range forces[5]. In this we define a cut of radius beyond which the long range interactions will be truncated to increase the computational efficiency. For consistency of PBC with minimum image convention cut-off distance must be less than or equal to the half of simulation box length, with this configuration any atom *i* interact with only

any other atom *j* from the same image and will not interact with its own image from different replica box.

- **Stochastic Boundary Conditions**

  The working molecule will be solvated in a solvent (water) sphere at 1atm pressure. The solvent molecules are submitted to an additional force field that restrained them in sphere form and also maintains a strong resemblance to bulk water

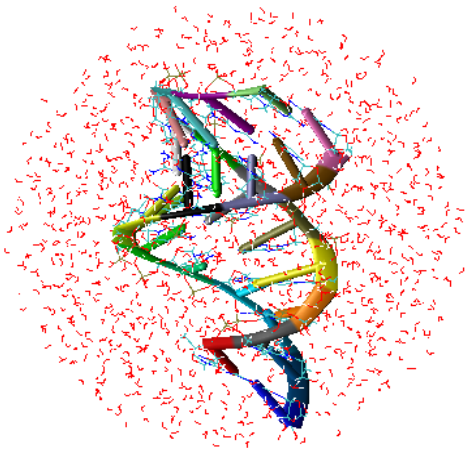

Figure 12 Solvation with SBC

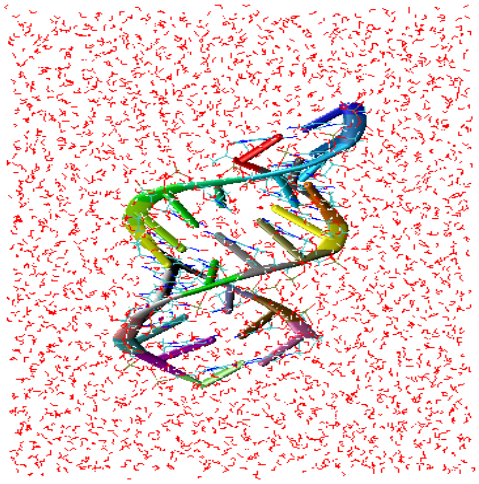

Figure 13 Solvation with PBC

1.10.2.Implicit Solvation

In this solvation, only solvent effect is modelled not the solvent molecules, we treat solvent molecules as a continuous medium of uniform polarizability and of fixed dielectric constant and because of this continuum it doesn't calculate short range interactions which makes it less accurate but simple and computationally easier.

Water has a property called dielectric which is one of the necessary properties with biological point of view. With this dielectric property water screens electrostatic interaction between charged particles so we can model charged particle as a dielectric continuum. In this way, we can calculate electrostatic forces of a

biological system with a bunch of differential equations which can be solved for electric field caused by those charges.

- **Generalized Born Implicit Solvation (GBIS)**

  GBIS is a model to do implicit solvation, the generalized born equation is an approximation of the Poisson-Boltzmann equation. In this model, atoms are considered as charged spheres whose internal dielectric is lower than that of surrounding environment[11]. This model has this functional form:

$$G_s = -\frac{1}{8\pi\varepsilon_0}\left(1-\frac{1}{\varepsilon}\right)\sum_{ij}^{N}\frac{q_i q_j}{f_{GB}}$$

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}} \qquad D = \left(\frac{r_{ij}}{2a_{ij}}\right)^2 \qquad a_{ij} = \sqrt{a_i a_j}$$

where

- ε_0 = the permittivity of free space
- ε = dielectric constant of the solvent
- q = electrostatic charge on particles
- D = distance between particles *i* and *j*
- a = effective born radius

1.11. Minimization and Equilibration

MD minimization and equilibration typically involves more than just one cycle. Usually initial coordinates in PDB file have high energies and forces, they are not very stable so to find a local energy minima we do minimization. But with minimization process system only reaches to local minima and there can be much more energy minima's present so to overcome the energy barrier to reach the other energy minima's, we will heat the system to a desired temperature gradually. Now for further relaxation of molecule after heating we will do equilibration. We can also fix some atoms during this process in starting cycles to save our computational efforts and can release them slowly in further cycles of minimization and equilibration. After this minimization and equilibration we can perform our simulation.

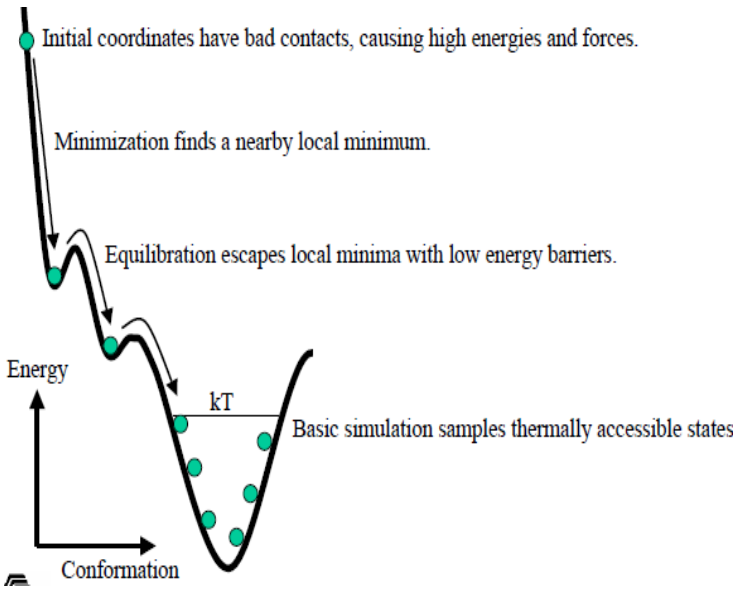

Figure 14 Minimization and Equilibration

1.12. Steered Molecular Dynamics (SMD)

SMD is a type of molecular dynamics simulation in which we will be applying external force to one or more atoms to serve our purpose along with, we can fix other molecules also and can study the behavior of interactions of other molecules with respect to the atom we are applying force on. Why do we do steered molecular dynamics…? It is a compliment to atomic force microscopy and optical tweezers experiment[9][12]. This SMD can be done in two ways: a constant force pulling method and a constant velocity pulling method.

In constant velocity pulling method, we need to connect a dummy atom with SMD atom using a spring and will apply all external forces to the dummy atom which will passes on to SMD atom and this is how SMD atom steers the whole molecule and the force between dummy and SMD atom will be measured using:

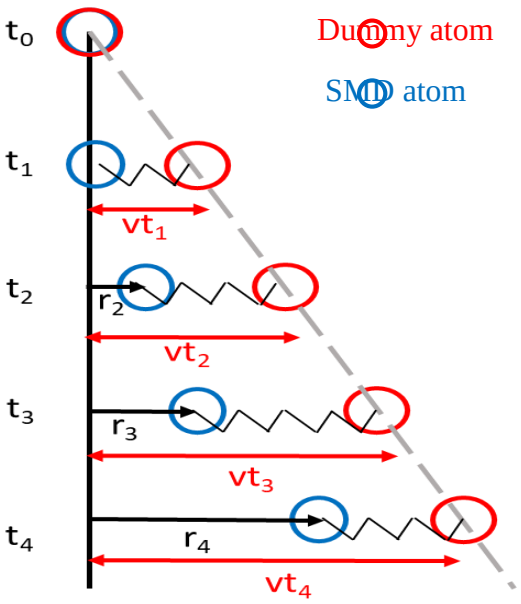

Figure 15 SMD pulling mechanism

$$\vec{F} = -\nabla U$$

$$U = \frac{1}{2}k[vt - (\vec{r} - \vec{r_0}) \cdot \vec{n}]^2$$

where

- U = potential energy
- k = spring constant
- v = pulling velocity
- t = time
- $\vec{r}$ = actual position of SMD
- $\vec{r_0}$ = initial position of SMD
- $\vec{n}$ = direction of pulling

In constant force pulling method, one atom will be fixed and the SMD atom will be pulled with a constant force without any spring or a dummy atom

.

1.13. System to study

➢ RNA (4JRT)

In this work, we will be working on RNA unzipping using SMD technique with constant velocity pulling method. This RNA is picked from an online rcsb server with an id of 4JRT with 12 base pairs having a fasta sequence of 5'-GGGUGGUGCGGG-3' with complementary base pairs 3'-CCUGCACUGCCC-5' consisting of 517 number of atoms. In RNA we will be fixing one atom, Phosphorus of one strand and applying force on another atom Phosphorus of the other strand called SMD atom. With SMD atom, a dummy atom will also be connected using a spring which will not be showing in visualization. So,

as we apply force on dummy atom and RNA will start unzipping by the breaking of H-bonds in between the base pairs. A comparison will be done with DNA unzipping (Prateek Meena's dissertation) to know the difference between the dynamics of these two.

- **DNA hairpin loop**

In another work, we took a DNA with hairpin loop having a fasta sequence of "5′-GCGAGCCATAATCTCATCTG GAAA CAGATGAGATTATGGCTCGC-3′" in both implicit and explicit solvation at high temperatures and also same DNA without hairpin loop for a comparative study. Another comparative study will be done on DNA hairpin loop at low temperature and high temperature (Prateek Meena's dissertation).

Chapter 2

Methodology

The method for performing SMD consists these major steps:

- PSF file generation using force fields
- Solvation in water box (PBC)
- Implicit solvation (GBIS)
- Steered Molecular Dynamics Simulations
- Analysis

2.1. Generating a PSF file

In order to generate a psf file, we require pdb structure file with topology and parameter file using "automatic psf builder" tool in VMD software. In VMD, CHARMM36 force fields applied on the pdb, topology and parameter file to generate a full psf file of a particular structure file.

2.2. Solvation in water box

Solvation is done in a water box with periodic boundary conditions. TIP3P water model is used for solvation. In case of DNA with hairpin, it is dissolved with 0.2 Molar litre^{-1} concentration using "autoionize" tool in while in RNA simply water box is added. Then minimization and equilibration is done to get the local energy minima at a certain temperature for further simulations.

- **RNA (4JRT)**

The RNA molecule consists of 542 atoms but after solvation the atom number increases to 9509. Then minimization was done for 8000 timesteps followed by heating at 300K temperature for 10 ps. Now to avoid strain of molecules after heating, for further relaxation equilibration was done for 20 ps in NPT ensemble.

- **DNA hairpin**

The DNA with hairpin molecule initially has 1398 atoms in with hairpin case and has 1268 in without hairpin case but after solvation the atom number increases to 157480 in hairpin case. Then minimization was done for 20000 timesteps followed by heating at 320K temperature for 100 ps. Now to avoid strain of molecules after heating, for further relaxation equilibration was done for 200 ps in NPT ensemble for hairpin.

2.3. Implicit Solvation (GBIS)

No solvation is done as GBIS only modeled the water effects not the molecules itself. GBIS is switched on in minimization and equilibration configuration file which has different parameters like alpha cutoff, solvation dielectrics, ion concentration etcetera.

- **RNA (4JRT)**

The RNA molecule having __ atoms. Then minimization was done for 8000 timesteps followed by heating at 300K temperature for 10 ps. Now to avoid strain of molecules after heating, for further relaxation equilibration was done for 20 ps in NPT ensemble.

- **DNA hairpin**

The DNA with hairpin molecule has 1398 atoms with hairpin and 1268 without hairpin. Then minimization was done for 8000 timesteps followed by heating at 320K temperature for 100 ps in both hairpin and without hairpin. Now, for further relaxation equilibration was done for 200 ps in NPT ensemble.

2.4. SMD simulation

To perform unzipping, we are using SMD technique in which we will be fixing one atom another will selected as SMD atom which is going to be connected with a dummy atom with the help of a spring and the dummy atom will pull with a constant velocity for unzipping of nucleic acids.

- **RNA (4JRT)** – The atom P from 2nd will kept fixed and SMD atom P selected from 24th (last) residue. Different pulling velocities we used are 0.001 & 0.003 Å/timestep with different spring constants 1, 2, 3, 5 & 7 kcal mol^{-1} Å^{-2} and simulation was done for 100ps in both implicit and explicit case.
- **DNA hairpin –** C3′ atom from 2nd used as fixed atom and SMD atom choses from 44th residue C3′. Pulling velocity of 0.0001 Å/timestep and 0.2 kcal mol^{-1} Å^{-2} spring constant for 6ns and pulling velocity of 0.00005 Å/timestep and 0.2 kcal mol^{-1} Å^{-2} spring constant for 12ns and same parameter used for without hairpin in implicit case. And 0.000011 Å/timestep and 0.26 kcal mol^{-1} Å^{-2} for 45ns in explicit case for with hairpin DNA.

2.5. Analysis

After SMD simulation, a trajectory file is generated of the whole simulation which is used for the analysis for bond distances between fixed and SMD atom and forces for bond breakage which is described in chapter 3 & 4.

Chapter 3

RNA unzipping

In this chapter, we will be analyzing the bond distances (in angstrom) between fixed and SMD atom with different SMD velocities and spring constant throughout the whole simulation (Bond Distance versus Time). Another observation is done on the forces (in picoNewtons) required to break those H-bonds (Force versus Time) throughout the simulation with the corresponding conformation of nucleic acid in both explicit and implicit environment at 300K temperature.

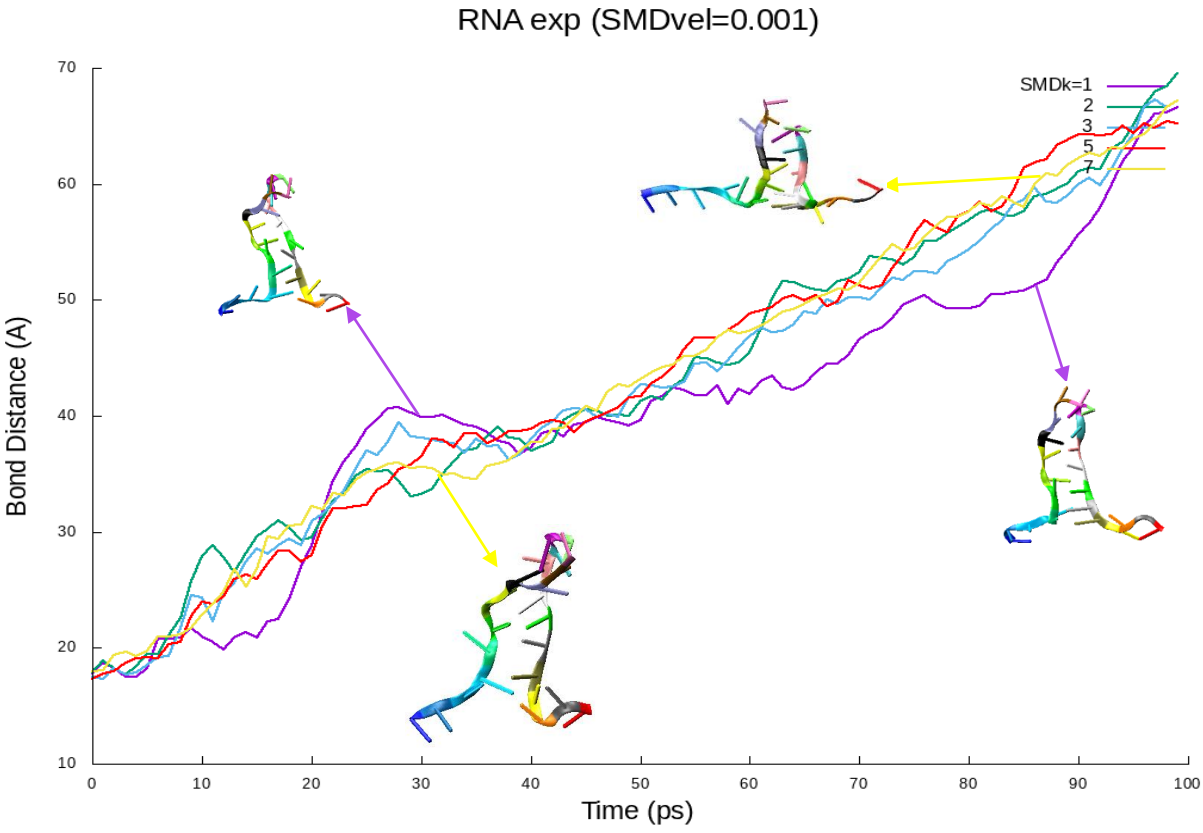

Figure 16 RNA SMDvel=0.001 Explicit Bond Distance

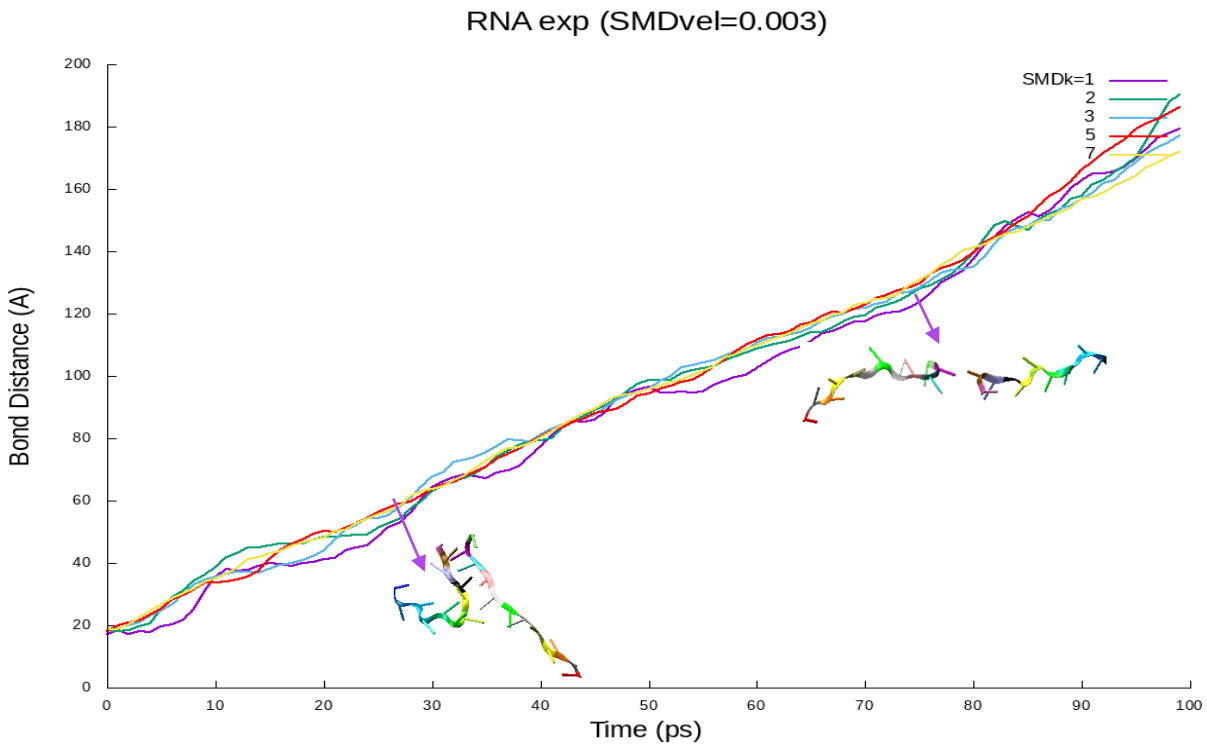

Figure 17 RNA SMDvel=0.003 Explicit Bond Distance

This is a Bond Distance (in Angstrom) versus Time (in picoseconds) for RNA in explicit solvation with pulling SMD velocity of 0.001 and 0.003 Å/timestep with different spring constant k values of 1, 2, 3, 5 and 7 kcal mol^{-1} $Å^{-2}$.

Analysis:

- At any point of time we can see that the bond distance is lower in low k value with respect to higher k values which represents lower k value will unzip RNA slower than higher k value. Also, the graph is getting much smoother with higher k values.
- The short jumps in graph (bond distance) shows the H-bonds breakages.
- With 0.003 pulling velocity, the bond distances are much higher (~190 Å) than at 0.001 pulling velocity (~70 Å).
- With the velocity 0.003 Å/timestep, the graph is even smoother than velocity at 0.001 Å/timestep, but at the same time it's hard to observe the bond break jumps.

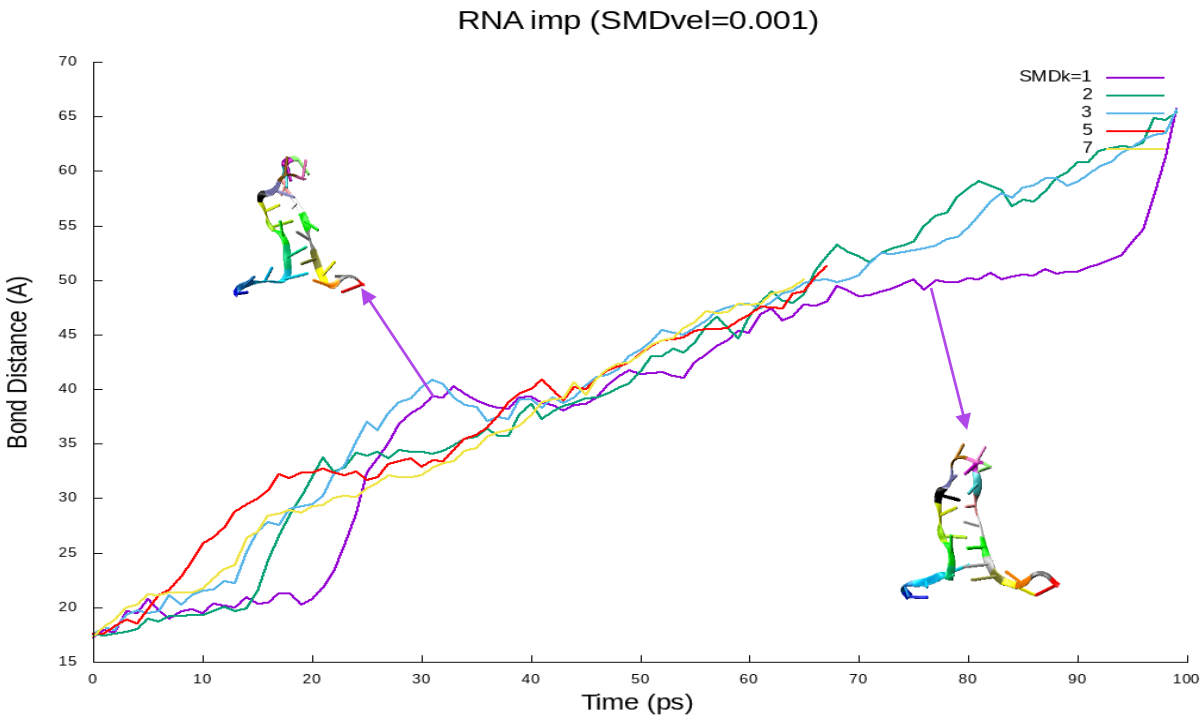

Figure 18 RNA SMDvel=0.001 Implicit Bond Distance

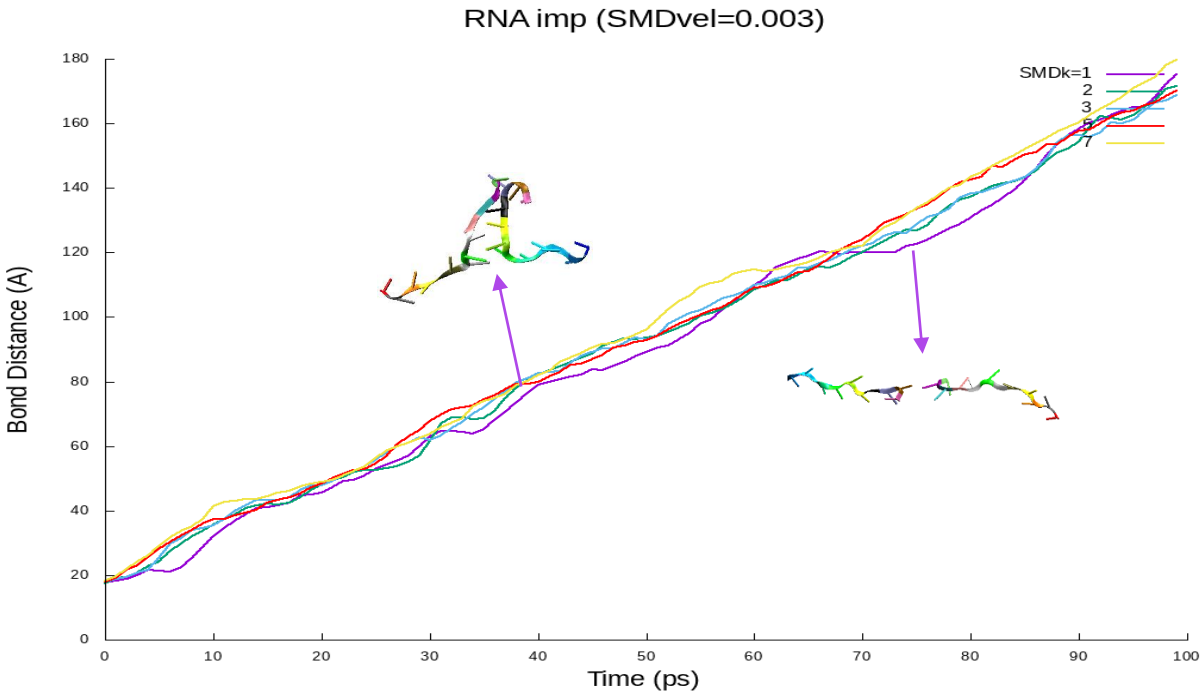

Figure 19 RNA SMDvel=0.003 Implicit Bond Distance

Bond distance versus time graph in implicit solvation with same velocities and k values as in explicit conditions.

Analysis:

- With increase in k values, the graph is getting linearized (smoother).
- The jumps are hard to analyze because of high velocity.
- With low k=1, takes maximum time for first jump (~7.5ps). For higher k, unzipping is faster.
- The bond distance is larger in case of high velocities and smaller in lower velocities.

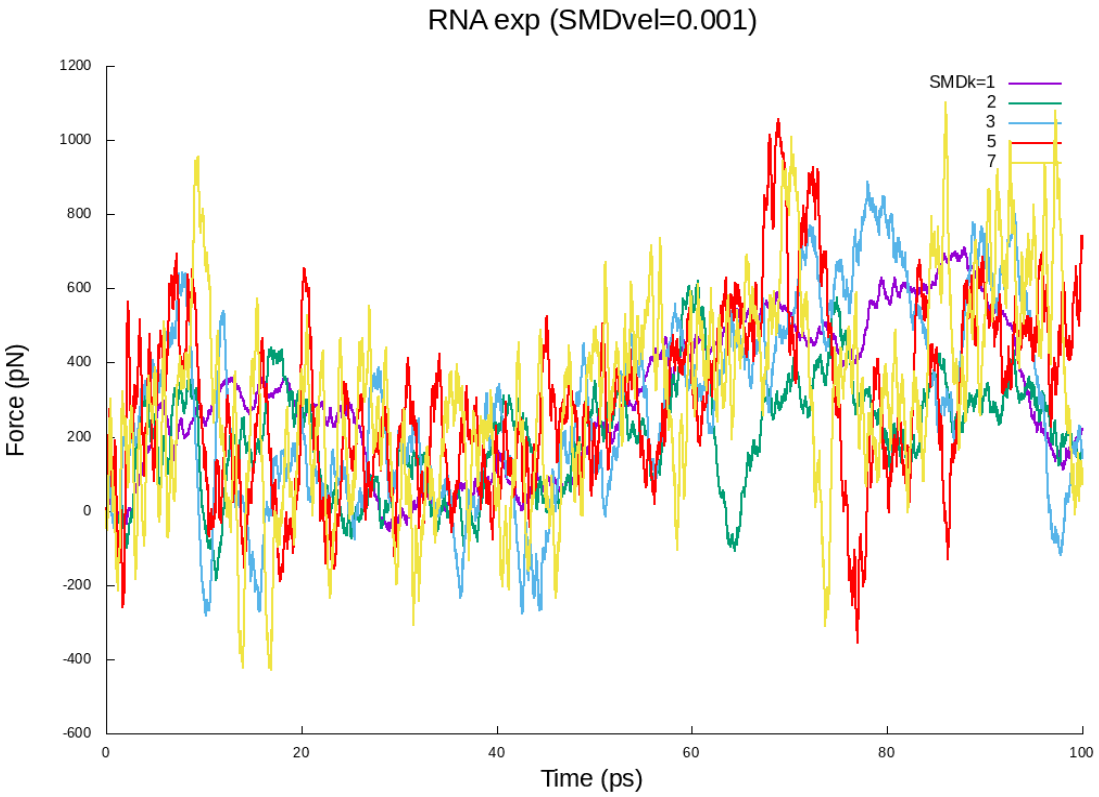

Figure 20 RNA SMDvel=0.001 Explicit Force

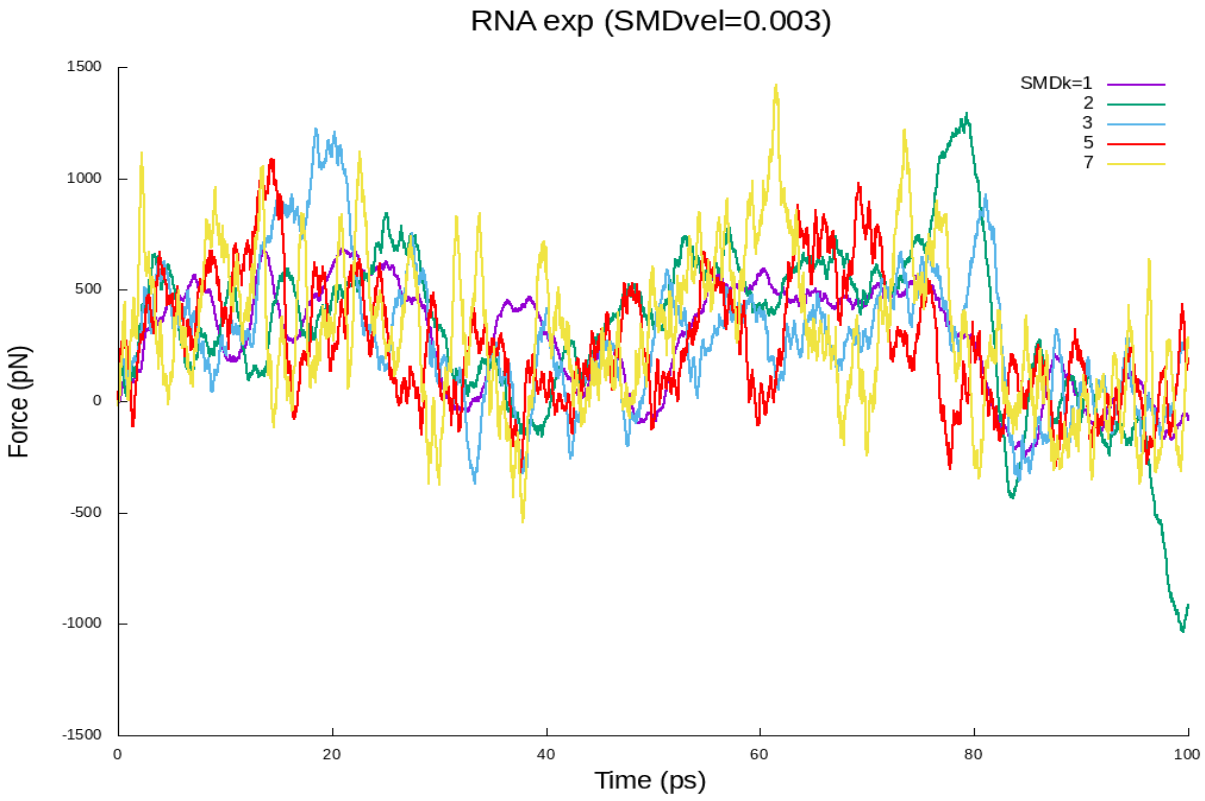

Figure 21 RNA SMDvel=0.003 Explicit Force

Graph between force (pN) and time (ps) with different k values in explicit solvation.

Analysis:

- In lower k values, fluctuations is less showing very low number of bond breakage happen.
- In higher k values, fluctuations in more showing more bond breakages in 100ps timescale.

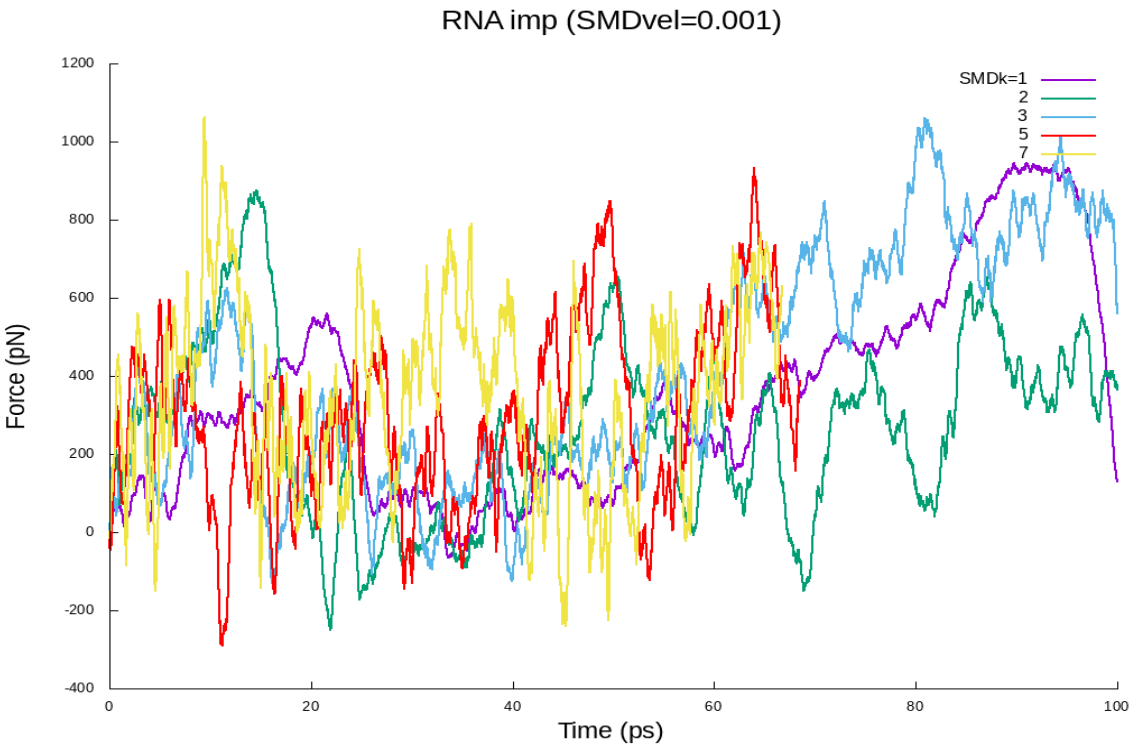

Figure 22 RNA SMDvel=0.001 Implicit Force

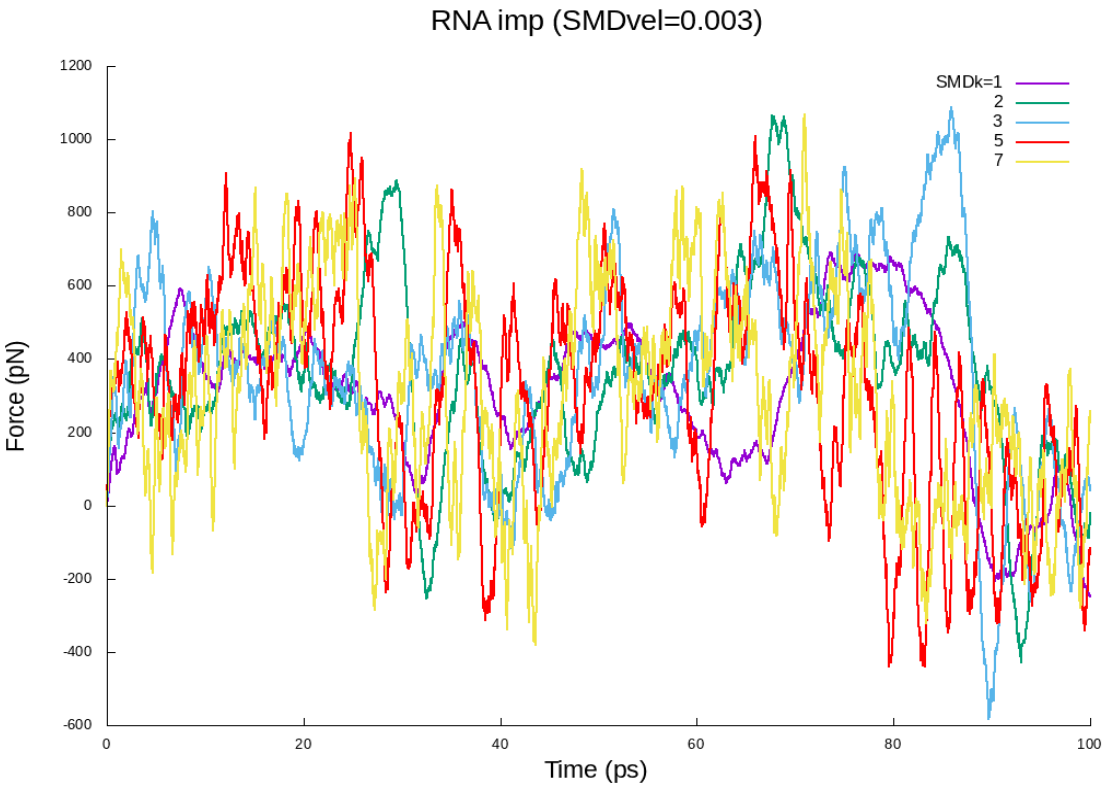

Figure 23 RNA SMDvel=0.003 Implicit Force

Graph between force and timescale in implicit solvation at different puling velocities.

Analysis:

- In lower k values, fluctuations is less showing very low number of bond breakage happen.
- In higher k values, fluctuations in more showing more bond breakages in 100ps timescale.

➢ Comparison with DNA (*Prateek Meena's dissertation*)

- With comparing to DNA, we found that unzipping with respect to bond distances doesn't have much difference to conclude anything.
- From force vs time data, we can say that RNA is less stable than DNA as it requires less force to unzip.

Chapter 4

DNA with hairpin loop

In this chapter, we will be analyzing the bond distances (in angstrom) between fixed and SMD atom with different SMD velocities and spring constant throughout the whole simulation (Bond Distance versus Time). Another observation is done on the forces required to break those H-bonds (Force versus Time) throughout the simulation and will be studying the effect of hairpin loop in this unzipping process in both explicit and implicit environment at 320K temperature.

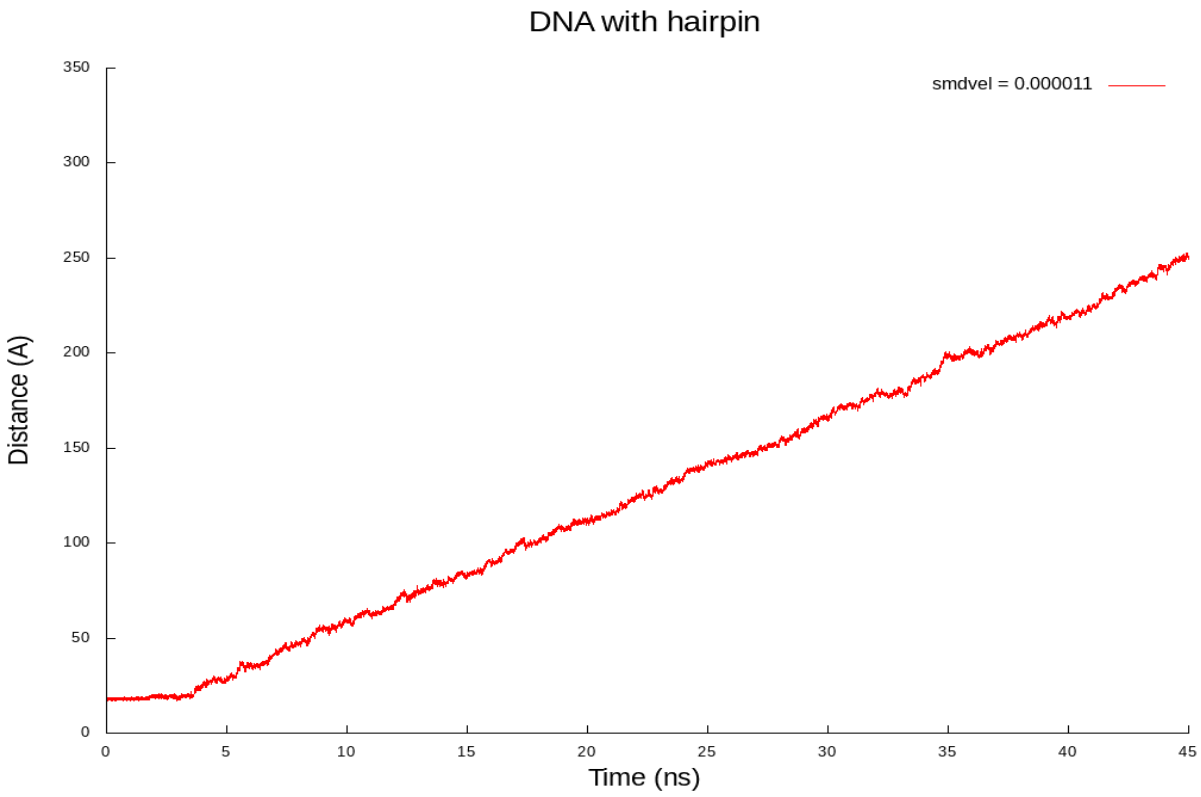

Figure 24 DNA hairpin Explicit Bond Distance

This graph shows bond distance versus time simulation which is in explicit solvation at v=0.000011 Å/timestep and k=0.26 kcal mol^{-1} $Å^{-2}$ at 320K temperature.

Analysis:

- The graph is following a linear trend not showing any jumps or pauses for bond breaking which may be because of high temperature, it's easier to unzip DNA at higher temperatures in explicit solvation method.

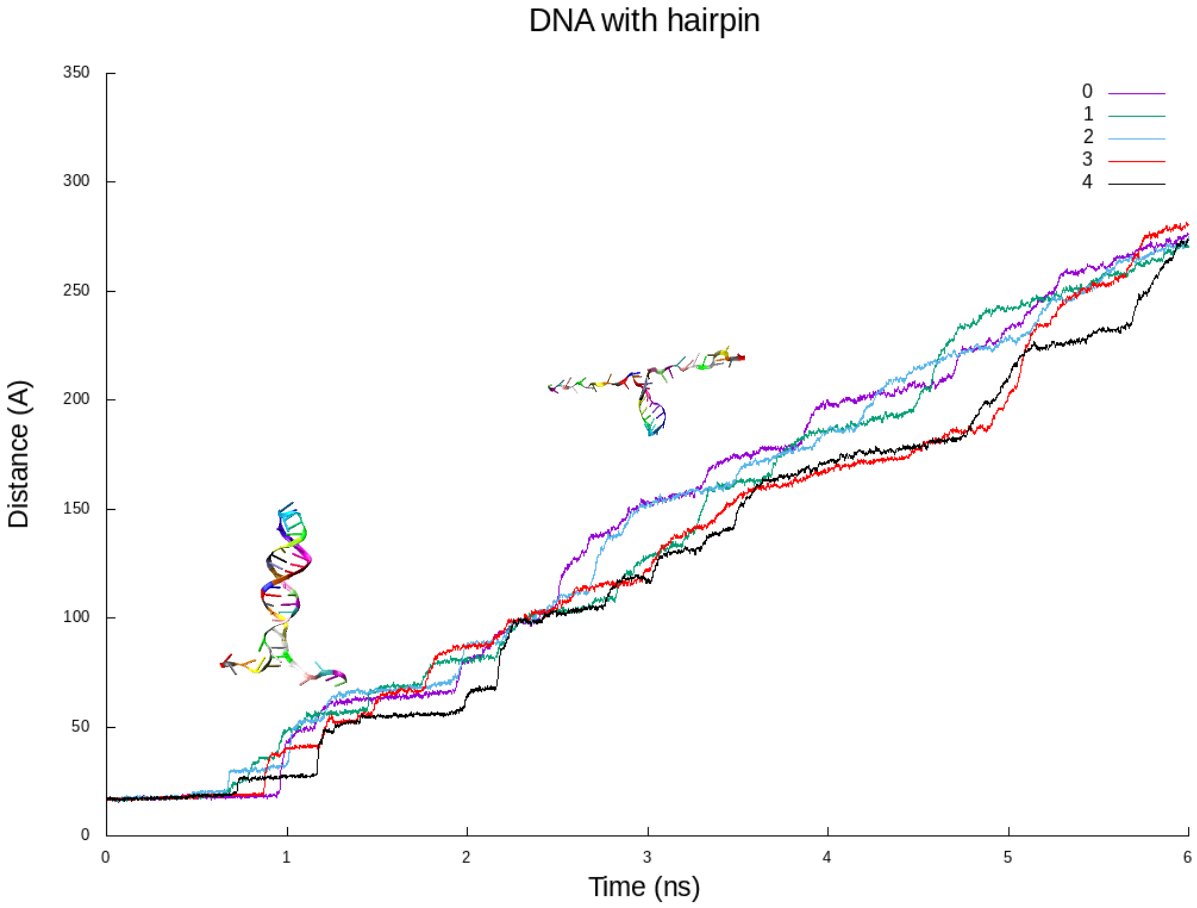

Figure 25 DNA hairpin 6ns Implicit Bond Distance

This simulation is done for 5 times to increase sampling for 6ns at v=0.0001 Å/timestep at k=0.2 kcal mol^{-1} $Å^{-2}$ in implicit solvation.

Analysis:

- This graph shows very clearly that 1st bond breakage happens at ~0.8ns than next is at ~1.2ns with sharp jumps.
- After ~2.5ns, the graph is not very sharp because of the DNA unstability after continuous bond breaking process.

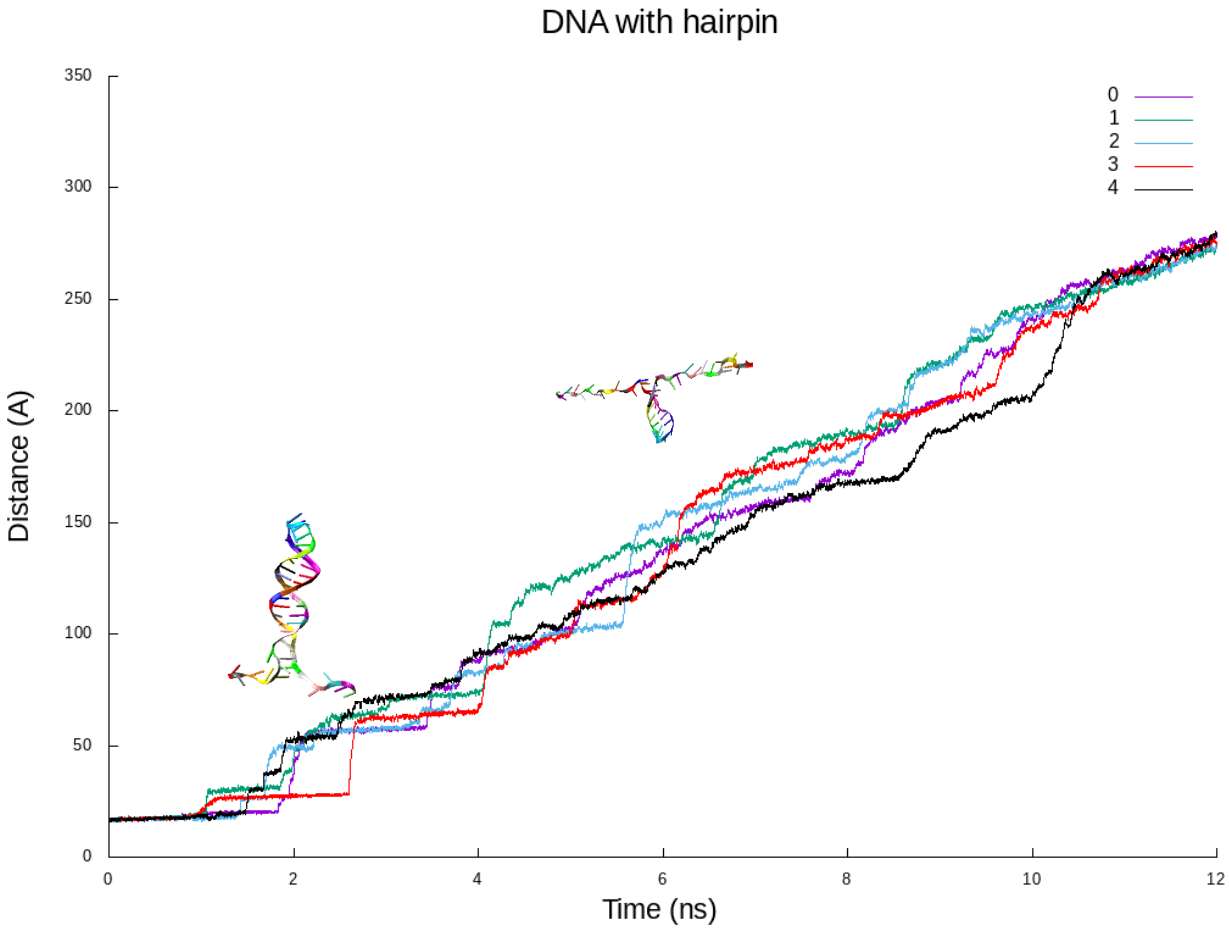

Figure 26 DNA hairpin 12ns Explicit Bond Distance

This simulation is done for 5 times for 12ns at v=0.00005 Å/timestep at k=0.2 kcal mol^{-1} $Å^{-2}$ in implicit solvation.

Analysis:

- This graph is sharper than of 6ns may be because of low velocity it is more precise to examine bond breakage (slow dynamics).
- From graph, we can say that 1^{st-}bond breakage happens at ~1ns than next is at ~2.0ns with sharp jumps.
- After ~4ns, this graph is showing losing its sharp jumps because of its losing stability as bond breakage continues.

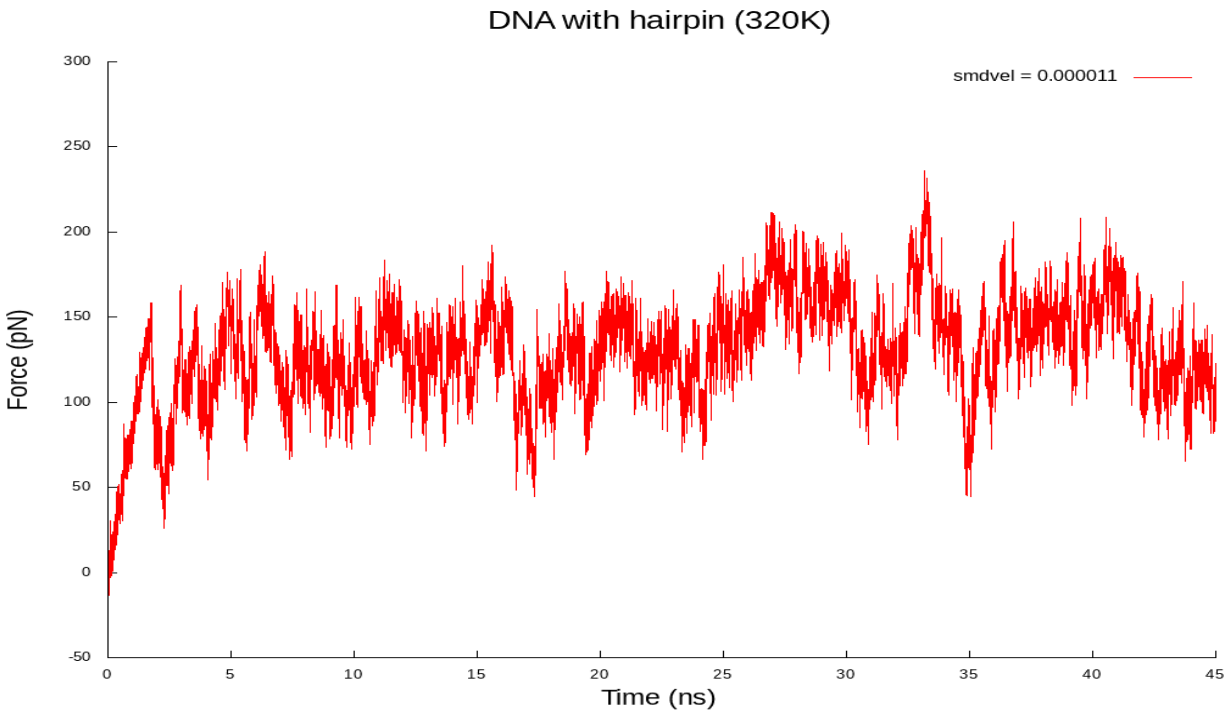

Figure 27 DNA hairpin Explicit Force

Analysis:

- The average force required to unzip this DNA at 320K temperature is ~130pN.
- Can't say about any bond breakage as ups and downs are not very discrete and clear.

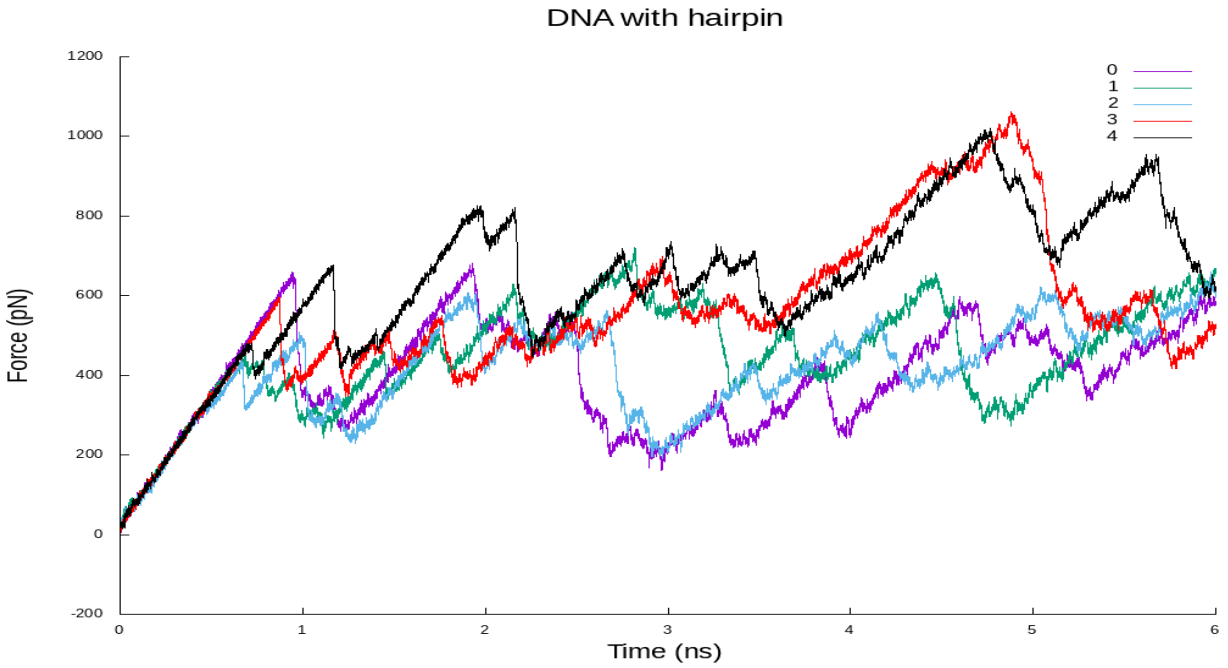

Figure 28 DNA hairpin 6ns Implicit Force

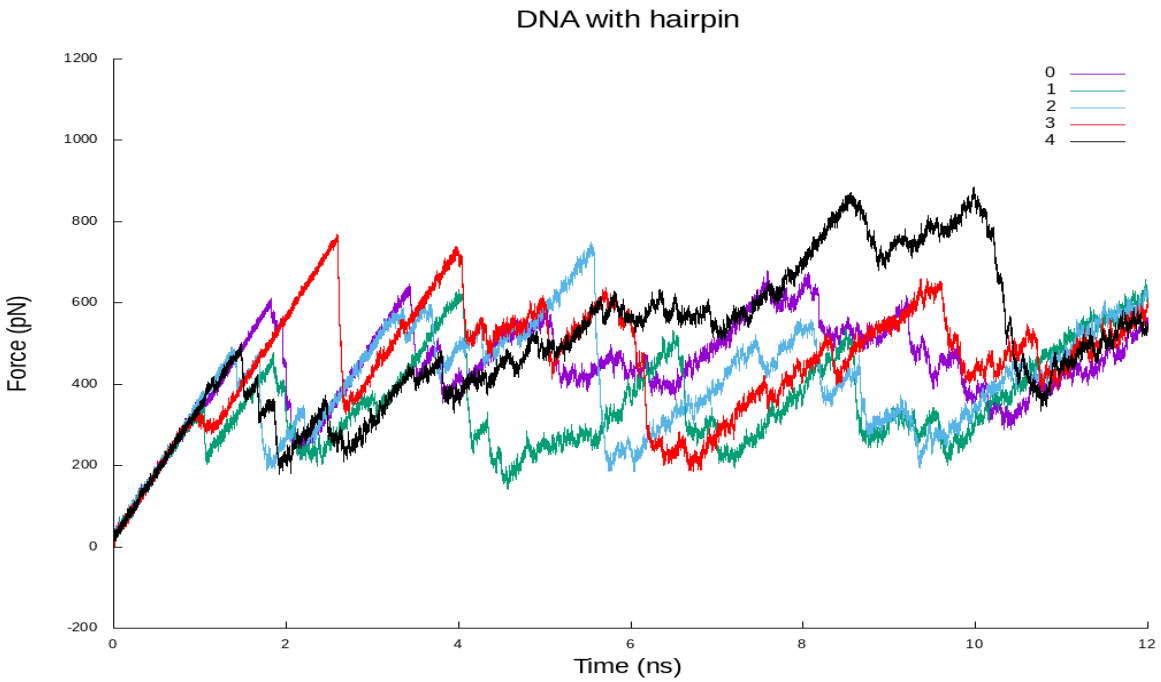

Figure 29 DNA hairpin 12ns Implicit Force

Analysis:

- In starting points, bond breakages is clearly visible with the sudden downs in forces.
- With the variable heights of ups and downs of forces, we can't say exactly that a certain number of bond breakage happen but can predict, more height more bond breakage.

➢ Comparison with DNA without hairpin loop

We took another DNA of same base pair but without any hairpin to study the effect of hairpin loop in this unzipping process. Simulations been done same as DNA with hairpin loop, for 6 and 12 ns at temperature of 320K with pulling velocities 0.0001 and 0.00005 Å/timestep respectively keeping spring constant at 0.2 kcal mol^{-1} $Å^{-2}$, for 5 times.

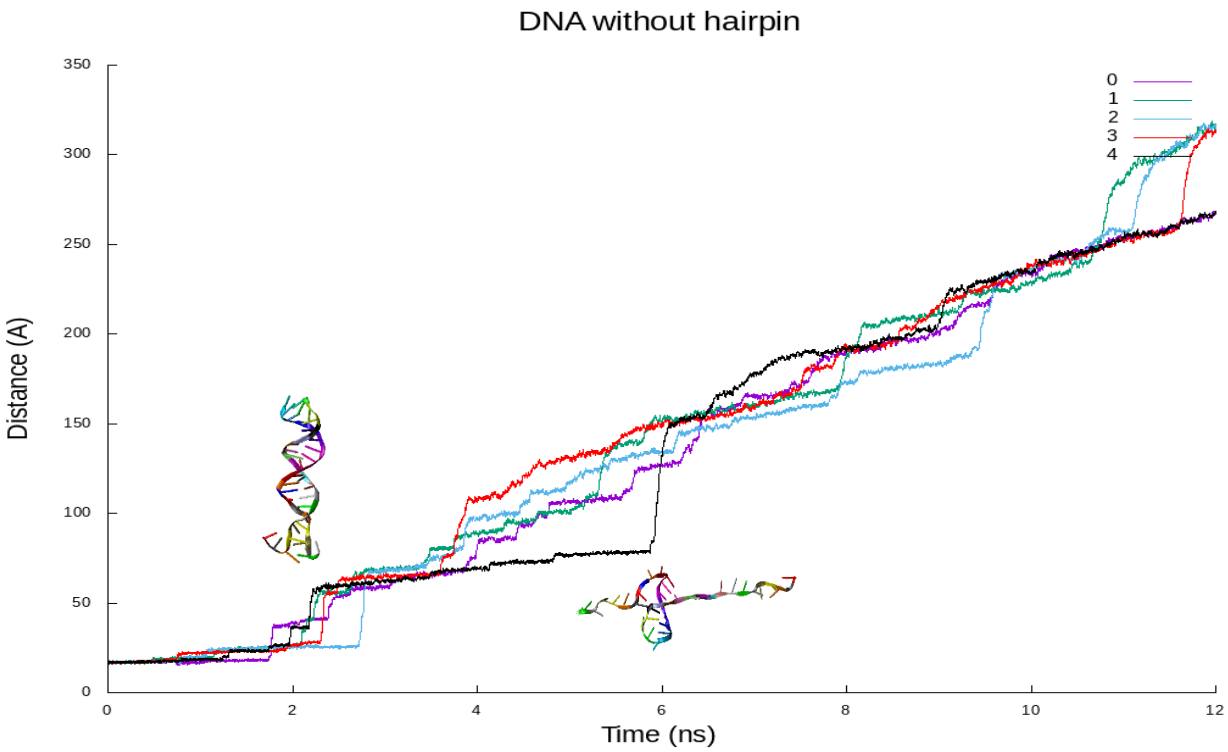

Figure 30 DNA without hairpin 12ns Implicit Bond Distance

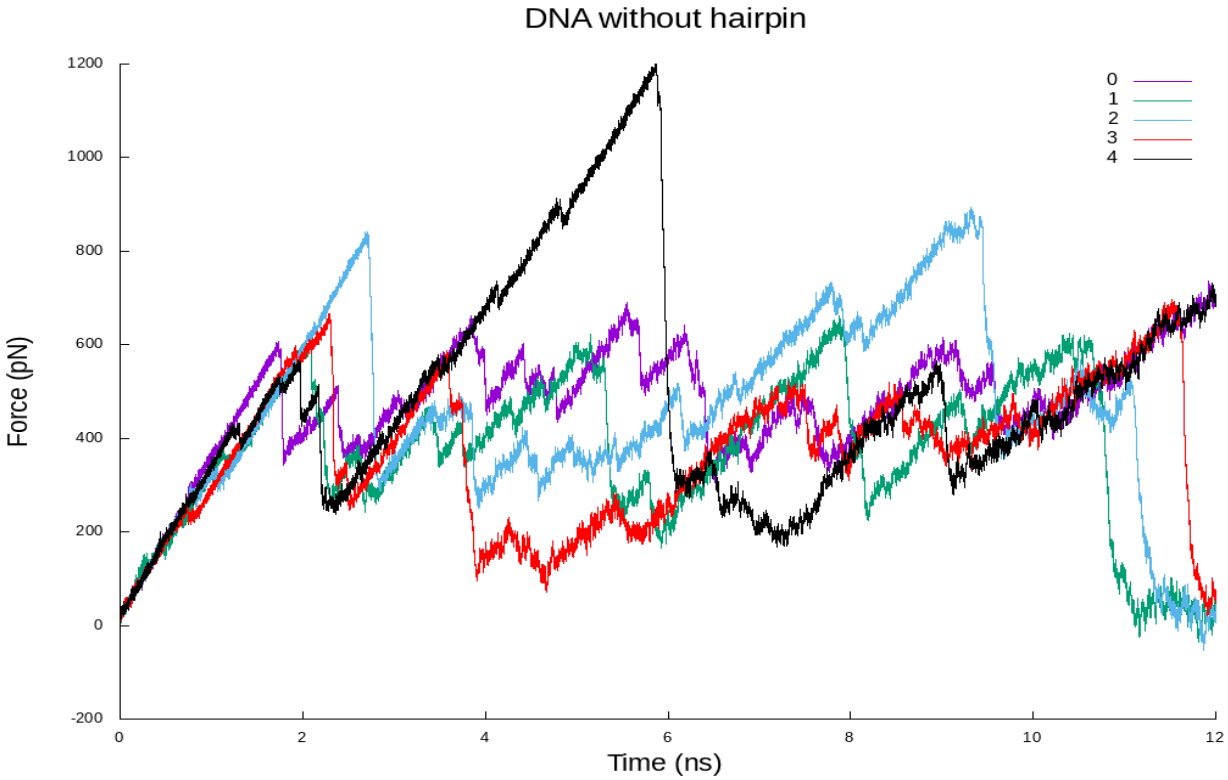

Figure 31 DNA without hairpin 12ns Implicit Force

In comparison with this data, we didn't found any major effect of hairpin loop in DNA unzipping. The dynamics of with and without hairpin are almost same in bond distance and forces data. The only thing differs is that after unzipping the DNA without hairpin splits in two single strands but DNA with hairpin is still a single stranded because of hairpin loop holding the ends of two strands.

➢ Comparison with low temperature at 300K (*Prateek Meena's dissertation*)

- In comparison with 300K, we found that unzipping starts earlier in case of high temperature (320K) than in low temperature.
- Bond breakage is easier at high temperature as there is no sudden jumps unlike at low temperatures. It is more, sort of, continuous bond breaking than at high temperature.
- From force vs time data, we can conclude that at high temperatures less force is required to unzip than at low temperature.

Chapter 5

Conclusions

In this chapter we will be concluding this whole dissertation of unzipping of Nucleic Acids in different solvation environments, different temperatures on different working systems: RNA & DNA, DNA with and without hairpin loops. In conclusion:

➢ DNA vs RNA

- SMD for nucleic acids can be done in both implicit and explicit methods.
- Implicit is a good model to get similar results as explicit.
- DNA is more stable than RNA because of absence of one hydroxyl group (OH).
- Unzipping is much more faster and smoother in high pulling velocities and high spring constant but also it is hard to show the bond breakages jumps in graphs.

➢ DNA hairpin and without hairpin

- A hairpin loop doesn't affect much in the unzipping of the DNA.
- After full unzipping DNA without hairpin separate into two individual strands but in case of hairpin it is simply become a single stranded because of the hairpin connects two ends of the DNA strands.

➢ DNA hairpin in different temperatures

- Explicit results are supposed to give better results than implicit. But we found implicit gives better results than in explicit solvation method.

- From VMD visualization and graphs, it is not fixed that at a time only 1 H-bond will break, it can be 2 or 3 or more at a time.
- As unzipping proceeds, nucleic acids are getting destabilize and sharp jumps & pauses[7] in graphs will vanishes gradually.

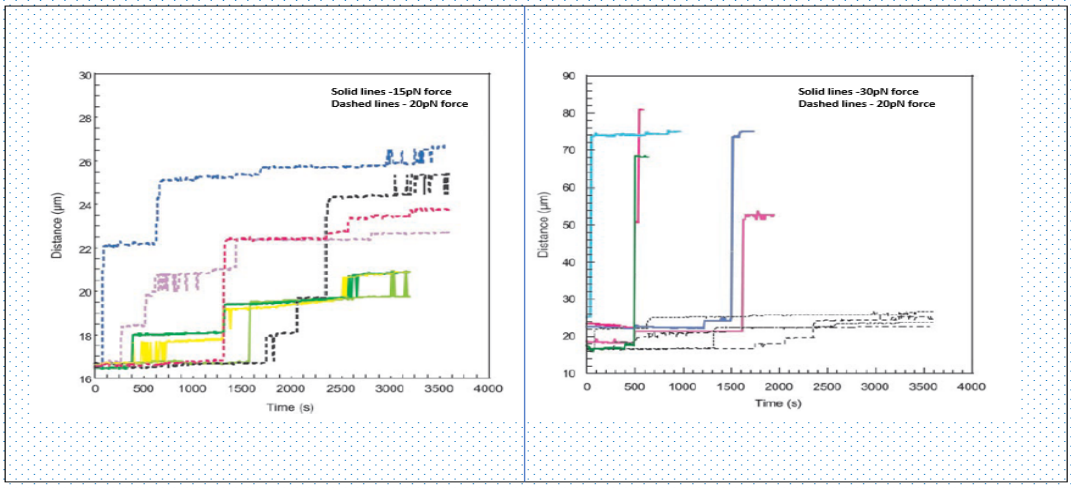

Figure 32 Experimental data Bond breakages [8]

- Our results follow the same trend as experimental work with different forces of pulling
- Temperature plays an important role in unzipping, with high temperatures unzipping is easier than low temperature as it requires more force to unzip at low temperatures.
- The forces required for breaking at different temperatures also follow the same trend as experimental data.
- Longer simulations with low pulling velocities always better to get a more detailed information about the dynamics.

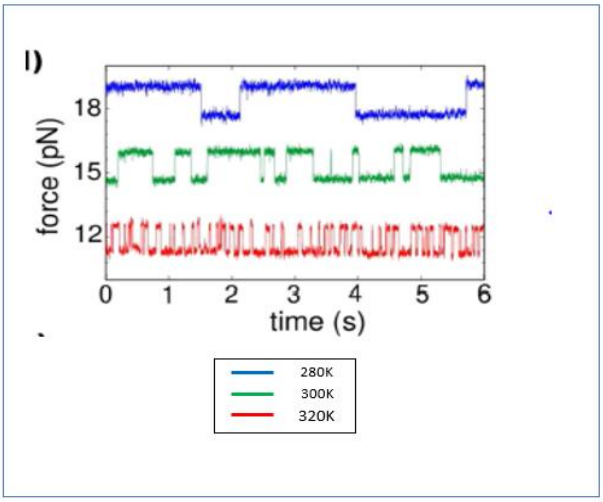

Figure 33 Experimental Data Forces at different temperatures [6]

➢ Future work

- To study the dynamics of AT and GC bonds separately, AT and GC rich nucleic acids will be taken.
- Longer simulation will be run but with slow dynamics for detailed analysis.

Bibliography

1. Online nucleic acid builders

   https://iith.ac.in/3dnus/DNA%20Mismatch.html

   http://structure.usc.edu/make-na/server.html

2. Online Protein Data Bank

   https://www.rcsb.org

3. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26* (16), 1781–1802.
4. Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38.
5. Ponder, J. W.; Case, D. A. *Protein Simulations Advances in Protein Chemistry* **2003**, 27–85.
6. Rico-Pasto, M.; Pastor, I.; Ritort, F. *The Journal of Chemical Physics* **2018**, *148* (12), 123327.
7. Weeks, J.; Lucks, J.; Kafri, Y.; Danilowicz, C.; Nelson, D.; Prentiss, M. *Biophysical Journal* **2005**, *88* (4), 2752–2765.
8. Danilowicz, C.; Coljee, V. W.; Bouzigues, C.; Lubensky, D. K.; Nelson, D. R.; Prentiss, M. *Proceedings of the National Academy of Sciences* **2003**, *100* (4), 1694–1699.
9. Neuman, K. C.; Nagy, A. *Nature Methods* **2008**, *5* (6), 491–505.
10. Cheatham, T. E.; Miller, J. L.; Spector, T. I.; Cieplak, P.; Kollman, P. A. *ACS Symposium Series Molecular Modeling of Nucleic Acids* **1997**, 285–303.

11. Bashford, D.; Case, D. A. *Annual Review of Physical Chemistry* **2000**, *51* (1), 129–152.

12. Lu, H.; Schulten, K. *Proteins: Structure, Function, and Genetics***1999**, *35*(4), 453–463.

13. Alder, B. J.; Wainwright, T. E. (1959). "Studies in Molecular Dynamics. I. General Method". *J. Chem. Phys*. **31** (2): 459.