# Simulating pulling of Nucleic acids using Implicit model

Prateek Meena

*A dissertation submitted for the partial fulfilment of*

*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 “Unzipping of nucleic acid using Constant velocity Steered Molecular Dynamics” submitted by Mr. Prateek Meena (Reg. No. MS14097) for the partial fulfilment of BS-MS dual degree programme 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: April 26, 2019

# Declaration

The work presented by the 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 the part of 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 source listed within have been detailed in the bibliography.

Prateek Meena

Dated: April 26, 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

# Acknowledgement

I would first like to express my sincere gratitude to my supervisor Dr. Monika Sharma for the continuous support in my thesis work, for her patience, motivation and immense knowledge. She gave me opportunity to work with them. I had learned many new things during my thesis work. I want to thank my committee members Dr. Shashi Bhushan Pandit and Dr. Ujjal K Gautam for their suggestions and guidance. I want to thank Naman Kumar Bharti for his support and help. A special thanks to my friends and my family for always being there for me and helping me throughout my thesis work.I want to thank IISER Mohali for providing me all the infrastructure and instruments for my thesis work

# List of figures

# Contents

## 4. Effect of implicit solvent on DNA hairpin and DNA without hairpin



## 5. Conclusion

# Abstract

Unzipping of Double Stranded DNA (dsDNA) into Single Stranded DNA **(ss**DNA) is a ubiquitous process which is central to many biological processes. In order to understand the dynamics and kinetics of these processes, Single molecule micromanipulation techniques has provided a new insight to understand these processes. These techniques include optical tweezers, atomic force microscopy (AFM) and micro niddle. In the present work, we use Steered Molecular Dynamic (SMD) to understand the dynamics. We used different solvation methods includes implicit and explicit solvation. In case of B-DNA dodecamer, A comparative study was done for different spring constant with different velocity and It is found that, at lower spring constant there are long pauses. Another study on DNA hairpin shows that, at high temperature it required less force than the low temperature. In comparison with implicit and explicit it was found that implicit shows better result than explicit. All of these are best explained in the next sections.

# Chapter 1

# Introduction

## 1.1 Theory

Molecular dynamic simulation is a computer simulation method used to studying physical movement of atoms and molecules. Molecular Dynamic simulation is a tool for simulating motions of a multi-body system's atoms. A classical system uses molecular dynamics to calculate the balance and transportation properties (viscosity, thermal conductivity, diffusion, reaction rate, protein folding time, structure and surface coating). Classical Newtonian mechanics can be used to simulate molecular dynamics. It can be used as a supplement to traditional experiments and can learn new things that would otherwise be impossible to find. There are two types of simulation techniques one in Monte Carlo (MC) and other one is Molecular Dynamics (MD). In our work we use Molecular Dynamics as a simulation technique because it gives a route to dynamical properties of the system like time-dependent responses to perturbation, transport coefficient and rheological properties of spectra.

Single molecule micromanipulation methods have provided a new insight into the study of macromolecules. These methods include optical tweezer, micro niddle, atomic force spectroscopy and various technique [(4)]. In single molecule experiment force is directly exerted on molecule and behaviour of the molecule as a function of time is observed. In single molecule micromanipulation we can study the DNA unbinding, unzipping and stretching. We can also study same for the RNA and protein. Among them unzipping of double - stranded (dsDNA) into single - stranded (ssDNA) is the ubiquitous process which is central to many cellular processes.

In our present work we use Steered Molecular Dynamics (SMD) to unzip the DNA. The basic idea behind any SMD simulation is to apply an external force to one or more atoms, which we refer to as SMD atoms. In addition, you can keep another group of atoms fixed and study the behaviour of your system under various conditions. SMD can be done by two ways one is constant force and other one is constant velocity pulling. In the previous experiments it was showed that separation of bases in the DNA is identical to temperature induced process commonly knows as DNA melting. In these processes due to temperature influence A-T rich sites opens first than the G-C bases [(5)]. That means separation of DNA is not sequential. But Molecular Dynamics make it possible to study sequential separation of base pairs.

## 1.2 System to study

we use DNA as a system to study consisting of 12 base pairs having the sequence 5’-CGCGAATTCGCG-3’ in the first part of the experiment. In another part we use a DNA with hairpin loop consisting of 44 base pairs having the sequence 5’-

GCGAGCCATAATCTCAATCTG GAAA CAGATGAGATTATGGCTCGC-3' with comparison to same sequence without hairpin. Deoxyribonucleic acid (DNA) is consists of molecules called nucleotides. Each nucleotide contains a phosphate group, a sugar group and a nitrogen base. The four types of nitrogen bases are adenine (A), thymine (T), guanine (G) and cytosine (C). These bases linked together by hydrogen bonds A-T have double bond between them while G-C have triple bond. As a result, it effects the dynamic and kinetic during the unzipping of DNA. In addition to this DNA have a coiled structure So, DNA requires two to three times more force to rupture when force is applied across opposite ends of the duplex, in a shearing geometry, as opposed to when forces are applied to the same end of the duplex to unzip it instead (ref 6).

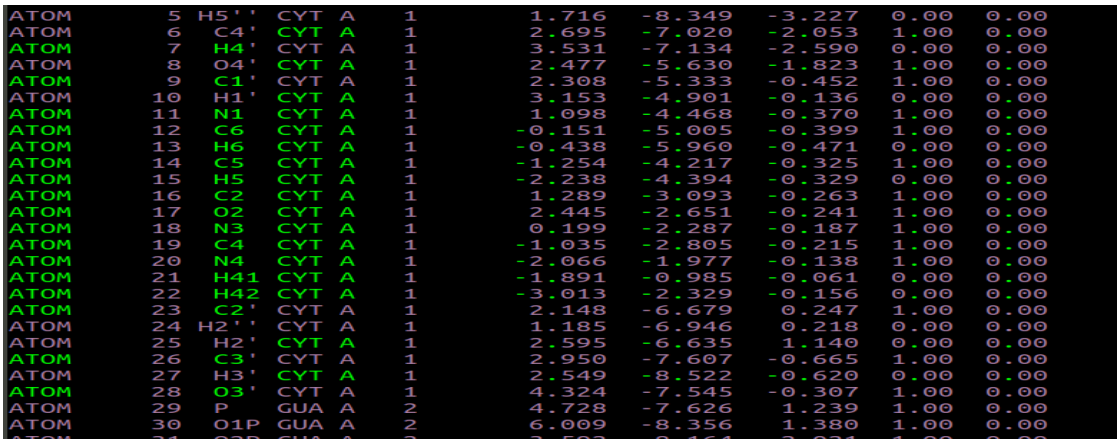

Figure 1.1 System to study (a) DNA dodecamer (b) DNA with hairpin (c) DNA without hairpin

The dodecamer of helical B-DNA with the sequence $d(CGCGAATTCGCG)_2$ was made through an online software (http://structure.usc.edu/make-na/) and the hairpin loop DNA and DNA without hairpin loop was made through an online software (https://iith.ac.in/3dnus/) to perform the further simulation. This generates a PDB file as a basis to start the simulation. In order to perform all the simulations, we required PDB file, PSF file, CHARMM topology file, CHARMM parameter file. All of these files are the basis for start the simulation.

### 1.3 PDB file

The term PDB can allude to the Protein Data Bank (http://www.rcsb.org/pdb/), to an information record gave there, or to any document following the PDB design. Documents in the PDB incorporate data, for example, the name of the compound, the species and tissue from which is was gotten, creation, amendment history, diary reference, references, amino acid arrangement, stoichiometry, auxiliary structure areas, crystal lattice and symmetry group, lastly the ATOM and HETATM records containing the directions of the protein and any waters, particles, or different heterogeneous atom in the crystal. Some PDB documents

incorporate different arrangements of directions for a few or all molecules. Because of the limit of x-beam crystallography and NMR structure examination, the directions of hydrogen molecules are excluded in the PDB.



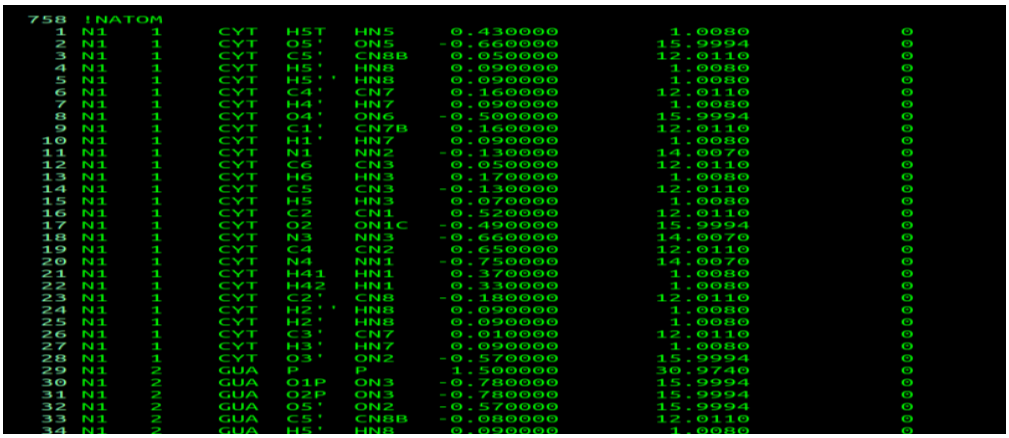
```
ATOM      5 H5''  CYT A   1       1.716   -8.349   -3.227  0.00  0.00
ATOM      6  C4'  CYT A   1       2.695   -7.020   -2.053  1.00  0.00
ATOM      7  H4'  CYT A   1       3.531   -7.134   -2.590  0.00  0.00
ATOM      8  O4'  CYT A   1       2.477   -5.630   -1.823  1.00  0.00
ATOM      9  C1'  CYT A   1       2.308   -5.333   -0.452  1.00  0.00
ATOM     10  H1'  CYT A   1       3.153   -4.901   -0.136  0.00  0.00
ATOM     11  N1   CYT A   1       1.098   -4.468   -0.370  1.00  0.00
ATOM     12  C6   CYT A   1      -0.151   -5.005   -0.399  1.00  0.00
ATOM     13  H6   CYT A   1      -0.438   -5.960   -0.471  0.00  0.00
ATOM     14  C5   CYT A   1      -1.254   -4.217   -0.325  1.00  0.00
ATOM     15  H5   CYT A   1      -2.238   -4.394   -0.329  0.00  0.00
ATOM     16  C2   CYT A   1       1.289   -3.093   -0.263  1.00  0.00
ATOM     17  O2   CYT A   1       2.445   -2.651   -0.241  1.00  0.00
ATOM     18  N3   CYT A   1       0.199   -2.287   -0.187  1.00  0.00
ATOM     19  C4   CYT A   1      -1.035   -2.805   -0.215  1.00  0.00
ATOM     20  N4   CYT A   1      -2.066   -1.977   -0.138  1.00  0.00
ATOM     21  H41  CYT A   1      -1.891   -0.985   -0.061  0.00  0.00
ATOM     22  H42  CYT A   1      -3.013   -2.329   -0.156  0.00  0.00
ATOM     23  C2'  CYT A   1       2.148   -6.679    0.247  1.00  0.00
ATOM     24 H2''  CYT A   1       1.185   -6.946    0.218  0.00  0.00
ATOM     25  H2'  CYT A   1       2.595   -6.635    1.140  0.00  0.00
ATOM     26  C3'  CYT A   1       2.950   -7.607   -0.665  1.00  0.00
ATOM     27  H3'  CYT A   1       2.549   -8.522   -0.620  0.00  0.00
ATOM     28  O3'  CYT A   1       4.324   -7.545   -0.307  1.00  0.00
ATOM     29  P    GUA A   2       4.728   -7.626    1.239  1.00  0.00
ATOM     30  O1P  GUA A   2       6.009   -8.356    1.380  1.00  0.00
```

Figure 1.2 - PDB file Format

In this figure on reading from left to right are record type, atom ID, atom name, residue name, residue ID, x, y, and Z coordinates, occupancy and temperature factor which is called beta.

### 1.4 PSF file

A PSF file is also known as protein structure file because it contains all the molecular specific information which can be used to apply a particular type of force field. In our case we used CHARMM force field as a force field parameter. This force field divided into two files one is topology and other one is parameter file. A topology file defines the atom type, atom names, bonds and partial charge of each residue and any patches which is useful to link or mutate these basic residues. A parameter file contains information about mapping between bonded and non-bonded interaction including various combination of atoms types and spring constants and similar parameter for all the bond, angles, dihedral, improper and van der Waals terms in potential function. Atoms, bond, angles, improper, dihedral and cross term are the main section of interest of PSF file.



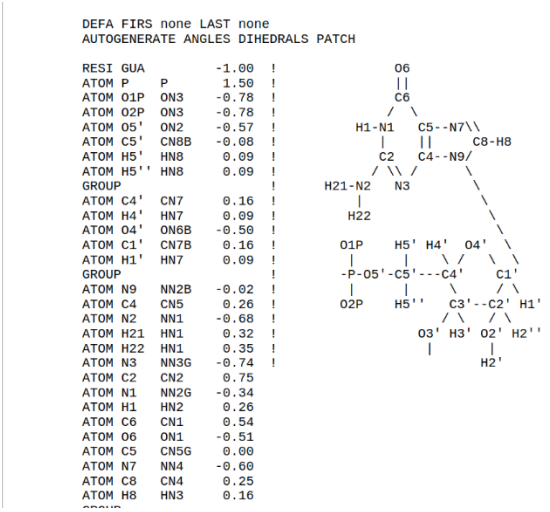
```
758 !NATOM
  1 N1    1       CYT   H5T    HN5     0.430000         1.0080           0
  2 N1    1       CYT   O5'    ON5    -0.660000        15.9994           0
  3 N1    1       CYT   C5'    CN8B    0.050000        12.0110           0
  4 N1    1       CYT   H5'    HN8     0.090000         1.0080           0
  5 N1    1       CYT   H5''   HN8     0.090000         1.0080           0
  6 N1    1       CYT   C4'    CN7     0.160000        12.0110           0
  7 N1    1       CYT   H4'    HN7     0.090000         1.0080           0
  8 N1    1       CYT   O4'    ON6    -0.500000        15.9994           0
  9 N1    1       CYT   C1'    CN7B    0.160000        12.0110           0
 10 N1    1       CYT   H1'    HN7     0.090000         1.0080           0
 11 N1    1       CYT   N1     NN2    -0.130000        14.0070           0
 12 N1    1       CYT   C6     CN3     0.050000        12.0110           0
 13 N1    1       CYT   H6     HN3     0.170000         1.0080           0
 14 N1    1       CYT   C5     CN3    -0.130000        12.0110           0
 15 N1    1       CYT   H5     HN3     0.070000         1.0080           0
 16 N1    1       CYT   C2     CN1     0.520000        12.0110           0
 17 N1    1       CYT   O2     ON1C   -0.490000        15.9994           0
 18 N1    1       CYT   N3     NN3    -0.660000        14.0070           0
 19 N1    1       CYT   C4     CN2     0.650000        12.0110           0
 20 N1    1       CYT   N4     NN1    -0.750000        14.0070           0
 21 N1    1       CYT   H41    HN1     0.370000         1.0080           0
 22 N1    1       CYT   H42    HN1     0.330000         1.0080           0
 23 N1    1       CYT   C2'    CN8    -0.180000        12.0110           0
 24 N1    1       CYT   H2''   HN8     0.090000         1.0080           0
 25 N1    1       CYT   H2'    HN8     0.090000         1.0080           0
 26 N1    1       CYT   C3'    CN7     0.010000        12.0110           0
 27 N1    1       CYT   H3'    HN7     0.090000         1.0080           0
 28 N1    1       CYT   O3'    ON2    -0.570000        15.9994           0
 29 N1    2       GUA   P      P       1.500000        30.9740           0
 30 N1    2       GUA   O1P    ON3    -0.780000        15.9994           0
 31 N1    2       GUA   O2P    ON3    -0.780000        15.9994           0
 32 N1    2       GUA   O5'    ON2    -0.570000        15.9994           0
 33 N1    2       GUA   C5'    CN8B   -0.080000        12.0110           0
 34 N1    2       GUA   H5'    HN8     0.090000         1.0080           0
```

Figure 1.3 PSF file format

In figure reading from left to right atom ID, segment name, residue ID, resudue name, atom name, atom type, charge, mass and an unused 0.

### 1.5 Topology file

A CHARMM forcefield topology file contains all of the information which is needed to convert a list of residue names into a PSF structure file. In crystal PDB file some of the coordinates of hydrogen and other atoms are missing. So, it contains internal coordinates that allow the automatic assignment of coordinates. We use current version of force field CHARMM36 which is good for nucleic acids.



```
DEFA FIRS none LAST none
AUTOGENERATE ANGLES DIHEDRALS PATCH

RESI GUA          -1.00  !
ATOM P    P        1.50  !
ATOM O1P  ON3     -0.78  !
ATOM O2P  ON3     -0.78  !
ATOM O5'  ON2     -0.57  !
ATOM C5'  CN8B    -0.08  !
ATOM H5'  HN8      0.09  !
ATOM H5'' HN8      0.09  !
GROUP                    !
ATOM C4'  CN7      0.16  !
ATOM H4'  HN7      0.09  !
ATOM O4'  ON6B    -0.50  !
ATOM C1'  CN7B     0.16  !
ATOM H1'  HN7      0.09  !
GROUP                    !
ATOM N9   NN2B    -0.02  !
ATOM C4   CN5      0.26  !
ATOM N2   NN1     -0.68  !
ATOM H21  HN1      0.32  !
ATOM H22  HN1      0.35  !
ATOM N3   NN3G    -0.74  !
ATOM C2   CN2      0.75
ATOM N1   NN2G    -0.34
ATOM H1   HN2      0.26
ATOM C6   CN1      0.54
ATOM O6   ON1     -0.51
ATOM C5   CN5G     0.00
ATOM N7   NN4     -0.60
ATOM C8   CN4      0.25
ATOM H8   HN3      0.16
```

Figure 1.4 Topology file format

### 1.6 Parameter file

A CHARMM forcefield parameter file contains all of the numerical constants needed to evaluate forces and energies, given a PSF structure file and atomic coordinates. The parameter file is closely related to the topology file that was used to generate the PSF file, and the two are typically distributed together and given matching names. So, what are the force field and how they used to perform the simulations? Force field is just a mathematical form which describe the dependence of the energy on a system of N particles [(9)]. It is composed of interatomic potential energy of system of N particle.

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

In this equation first four terms which includes bond stretching, bending, dihedral and improper torsion are called intramolecular term and last two terms are the van der Waals interaction and the coulombic interaction.

### 1.7 Software used

To perform all the simulations, we use NAMD and VMD. NAMD is a parallel programming code which is designed for high performance simulation of biomolecules. It uses the popular graphic program VMD for simulation set up and for analyzing trajectory. It is file compatible with many other program like AMBER, CHARMM and X-PLOR [6,7]. In order to do so it uses various algorithms and concept to perform the dynamics.

### 1.8 Water models

We need to perform different simulation on the system which includes Solvation, minimization and equilibration (implicit and explicit) and then Steered molecular dynamics. System needs to be solvated to closely resemble the cellular environment. In order to do that we put our system in a water box using VMD tk console or this can be done manually. But, In our case we uses *solvate* package in VMD to solvate the system. It is very important to choose water box dimension wisely because we need a water box of size much more than our system and make sure that our system does not go outside during simulation. We use TIP3P water model which is default in NAMD. There are various other type of water models like SPC, TIP4P, TIP5P depending upon your choice. In general, 3-site models are widely used in many applications of molecular dynamics. TIP3P is a 3-site model which means it have three interaction points corresponding to the three atom of water molecule. Each atom has a point charge and the site correspond to the oxygen atom also has the lennard-Jones parameter. What model we are choosing also effect the computational cost for example if we are choosing 3-site water model that means 9 distances are required for each pair of water molecules. In case of 4-Site model 10 distances are required. So, the computation cost of a water simulation increases with number of interaction site in the water model. The potential for model such as TIP3P and TIP4P given by this equation:

$$E_{ab} = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{\text{OO}}^{12}} - \frac{B}{r_{\text{OO}}^{6}},$$

Where $k_c$ is the electrostatic constant has a value of 332.1 Å·kcal/(mol·e $^2$) in the unit commonly used in molecular modelling, $q_i$ and $q_j$ are the partial charges relative to charge of electron, $r_{ij}$ is the distance between two atoms or charged sites and A and B are the Lennard-Jones parameters.

The table shown below showing the difference between different types of water model with respect to their bond length and bond angle

| | TIPS | SPC | TIP3P | SPC/E |
|---|---|---|---|---|
| r(OH), $A^o$ | 0.9572 | 1.0 | 0.9572 | 1.0 |
| HOH, deg | 104.52 | 109.47 | 104.52 | 109.47 |
| A, $10^3$ Kcal $A^{o12}$/mol | 580.0 | 629.4 | 582.0 | 629.4 |
| B, kcal $A^{o6}$/mol | 525.0 | 625.5 | 595.0 | 625.5 |
| q(O) | -0.80 | -0.82 | -0.834 | -0.8476 |
| q(H) | 0.40 | 0.41 | 0.417 | 0.4238 |

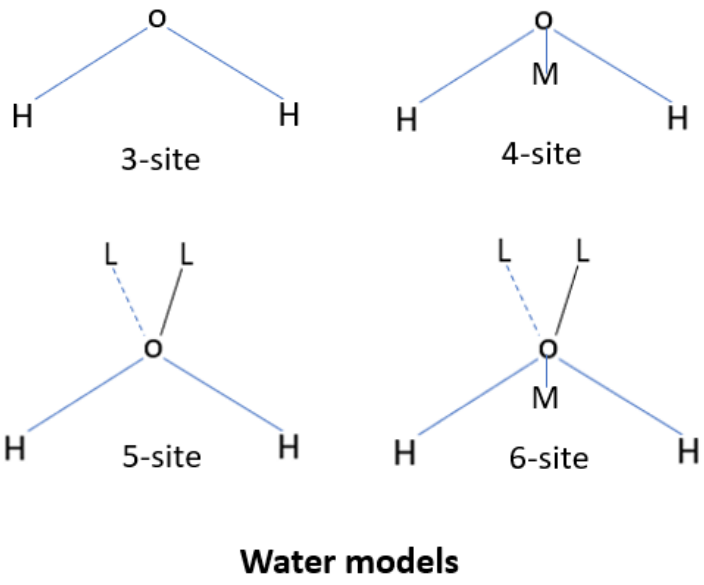



Figure 1.5 water models

These are the general shapes of water models but the exact geometrical parameters like OH distance and HOH angle vary depending on the model.

In addition to this all of these simulations were carried out in different solvation methods. These methods named as explicit solvation and implicit solvation. In explicit solvation force field calculation calculate all the interaction including all the water molecules but in implicit

solvation it treats water as continuum and does not consider water molecule interaction resulting in less simulation time.

### 1.9 Explicit Solvation

In the explicit case we use Periodic Boundary Conditions, In molecular dynamics we can simulate the system of N particles in isolation surrounded by vacuum but we are interested on the bulk properties of the system. For this we need to impose PBC on the system. We can also use rigid walls but the surface effect would blur the real bulk physics (ref guide). Because the fraction of atom near the walls is proportional to $N^{-1/3}$ which is negligible in a macroscopic a sample but if we considered the case in a typical simulation it can go from 0.06 for $N = 10^6$ atoms to 0.49 for $N = 10^3$. In the PBC the simulation box is surrounded by an infinite number of its replicas but the interesting parts is only the N atoms inside the main cell are considered explicitly [(8)]. As soon as the atoms leaves the cell, an image particle enters from the opposite side to replace it.

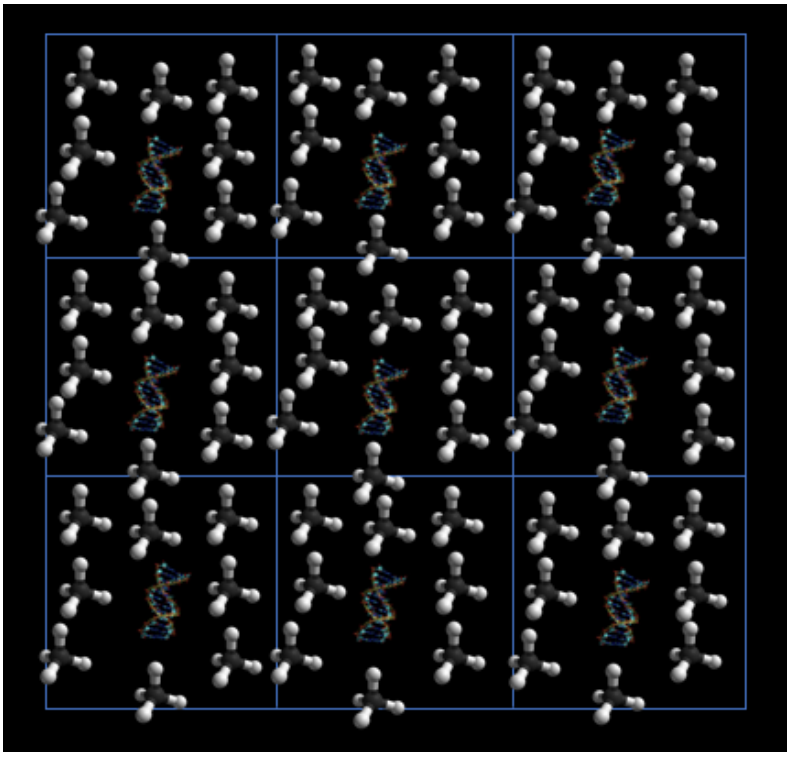


Figure 1.6 Periodic Boundary Condition

In this figure these are the replicas of the system to explain the periodic boundary conditions.

PBC can be used together with the minimum image convention that means only the interaction with the minimum image are considered [8]. But this is not the best solution because the time needed to compute the forces will scales as $N^2$. In case of large system where the interatomic potential has decayed to a negligible value for distances much smaller than l/2 where L is the length of simulation box then there is no need to compute all the interactions. Another reason is because the potential is not constant on the surface of a cube around a given particle. By employing a spherical truncation such that the only interaction between particle separated by a distance smaller than $R_c$ are taken into account. $R_c$ is the cutoff radius which we set in the NAMD configuration file.

Cutoff indicates the distance in $A^o$ beyond which electrostatic interactions are cut-off. If we applying a fast solvar like Particle Mesh Ewald or Multilevel Summation Method then the cutoff parameter will define the splitting distance for the 1/r interaction potential where short-range part is evaluated exactly between atoms within cutoff distance [8].

In addition to periodic boundary conditions there are other parameters which includes adjustable parameters, simulation parameters, forcefield parameters, Integrator parameters, Constant temperature and electrostatic with PME. Simulations parameters includes CHARMM force field (whether or not use force field) and temperature. Force field parameter includes 1-4 Scaling, cutoff, switching, switch distance and pair list distance. Integrator parameter includes timesteps, rigid bonds, non-bonded frequency and steps per cycle. Constant temperature includes langevin dynamics. Electrostatics with PME includes PME grid spacing [9]. Constant pressure controls the parameters like langevin piston, langevin piston temperature.

### 1.10 Implicit Solvation

On the other side, we also use implicit solvation to compare it with the explicit solvation. In case of implicit solvation, we use Generalized Born implicit solvation method. Implicit solvation is also called continuum solvation method because it considered solvent as a continuous medium instead of individual explicit solvent molecule [10]. Water have a property of dielectric. As a result, water screens electrostatic interactions between charged particles. So, water can be modeled as a dielectric continuum. From this we can calculate the electrostatic forces of a biological system using differential equation which can be solved for electric field caused by the collection of charges. In comparison with explicit solvent this method is less time consuming than explicit. This method is most often used in molecular dynamics and also in molecular mechanics. We can calculate free energy of solute-solvent interaction in various chemical and structural processes such as folding and unfolding of protein, DNA, RNA and polysaccharides. There are various types of implicit solvation which includes Accessible surface area-based method (ASA) and Generalized Born Implicit Solvation method (GBIS).

GBIS model is an approximation to the Poisson-Boltzmann equation because it is computationally expensive to calculate without approximation [11]. So, it models all the atoms in the system as a charge sphere whose internal dielectric is lower than the environment. The screening which each atom, i , experience depend on the local environment. If it is surrounded by more atoms then the less its electrostatic will be screened because it is more surrounded by the low dielectric. There are different Generalized Born (GB) model which calculates atomic

descreening differently. This descreening is used to calculate born radius of each atom. A large born radius means there is small screening (strong electric field) as if the atom were in vacuum. Similarly, A small born radius means large screening (weak electric field) as if atom were in bulk of water. This is the general equation of GBIS-

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

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}}$$

$$D = \left(\frac{r_{ij}}{2a_{ij}}\right)^2$$

$$a_{ij} = \sqrt{a_i a_j}$$

where, $\boldsymbol{\varepsilon_0}$ = Permittivity of free, $\boldsymbol{\varepsilon}$ = Dielectric constant, $\boldsymbol{q_i}$ = Electrostatic charge on particle *i* and *j*, $\boldsymbol{a_i}$ = Effective Born radius, $\boldsymbol{r_{ij}}$ = Distance between particle *i* and *j*.

The Generalized Born with accessible surface area model deal with the hydrophobic solvent accessible surface area term is GBSA [(11)]. GBSA is among the most commonly used implicit solvent combinations. This can be used in molecular mechanics termed as MM/GBSA.

If we talk about the accuracy of implicit solvent then it was found that implicit solvents are less accurate than explicit because of the elimination of the water molecule and represent water in an average manner. So, it is better to use caution when employing the implicit solvent for molecular dynamic research. In order to perform minimization and equilibration in implicit solvent we need NAMD configuration file with some additional parameters and changes for the implicit solvation and these parameters includes structure and coordinate, gbis, cutoff, alpha cutoff and ion concentration. All of these parameters are related to each other like alpha cutoff value should be less than cutoff as it determines the born radius of each atom. In case of cutoff value, this value should be higher than the PME simulations because there are no long-range interaction. Ion concentration must be set wisely because increase in ion concentration result in increasing electrostatic screening. we have to run simulation at different values of these parameters and check the stability of the system after minimization and Equilibration.

### 1.11 Steered Molecular Dynamics

There are various techniques which includes optical tweezers and AFM which can be used to study the unzipping, unwinding, untwisting and shearing of nucleic acids. These biological

processes involve the transition from one state to another. In order to go from one state to another they need to cross the energy barrier. These events are difficult to reproduce on molecular dynamics time scales which is only the order of tens of nanoseconds [(7)]. Single molecule micromanipulation techniques provide insight to understand these issues by applying external force on the system to unzip the system and understand the mechanical properties. But there are some limitations of these techniques as we discussed in the introduction that in some experiment like DNA melting bases with A-T rich sites open first then G-C rich sites due to temperature influence [(5)]. So, Unzipping of Double Stranded (dsDNA) to Single Stranded (ssDNA) is central to many cellular processes.

We use Steered Molecular Dynamics to unzip the DNA. Steered Molecular Dynamics is a simulation technique in which external force is applied on the system in a simulation to study their mechanical properties as well as the kinetics of the system. Steered molecular dynamic can be done in two ways one is constant force pulling and other one is constant velocity pulling.

In case of constant force pulling we applied constant force directly on the SMD atom which we select in the PDB file. But in constant velocity pulling, we attached SMD atom to a dummy atom through a spring and the velocity is exerted on the dummy atom. Force between SMD atom and dummy atom is calculated. We can vary the stiffness of the spring or the position of the restrain to pull the atom along. We can apply various types of force like torque, rotation on the system. It depends on how we want to work on the system. For examples we want to unzip the DNA followed by untwist. This can be done using COLVARS (collective variable). It can be done easily because NAMD provides facilities for applying different types of external force. But in our Case we just did only unzipping by using constant velocity Steered Molecular Dynamics.

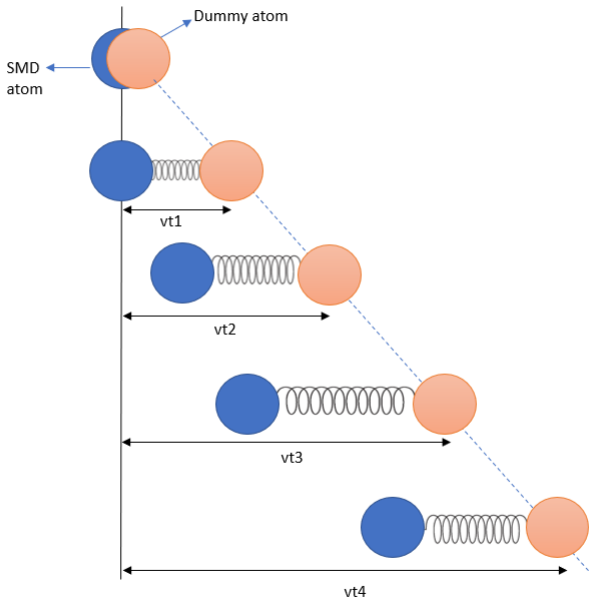



Figure 1.7 - SMD and Dummy Atom

The force between dummy atom and SMD atom calculate using the following equation:

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

$$U = \frac{1}{2} k[vt - \left(\vec{R} - \vec{R_0}\right) \cdot \vec{n}]^2$$

- U = Potential energy
- K= Spring constant
- v = pulling velocity
- t = time
- $\vec{R}$ = Actual position of the SMD atom
- $\vec{R_0}$= Initial position of the SMD atom
- $\vec{n}$ = Direction of pulling

A comparative study was done between implicit and explicit solvation which is discussed in the next chapters.

# Chapter 2

# Methodology

In order to unzip the DNA, we use Steered Molecular Dynamics (SMD) by applying external force on the SMD atom by fixing another atom using CHARMM36 force field in VMD and NAMD. But this includes many other simulations as we discussed in the introduction part. All of these steps are as follows:

- Generating a Protein Structural File (PSF) file
- System in water box (simulation with periodic boundary conditions)
- System in Generalized Born Implicit Solvent
- Simulation with Steered Molecular Dynamics

## 2.1 Generating a Protein Structural file (PSF) file

A PSF file can be generated through a PDB file using VMD autopsf module using psfgen plugin. This autopsf module uses CHARMM36 forcefield topology file to generate PSF file. The X-ray structure from protein data bank does not contain hydrogen atoms in the system So, The PDB and PSF file generated through psfgen contain guessed coordinates of hydrogen atom. Later, energy minimization was done which helps to make their position reasonable.

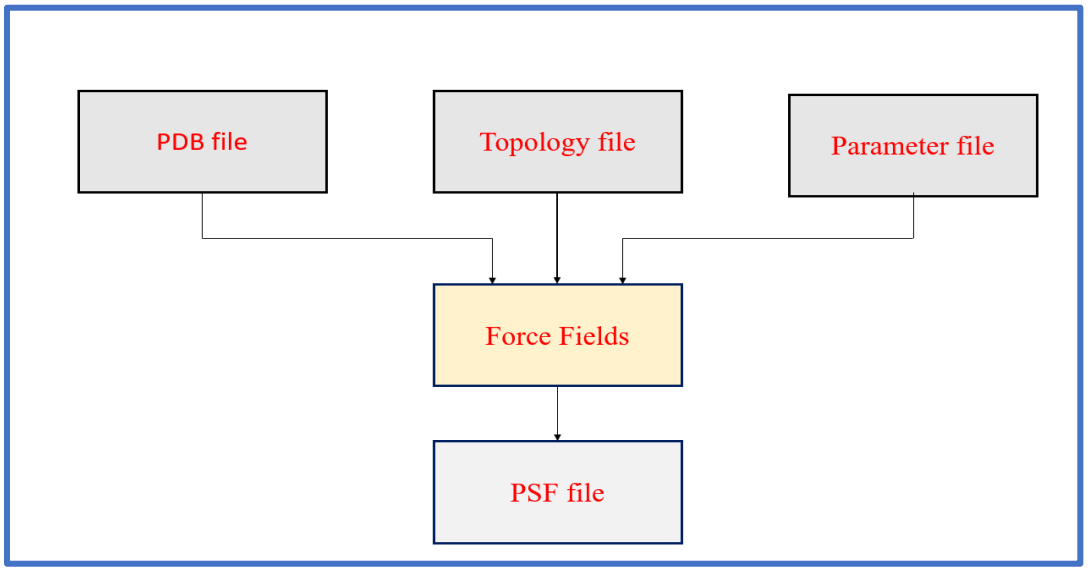



Figure 2.1 Generation of PSF file

## 2.2 System in water box (Simulation with Periodic Boundary Conditions)

We need to minimize and equilibrate our DNA-water system so the system become stable to perform the further simulations. To do this we need a NAMD configuration file which have the information about how we will minimize and equilibrate the system. Generally, configuration file includes information about the temperature control, pressure control and the periodic boundary conditions. We use TIP3P water model to solvate the system. In case of DNA hairpin, we added ions to the solvation box. We added 0.2 M/l of NaCl ions in the solvation box. The size of the water box was large enough so that DNA will not go outside during unzipping.

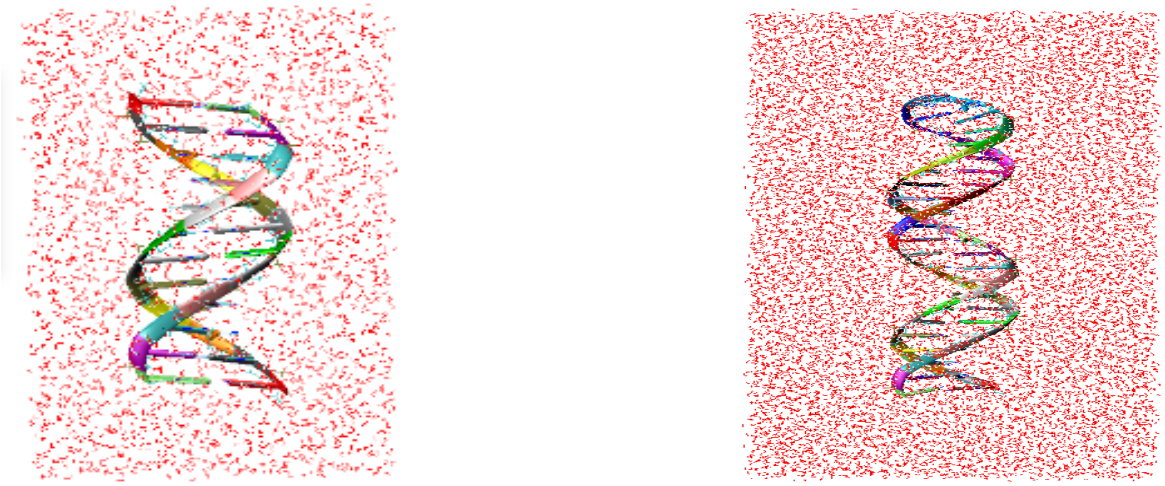


Figure 1.6 B-DNA in water box (a) Dodecamer B-DNA (b) DNA hairpin

In case of B-DNA dodecamer with sequence 5'-CGCGAATTCGCG-3' the number atoms were 9128. The minimization was done for 8000 steps after that system was heated for 10ps. Equilibration was done for 20ps. All of these simulations were done in the NPT ensemble for the temperature 300K.

In case of DNA hairpin loop with sequence 5'-GCGAGCCATAATCTCAATCTG GAAA CAGATGAGATTATGGCTCGC-3' the number atoms were 157480. The minimization was done for 20000 steps after that system was heated for 100ps. Equilibration was done for 200ps. All of these simulations were done in the NVT ensemble for the temperature 300K.

## 2.3 System in Generalized Born Implicit Solvent

We need a different configuration file which includes additional information about implicit solvent techniques. In the configuration file we set gbis parameters like alpha cutoff, ion concentration, solvent dielectric, intrinsic radius offset, GBIS delta, GBIS beta, GBIS Gamma,

SASA and surface tension. A different constrain atom and fixed atom file was made through PDB and PSF.

In case of B-DNA dodecamer with sequence 5'-CGCGAATTCGCG-3' the number atoms were 768. The minimization was done for 8000 steps after that system was heated for 10ps. Equilibration was done for 20ps. All of these simulations were done in the NPT ensemble for the temperature 300K.

In case of DNA hairpin loop with sequence 5'-GCGAGCCATAATCTCAATCTG GAAA CAGATGAGATTATGGCTCGC-3' the number atoms were 1398. The minimization was done for 8000 steps after that system was heated for 10ps. Equilibration was done for 20ps. All of these simulations were done in the NPT ensemble for the temperature 300K.

In case of DNA without hairpin loop with sequence 5'-GCGAGCCATAATCTCAATCTG-3' the number atoms were 1268. The minimization was done for 8000 steps after that system was heated for 10ps. Equilibration was done for 20ps. All of these simulations were done in the NPT ensemble for the temperature 300K.

Minimization and equilibration contrast from one another by the nature in which they apply the molecular dynamics forcefield. Energy minimization involves searching the energy landscape for local minimum i.e., place in which the molecule is relaxed, by methodically fluctuating the position of atom and computing the energy. Equilibration includes molecular dynamics where Newton Second Law is illuminated for each atom in the system and calculate its trajectory. Accomplishment of equilibrium judged by how well velocities, pressure and so on are disseminated in the framework over a given measure of time.

These simulations will generate output files which we will use for the further simulations.

**2.4 Simulations with Steered Molecular Dynamics:**

We need a different configuration file to perform SMD simulations, these configuration files include information about the SMD atom, Fixed atoms, spring constant and the pulling velocities. In case of B-DNA we fixed phosphorus atom of residue 2, and for SMD atom we set phosphorus atom of residue 12. This was done by the tkconsole in the VMD by changing beta column and occupancy column values. So, the NAMD will be able to recognize fixed atom and SMD atom. In case of DNA with hairpin and DNA without hairpin set phosphorus atom as SMD atom and fixed atom (Implicit). A pulling direction was set to unzip the DNA.

In case of B-DNA dodecamer with sequence 5'-CGCGAATTCGCG-3' SMD was done for 100ps with pulling velocity 0.001, 0.002 and 0.003 in $A^0$/timestep by varying spring constant K = 1,2,3,4 and 5 in (*kcal/mol/Å²* ).

In case of DNA hairpin loop with sequence 5'-GCGAGCCATAATCTCAATCTG GAAA CAGATGAGATTATGGCTCGC-3' SMD was done for 12ns for implicit and 45ns for explicit with pulling velocity 0.0001 $A^0$/timestep (Implicit) and 0.000011 $A^0$/timestep (Explicit). The spring constant K is equal to 0.2 *kcal/mol/Å²*

In case of DNA without hairpin loop with sequence 5'-GCGAGCCATAATCTCAATCTG-3' SMD was done for 12ns with pulling velocity 0.0001 $A^0$/timestep (Implicit) and spring constant K is 0.2 *kcal/mol/Å²*

These simulations will create a log file and other output files of the simulation which includes all the information about velocities, force and energies. We will use these files in the analysis and conclusion part to visualize the trajectory of the DNA unzipping.

# Chapter 3

# Effect of Implicit Solvent on B-DNA dodecamer unzipping

In this chapter we will analyze bond separation distance vs time step in the first part then we will analyze the force vs time in the second part. In addition to this, A comparative study was done between the effect of implicit solvent and explicit solvent.

### 3.1 Bond separation vs Time

In the VMD, we loaded dcd files and psf files which was generated after the SMD simulations. To obtain the bond seperation vs time dat file, we select SMD and Fixed atom in the visualize structure. A dat file was generated using VMD plot module. Gnupot was used to plot all the graphs.

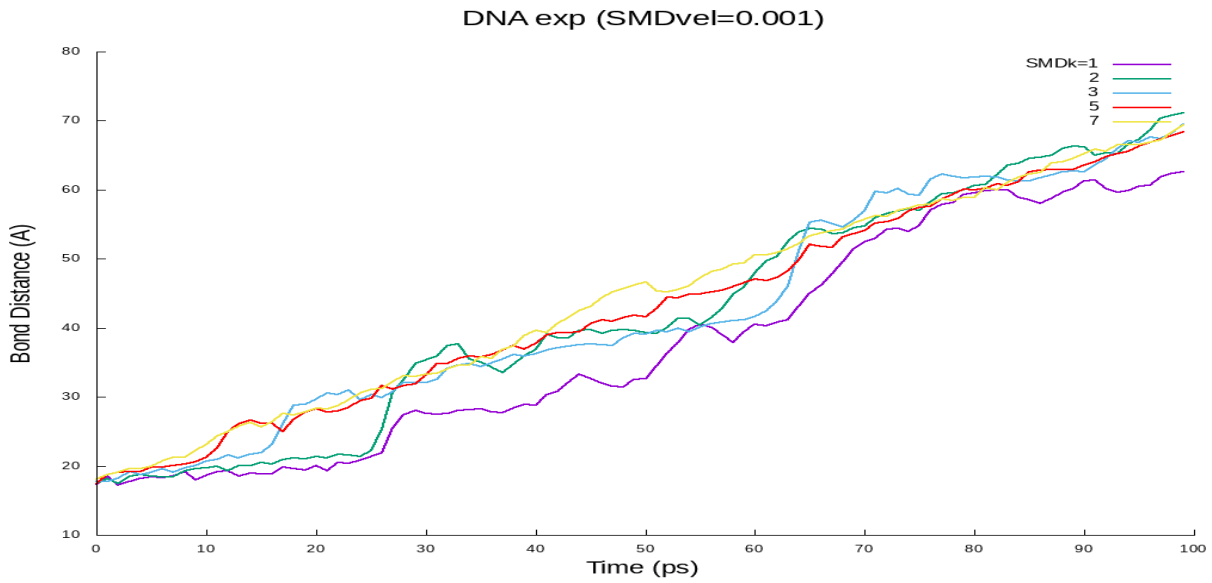



Figure 3.1 -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.001 on the bond separation distance

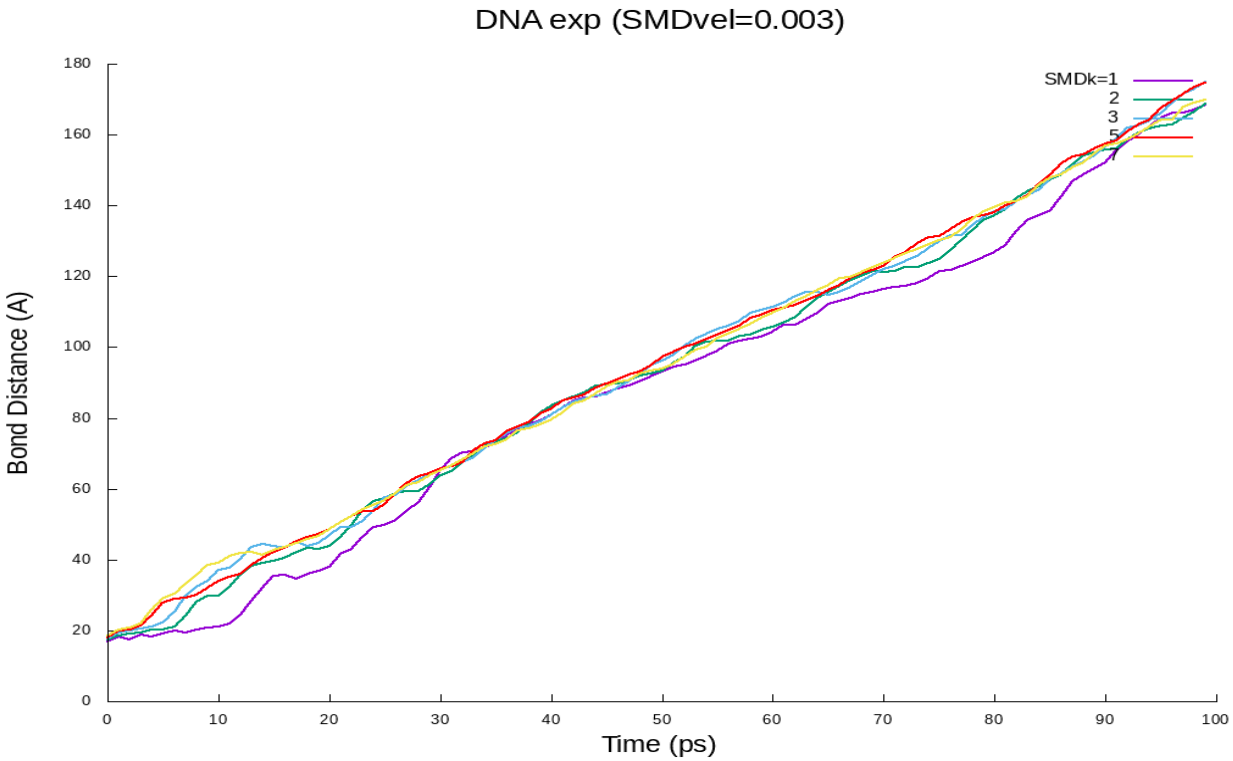



Figure 3.2 -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.003 on the bond separation distance.

- In figure (3.1) effect of different spring constant was studied with low pulling velocity and it was found that at low spring constant there are long pauses than high spring constant. In case of k =7 the graph is uniform with respect to time but it takes around 28ps in case of K=1 to rupture the first base pair of B-DNA. On increasing the value of K, uniformity increases in the graph.
- In figure (3.2), As we increase the pulling velocity to 0.003, it was found that graph is more sharp than figure (a) with less pauses and jumps. Effect of spring constant was observed only for the K =1.
- On comparing the Bond separation distance, it was observed that at low velocity pulling DNA unzipping (Bond distance) is less than the high velocity pulling. In figure (a) bond separation distance reaches only to 70A$^{o}$ but In figure (b) it reaches to 180A$^{o}$ .

**Bond separation Vs Time (Implicit)**

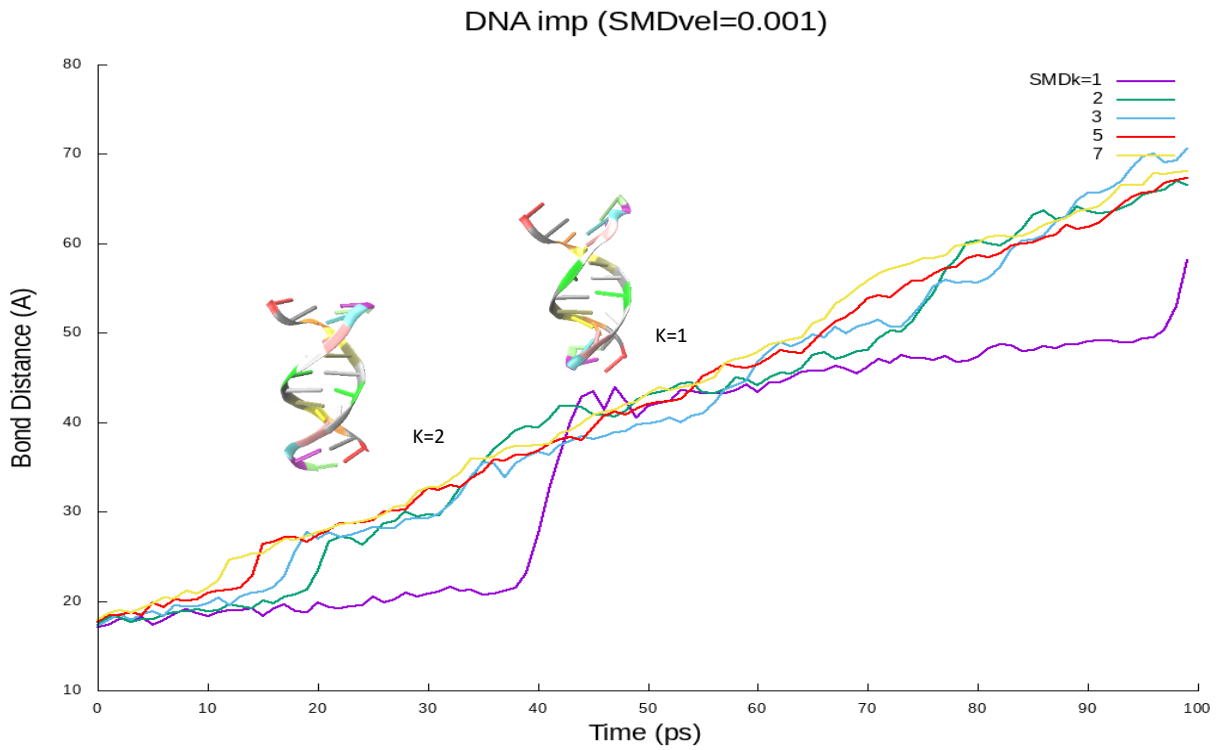



Figure (3.3) -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.001 on the bond separation distance.

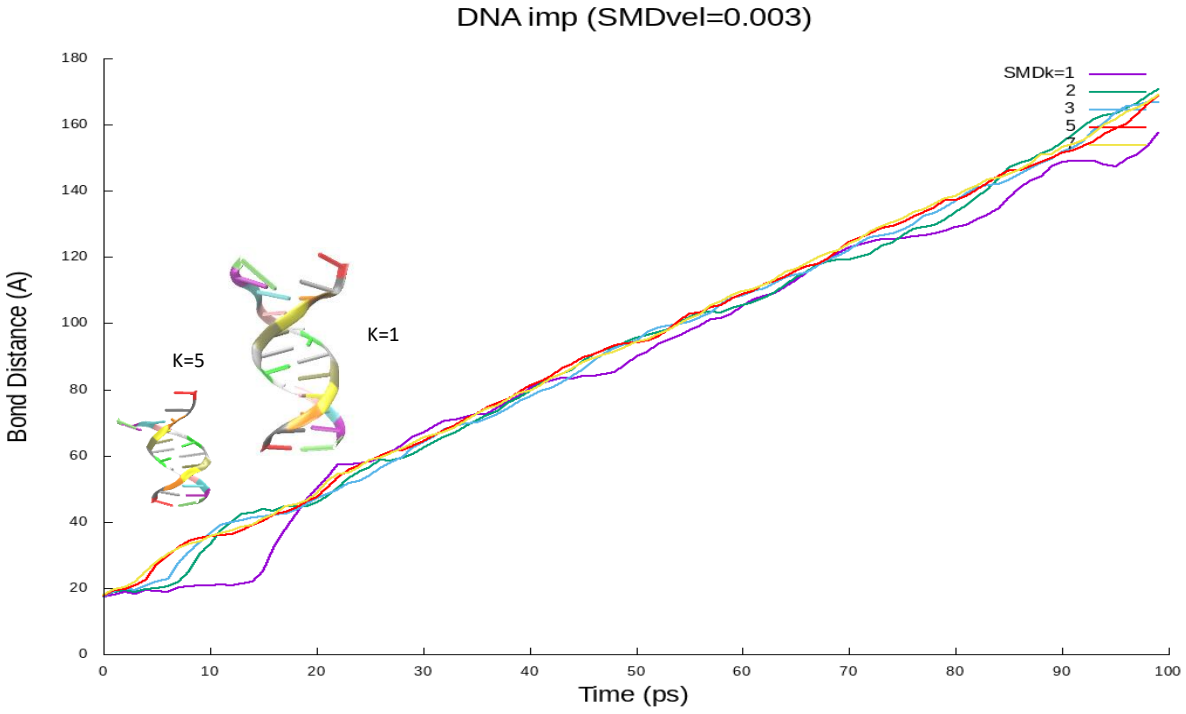



Figure (3.4) -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.002 on the bond separation distance

- In figure (3.3) also shows the same trend of pauses and jumps in the graph but the effect of spring constant shows a little variation in this case (implicit). It took around 40ps to rupture the first base pair but after that there is a constant increase in the Bond separation distance. Because of the more than one rupture of base pairs.
- In figure (3.4), As we increase the pulling velocity to 0.003, it was found that graph is more sharp than figure (a) with less pauses and jumps. Effect of spring constant was observed only for the K =1.
- Implicit shows the same results as explicit. We can consider implicit as a solvation method with less time consuming.

### 3.2 Force Vs Time (Explicit)

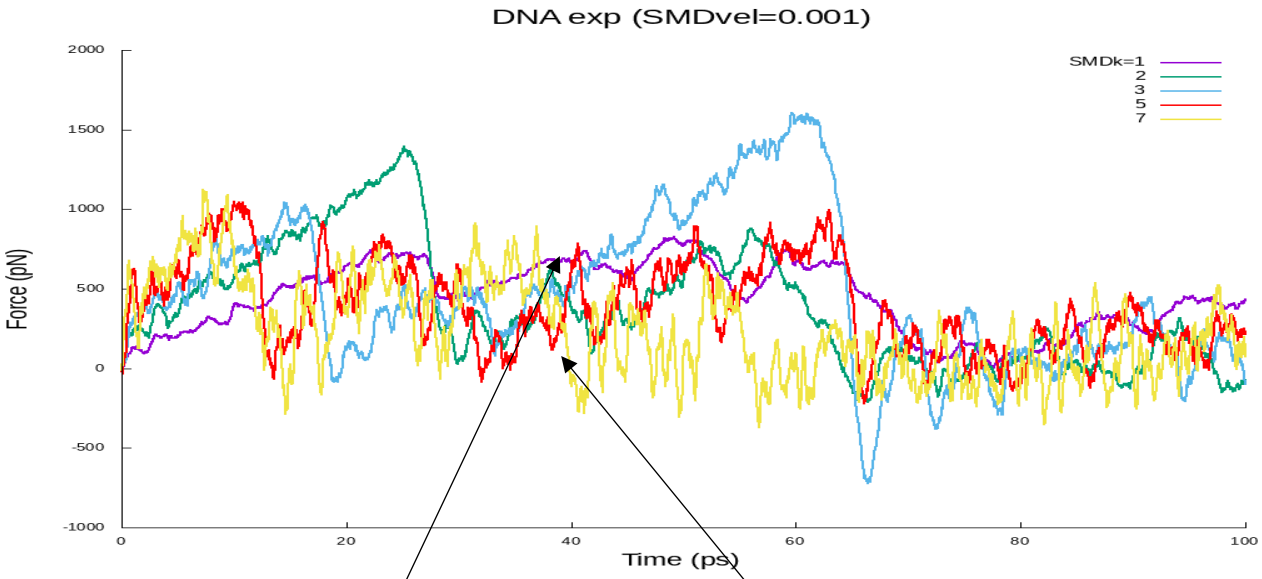
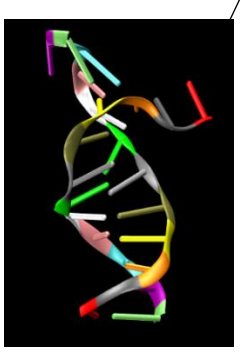



Figure (3.5) -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.001 on the force.



This figure shows the conformation of B-DNA at 40ps for K =1

This Figure shows the conformation of B-DNA at 40ps for K =7

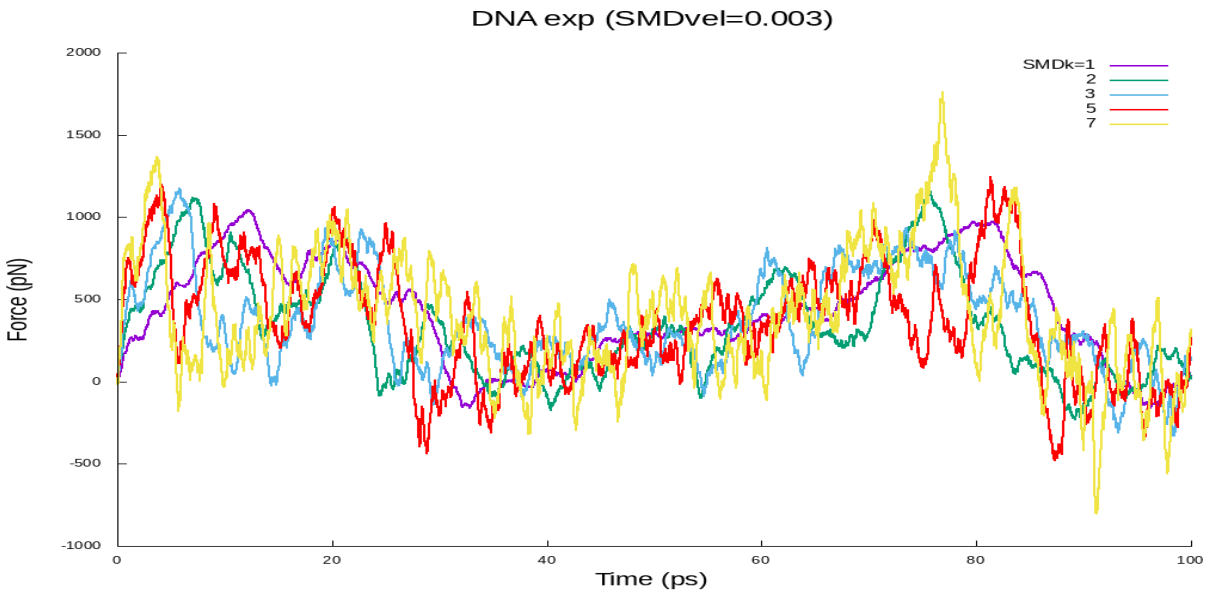



Figure (3.6) -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.003 on the force.

- From the graph, we can see the effect of the different spring constant. For example, in case of K =1 it took around 25ps to initiate the rupture but in the case of K = 7 it took less than 10ps to initiate.
- In figure (3.5) we can see from the conformations at 40ps, at low K value unzipping is less than the high K value.
- In figure (3.6) After t >70ps, force reaches to 1700pN which is may be because of DNA become completely unzipped.

**(Implicit)**

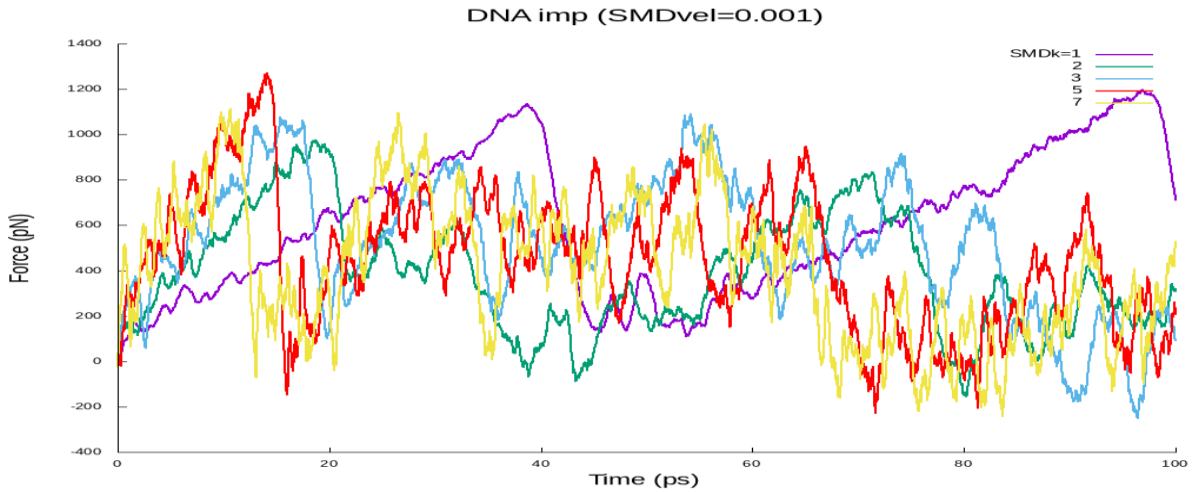



Figure (3.7) -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.001 on the force.

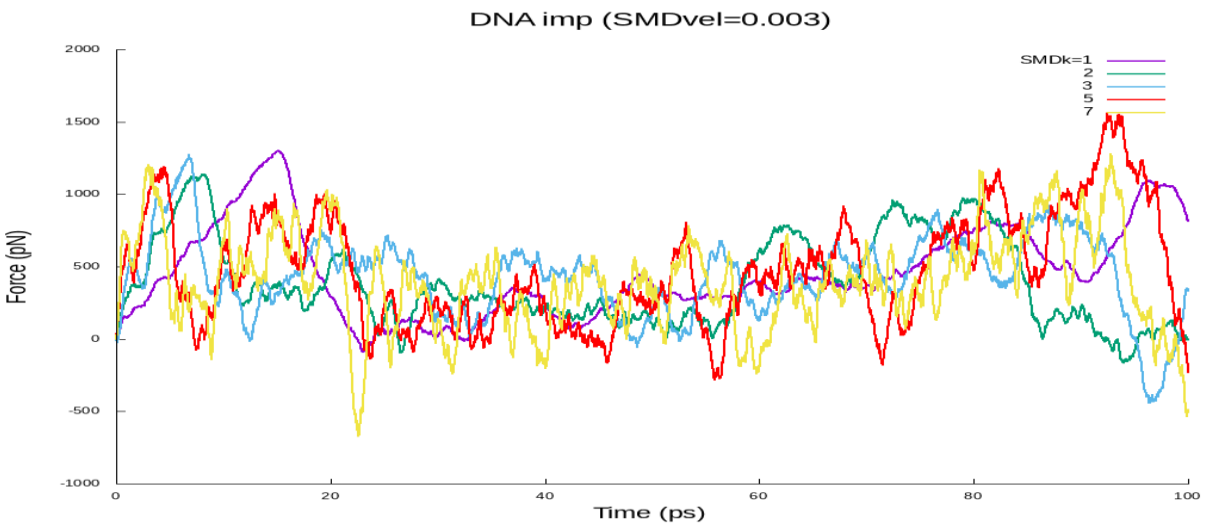



Figure (3.8) -This graph shows the dependence of different spring constant which is K = 1,2,3,5,7 with velocity 0.003 on the force.

- All the simulations on different spring constant require around 1200pN force to initiate the unzipping but they have different timescales. For example, in case of K=1 it requires around 40ps.
- It requires around 1200pN force to initiate the rupture in both the cases (figure 3.7 and figure 3.8).
- On comparing implicit and explicit, both are showing the same trend in graphs except for the value of rupture force which is different in explicit and implicit. In case of Implicit rupture force is high than the explicit. This may be because of the interaction of water molecule with the system make it smooth to unzip than the case of implicit.

It was found that, effect of implicit solvent on B-DNA dodecamer is same as the explicit solvent. Trend in unzipping was only affected by the change in spring constant values and the pulling values.

## DNA Vs RNA

A comparison was done between B-DNA and A-RNA, the data for the RNA was collected from some Other source (Naman Kumar Bharti thesis work). It was found that DNA requires less force than the RNA. In case of B-DNA unzipping starts at 40ps for K =1 with pulling velocity 0.001 but in case of A-RNA unzipping starts at 20ps. In case of A-RNA force require to unzip it was around 600pN which is very low than the DNA (1400pN) for K=5 and pulling velocity 0.002 $A^0$/timestep.

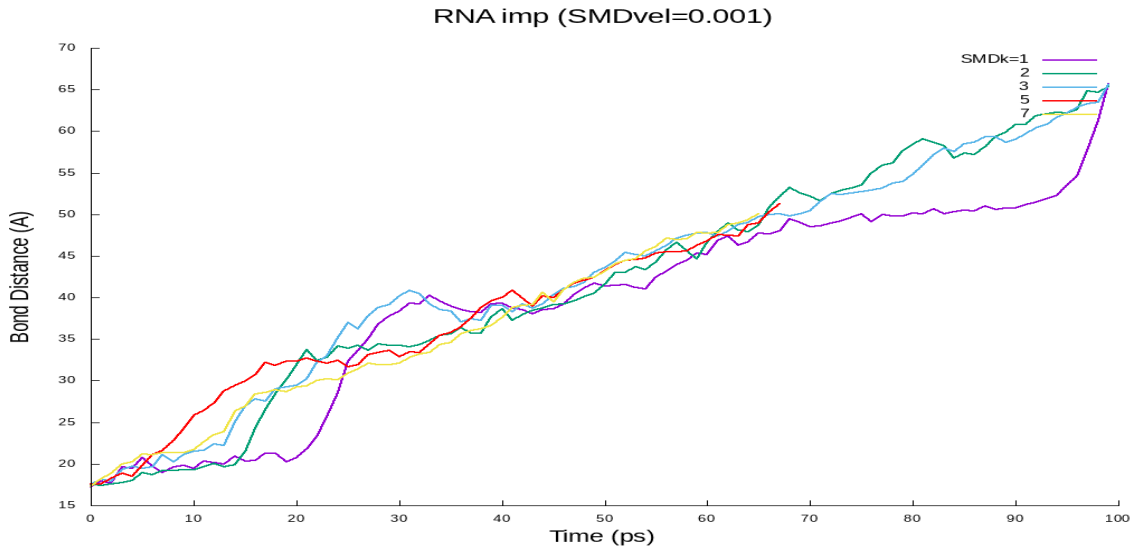



Figure 3.9 A-RNA with velocity 0.001 for different spring constant

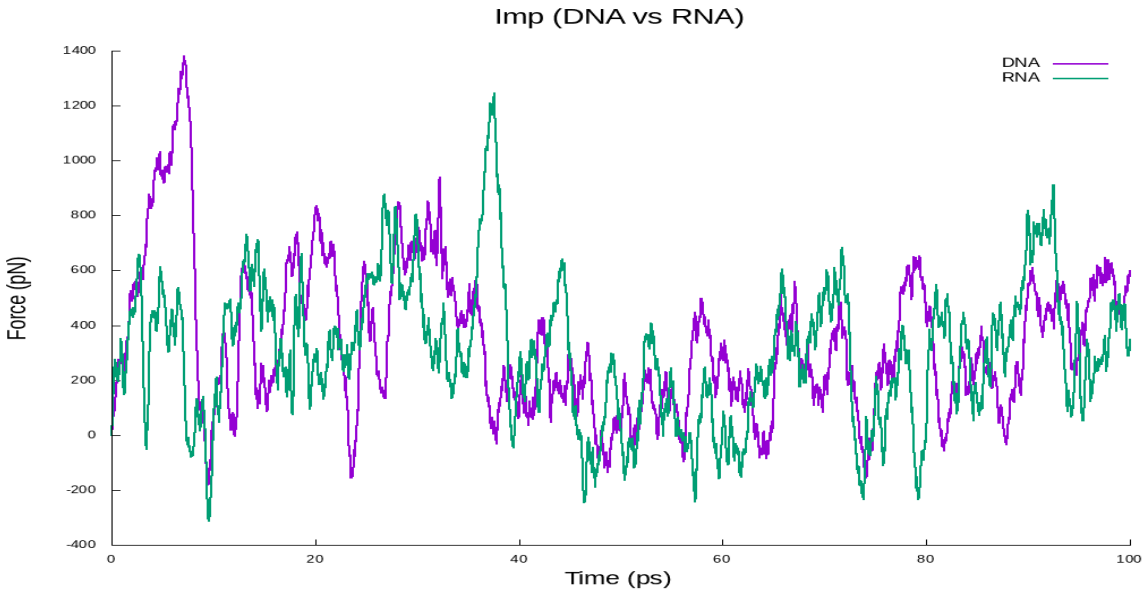



Figure (3.10) – Comparison of DNA vs RNA (implicit) for pulling velocity 0.002 and K = 5

# Chapter 4

# Effect of Implicit solvent on DNA hairpin and DNA without hairpin

### 4.1 Bond and Force analysis (Implicit)

(A DNA hairpin with sequence 5'-GCGAGCCATAATCTCAATCTG GAAA CAGATGAGATTATGGCTCGC-3') and (DNA without hairpin with sequence 5'-GCGAGCCATAATCTCAATCTG-3')

We performed simulation for both implicit and explicit solvation. In case of implicit solvation simulation were performed for 6ns and 12ns at 300K temperature. In case of explicit solvation simulation was performed for 45ns for 300K for DNA with hairpin. We did multiple runs for the simulation.

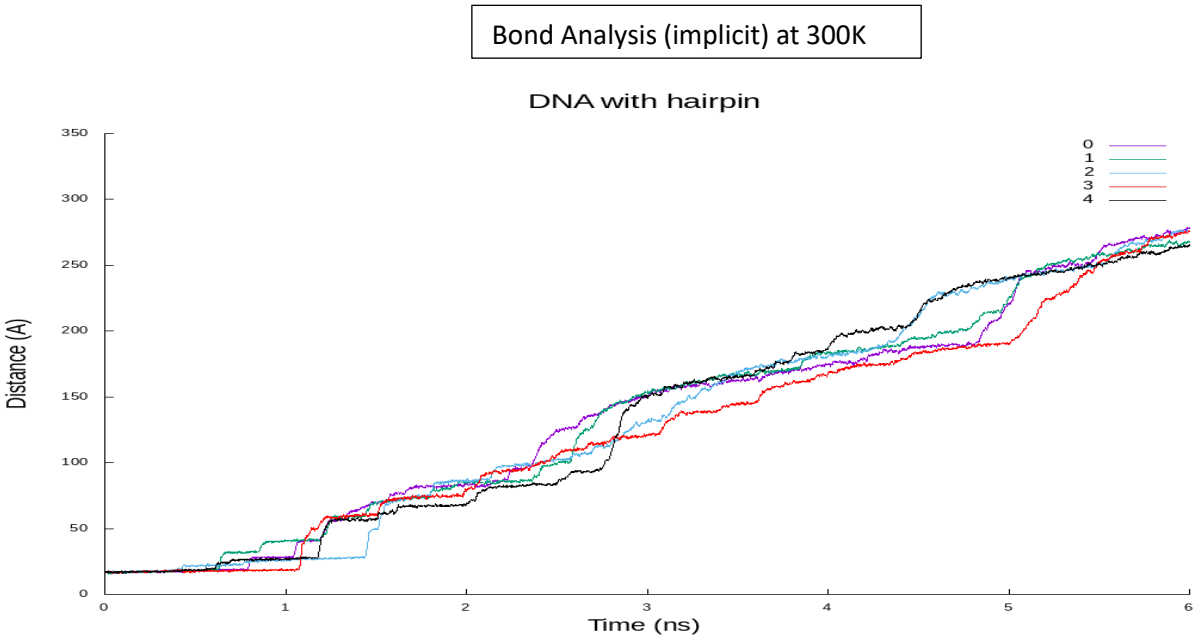



Figure 4.1 DNA with hairpin bond separation vs time for multiple runs at 6ns

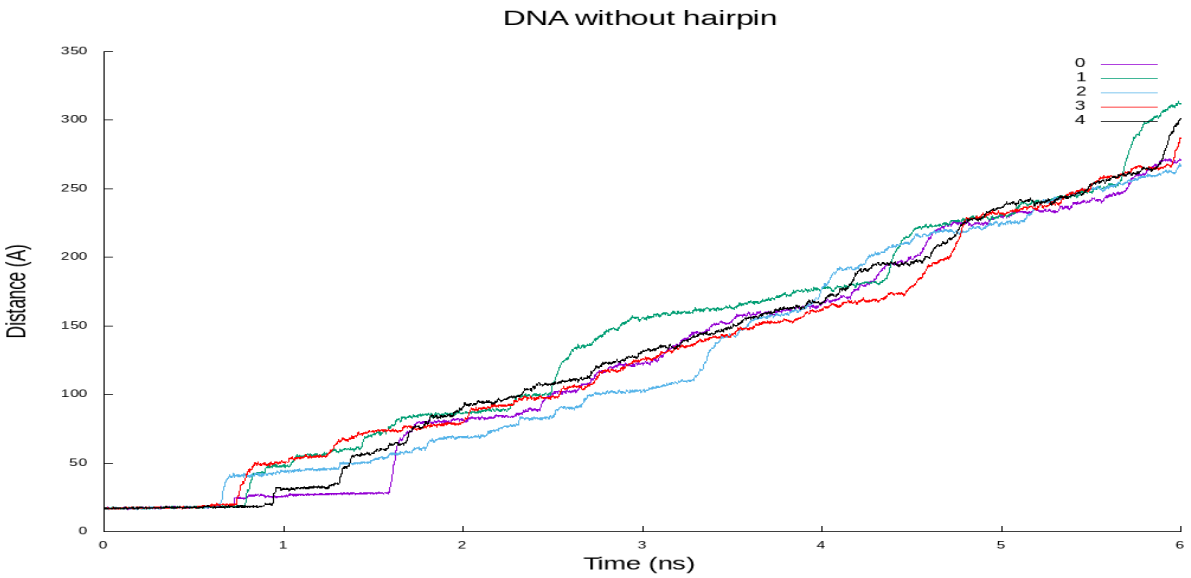



Figure 4.2 DNA without hairpin bond separation vs time for multiple runs at 6ns

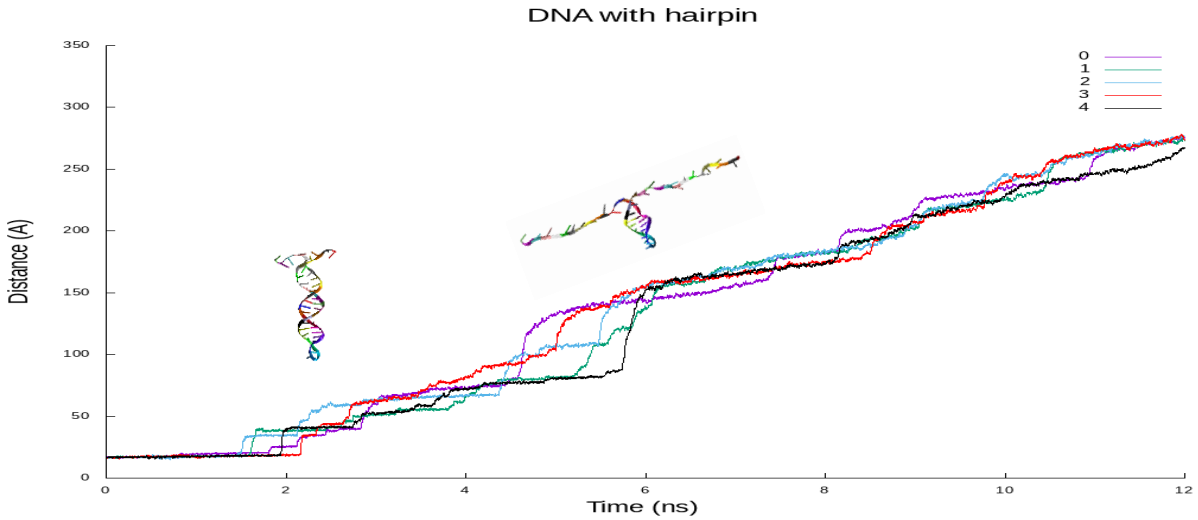



Figure 4.3 DNA with hairpin Bond separation vs time for different runs at 12 ns

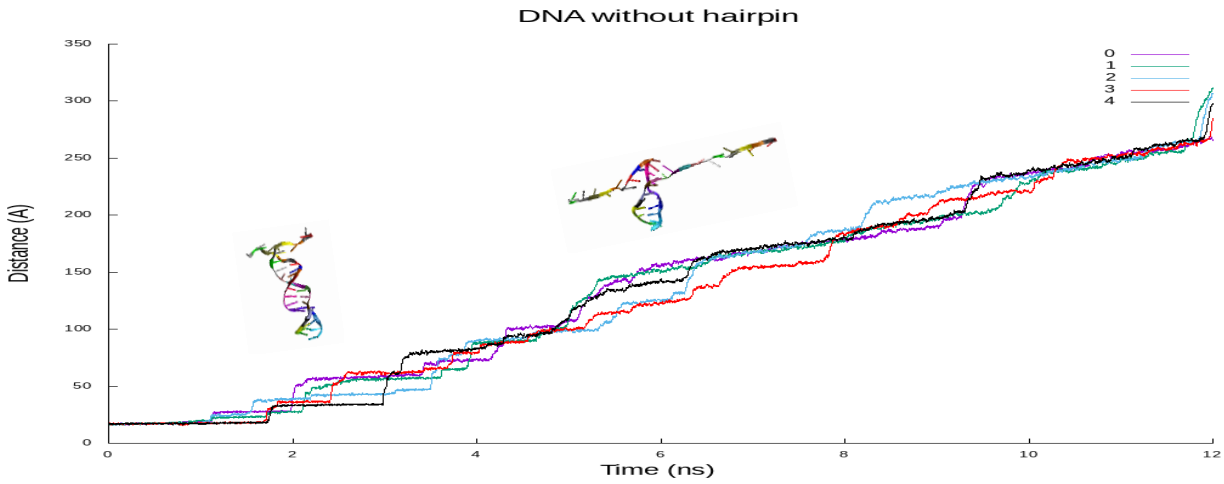



Figure 4.4 DNA without hairpin Bond separation vs time for different runs at 12ns

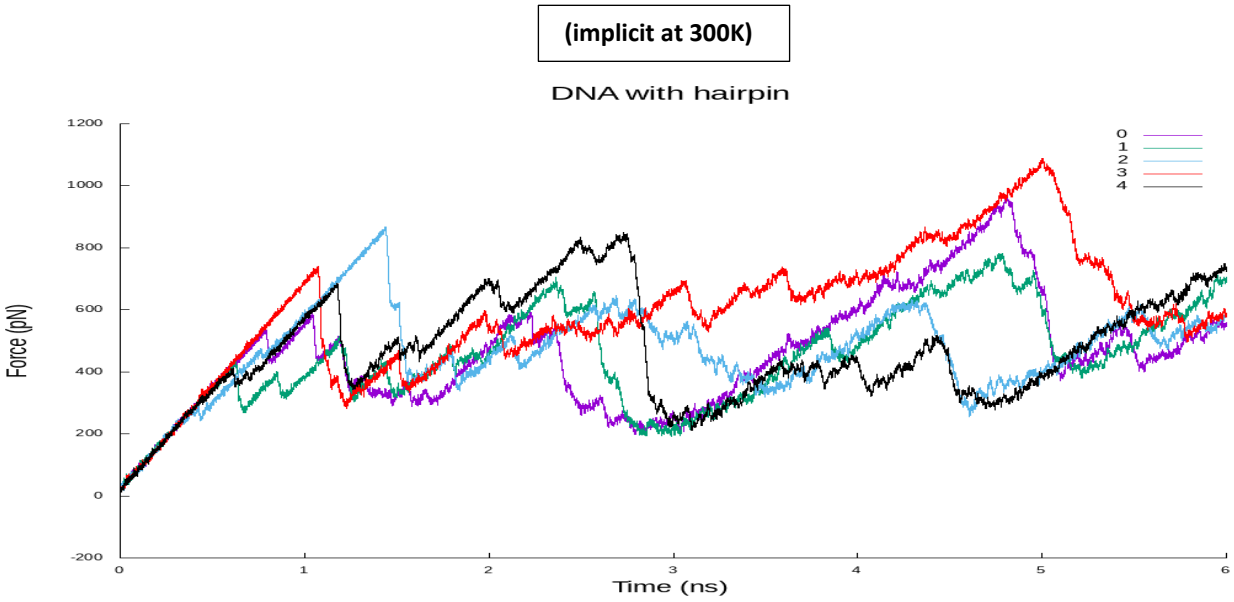



Figure 4.5 DNA with hairpin force vs time for different runs at 6ns

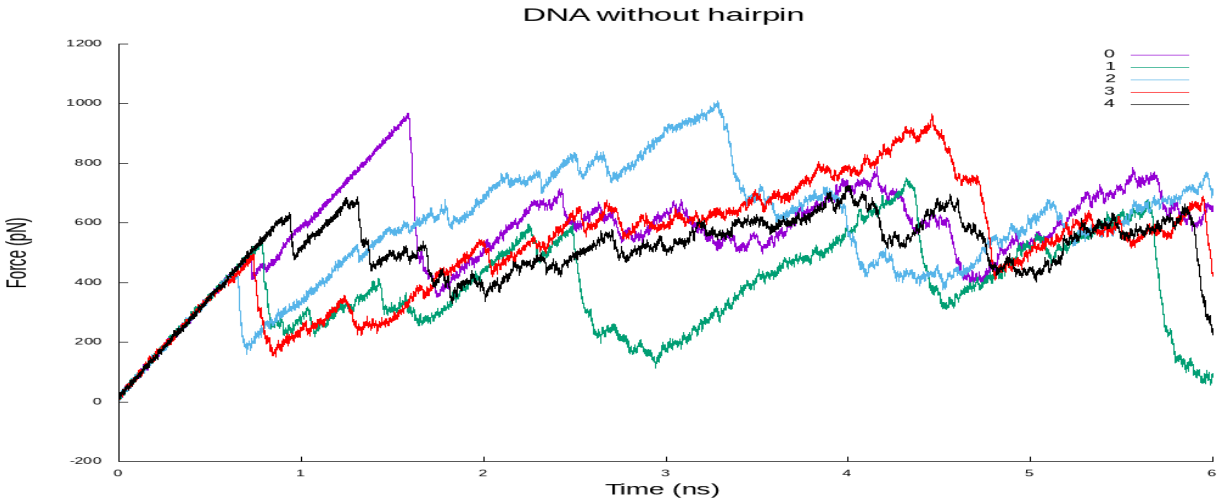



Figure 4.6 DNA without hairpin force vs time for different runs at 6ns

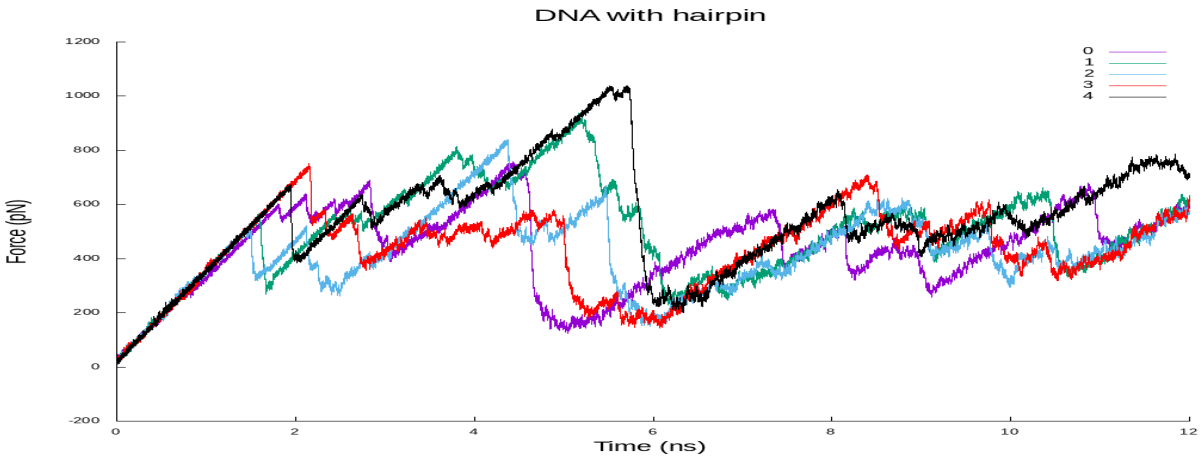



Figure 4.7 DNA with hairpin force vs time for different runs at 12ns

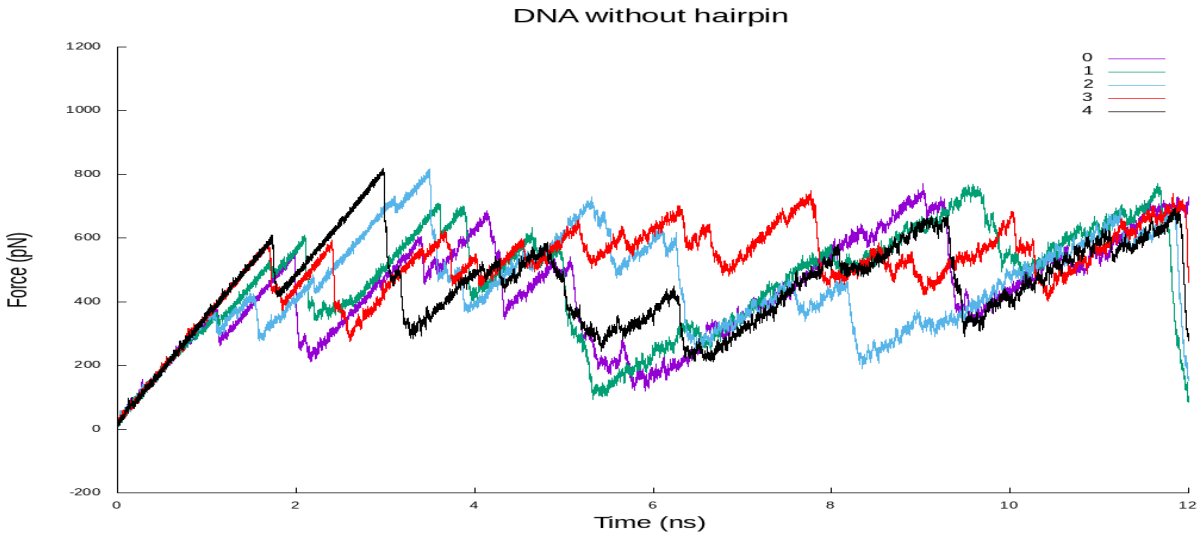



Figure 4.8 DNA without hairpin force vs time for different runs at 12 ns

## 4.2 Bond and Force analysis (Explicit)

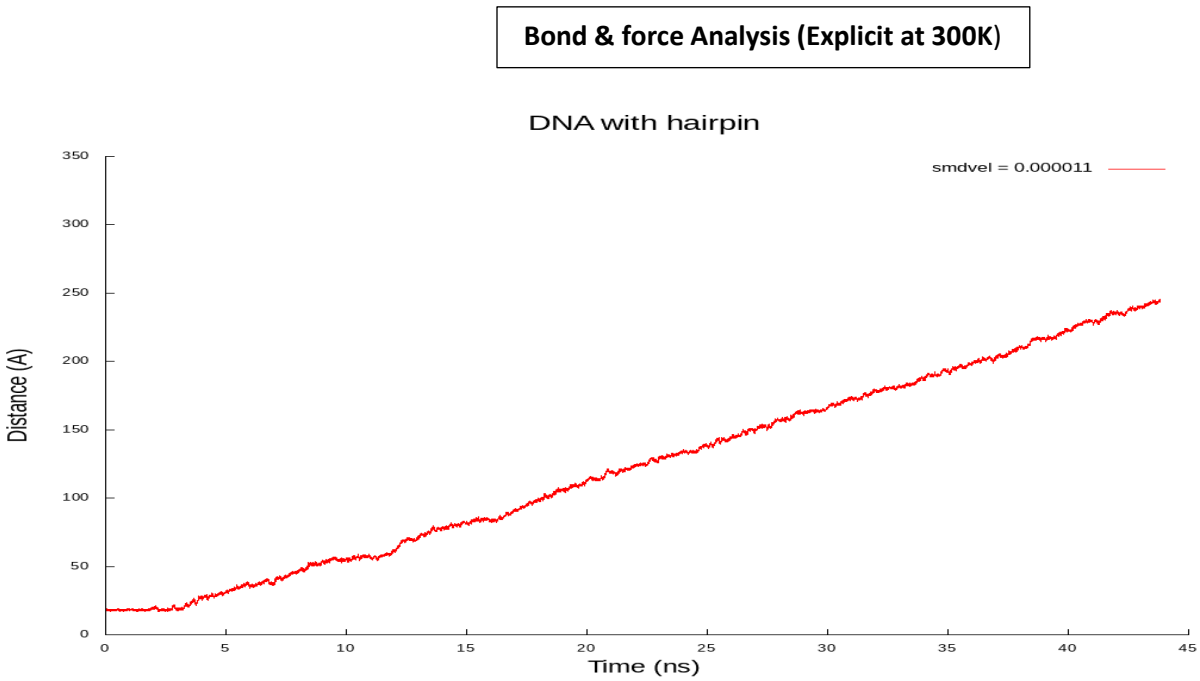



Figure 4.9 DNA with hairpin (Bond analysis) Explicit

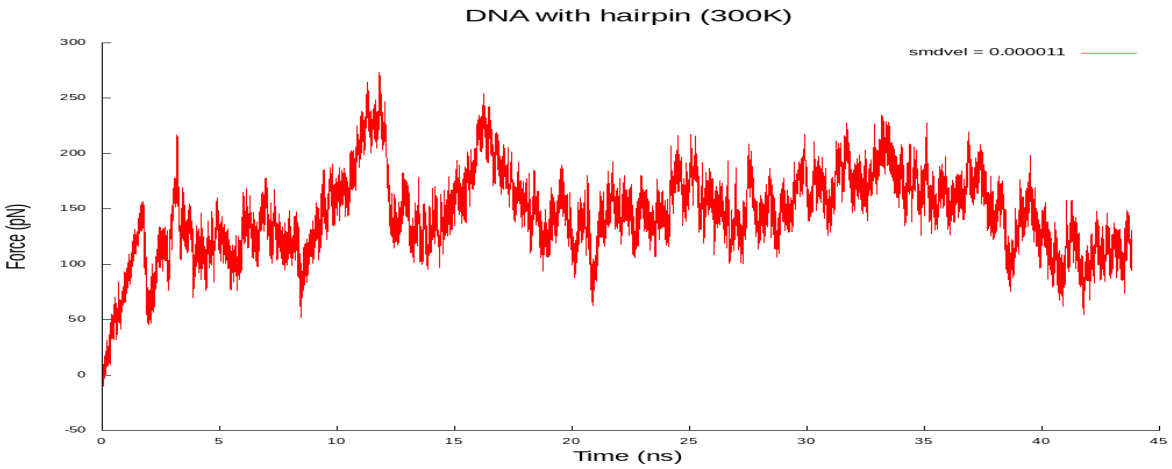



Figure 4.10 DNA with hairpin force analysis in case of explicit

- We did not find any difference between DNA with hairpin and DNA without hairpin in terms of Bond separation and Force. Graphs are looking similar. But under constant velocity pulling the unzipping of DNA is followed by pauses and jumps in the spectra which validate SMD qualitatively and quantitively.
- Implicit shows better results than explicit, this contradicts the fact that explicit is much better than implicit because of water interactions.
- It was observed that force required to unzip the DNA is less in case of explicit ~150pN but in case of implicit value goes around ~700pN.

**4.3 High temperature vs low temperature:** *A comparative study was done at higher temperature for both the case. The data for simulation at 320K was collected from (Naman Kumar Bharti thesis work).*

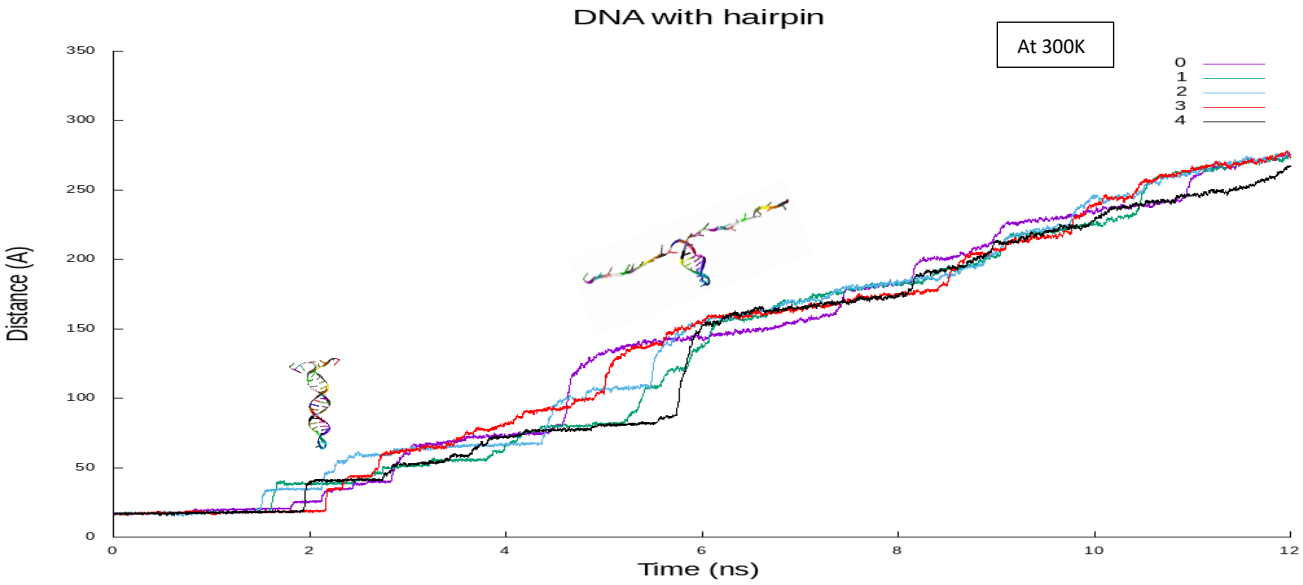



Figure 4.11 DNA hairpin at 300K (Bond separation vs time)

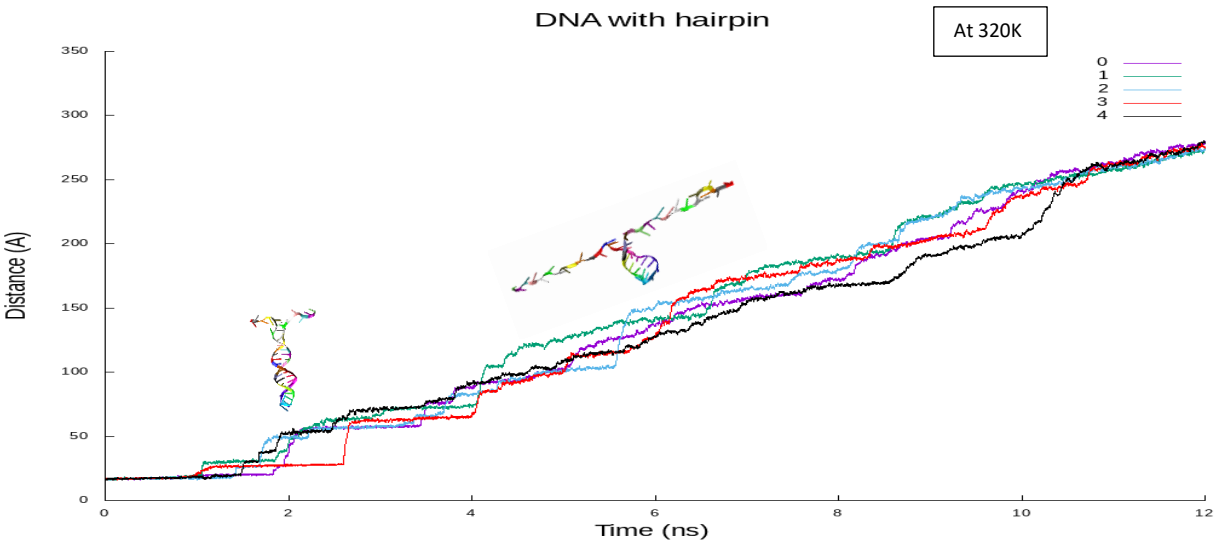



Figure 4.12 DNA with hairpin at 320K (Bond separation vs time)

- On comparing DNA with hairpin at 300K and 320K it was found that, on increasing temperature DNA become less stable. As we can see, In case of low temperature it starts unzip around ~2ps but in case of high temperature it started unzip around ~1ps.

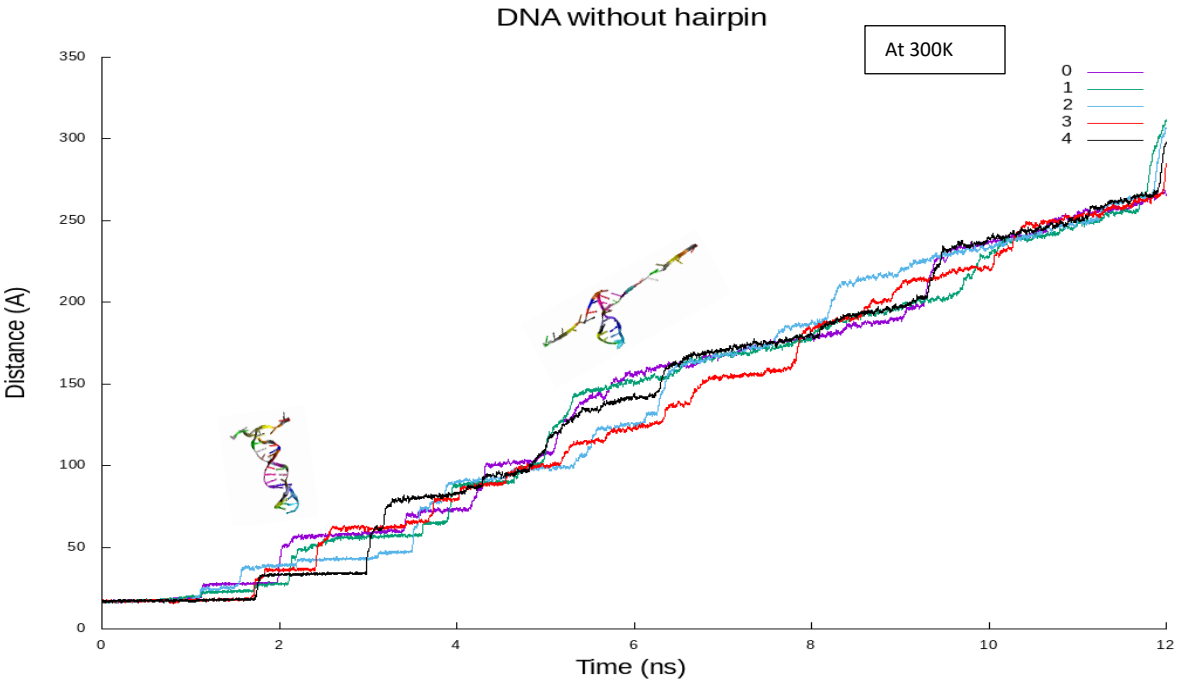



Figure 4.13 DNA without hairpin at 300 K (Bond analysis)

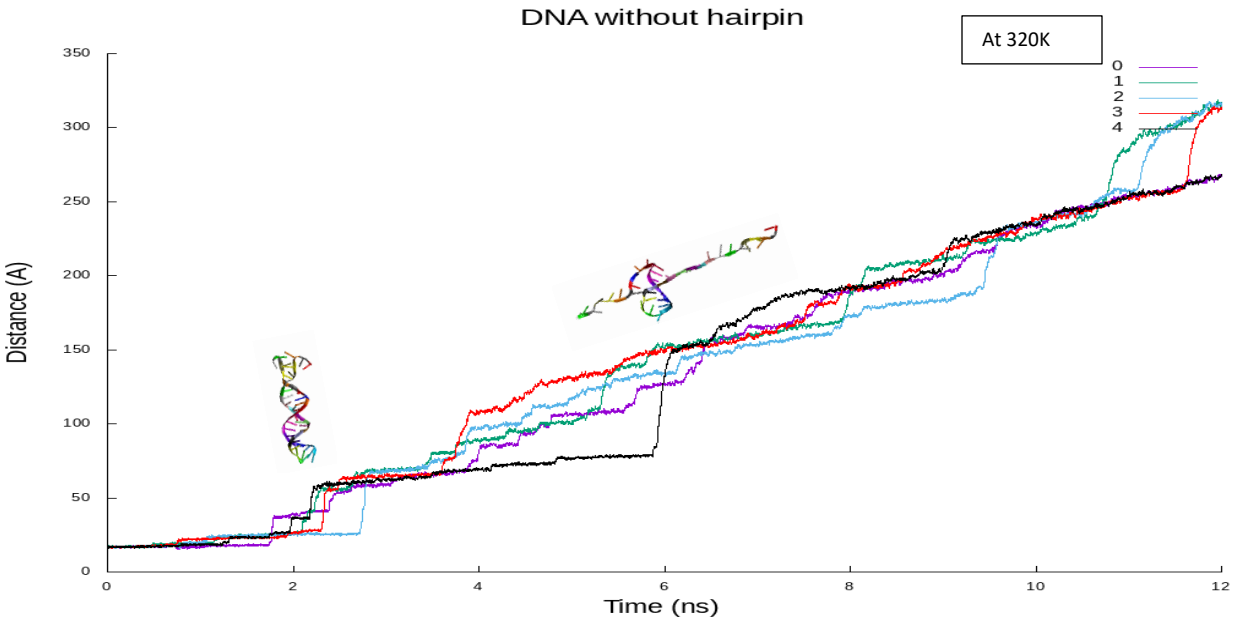



Figure 4.14 DNA without hairpin at 320 K (Bond analysis)

- Same effect was observed in this case (DNA without hairpin), As we can see from the conformation at different time scales that, At high temperature DNA unzipping more than low temperature.

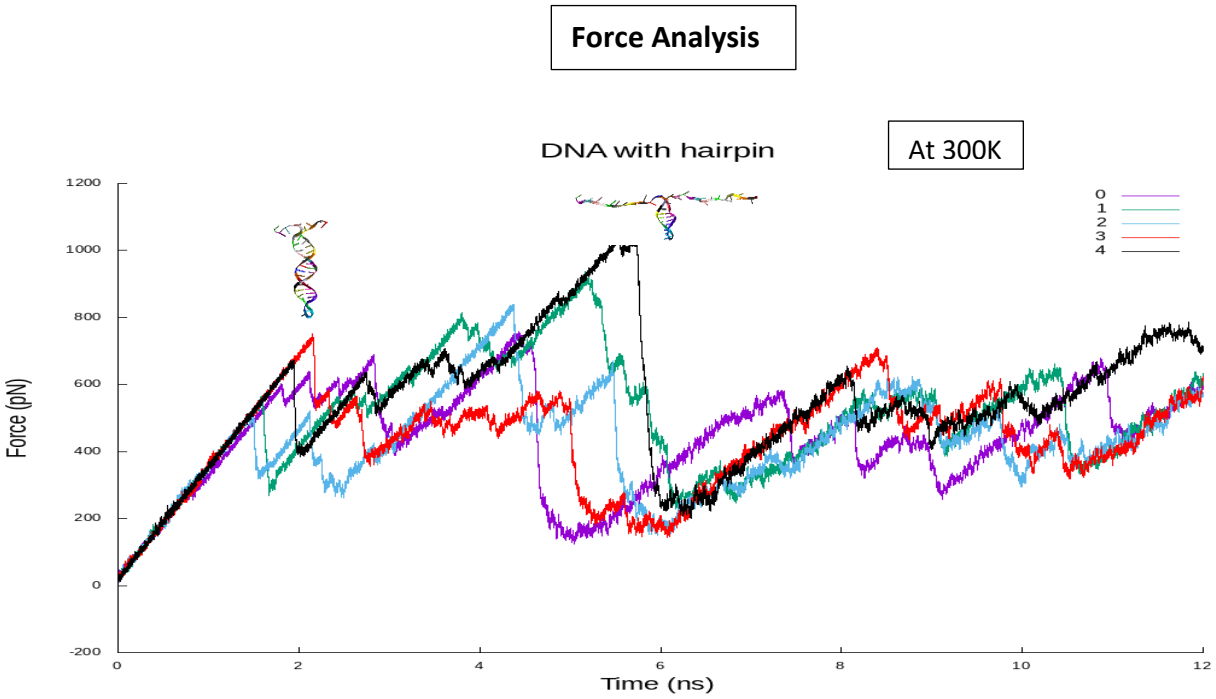



Figure 4.15 DNA with hairpin at 300K (force analysis)

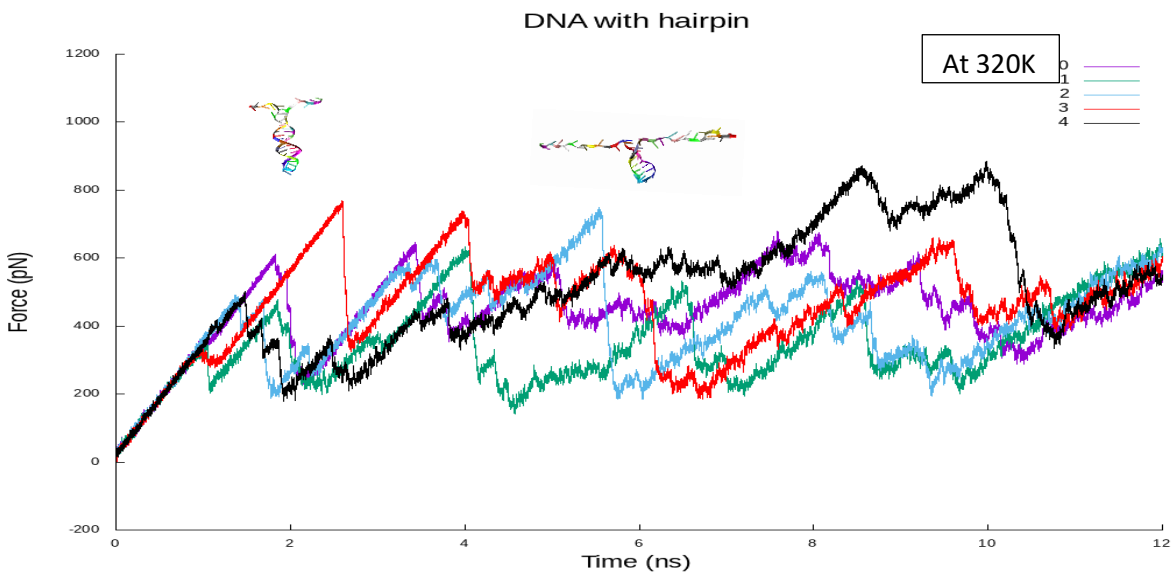



Figure 4.16 DNA with hairpin at 320K (force analysis)

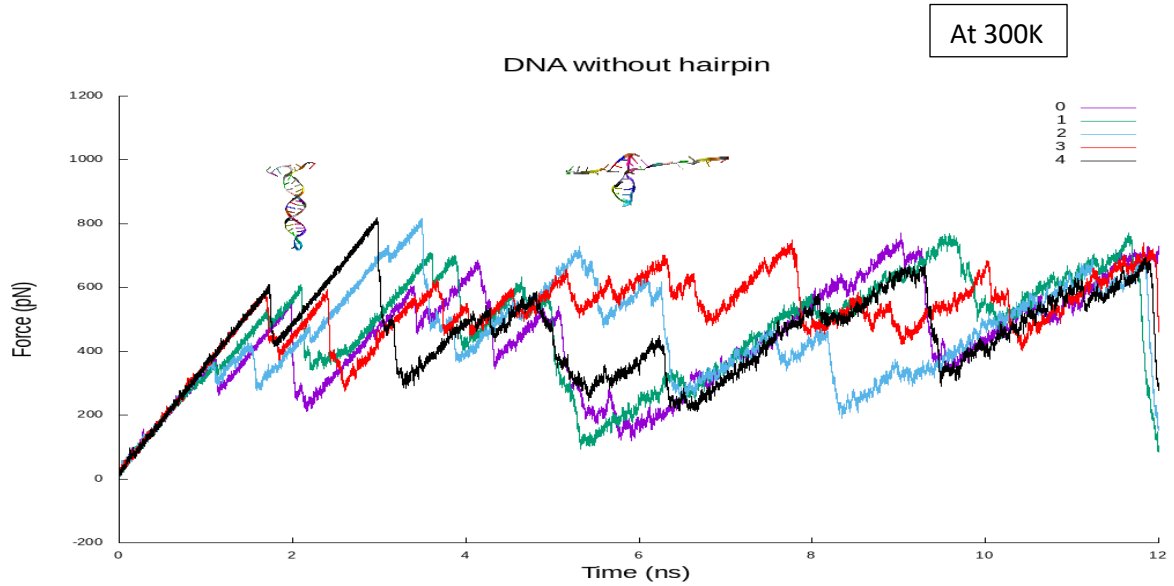



Figure 4.17 DNA without hairpin at 300K (force analysis)

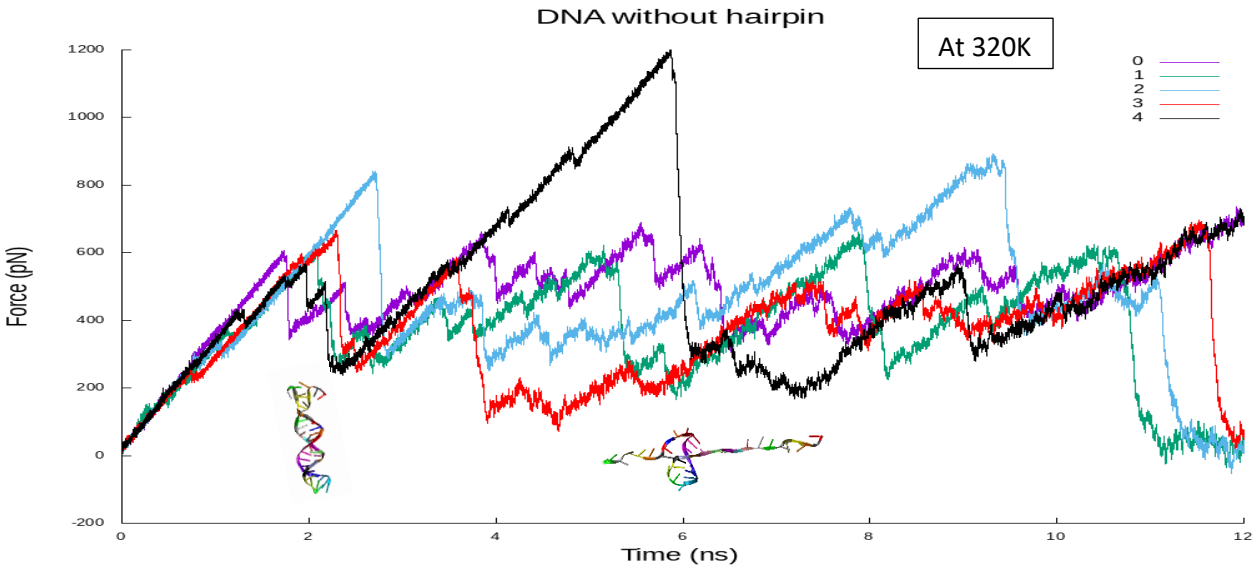



Figure 4.18 DNA without hairpin at 320K (Force analysis)

- On comparing both the system, DNA with hairpin and DNA without hairpin with respect to temperature. It was found that both the system requires less force at high temperature to initiate the rupture.
- In case of DNA with hairpin the value of rupture force goes to 600pN for low temperature and around ~450pN for the high temperature.
- If we look at the overall force with respect to time, we can see that DNA become unstable after some time(~8ns) which results noises in the spectra.

From the overall analysis, it was found that implicit shows much better results than the explicit for both the temperature. But this contradict the fact that explicit shows better results than implicit because there is no water interaction and we use GBIS model which is an approximation to the Poisson Boltzmann equation. We were assuming same effect in both the solvation. But from the results we can conclude, Implicit is much better than the explicit with less simulation time.

# Chapter 5

# Conclusion

### 5.1 System 1(B-DNA dodecamer with sequence 5'-CGCGAATTCGCG-3')

- Steered molecular dynamics of nucleic acid can be carried out in both implicit and explicit solvent models. We can suggest implicit as a solvent which can be used in place of explicit. Both shows the same effect. Other advantage is implicit is less time consuming than the explicit.
- Effect of different force constants at different pulling velocities are studied, suggesting that increasing force constants and pulling velocities results in faster unzipping as expected.
- Our preliminary simulation results indicate that unzipping of A-RNA requires less energy than unzipping of B-DNA, indicating stability of the latter over former. This is because of the stability of B-DNA over A-RNA. RNA contains the ribose sugar which has one more hydroxyl group than deoxyribose (In case of DNA). This results in less stability of RNA.
- Groove size in case of RNA is large than the DNA. So, this can be another reason for requiring less energy to unzip.

### 5.2 System 2 (DNA hairpin and DNA without hairpin)

(A DNA hairpin with sequence 5'-GCGAGCCATAATCTCAATCTG GAAA CAGATGAGATTATGGCTCGC-3') and (DNA without hairpin with sequence 5'-GCGAGCCATAATCTCAATCTG-3')

- Under constant velocity pulling, the separation of double-stranded DNA into two single strands is known to proceed through a series of pauses and jumps [(1)].
- At higher temperature it requires less force than the lower temperature.
- As the separation distance increases, it was found out that there are noises in the graph due to decrease in stability of the system.

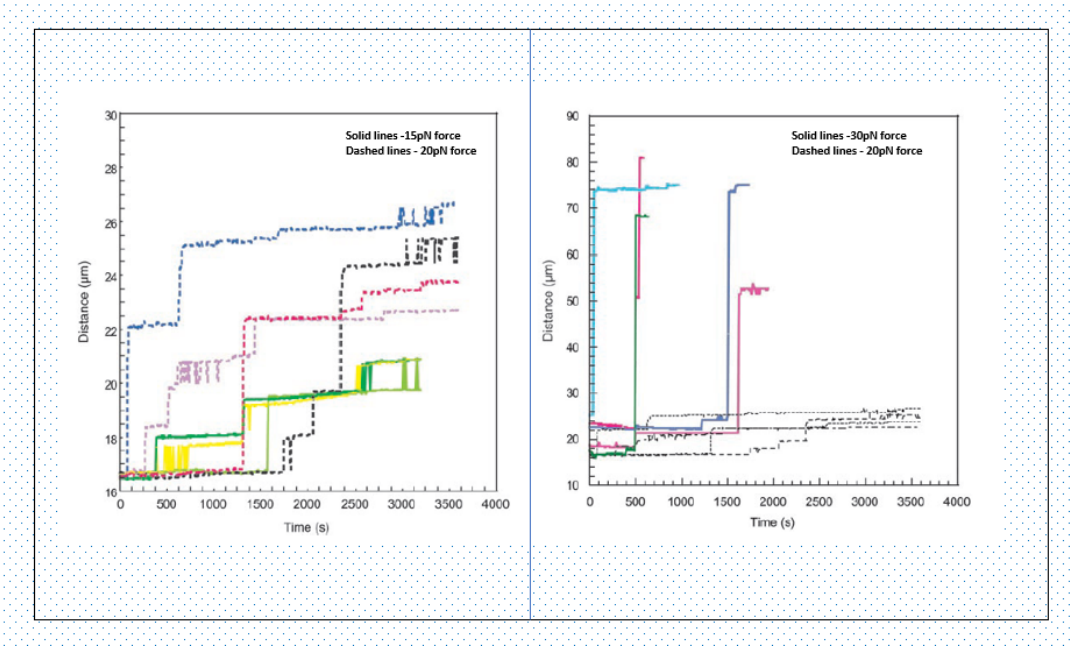



Figure (5.1) – Bond separation at different pulling forces [(2)]

- From the experimental data it was found that our simulation shows the same trend in unzipping through a series of pauses and jumps.

- As we can see in experimental figure 5.1, Dashed line correspond to high force and solid line corresponds to the low force in the left image[(2)]. It was observed that at high force DNA starts unzipping much rapidly than the low force.

- On increasing pulling force, it shows the same effect as increasing pulling velocity. We can see that, at high force it started unzip faster at a short time scale.

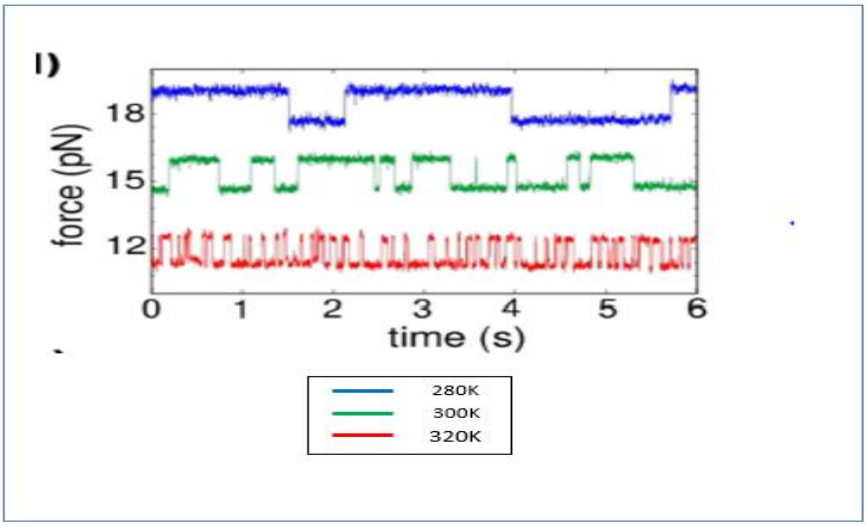



Figure 5.2 - Forces at different temperature- Red profile for 320K, green profile for 300K and blue profile for 280K [(3)]

- Figure (5.2) shows the effect of the temperature, we can see that at high temperature it requires 12pN force but as we decrease the temperature unzipping force increases[(3)]. This figure is comparable to our data because they used DNA hairpin with same sequence.
- Experimental data confirms the validity of molecular dynamic simulation by observing experimental data. It was found that, high temperature requires less force than low temperature.
- On comparing Implicit with explicit, Implicit shows better results with less simulation time. This contradicts that explicit is much better than implicit. So, we can use implicit as a solvation method which is more efficient and less time consuming than explicit.

**5.3 Future work**:

- AT and GC rich nucleic acids will be studied separately.
- Longer simulations will be carried out in both explicit and implicit solvent and comparative analysis with reported experimental pulling profiles will be done.
- In addition to this we can check the effect of hybrid model which includes both implicit and explicit.

# Bibliography


1. Weeks, J.; Lucks, J.; Kafri, Y.; Danilowicz, C.; Nelson, D.; Prentiss, M. *Biophysical Journal* **2005**, *88* (4), 2752–2765.
2. 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.
3. Rico-Pasto, M.; Pastor, I.; Ritort, F. *The Journal of Chemical Physics* **2018**, *148* (12), 123327.
4. Neuman, K. C.; Nagy, A. *Nature Methods* **2008**, *5* (6), 491–505.
5. Vlijm, R.; Torre, J. V.; Dekker, C. *Plos One* **2015**, *10* (10).
6. Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38.
7. 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.
8. Ponder, J. W.; Case, D. A. *Protein Simulations Advances in Protein Chemistry* **2003**27–85.
9. 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.
10. Steinmann, S. N.; Sautet, P.; Michel, C. *Physical Chemistry Chemical Physics* **2016**, *18* (46), 31850–31861.
11. Bashford, D.; Case, D. A. *Annual Review of Physical Chemistry* **2000**, *51* (1), 129–152.
12. http://structure.usc.edu/make-na/
13. https://iith.ac.in/3dnus/