Thermodynamic Studies on DNA Unzipping

THESIS

Submitted in partial fulfilment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

Amar Singh

Under the supervision of

Dr. Navin Singh

BITS Pilani

Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

2015

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE PILANI 333 031 (RAJASTHAN) INDIA

CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled **"Thermodynamic Studies on DNA Unzipping"**, submitted by **Amar Singh**, ID.No. **2010PHXF006P** at Physics Department, BITS-Pilani, Pilani Campus for the award of the degree of Doctor of Philosophy (Ph.D.), is a bonafide record of his original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Signature: .....

Dr. Navin Singh

Assistant Professor,
Department of Physics,
BITS-Pilani, Pilani Campus

Date: .....

Dedicated to My Parents

Acknowledgements

I feel privileged to thank first and foremost my supervisor, Dr Navin Singh, although not for the common reasons but rather for the uncommon ones. I am most thankful to him for providing a stress-free environment and for his great support during the thesis work. In terms of research, he used to say that I am his supervisor. If this is true, it is only because he unsuspectingly encouraged me to remain on the conventional pathways while doing nonconventional research. I am grateful that he has been fully supportive even when mistakes were made and the progress was slow. He has also put a great amount of time on careful reading of the thesis, and valuable comments upon it, for which I am grateful to him.

I would like to offer a special gratitude to the erudite members of my Doctoral Advisory Committee Prof. S N Karbelkar and Prof. Anshuman Dalvi for their support, critical review and suggestions during the progress review and to review my draft thesis. I sincerely thank Prof. D D Pant, Head of Department and Prof. Rakesh Choubisa, DRC Convener for their persistent support.

I am thankful to Prof. R R Mishra and Dr. Kunal Bhattacharya for introducing me to the field of statistical mechanics and numerical simulations during my PhD course work. I would like to thank Prof. Shibasish Chowdhury, Department of Biological Science, for his first advising year and my first course in biophysics, where for the first time I learned about proteins, DNA and RNA structures. Also, I wish to express my gratitude to all the faculty members of physics department for their kind support time-to time. The technical staff of the department, Rajeev Ji & Srikant Ji are also gratefully thanked for their help.

I would humbly express my regards to the Vice-Chancellor, Director and Deans of Birla Institute of Technology & Science (BITS), Pilani for providing me the opportunity to pursue my doctoral studies by providing necessary facilities and financial support. I would like to thank Dr. Anshuman, Unit cheif Estate management for helping me with a pleasant stay in campus.

I express my gratitude to Prof. S K Verma (Dean ARD) and Prof. Hemant Jadav (Associate Dean, ARD), for their constant official support in organizing my research work. All other staff members of ARD, Raghuveer Ji and Mahipal Ji are also gratefully thanked for their help.

I would like to thank all my seniors, in particular Dr. Muninder Jain and Dr. Rajesh Bhatt for their support and motivation in all my endeavours. I would like to extend my warm thanks to my lab-mates Dr. Amrit Sarmah, Dr. Rituparna Bhattacharjee, Captain R Singh, Mr. Arghya Maity and others for their refreshing company and continuous support. I also thank the co-scholars, in particular Dr. Munesh, Mr. Jitendra, Mrs. Keerti, Mr. Ravi, Mr. Surender, Mr. Tridev and others whom I could not mention for space, for moral support, diversified help, and several light-hearted moments which will always be remembered.

I have enjoyed the company of my friends Mr. Vikas Sharma, Mr. Nitish Yadav and Dr. Vijay Kumar who have been great friends and with whom I have shared many memorable moments, I would like to thank them for their friendship and providing me the reference papers whenever required.

The Financial Assistance from UGC, New Delhi in the form of BSR Fellowship and travel grant from ICTP, Italy to participate in a conference, are gratefully acknowledged.

Finally, I owe my heartfelt gratitude to my parents who have always encouraged me to follow my heart and inspired my inquisitive mind throughout my childhood and study career. I wouldnt be here without my family and their moral support and blessings, I am grateful to them for who I am today. Last but not least, I thank my wife Monika, for her patience, help as well as encouragement. It goes without saying that her sterling efforts to keep me cheerful in the tough times, are greatly appreciated.

Amar Singh

Abstract

Prediction of DNA duplex stability and thermodynamics is invaluable for many molecular biology applications. Sequence dependent stability of duplex DNA plays a major role in fundamental processes of the living cell, such as replication and transcription. In these fundamental processes, DNA must undergo dynamical and structural changes. Since these processes require a local opening of the DNA molecule, they resemble DNA denaturation, or DNA melting, which is a considerably simpler process to study theoretically and experimentally. For this, *In-vitro* we need to understand how the CELL environment affects the equilibrium properties and the dynamics of the DNA molecule. After the discovery of the helix structure, the attention was for a long time directed at the DNA denaturation, specifically the effect of chain sequence and length. In past few years a particular interest has been devoted to the study of mechanical strand separation shading light on the role of thermal fluctuations and sequence. In this work, the denaturation and mechanical unzipping of DNA were explored by numerical simulations applying the PBD model. Although DNA dynamics has gained increased interest during the last decades, there is still a need of more insight and knowledge within this field. This thesis work contains the first quantitative study of DNA denaturation as well as the DNA unzipping in concentrated and crowded solutions through the well-known PBD mesoscopic model. We study the thermal melting of short heterogeneous DNA sequences as well as the complete unzipping of λ-phase DNA at different salt conditions. We also study the unzipping of short DNA molecules having defects or mismatches at different locations in the sequence. The outcome elucidate the role of loop entropy and bubble formation in opening of DNA molecules. In case of force induced unzipping not only the magnitude of the applied force is important, but the nature of force, how and where force is applied, is also important in the understanding of unzipping of dsDNA. How and up-to which length the applied force propagates, when it is applied at different sites in the chain, is introduced. We also find the thermal stability of homogeneous DNA molecule in crowded solution. Through the density profile we clearly show how the crowders block the

propagation of bubbles that are created due to thermal fluctuations. This is attributed to the volume occupied by crowder molecules, which inhibits the entropy increase necessary for DNA melting. Finally, the work presented in this thesis can be extended to find practical applications of DNA unzipping, such as sequencing of DNA by critical force and melting temperature and measurement of thermodynamic properties of DNA in conditions not accessible by bulk methodologies.

Preface

About the thesis

The thesis presents a theoretical study on the base pair separation in the DNA molecule in concentrated solutions. For the studies we considered the Peyrard Bishop Dauxois model, popularly known as PBD model. In order to understand the dynamics of the transcription and replication processes, it is very important to study the localized fluctuations in DNA molecule. The thesis work is deeply motivated by the recent experimental studies that demonstrate the thermal melting or forced induced melting in crowded and concentrated solutions. The focus of thesis is to understand the nature of transition from double stranded DNA (dsDNA) to single stranded DNA (ssDNA) in crowded and concentrated solutions which is an important part of these transitions but somehow unattended. The defects in the DNA molecules play a crucial role in the various functioning of the molecule. These defects in the molecule may cause cancer or sometimes the cell death. We studied the thermal as well mechanical response of DNA molecule that is having defects on different locations.

To the reader

The reader is assumed to be associated or working in soft condensed matter physics. The main prerequisite for the reader is to have basic knowledge of statistical mechanics and numerical techniques. The thesis requires some basic biology understanding which may make reader comfortable in relating our studies with the real biological processes.

The thesis is written in "we" form. Sometimes this may appear strange to the reader, as the thesis is of a single author. However, in order to have continuity and flow in the text, I stick to use *we* instead of *I* throughout the thesis.

Thesis outline

The thesis is divided into eight chapters, where the first two, respectively are meant as background to the particular fields in biology and physics studied. In chapters 3-7 we discuss the research work that has been carried out. Here is brief discussion about the organization of the thesis.

- In **chapter 1** we discuss the basic structure and dynamics of the DNA molecule. A detailed literature review of the theoretical as well as the experimental studies that is related to our work is also present in this chapter.
- The details of the statistical model and the methodology that is used to study the research problems is discussed in **chapter 2**.
- In **chapters 3 & 4** the phase transitions in DNA molecule in the concentrated solutions are discussed. How the salt that is present in the solution containing DNA or the solvent molecules affect the DNA stability are the focus of these chapters. The experimental studies on the force induced unzipping are performed in two different ensembles. The equivalence of these ensembles is also discussed in 3^{rd} chapter.
- In **chapter 5** we explore the role of defects in DNA molecules on the thermal as well as on the mechanical stability of the molecules.
- **Chapter 6** discuss an important but unattended area of the thermal melting of DNA. Here we study thermal melting of DNA in crowded environment.
- The force that is applied on a DNA chain has a gradient along the chain. Is there a point above which the chain does not feel the force is explored in **chapter 7**.
- Finally, the research work is summarized in chapter "Conclusions and Future work". Here we outline some conclusions and the future scope of the work.

Contents

List of Figures

LIST OF FIGURES

List of Tables

Nomenclature

Symbols

η	Crowder density
γ	Solution constant
κ	Single strand elasticity
λ	Solution constant
ϕ	Fraction of open pairs
ρ	Anharmonicity in stacking
θ	Fraction of intact pairs
a	Inverse of potential width
b	range of anharmonicity
C_v	Specific heat
D	Potential depth
f	Free energy
F_c	Critical force
H	Hamiltonian
k_β	Boltzmann constant
S	Entropy
T_m	Melting temperature
V_M	Morse potential

Z	Partition function
Z_c	Configurational part of the partition function
Z_p	Momentum part of the partition function

Acronyms

A	Adenine
AFM	Atomic force microscop
C	Cytocine
CEE	Constant extension ensemble
CFE	Constant force ensemble
DNA	Deoxyribonucleic Acid
dsDNA	Double stranded DNA
G	Guanine
MC	Monte Carlo
MD	Molecular Dynamics
mRNA	Messenger RNA
MT	Magnetic tweezer
OT	Optical tweezer
PBD	Peyrard Bishop Dauxois
RNA	Ribonucleic acid
SMFE	Single molecule force experiment
SMFS	Single molecule force spectroscopy
ssDNA	Single stranded DNA
T	Thymine
tRNA	Transfer RNA

Chapter 1

Introduction

Deoxyribonucleic acid (DNA) is a biomolecule that contains genetic instructions used in the development and functioning of all known living organisms [1, 2]. The main role of DNA molecules is the long-term storage of genetic information. DNA is often compared to a set of blueprints, like a recipe or a code, as it contains the instructions needed to construct other components of cells, such as proteins and ribose nucleic acid (RNA) molecules. The DNA molecule carries the description of each protein in terms of nucleotide sequence. The segments of DNA chain that carry these genetic informations are called GENES. Rest part of the chain have the structural purposes [3–6]. They are involved in regulating the use of these genetic informations through the two central and vital processes, the *transcription* and *replication* of DNA. In both the processes one thing is common that is opening of the helix, either to decipher the code (DNA transcription) or to serve as a template for daughter DNA (DNA replication). *In-vitro*, these are known as *denaturation* of DNA. In order to understand these important biological processes in depth, one has to study the opening of the double helix DNA and the separation of the two strands in DNA [7–9]. In this thesis we investigate statistical and thermodynamic properties of DNA denaturation under different physiological conditions.

In this chapter, we briefly discuss the structural and dynamical behavior of DNA. A brief literature survey on theoretical as well as on experimental studies on denaturation of DNA is also presented.

1.1 The structure of DNA

The DNA molecule is an assembled form of two large linear polymers which are known as the *strands*. The monomer or the repetitive unit of these strands is

known as nucleotide. The constituents of a nucleotide are:

1. Phosphate,
2. de-oxy ribose, sugar and
3. hetrocyclic carbonic bases

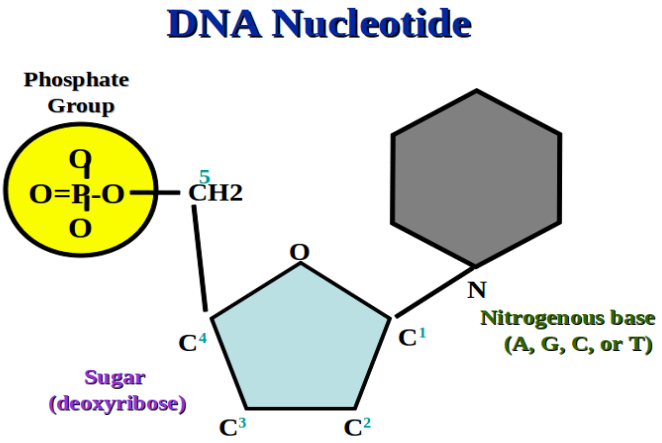


Figure 1.1: The monomer of DNA strand; the nucleotide. The phosphate and carbonic base are covalently bonded with the sugar.

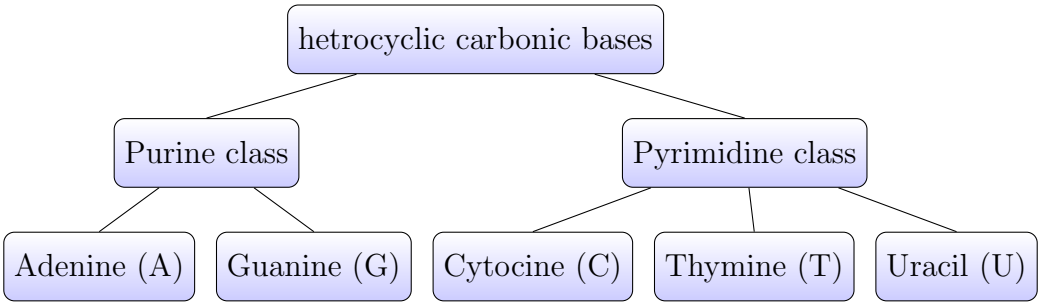


Figure 1.2: Classification of carbonic bases

In nucleotide, the phosphate group and carbonic base are covalently bonded with the sugar, as shown in Figure 1.1. These bases are broadly classified under two categories: Purines and Pyrimidines, as shown in Figure 1.2. Purines are relatively bigger molecules[1]. The chemical structure of these bases is shown in Figure 1.3. In the *strands* of DNA, the nucleotides are joined through the ester bonds between the sugars and phosphate group as shown in Figure 1.3. The phosphate groups

[1]Uracil (U) is only for RNA molecule, instead of thymine (T) in DNA.

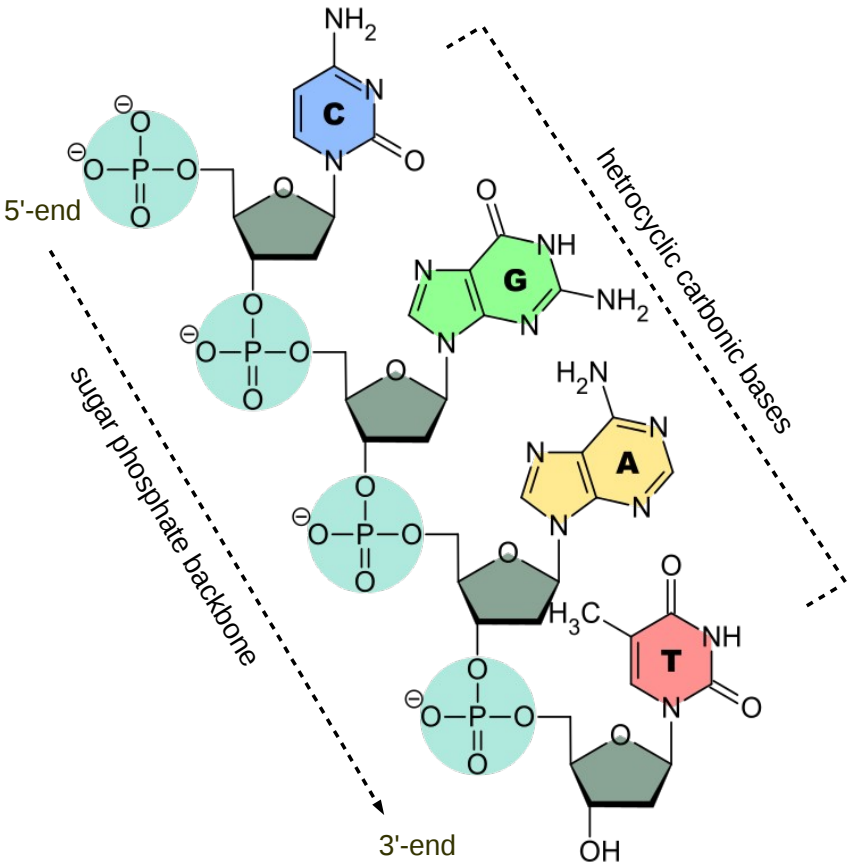
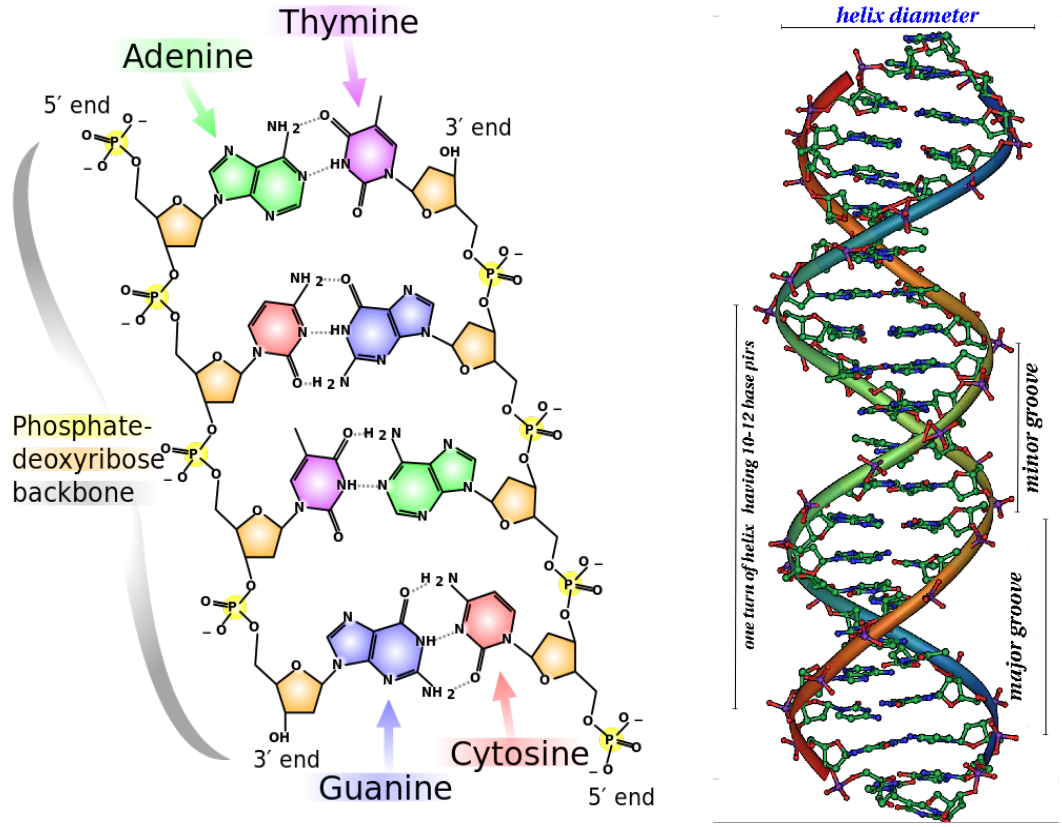


Figure 1.3: Primary structure of DNA, (single *strand*). The nucleotides are joined by ester bonds between sugars and phosphate group and make the sugar-phosphate backbone. The end with a terminal hydroxyl group, called 3-end, and the end with a terminal phosphate group, called 5'-end. *Image taken from Wikipedia Commons*

that form phospho-diester bonds with the 3^{rd} or 5^{th} carbon atoms of adjacent sugar rings, give directionality to the DNA strand. The end with a terminal hydroxyl group is called a 3-end while the end with a terminal phosphate group is called a $5^{'}$-end as shown in Figure 1.3 [2, 8–12]. The carbonic bases are covalently bonded with sugar. Thus the single strand of DNA is a quasi regular structure. The regular part is formed by alternating sugar-phosphate groups, and irregular part is the bases in the chain [1, 2, 8, 10, 13]. The sequence of these four bases, along the backbone[1], encodes the information that is known as **genetic code**. This code specifies the sequence of the amino acids. Any changes in the sequence of these bases can lead to crucial changes in the properties of the organisms and in its functionality [2, 6, 7, 11]. The sequence of nucleotides in a DNA strand gives the molecule's primary structure.

[1] by convention, sequences are usually presented from the 5'-end to the 3'-end



(a) H-bonding between complementary strands of DNA (b) DNA double helix structure

Figure 1.4: The structure of DNA. (a) Two strands of nucleotides with covalent bonds shown as solid lines and hydrogen bonds between bases as dashed lines. (b) Double helical structure of DNA in B-form. *Image taken from Wikipedia Commons*

The secondary structure of DNA molecule is formed when two single strands of DNA are coupled with hydrogen bonds between the bases as shown in Figure 1.4(a). The bonding between these bases defines a *base pair*. These two strands of DNA run in opposite directions and are therefore anti-parallel to each other. In most cases, all bases of one strand pair with a complementary base on the other strand. The geometry and chemical composition of the bases allow the most stable pairing when A bonds with T and G bonds with C. That is why these are called as complementary bases. The H-bond strength for these two base pairs is different. The A-T base pairs is formed by two H-bonds, and the C-G base pairs bind via three H-bonds. There is also a small heterogeneity in the masses of different base pairs, but it differs only by about 3% [1, 3–6].

The nature of the bases and sugar-phosphate backbone is quite different towards water. While bases are hydrophobic (water hating) in nature, the sugar-phosphate are hydrophilic or hydro-soluble molecules. Therefore the bases along

the strand tend to come together (thus keeping water away from the core of DNA) and the highly negatively charged sugar-phosphate get pushed out (thus protecting the core). Hence the orientation of the base pairs is such that each base pair gets a rotation of around 35° with respect to the previous one, around a common central axis. As a whole this forms a double-helix structure [2, 8, 10, 11]. In physiological conditions, the double-helix structure of DNA is very stable. This stability is the result of the following two contributions. Firstly, the helix structure of DNA molecule is formed via the H-bonding between the bases of single strands. Secondly, the double-helix DNA is stabilized by base-pair stacking interactions between adjacent nucleotide base-pairs. The base-pair stacking potential is a short-ranged attractive interaction[1]. It is attributable on one hand to attractions between π orbitals of aromatic rings in successive bases and on the other hand to hydrophobic interactions that tend to push bases together. Such a structure has the benefit of prohibiting genetic information from being lost or destroyed, but it causes severe constraints on the access of genetic information and the transfer of it from one generation to another [1, 2, 8, 16].

There are several possible conformations for the double helix that are differ by spatial positions & orientations of the atoms and the direction of the helix turn. The most abundant form in living cells is B-DNA. In this conformation, a turn of the double helix consists of about ten nucleotides and the diameter of the helix is 22 to 26 Å. B-DNA is a right-handed double-helix, each pitch contains about 10 - 11 base-pairs, and two neighboring base-pairs are separated vertically by about 3.4 Å, the pitch length of the double-helix along the central axis is about 36 Å. Similarly A-DNA, Z-DNA, S-DNA etc. are the other forms of DNA. According to Chargaffs rule the amount of Guanine is equal to Cytosine and the amount of Adenine is equal to Thymine in all these conformations of DNA molecule [17, 18]. In biological cells, DNA is usually very long[2], it's length ranges from millimeters (bacteria) to meters (human) and even approaches kilometers (salamander) [19, 20].

1.2 Importance of DNA in Cell

Every life form, reproduce itself by cell division and within cells, DNA is organized into long structures called chromosomes. Within the chromosomes, chromatin

[1] the stacking interactions are sequence-dependent: different sets of bases have distinct stacking energies [14, 15]

[2] the total number of base pairs in one cell of our body is of the order of ten billion (10^{10})

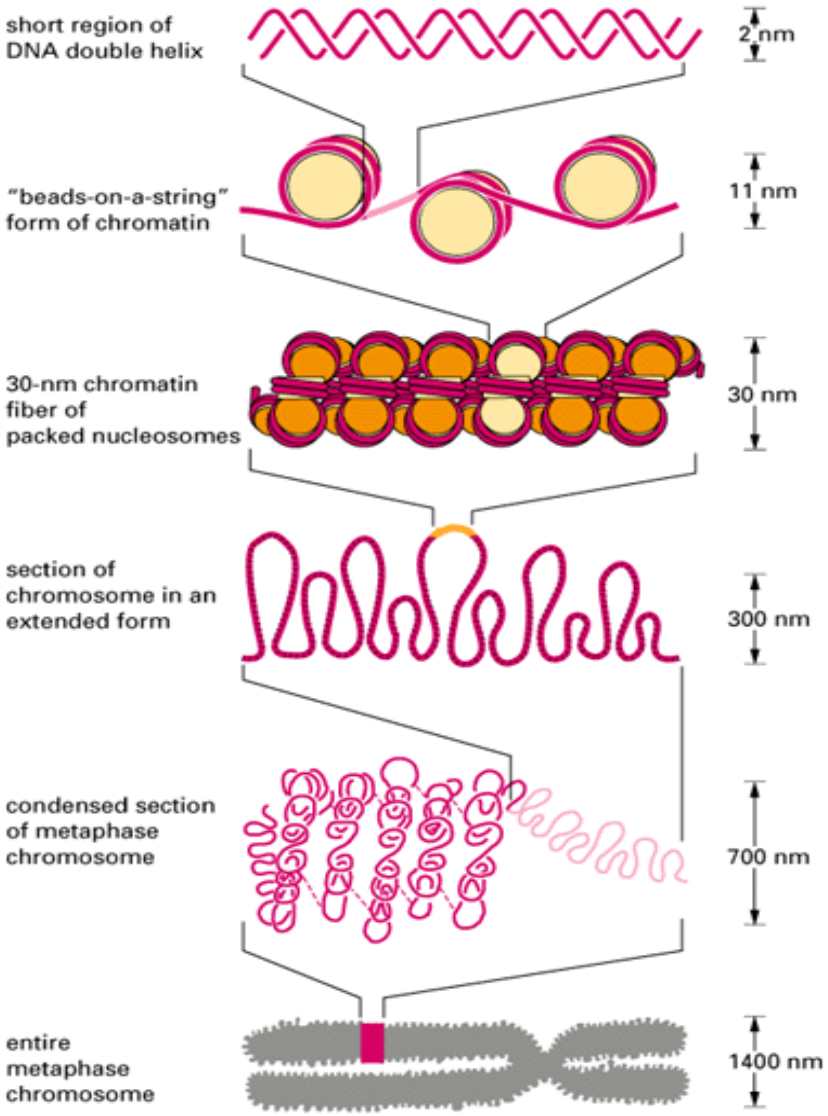


Figure 1.5: The organization of DNA within the chromosome structure. Image taken from Reference [21]

proteins such as histones compact and organize the DNA molecule[1] as shown in Figure 1.5. The first step in the study of important biological processes consists of understanding the properties of the DNA double helix and how it interacts with proteins. Among these processes the *transcription* and *replication* of DNA, are the two central and vital processes for sustaining life, which involve a dynamical alteration of the molecule. In this section these two are briefly discussed here.

[1] the purpose of this packing is to allow the long DNA molecule to fit in a nucleus.

DNA replication

During cell division, DNA is transferred from parent cell to its daughter cells through a *replication* process. First DNA is replicated and then carried to the next generation of cells. The replication process, shown schematically in Figure 1.6

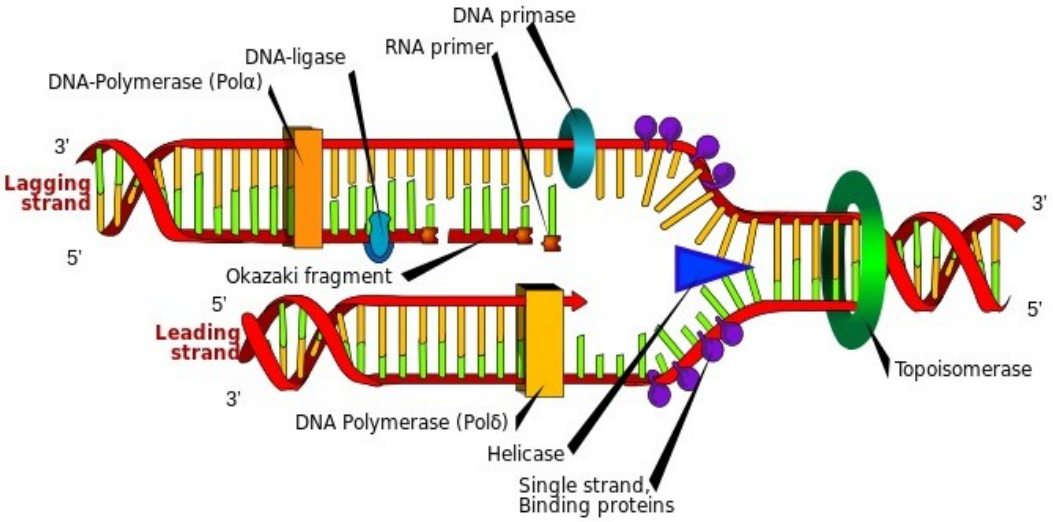


Figure 1.6: Schematic representation of DNA replication, it is the process of copying a double-stranded DNA molecule. *Image taken from Wikipedia Commons*

starts by the building of a short single stranded ribose nucleic acid (RNA) molecule, which acts as a primer for the duplication process[1] [2, 6, 10, 11]. This practically means that before each cell division, the DNA sequence must be duplicated. The replication process is initiated by an enzyme called DNA-polymerase. It starts by producing a local opening of the base pairs, separating the two strands from each other. The local opening then traverse through the chain like Y-fork. The bases are matched to synthesize the new partner strands known as leading strand and lagging strand. Many enzymes are involved in the DNA replication fork.

In the last few decades our knowledge about DNA replication has greatly improved and this process is paramount to all life. Scientists used a variety of experimental approaches to identify the genes that are crucial in copying of the DNA. It is clear that the genetic content of the DNA molecule is kept protected inside the helix, and that any read-out of the sequence needed a strand separation. Scientists thus started to search for conditions that would disrupt the hydrogen bonds which connect the two strands.

[1] this primer binds to each of the open strands.

DNA transcription

The other important phenomenon that involves the opening of DNA base pairs, is *transcription*. In this process the informations contained to form a gene, are read. During each cell cycle, the DNA sequence is read by an enzyme called RNA-polymerase and genetic information is transcribed into a messenger RNA (mRNA). This process is called *transcription* [2, 6, 10, 11]. The mRNA is then transmitted to another part of the cell and read by a protein synthesis machine called the ribosome[1]. In order to initiate this process, RNA polymerase has to

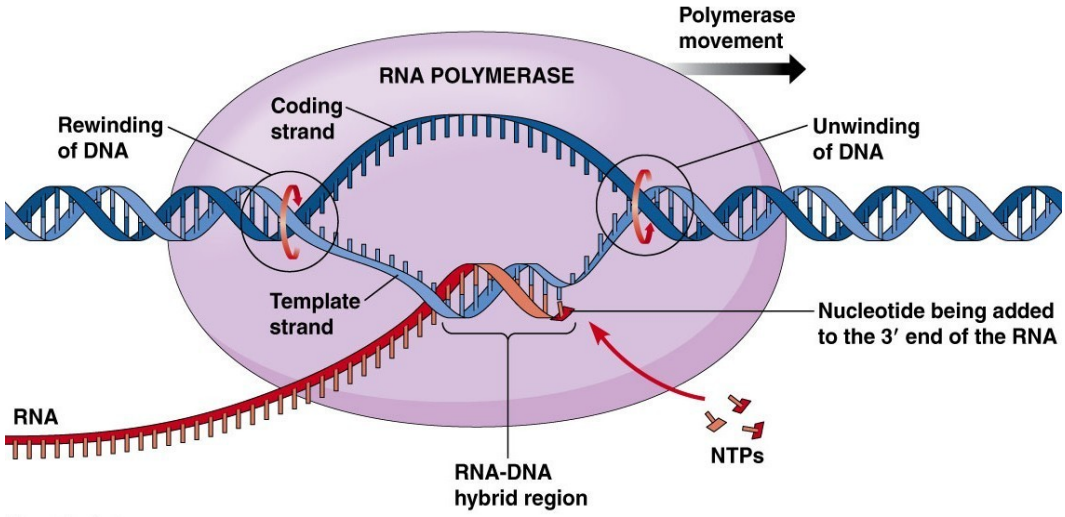


Figure 1.7: The figure depicts the process of DNA transcription. The coding strand of DNA is used as a template for the mRNA. *Image taken from Wikipedia Commons*

recognize and connect to a specific site on double stranded DNA[2], that originates a local opening of the base pairs in DNA chain. It is therefore necessary that base-pairing is only marginally stable, so that the helix can be opened and the sequence read. RNA polymerase has the ability to close the bases after copying the sequence. The process is done in a very coordinated manner with a speed of several tens to hundred base pairs per second [10, 20]. The process is shown in Figure 1.7.

From the above structural and functioning discussion of DNA, one finds that the double helix DNA has several advantages. Here some of the most relevant aspects are listed down:

[1] in the ribosome a process called translation begins, where the mRNA is used as a template to synthesize a protein.

[2] transcription process is controlled by so-called "transcription factors", which form a pre-initiation complex.

- Genetic information is coded twice in the two complementary strands. This allows to store the "information" and a check for the errors during the replication process.

- The external phosphate backbone protects the internal base-pairs from any irreversible damage, as these base-pairs are the carrier of the genetic information .

- The lineal or stacked arrangement of the bases along the longitudinal axis of the DNA allows proteins to directly access the fragment of the sequence.

- The process of opening (unwinding) and closing of the two strands of DNA is reversible. That is why the replication and the transcription processes carried out without damaging the original molecule.

1.2.1 DNA helix opening in-vitro

One can separate the two strands of DNA by increasing the temperature of the solution that contain the DNA molecule. This is known as "thermal melting" or "temperature induced DNA melting". In this process the hydrogen bonds between bases are broken and the two strands separate from it's helix structure. It initiates in the chain with the formation of local "denaturation bubbles" in the same way as in transcription and replication. Thus, DNA denaturation is seen as a valid approach on the way of understanding the mechanism of transcription and replication[1]. The other way is by pulling one of the DNA strand keeping other strand attached to a glass slide. This is known as "force induced DNA melting" or "DNA unzipping".

Soon after the discovery of it's structure in 1953, attempts were made to study the denaturation process. Although it is a field in a continuous evolution, there are still many open questions regarding DNA: for example, how RNA polymerase search and connect to a specific site on double stranded DNA, how this can be realized *in vitro* etc.? Moreover, recent developments in experimental techniques have made it possible to manipulate DNA for nanotechnology [22–24], for storing molecular memory [25, 26]. Hence, it creates an interest to study DNA properties in conditions that are even not necessarily physiologically relevant.

[1] although it differs from the base pairs opening that occurs during transcription, its understanding can still bring a lot of useful information on what happens in this process.

Literature review

In order to understand the DNA denaturation or DNA melting, various interactions and mechanism *in-vivo* or *in-vitro*, many experiments have been done and various theories have been proposed. As the DNA research is very vast and continuous evolutionary, it is difficult to put all together. Hence we focus on the literature/work that is related to the problems, investigated in this thesis. The next two sections cover these studies on DNA melting and unzipping.

1.3 Thermal Melting of DNA

Melting of DNA, (also called *denaturation*) is a process in which the hydrogen bonds between bases are broken and the two strands separate from it's helix structure due to thermal fluctuations. The H-bonding between the bases on opposite strands, is much weaker than the covalent bonds in the molecule. The energy needed to break a hydrogen bond is about 10^{-2}-10^{-1} eV, while to break a covalent bond it is 5-10 eV. Therefore, in the temperature ranges for strand separation, melting does not affect the primary structure of DNA. It is is a reversible process. In the thermal melting of DNA, the thermal energy provided to the system transforms the ordered helical structure into the state of a disordered coil [27–30]. This is known as *helix-coil* transition. The helical state is energetically favorable state while the coil state is more entropic. Physicists are now quite interested in the physical significance of such dynamical behavior of DNA double-helix induced by thermal energy [31–35].

The internal dynamics of DNA has been studied by different experimental methods. The most widely used method for the experimental study of DNA denaturation is UV absorption spectroscopy [2, 3, 6]. An increase in temperature causes a sudden opening of base pairs, which creates *bubble(s)* in the sequence. Once the bubble is formed, it may grow and hence break the other base pairs. This process is similar to the nucleation and propagation of bubbles in crystal structure. The change in stacked bases consequently accompanied by an abrupt increase in the absorption[1] as shown in Figure 1.8. The temperature at which half of the base pairs are open in a DNA chain, is known as the melting temperature, T_m [2, 3, 36]. As a whole, DNA denaturation process is an order-disorder transition, the order state being given by the intact bases, while the disordered

[1] the intact base pairs in double stranded DNA absorb less ultraviolet light than the bases in a single stranded chain.

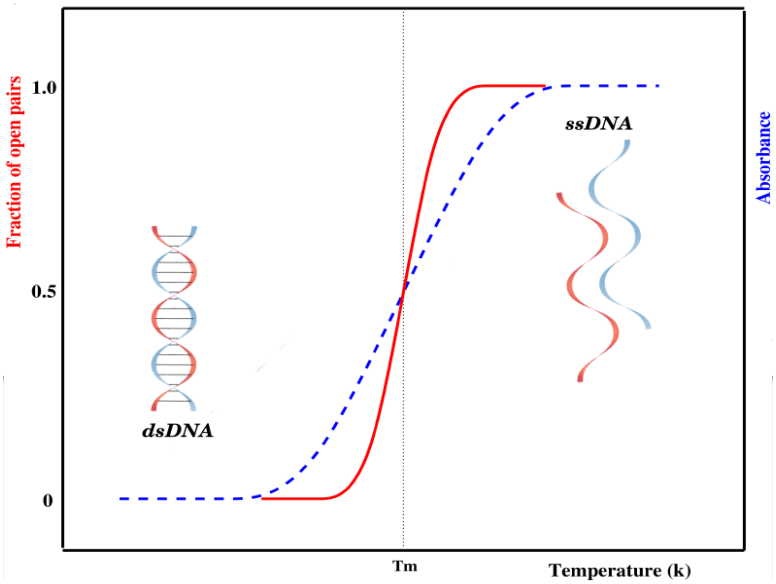


Figure 1.8: Figure showing increase in fraction of open base pairs (solid line) with temperature compared to increase in normalized UV Absorption (dotted line).

state corresponds to loops formed by broken base pairs [31, 37, 38]. The melting temperature of a DNA molecule depends on several factors such as: length of the molecule, specific base pair sequence[1], number of mismatches or defects in the sequence, etc. The solution conditions like: buffers, pH, salt concentration, hydrophobicity, or even surfactants may also affect the melting transition. Excluded volume[2] is another factor that can influence the stability of DNA molecule. Depending on these specific conditions the melting temperature, T_m of DNA typically ranges between 10°C and 100°C [33, 39]. In this section, we focus the discussion related with thermal melting of DNA in presence of salt (cation), in the presence molecular crowders and in presence of mismatch or bubble in the sequence.

1.3.1 Role of salt concentration

Experimental studies:

In the mid 1950s several experiments demonstrated that heating DNA solutions resulted in a cooperative unraveling of the double strand helix to single strands [40, 41]. DNA melting has been used as a measure of DNA stability since 1980s when melting analysis was first combined with polymerase chain reaction (PCR) [42].

[1]GC pairs are formed of three hydrogen bonds, while AT ones have only two, so the latter opens at a lower temperature.

[2]excluded volume refers to the area in a solvent that is occupied by another molecule.

This technique is used to amplify a single copy or a few copies of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence. It is also used to determine the effect of solution conditions. It has been found that the salt (in form of $NaCl$ or $MgCl_2$) concentration of the solution (or ionic solution), influences the melting temperature [43–45]. Salt dissolved in the solution release Na^+ or Mg^{2+} cations. As the DNA molecules are strong polyelectrolytes, having negatively charged phosphate groups, researchers found that it would be interesting to analyze the role of cations (Na^+ or Mg^{2+}), in the melting transition of dsDNA.

The electrostatic free energy change in transition from double helix to single strand, is related to the electrostatic repulsion between the phosphate charges on the opposite strands of DNA. The high concentration of the counterions screens the negative charges on the two strands. [43, 46, 47]. Experimental data relating to the dependence of the melting temperature on GC content and concentration of Na^+ ions in solution were obtained by Owen *et al.* [48]. The analysis of experimental findings support the hypothesis that the change in T_m with salt concentration is due to changes in the screened interactions between the negatively charged phosphate groups. For example, the T_m of 92 different DNA sequences were measured over a wide ionic concentration range by R. Owczarzy *et al.* [49]. They derived a relationship for scaling the T_m of DNA duplex oligomers between different ion concentrations. Later on this group predicted stability of DNA duplexes in solutions containing magnesium and monovalent cations [50]. DNA double helix is also stabilized by stacking interactions between base pairs. P. Yakovchuk *et al.* studied the salt dependent base-pairs stacking contribution to the DNA duplex stability [51]. More recently Vuletić *et al.* [52] explained that how the DNA conformations changes with a decrease in DNA concentration in the very low added salt environment. They also validated the Manning condensation and conductivity theories devised for dilute aqueous polyelectrolytes in the absence of added salt.

Several groups observed the effect of salt on the melting temperature of short synthetic DNA *hairpins* [53–55]. DNA hairpin is formed when self-complementary DNA sequence have the potential to form stem-loop structures. The formation of hairpin structure is shown in Figure 1.9. These hairpin molecules are useful for thermodynamic studies, as they form stable duplexes that are partially paired over a convenient temperature range [54]. The effects of sequence and loop size on hairpin stability have been reported by different groups [56–58]. More recently, the effect of monovalent cations size on the thermal stability of DNA hairpin is

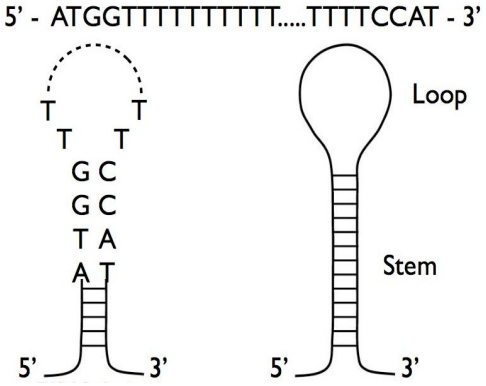


Figure 1.9: The formation of DNA hairpin. The segments of the sequence can pair with itself, by forming hydrogen bonds between complimentary bases and form a stem-loop structure, known as DNA hairpin.

measured by Earle Stellwagen *et al.* [59]. The melting temperature decreases with larger size cations as the larger cations are less effective in shielding the charged phosphate residues in duplex DNA. This may be due to the fact that the larger cations cannot approach the DNA backbone as closely as smaller cations. Despite various attempts, the understanding of *helix-coil* transition in an ionic environment is still far from any conclusion.

Theoretical studies:

Soon after the discovery of DNA double helix structure, scientists proposed or developed some statistical models to understand the DNA denaturation. The helix-coil transition in DNA has been a subject of very intensive theoretical investigations since 1960s [4, 31, 60–63].

Several group of researchers proposed some empirical relations or formulas to predict the experimentally observed salt dependent melting temperature of a DNA molecule. Frank-Kamenetskii [44] presented a relation to predict the experimental data of Owen *et al.* related to the salt dependent melting temperature. J. Santalucia Jr., compared the different nearest neighbor parameters that were used to study the salt effect on the melting transition in DNA [64]. These different studies were presented in different formats of nearest neighbor (NN) thermodynamics. The NN model for nucleic acids assumes that the stability of a given base pair depends on the identity and orientation of neighboring base pairs. In his studies a unified set of NN parameters were proposed.

Using Poisson-Boltzmann (PB) and counterion condensation (CC) polyelectrolyte theories, N. Korolev *et al.* calculated the electrostatic free energy differ-

ence between the single-stranded and double-helical DNA [65]. The results clearly indicate that the stability of DNA in solutions containing $NaCl$ and $MgCl_2$ can be explained by the change in the electrostatic component of the free energy of double helix and single strand conformations. Later on, the base pair opening probabilities in B-DNA were calculated at different salt concentrations by Chen and Prohofsky [66]. They used a model proposed by Prohofsky and co-workers [67, 68]. It was based on self consistent phonon approximation (SCPA) theory and real configurations of DNAs were considered. The H-bonding of bases were represented by Morse potentials as a function of the distance between paired bases, while the other covalent bonds assumed as harmonic potentials with appropriate parameters.

Krueger *et al.* studied the base pair opening probabilities [69] by calculating the partition function using the unified nearest neighbor parameters. They proposed that the open base pair proceeds through a formation of a highly constrained small loop, or a ring. An entropic penalty parameter, "ring factor" was included for the unfavorable positioning of the unbound base. The values for this factor were estimated from the comparison of theoretical probabilities with the probabilities measured for by NMR experiments. Zhi-Jie and Shi-Jie used the tightly bound ion theory (TBI) [70] to investigate how the metal ions affect the folding stability of B-DNA helices. The basic idea of the model was to separate the tightly bound ions from the diffusive ions in solution. The model explicitly accounts for the discrete modes of ion binding and the correlation between the tightly bound ions, and treat the bulk solvent ions using the Poisson-Boltzmann theory. They quantitatively evaluated the effects of ion concentration, ion size and valence, and helix length on the helix stability. Moreover, they derived practically useful analytical formulas for the thermodynamic parameters as functions of finite helix length, ion type, and ion concentration. they found that the helix stability is additive for high ion concentration and long helix and non-additive for low ion concentration and short helix [71].

The process of thermally induced DNA melting has been studied extensively over the last few decades in the framework of the Poland-Scheraga model (PS model) [31, 72, 73], which was proposed in 1966. In this model the DNA molecule is considered as being composed of an alternating sequence of bound and denaturated states. The bound state is energetically favored over unbound states while the loop segment or open states are entropically favored. Jost and Everaers [74] introduced a local sequence-dependent salt correction of the nearest-neighbor pa-

rameters in this model. They compared the predictions of this unified PS model with the experimental data. The melting behavior and probabilities for single base-pair opening of long DNA chains were examined by N Theodorakopoulos [75] within the framework of Peyrard-Bishop-Dauxois (PBD) model. In this model the hydrogen bonding between bases is represented by Morse potential while stacking energies are presented by a anharmonic potential [76, 77]. The calculated probabilities for single base-pair opening were consistent with values obtained from the experiments.

Y. Lui [78], developed a thermodynamic model to predict DNA melting in ionic solutions. The predicted melting temperatures and melting curves through their model, capture the general feature of DNA melting and are in good match with the available simulation and experimental results. They found that T_m increases with increasing ionic concentration, in particular, sharply at low ionic concentrations and much less at high ionic concentrations. The sharp increase of T_m at low $NaCl$ concentrations is because of the significant increase in the number of counterions near the DNA as a result of electrostatic attractions. As there should be a threshold for the presence of counterions, increasing more the salt may not lead to the significant increase in the T_m.

1.3.2 DNA in a crowded solution

Experimental studies:

This is a known fact that the core of the cell is having large number of biomolecules that reduces the free volume of the cell. About 20-40% of the total volume (∼50-400 g/l) of the cell is occupied by the nucleic acids, protein, lipids, saccharides etc [79–82]. Presence of such biomolecules defined as *molecular crowding*. The presence of crowders in the cell restrict the movement of individual molecules that suppresses the thermal fluctuations. It is important to note that the DNA melting and other important *in-vivo* cellular processes occur in this crowded environment [83–86].

In recent years, there has been a growing appreciation of the impact of molecular crowding on DNA duplex stability. The obtained results by different groups are quite interesting which shows that melting temperature of the DNA chain increases about $2 - 20°C$ in crowded environment. *In-vitro*, Poly(ethylene) glycol (EG or PEG) and dextrans are the most commonly used molecules as a cosolute[1]

[1]the term cosolute is refer to polymer molecules used for creating molecular crowding, which behave like solutes rather than solvents of biomolecules.

in aqueous solution to mimic cellular environments. The prime reason for their use is that they are inert with nucleotides and they are available in different molecular weights. Osmolytes[1] are also used to study the base pair stability in the crowded solution [87]. Harve *et al.* [88], found a variation of 0.5-2.5 K in the increase of melting temperature of 20 oligomers DNA in crowded solution. Nakano *et al.* group reported that there is a decrease in T_m with low molecular weights PEGs or EGs concentrations while they observed an increase in T_m with high molecular weight PEGs [89]. As the large cosolutes are having large steric hindrance which favor the reactions that decrease the net volume hence there will be an increase in T_m. While the short cosolutes (ethylene glycol) are having lower steric hindrance and hence there will be a decrease in the T_m. In the light of a reduced water activity the role of excluded volume or crowders on the stability of duplex form of DNA can be explained [90, 91].

The DNA duplex stability studied in ionic crowded environment which suggests, the stability of the DNA at high polyanion concentration is significantly increases compared to any other similar ionic strength conditions involving just $NaCl$ or a mixture of $NaCl$ with PEG [92, 93]. This results in an additional electrostatic contribution on top of the excluded volume effect. The decrease in T_m, at higher ionic concentration is attributed to an increased electrostatic repulsion among the DNA phosphates and the modification of the electrostatic interactions with counterions. Recently, the effects of loop length on the conformation, thermodynamics stability and the hydration of DNA Guanine rich quadruplexes[2] under molecular crowding conditions have been investigated [94]. The effects of molecular crowding with PEP-Na (ploy ethylene sodium phosphate) on the thermodynamics of DNA duplexes, triplexes and G-quadruplexes were systematically studied [95]. Thermodynamic analysis demonstrated that PEP-Na significantly stabilized the DNA structures. The effect of polymeric solutes on the thermal denaturation behavior of DNA-Gold nanoparticle assemblies also studied by Goodrich *et al.* [96]. Polymeric solutes can dramatically affect biochemical reactions via molecular crowding. However, both PEG and dextran increase the stability of DNA-Gold aggregates, melting transition temperatures in the presence of PEG, are affected more significantly. For a high ($\sim$15%) weight percent of PEG, aggregation was

[1] Osmolytes are a class of compounds ubiquitously used by living organisms to respond to cellular stress or to fine-tune molecular properties in the cell

[2] G-rich strand of human telomeric DNA can fold into a four-stranded structure, that is called as G-quadruplex.

observed even in the absence of complementary oligonucleotides[1]. All these studies indicate the importance of molecular crowding and their impact in designing the bio-sensors or the drug design etc.

Theoretical studies:

Theoretical studies on the effect of molecular crowders on the dynamics of DNA is very limited. However, last few years witnessed a growing interest in the theoretical studies related to the DNA dynamics in presence of crowders. Some of the main area where people focus are the DNA-protein interaction and DNA melting [97–100]. All these studies predict that molecular crowding can increase the DNA duplex stability. The level of stabilization can be different depending on the representation of denatured state.

Harve *et al.* through atomistic molecular dynamics simulations, elucidated that molecular crowding stabilizes hydrogen-bonding between complementary nucleotides [88]. Y. Lui *et al.* [78] developed a thermodynamic model to predict DNA melting in ionic and crowded solutions. There findings predict an increase of 8°C in the melting temperatures dsDNA and DNARNA hybrids which are in good match with the available simulation and experimental results. In this model the base pairs are represented by two types of charged Lennard Jones spheres. It was observed that T_m increases in the presence of crowder is due to volume occupied by the crowder molecules which suppresses the entropy in the process of DNA melting. Their calculations showed that at a given concentration, a larger crowder exhibits a greater suppression of entropy which results in a higher T_m of DNA. In this model, the water molecule were treated as a background which implies that the water molecules do not occupy any volume in the system. This may lead to an inaccurate description of the molecular crowder in the solution. As a consequence, a revised model was proposed and the water molecules were represented explicitly as neutral spherical particles. The interactions between all species like dsDNA and ssDNA segments, water, crowders and ions were described by Lennard Jones and Coulombic potentials [101]. In the presence of crowders, the changes in the Gibbs energy, entropy and enthalpy predicted from the model. All these were found in good agreement with the experimental observations.

In a recent work of Brackley *et al.* [102], using coarse grained Brownian dynamics (BD) simulations reported the effect of molecular crowders on the protein-DNA target search process. They studied the effect of presence of crowders in cytosol

[1]Oligonucleotides are short DNA molecules, synthesized in the laboratory.

and along the DNA molecule. Their findings suggest that a proper account of the crowded cellular environment is crucial for a complete understanding of the protein-DNA target search. Matthias and Feng Gai [103] considered the melting of ubiquitin protein in the presence of the molecular crowders. They found a power law behaviour between the concentration of crowders and the melting temperature of the system. These theories predict that an inert molecular crowders destabilize the unfolded state of proteins via a repulsive depletion force that arises from the volume occupied by the molecular cosolutes. A theoretical method was presented by Sanbo and Zhou [97, 104] for modeling of the crowders at the atomic level in order to understand the protein folding and binding stability in presence of crowders. The method is rich enough to achieve exhaustive conformational sampling in modeling of crowding effects. It covers the challenges posed by large protein oligomers and by complex mixtures of crowders. In this model, instead of calculating the free energy differences between two end states in the absence and presence of crowders, they calculated the transfer of free energies of protein molecules from a dilute solution to a crowded solution. These simulations and atomistic modeling of protein folding stability, provide results consistent with experimental studies.

These studies reveal that the denaturation and folding process in a crowded environment is influenced by both entropic and enthalpic effects. However, DNA melting or denaturation in crowded solution is still far from complete understanding.

1.3.3 DNA having mismatch or bubble in the sequence:

Experimental studies:

With the recent advent of novel experimental techniques, it is possible now to obtain more detailed information on the microscopic configurations of the fluctuating DNA. The neutron scattering experiments on the melting transition of DNA explained the spatial correlation along the molecule through the melting transition [105]. The fluorescence correlation spectroscopy techniques is used to estimate the time scale of the opening and closing of loops of denaturated segments [106, 107]. These experiments provide the experimental investigations of the DNA breathing dynamics [106]. In the earlier experiments it was shown that the large amplitude breathing in DNA occurs well below the denaturation temperature [108, 109]. It initiates in the chain with the formation of local "denaturation bubbles". A single *mismatch or defect*[1] in the DNA sequence form nucleation sites for these bubbles

[1]the absence of H-bonding between the complementary bases called a *defect site*. This can

and reduces the overall stability of the duplex [110]. Bonincontro *el al.* studied the thermodynamic and dielectric properties of a short DNA molecule of mismatched sequences [111]. They measured melting temperature by evaluating ultraviolet absorption as a function of temperature and compared with those of the samples with defects in the sequence.

A Short sequence of DNA oligonucleotide has been studied by Bonnet *et al.* [112]. They measured the change in the melting temperature with one defect placed on different locations of the base-pairs. They found the melting temperature, $T_m = 315$ K for defect less chain and a decrease in T_m about 8 K with one defect in sequence. N. Nelson *et al.* measured the T_m of 13 different mismatched duplexes and developed a method for detecting all types of single-base mismatches in oligonucleotide sequence [113]. These experiments indicate a thermal destabilization of the DNA oligomers containing the defects or mismatch. The defect or mismatch in the sequence originates the local opening or bubble in the sequence and hence the decrement in melting temperature. Under physiological conditions, this local denaturation occurs spontaneously due to fluctuations, the *breathing* of DNA, which opens up few tens of base pairs (bubbles) like DNA opens in the initial stage of the transcription process. Choi *et al.* proposed that the formation of bubbles is directly related to the transcription sites. In particular, their results indicate that in the absence of proteins, the transcription start site could be predicted on the basis of the formation of bubbles of ten or more base pairs. Hence, the secret of the transcription initiation site is not in the protein that reads the code but really is an characteristics of DNA. They quote: *DNA directs its own transcription* [114]. The experimental observation of DNA breathing in real time is difficult in ensemble measurements due to the low frequency and short duration of base pair opening (20-100 μs) [115].

Theoretical studies:

The most important feature of the DNA denaturation is the cooperativity in the system which is a manifestation of long-range interactions along the chain. The breaking of the H-bonding between strands decreases the overlapping between the π-electron orbitals of the organic rings in the bases. This favors the unstacking of adjacent bases which results in the instability of the next hydrogen bond. With the continuation of this breaking, a bubble in the double helix is formed. This bubble makes the system more entropic which favors the complete denaturation of

be a mismatch or deletion of the bases on the complementary strand.

the molecule. Therefore DNA denaturation is a highly cooperative phenomenon and hence a single mismatch or defect in the sequence can lead the opening of the double helix. Several theoretical studies have proved the existence of this cooperative phenomenon [51, 116–119].

The effect of defects or mismatches on the melting profile of short heterogeneous DNA chains was calculated by N. Singh and Y. Singh [120]. They modified the on-site potential and stacking energy for defected site and found good agreement with the experiment results. A. Hanke and R. Metzler extensively used the PS model to establish a general framework for studying the bubble dynamics in dsDNA [39]. By calculating the bubble free energy using the PS model, they further used Fokker-Planck equation to describe the bubble dynamics, both, below and at the melting temperature of DNA. Alexandrov *et al.* [121] studied the bubble statistics and dynamics in dsDNA using Langevin dynamics. Monte Carlo simulation is also a good approach to study the DNA melting transitions. Using this approach in PBD model, bubble nucleation and cooperativity in DNA melting was studied by S. Ares it et al. [122]. The PBD model allows the local melting of Hydrogen bonding and formation of denaturation bubbles in the chain.

A mechanism connecting the local untwisting and opening of DNA double helix was proposed by T. Lipniacki [32]. S. Ares *et al.*, [122] studied the formation and role of bubble states in the pre-melting regime with no further parameters or fitting in the model. They accurately reproduced experimental findings. Moreover, T. S. van Erp *et al.* [123] and Z. Rapti *et al.* [124], gave a thorough analysis of the bubble-statistics as a function of position and bubble size for several DNA sequences using PBD model. N. Theodorakopoulos [125] shown that bubble statistics is very sensitive to the presence of nonlinear stacking interactions. Recently, Martin & Kevin [126] used Brownian dynamics to study the kinetics and thermodynamics of single-stranded DNA hairpins, gaining insights into the role of stem mismatches and the kinetics rates underlying the melting transition. Through these studies one can interprets thermal destabilization of mismatched DNA oligonucleotides. Mean bubble formation time in DNA denaturation was investigated by K P N Murthy and G. M. Schutz [127] using the Poland-Scheraga free energy of the bubble size in a double-stranded DNA. The investigations of bubble formation in circular DNA was studied by M. Zoli using the path integral technique [128].

The dynamics of defects is also important in the replication process. DNA replication is initiated at distinct sites called replication origins, where pairs of

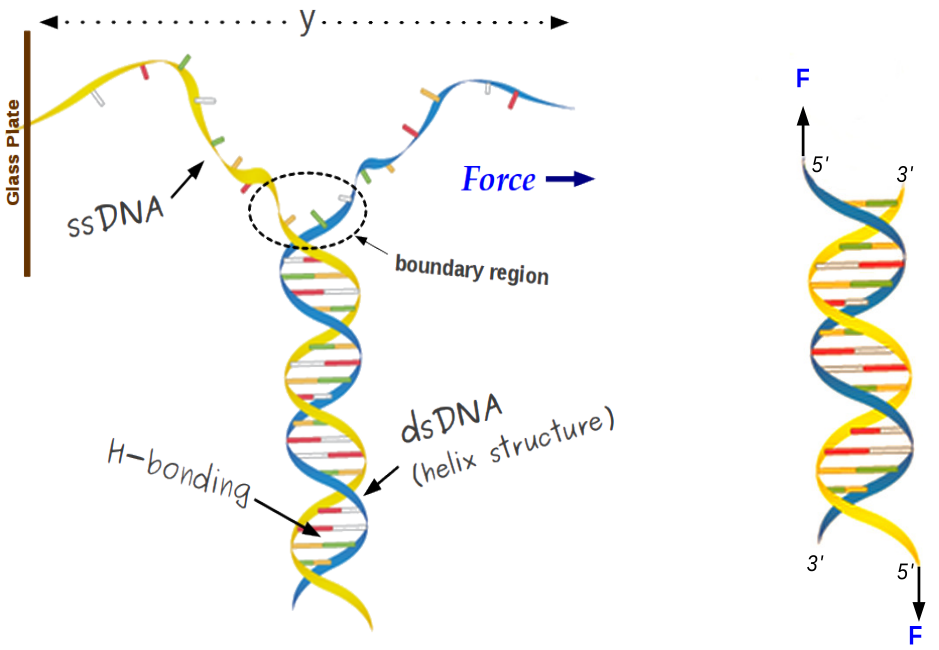
replication forks begin to duplicate DNA bidirectionally outward from the origin site. The defects or mismatch in the sequence may delay or stall the replication process. Gauthier *et al.* [129] introduced a rate-equation formalism to study DNA replication kinetics in the presence of defects. Through their proposed model they found a crossover between two regimes: a normal regime, where the influence of defects is local, and an initiation-limited regime, where the defects have a global impact on replication. As the density of defects increases the replication fork stalls lead to a crossover threshold between these two regimes. The role of the defects in the design of a molecular motor has been discussed by McCullagh *et al.* [130]. From a biological perspective, their results highlighted the remarkable ability of the DNA design to mitigate the propagation of damage, hence limiting the detrimental effects that this may have on healthy cell function. How the defects affect the melting, elastic, and other properties are problems of scientific interest.

1.4 Force induced melting of DNA

Force induced separation of dsDNA strands is more closer to the unzipping process occurring in living organisms[1], where enzymes or proteins attach to the molecule and pull it to originate local opening of the base pairs [2, 131]. During the replication and transcription processes, the DNA molecule is severely deformed: the double-helix is untwisted, stretched, compressed, and the base-pair patterns are locally destroyed. These reversible structural transitions in response to various external conditions are very important for the biological function of DNA, as it is a hereditary material [6, 131, 132]. *In vitro*, at relatively low temperature where thermal DNA melting does not occur, the two strands of DNA could be separated by applying an opposite force on the two strands of DNA molecule [133–135]. This phenomenon is called *force-induced DNA melting* or *DNA unzipping*. In the force induced melting the force can be applied either perpendicular to the helix axis or parallel to the helix axis. That is why the force induced DNA melting is a directional melting. In case when the force is applied perpendicular, it is commonly referred as *unzipping* while for the parallel applied force, it referred as *rupture* [136–142] (see Figure 1.10).

In the past two or three decades, the development of experimental techniques and Single Molecule Force Spectroscopy (SMFS) experiments enhanced our understanding in determining the functions and structures of DNA as well as other

[1]Double-stranded DNA is thermally stable at body temperature

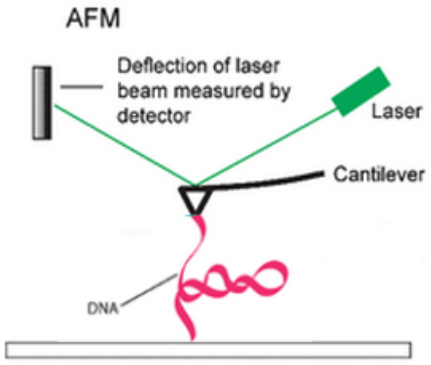


(a) Pulling force is perpendicular to helix-axis (b) Pulling force is along to helix-axis

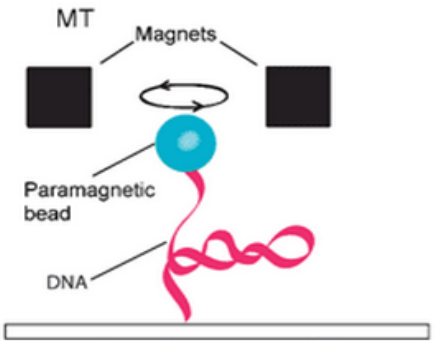
Figure 1.10: Mechanical unzipping of dsDNA, two strands of DNA are pulled in opposite direction. The strand separation for the unzipping (a) and the rupture (b) is completely different.

biomolecules [131, 143]. The main advantage of SMFS is its ability to detect molecular interactions which are responsible for the mechanical stability of their structures. Using SMFS one can separate out the fluctuations of individual base-pairs or molecules from the ensemble average behavior observed in traditional bulk experiments. There are number of techniques for manipulating single molecules like atomic force microscopy (AFM) [144] (range of forces between 10 and 10^4 pN), optical tweezers (OT) [145] (force range of 10^{-1} to 100 pN), magnetic tweezers (MT) [146] (force range between 10^{-2} to 100 pN), bio-membrane force-probe [147], and many other [131, 148]. A schematic representation of these experimental set-ups is shown in Figure 1.11.

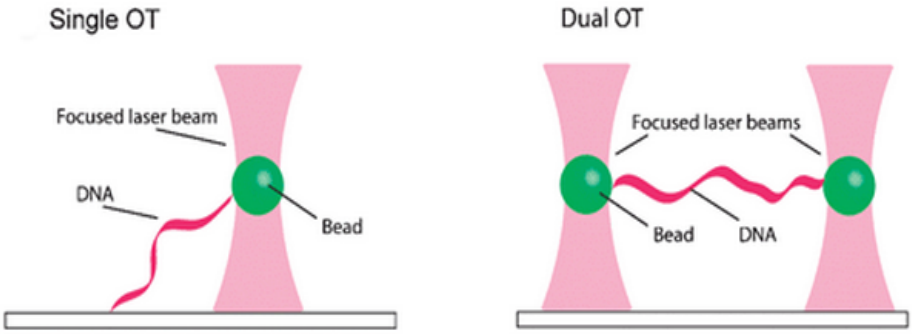
In all these techniques, a DNA molecule is attached to a surface on one end and to a force sensor on the other. The force sensor is usually a trapped micron sized bead or a cantilever. The displacement of the cantilever measures the force. In a typical unzipping experiment one of the strands is attached to a glass slide while the other strand is pulled as shown in Figure 1.10(a). There are two main possible ways to separate two strands. Either the strand is pulled with a constant velocity (constant extension ensemble) or with a constant force (constant force ensemble)



(a) Pulling DNA through AFM



(b) Pulling DNA through Magnetic Tweezers



(c) Pulling DNA through Optical Tweezers (OT)

Figure 1.11: Mechanical stretching of DNA molecule with different techniques, the strand of DNA is pulled to stretch it. Figures are taken from [149]

[131, 150, 151]. The microscopic features of unzipping may vary with the choice of ensemble. In thermodynamic limit these two ensembles are equivalent [152].

Constant Extension Ensemble (CEE)

In this ensemble, the separation between the end base pair of one of the ends in the dsDNA molecule is kept fixed and the average force needed to keep this separation is measured [133, 152, 153]. The experimental setups, optical tweezers [145] and atomic force microscope (AFM) [144], essentially control the position of the end bases where a force is applied. In AFM, the force is applied using a linear ramp protocol with very small velocity. In such cases the CEE is the appropriate ensemble because the changes that take place due to the small variation in the velocity are very small. When the displacement of the base pair is held constant, the force adjusts to compensate for the different average binding energies in AT-rich and GC-rich regions. Hence, one does not expect large jumps and metastable states in this case.

Constant Force Ensemble (CFE)

In constant force ensemble the force is applied on the end base pair of the strand at constant rate, keeping base pair on the opposite strand fixed. The distance between two bases in a pair on which force is applied, is allowed to fluctuate [134, 154, 155]. This approach is more easily modeled theoretically and may be more closely related to strand separation in cells. This ensemble is predominantly used in the study of the non equilibrium thermodynamics of small systems, where the average extension is taken as the control parameter [156]. Magnetic tweezers [146] provide a constant force ensemble in which constant force is applied at the end bases of the DNA molecule. For homogeneous DNA, the unzipping transition is expected to occur continuously and completely at a constant rate once the constant applied force exceeds the threshold for separating the single base pairs. On the other hand, for heterogeneous sequence, certain pauses and jumps are observed instead of continuous opening at a constant rate, that depend on the applied force as well as the base sequence [155]. These pauses and jumps in the opening of dsDNA correspond to the weaker and stronger sections of the chain.

It is important to know the range of forces encountered at the molecular level before studying the structural and dynamical properties of biomolecules. The smallest force on a molecule is due to the thermal agitation. It sets a lower limit on force measurements and above these forces there is the entropic forces that result from a reduction in the number of possible configurations of the system consisting of the DNA molecule. For example, DNA molecule in solution adopts a random coil configuration that maximizes its configurational entropy [148, 157]. When the DNA is stretched, the molecular entropy is reduced. The work done by this applied force can be calculated by measuring the change in the displacement of the molecule. The typical energies involved are of order of few k_BT and the typical lengths are of the order of a nanometer, hence the entropic forces are of order $k_BT/nm = 4$ pN [158]. The elastic forces are of order ~ 160 pN. These are typically the forces necessary to break ligand bonds or to deform the internal structure of the molecule. The strongest forces encountered at the molecular scale are those required to break a covalent bond of the order of ~ 1600 pN [148]. The range of these forces strongly depends on the chain length, sequence, temperature and the solution conditions, (*i.e.* salt, molecular crowding, etc.). Various groups have investigated these effects on DNA unzipping. In this section we review the experimental and the theoretical studies that are done to understand structural response of DNA due to applied force.

1.4.1 Role of salt concentration

Experimental studies:

The first measurements of the entropic elasticity of a single DNA molecule were reported by Smith *et al.* in 1992. They used a combination of magnetic fields and hydrodynamic drag to pull DNA molecule with a force F on small superparamagnetic beads attached to the molecule. The excellent agreement between the elastic theory of an ideal polymer and the measurements on DNA has provided a benchmark, calibration and framework for all single molecule work on DNA. It was found that in the low force regime (<10 pN) the elasticity of dsDNA is entropy dominated, where the molecule behaves as an ideal polymer of persistence length[1] about 50 to 100 base pairs. It's elastic behaviour is purely due to a reduction of it's entropy upon stretching. From 10 pN and up to about 60 pN, DNA stretches elastically as does any material. (*i.e.* obeys Hookes law). However, at force above 60

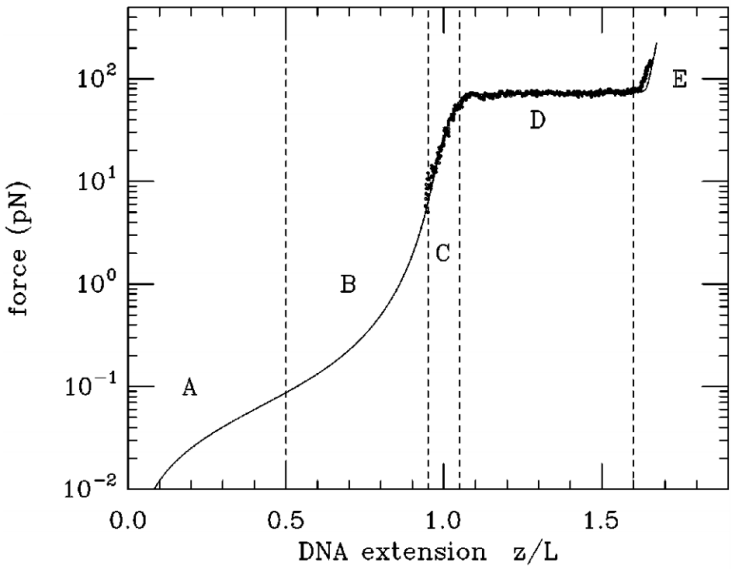


Figure 1.12: Force-Extension curve for dsDNA taken from [159] (A) For forces less than 0.1 pN, random-walk fluctuations are being suppressed. (B) For forces between 0.1 and 10 pN the small curvature fluctuations reduce the length of the DNA. (C) For forces between 10 and 60 pN, simple linear elasticity of DNA is observed. (D) Force at about 65 pN, a force plateau is observed, corresponding to coexisting regions of normal and overstretched DNA. (E) Force above 70 pN, DNA is entirely converted to its overstretched form.

pN, a surprising transition was discovered (see Figure 1.12), where DNA stretches to about 1.7 times to length and a phase transition occurs from the B-form to a stretched or S-form [141, 148, 158–160]. This transition is highly cooperative, *i.e.*

[1] the persistence length is a basic mechanical property quantifying the stiffness of a polymer.

a small change in force results in a large change in extension. An explanation of this transition is attributed to the short range nature of base pair stacking interactions. At higher forces the stacking potential can no longer stabilize the B-form configuration of dsDNA and the stacked helical pattern becomes distorted [131].

Experiments done by Bloomfield's group, predict that the DNA melts during the over-stretching transition [161, 162]. They proposed that this transition takes place in two stages. First it undergoes a cooperative transition in which an equilibrium melting process occurs containing a non-zero fraction of bound pairs. In the second stage the remaining base pairs are broken. Thus, it is the enthalpic energy rather than the free energy, that determines the stretching force. Hence the force required to completely separate the two strands is much greater than the over-stretching force. In a careful set of experiments, this group has shown that changes in solution conditions, i.e. salt concentration, pH, etc., which favor DNA denaturation also reduce the critical force needed to overstretch the molecule [162–165]. They found the persistence length of DNA is reduced in high salt concentrations by electrostatic screening of the repulsive charge along the backbone [166]. The elastic response of lambda phage DNA molecules was also probed using optical tweezers at concentrations of trivalent cations that provoked DNA condensation in bulk [167].

This mechanical unzipping of DNA using force was first observed by Bockelmann *et al.* [133, 153, 168]. They used λ-phage DNA and in this experiment one strand of DNA is attached to a glass slide and the other strand to a micro-needle. The other end of helix is capped with a hairpin molecule to avoid complete separation (see Figure 1.11). The deflection of the tip was measured on a video image as a function of displacement. They found the typical unzipping forces are in the range of 12-15 pN. Unzipping of dsDNA strongly depends on the ionic strength of the solution. The high salt conditions, favor the formation and stabilization of secondary structure of DNA. To understand the conformational behavior of a giant duplex DNA chain in a mixed solution with various biopolymers with different state of ionization, the structure of the DNA chain was analyzed in the presence of polycations, polyanions, and neutral polymers as a model for cellular environment [169]. Concentrated medium with neutral polymer induces the discrete folding transition of the DNA and the addition of small amounts of either the polycation or the polyanion causes structural changes in the folded DNA. The forces exerted by single stranded binding (SSB) proteins in maintaining the open regions of ssDNA have been measured directly by different groups [170, 171].

Recently Huguet *et al.* [172], experimented single molecule force unzipping in a wide variety of conditions, including salt concentration, pH, and temperature. They determined the 10 unique nearest neighbor base-pair free energies at different salt concentrations and found that AA/TT stacking energies are strongest while CC/GG are the weakest stacking energies. They also measured the unzipping forces for different chain lengths of DNA in a wide range of salt concentrations, 0.01 to 1 M Nacl. The unzipping force show a logarithmic increase with increasing salt concentration of the solution. More recently, the elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions are studied by B. Alessandro *et al.* [173]. For both monovalent and divalent salts, they found that the electrostatic contribution to the persistence length, is proportional to the Debye screening length and it varies as the inverse of the square root of cation concentration. The intrinsic persistence length is about 0.7 nm for both types of salts. The divalent cations are found to be more suitable for screening the electrostatic interactions compared with monovalent cations.

These experiments show direct analysis of paired nucleotide sequences and predict a more quantitative understanding of the interaction energies which stabilize the DNA double helix [131, 148, 170, 174, 175].

Theoretical studies:

The above discussed single molecule experiments have revealed the potential of the technique for the study of DNA unzipping that inspired to develop a theoretical understanding of the problem. The theoretical studies of DNA molecules under mechanical tension treat the transport of the applied stress in DNA, the role of base stacking and structures of stretched DNA in dilute as well as the concentrated solutions. M. Kosikov *et al.* [176] optimized the configurations of both poly(dA).poly(dT) and poly(dG).poly(dC) homopolymers under high and low salt conditions. In order to sample the conformational space as widely as possible, they adopted an implicit representation of the chemical environment with a distance dependent dielectric constant and atomic charges modified to mimic counterion condensation. They found the energetic and structural changes in high salt regime, which is thought to mimic the natural cellular environment of the double helix, persist as well under simulated low salt conditions. The computed behavior of poly(dG).poly(dC) and poly(dA).poly(dT) polymers were found identical in their findings. This is because they used the simplified nature of homopolymer model. Podgornik *et al.* [177] presented a theory which gave the stretching and bending moduli renormalization in the presence of salt. They introduced a

mesoscopic model of the self interacting polymer chain including the bending and stretching elasticity. At low ionic strength of the solution their results are found to in close agreement with the experiments. They found that not only the persistence length but also the stretching modulus depend on the salt present in the solution. A. Wynveen and C.N. Likos [178] using molecular dynamics simulations and theoretical calculations showed that increasing salt concentrations of the solution reduces the range of the forces between the molecular brushes at a given separation between the DNA chains. The theoretical approach was based on the two dimensional cylindrical cell model. Their findings were found to be consistent with the experimental studies on the salt dependent investigations of DNA grafted colloids.

Romano *et al.* [179] studied the DNA overstretching transition using recently developed coarse-grained model. They found that the overstretching at 23°C occurs at 74 pN, about 6-7 pN higher than the experimental value at equivalent salt conditions. The 6-7 pN overestimation is because of underestimating the extension at larger forces. Furthermore, they reproduced the temperature dependence of the overstretching force as well. Recently Snodin *et al.* [180] introduced an extended version of this coarse grained model to capture the thermodynamic, structural, and mechanical properties of single and double stranded DNA. This model can be used for a range of salt concentrations including those corresponding to physiological conditions (the previous model was parameterized to just one salt concentration).

1.4.2 DNA in a crowded solution

Experimental studies:

It is a well known fact that the molecular crowders can influence the stability, dynamics, and function of DNA molecule. This is believed that the nucleic acids, protein, lipids, saccharides etc. occupy about 20-40% of the total volume (∼50-400 g/l) [80]. This indicates that the volume excluded by these biomolecules and chemical interactions are critical for determining the structure and stability of DNA molecule. It is an interesting work to probe that how the molecular crowding affects the structure and stability of highly ordered DNA structures. This is believed that the highly crowded environment strongly promote the DNA self assembly processes. This leads to extremely condensed and thermodynamically stable DNA aggregates. The G-rich strand of human telomeric DNA can fold into a four stranded structure called *G-quadruplex*. Such structural transitions have been observed *in vivo*, as well as *in vitro* [86]. It is understood that the crowding particles

decrease the entropy of the folded or unfolded state as well and hence stabilize the ordered structure. The effects of molecular crowding on the mechanical stability of protein molecules studied by Yuan *et al.* [181]. It was found that the mechanical stability of ubiquitin molecules was enhanced by molecular crowding. However, there is a lack of reports on the mechanical stabilities of DNA molecules under macromolecularly crowded conditions, which prevents the interpretation of the biological relevance of these molecule structures. Nevertheless, together with new theoretical studies on the biomolecules under mechanical tension, single-molecule force spectroscopy continue to generate new insights into force and its relation to structure and functions of DNA in crowded environment.

Theoretical studies:

Over the last few years theoretical challenges have arisen in the area of unzipping of biopolymers in crowded environment. As the cells are having a very crowded environment because they are composed of many different kinds of biomolecules that may occupy a large fraction ($\sim 40\%$) of the total volume. In section 1.3, we noticed that crowding can influence the stability, dynamics, and functions of DNA molecules. However, the response of a polymer to a applied force remains an elusive problem. Very limited theoretical results investigating the effect of mechanical force on DNAs or proteins in a crowded environment are available.

By considering a linear polymer chain in a disordered environment[1], A. R. Singh *et al.* [182], studied effects of molecular crowding on stretching of polymers in poor solvent and derived force-extension curves which exhibit oscillations in extension. The disordered environment induces intermediate states because of the heterogeneity in the interaction and the transitions are discontinuous for finite chains. D. Thirumalai [85] demonstrated for the first time that the critical force for unzipping a biopolymer under tension obeys a nonlinear dependence on the volume fraction of crowding particles. Mechanical unfolding of DNA quadruplexes in concentrated solution has been studied computationally with atomic detail in a few works [183, 184]. Li *et al.* studied the unfolding of the parallel human telomeric DNA and estimate the unfolding free energy through Jarzynski equality [185]. They showed that the unfolding pathway depends on the pulling-force application site. Likewise, Yang *et al.* studied the unfolding of the DNA quadruplexes in presence of cations and showed that the largest contribution to the unfolding free energy comes from the removal of the central ion. Role of these central cations

[1] this disordered environment was meant to roughly represent molecular crowding as seen in cells

in the mechanical unfolding of DNA and RNA G-quadruplexes have been studied very recently by Bergues-Pupo *et al.* [186]. They use all-atom molecular dynamics simulations with explicit solvent to analyze the mechanical unfolding of representative intra-molecular G-quadruplex structures in the presence of stabilizing cations.

1.4.3 DNA having mismatch or bubble in the sequence:

Experimental studies:

Understanding the role of bubbles or loop entropy in the transcription or replication process is an important area where single molecule force studies have a vital role to play. The detailed analysis of these processes *in-vivo*, one can understand by analyzing it *in-vitro* via performing the experiments in presence of the environment similar to *in-vivo*. Several groups performed the experiments at different temperatures and proposed an experimental phase-diagram of DNA unzipping. One of the most interesting features of these experiments are the observation of certain "pauses and "jumps in the opening of dsDNA corresponding to the weaker and stronger sections of the chain. The earlier studies on DNA stretching and unzipping established the fact that the stretching and unzipping strongly depend on the base sequences and lengths.

Changhong *et al.* [187] presented and discussed the elasticity profiles of single-stranded poly(dA) and poly(dT) obtained by AFM-based single-molecule force spectroscopy. Their results showed that poly(dT) exhibits the expected entropic elasticity behavior, while poly(dA) displays two over-stretching transitions at ∼23 pN and ∼113 pN , which have not been previously observed. They proposed that these two force plateaus are due to breaking the stacking interactions between consecutive adenines (A). However, the molecular events that occur during the stretching of poly(dA) through the two plateau phases remain unclear and warrant further studies. In the other hand the unzipping of dsDNA strongly depends on the approach to unzip the DNA molecule. It has been suggested that the structure that results when dsDNA is pulled from the $3' - 3'$ ends differs from that which results when it is pulled from the $5' - 5'$ ends. The experiments by Danilowicz *et al.* [188] demonstrate that the overstretched states of $\lambda-$phage dsDNA show different properties, suggesting that they correspond to different structures. For $3' - 3'$ pulling versus $5' - 5'$ pulling, the following differences were observed: (i) the forces at which half of the molecules in the ensemble have made a complete force-induced transition to single stranded DNA are 141 ± 3 pN and 122 ± 4 pN,

respectively; (ii) the extension versus force curve for overstretched DNA has a marked change in slope at 127 ± 3 pN for $3' - 3'$ and 110 ± 3 pN for $5' - 5'$; (iii) the hysteresis in the extension versus force curves at 150 mM NaCl is 0.3 ± 0.8 pN μm for $3' - 3'$ versus 13 ± 8 pN for $5' - 5'$. These experiments support the role of the end entropy in the opening of dsDNA molecule performed either in force or distance ensemble.

Theoretical studies:

The formation of local denaturation zones or bubbles in double-stranded DNA is an important study in the field of conformational changes of biological macromolecules. Several groups investigated the role of loop and bubble entropies in unzipping of dsDNA molecule. In addition researchers also investigated the different approaches to unzip, *i.e.* the pulling or shear unzipping of the DNA molecule.

R Kapri [189] introduced the problem of a randomly forced DNA. One of the ways to elucidate such a problem is to apply random forces at different sites and study the phase diagram. The fluctuating force unzips the DNA by a gradual increase of bubble sizes. The obtained phase diagram resulted in important information about the existence of an Eye-phase (a phase consisting of configurations of eye-type conformations). D Giri [190] also studied the DNA unzipping when a force is applied to the interior of the chain. The effects of Eye-phase or bubbles in the double-stranded DNA and its role during the DNA unzipping were investigated in constant distance (or extension) ensemble. The purpose of this study was to provide the exact results of a semimicroscopic model of short chains by incorporating the directional nature of hydrogen bonds. They proposed a method to study the effect of molecular interactions right at the individual base pair level and their role on unzipping profile [190]. The thermodynamics and finite size behaviour of DNA unzipping are investigated by Kapri [191]. They investigated the temperature force phase diagram in two different ensembles, CFE and CEE.

N Singh [192] obtained force elongation behavior for short DNA molecule by changing it's sequence in the CEE for two different temperatures. These results suggest that the DNA sequence has a very specific effect on the unzipping of a dsDNA molecule and can therefore be used to find this sequence by determining the force extension curve in the CEE. This group has also studied the effects of defects[1] or mismatch in the sequence on unzipping of dsDNA. The peak forces in CEE were investigated in presence of these defects in the sequence including

[1] the absence of H-bonding between the complementary bases called a *defect site*

the one being stretched [152]. Many interesting features of DNA overstretching and unzipping transition using fully atomistic molecular dynamics simulations was shown by Maity *et al.* [140]. They showed that the conformational changes of double-strand DNA during the stretching of the strands along or perpendicular to the helix axis, can be viewed as force induced DNA melting. Their findings predict that the melting force in the case of unzipping is smaller compared to the melting force when the DNA is pulled along the helical axis.

The opening profiles for mechanical stretched DNA with constrained imposed on the ends of DNA were studied by Sikha *et al.* [193] using PBD model and by A R Singh [194] using mutually attracting-self-avoiding walks. The opening of a ds-DNA depends not only on the sequence but also on the constraints on the chain in the experimental setups. Their results clearly demonstrate that the force-induced melting of dsDNA, whose one of the ends is constrained, is significantly different from the thermal melting. Whereas more recently, using Langevin dynamics simulations R. K. Mishra *et al.* [142] showed that the rupture force also depends on the ends where the shearing force is applied (*i.e.* force on 3'-3'-end or 5'-5'-ends of the strands). In the series of the DNA rupture dynamics, recently Nath *et al.* [195] studied the DNA rupture event of the same base sequence studied in the experiments by Hatch *et al.* [196] For a short chain, they found that the rupture force increases linearly and saturates for a longer chain, which is consistent with the experiment results. Through the analytical calculations this saturation of rupture force in verified via a DNA ladder model [142]. M. Mosayebi *et al.* [197] introduced the time scale of observation in the critical shearing force required to rupture a duplex. They argued that rupture must be understood as an activated process, where the duplex state is metastable and the strands will separate in a finite time that depends on the duplex length and the force applied.

The model we are using throughout the thesis is the PBD model, proposed by M Peyrard and A Bishop in 1989 [77]. It was modeled as a first attempt to model DNA transcription, but found to be successful in investigating the melting transition in DNA sequences. Later on this model showed its strength in explaining the force induced unzipping of DNA. In Chapter 2, we have discussed this model in detail.

1.5 Existing research gap

The experimental and theoretical investigations on various aspects of DNA denaturation or DNA melting and various interactions have been done. These studies enable us to understand various interactions and mechanisms involved in the process. However, the complexity of the problem offers challenge to understand more. This motivates us to make continuous efforts to study different aspects of transition from double stranded to single stranded configuration. The prime objective of the thesis is to fill the gap to an extent, in existing investigations. Some of the research gaps which are addressed in this thesis. are listed here. .

- Researchers from several groups have investigated thermal melting of DNA in ionic solutions and proposed some empirical formulas to predict melting temperature, T_m at different salt concentrations of the solution. However, the role of salt concentration on the mechanical unzipping or stretching is still not very clear. In most previous studies the salt effects or the ionic nature of the solvents were assumed to have a constant value. Some of the investigations discussed the role of salts in the stability of dsDNA however, their approaches were primarily focused on the thermal stability of the dsDNA molecule. The understanding of DNA unzipping due to an applied force in an ionic solution is important in order to get a precise idea of the physical processes. In past few years, some experimental groups have investigated the stability of DNA molecule in ionic solution. Nevertheless there is inadequacy in theoretical investigations, that can predict the critical force at low as well as high salt concentrations.

- Interestingly in most of the previous studies (experimental, theoretical and numerical simulation), the effect of the cellular environment has also been ignored particularly in the area of DNA unzipping in crowded environment. Some recent studies have shown that crowding can influence the stability, dynamics, and functions of DNA molecules. The results obtained by different groups are quite interesting which shows that melting temperature of the DNA chain increases about $2 - 20^{\circ}$C in crowded environment. However, the response of a polymer to a applied force remains an elusive problem. There have been very few theoretical studies investigating the effect of mechanical force on DNAs in a crowded environment, which prevents the interpretation of the biological relevance of DNA molecule structures.

- How the defects, mismatches or bubbles in a sequence affect the melting, elastic, and other properties are problems of intense scientific interest in past few years. These defects or mismatches create the bubble in the sequence that will lead the DNA breathing during DNA helix-coil transitions. The experimental observation of DNA breathing in real time is difficult in ensemble measurements due to the low frequency and short duration of base pair opening (20-100 μs). Also, for long sized bubbles, there is a possibility of forming hairpin like structure in the bubble part of DNA chain. Several groups have investigated the role of bubble formation and bubble size in DNA melting. However, there are still open questions, like; (a) how the defect location in sequence affects the opening of DNA molecule, (b) what role the loop entropy plays in opening of DNA molecule, (c) is there a critical length which saturates the critical force as DNA length increases?

We expect future research on above discussed issues that will elucidate the response of DNA molecule to the applied force in presence of salt, in presence of crowders and in presence of defects in the sequence.

1.6 Objectives of the present work

Objectives of the present work are the following.
To investigate:

1. the phase diagram of mechanically stretched DNA in ionic solution.
2. role of cations on thermal as well as mechanical stability of dsDNA molecule.
3. influence of defects or mismatches on DNA melting and DNA unzipping of short DNA molecules.
4. the DNA denaturation in crowded environment.
5. how the applied force propagates in DNA chain, when the force is applied on different sites in the chain?

In order to achieve above objectives, we use the Peyrard-Bishop Dauxois model (PBD). It was modeled as a first attempt to model DNA transcription, but have had good success in its predictions of melting temperature and transition curves of DNA sequences. It seems reasonable that the PBD-model could serve as a first step towards description of the transcription and replication of DNA, as well. In Chapter 2, this model is discussed in detail.

Chapter 2

Peyrard Bishop Dauxois (PBD) Model

In 1989, Michel Peyrard and Alan Bishop [77] proposed a statistical model of DNA to explain the strand separation in the molecule at the base pair level. An objective of the model was to explain thermal denaturation of DNA molecule. The model successfully explain the thermal denaturation of long as well as short chains. With some modifications this model explain many features of force induced unzipping as well as the stability as a function of salt content of the solution. In many aspects the model is similar to the Ising model for ferromagnet as it assumes the interaction at the nearest neighbor level and consider the states of a base pair as open or close. Initially the model treated the nearest neighbor coupling as a harmonic spring which was later modified to incorporate the anharmonicity and then it is popularly known as PBD model[1], Various groups have analyzed the statistical mechanics of the DNA denaturation transition using a transfer integral technique [47, 193], path integral method [198, 199] and have calculated the inter-strand separation as a function of the temperature. Numerical results by different groups verify that the denaturation temperature is in good agreement with the experimental findings [47, 75, 200–204]. The PBD model allows the local melting of hydrogen bonding and formation of denaturation bubbles in the chain. Later several groups used the PBD model to study the DNA unzipping process as well [47, 193, 205–208].

[1]This model is known as PBD model after introducing anharmonicity by T. Dauxois [76]

2.1 Basic assumptions in the PBD-model

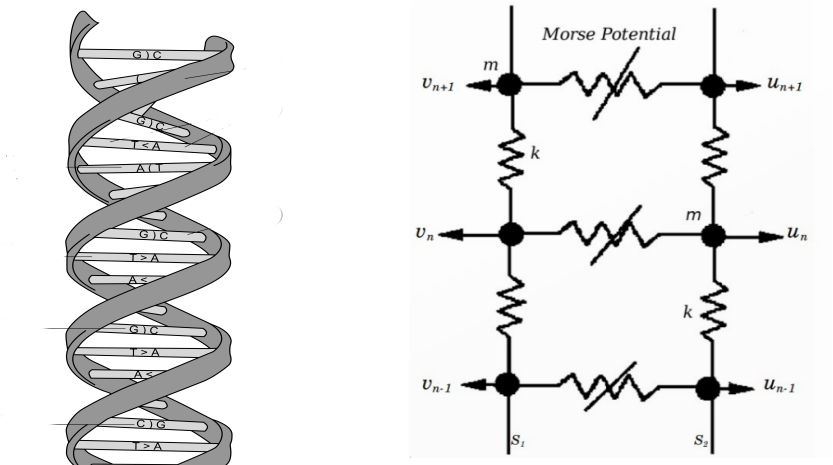
The objective of the PBD model is to understand the phenomenon of thermal denaturation of dsDNA and to get the dynamics of molecule at the base pair level. The model works in center of mass frame and ignore the three dimensional motion of base pairs. In a sense it is a quasi-one dimensional model. The basic features of the model are:

- The model initially considered DNA molecule as a homogeneous[1] system. However, the model later on proved to be useful for the heterogeneous sequence too.
- The model neglects any kind of winding and knotting in the DNA. The base pairs in the sequence were assumed to be in a plane. Later on the model was modified to incorporate the helicoidal geometry [19, 209, 210].
- A base pair interacts with its nearest neighbors only through a harmonic potential. However the harmonic coupling was later on modified [76, 211].
- The hydrogen bonding between two nitrogeneous bases on opposite strands is represented by Morse potential which represents not only the attraction between the bases in a pair but also the repulsion between the two phosphate groups on the opposite strands [212].
- The longitudinal displacements of bases are neglected, as their amplitude of vibrations is much smaller than the amplitude of vibration of transverse motion.

2.2 Model Hamiltonian

The motions of the base pairs in the PBD model is shown in Figure 2.1(b). For each base pair two degrees of freedom is assumed. For the n^{th} base pair these are denoted by u_n for one chain and v_n for the second chain which correspond to the displacement of bases from their mean position. The two bases in a pair are connected through hydrogen bond and in the model it is represented by a Morse potential. In addition to this the base pair along the strands are coupled through

[1] homogeneous in the sense that all nucleotides have equal mass



(a) DNA as double helix

(b) The ladder structure of DNA that shows the motions of bases and pairs.

Figure 2.1: Graphical presentation of PBD model assumptions. DNA double helix assumed as a long straight ladder in which nucleotides (filled circle) are placed in each intersection and transverse motions are only considered.

a harmonic (which was later modified to anharmonic) potential. The complete Hamiltonian of the system is [77, 213, 214],

$$H = \sum_n \left[\frac{1}{2} m \{ \dot{u}_n^2 + \dot{v}_n^2 \} + \frac{1}{2} k \left(u_n - u_{n+1} \right)^2 + \frac{1}{2} k \left(v_n - v_{n+1} \right)^2 + V(u_n, v_n) \right] \quad (2.1)$$

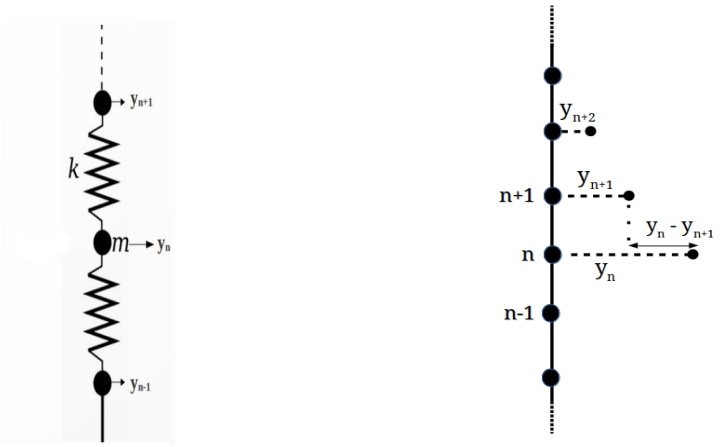
Here m is the reduced mass of the base pair and k is the elasticity of the DNA strand. In the model, the sequence heterogeneity was ignored and a common mass m and elastic constant k along each strand[1] was considered. The first term in the Hamiltonian represents the kinetic energy of the base pair while the next two terms represent the stacking potential which arises due to coupling between two base pairs along the strand. The last term in the Hamiltonian represents the hydrogen bonding between the bases in a pair and is represented by Morse potential. The motions which are responsible for motion of base pairs are either *in-phase* or *out-of-phase* motions. Mathematically they are:

$$x_n = \frac{u_n + v_n}{\sqrt{2}} \quad ; \quad y_n = \frac{u_n - v_n}{\sqrt{2}} \quad (2.2)$$

and the Hamiltonian can be written as,

$$H = \sum_n \left[\{ \frac{1}{2} m \dot{x}_n^2 + \frac{k}{2} \left(x_n - x_{n+1} \right)^2 \} + \{ \frac{1}{2} m \dot{y}_n^2 + \frac{k}{2} \left(y_n - y_{n+1} \right)^2 \} + V(y_n) \} \right] \quad (2.3)$$

[1] sequence-dependent stacking term has been introduced in model by Alexandrov *et al.* [215]



(a) (b)

Figure 2.2: (a) PBD model can be viewed as a model of a one-dimensional monotonic lattice with each atom having mass m and nearest neighbor interaction W. (b) displacements of nucleotides from the equilibrium positions.

Out of these, the motion that is responsible for the hydrogen bond stretching is the *out-of-phase* motion. The *in-phase* motion contributes to the potential energy of the system and has nothing to do for the breaking of hydrogen bonds. Thus, the status of each base pair is described by a scalar variable y_n, which represents the transverse stretching of the hydrogen bonds connecting the two bases in a pair see Figure 2.2. Thus, the on-site potential V, is function of variable y_n only. The part of the Hamiltonian which depends on the variable x_n is decoupled from stretching part and corresponds merely to a harmonic chain without on-site potential and hence can be ignored while studying the opening of the DNA helix. Thus, the effective Hamiltonian is written as,

$$H = \sum_{n=1}^{N} \left[\frac{p_n^2}{2m} + V(y_n) \right] + \sum_{n=1}^{N-1} \left[W(y_n, y_{n+1}) \right] \tag{2.4}$$

where

$$V(y_n) = D(e^{-ay_n} - 1)^2 \quad \& \quad W(y_n, y_{n+1}) = \frac{k}{2}(y_n - y_{n+1})^2 \tag{2.5}$$

Here y_n is the displacements of the atoms from equilibrium, Figure 2.2(b). D measures the depth of the potential and a controls the width of the potential (the smaller a is, the larger the well), Figure 2.3. It includes a strong repulsive part for $y < 0$, corresponding to the steric hindrance and it has a minimum at the equilibrium position $y = 0$. It becomes flat for large y, giving a force between the bases that tends to vanish, as expected when the bases are very far apart; that

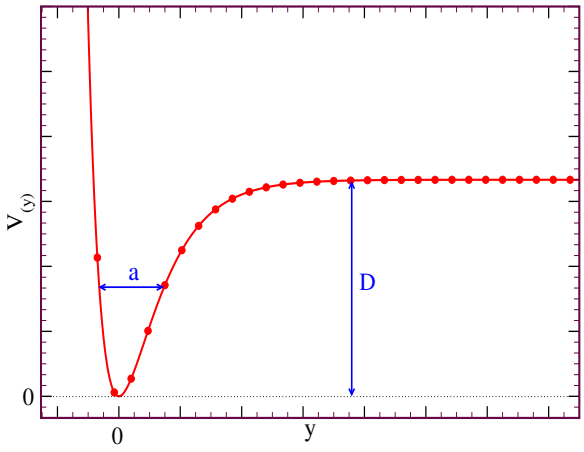


Figure 2.3: The shape of on-site potential for the base pair interaction of the strands of DNA. D is the depth of the potential and a is width of the potential.

confirms complete dissociation of the base pair. In a homogeneous sequence of DNA, D is taken to be site independent, but in heterogeneous DNA the value of D depends on whether the n^{th} base pair is AT or CG. Any other potentials, like Lennard-Jones potential, can be used to represent the on-site potential in DNA, but Morse potential is found to be a good representation of the energy as a function of end atom distance when the bonds are hydrogen bonds [216]. The stacking potential is a key factor in the PBD model. It represents the stacking interactions that are contributed by dipole-dipole interactions and in water, the solvent-solvent induced hydrophobic interactions. The resultant of these is a complex interaction pattern between overlapping base pairs. Later on Dauxois *et al.* [76, 211] modified this potential and a nonlinear term was introduced. This nonlinear term takes care of the cooperativity of melting. Since stacking energy is not the property of individual bases only but is due to the pairing between the bases. When the hydrogen bonds between the bases stretched or break, the electronic configuration of the bases modifies which reduces the coupling between the neighboring bases. The following anharmonic potential model mimics the essential features of the stacking energy

$$W(y_n, y_{n+1}) = \frac{k}{2}(y_n - y_{n+1})^2[1 + \rho e^{-b(y_n + y_{n+1})}] \tag{2.6}$$

where k, the force constant is related to the stiffness of a single strand and the second term is the anharmonic term. The range of anharmonicity is defined by the parameters ρ and b. As we see in Eq.2.6, for the zipped state, $y_n = 0$, force constant is equal to $k(1 + \rho)$. The stretching of the two interacting base pairs

decreases this force constant from $k(1+\rho)$ to k. A decrease in the force constant in the unzipped state provides large entropy and hence favors unzipping either at high force or high temperature. The difference in the force constants between the zipped and the unzipped state of base pairs, create an energy barrier, which depends on the parameters ρ and b.

Thus, the complete Hamiltonian of the system is,

$$H = \sum_{n=1}^{N}\left[\frac{p_n^2}{2m} + D(e^{-ay_n}-1)^2\right] + \sum_{n=1}^{N-1}\left[\frac{k}{2}(y_n - y_{n+1})^2[1+\rho e^{-b(y_n+y_{n+1})}]\right] \quad (2.7)$$

The Hamiltonian has five parameters ρ, k, b, a and D,. The potential depth D and the inverse of the potential width a are the site dependent parameters and rest are the site independent. The model has been successfully applied to study the thermal denaturation and the unzipping of dsDNA. Though, PBD model ignores the helical structure of dsDNA, it is found to have enough details to analyze the behavior at a few Å scale relevant to molecular-biological events [217].

2.3 Partition Function

Using the model Hamiltonian of the Eq.2.4, the structural transition and associated properties of the system can be investigated. The Hamiltonian can be used to calculate the canonical partition function as,

$$Z = \int \prod_{n=1}^{N} \{dy_n dp_n \exp(-\beta H)\} = Z_p Z_c, \quad (2.8)$$

where Z_p corresponds to the momentum part of the partition function while the Z_c contributes as the configurational part of the partition function. N is the number of base pairs in the chain and $\beta = \frac{1}{k_B T}$. Since the momentum part is decoupled in the integration, it can be integrated out as a simple Gaussian integral, that will give the familiar kinetic-energy factor for N particles:

$$Z_p = (2\pi m k_B T)^{N/2}, \quad (2.9)$$

here m is the reduced mass of the base pair and k_B is Boltzmann constant. The configurational partition function, Z_c, is

$$Z_c = \int_{-\infty}^{\infty}\left[\prod_{n=1}^{N-1} dy_n \exp\{-\beta[W(y_n, y_{n+1}) + V(y_n)]\}\right] dy_N V(y_N), \quad (2.10)$$

where y_n is the dislocation of n^{th} base pair from equilibrium. As the configurational part of the partition function has coupled terms, the calculations are little tricky. To solve the equation for Z_c one can define a Kernel $K(x, y)$, [202] as,

$$K(y_n, y_{n+1}) = \exp\left[-\beta H(y_n, y_{n+1})\right] \tag{2.11}$$

For the homogeneous chain, we have $K(x, y) = K(y, x)$. Thus, the partition can be written as,

$$Z_c = \int_{-\infty}^{\infty} \prod_{n=1}^{N} dy_n K(y_n, y_{n+1}) \tag{2.12}$$

The equation for partition function can be solved by introducing the integral equation

$$\int dy K(x, y)\phi(y) = \lambda\phi(x) \tag{2.13}$$

because of symmetry and the fact that $K(x, y) > 0$, if one assumes

$$\|K(x, y)\| = \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{K(x, y)\}^2\right]^{1/2} < \infty$$

the integral equation contains positive Hilber Schmidt type kernel and will be having positive eigenvalues and orthonormal eigenfunctions. If the eigenvalues are denoted by λ_i and eigenfunctions as ϕ_i, the orthonormality condition demands that,

$$\int dx\phi_n(x)\phi_m(x) = \delta_{nm} \qquad \text{and} \qquad \sum_{n=1}^{\infty}\phi_n(x)\phi_n(y) = \delta(x - y). \tag{2.14}$$

In terms of these orthonormal eigenfunctions, the kernel $K(x, y)$ can be expanded as [218]

$$K(x, y) = \sum_n \lambda_n\phi_n(x)\phi_n(y) \tag{2.15}$$

Substituting the expression of $K(x, y)$ in terms into Eq.2.12 and applying the orthonormality conditions (Eq.2.14), the configurational partition function will be,

$$Z_c = \sum_{n=1}^{\infty} \lambda_n^N \tag{2.16}$$

The above expression is valid for the periodic boundary condition. For the open boundary conditions, we have,

$$Z_c = \sum_n \left(\int dy \phi_n(y) \exp\left\{ -\beta \frac{V(y)}{2} \right\} \right)^2 \lambda_n^{N-1} \tag{2.17}$$

The eigenvalues and corresponding eigenstates can be found by diagonalizing the resulting matrices that appear in Eq.2.11

However the calculations of partition function is not straight forward once we consider the heterogeneity in the sequence of a DNA chain. In this case the n^{th} site is not of the same nature as $n-1$ and $n+1$ sites. Different approaches have been proposed and used by different groups to overcome this problem. Y Zhang *et al.* [202] proposed extended transfer matrix approach (ETMA) where the different kernels can be expanded in terms of same or different bases. Another approach was proposed by Cule [219] where the heterogeneity was introduced by quench disorder. For a chain with open boundaries the configurational part of the partition function can be evaluated numerically with the help of matrix multiplication method [47, 193, 200].

The important part in the integration of the configurational space is the selection of proper cut-offs for the integral appearing in Eq.2.10. A larger cut-off implies that the probability of the molecule to be completely denatured increases for finite chain. This probability becomes 1 in the limit of infinite cut-off. This is related to the fact that the integral of Eq.2.10 is divergent in nature. To avoid this divergent nature of the partition function in numerical calculations a double-stranded ensemble was introduced by T S Van Erp *et al.* [123]. In a detailed work they showed that the upper cut-off can be ≈ 144 Å for a set of model parameters at $T = 300$ K while the lower cut-off is -0.4 Å.

We are using the matrix multiplication method to calculate the partition function by defining different kernels corresponds to the sequence of the chain. The integration in the Eq.2.10 can be written as:

$$Z_c = \int_{-\infty}^{\infty} \exp\left(-\beta \frac{V(y_1)}{2} \right) dy_1 \prod_{n=1}^{N-1} dy_n \exp\left[-\frac{\beta}{2} \{ V(y_n) + V(y_{n+1}) + 2W(y_n, y_{n+1}) \} \right] \exp\left(-\beta \frac{V(y_N)}{2} \right) dy_N \tag{2.18}$$

The accuracy of these approaches is limited first by discretization errors in the integrations and by the need to numerically evaluate integrals over an infinite

domain. This can be partially lifted by a finite size scaling analysis, which involves a properly controlled approach to infinity. Integration methods, such as the Gauss-Legendre quadrature [220] which select appropriate absissa for the evaluation of the function according to the number of points involved in the calculation, are useful to integrate over a large domain with a reasonable number of points [221]. We choose the order of the Gauss-Legendre node to be equal to the number of the points used to discretize the integration (this will be the dimension of the matrix appeared in the Eq.2.11). In our calculations, we found that in order to get a precise value of melting temperature (T_m) one has to choose the large grid points. The dimension of 900×900 was found to be suitable for your analysis.

After a discretization of the coordinate variables and the introduction of a proper cutoff on the values of the y_n, the task is reduced to discretized the space to evaluate the integral numerically.

The calculated partition functions, Z_p (Eq.2.9) and Z_c (Eq.2.10), can be used to calculate the other thermodynamic quantities of interest by evaluating the Helmholtz free energy of the system. The free energy per base pair is,

$$f(T) = -\frac{1}{2}k_B T \ln\left(2\pi m k_B T\right) - \frac{k_B T}{N} \ln Z_c. \tag{2.19}$$

The other thermodynamic quantities like entropy S, specific heat C_v are evaluated using the following relations,

$$S(T) = -\frac{\partial f}{\partial T} \tag{2.20}$$

and

$$C_v(T) = -T\frac{\partial^2 f}{\partial T^2}. \tag{2.21}$$

2.4 Limitations of PBD model

The PBD model is found to have enough details to explain many quantitative description of DNA despite its simplified character. Most importantly, it allows one to study the statistical and dynamical behavior of very long as well as short homogeneous as well as heterogeneous DNA sequences. However, the model can not be said as complete as it lacks many structural and dynamical features of DNA molecule. The model does not account the three dimensional structure of the molecule and its role in the structural transition of the molecule [222]. Also, the microscopic partition function like in Eq.2.10, is in principle impaired because it does not take into account that an open state of the base pair. The open state

(or base) will have enormous number of micro-states because of the flexibility of base and its interaction with water molecules and ions. The water and the solution conditions are considered as the background effect [34]. The effect of hydration of bases in the open state is pretty obvious and it leads to the activation barrier for base pairs opening and closing. There were attempts in the literature to incorporate such a barrier into the PBD model [47, 207, 223, 224].

Chapter 3

Phase diagram of DNA in concentrated solution

3.1 Introduction

In this chapter we discuss the stability of double stranded DNA (dsDNA) in a concentrated solution. The concentrated solution means the salt present in the solution is not constant now. This is known that the stability of the dsDNA molecule is primarily due to the hydrogen bonding between the bases of the complimentary strand. In most of the previous studies, the hydrogen bond interaction and the effects of surroundings, such as salt concentration of the solution, are taken as constant [76, 225]. However, the DNA molecules are strong polyelectrolytes, having negatively charged phosphate groups, this would be interesting to analyze the role of cations/salt (Na^+ or Mg^{2+}), in the melting or unzipping profile of DNA. To stabilize a DNA molecule one should require a counter-ion (positive ion) to screen the repulsion between the two negative strands of the DNA molecule. In the buffer solution we have the cations like sodium or magnesium ions which are derived from the salt present in the solution. The concentration of these ions contribute not only to the stability of the molecule but also play an important role in the folding kinetics of the molecule [226, 227].

Several groups [170–172, 228] have experimentally measured the force required to destabilize the dsDNA as a function of concentration of salt in the solution. To understand the mechanism theoretically, the counter-ion condensation model (based on the two state ion distribution) [46, 229], the Poisson-Boltzmann model (based on mean field calculations) [230, 231] have been used. Recently, the tight bonding approximation (TBA) [70, 71], Poland-Scheraga (PS) [232, 233], Coarse-

Grained Model [234–238] and Peyrard Bishop Dauxious (PBD) [75] models are also used to study the helix coil transition in these molecules. Most of these studies focused primarily on the thermal stability of the dsDNA molecule as a function of salt concentration of the solution.

We have investigated the effect of salt present in the solution, on the thermal melting and on forced induced unzipping of a heterogeneous dsDNA molecule using PBD model [76].

3.2 Modified Model Hamiltonian

Experimental observations predict that the melting temperature of dsDNA scales logarithmically with the salt present in the solution [49, 239]. Since the two strands of the dsDNA are negatively charged, the presence of cations like sodium or magnesium ions, shield the repulsion between the phosphate groups of each strands. Therefore, at higher concentration system prefers to be in less entropic state, more thermal or mechanical energy will be required to break the hydrogen bonds. In addition to this, the melting temperature has been found to have a linear dependence on the value of potential depth. In the PBD model, the stability in hydrogen bond is represented by the depth of Morse potential, D_n. Thus, this parameter should be a function of salt concentration of the solution. Keeping these factors in the background, we modify the potential depth D_0 in PBD model as,

$$D_n = D_0 \left[1 + \lambda_1 \ln\left(\frac{C}{C_0}\right) - \lambda_2 \ln^2\left(\frac{C}{C_0}\right)\right] \tag{3.1}$$

Here, the concentration, C is expressed in moles per liter and C_0 is the reference concentration chosen to be 1 mole/liter. The λ_1 and λ_2 appearing in the potential are solution constants [75, 240, 241].

An additional term in the Hamiltonian is the solvent term which simulates the formation of hydrogen bonds with the solvent, once the hydrogen bonds are stretched by more than their equilibrium values. We adopts the solvent term from the refs. [241–243].

$$V_{sol}(y_n) = -\frac{1}{4} D_n \left[\tanh\left(\gamma y_n\right) - 1\right] \tag{3.2}$$

The "tanh"term in the potential enhances the energy of the equilibrium configuration and the height of the barrier below which the base pair is closed. The small barrier basically determines the threshold stretching of hydrogen bond about which a base pair may be temporarily broken, re-bonded and then fully broken.

Of course, this come to the broken state at a length greater than $\sim$ 2Å. As solvent role is to stabilize the denatured state, this form of potential can be a good choice. The term, γ is the solvent interaction factor and it reduces the height of the barrier appears in the potential [210, 242–244]. The effective potential in model Hamiltonian will be

$$V_{\text{eff}} = V_M + V_{\text{sol}} \tag{3.3}$$

Where V_M is Morse potential given by Eq.2.4. We tune various values of γ from 0.1 to 1.0 and plotted the effective potential, as shown in Figure 3.1. We found

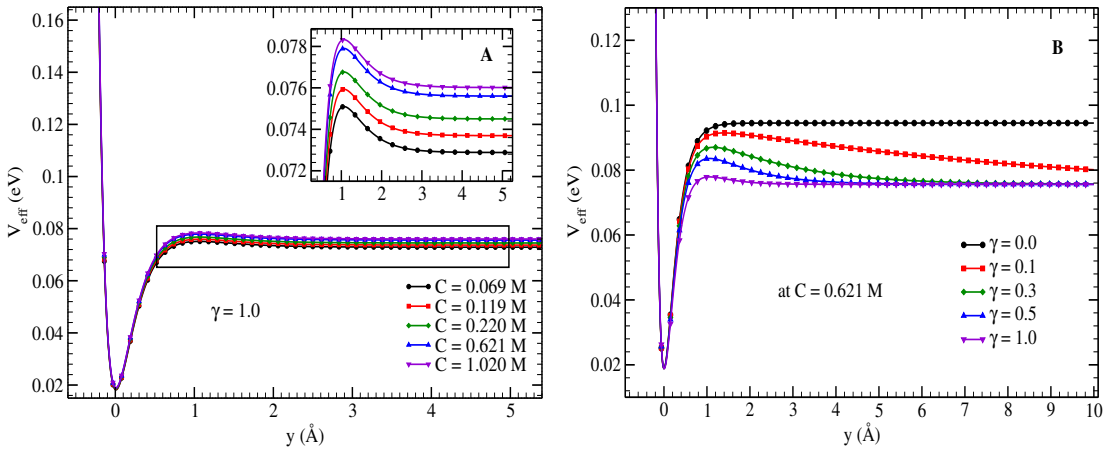


Figure 3.1: Plot of effective potential as a function of base pair stretching (in Å). In figure A, the effect of increase in the salt concentration of the solution on the potential depth is shown. The variation in the barrier height with the increase in the solvent interaction factor γ is shown in figure B.

that for larger values of γ, unzipping transition is more favorable. As the broken state occurs $\sim$ 2.0 Å, the value of γ should be chosen which reflect the breaking around 2.0 Å. We found $\gamma = 1.0$ Å^{-1} as a suitable choice for our calculations. Thermodynamics of the transition can be investigated by evaluating the expression for the partition functions, given in Eq.2.9 and Eq.2.10. The calculated partition functions, Z_p and Z_c , are used to calculate the other thermodynamic quantities of interest by evaluating the Helmholtz free energy of the system, Eq.2.19.

In *vitro*, the fraction of open pairs, ϕ as a function of temperature is monitored by using various spectroscopic techniques. Theoretically, we can calculate ϕ, using the partition function. For long chains, when the fraction ϕ of open base pairs goes practically from 0 to 1 at the melting transition, the two strands are not yet completely separated. At this point, the great majority of the bonds is disrupted and the dsDNA has denaturated, but the few bonds still remain intact, prevent the two strands going apart from each other. The real separation occurs only

at high temperatures. For very long chains, the double strand is always a single macromolecule, and hence one need to calculate the fraction of intact or broken base pairs only. The details of process of separation of short chains are different from that of the long dsDNA chains. In case of short chains, the end entropy contributes significantly in addition to the loop entropies. Hence the breaking of few bonds as well as the strand separation happens to be in a very narrow range of temperature. Thus average fraction of bonded base pairs, $\theta(=1-\phi)$ is defined as [33, 120, 200],

$$\theta = \theta_{\text{ext}}\theta_{\text{int}} \tag{3.4}$$

θ_{ext} is the average fraction of strands forming duplexes, while θ_{int} is the average fraction of unbroken bonds in the duplexes. The dissociation equilibrium can be neglected in the case of long chains and thus θ goes to zero while θ_{ext} is still practically 1. To compute θ_{int}, one has to separate the configurations describing a double strand on the one hand, and dissociated single strand on the other [200]. Since a bond between two bases is said to be broken if their separation is greater than the average separation between two bases. Therefore, n^{th} bond is considered to be broken if the value of y_n is larger than a cutoff y_0. One can therefore define θ_{int} as:

$$\theta_{\text{int}} = \frac{1}{N}\sum_{n=1}^{N}\langle\vartheta(y_0 - y_n)\rangle \tag{3.5}$$

where $\vartheta(y)$ is Heaviside step function and the canonical average $\langle . \rangle$ is defined considering only the double strand configurations. As discussed above, when θ goes practically from 1 to 0 at the melting, the two strands may not get completely separated, while for short chains, the single bond disruption and strand dissociation occur in a very narrow range of temperature. Therefore, one need to compute both θ_{int} and θ_{ext}. For θ_{ext} we use the expression [120, 200] :

$$\theta_{\text{ext}} = 1 + \delta - \sqrt{\delta^2 + 2\delta} \tag{3.6}$$

where

$$\delta = \frac{Z(s_1)Z(s_2)}{2N_0 Z(ds)} \tag{3.7}$$

here $Z(s_1)$, $Z(s_2)$ and $Z(ds)$ are the configurational isothermal isobaric partition functions of systems consisting of molecular species single strand s_1, single strand s_2 and the double strand configuration dsDNA respectively. The partition function for each species can be factored into internal and external part and can be written

as,

$$\frac{Z(s_1)Z(s_2)}{2N_0Z(ds)} = \frac{Z_{int}(s_1)Z_{int}(s_2)}{a_{av}Z_{int}(ds)} \frac{a_{av}Z_{ext}(s_1)Z_{ext}(s_2)}{2N_0Z_{ext}(ds)} \tag{3.8}$$

where $a_{av.} = \sqrt{a_{AT}a_{GC}}$. In analogy to what has been proposed for the Ising model on the basis of partition function of rigid molecules, one makes the following choice [200]

$$\frac{a_{av.}Z_{ext}(s_1)Z_{ext}(s_2)}{2N_0Z_{ext}(ds)} = \frac{n^*}{n_0}N^{-p\theta_{\rm int}+q} \tag{3.9}$$

and then,

$$\delta = \frac{Z_{int}(s_1)Z_{int}(s_2)}{Z_{int}(ds)} \frac{n^*}{a_{av}n_0}N^{-p\theta_{\rm int}+q} \tag{3.10}$$

where n^* is a chosen reference concentration as $1\mu M$ while n_0 is the single strand concentration which we have chosen as 3.1 μM. p and q are the parameters which can be calculated using experimental results [120, 200]. We have taken the parameters p and q, as p=29.49 and q=27.69. Using this equation we calculate the $\theta_{\rm ext}$ and hence the fraction of open pairs, ϕ as a function of temperature.

3.3 Temperature induced transition

In this section we investigate the role of salt concentration on the thermal stability of the dsDNA molecule. For most of the thermal denaturation studies this effect has been ignored. Here we extend the previous studies on thermal denaturation of dsDNA using modified PBD model and reproduce the experimental findings. Numerous experimental studies have investigated DNA melting. For example, the T_m of 92 different DNA sequences were measured over a wide ionic concentration range [49]. We choose three different DNA chains of short length out of these DNA sequences. and the sequence vary in terms of the fraction of GC & AT base pairs. We call them as 30% GC, 50% GC and 75% GC chains. The chain sequences are,
(a) 5'-TGATTCTACCTATGTGATTT-3' (30% GC)
(b) 5'-TACTTCCAGTGCTCAGCGTA-3' (50% GC)
(c) 5'-GTGGTGGGCCGTGCGCTCTG-3' (75% GC)
We adjust the model parameters to match our calculated results with the experimental results by [49]. It is found that the stiffness of the chain plays a crucial role in the melting or denaturation of the dsDNA molecule, in addition to the bond energy. While bond energy is represented by Morse potential, the strand elasticity is represented by stiffness parameter k in the anharmonic stacking term. Thus along with the value of D_0, we check various values of k and ρ to get close match

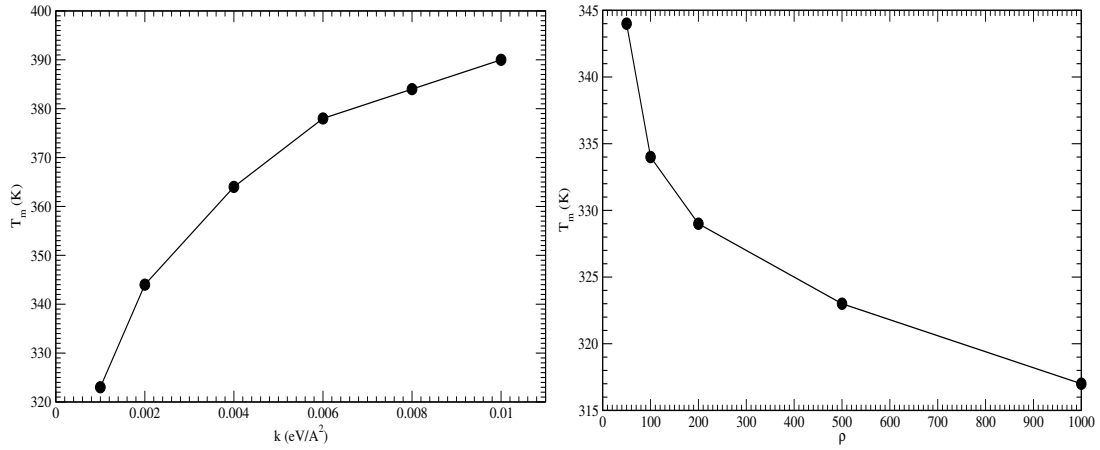


Figure 3.2: The figure in left showing effect of anharmonicity, ρ on the melting temperature of a DNA chain, effect of chain stiffness, k on melting temperature T_m in figure on right [245]

the experimental results. We found the values of $\rho = 1.0$ and $k = 0.01$ eV/Å^2 as suitable choice for the current investigation.

To adjust the melting temperature within the range of the experimental observations, we finally tune the values of D_0, keeping other parameters same for all the three chains. With the value of D_0 (for AT base pair) as 90.3, 89.1 and 87.5 meV for 30% GC, 50% GC and 75% GC chains respectively, we found close match with the experimental results for all the concentrations. The values of D_0 for GC base pair is 1.5 times of the value for AT base pair. In the present investigation the stiffness parameters is considered as site independent, however, elasticity of the strand is suppose to depend on the distribution of different bases along the strand[1] [69, 239]. The list of all the model parameters are given in Table 3.1,

We calculate the free energy and the specific heat per base pair of the system using Eq.2.19 and Eq.2.21. At the temperature when the system gets the sufficient amount of energy that is needed for transition from double stranded configuration to single stranded configuration the free energy shows a kink. For better visualization, we show the transition through the specific heat per base pair as a function of temperature. At the transition point, this is shown by a peak. In order to avoid the overflow we show the curve in Figure 3.3 for only one value of concentration (0.621 M) for all the three chains. Although the nature of all the three curves are different, they show the transition from double stranded configuration to single stranded configuration of DNA molecule. For short chains, the breaking of hydrogen bonds and the strand dissociation occurs in the same temperature range, thus,

[1]The effect of the stacking heterogeneities are discussed in next chapter.

Table 3.1: The complete set of model parameters.

parameters	**values**
inverse of potential width, a	4.2 (AT) & 6.2 (GC)
chain stiffness, k	0.01 eV/Å^2
anharmonicity, ρ	1.0
anharmonic range, b	0.35 Å^{-1}
solution constant, λ	$\lambda_1 = 0.01$ & $\lambda_2 = 0.002$
solution constant, γ	1.0
opening cutoff, y_0	2.0 Å

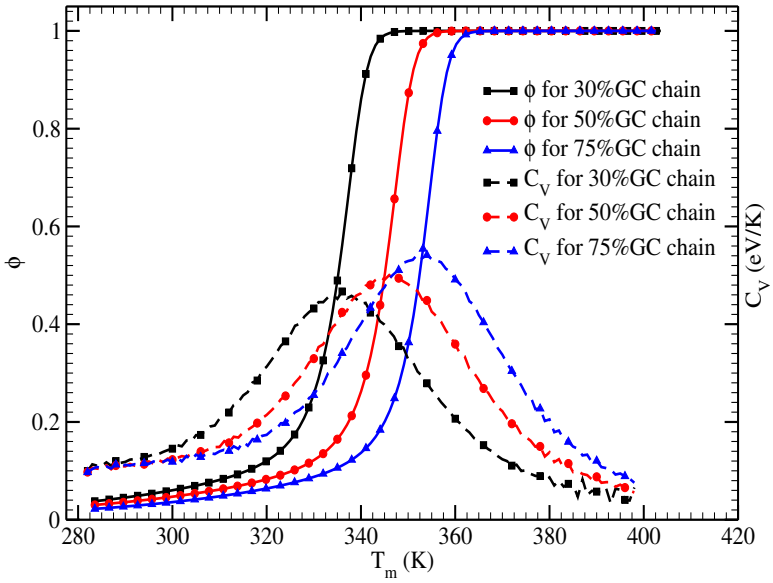


Figure 3.3: Specific heat, C_v and fraction of open pairs, ϕ plotted as a function of temperature for salt concentration of 0.621 M for all the three chains. As these are short chains, we calculate $\theta = \theta_{\text{ext}}\theta_{\text{int}}$. The value of C_v is scaled to show that the peak position and 50% of the open pairs meet at the same point

one has to calculate the θ_{ext} as well as θ_{int} [200]. The chain is said to be denatured when 50% of the base pairs are in open state. The temperature corresponding to $\theta = 0.5$ is same as we get from the calculation of specific heat (Figure 3.3).

We further obtained the value of melting temperature, T_m, for all the five concentrations 0.069 M, 0.119 M, 0.220 M, 0.621 M & 1.02 M, for which the experimental results are available. The results are shown in Figure 3.4. The phase diagram in Figure 3.4, shows the variation in the melting temperature as a function of salt concentration of the solution. It has been observed, experimentally as well

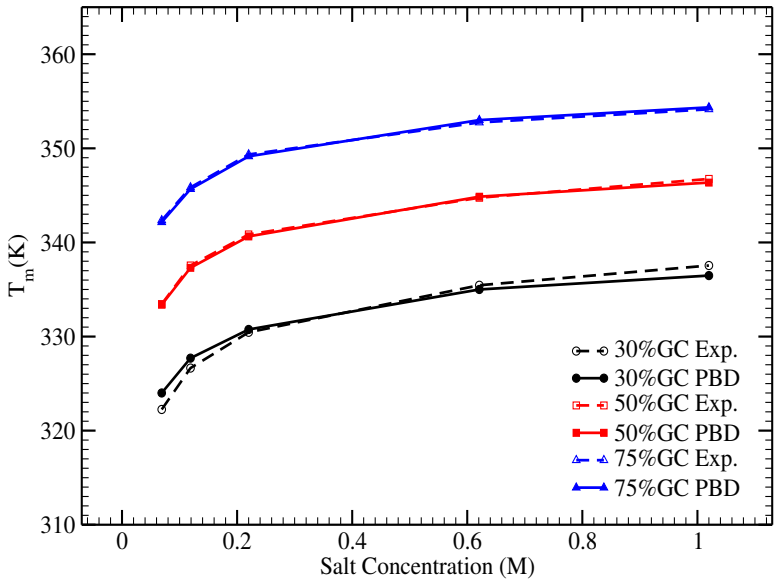


Figure 3.4: Temperature-salt phase diagram showing the variation in T_m as a function of salt concentration for all the three chains. We compare the results with the experimental findings [49].

as theoretically, that the melting temperature has a logarithmic dependence on the concentration of Na^+ in the solution. The value of potential depth, D, is tuned to get the proper match with the experimental data. The results shown are close to the experiments, however at low value of salt concentration our data points have slight deviation ($\sim$ 2-3 K) from the experimental data. We observe that the deviation from the experimental data at low concentration, is most for weaker chain (30% GC). At the lower concentration the melting of AT & GC pairs differs significantly [239] and the stiffness parameter may play an important role.

3.4 Force induced transition

In this section we investigate the role of salt concentration on the mechanical unzipping of dsDNA molecule. We have discussed in detail, the forced induced unzipping in Chapter 1.4. In *vitro*, the double stranded DNA is pulled mechanically, keeping other end fixed. These experiments are performed either at constant displacement of end pairs or at varying loading rates. Although both the set-ups give the same critical force for an infinite chain, the microscopic and dynamic behavior of unzipping of dsDNA in the two ensembles are different. The force induced unzipping is studied in both constant extension ensemble (CEE) and constant force ensemble (CFE). We found that the results are independent of choice

of the ensembles and are in good agreement with the experimental results.

3.4.1 Constant force ensemble

We calculate the unzipping force in the constant force ensemble by pulling one end of the DNA molecule keeping the other end fix [154, 246–248]. We take the same sequence of 20 base pairs which we considered for thermal studies, however we repeat them to a length which can be considered as an infinite chain. We found that a length of about 600 base pairs is sufficient to be considered as an infinite chain, see Figure 3.5. This length may be model dependent and one may get

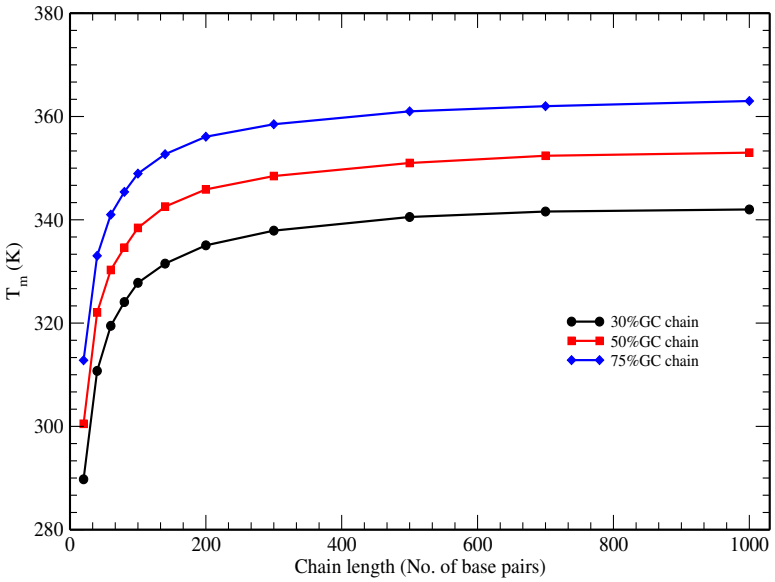


Figure 3.5: melting temperature as function of chain length for different GC content in the chain. We take the value of potential depth as $D_{AT} = 0.076$ eV for all the three chains.

different number of base pairs in a chain that can be considered as infinite chain. The modified Hamiltonian of the system is as:

$$H_f = \sum_n \left[\frac{p_n^2}{2m} + W(y_n, y_{n+1}) + V(y_n) \right] - F \cdot y_i \tag{3.11}$$

Where F is the applied force at the i^{th} base pair in the chain. y_n is the separation of the bases from the equilibrium. We have included a term $F \cdot y_i$, which is basically the external energy in the system. We apply the force at the end base pair of the chain. The model parameters are taken same as in the previous section, however, we tune the value of potential depth D_0 in such a way that we get closer to the experimental results [172] for 50% GC chain. We take the value of

potential depth as $D_{AT} = 0.076$ eV for all the three chains. As the force induced unzipping experiments are performed at room temperature (∼300 K), the melting temperature T_m of the system should be much higher than 300 K, in order to ensure that at 300 K, base-pair opening is not due to thermal fluctuations. With these model parameters, T_m is approximate 350 K for 50% GC chain. We calculate

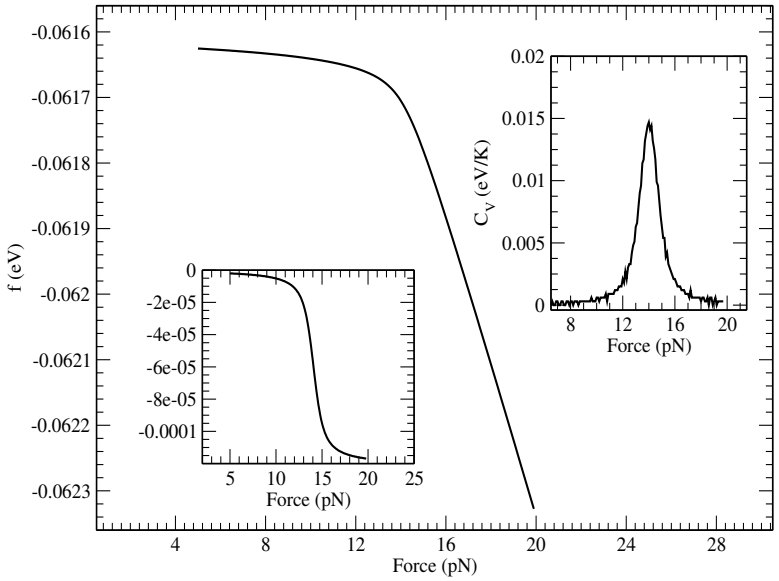


Figure 3.6: Free energy as a function of the applied force on the end of 30%GC infinite chain for the salt concentration of 0.621 M. The first derivative, Entropy and the second derivative, Specific heat C_v are also plotted in insets.

the free energy of the system as a function of applied force. The force is applied on the 3′-end of the chains at temperature 300 K. At the critical force we observe a kink in the free energy which gives rise to the peak in the specific heat at constant force as shown in Figure 3.6. This is the point where the system transform from close state to open state. We obtain the value of critical force for all the three chains by locating the peak in specific heat as a function of applied force. We also calculate the average number of open pairs in a chain as a function of the applied force through Eq.3.5. As the calculation for force induced unzipping is for infinite chain, we calculate the $\phi = 1 - \theta = 1 - \theta_{\rm int}$. From the Figure 3.7, it is clear that the $\phi = 0.5$ at the same value of force as predicted by specific heat as a function of applied force. We calculate the critical force for different salt concentration varying from 30 mM to 1.02 M. The phase diagram in Figure 3.8, shows the two phases in case when the dsDNA is forced to unzip mechanically as a function of salt concentration. We compare our results with the experimental phase diagram [172]. The results reported here are in good match with the experimental results

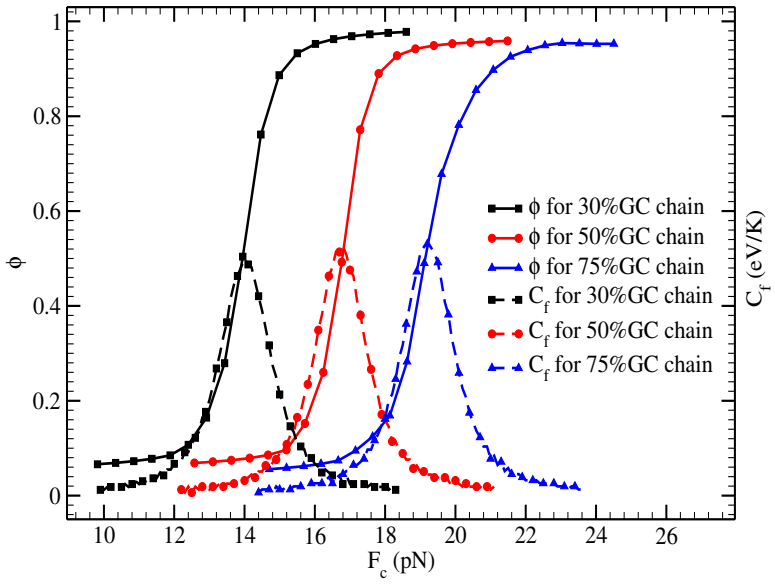


Figure 3.7: The average number of open pairs as a function of the applied force on the chain for the salt concentration of 0.621 M. As this is for an infinite chain, the $\theta \approx \theta_{\rm int}$. Again value of specific heat is scaled to get the two curves on the same plot.

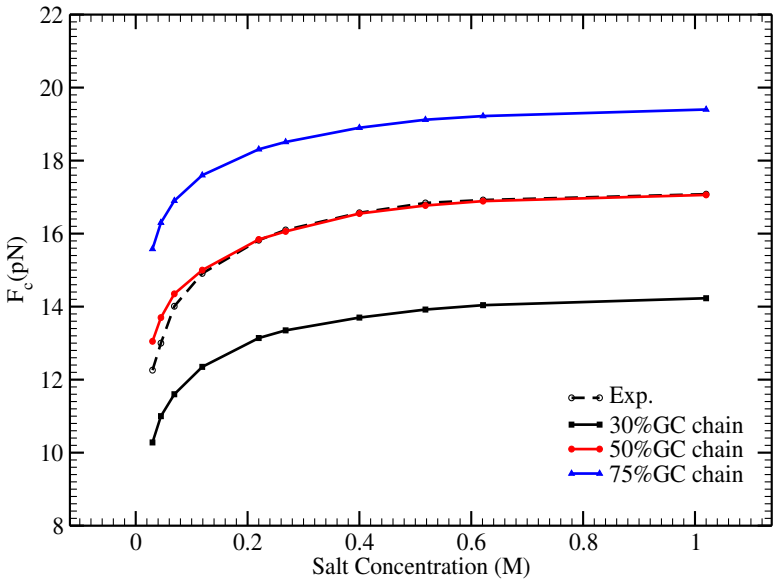


Figure 3.8: The phase diagram showing the dependence of critical force on the salt concentration of the solution. For comparison we show the experimental results obtained by [172], in the diagram.

except for low concentrations, where the slight deviation has been observed ($\sim$ 0.3-0.7 pN). As discuss earlier the stacking interaction contributes significantly at lower salt concentration during forced induced unzipping of DNA chain.

3.4.2 Constant extension ensemble

In this ensemble, the separation between the end base pair of one of the ends in the dsDNA molecule is kept fixed. Average force needed to keep this separation is measured [133, 153] by evaluating the change in free energy in displacing the first (end) base pair from a position to its next. Let us consider a chain of N base pairs of which the 1st base pair is stretched to a distance y and the N^{th} base pair is fixed to zero separation, $y_N = 0$. The configurational part, Z_c of the partition function with open boundary condition, defined as [192]:

$$Z_c(y) = \int \prod_{n=1}^{N-1} dy_n \delta(y_1 - y) \exp\left[-\frac{\beta}{2}\left\{V(y_n) + V(y_{n+1}) + 2W(y_n, y_{n+1})\right\}\right] \quad (3.12)$$

The work done in stretching the base pair to y distance apart is

$$w(y) = \frac{1}{2}V_1(y) - k_B T\left[\ln Z_c(y) - \ln Z_c\right] \quad (3.13)$$

The first term, $\frac{V(y)}{2}$ in above equation is residue of the on-site potential while constructing the transfer integral that connects the base pair 1 with the base pair 2, only $\frac{V(y)}{2}$ has been taken into account for the stretched base pair[1]. The derivative of $w(y)$ with respect to y gives the average force $F(y)$ that is needed to keep the extension equal to y. We plot the value of $F(y)$ as a function of extension y at temperature 300 K with salt concentration of 0.621 M in Figure 3.9. We observe that at beginning a very high force require to initiate the unzipping of the chain, this force is known as peak force. We calculate the peak force as well as the critical force require to unzip the chain. In all the cases we apply the force on the 3′-end of the chain.

In CEE, we investigate the effect of salt concentration on the peak force in addition to the critical force. We calculated F_C and F_P at different salt concentrations. The variation in peak force with salt concentration shown in Figure 3.10. The dependence of peak force on the salt concentration is in the same manner as for the critical force[2]. From Figure 3.11 We found that both the ensembles give the same critical force for an infinite chain, although the microscopic and dynamic behavior of unzipping of dsDNA in these ensembles are different. We observe that the variation in critical force with salt concentration is equivalent in

[1] this is because we are employing with open boundary conditions, when the periodic boundary condition is imposed there is no end term appears.

[2] The peak force for 75%GC is more than the 30%GC and 50%GC, it is because of the end base pair is GC in 75%GC chain.

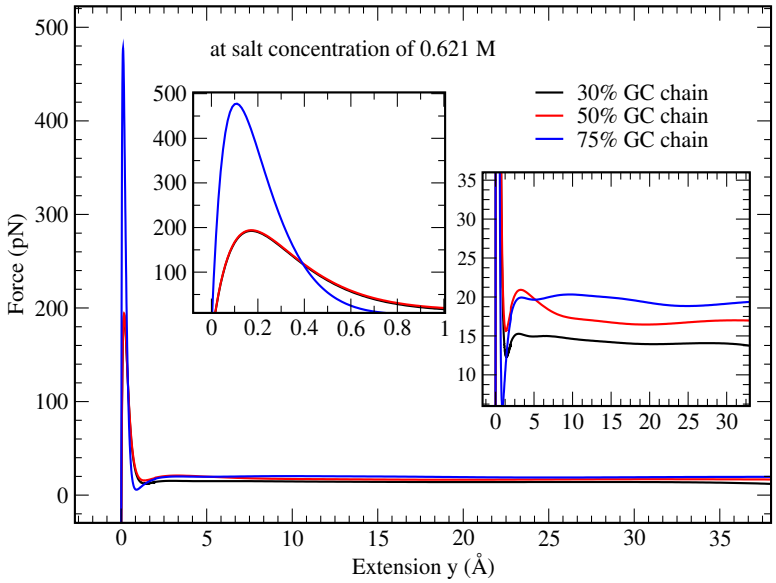


Figure 3.9: $F(y)$ as a function of extension y at temperature 300K with salt concentration of 0.621 M. Figure shows peak force and the critical force for 30%, 50% and 75% GC chain.

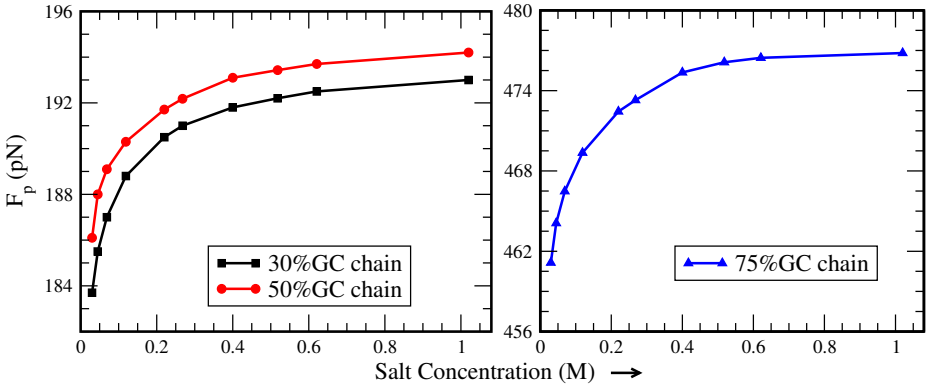


Figure 3.10: The graph showing the variation in peak force F_p as a function of salt concentration of the solution

both the ensembles and having good match with the experimental values [150].

3.5 Conclusions

We have investigated the role of salt concentration on the thermal as well as on the mechanical unzipping behavior of heterogeneous dsDNA molecule. The PBD model is modified to incorporate the salt as well as the solvent effect of

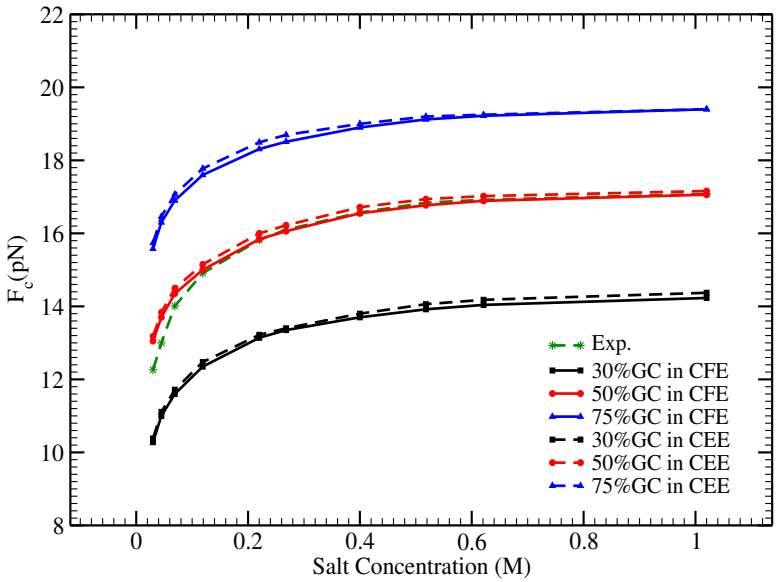


Figure 3.11: The graph showing the variation in critical force F_C as a function of salt concentration of the solution in both ensembles, Constant force and constant extension ensemble.

the system. Our results indicate a close match with the earlier observations on thermal denaturation of dsDNA, which shows that the melting temperature varies non-linearly or logarithmically with the salt concentration of the solution. In accordance with Manning's counter-ion condensation theory [46], this is due to the layer of condensed counter-ions on the DNA surface that neutralizes the phosphate charges. This decreases the inter-strand electrostatic repulsion and the overall stability of DNA molecule increases and hence system need more thermal energy to break or denature. The melting temperature is found to vary with the salt concentration as well as with the GC content of the chain.

We have investigated the role of salt present in the solution on the mechanical unzipping behavior of the dsDNA molecule. The mechanical unzipping of dsDNA molecule have been studied in two different force ensembles, Constant force and Constant extension ensemble. We found that the critical force (force needed to completely unzipped the molecule) increases with the increase in the salt concentration of the solution. The addition of salt in the solution basically shields the repulsion between the phosphate groups in the dsDNA chain and the interaction of hydrogen bond increases. This simply means that system needs more force to to unzip the dsDNA molecule. Our results are found to be in close agreement with the experimental phase diagram [172] in both the ensembles. Although in both the ensembles the microscopic and dynamic behavior of unzipping of dsDNA are

different. The peak force F_p in CEE also increases logarithmically as a function of salt concentration of the solution. In phase diagram there is some deviation at the lower salt concentration it may be because of the stacking heterogeneity, which has been taken as constant in the current investigation, might be the key factor at low concentration.

We conclude that the PBD model, although a quasi one dimensional model in nature, can be a good choice to investigate the presence of salt in the solution and its effect not only on the thermal denaturation of dsDNA chain, but also on the mechanical unzipping of the chain.

Chapter 4

Effect of salt concentration on the stability of heterogeneous DNA

4.1 Introduction

This chapter is an extension of the Chapter 3, to investigate the effect of solution environment on the stability of a heterogeneous DNA molecule. The double helical structure of DNA molecule is stable due to the hydrogen bonding between the bases on the complimentary strands as well as due to the stacking interaction between the bases along the strand [8]. The double stranded state of these molecules strongly depends on the base pair composition, the sequence, and the ionic nature of the solution. In Chapter 3, we have studied that the melting temperature and the critical force increase non-linearly with the total salt concentration of the solution. Some semi-empirical calculations show that the thermal stability of the DNA molecule depends on the sequence heterogeneity of the molecule [69]. However not much attention has been paid to understand the role of cations to the stability of heterogeneous DNA molecule that is stretched by a force. As the bonding nature of $A-T$ is different to the $G-C$, the response of any change in the salt concentration may be different for these base pairs, it would be interesting to study the different solution constants for AT and GC base pairs and its effect on the transition from double stranded to single stranded state. In this chapter, we study the solvent effect and the sequence heterogeneities which concerns the salt concentration of the solution. We describe the modified model Hamiltonian which deals with the sequence heterogeneities in terms of salt concentrations.

4.2 Modified Model Hamiltonian

For the current investigation we use PBD model [76] which is a Hamiltonian based model as discussed in Chapter 2. The model considers the stretching between corresponding bases only. The hydrogen bonding between the two bases in the n^{th} pair is represented by the Morse potential and this on-site potential is modified to incorporate the solution effect in Chapter 3.2 by the Eq.3.3. The heterogeneity in the base pair sequence is taken care by the values of D_n and a_n.

It is found that the stiffness of the chain plays a crucial role in the melting or unzipping of the dsDNA molecule, in addition to the bond energy of the bases. The strand elasticities represented by stiffness parameter k in the anharmonic stacking term as given in Eq.2.6. The stacking interaction $W_S(y_n, y_{n+1})$ is independent of the nature of the bases at site n and $n+1$ as these parameters are assumed to be independent of sequence heterogeneity. However, the sequence heterogeneity has effect on the stacking interaction along the strand. This can be taken care through the single strand elasticity parameter k as sequence dependent.

As the bonding nature of $A-T$ is different to the $G-C$, the response of any change in the salt concentration may be different for these base pairs, it would be interesting to study the variable λ^n for AT and GC base pairs and its effect on the transition from double stranded to single stranded state. The modified potential depth which incorporate the salt effect given by:

$$D_n = D_0 \left[1 + \lambda_1^n \ln\left(\frac{C}{C_0}\right) - \lambda_2^n \ln^2\left(\frac{C}{C_0}\right)\right] \tag{4.1}$$

here, the solution constants λ_1 and λ_2 are site dependent. This is how we are introducing the sequence heterogeneities for solvent effect. Thermodynamics of the transition can be investigated by evaluating the partition function.

4.3 Temperature induced transition

The hydrogen bonding between bases of complementary strands and the stacking between neighboring bases stabilize the double-stranded structure of DNA molecule. These base stacking interactions are of the order of magnitude of a few k_BT. Thermal fluctuations can cause the disruption of base pair bonds which ultimately leads to the unzipping of double helix and hence, this opening of bases is entropic in nature. Therefore the stiffness of the chain contributes a lot to the melting or denaturation of the dsDNA molecule [249]. In this work, we consider

the stiffness parameter, k, as well as the solution constants, λ_1 & λ_2 as sequence dependent, *i.e.*, their values depend on the nature of base or base pair in the sequence. This means that the shielding due to cations will depend whether, along the sequence, there is an AT or a GC base pair. We calculate the melting temperature of the homogeneous AT chain and homogeneous GC chain to calibrate our results. First, we consider two kinds of homogeneous sequence of 1000 base pairs, one that is having all AT pairs and the other that is having all GC base pairs. In Figure 4.1, we show the nature of denaturation of AT and GC chains and their dependence on the concentration of salt in the solution. We compare our results with the empirical calculations by Krueger *et al.* [69]. For pure GC chain our model based calculations are in good match with Krueger results. However, there is a difference in the behavior of pure AT chain at lower salt concentration. To investigate the sequence effect of salt concentration on the stability of DNA

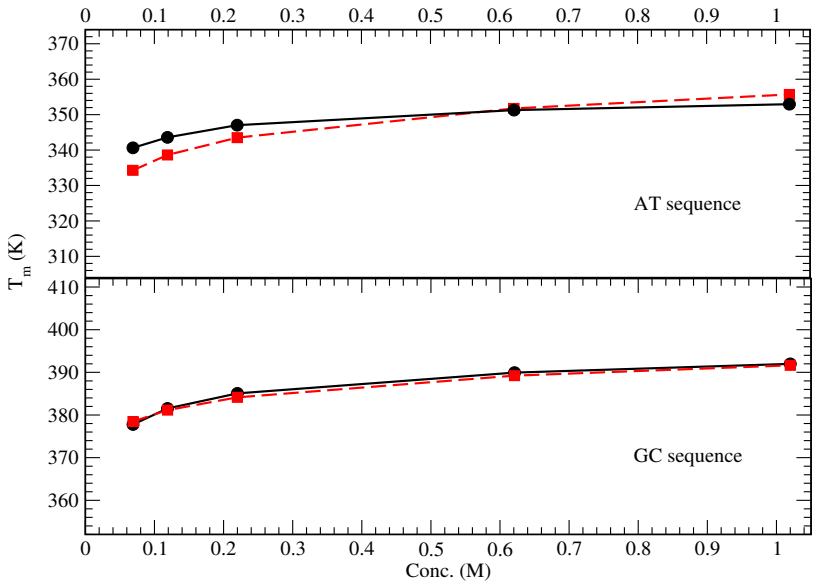


Figure 4.1: The phase diagram calculated using the PBD model (black circle) and from the empirical relation given by Krueger *et al.* [69] (red square).

molecule, we consider a chain of 1000 base pairs. The respective location of AT & GC pairs along the chain is random. We consider a chain that is having 100 AT and 100 GC pairs in a block and these blocks appear in the sequence alternatively (chain 1). We consider another chain of 1000 base pairs that is a segment of λ phase DNA (chain 2) [250]. For our investigation, we consider four different cases:

(a) constant k with $\lambda_{AT} = \lambda_{GC}$ (b) constant k with $\lambda_{AT} \neq \lambda_{GC}$

(c) variable k with $\lambda_{AT} = \lambda_{GC}$ (d) variable k with $\lambda_{AT} \neq \lambda_{GC}$

We take the elastic constant k in the stacking energy term, from refs. [14, 15, 215, 251]. We take the average of the values and for each base we calculate the change in elastic energy and scaled it to the elastic constant k_j, where, j is the index for any of the 16 possible combination as shown in Figure 4.2. We adopt the elastic

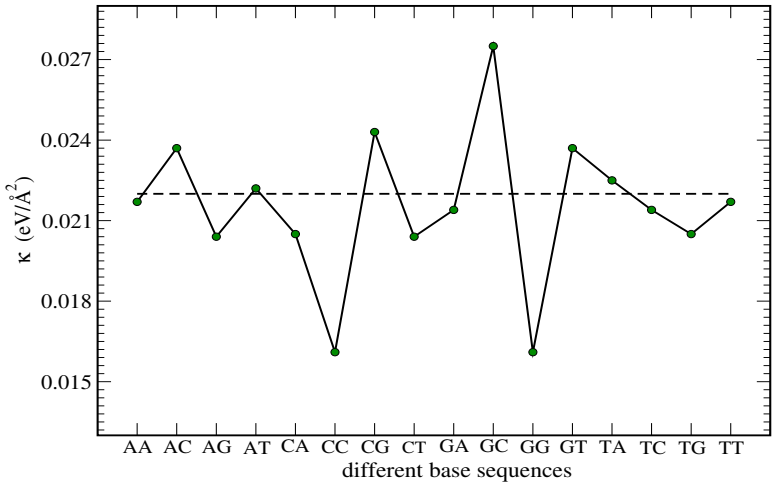


Figure 4.2: Values of all 16 different stacking constants which are calculated on the basis of results of Svozil *et al.* [14]. The stacking constant appears to be strongest for the GC pair while it is showing the lowest value for GG/CC stacking.

constant as a function of base sequence as, wherever, AG stacking is mentioned, it means the base sequence will be from $5' - A - G - 3'$ and the reverse sequence will be on the complementary strand from $3' - 5'$. That is why we have only 10 unique stacking energies as shown in Figure 4.2. The lowest value of the stacking constant is for GG and CC stacking while it is highest for GC stacking. For cases (a) & (b) the elastic constant is the average value of the variable k and it is 0.022 eV/Å^2. The value of solution constant λ is also different for AT and GC pairs. At different salt concentrations varying from 30 mM to 1.02 M, we calculate melting temperature of the chains. The results obtained for all the four cases are shown in the Figure 4.3. From the figure this is clear that when we take constant stiffness parameter (k), the melting temperature increases by $\sim$ 10 K for any amount of salt in the solution. However, the solution constants λ have significant effect as far as sequence dependence is concerned. If we compare either for constant k, [case(b)] or for variable k, [case(d)], the solution constant λ affect the melting transition in the lower concentration range. While for higher concentrations, there is no significant change in melting temperature, in the lower range of concentration, not only the T_m values ($\Delta T_m \approx 1$ K) are different but also the slope of the curve is different. This is an important point that was missing in the previous studies [47, 69]. The

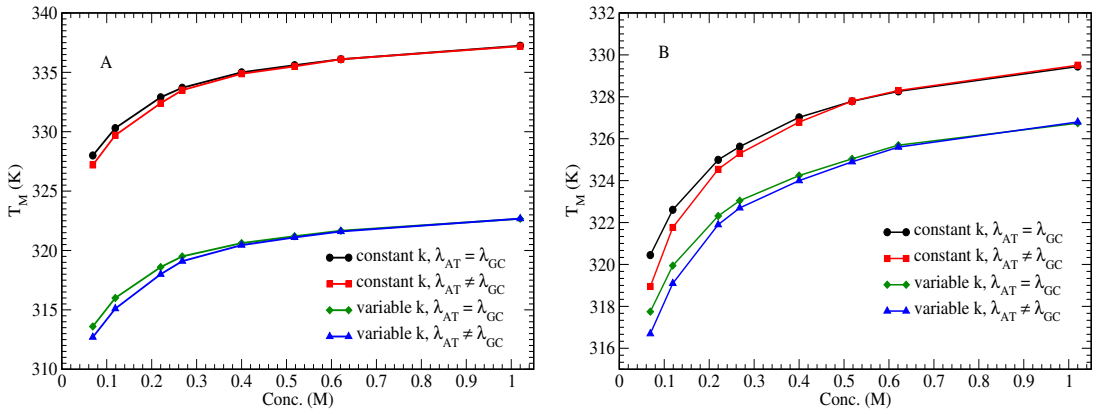


Figure 4.3: The phase diagram for temperature induced transition in DNA for four different cases. Figure A is for a chain that is having alternate AT & GC pairs while figure B is for a section of λ-phage DNA.

nature of the curves at lower concentrations indicates the difference in the opening or denaturation of different heterogeneous chains, it is because of the heterogeneity in the solution constant.

4.4 Forced induced transition

The main objective of this chapter is to investigate forced induced transition in heterogeneous DNA molecule that is surrounded by the cations. In this section, we investigate unzipping of DNA for all the (a)-(d) cases. The force that is required to unzip the chain, the critical force (F_c), is calculated in the constant force ensemble (CFE) for different salt concentrations. All the calculations are done at room temperature, *i.e.* 300 K. The model Hamiltonian for the force that is applied at either of the ends, is given by Eq.3.11, We consider two kinds of heterogeneous chains as discussed above. For the forced induced unzipping investigations we take the elastic constant, k, as an average value 0.022 eV/Å^2. The other important parameters like the potential depth, D_0, the solution constants, λ_i are tuned in order to get a good match with the experimental results. The set of values for which our results are found in better agreement with experiments are listed in Table 4.1 [1]. The melting temperature of the 1000 bps chain is around 319 K and 330 K for 0.030 M & 1.020 M respectively with these set of parameters. The transition from close to open state is different for mechanical unzipping and

[1]We have chosen these particular values of λ_n^i, because the average of λ_{AT} & λ_{GC} is equal to the values of λ in our previous investigations [47]

Table 4.1: The complete set of model parameters.

Parameters	Values
Potential Depth, D_0	0.064 eV (AT) & 0.096 eV (GC)
Inverse potential width, a	4.2 (AT) & 6.3 (GC)
Anharmonicity, ρ	1.0
Anharmonic Range, b	0.35 $Å^{-1}$
Average chain stiffness, k	0.022 eV/$Å^2$
Solution constant, λ_1	0.012 (AT) & 0.008 (GC)
Solution constant, λ_2	0.0024 (AT) & 0.0016 (GC)
Solution constant, γ	1.0

temperature denaturation. In force induced transition the opening is primarily due to stretching of hydrogen bond that causes a pair to break [140, 252]. In this case, there is an interface of open and close state which forces system to move from one state to another depending on the value of applied force [152]. We consider the four different cases as mentioned in Section 4.3 in order to get more details of the force induced transition in DNA. We calculate the critical force for each chain at different salt concentrations and plotted in Figure 4.4 In the

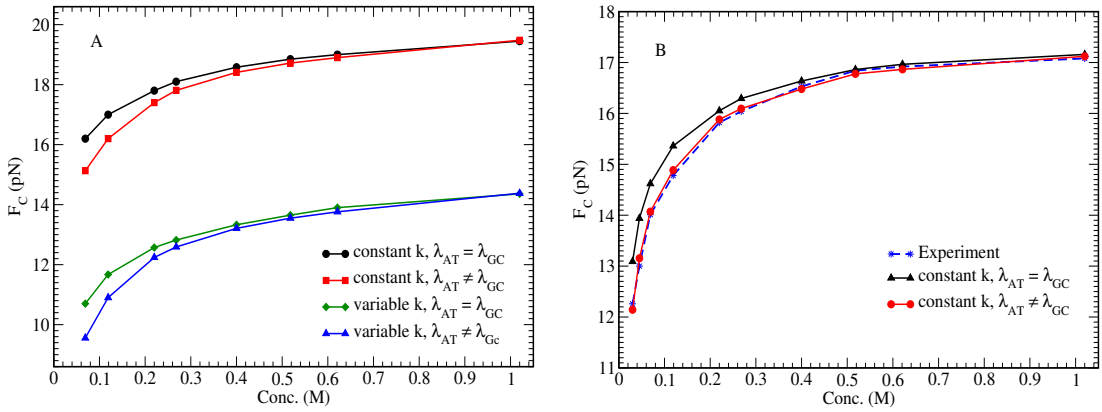


Figure 4.4: The phase diagram of DNA chain when the DNA is pulled from one of the ends. Figure A is for a chain that is having alternate AT & GC pairs while figure B is for a section of λ-phage DNA. Here we compare our results with the experimental results found by Huguet *et al.* [172].

force induced transition, we observe that there is smaller contribution of variable stiffness parameter, k. We find that the solution constant λ_i has more prominent

effect on the phase diagram, as shown in Figure 4.4. At lower concentration the phase diagram shows the difference for case of different solution constant, λ_i for AT and GC base pair. However at concentrations higher than 0.06 M, the sequence effect diminishes. When we take constant k and variable λ_i , case (c), the obtained phase diagram, in Figure 4.4B, shows a good match with the experimental phase diagram obtained by Huguet *et al.* [172]. The earlier result [47] in Chapter 3, that showed some mismatch at the lower concentrations is showing better match by including the sequence dependent solution constant. This means at lower salt concentration the behavior of AT and GC is different towards the cations present in the solution. To get more insight of the sequence dependent salt effect on the force induced transition in DNA, we take different segments of λ-phage DNA that is having different GC content. The sequences that we take, are [1]:
(i) 1^{st} 1000 bp having 51.30%GC,
(ii) 2^{nd} 1000 bp having 53.60%GC,
(iii) 3^{rd} 1000 bp having 54.50%GC and
(iv) a segment of 1000 bp from the mid having 58.90% of GC pairs.
We observe some interesting features of the transition from close to open state for different sequences as shown in Figure 4.5. However, the best match we are

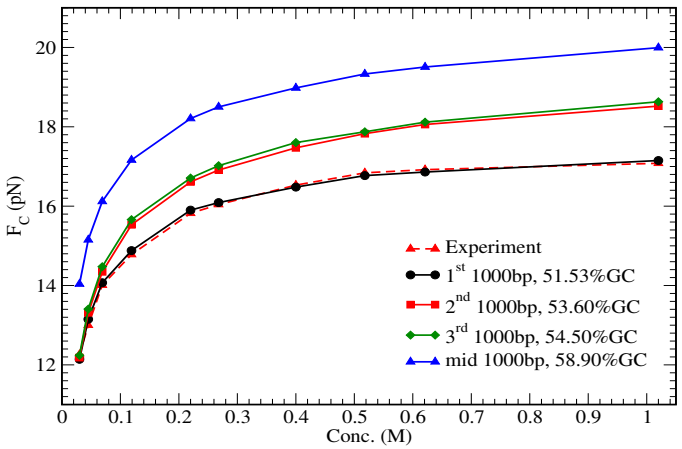


Figure 4.5: Variation in the critical force of λ-phage DNA molecule with salt concentration of the solution. Here we take different segments of the chain that varies not only in %GC content but also in the sequence of the bases along the strand. Again we compare our results with the experimental findings of Huguet *et al.* [172].

getting for the chain that is having 51.53% of GC content. This is in accordance with the fact that real sequences of DNA is having about 50% GC pairs randomly

[1]We take the sequence from the url:
http://www.cf.ac.uk/biosi/staffinfo/ehrmann/tools/dna/PhageLambda.html [250].

distributed along the chain. The critical force for different DNA sequence varies as the GC content varies, as shown in Figure 4.5

4.5 Conclusions

Sequence dependent salt concentration effect on the stability of DNA molecule have been studied in this chapter. As the interaction between AT and GC pairs are not same, the interaction of cations with these molecules may not be the same. Therefore we have modified the on-site potential to take care of the heterogeneity in the sequence (*i.e.* the presence of AT or GC pair). Using PBD model, we have investigated the role of cations on the melting temperature as well as on the critical force that is required to unzip the chain from one end. In the mechanical unzipping case, we have considered different chains that are having different GC content. Our results are in good agreement with the experimental results for the chain that is having 51% GC content. This is consistent with the fact that for the experiments, in general, the sequence that is used, are having $\sim 50\%$ GC content. The sequence heterogeneity can be introduced by variable bond strength of the pairs and by choosing variable stacking energies for possible 16 conformations of the base stacking in the PBD model. Hence, we chose variable values of chain stiffness k as well as the variable solution constants λ_i to analyze the sequence effect on the stability of the molecule. We found that the cations have more significant effect on the H-bonding (and hence on the stability of the molecule) than the variable base stacking of the sequence . This may be due to fact that the hydrogen bonding play more crucial role than the stacking energy in the overall stability of the molecule. However, this should be noted here that our calculations are based on the average value of stacking as well as bond energies.

Since our investigations based on PBD model, reproduce experimental results very well, we conclude that the this model is a good choice to investigate the presence of salt in the solution and its effect on the mechanical unzipping of the of dsDNA molecules. One can predict melting temperature as well as the critical force even at low salt concentrations of the solution by introducing some modification into the model.

Chapter 5

Pulling short DNA molecules having defects on different locations

5.1 Introduction

As it is known that the allowed pairing in the two complementary strands of DNA follow a simple rule, that is, Adenine (A) can form a hydrogen bond with Thymine (T) while Guanine (G) can form a hydrogen bond with Cytosine (C) [8, 36]. The strength of these two kind of base pairing is not same as the AT base pairs has two hydrogen while GC base pair has three hydrogen bonds. The approximate ratio of GC and AT bond strengths varies from 1.2 to 1.5 as mentioned by various research groups [123, 193, 228, 253]. In absence of the complementary base on the opposite strand, the pairing between the two bases is absent. This site is called a *defect site* as there will be an unstable (or non-existing) hydrogen bond between two bases on the opposite strands. Defect in the DNA molecule plays a crucial role in the biological processes like replication and also a single defect can significantly change the conformational fluctuations [110, 243, 254], because replacing a paired base with a mismatched base allows flexibility in the strand. The presence of these defects offers many interesting physics and biochemistry of the molecule. The dynamics of these defects may delay the replication process and hence leads to the breathing dynamics of opening of the chain [129]. It is predicted that in embryonic cells, these delays may cause the *cell death* while in mature cells like somatic cells, this damage (defect) may be an initiation step in the development of cancer [255–257]. These defects are present in the DNA based actuators. The

role of these defects in the designing of molecular motor has been discussed by McCullagh *et al.* [130]. How these defects affect the melting, elastic properties etc., are problem of scientific interest. The main objective of the current study is:

- Effect of defect(s) location and density on the denaturation of DNA molecule.
- Investigate the opening of DNA molecule having different sequence of AT/GC with defected or mismatched pairs.
- Comparison of the results of both the ensembles, thermal and force induced opening.

5.2 Modified Model Hamiltonian

We use Peyrard Bishop Dauxois (PBD) model to investigate the thermal and mechanical denaturation of DNA molecules in presence of defects. This model considers the stretching between corresponding bases only and the hydrogen bonding between these two bases is represented by the Morse potential. If a pair is defected it means there is an absence of hydrogen bond and thus we need to modify the on-site potential for any defect in the sequence. We have followed the previous approach by Singh and Singh [120, 258] where the defect in the model was introduced via Morse potential that represents the hydrogen bonding with zero potential depth. We consider only the repulsive part of the potential in order to avoid the crossing of two bases in a pair. The Morse potential for defected site is presented in the Figure 5.1 Thermodynamics of the transition can be investigated by evaluating the expression for the partition functions, given in Eq.2.9 and Eq.2.10 with this modified potential. As we have discussed in Chapter 2.3, there are different ways to evaluate the configurational partition function. The important part in the integration of the configurational space is the selection of proper cut-offs for the integral appearing in Eq.2.10 to avoid the divergence of the partition function for finite chain. The method to identify the proper cut-off has been discussed by several groups [123, 202, 222, 259]. The calculations done by T.S. van Erp *et al.* show that the upper cut-off will be $\approx$ 144 Å with the our model parameters at $T = 600$ K while the lower cut-off will be -0.4 Å. In an earlier work by Dauxois and Peyrard it was shown that the T_m converges rapidly with the upper limit of integration [259]. In that work they considered infinite homogeneous chain and evaluated the partition function using TI method. For short chain, we calculate T_m for different values of upper cut-offs which is shown

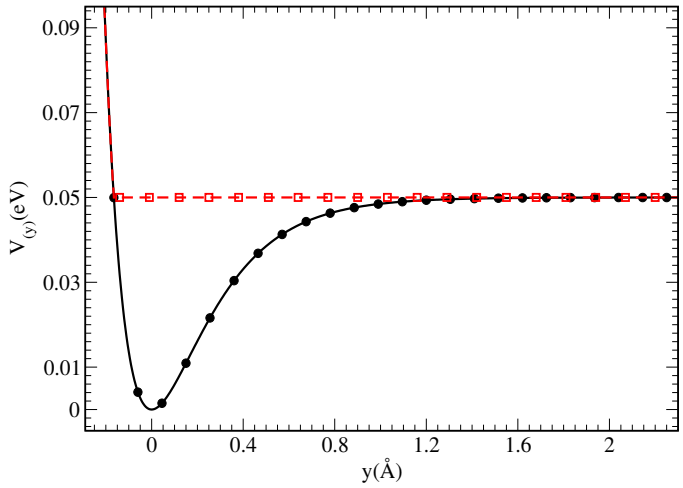


Figure 5.1: The on-site potential for the defect in a pair is shown by square symbol and dashed (red) line. There is no minimum of potential for the defect site. While for the bases in a pair is represented by the depth of the potential (solid line with black circles).

in Figure 5.2. From the plot it is clear that the choice of 200 Å is sufficient to avoid the divergence of partition function. Thus the configurational space for our calculations extends from -5 Å to 200 Å. Once the limit of integration has been

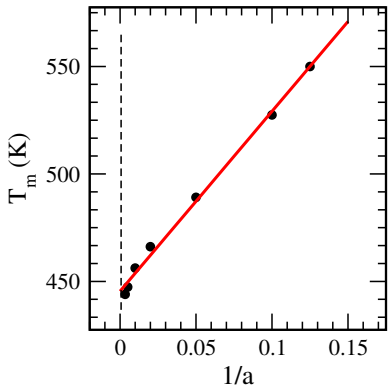


Figure 5.2: The melting temperature T_m calculated for different values of upper cut-off (a) for homogeneous chain. The best straight line fit for this plot is found for $1/a$. The different cut-offs are 8, 10, 20, 50, 100, 200 and 300 Å.

chosen, the configurational partition function Z_c, can be evaluate by using matrix multiplication method [120]. The calculated partition functions, Z_p (Eq.2.9) and Z_c (Eq.2.10), are used to calculate the other thermodynamic quantities of interest by evaluating the Helmholtz free energy of the system, Eq.2.19. One can calculate the probability of opening of the i^{th} pair through the model Hamiltonian, which

is defined as: [193]:

$$P_i = \frac{1}{Z_c}\int_{y_0}^{\infty} dy_i \exp\left[-\beta H(y_i, y_{i+1})\right] Z_i \tag{5.1}$$

where

$$Z_i = \int_{-\infty}^{\infty} \prod_{j=1, j\neq i}^{N} dy_j \exp\left[-\beta H(y_j, y_{j+1})\right] \tag{5.2}$$

while Z_c is the configurational part of the partition function defined as in Eq.2.10. For y_0, we have taken a value of 2 Å. We calculate the opening profile of a chain with different positions of defect(s).

Another quantities like the average fraction $\theta(=1-\phi)$ of bonded (or open) base pairs can be calculated by introducing the dsDNA ensemble(dsDNA) [123] or using the phenomenological approach [120, 200] which we have discussed in Chapter 3.2.

5.3 Temperature induced transition

In experiments, researchers synthesize and/or characterize the samples to decipher the information stored by that sample. Similarly, we choose four samples of DNA molecules with equal chain length of 16 base pair. All these samples are having different number and distribution of AT/GC pairs. We identify all these molecules according to the distribution of base pairs and named them as:

Chain 1: *3'-AAAAAAAAAAAAAAAA-5'* (homogeneous)
Chain 2: *3'-AGAGAGAGAGAGAGAG-5'* (alternate AT/GC pairs)
Chain 3: *3'-AAAAAAAAGGGGGGGG-5'* ($50\%AT{+}50\%GC$)
Chain 4: *3'-TCCCTAGACTTAGGGA-5'* (random sequence)

As the nature of these chains are different, the number and location of the defect(s) should have significant effect on the melting profile of the chain. We consider the defects (varying in number from 1 to 4) and their effect on the melting temperature of the system. As we use the force ensemble also to study the defect location and density effect, one has to insure that the opening of dsDNA is due to the applied force only, not because of the thermal fluctuations. Hence, we choose the model parameters such that these chains are having melting temperature above 300 K. The complete set of model parameters are listed in Table 5.1 We calculate the melting temperature of the above four chains. At the temperature when the system gets the sufficient amount of energy that is needed for transition from double

Table 5.1: The complete set of model parameters.

Parameters	**Values**
Potential Depth, D_0	0.10 eV (AT) & 0.15 eV (GC)
Inverse potential width, a	4.2 $Å^{-1}$(AT) & 6.3 $Å^{-1}$ (GC)
Anharmonicity, ρ	5.0
Anharmonic Range, b	0.35 $Å^{-1}$
Average chain stiffness, k	0.0215 eV/$Å^2$
Opening cutoff, y_0	2.0 Å

stranded configuration to single stranded configuration the free energy shows a kink. For better visualization, we show the transition through the specific heat per base pair as a function of temperature. Figure 5.3, shows the transition point

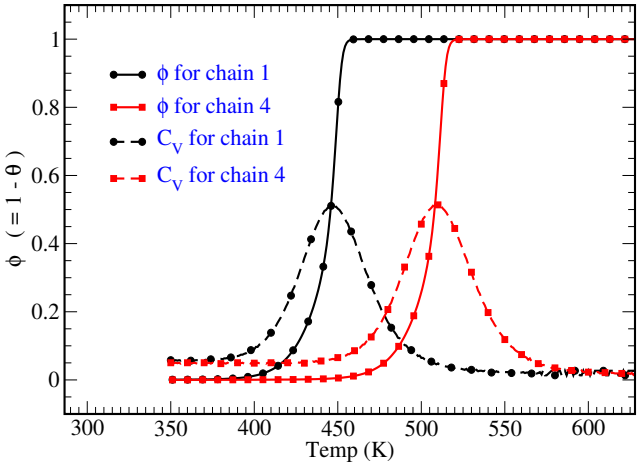


Figure 5.3: The melting temperature, T_m, calculated by specific heat and fraction of open pair θ for the *chain 1* (homogeneous) and *chain 4* (random). The parameters p and q are adjusted in order to get precise match with peak in specific heat. The values are $p = 12.0$ and $q = 10.0$. The value of C_v is scaled to show that the peak position and 50% of the open pairs meet at the same point (temperature).

for above said chains without any defect in the chains. The melting temperature, T_m is calculated by the peak in the specific heat as well as from θ as given in [120, 200]. The melting temperature for chain *1, 2, 3,* and *4* without any defect is 447.5 K, 508.8 K, 511.0 K and 509.8 K, respectively. Now we introduce the defect in the chain at different locations and investigate the melting of these four different DNA chains with different defect densities.

Homogeneous chain

Figure 5.4 represents the change in T_m for different defect density in the chain. We introduce defect in the chain varying number from 1 to 4. The x-axis of the

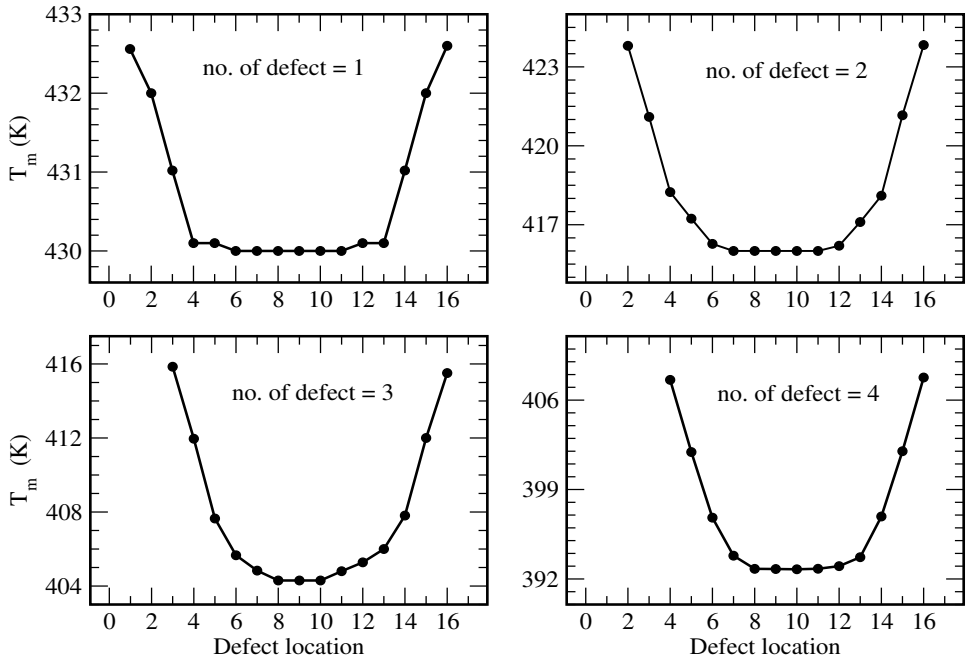


Figure 5.4: The melting temperature, T_m vs the defect locations for homogeneous chain. We take defects in number from 1 to 4 in the sequence.

plot is the defect location in the chain. As we introduce the defect in the chain, there are effectively fewer base pairs to break, and hence we need less energy to melt the DNA chain. For the case of one defect, the defect location is mentioned as where the defect is present. Now we take two continuous defects in the chain and change the location from 3′-end to 5′-end. So the defect location in this plot is the second defected site varying from 2 to 16. We further increase the defects in the chain, in case of three defects, the defect location vary from 3 to 16, as we take continues three defects in the chain, while in case of four defects it vary from 4 to 16. For example with four defects from 7 to 10 is located at 10 in the plot and if two defects at 5 & 6, it is located at 6.

In case of one defect, when the first site (3′-end) is a defect pair, the melting temperature is about 433 K. The melting temperature further reduces to 432 K if the II pair is the defect pair. However there is something interesting to note after this. As the location of defect is moved to 4^{th} pair onward the melting temperature reduces to 430 K and remains constant till we reach to 5′-end at 13^{th} site. As we proceed towards 5′-end, the T_m increases and the complete cycle displays a necklace

kind of plot as shown in Figure 5.4. As we increase the number of defects from 1 to 4 the melting temperature decreases up-to $\sim$ 408 K for defects at the ends, and up-to $\sim$ 393 for defects at the mid of the chain. Qualitatively, we may argue that the defect introduces two extra ends in the middle of the molecule, which becomes effectively weaker chain and thus we get low melting temperature when defects are at mid. We observe that all four plots preserve the necklace kind of pattern for complete cycle. It seems interesting that the locations of defect(s) at the middle of the chain having equal melting transition of DNA molecule. As the number of defects in the chain is increased from 1 to 4 this width decreases from 10 base pair to 6 base pairs.

Chain having alternate AT/GC pairs

We consider Chain 2 with alternate AT/GC base pairs. Now it matters whether the location of defect site is AT or GC pair. In this chain, the energy landscape of the chain is not smooth over the complete region as shown in Figure 5.5. This is

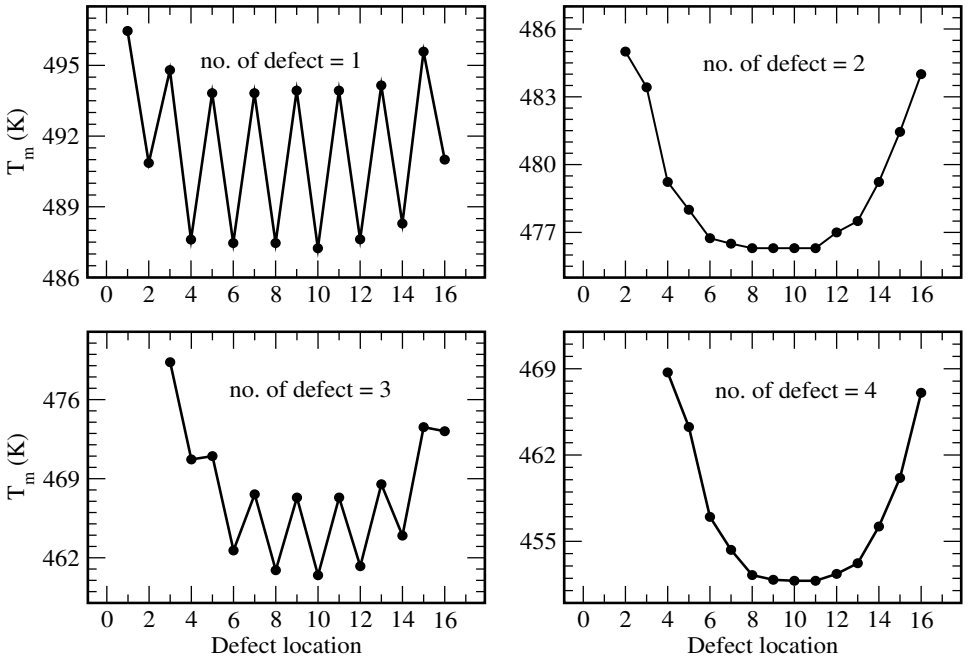


Figure 5.5: The melting temperature, T_m vs the defect locations for a chain having alternate AT/GC pairs for different defect densities.

due to the difference in the dissociation energies of AT and GC pairs. Therefore T_m for GC defected site reduces more than the AT defected site and hence for one defect that move from 1-16, the T_m shows a zig-zag pattern, the variation in T_m is from 496 K to 487 K. When we consider two defects in this chain this pattern

is lost, this means with two defects the sequence heterogeneity is lost and there is an average of AT and GC pair's dissociation energies. However, the necklace pattern obtained for this case is not as symmetric as for the homogeneous chain. It is because of the ends are different, at 3′-end there is an AT pair while on 5′-end, there is a GC pair. Moreover this pattern will be symmetric with odd number of base pairs in the chain (ends will have same kind of bases). With three defects in the chain, we obtain again a zig-zag pattern for the values of T_m at different defect locations but the T_m is lower as compared to the chain with one defect. Further increase in the defect, with four defect we retain the asymmetric necklace pattern for T_m at different defect locations, see Figure 5.5.

Chain having 50% AT & 50% GC base pairs

For chain 3, we obtain a hook kind of structure in the plot for T_m at different defect locations with one defect, see Figure 5.6. As this chain is having 50%AT+50%GC pairs in the sequence, in the middle of the chain (on 8 & 9 pair) there is an interface of GC & AT pair. Therefore we observe a sudden drop in the T_m, as the next defected pair is a GC pair. The smoothness at the interface

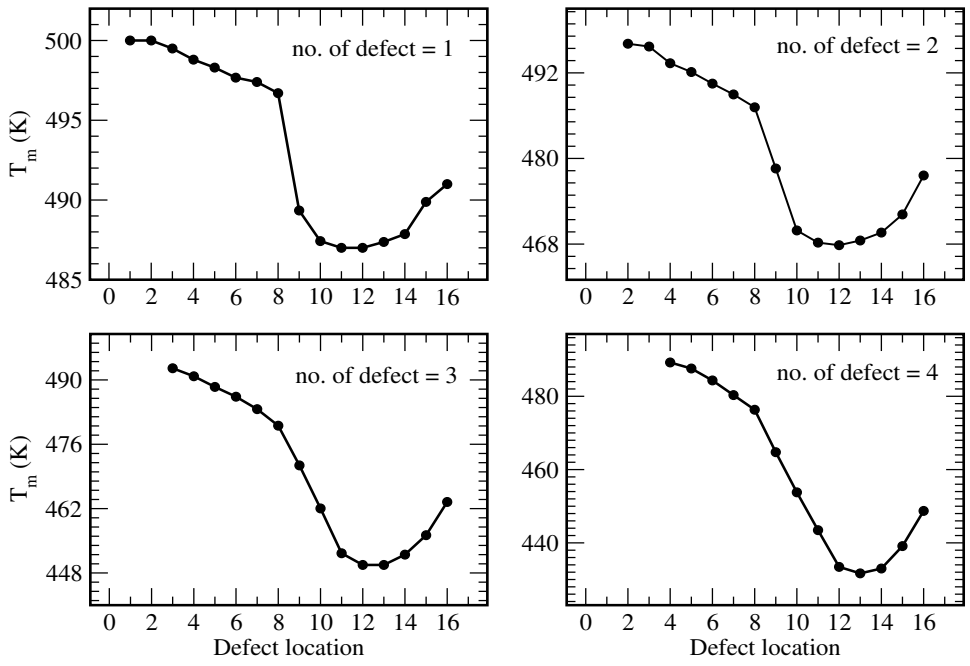


Figure 5.6: The melting temperature, T_m vs the defect locations for a chain having 50%AT and 50%GC pairs. for different defect densities.

increases with increase in the number of defects in the chain as shown in Figure

5.6. As the number of defect is more, on the interface the effect of presence of AT and GC pairs diminishes.

Random sequence of AT & GC

We take a chain with equal ratio of AT & GC but random distributed over the chain. The difference in energies of AT and GC base pairs can clearly be observed with the defect of these bases in the chain as this is reflected from the fluctuation in the values of T_m in this chain with one defect, Figure 5.7. The random pattern

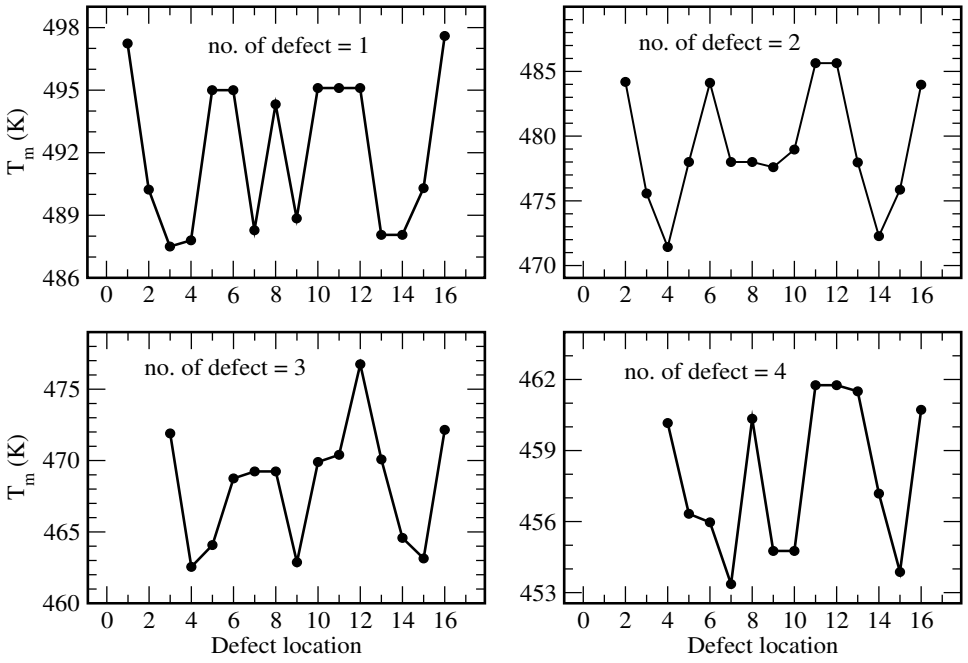


Figure 5.7: The melting temperature, T_m vs the defect locations for a chain having random distribution of AT & GC pairs. for different defect densities.

of the plot varies with number of defects in the molecule. consider, single defect on 4^{th}, 5^{th} & 6^{th} site. While T_m is 488 K for 4^{th} site, it is 495 K for 5^{th} & 6^{th} sites. For two defects in the molecule, this is averaged to 471 K, *i.e.* the distinction of AT and GC pair is lost. Similarly, for three defects in the molecule, the high barrier on 10^{th} & 11^{th} sites is lost. As the defect location at the mid gives less T_m, the chain is having lowest melting temperature with four defects from 4^{th} to 7^{th} and two of them are GC defect. From the pattern of T_m with four defects we observe that the T_m is less when 2 defect are GC bases out of the four defects. With one defect T_m is less with GC defect.

5.4 Force Induced transition

Physics of opening of chain due to thermal fluctuation and mechanical forces is completely different. Whereas, in the case of thermal denaturation, the opening is due to increase in the entropy of the system, in the case of mechanical unzipping, opening is enthalpic. All the base pairs of dsDNA that is kept in a thermal bath share equal amount of energy. In the case when the chain is pulled from an end, the amount of force decreases from the pulling point to the other end of the chain. Thus the location of defect(s) should have different impact on the opening of the chain that is subjected to a mechanical pull on an end. As it is also known that replication process is initiated due to force exerted by DNA polymerase on a segment of DNA chain, it would be interesting to investigate the role of the defect(s) in the pair on the replication process. Mathematically one can model the replication as the force applied on an end of the DNA chain We investigate here the mechanical unzipping of the same four chains that we considered for thermal denaturation investigations.

We will show here that PBD model successfully describes many of the unzipping properties of DNA with a mismatched/defected sequence also. We study the mechanical unzipping in CFE. We calculate the critical force of the chain at different defect locations and densities in the chain via calculating the free energy of system through Hamiltonian given in Eq.3.11. Figure 5.8, shows the specific heat

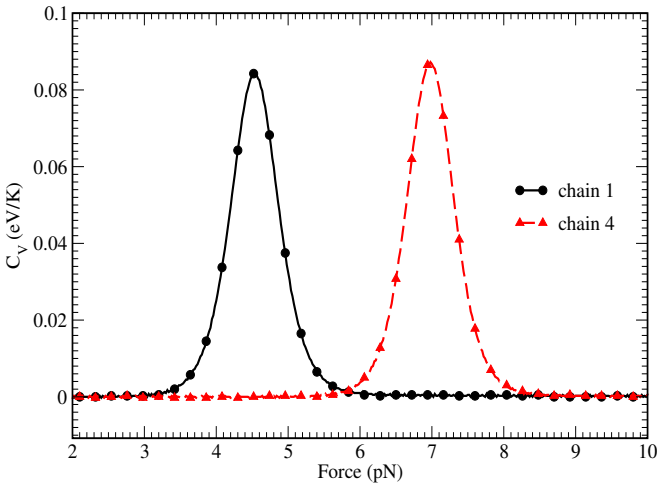


Figure 5.8: The critical force, F_c calculated by specific heat for the *chain 1* (homogeneous) and *chain 4* (random) with out any defect in the chain. The peak in specific heat corresponds to the critical force of the chain.

as a function of force for the above said chains. The force is applied on 3'-end

of the chain. The critical force for pure chain *1, 2, 3* and *4* is 4.54, 6.96, 6.98, and 6.97 pN respectively. Like this we calculate the critical force for each chain with the defected or mismatched pairs in the chain and plotted that critical force with defect location in the chain. We also monitor the applied force directions *i.e.* from $3'$ to $5'$-end or $5'$ to $3'$-end and investigate the change in F_c for both force directions.

Homogeneous chain

As shown in the Figure 5.9 when the force is applied on $3'$-end and the defect site is 1^{st} pair, the critical force reduces to 4.04 pN from 4.45 pN. This value further decreases to 3.99 pN when the defect site is 2^{nd} pair. When the defect is

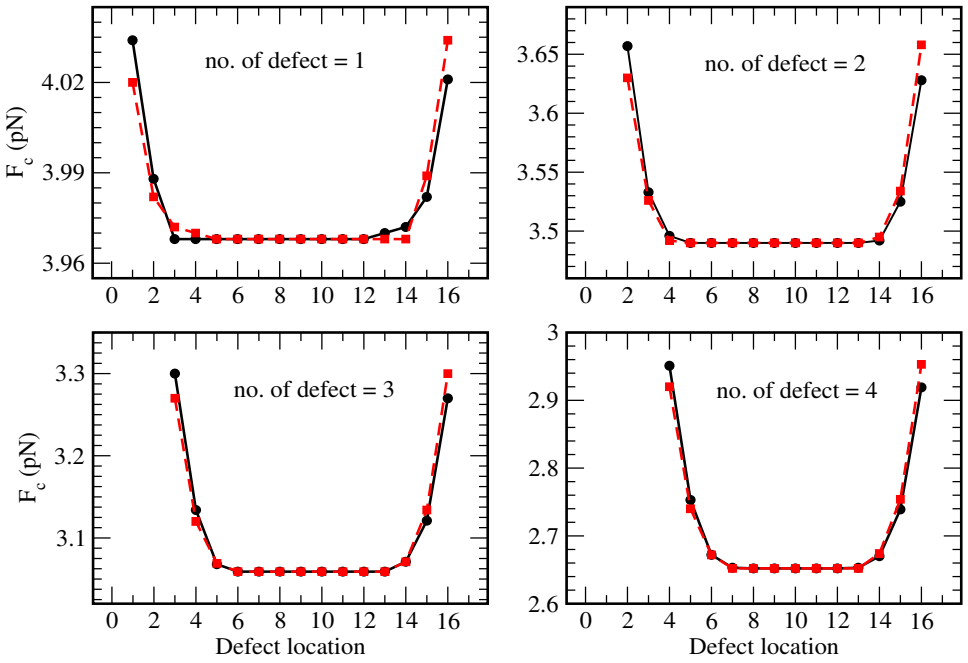


Figure 5.9: The critical force, F_c, for homogeneous chain with different defect locations, We show both the cases when force is applied at $3'$-end (black solid line) and at the $5'$-end (red dotted line).

located anywhere between 3-13 sites, critical force remains equal for different defect locations, it is 3.97 pN. This means that the base pairs (defects) in this section of the chain have similar responses to the applied force, irrespective of their location. The defect pair means a loop in the chain which will increase the entropy of the chain. From the results, it is clear that the loop contributes to the opening of the chain in addition to the applied force and end entropy. As the location is between 14-16 the end entropy suppressed by the defect and hence there is an increase in

the value of critical force in this region. The critical force is for this case is 4.02 pN. In this case too, we observe a necklace pattern. The pattern obtained here is not as symmetric as that observed for thermal melting of the same chain with a single defect (see Figure 5.4).

We consider now the opening of the chain in another condition: The force is applied on the 5′-end and defect pair is at the 3′ end. In this case, the critical force is 4.02 pN which is less than that for the previous case (where the force is on 3′-end and defect pair is also at the 3′-end), where F_c is 4.04 pN. The difference is about 0.02 pN. The reason for this reduction is the difference in the end entropies for these two cases. In case when force is applied on 5′-end and the defect location is at 3′-end, the end entropy of this open end contributes to the opening. In contrast to the first case, when the defect end is the 3′-end (the force is also on this end), the contribution from the 5′-end will be less as it is an intact pair. Hence we need slightly higher force to open the chain for the first case. We obtained similar results for this chain with more defects (m = 2,3,4). For all the investigations, whenever m > 1 all the defects are consecutive defects. As the number of defects increases in the chain, the difference in F_c for the two different cases is greater. To validate our arguments, we calculate the probability of opening for different defect locations in the chain using Eq.5.1. To avoid the overflow of figures, we

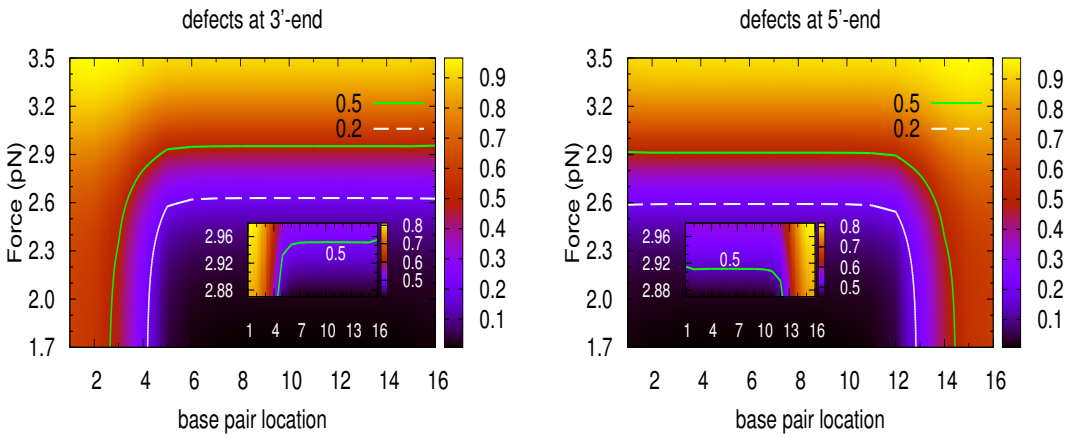


Figure 5.10: The density plots to show the difference in the opening of a homogeneous DNA molecule (chain 1 with four defects) when force is applied on the 3′-end. (Left) When defect pairs are 14 (3′-end). (Right) When the four defect pairs are on the 5′-end (1316). The difference in the critical force F_c for the two cases is observed here (more clearly in the zoomed version). In order to open 50% of the base pairs, the F_c for first case (left) is 2.96 pN while for the second case (right) it is 2.92 pN [260].

display the surface plot for four defects at 3′-end & 5′-end in Figure 5.10. The finite entropy of homogeneous chain is important here. We compare these two cases in the surface plot. The opening profile verify the suppression of end entropy when the defects are at 3′-end while force is also applied at 3′-end. The case is just opposite when the defects are at 5′-end, here the end entropy due to this 5′ open end, will contribute in the opening and hence the critical force decreases more than the case when defects are at 3′-end. We observe that the difference in the critical force for these two cases is ∼0.04 pN We obtained same pattern for decrease in F_c with defect location while changing the applied force direction at 5′-end. It is apparent as the sequence is homogeneous in nature.

Chain having alternate AT/GC pairs

In this case with one mismatch in the chain, the energy landscape of the chain is not smooth over the complete region as shown in Figure 5.11. This is due to the difference in the dissociation energies of AT and GC pairs. The energy loss of a disruption of a GC pair is more than that of an AT pair. Hence, a GC mismatch will likely result in an overall decreased F_c. Therefore, for one defect that move from 1-16, the plot for F_c shows a zig-zag pattern. In this case also, we have

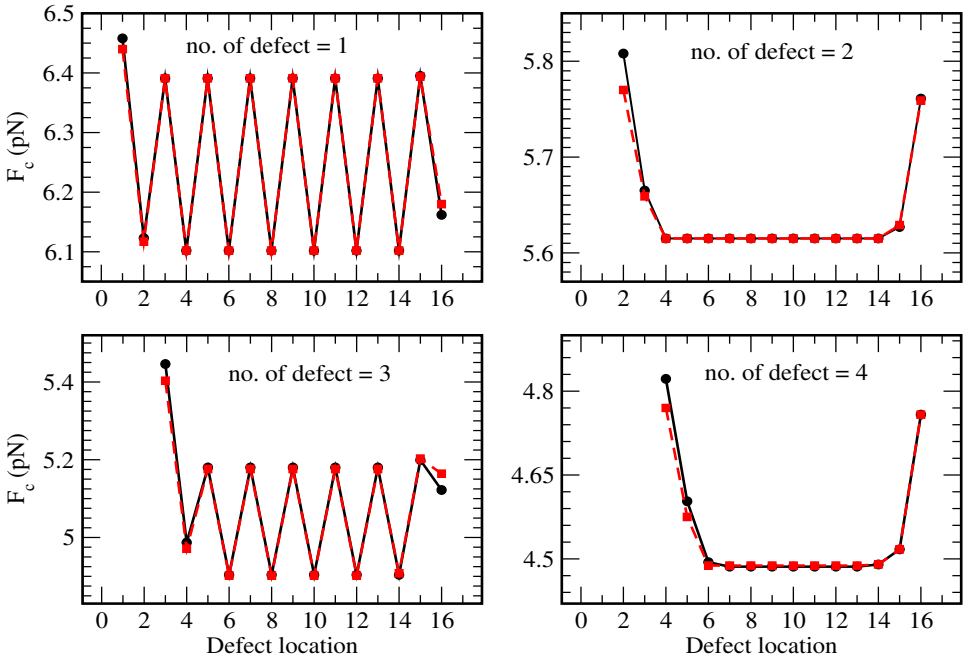


Figure 5.11: The critical force, F_c, for different defect locations for a chain having alternate AT & GC pairs, We show both the cases when force is applied at 3′-end (black solid line) and at the 5′-end (red dotted line).

four possible combinations of force and defect locations. The first is when force is

applied on the 3′-end and the defect is also at the 3′-end. The second one is when force is applied on the 3′-end and the defect is at the 5′-end. The other two cases are the alternate combinations of these two. The end entropy effect (when the defects are at the ends) is clearly visible as we change the applied force direction. If we fix the location of applied force at the 3′-end and change the defect locations, we find that for a single defect the difference in F_c is ~ 0.3 pN. This is because of the difference in the entropy contribution from the two ends. In one case the end is AT while in another case it is GC. If we fix the defect location and change the applied force locations from the 3′-end to the 5′-end, we find that the difference in F_c is $\sim$0.02 pN. The same argument which we gave for chain 1 with a single defect is valid here too.

Further more, with two defects in sequence, the sequence heterogeneity is lost and there is an average decrease in F_c of AT and GC pair's energies. When the force is applied on the 3′-end and the defect locations are at the 3′-end and the 5′-end, the difference in F_c is $\sim$0.04 pN. This chain with two defects can be thought of as a homogeneous chain (of an AT+GC block) with a single defect. However, the ends in this chain can be either AT or GC and hence we get a different pattern at the ends as compared to chain 1. For this chain, in addition to the ends the interface of a defect and intact pair affects the opening of the chain. The interfaces for this chain with two defects are either of AGA or GAG kind at the ends. Hence there is a difference in the critical force for the four cases. Similar kind of feature is observed with four defects for the same chain .

Chain having 50% AT & 50% GC base pairs

This chain has AT pairs at 3′-end and GC pairs at 5′-end, therefore the change in F_c due to defects at 5′-end is more than the defects at 3′-end.[1] Hence we get a hook kind of structure in the plot for change in F_c as shown in Figure 5.12. In this case, it is important on which end the force is applied. When the force is applied on the 3′-end and the defect locations are at the 3′-end and the 5′-end the difference in the F_c is $\sim$0.32 pN. We see the pattern is more smooth for T_m in comparison with the pattern for F_c with the defect locations, since the opening is entropic in nature in thermal denaturation while it is enthalpic in mechanical unzipping. As we increase the defects in the chain the smoothness at the interface, (on 8^{th} & 9^{th} pair) increases.

[1] This is due to the difference in the dissociation energies of AT and GC pairs along with the end entropies.

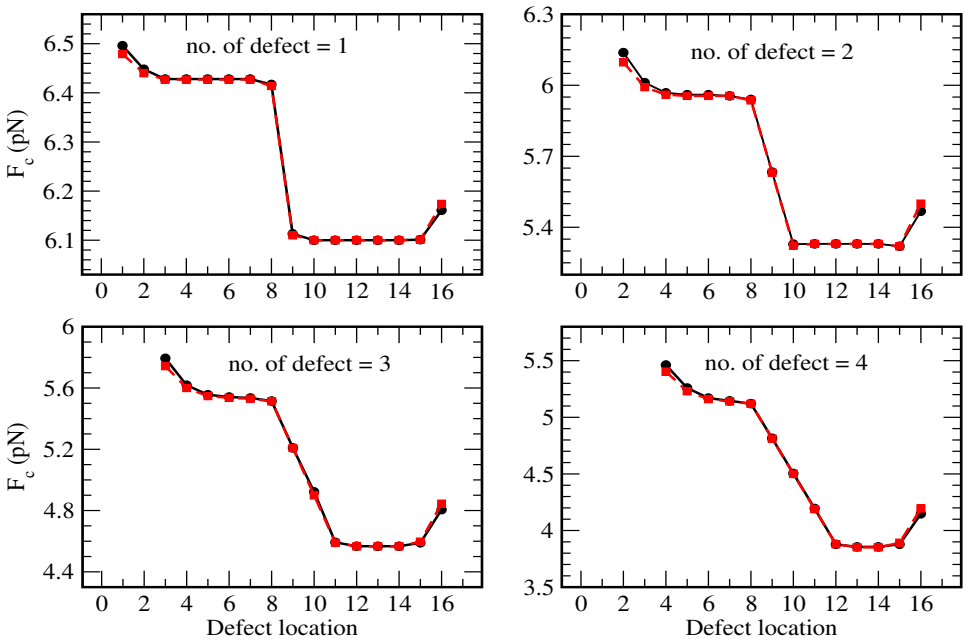


Figure 5.12: The critical force, F_c, for different defect locations of 50% AT & 50% GC pairs. We show both the cases when force is applied at 3′-end (black solid line) and at the 5′-end (red dotted line).

Random sequence of AT & GC

Let us investigate the chain 4, random sequence of AT & GC pairs with 1-4 number of defects. With one defect at the middle of the chain, we obtain decrease in F_c according to the defected pair's nature, see Figure 5.13. The pairing of bases at 7^{th} & 9^{th} site is GC, therefore we get lower value of F_c for mismatch of these bases, while the other sites at mid are AT in nature, the critical force, F_c is high for mismatch at these sites. With more number of the defects in the chain, the decrease in F_c is according to the average energy of the defected sites, just as we get equal value of F_c for two defects at sites from 7^{th} to 10^{th}. The unzipping behavior of this chain displays some of the features of all the three chains that we considered above. In case when the defect(s) are in the middle of the chain, the change in the value of critical force is negligible, i.e., it does not matter from which end the chain is pulled The pattern for decrease in F_c with defect locations for this chain is more or less equal to the pattern for T_m except with four defects in the chain. The end entropy effect is also visible here for applied force at 3′-end & 5′-end.

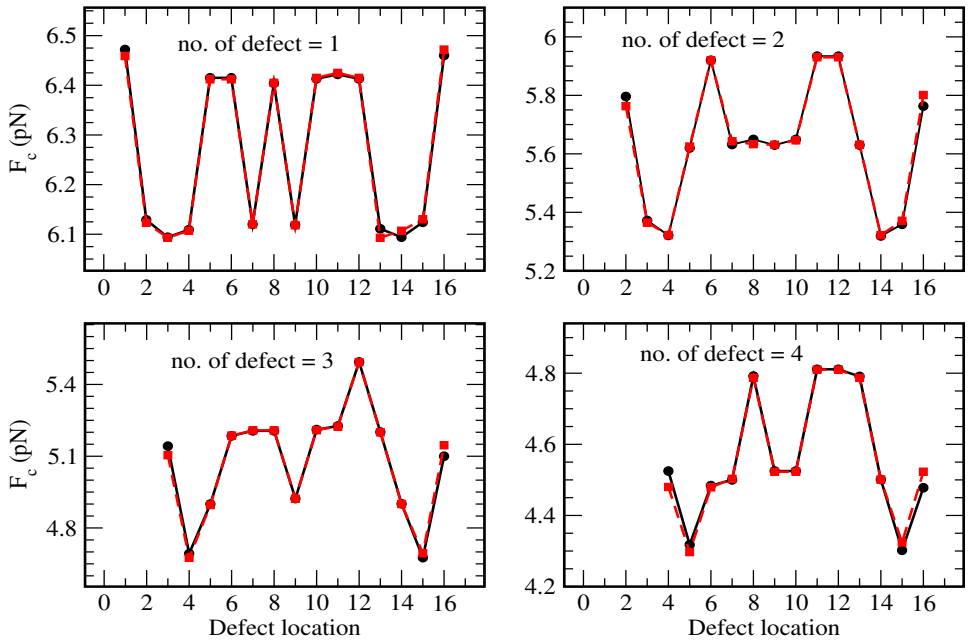


Figure 5.13: The critical force, F_c, with different defect locations for a chain having random distribution of AT & GC pairs, We show both the cases when force is applied at 3′-end (black solid line) and at the 5′-end (red dotted line).

5.5 Conclusions

We have studied the role of defects on the thermal as well as mechanical denaturation of DNA molecules of short length. It is known that the defects delay the replication process which may further cause cell death and hence may lead to initiation of cancer. Motivated by experimental studies, we considered four different kinds of DNA molecules. These molecules have different numbers of AT and GC pairs and the distribution of these pairs along the chain is also different. We have monitored the defect density as well as the defect locations in the chain and calculated the melting temperature as well as the critical force for these different chains. The mechanical unzipping was performed in constant force ensemble. We have employed the PBD model for these investigations and the on-site potential for defect site is modified by a potential that has only a short-range repulsion and a flat part without well of the Morse potential.

We found some interesting results for opening of DNA molecule which has mismatch in the sequence. The effects of mismatches were found to depend on the specific base pair, the sequence, and the location of the defects. For homogeneous chain, we found that there is a segment (4-13) of the chain, where no matter where the defect is present in the chain the T_m does not change. In case of heterogeneous

chain, there is no plateau but it matters on a location whether there is an AT pair or a GC pair. When we compare opening in two ensembles for homogeneous chain we observe that there is a striking difference. While for the homogeneous chain we obtained a symmetric necklace kind of plot for decrease in T_m in thermal ensemble, this was missing in force ensemble. This indicate the role of finite end entropy of the homogeneous chain in the denaturation of the DNA molecule. As in the case of thermal melting the ends were not of much importance because each of the base pair share the same amount of thermal energy. There only the sequence of AT & GC pairs matter.

For mechanical unzipping case, it is important for all kind of chains whether the force is applied on $3'$-end or $5'$-end. For unzipping in the constant force ensemble we considered four possible cases. The first two are when force is applied on the $3'$-end and the defect locations are either on the $3'$ or $5'$-end. Similarly, the other two combinations are when the force is applied on the $5'$-end and defect locations are either on the $3'$ or $5'$-end. In all these cases, the nature of the end pair is important. To show the importance of ends in the opening we calculated the probabilities of opening for the homogeneous chain with four defects. Here we considered first two cases. When the defect location is the $3'$-end, the end entropy is suppressed and hence we obtained a slightly higher critical force for this case. For a sequence of alternate AT & GC pairs, the qualitative features of the pattern of T_m & F_c are similar except the defect location at the ends. For this chain, we observed that, in addition to the ends the interface of a defect and intact pair affects the opening of the chain. We believe that the location of the defect, which influences the local binding energy distribution, has an impact on the overall stability of the DNA molecule.

In conclusion, our results demonstrate that DNA denaturation is sensitive to the defects in DNA base-pairing. This is an attempt to understand the defects and their effect on the replication process. This may have implications for the design of DNA detection technology. Once fully characterized, the results may be used to detect DNA sequences with defects. The system also provides an opportunity for investigating how a local, microscopic perturbation affects the macroscopic properties of the opening of double helix DNA.

Chapter 6

DNA denaturation in crowded environment

6.1 Introduction

Since the first experimental evidence that the double stranded configuration of DNA (dsDNA) can be transformed into two single strands upon heating the solution, many interesting features of DNA denaturation have been reported [34, 124, 131, 261, 262]. In most of these studies, the DNA is assumed to be in a diluted conditions. However, the core of the cell is having large number of biomolecules and the entire volume of the cell is occupied by the soluble as well as by non-soluble biomolecules. Thus the cell is called a crowded solution. One can not treat this as a concentrated solution because diverse kinds of molecules are present. Consequently the DNA melting and other important *in-vivo* cellular processes occurring in crowded environment are quite different [83–85]. Among all the activities, one of the very important activity is DNA melting or DNA unzipping. This physical crowding of the DNA environment in the cell can significantly alter the DNA melting process, and few experimental studies have examined crowding effects on DNA melting [86, 92, 93, 95, 263]. The expected change in the melting temperature T_m of DNA varies with the nature of these crowders. While the experimental results by Harve *et al* [88], shows a variation of 0.5-2.5 K in the increase of T_m, Nakano group reported that there is a decrease in T_m at high PEG concentrations [89]. The studies related with DNA melting are very important not only to understand the biological activities like replication, conformational stitches etc., but also to fabricate the bio-sensors or the drug design etc. [94, 96, 101, 264–266].

In the present work, we investigate the role of the molecular crowders on ther-

mal denaturation of DNA. The thermodynamics of the melting and the opening profile of base pairs thoroughly discussed in the presence of these crowders with the diversity in the crowder density and the crowder locations.

6.2 Modified Model Hamiltonian

We use PBD model which is discussed in details in Chapter 2. The model assumes the solvent as a background effect and it appears in the parameterization of the model parameters. As the PBD model is a quasi one dimensional model, the only fluctuation that is allowed to the base pairs is along the y direction. If the DNA is in dilute solution, the configurational space for the base pair will be from $-\infty$ to ∞. Presence of the crowders will restrict the movement of base pair along y direction [101]. This is a physical situation. How the crowders can be realized in a theoretical model? We model the crowders by the following arguments. The base pair that is surrounded by a crowder will need very high amount of energy to break the hydrogen bond. We modify the potential terms of the model to incorporate the presence of crowders and the inert nature of these molecules is assumed. As the crowders are nothing but the biomolecules which are not static but have constant motion in the space. The presence of crowders in the cell restricts the movement of base pairs and hence suppresses the thermal fluctuations. In addition to this, the thermal fluctuations in the DNA molecule (that is the system for us) increase

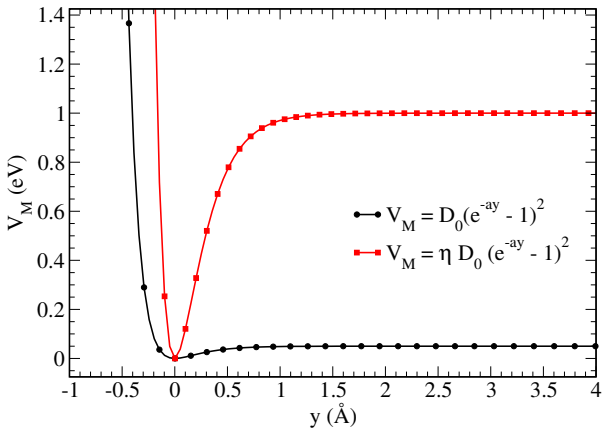


Figure 6.1: The on-site potential for the crowded and free sites. The fluctuation or movement of the bases in a pair will be restricted due to the presence of the crowder, hence the bases in a pair need higher energy to break the H-bonding.

the system pressure that may displace the crowder from its original position. Thus these crowders can not be treated like a *hard* wall, but should be treated as *soft* wall

which can be displaced or can be crossed. Mathematically, they can be as a high potential barrier. To the zeroth order approximation, we can argue that for the site where the crowder is present the base pair require very high amount of energy to overcome the potential barrier and hence to break the hydrogen bond. The hydrogen bonding between the bases in a pair is presented by the Morse potential in PBD model. Therefore we modify the potential barrier for the site where the crowder is present as shown in Figure 6.1. The Morse potential for crowded site is presented as :

$$V_M = \eta D_0 (e^{-ay} - 1)^2 \tag{6.1}$$

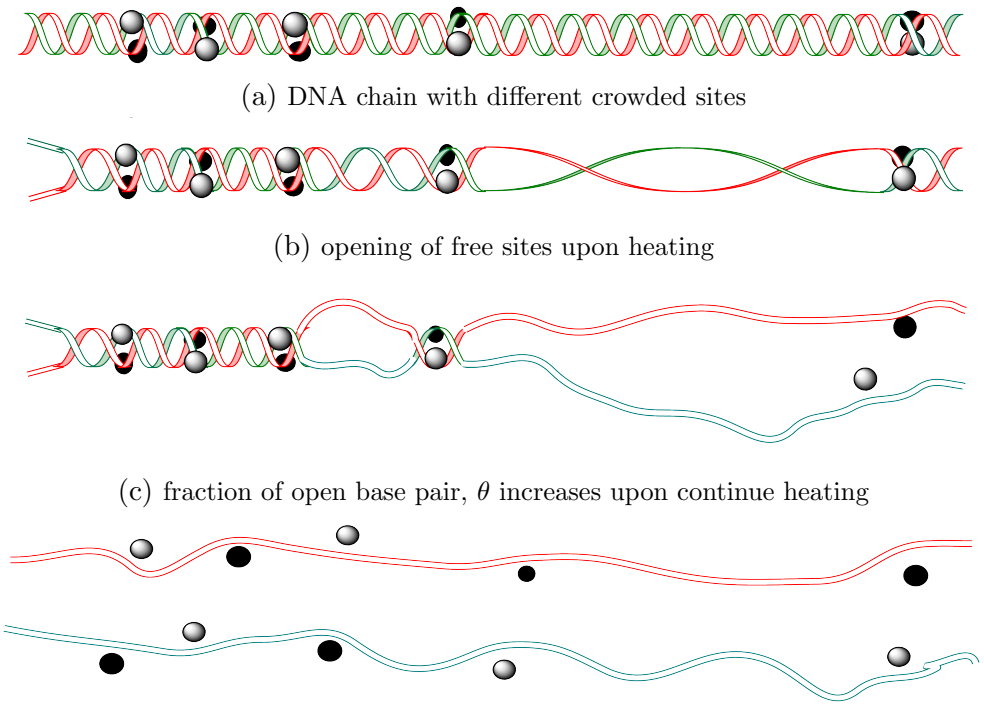
where, η is the scaling factor for the dissociation energy of the crowded base pairs which corresponds to the crowder strength. We evaluate the partition function with the help of Eq.2.10 and calculate the free energy per base pair as given in Eq.2.19.

6.3 Temperature induced transition

The presence of crowders in the cell restrict the movement of base pairs and hence suppresses the thermal fluctuations in nucleotides. We assume that some of the sites or base pairs in double helix DNA chain are crowded as shown in Figure 6.2. The threshold energy to overcome the H-bonding between the bases in a pair surrounded by the crowders is comparatively higher than the free site. With the increase in the thermal energy, system may transform from ordered state to disordered state. Hence, fraction of open base pair increases with increasing temperature. Since the thermal energy is also available to the crowders, the increase in thermal energy may displace the crowder from its original position. As a consequence, the site that is surrounded by crowder, may become a free site.

6.3.1 Melting temperature of DNA chain

We consider a homogeneous chain of 100 base pairs and using the modified form of model we investigate the microscopic details of the effect of crowders on the melting profile of DNA. The complete set of model parameters is given in Table 6.1. We vary the number of crowders in the chain to cover the whole range of crowder densities, we also alter the crowder locations for a fix crowder density to investigate melting profile of the DNA molecule. To study the effect of crowders on melting temperature we consider different cases of dsDNA chain. All these cases



(a) DNA chain with different crowded sites

(b) opening of free sites upon heating

(c) fraction of open base pair, θ increases upon continue heating

(d) complete separation of the chains

Figure 6.2: Graphical presentation of DNA double helix opening in crowded environment.

Table 6.1: The complete set of model parameters.

Parameters	**Values**
Potential Depth, D_0	0.05 eV
Inverse potential width, a	4.2 Å^{-1}
Anharmonicity, ρ	50.0
Anharmonic Range, b	0.35 Å^{-1}
Average chain stiffness, k	0.01 eV/Å^2
Opening cutoff, y_0	2.0 Å

can be broadly classified in three category.
(i) A chain in diluted solution,
(ii) A chain in presence of crowders that are present in a segment, *i.e.* sites 1-20
(iii) A chain in presence of crowders that are randomly distributed along the chain.
It is clear from the plots in Figure 6.3, there is considerable effect of crowders on the melting temperature of the chain. While there is an increase of $\sim$ 23 K in the T_m for case (ii), it is $\sim$ 68 K increase for case (iii). The crowders occupy the space, hence the accessible states decreases for the chain, and there will be less entropic

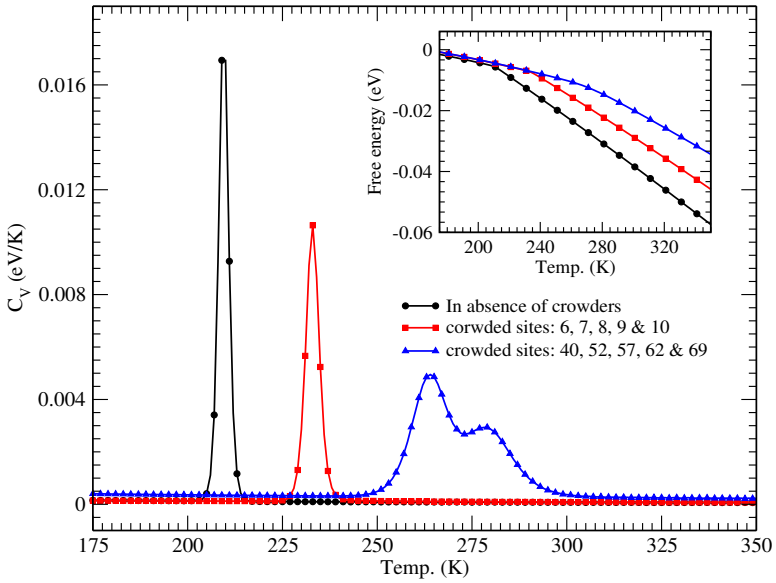


Figure 6.3: The specific heat as a function of temperature with $\eta = 20$. The three curves show the effect of crowders on the change in the melting temperature of the system. T_m = 210, 233 & 278 K respectively for diluted chain (black circle), 6-10 crowded sites (red square) & randomly crowded sites with 5% crowder density in the chain (blue triangle).

contribution in the opening of the chain. As a consequence, we get higher T_m in presence of the crowders. Interesting point to note that the only difference in case (ii) and case (iii) is the distribution of the crowded site. The number of crowders in both the cases is five. This means that the distribution of these crowders play an important role in the melting transition. If the same number of crowders are distributed along the chain, the system needs more nucleation sites. This results in higher melting temperature of the system. However, before reaching to any conclusion we need to investigate the role of different distributions of the crowders on the melting transition.

6.3.2 Fraction of open base pairs

It is described in Section 3.2, for long chains, the double strand is always a single macromolecule, and hence one needs to calculate the fraction of intact or broken base pairs only. The fraction of closed base pairs, θ_{int} is defined as:

$$\theta_{\text{int}} = \frac{1}{N}\sum_{n=1}^{N}\langle\vartheta(y_0 - y_n)\rangle \tag{6.2}$$

where $\vartheta(y)$ is Heaviside step function and the canonical average $\langle . \rangle$ is defined considering only the double strand configurations. The threshold value for y_0 is chosen as 2 Å. We calculate the fraction of open base pairs of a homogeneous chain of 100 base pairs in absence of molecular crowders. The chain is said to be denatured when 50% of the base pairs are open. The temperature corresponding to $\theta = 0.5$ is same as we get from the calculation of specific heat. The presence of crowders change the opening process of dsDNA. As the temperature of the system increases, there is an increase in the thermal activity of the molecule. However, the presence of the crowder on a particular site restrict the motion of the base pair. Thus the transition from dsDNA to ssDNA occurs at higher temperature. While the transition for dilute solution dsDNA the transition is smooth, the opening of crowded dsDNA has several pauses. We generate large number of distributions of

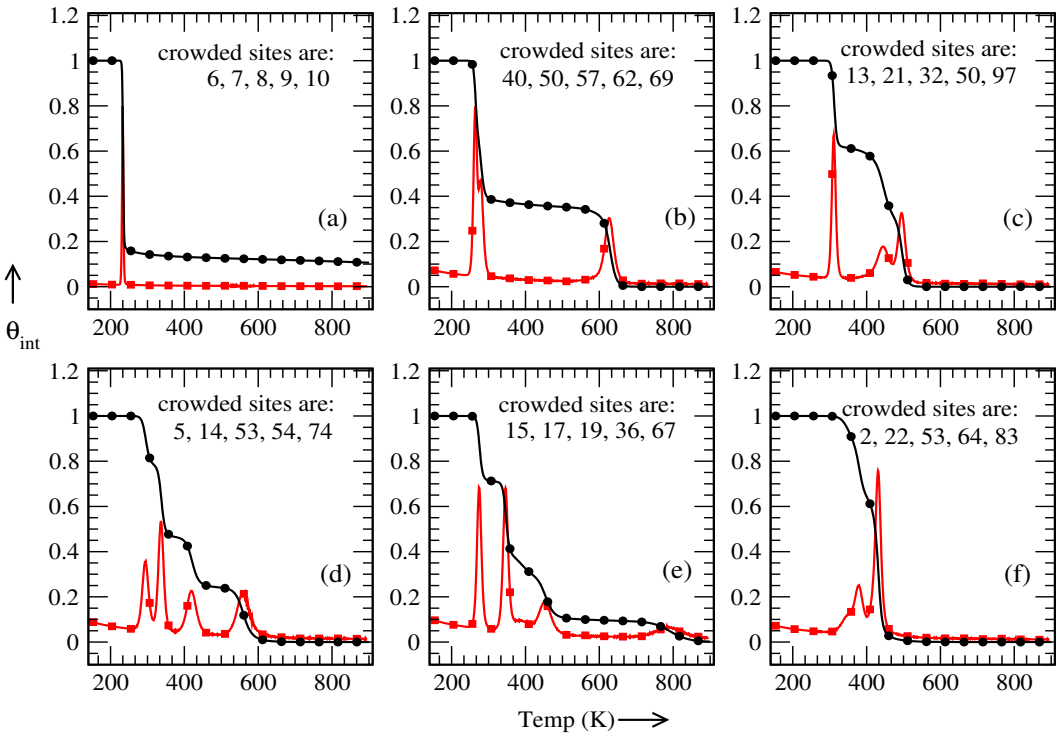


Figure 6.4: Figure shows variation in specific heat (red square) and fraction of intact base pairs θ (black circle) as a function of temperature. Figures (a)-(f) show the variation when different distribution of crowded sites in a homogeneous chain of 100 base pairs is considered. The specific heat values are scaled such that it can be shown in same figure.

crowder locations for a fix crowder density. To avoid excessive figures, we choose only six different cases to display in Figure 6.4. In all these cases, the distribution of the crowders along the chain is different keeping equal crowder density, *i.e.* 5%. Specific heat, C_v and fraction of closed pairs, θ plotted on a same graph as a

function of temperature. The value of C_v is scaled to compare the peak positions with change in θ. We observe some interesting feature of the transition from close to open state of DNA molecule, (as shown in Figure 6.4). Although the number of crowders and the composition of the chains are same, there is a striking difference in the transition. For a chain that is having 5 crowders on one of the ends *(6-10)* there is an increase in the T_m but the transition is smooth as shown in Figure 6.4(a). As the sites from *(6-10)* are crowded, these sites are bound to open, hence we get a pause in plot of fraction of intact base pairs at $\theta = 0.1$ (as the 0.1% of total base pairs are crowded or bound to open). For all other cases, the distribution of these crowded sites are random along the chain, we have multiple peaks in the specific heat and pauses in fraction of intact base pairs, θ.

Let us consider the distribution of crowded sites in case (c) as: *13, 21, 32, 50, 97* (Figure 6.4(c)). In this case, one of the crowded site *97* is located on one of the ends of the chain while the rest four crowders are on the other side. The thermal fluctuation will force the chain to open from one end where only one site has crowder. Thus, the highest peak corresponds to $\sim$ 310 K in the C_v and fraction of intact base pairs θ decreases sharply (1$\sim$ 0.6). However two more peaks appear, one at 445 K while other at 496 K. Because of the crowders which are covering/occupying the free space available to those crowded sites, there is a pause in the continues breaking of the base pairs. Therefore these peaks correspond to the further opening of the crowded sites in the chain at higher temperature. The peaks and the fraction of intact pairs shows that system need very high energy to separate the DNA strands. Similarly, for case (d), we observe four peaks. In this case, the crowders are located on *5, 14, 53, 54, 74*, *i.e.* they are distributed from one end to other with a concentration of two *53, 54*, in the middle. This opening shows the bubble formation in the sequence and the peaks are corresponding to these bubbles in the chain. The propagation of these bubbles is blocked due to presence of crowders in the sequence. Once the temperature cross the threshold energies of all the crowded sites all the base pairs in the chain open and we get a transition from dsDNA to ssDNA and θ goes to zero. For all the distributions as shown in Figure 6.4(b-f) we observe striking differences in the way of opening of the chain and it is on account of presence of crowder.

6.3.3 Opening profile of DNA molecule

To have the microscopic details of opening of DNA due to thermal fluctuations, we further calculate the probability of open pairs. The probability of opening of

the i^{th} pair, in a sequence is defined as [193]:

$$P_i = \frac{1}{Z_c} \int_{y_0}^{\infty} dy_i \exp\left[-\beta H(y_i, y_{i+1})\right] Z_i \tag{6.3}$$

where

$$Z_i = \int_{-\infty}^{\infty} \prod_{j=1, j\neq i}^{N} dy_j \exp\left[-\beta H(y_j, y_{j+1})\right] \tag{6.4}$$

while Z_c is the configurational part of the partition function defined as in Eq.2.10. The opening of individual base pair is more clear from density plot as a function of temperature, see Figure 6.5.

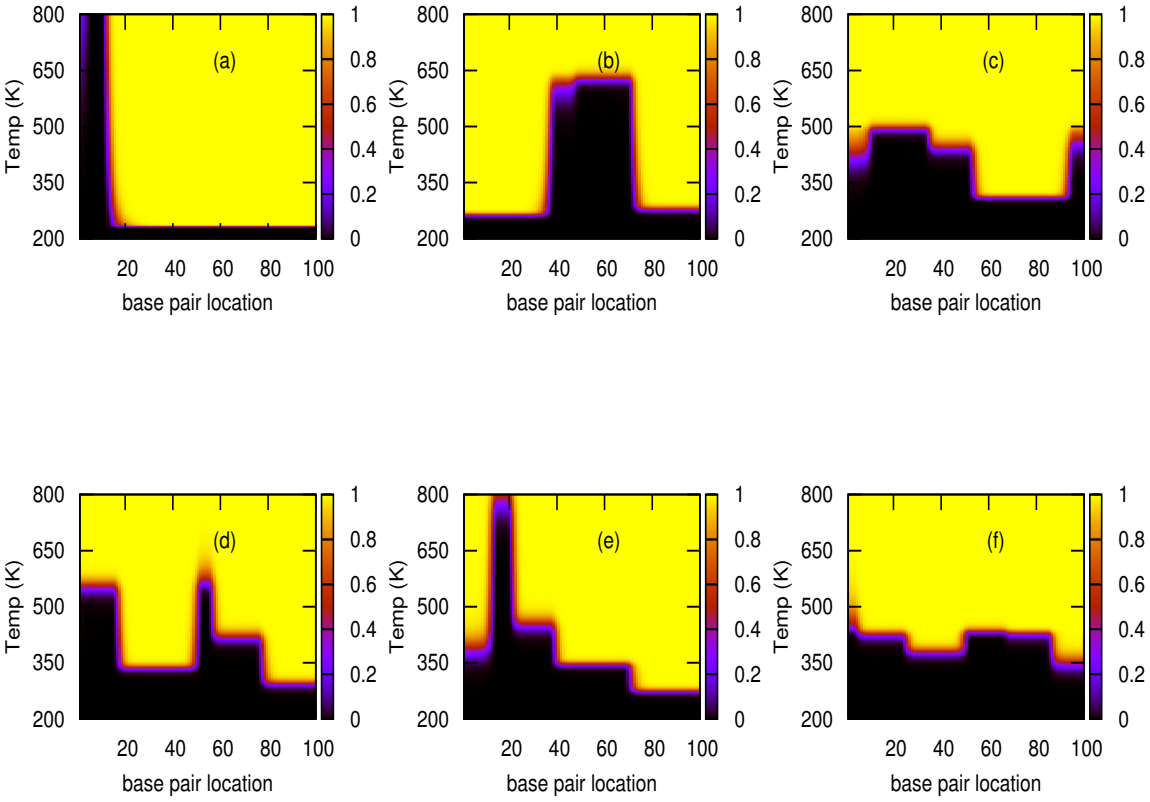


Figure 6.5: The opening profile of dsDNA in the presence of crowders. The site location or the base sequence is on x-axis while y-axis is temperature. The plots (a) to (f) follow the same distribution of crowders that is shown in Figure 6.4(a-f).

We calculate the opening profile of a chain of 100 base pairs with different positions of the crowders in the chain. The opening profile is shown in the density

plot in Figure 6.5. We present here the same 6 sequences that we discussed above. The site location or the sequence is shown on x-axis while y-axis displays the temperature. From the density plot the picture of opening of the sequence is very clear. The plots display different colors that show the fraction of open pair(s) as a function of temperature. When the crowders are dense somewhere in the middle of the chain, Figure 6.5(b), a big black patch shows intact pairs. This section of the chain, say *35-75*, has less degree of freedom, hence it prevents the complete opening of the chain. While the ends are open (as displayed by yellow color), the middle section of the chain is still bound. This means that even the crowders are not very close, there collective effect is enormous on opening. Similarly, for case(f), as the location of crowders are spanning over the chain from *2 to 83*, there is a spread of dark patch along the sequence. A segment of pairs at the last end & between crowded sites 22^{nd} and 53^{rd}, open initially and further increase in temperature will separate the two strands of the chain. This may be reason of getting two peaks of same height in the θ vs temperature plot for the same case. A small patch about *12-20* sites can be seen in Figure 6.5(e), because of the crowders at $15^{th}, 17^{th}$ and 19^{th} sites. The last end of this chain holds about 30% free pairs, therefore we get a peak in C_v which corresponds to $\theta = 0.7$ and the maximum probability at of opening in density plot. As we observe in all density plots, (a)-(d) the opening profile validates the multistage transition in presence of the crowders.

6.3.4 Influence of crowder density and locations

We found that the distribution of crowders play an important role in the melting transition. Now a genuine question arises, is there any correlation between the crowder density and the melting temperature of the molecule? We investigate now the effect of crowder density on the melting profile of DNA molecule by choosing a wide range of crowder densities varying from 3% to 30 %. We generate large number of distributions of crowder locations for different crowder density. We choose two different cases:

(i) presence of crowders in a segment of the chain

(ii) crowders are spanning over the chain.

Presence of crowders in a segment of the chain

We split the 100 base pair chain in five segment of 20 pairs. The crowder locations are randomly selected in each segment, and we calculate fraction of intact

pairs, θ. As per definition, T_m of the chain is the point in plot where $\theta = 0.5$, accordingly we find the melting temperature of the chain for each distribution. Table 6.2, represents the average T_m for different segments of the chain. When we

Table 6.2: Average T_m for crowders in a segment at different crowder densities with $\eta = 20$.

Crowder density (%)	$T_m(K)$ for crowders in a section				
	1-20	21-40	41-60	61-80	81-100
3	236.20	246.52	261.50	246.80	236.08
5	236.35	248.25	262.05	248.45	237.12
7	237.17	249.42	263.14	249.86	237.58
10	237.58	249.89	264.68	249.56	237.93
in absence of crowders, $T_m = 210.60$ K					
when 3′-end is restricted, $T_m = 231.50$ K .					

compare the values for T_m at a fix density in the Table 6.2, we find the presence of of crowders in a segment increases the T_m and holding higher T_m for crowders at the middle segment. This can be explained by the contribution of the finite end entropy. While the crowders are at the ends, the chain opens from the other end

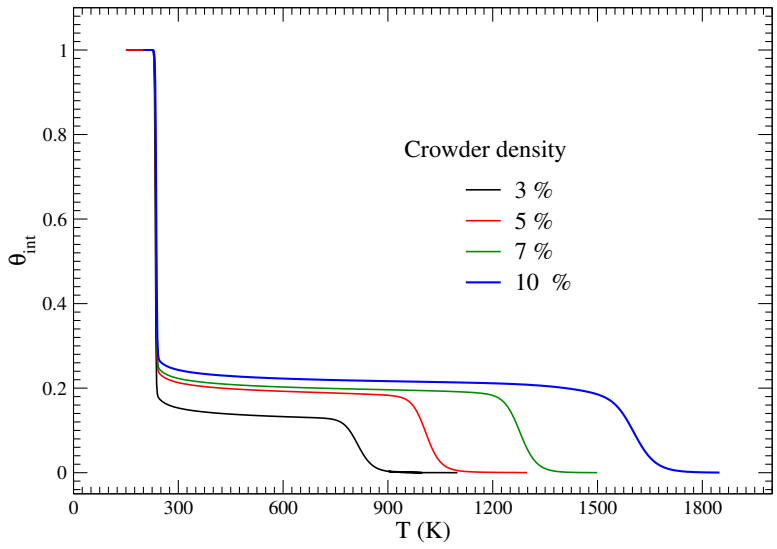


Figure 6.6: Variation in θ_{int} and C_v as function of temperature for different crowder densities in a segment of 1-20 pairs. Again we scaled the y-axis for C_v to compare it with θ

which has higher entropy, crowders at the middle segment restrict the propagation

of opening from the ends, it will increase the T_m. As we are considering the homogeneous chain, the first (1-20) and last (81-100) segments have same behavior in presence of crowders. In these two cases, the chain will open from either of the ends and this is validated from the end restricted chain which has the melting temperature, $T_m = 231$ K. The observable thing is, increase in crowder density for a segment, does not change the T_m significantly (the maximum difference is $\sim$ 2.5K). This clearly indicates that if the crowders are confined in a particular segment of the chain, the chain is forced to open from the ends. To get a clearer picture, we plot θ for a segment of 1-20 pairs at different crowder densities in Figure 6.6. The crowders are located at the 3′-end of the chain, more than 50% of pairs are free at the 5′-end, the opening of pairs initiated from the other end, hence we get a sharp peak in C_v and sharp decrease in θ. increase in crowder density only effects the complete opening of the chain as shown in Figure 6.6.

Random distribution of crowders over the chain

In Figure 6.3, we found the distribution of crowders play an important role in the melting transition. We have done a systematic quantitative and study to cover the whole possible range of crowder densities and locations in the chain. We randomly generate the crowded sites and calculate the θ for that sequence. We note values of T_m from plot of θ. Table 6.3, represents the average T_m for different density of the crowders in the chain. It is known that at a given density, a larger

Table 6.3: Average melting temperature (in Kelvin) for different crowder densities with $\eta = 20$. For each density we generate 50 random distributions.

Crowder density (%)	T_m^{avg} (K)	T_m^{min} (K)	T_m^{max} (K)
3	308.48	247.50	379.40
5	381.76	279.00	477.50
7	480.20	355.20	627.36
10	599.45	425.15	765.28
14	761.35	639.65	922.87
20	996.28	902.35	1128.15
30	1331.58	1152.00	1454.60

crowder exhibits a greater suppression of available space to DNA molecule. We consider the crowder strength via the potential height in Morse potential as shown

in Figure 6.1. As the crowder strength increases, we take the higher potential height. Again we calculate the T_m, for different crowder strengths for the crowder densities that are shown in Table 6.3. For an average value we generate 50 different distributions for each crowder strength and density. Now we have sufficient arguments and data to explore any power law behavior in the melting temperature as a function of crowder density. We plot the average melting temperature for different crowder densities at different potential heights which corresponds to the strength of the crowder (as shown in Figure 6.7). We found the T_m increases monotonically

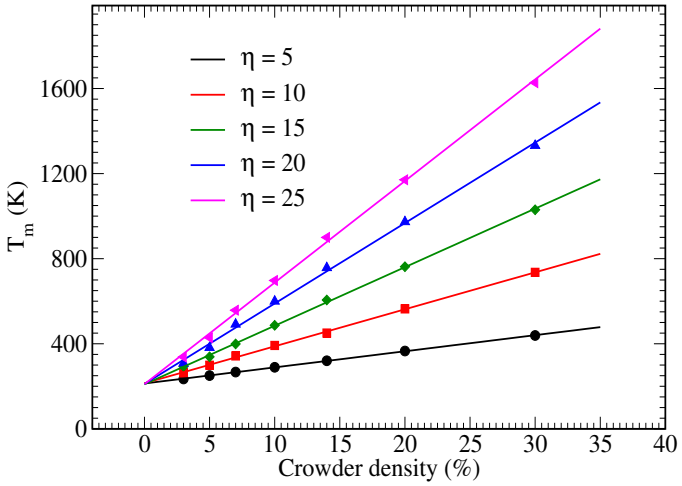


Figure 6.7: The variation in the average melting temperature of the chain as function of crowder density for different values of potential heights η. The points are corresponding to T_m, while the solid lines are best fit for these points at a particular scaled potential. The melting temperature of free chain is 210 K.

as the crowder density increases in the solution which is in agreement with the results of Y Liu *et al.* [101]. From the intercept of the fitted lines we found the melting temperature of the free chain (in absence of any crowder) is ∼210 K which is very close to the direct calculated value of T_m = 210K. Similar kind of power law behavior for the melting of ubiquitin protein has been shown by Waegele and Gai [103]. Our power law is little bit different from their results. This difference is obvious as melting or denaturation of proteins and DNA is quite different. Although there is a hydrogen bonding between the monomers in both the cases, the geometry of coil and helix conformations in both the cases are different [103, 267]

6.4 Conclusions

We have studied the effect of crowders on the melting profile of homogeneous ds-DNA chain. A crowded environment produces a marked stabilizing effect on double helix DNA structures. Since the cell contains a vast array of macromolecules that can exclude a volume fraction, it is likely that the physical and chemical behavior of DNA within this crowded space will be affected. With suitable modifications in the PBD model for the sites where crowders are present we show how these crowders affect the melting profile of DNA, the inert nature of the crowders is assumed.

The data and analysis presented in this chapter, confirm that in addition to fundamental interactions between base pairs, conditions in solution, *i.e.* presence of crowders, play an important role in the stability of dsDNA molecule. While the DNA transform from double strand configuration to single strand configuration smoothly in dilute solution, the transition is showing multiple peaks of different heights in case of crowded solution. The density profile clearly show how these crowders block the propagation of bubbles that are created due to thermal fluctuations. This is attributed to the volume occupied by crowder molecules, which inhibits the entropy increase necessary for DNA melting. While for a homogeneous chain, we have probably one nucleation site, for the crowded DNA chain, we may have multiple nucleation sites that require high amount of energy to transform the stable state of DNA.

The predominant out come of the study is the influence of crowder locations, spanning over the chain on the opening of DNA strands. We examined different cases where the distribution of crowders along the chain are different. The dsDNA helix is more stable for randomly distributed crowders than the crowders which are present in a segment. The entropic contribution for randomly crowded sequence is less and discontinues, while the crowders in a segment will suppress only segment's entropy and the opening of the free sites will pursue the separation of 50% pairs. By varying the crowder density and there distributions we show that the melting temperature increases monotonically with crowder density. We also found that at a given density, a larger crowder exhibits a greater suppression for DNA melting and hence a higher T_m, as reported by Y Liu *et al.* [101].

Although the temperature changes occurring *in vitro* seem to be smaller than those observed in this work, the results display interesting features of opening of DNA molecule in crowded environment. This is an initial attempt to understand the complex dynamics of the melting transition of the dsDNA molecule in

molecular crowded environment through the PBD model.

Chapter 7

Critical length in mechanical unzipping of DNA

7.1 Introduction

DNA unzipping, the separation of double helix into single strands, is crucial in analyzing many biophysical systems. Although the large-scale separation of double-stranded DNA has been studied with a variety of theoretical and experimental techniques [131, 142, 170–172, 228, 268, 269], the minute details of the very first steps of unzipping are still unclear. The unzipping or breaking of hydrogen bonds in dsDNA reveals many interesting features that depend whether the force is applied along the axis of the molecule [136, 142, 195, 270–272] or the force is applied perpendicular to the helical axis [133, 138, 168, 172, 193]. When a force is applied perpendicular to the helical axis, bases are sequentially stretched during the unzipping. On the other hand, in the process of unzipping due to shear force as shown in Figure1.10, the stretching is spread over many base pairs.

All these studies in both the approaches, revealed that not only the magnitude of the force but also the location and the direction of applied force, play pivotal role in the unzipping of dsDNA. At zero temperature, deGennes [273] modeled the infinite DNA chain by a ladder and calculated the shear forces to separate the two strands of DNA. One of the interesting outcome of his calculations was the saturation in the value of critical force. His calculations showed different response for the short (or finite) chain and long (or infinite chain). The shear stress relaxes over a distance (number of base pairs) as a function of the elastic constants of the chain on either side of the chain. The harmonic spring model was later modified by Chakraborti and Nelson [274] by representing the interaction between the bases by

Lennard-Jones potential. They showed that the strain, that is needed to unzip the chain, is localized over a narrow range on either side of the chain. Using Langevin dynamics simulation [142, 195], extended PBD model [275] and coarse-grained modeling [197] several workers found similar kind of trend at finite temperature. In an experiment [171, 276] the critical forces for several dsDNA molecule of 12 to 50 base pairs are measured at room temperature. They found a linear increase in the value of critical force with the number of base pairs in a sequence. The surprising feature of their observation is the saturation in this value around 20 base pairs. The critical force approaches an asymptotic value beyond this number.

In this chapter, we discuss the force induced unzipping of dsDNA when force is applied in the perpendicular direction to the axis of the molecule. We calculate the value of critical force, F_c for chains that varies in the number of base pairs. For our investigation, we adopt the PBD model, [76, 122, 225, 253] which is discussed in Chapter 2.

7.2 Chain pulled from an end

Here we consider heterogeneous DNA molecules and investigate the effect of chain length on the forced induced unzipping of the molecule. This force calculations are done in constant force ensemble (CFE). The modified Hamiltonian of the system is as:

$$H_f = \sum_n \left[\frac{p_n^2}{2m} + W(y_n, y_{n+1}) + V(y_n) \right] - F \cdot y_n \tag{7.1}$$

where F is the applied force at the n^{th} base pair in the chain and y_n is the separation of the bases from the equilibrium. For current study, we use the modified on-site potential, $V_M(y)$ same as we have used in Chapter 3. We also choose three different DNA chains which we have considered in Chapter 3. These sequences vary in terms of the fraction of GC & AT base pairs. We call them as 30% GC, 50% GC and 75% GC chains. The chain sequences are,
(a) 5'-TGATTCTACCTATGTGATTT-3' (30% GC)
(b) 5'-TACTTCCAGTGCTCAGCGTA-3' (50% GC)
(c) 5'-GTGGTGGGCCGTGCGCTCTG-3' (75% GC)
To make chains of larger length we repeated the above sequences. The force response of the short and long chains are different because of the entropy contribution from the ends. This part of the study is aimed to find out the length of the chain which can discriminate the short and long chain limits. For this we

choose two different cases: (i) when one of the ends is restricted and other end is open, (ii) both the ends are open. First we calculate the melting temperature, T_m of the chain by evaluating the free energy and specific heat for all the three chains using the model parameters listed in Table 7.1. The calculated specific

Table 7.1: The complete set of model parameters.

Parameters	**Values**
Potential height, D	0.06 eV (AT) & 0.09 eV (GC)
Inverse of potential width, a	4.2 (AT) & 6.2 (GC)
Chain stiffness, k	0.01 eV/Å^2
Anharmonicity, ρ	1.0
Anharmonic range, b	0.35 Å^{-1}
Solution constant, λ	$\lambda_1 = 0.01$ & $\lambda_2 = 0.002$
Solution constant, γ	1.0
Salt concentration, C	0.621 M
Opening cutoff, y_0	2.0 Å

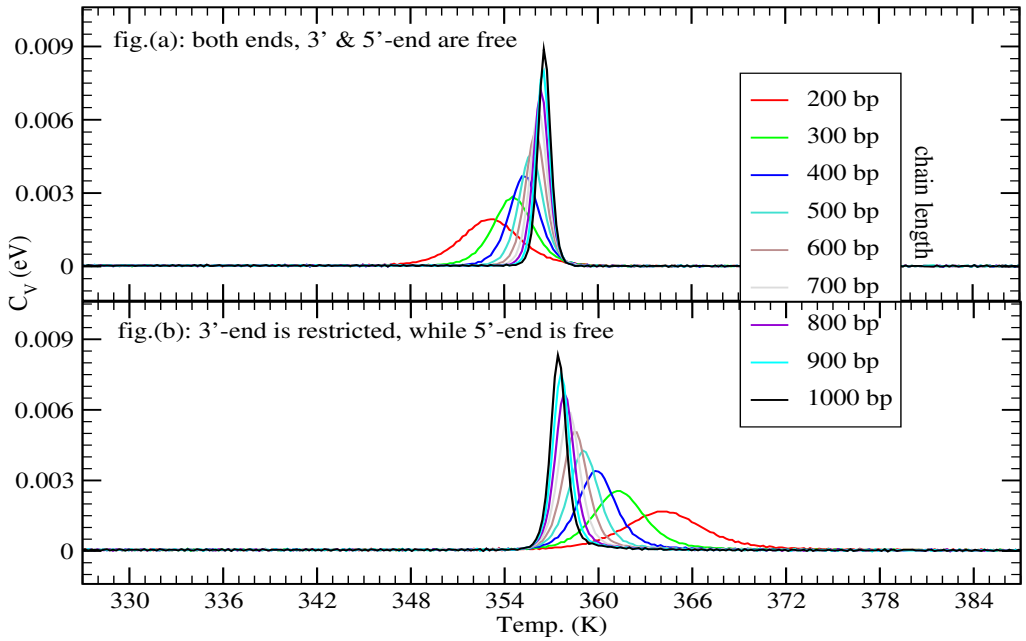


Figure 7.1: Variation in specific heat with increasing temperature at different chain lengths for the two cases: (a) both ends, 3′ & 5′ are free, (b) one end, 3′ is restricted and the other end, 5′ is free for 50% GC chain.

heat is plotted with increasing temperature for different chain lengths, as shown

in Figure 7.1. We consider two cases; either both the ends of the chain are free or one of the ends is restricted. The melting temperature, T_m for different chain lengths in these two cases is calculated through the peak in specific heat. How the melting temperature of the DNA molecule varies with the chain length is shown in Figure 7.2. When both the ends of the chain are free, the T_m increases with the

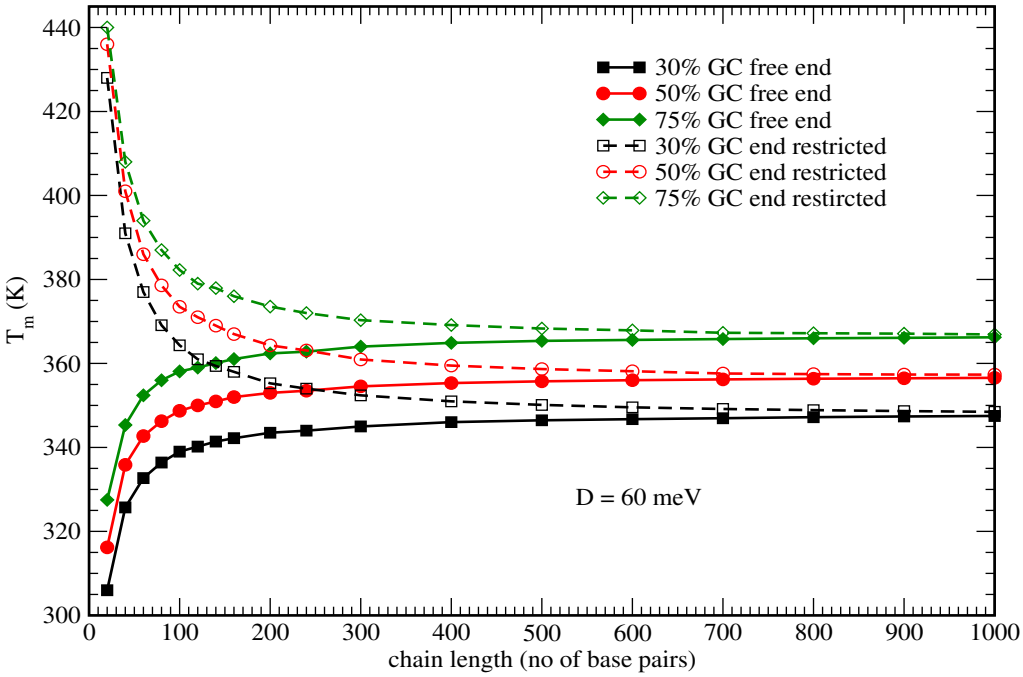


Figure 7.2: Variation in melting temperature with increasing chain length for two cases: (i) both ends, 3′ & 5′ are free (solid lines), (ii) one end, 3′ is restricted and the other end, 5′ is free (dashed lines), for all three 30%GC, 50%GC & 75%GC chains [277].

length. For the chain with one of the end restricted T_m decreases when the chain length increases. The prime factor for the contrasting behaviour of the chain is the end entropies in the stability of short chains. When the end is restricted, the contribution of the entropy of this end in opening, is suppressed that leads to the stability of the molecule. In addition to this, the restricted end exerts a binding force to the molecule which resists the opening. When the chain length increases the ends are far and the loop entropies play more important role in the opening of the chain. The loops are created due to random fluctuations of the weaker segment of the chain. For longer chains, we find that there is no change in the melting temperature of the chain with open ends or chain with one of close ends. As reflected from the graph, a molecule having 600 base pairs can be considered as a chain which is unaffected by the status of the ends. This length we call as infinite chain limit length, *i.e.* chain that is having 600 or more base pairs can be

considered as infinite chain.

Next we investigate the mechanical unzipping of DNA in constant force ensemble at 300 K. We evaluate the critical force of the chains with varying number of base pairs for two cases; (i) both the ends are open and the force is applied at 5′-end. (ii) when one of the ends, 3′ is restricted while force is applied on the other end, at 5′ end.

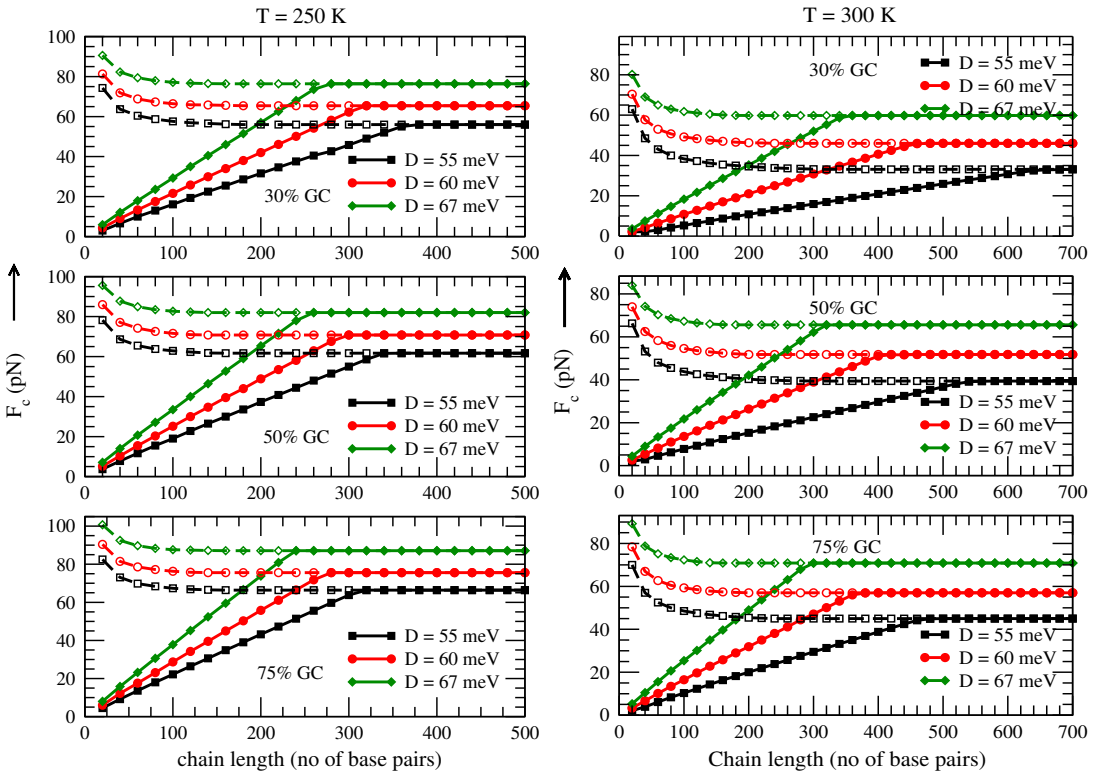


Figure 7.3: Variation in critical force as a function of chain length at T = 250 K (figures on left) & 300 K (figures on right) for two cases: (i) both ends are free (solid lines), (ii) one end is restricted (dashed lines), for all three 30%GC, 50%GC & 75%GC chains. For a chain we also vary the potential height in Morse potential [277].

To study the role of thermal fluctuations in achieving the infinite chain limit, we investigate all the three chains at two different temperatures T = 250K & T = 300K. Again the behaviour of short chains and long chains are different as shown in Figure 7.3. Beyond a certain chain length, the critical force for the two different cases is same. When the end is restricted, it exerts a force which resists the opening from the other end. This is more pronounced in case when chain is of shorter length. However, the effect diminishes as the chain length increases. It is because of the fact that, when a force is applied on an end, there is a length up-to which the effect of applied force sustains. After that the base pairs have no

information about the applied force. As the force applied on an end is increases, more number of base pairs are in open state which increases the entropy of the system. The increase in entropy, drive the system in disordered state. However, the binding force on the other end, in case of restricted chain, increases the enthalpy of the system. Hence the system has to overcome a large barrier to open the chain. With the increase in the chain length these two forces diminishes along the chain. To verify the argument, we considered the chains of same GC content but vary the value of potential depth, D. For the larger value of D, the chain will be more stronger. We find that, for a chain with higher potential depth, the infinite chain limit reduces. For example, for 30% GC chain, one need about 650 base pairs chain to be considered as infinite chain for $D = 55$ meV while for $D = 67$ meV, we get this limit at 375 base pairs.

7.3 Chain pulled from interior of the chain

If the DNA chain is pulled from any location other than its ends, how the response of the dsDNA will be differed is the question we discuss in this section. We consider homogeneous chains of different length and apply the force on different locations in the chain. Again we calculate the critical force in constant force ensemble. For this study we adopt the model Hamiltonian, H which is discussed in Chapter 2. The model parameters that are used for the calculations are shown in Table 7.2.

Table 7.2: The complete set of model parameters.

Parameters	**Values**
Potential Depth, D	0.07 eV
Inverse potential width, a	4.2 Å^{-1}
Anharmonicity, ρ	1.0
Anharmonic range, b	0.35 Å^{-1}
Average chain stiffness, k	0.01 eV/Å^2

We consider a dsDNA chain of 400 base pairs of homogeneous sequence. When we apply force on one of the ends, the chain opens as **Y**-fork. This feature has been discussed in past in detail by several workers [191, 192]. We then apply the force on various sites along the chain and reach on the middle of the chain (200 base pairs). Till this point, we observe slight variation in critical force, when force

is applied on different sites along the chain. The critical force that is needed to open the chain from one end is 21.5 pN and increases as we move along the chain as 23 pN for the case when force is applied on 50^{th} site, 29 pN for 100^{th} site, and reaches 30 pN for 200^{th} site. When we apply the force at mid, it propagates bidirectionally in the chain, hence the critical force at mid, F_c^{mid} should be twice of the critical force at the end, F_c^{end}. For the chain of 400 bps, the value of F_c^{mid} is less than the twice of the F_c^{end}, Figure 7.4(a). Assuming that this may be finite size effect, we then further increase the system size. We found that as the system size reaches 600 base pairs, the variation in the critical force for a site is significant. While the critical force at the end is still 21.5 pN, it is 23 pN for 50^{th}, 29 pN for 100^{th}, 36.5 pN for 200^{th} and 42.5 pN for 300^{th} site! For this chain length we are

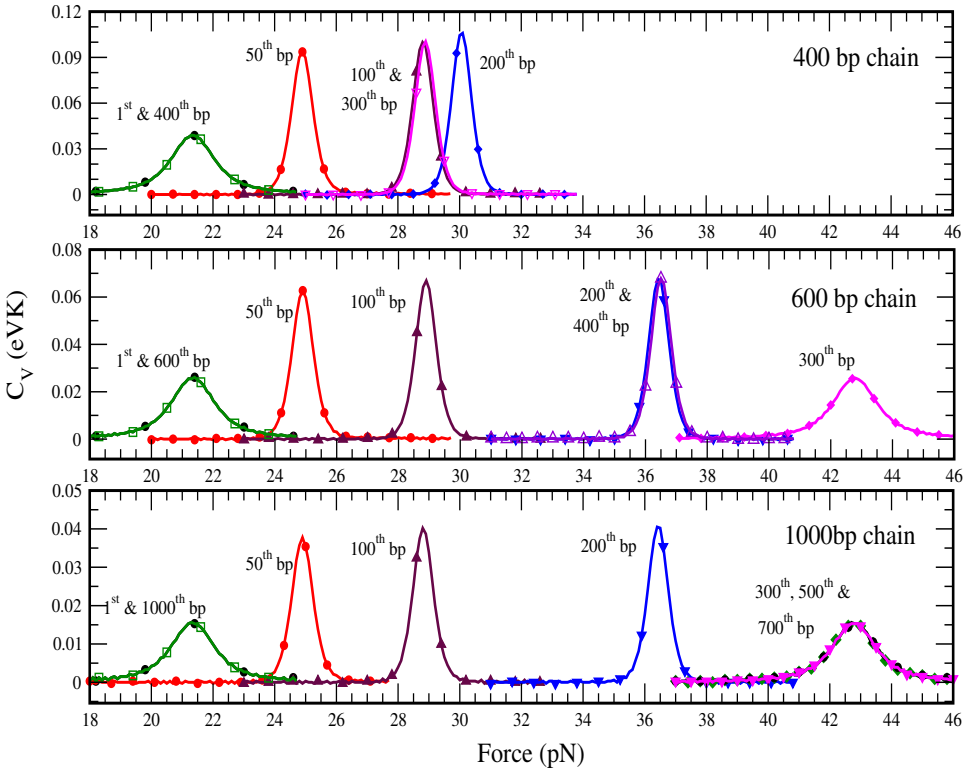


Figure 7.4: The specific heat as a function of applied force at different locations in the chain. The distribution of the critical forces (peak in specific heat) is shown along the chain for finite and infinite chains.

now getting critical force at the mid point twice that of the critical force at the ends, *i.e.* $F_c^{\text{mid}} = 2 * F_c^{\text{end}}$. As we move further along the chain, the critical force decreases from 42.5 pN to 21.5 pN as shown in Figure 7.4(b). The increase in the force value from 1^{st} to 300^{th} and decrease from 300^{th} to 600^{th} is found to be linear and of same slope. The critical force as a function of site location is perfectly

symmetric about the mid-point of the chain as shown in Figure 7.5. Thus, one can say that for PBD model with these set of parameters, the chain that can be assumed to behave as an infinite chain is for $N = 600$.

How the transition will be when one take system size $N > 600$? To investigate further, we increase the chain length from 600 to 700, 1000. We found some interesting transition here. No matter what is the length of the chain, $F_c^i = 2 * F_c^{\text{end}}$ at $i = 300$ and after that force get saturated. This saturation sustains till we reach the site position $(N - 300)$, where $N = 700, 1000$. After this position, the F_c again decreases and we get the same force to unzip the DNA from the other end. Again the slop of the curve on both side of the plateau is same and linear, see Figure 7.5. We have evaluated the F_c as a function of N for 2000 & 3000 base pair chain also and found that there is no change in the point where the relation $F_c^i = 2 * F_c^{\text{end}}$ holds, however, the length of the plateau region increases, and the length of the plateau region is $(N - 600)$, where N is the number of base pairs in the chain. This simply means that there is a region in a dsDNA chain where the force required to unzip the chain, is twice of that is required to unzip from the ends. This kind of behaviour has been observed by deGennes and other workers in the case of shear unzipping of the dsDNA chain.

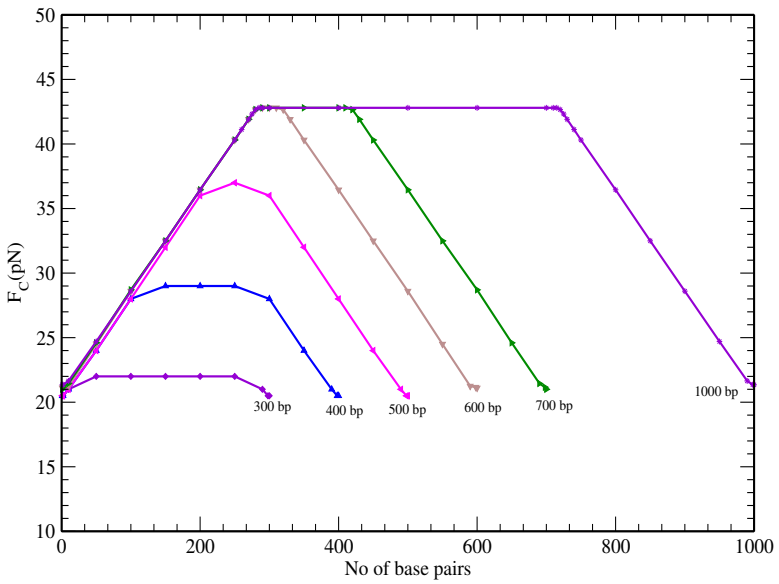


Figure 7.5: Values of the critical forces along the chain for increasing chain lengths when force is applied at the different locations in the chain. The points in the plot correspond to the point of force location in the chain.

In order to investigate more details of the distribution of forces along the chain we apply force on the different and random locations along the chain. We distribute

the applied force like, the force that is applied on 1st pair is now distributed over 25 sites. This distribution we have done in two different ways. (i) In a stretch we select 25 sites and the force is randomly distributed over these 25 sites. (ii) In a stretch we select 25 sites and the force is equally distributed over these 25 sites. The condition we applied that in all the two cases, $\sum_{i=1}^{25} f_i = f_e$, where f_e is the one site force that is applied on a particular site or end. Again, we consider three

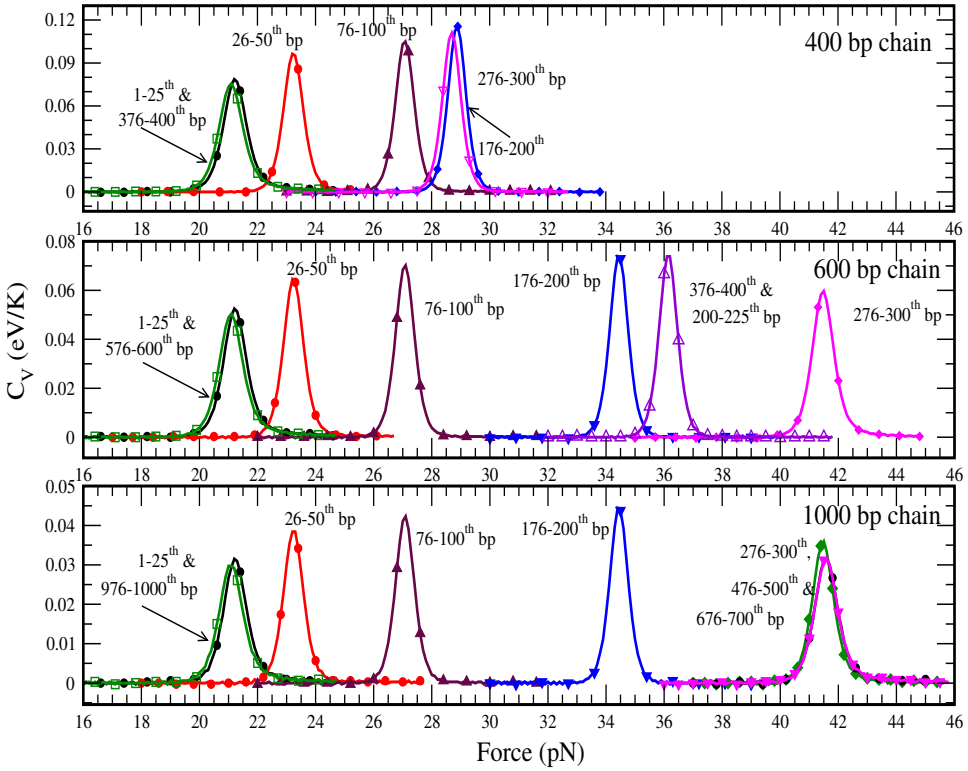


Figure 7.6: The distribution of the critical forces along the chain for short and long chains. The force is applied to the chain in a section of 25 base pairs and force is randomly distributed over these 25 sites.

chains that was discussed earlier, 400, 600 & 1000 base pairs chain. For the 400 base pair chain, the critical force that is required in the 1-25 segment is nearly same as the critical force, that is required to unzip the chain, in the 376-400 segment. These two segments can be considered as the end segments. The critical force in the 25-50 and 76-100 segment increases by the same amount as they were increased for the one site force. The critical forces when chain is pulled from end segment is 21 pN, while it is 23.5 pN for 25-50 segment and 27 pN for 76-100 segment. Whether the force is applied on a particular site or that force is distributed over a length, the same amount of force is required to unzip the chain. Similarly, the critical force to unzip the chain from 200th site was 30 pN, for the segment 176-200,

we get 29 pN. However, there is a slight variation in the force of about 1 pN in the middle of the chain. If we compare the variation for infinite chain, *i.e.* the 600 and 1000 base pairs chain, there is no significant change observed till 100^{th} site or segment. If we closely look the segments, 176-200 and 200-225, there is a difference in the critical force of 1.5 pN (34.5-36 pN). The reason for this deviation is the variation in the force that is required to unzip the chain from 200^{th} site to 225^{th} site as shown in Figure 7.6, while this was almost constant for finite (400 bps) chain.

The comparison of the critical forces for the infinite chain of 1000 bps as a function of applied force location is shown in Figure 7.7. From the figure it is clear that the force required to unzip the chain from one point (site) is slightly higher than the case when the same force is randomly distributed over a length of 25 sites in the plateau region. While is slightly lower than the case when the same force is equally distributed over a length of 25 sites in the plateau region, as shown in Figure 7.7.

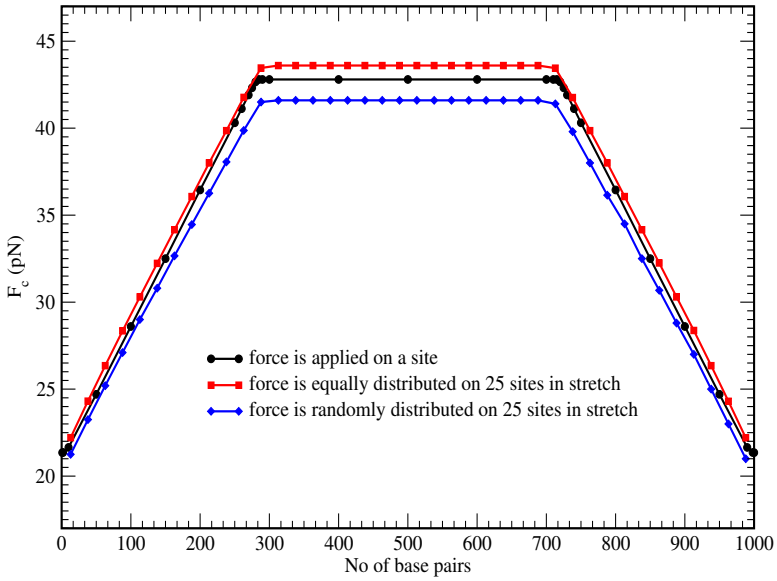


Figure 7.7: Values of the critical forces along applied force locations for long chains. The points correspond to the applied force locations. When force is applied on 25 sites, the point is shown by the mid number of that section.

We also compare the other possible cases to unzip an infinite chain. The cases and the critical force corresponds to these cases are shown in Table 7.3. We found that the critical force varies significantly in all the these cases. The reason for this change is the distribution of the force. As the force induced unzipping if more due to change in enthalpy of the system than the change in the entropy of the system,

one need higher amount of force when force is equally applied on the complete chain. Also the nucleation energy for the randomly distributed force case will be higher than for the case when force is equally distributed.

Table 7.3: Critical force of an infinite chain for different cases

Force applied as	**Critical force**
force is applied at a point or site	42.3 pN
force is randomly distributed on 25 sites in a section	41.4 pN
force is equally distributed on 25 sites in a section	43.2 pN
force is randomly distributed over the chain	66.7 pN
force is equally distributed over the chain	78.9 pN

7.4 Conclusions

We have investigated the effect of chain length on mechanical stability of dsDNA molecules. We adopt different approaches to pull the DNA molecule, either pull the end base pair or pull the base pairs from the interior of the chain. The critical force is calculated for free chain and the end restricted chain. We considered different heterogeneous chains of different lengths and calculated the melting temperature as well as the critical force. We found a minimum length where the end entropy or the chain length does not contribute significantly in the opening of chain. In case of a chain having unrestricted ends the critical force and the melting temperature T_m , increase with increasing chain length. This is due to the free end entropy that acts as an additional force on the chain; while critical force and T_m for the end restricted chain, decreases with increasing chain length. As the restricted end applies a binding force to the chain, it is quite natural that as the chain length increases the end where force is applied cannot experience the force applied by the restricted end on the other side of the chain. At the point where the chain becomes infinite the end effect nullifies and the critical force in both the cases is found to be same. We found that as the chain is stronger (high value of dissociation energy or more GC content) the critical force is more and it saturates at lower length in chain.

When the chain is pulled from interior section or mid of the chain, we observed interesting features of the opening. For this case also the end entropy plays an

important role in the opening of the chain. We found a length beyond that the end entropy does not contribute to the applied force which propagates towards both the ends of the chain. At selected model parameters we found that a homogeneous chain of 600 base pairs behaves as an infinite chain where the $F_c^{\rm mid}$ is twice of the $F_c^{\rm end}$. If the force is applied at either of the ends, it propagates up-to 600 bps while it propagates up-to 300 bps when the force is applied at the mid. When the chain length is less than this limit, the end entropy contributes to the applied force and hence the critical force decreases in both the cases. While for the chain length more than this limit, the critical force remains constant and we get a plateau region in force-length phase diagram.

The distribution of the applied force significantly affects the opening or unzipping of DNA. As the force induced unzipping if more due to change in enthalpy of the system than the change in the entropy of the system, one needs higher force when force is equally applied on the complete chain. Also the nucleation energy for the randomly distributed force case will be higher than the case when force is equally distributed. It is possible to extend these studies in understanding the mechanism involved in intra-molecular processes such as loop propagation in sequences, ligand binding at molecular level and DNA protein interactions. The molecular dynamics can provide the life time of these interactions. At this stage, we must point out that though the model ignores the semi-microscopic detail of dsDNA, e.g., orientation and inclination of base-pairs, helical structure of dsDNA, etc., even if, it captures some essential physics of the unzipping mechanism of DNA. It would be interesting to perform all-atom simulations, where the above shortcomings of the model can be avoided, and one can get a better description of unzipping mechanism.

Conclusions and Future Scope

The response of double stranded DNA molecule under different physiological conditions is a subject of vital interest. As a system dsDNA is complex yet an interesting candidate which offers many challenges to unfold. For example the opening of the molecule during transition and replication process is yet to be understood. Scientists and researchers have partial success in revealing some features of this process by some beautiful and sophisticated experiments followed by interesting theoretical and computational results. In the current work, our aim was to understand the response of dsDNA molecule in concentrated as well as in crowded solutions.

Using a theoretical model, we attempt to understand the denaturation of long as well as short DNA molecules in different environments. We conclude our findings point wise as:

- Role of cations, that are present in solution in form of Na^+ or Mg^{2+} to stabilize the double helical structure of DNA: As the two strands of DNA molecule are negatively charged, these cations act as an shielding agents to reduce the repulsion between the two strands. Different concentration of these cations affect the stability of DNA molecule in a non-linear manner. In general, the concentrations of these cations *in vivo* are constant, however, it has been observed that when cancer cells are in action, the concentrations of these cations vary. In most of the theoretical results it has been observed that the melting temperature, T_m varies logarithmically with the concentration of cations in the solution. By comparing our results in thermal as well as force ensemble we found that the dependence is not purely logarithmic. In addition to this, we investigate the role of sequence of DNA molecule in the stability of the molecule whose solution condition is not constant. Our studies reveal that the dominant player in the stability of the molecule is the hydrogen bonds. This may be considered as a little attempt to understand the role of cations in the stability of this complex molecule.

- We have studied the role of defect(s) on the thermal as well as mechanical denaturation of DNA molecule. It known that the defects or mismatch in the pair delay the replication process which may further cause the cell death and hence may lead to initiation of cancer. We considered four different kind of DNA molecules that are having different numbers of AT and GC pairs and the distribution of these pairs along the chain is also different. We have considered all the chains with m number of defects, where m varies from 1-4 and for $m > 1$ all the defects are in a block. We found that there is a segment (4-12) of the chain where T_m is unaffected by the location of the defect in the chain. In case of heterogeneous chain, there was no plateau but it matters on a location whether there was an AT pair or a GC pair. When we compared the opening in two ensembles for homogeneous chain we observed that there is a striking difference. While for the homogeneous chain we obtained a symmetric necklace kind of plot in thermal ensemble, this was missing in force ensemble. This validates the role of finite end entropy of the homogeneous chain in the denaturation of the DNA molecule. For the thermal melting the ends have less impact on the opening because of the fact that each base pair shares the same amount of thermal energy. There only the sequence of AT/GC pairs matters. For the chain that is pulled from an end by some force, it is important for all kinds of chains (with defect) whether the force is applied on $3'$-end or $5'$-end.

- The molecular crowders present in the solution affect the dynamics of the DNA molecule. It has been found that around 5-20% of the cell is crowded with nucleic acids, proteins, saccharides etc. In compare to the diluted solution, the free volume for DNA in crowded solution in less. This restrict the thermal fluctuations/movements of base pairs, which in result, make molecule more stable. The prime source of stability in DNA is represented by hydrogen bonding in PBD model. Hence we represent the crowders as potential barrier for the bases in a pair in the model where the inert and static nature of the crowders are assumed. The denaturation of DNA in presence of crowders shows multi-step transition. The multiple peaks of different heights in the denaturation profile, correspond to the crowders locations in the chain.

- The force applied on a chain decreases from the applied point. The applied force propagates to the other monomers of the chain through the backbone of the molecule. Thus, a monomer which is situated far from the applied point

may not feel the applied force. The question we addressed here is what is that distance where the monomer will not be feeling that force. For a DNA molecule that is subjected to a force along the axis of the molecule, there is saturation length found where the monomers do not feel the applied force on the system. This length is known as deGennes length. By applying force on the random sites of homogeneous DNA molecule, we have found the saturation length of the molecule. Here the force is applied in the perpendicular to the helical axis. Earlier results showed that if one apply a force on the end of a chain (F_{end}) and in the mid of the chain (F_{mid}), the ratio of these two is 0.5. Our calculations showed that this ratio is not universal but depends on the length of the molecule. If the length of the chain is larger than a certain length (can be called as *critical length*), we get this ratio 0.5. The knowledge of this saturation length may help to understand the mechanism involved in intra-molecular processes such as loop propagation in sequences, ligand binding at molecular level and DNA protein interactions.

Future Scope

In spite of sincere efforts in past years to understand the phenomenon like base pair breaking, dynamics and various secondary and tertiary structure formation etc. in DNA, we are still could not get close to the phenomenon that occurs *in vivo*. The complex structure and the various interactions that are involved in the functioning of the molecules are some of the main hurdles to these investigations. Many sophisticated experiments *in vitro* and numerical as well as theoretical calculations unravel some of the interesting features of helix to coil transition in DNA. Some of notable achievements can be cited as the description of phase transition in these molecules, the hysteresis in the zipping-unzipping cycles, presence of salts in the dynamics of the molecule etc. One of the natural extension of the presented work is to study the time evolution of DNA molecules in the concentrated and crowded solutions. This is a known fact that the cellular environment is crowded due to presence of various biomolecules. The day by day advancement in the experimental techniques [278] empowering us in the correct estimation the intracellular activities. In addition, the computer simulations [279–281] that can take care the cellular environment are also improving our understanding of the interactions and the force response of nucleic acids. Systematic efforts around the globe are going on to study the role of these crowders in the formation of G-quadrapluxes [282].

The formation of triple-helical DNA during DNA denaturation, is also a topic of interest because of it's possible implications in the field of molecular biology. This

triple-helix formation affects the activities like DNA transcription, gene expression and others requiring DNA opening. Near or at the critical melting of duplex DNA, this triple helix DNA exhibits an Efimov like state; an analogy with the Efimov effect in three-body quantum mechanics. This third strand in triple helix DNA, can be used to identify the base sequences in double helix without opening the double helix [283].

In mechanical separation of dsDNA molecule, time-dependent forces influence the dynamics of separation, such as mean opening time, mean opening force, and mean critical opening force. Recent years witnessed some simulation results that displays interesting features of during the unzipping of dsDNA molecule under the influence of an external oscillatory fields [284–287]. How the melting transition and bubble formation in the chain are modified under certain fields, are an important investigations to observe.

References

[1] B. Alberts *et al.*, *Molecular Biology of the Cell, Third Edition* (Garland Science, 1994).

[2] C. R. Calladine, H. Drew, B. Luisi, and A. Travers, *Understanding DNA, Third Edition: The Molecule and How it Works* (Academic Press, 2004).

[3] P. Nelson, *Biological Physics: Energy, Information, Life*, 1^{st} ed. (W. H. Freeman, 2003).

[4] M. V. Volkenshtein, *Biophysics*, Rev. from the 1981 russian ed. (Mir, 1983).

[5] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry, Part 2: Techniques for the Study of Biological Structure and Function*, 1^{st} ed. (W. H. Freeman and Company, 1980).

[6] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry: Part III: The Behavior of Biological Macromolecules (Their Biophysical Chemistry)*, First edition ed. (W. H. Freeman, 1980).

[7] A. Vologodskii, *Biophysics of DNA*, 1^{st} ed. (Cambridge University Press, 2015).

[8] C. F. H. C. Watson J. D., Nature **171**, 737 (1953).

[9] R. E. Franklin and R. G. Gosling, Nature **171**, 740 (1953).

[10] V. A. Bloomfield *et al.*, *Nucleic Acids: Structures, Properties, and Functions*, 1^{st} ed. (University Science Books, 2000).

[11] R. R. Sinden, *DNA Structure and Function*, 1^{st} ed. (Academic Press, 1994).

[12] R. Kavenoff and B. H. Zimm, Chromosoma **41**, 1 (1973).

[13] M. Egli and W. Saenger, *Principles of Nucleic Acid Structure (Springer Advanced Texts in Chemistry)* (Springer, 1988).

[14] D. Svozil, P. Hobza, and J. Šponer, The Journal of Physical Chemistry B **114**, 1191 (2010).

[15] V. R. Cooper *et al.*, Journal of the American Chemical Society **130**, 1304 (2008).

[16] L. V. Yakushevich, *Nonlinear Physics of DNA*, 1^{st} ed. (Wiley, 1998).

[17] E. Chargaff, R. Lipshitz, and C. Green, Journal of Biological Chemistry **195**, 155 (1952).

[18] D. Elson and E. Chargaff, Experientia **8**, 143 (1952).

[19] G. Gaeta, C. Reiss, M. Peyrard, and T. Dauxois, La Rivista Del Nuovo Cimento Series 3 **17**, 1 (1994).

[20] G. Gaeta, Journal of Biological Physics **24**, 81 (1999).

[21] G. Felsenfeld and M. Groudine, Nature **421**, 448 (2003).

[22] K. Komiya, M. Yamamura, and J. A. Rose, Nucleic Acids Research **38**, 4539 (2010).

[23] Y. Chen, S.-H. Lee, and C. Mao, Angewandte Chemie International Edition **43**, 5335 (2004).

[24] M. Takinoue and A. Suyama, Small **2**, 1244 (2006).

[25] M. Takinoue and A. Suyama, Chem-Bio Informatics Journal **4**, 93 (2004).

REFERENCES

[26] D. Liu and S. Balasubramanian, Angewandte Chemie International Edition **42**, 5734 (2003).

[27] R. Steinert and B. Hudson, Journal of Chemical Education **50**, 129 (1973).

[28] M. Perelroyzen, V. Lyamichev, Y. Kalambet, Y. Lyubchenko, and A. Vologodskii, Nucleic Acids Research **9**, 4043 (1981).

[29] V. Morozov, E. Mamasakhlisov, S. Hayryan, and C.-K. Hu, Physica A **281**, 51 (2000).

[30] V. F. Morozov *et al.*, Physica A **348**, 327 (2005).

[31] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers; Statistical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules* (Academic Press, 1970).

[32] T. Lipniacki, Phys. Rev. E **64**, 051919 (2001).

[33] R. M. Wartell and A. S. Benight, Physics Reports **126**, 67 (1985).

[34] M. D. Frank-Kamenetskii and S. Prakash, Physics of Life Reviews **11**, 153 (2014).

[35] M. D. Frank-Kamenetskii, Nat Phys **328**, 17 (1987).

[36] L. Stryer, *Biochemistry*, 4^{th} ed. (W.H. Freeman & Company, 1995).

[37] B. H. Zimm and J. K. Bragg, The Journal of Chemical Physics **31**, 526 (1959).

[38] B. H. Zimm, The Journal of Chemical Physics **33**, 1349 (1960).

[39] A. Hanke and R. Metzler, Journal of Physics A: Mathematical and General **36**, L473 (2003).

[40] R. Thomas, Biochimica et Biophysica Acta **14**, 231 (1954).

[41] S. A. Rice and P. Doty, Journal of the American Chemical Society **79**, 3937 (1957).

[42] C. T. Wittwer, Human Mutation **30**, 857 (2009).

[43] C. Schildkraut and S. Lifson, Biopolymers **3**, 195 (1965).

[44] M. D. Frank-Kamenetskii, Biopolymers **10**, 2623 (1971).

[45] G. Weber, EPL (Europhysics Letters) **73**, 806 (2006).

[46] G. S. Manning, Macromolecules **34**, 4650 (2001).

[47] A. Singh and N. Singh, Physica A **392**, 2052 (2013).

[48] R. J. Owen, L. R. Hill, and S. P. Lapage, Biopolymers **7**, 503 (1969).

[49] R. Owczarzy *et al.*, Biochemistry **43**, 3537 (2004).

[50] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder, Biochemistry **47**, 5336 (2008).

[51] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, Nucleic Acids Research **34**, 564 (2006).

[52] T. Vuletić *et al.*, Phys. Rev. E **83**, 041803 (2011).

[53] G. Bonnet, O. Krichevsky, and A. Libchaber, **95**, 8602 (1998).

[54] D. Bikard, C. Loot, Z. Baharoglu, and D. Mazel, Microbiology and Molecular Biology Reviews **74**, 570 (2010).

[55] V. pavlov, Y. Lyubchenko, A. Borovik, and Y. Lazurkin, Nucleic Acids Research **4**, 4053 (1977).

[56] P. M. Vallone *et al.*, Biopolymers **50**, 425 (1999).

[57] M. J. J. Blommers *et al.*, Biochemistry **28**, 7491 (1989).

[58] V. P. Antao, S. Y. Lai, and I. Tinoco, Nucleic Acids Research **19**, 5901 (1991).

[59] E. Stellwagen, J. M. Muse, and N. C. Stellwagen, Biochemistry **50**, 3084 (2011).

[60] J. H. Gibbs and E. A. DiMarzio, The Journal of Chemical Physics **30**, 271 (1959).

[61] C. Kittel, American Journal of Physics **37**, 917 (1969).
[62] Y. S. Lazurkin, M. D. Frank-Kamenetskii, and E. N. Trifonov, Biopolymers **9**, 1253 (1970).
[63] O. Gotoh, Advances in Biophysics **16**, iii (1983).
[64] J. SantaLucia, Proceedings of the National Academy of Sciences **95**, 1460 (1998).
[65] N. Korolev, A. P. Lyubartsev, and L. Nordenski Biophysical Journal **75**, 3041 (1998).
[66] Y. Chen and E. Prohofsky, Biophysical Journal **64**, 1394 (1993).
[67] Y. Gao and E. W. Prohofsky, The Journal of Chemical Physics **80**, 2242 (1984).
[68] Y. Gao, K. V. DeviPrasad, and E. W. Prohofsky, The Journal of Chemical Physics **80**, 6291 (1984).
[69] A. Krueger, E. Protozanova, and M. D. Frank-Kamenetskii, Biophysical Journal **90**, 3091 (2006).
[70] Z.-J. Tan and S.-J. Chen, The Journal of Chemical Physics **122**, 044903 (2005).
[71] Z.-J. Tan and S.-J. Chen, Biophysical Journal **90**, 1175 (2006).
[72] D. Poland and H. A. Scheraga, The Journal of Chemical Physics **45**, 1464 (1966).
[73] D. Poland and H. A. Scheraga, The Journal of Chemical Physics **45**, 1456 (1966).
[74] D. Jost and R. Everaers, Biophysical Journal **96**, 1056 (2009).
[75] N. Theodorakopoulos, Phys. Rev. E **82**, 021905 (2010).
[76] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
[77] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[78] Y. Liu *et al.*, The Journal of Physical Chemistry B **114**, 9905 (2010).
[79] A. B. Fulton, Cell **30**, 345 (1982).
[80] S. B. Zimmerman and A. P. Minton, Annual Review of Biophysics and Biomolecular Structure **22**, 27 (1993).
[81] D. Miyoshi and N. Sugimoto, Biochimie **90**, 1040 (2008).
[82] R. Hancock, Frontiers in Physics **2**, 53 (2014).
[83] R. Ellis, Trends in Biochemical Sciences **26**, 597 (2001).
[84] R. Ellis, Current Opinion in Structural Biology **11**, 114 (2001).
[85] D. L. Pincus and D. Thirumalai, The Journal of Physical Chemistry B **117**, 13107 (2013).
[86] S. Nakano, D. Miyoshi, and N. Sugimoto, Chemical Reviews **114**, 2733 (2014).
[87] C. H. Spink, N. Garbett, and J. B. Chaires, Biophysical Chemistry **126**, 176 (2007).
[88] K. S. Harve, R. Lareu, R. Rajagopalan, and M. Raghunath, Nucleic Acids Research **38**, 172 (2010).
[89] S. Nakano, H. Karimata, T. Ohmichi, J. Kawakami, and N. Sugimoto, Journal of the American Chemical Society **126**, 14330 (2004).
[90] C. H. Spink and J. B. Chaires, Biochemistry **38**, 496 (1999).
[91] C. H. Spink and J. B. Chaires, Journal of the American Chemical Society **117**, 12887 (1995).
[92] I. Khimji, J. Shin, and J. Liu, Chem. Commun. **49**, 1306 (2013).
[93] H. Karimata, S. Nakano, and N. Sugimoto, Bulletin of the Chemical Society of Japan **80**, 1987 (2007).
[94] T. Fujimoto, S. Nakano, N. Sugimoto, and D. Miyoshi, The Journal of Physical Chemistry B **117**, 963 (2013).
[95] R. Moriyama, Y. Iwasaki, and D. Miyoshi, The Journal of Physical Chemistry B **119**, 11969 (2015).

[96] G. P. Goodrich, M. R. Helfrich, J. J. Overberg, and C. D. Keating, Langmuir **20**, 10246 (2004).
[97] S. Qin and H.-X. Zhou, Phys. Rev. E **81**, 031919 (2010).
[98] J. Mittal and R. B. Best, Biophysical Journal **98**, 315 (2010).
[99] D. Tsao and N. V. Dokholyan, Phys. Chem. Chem. Phys. **12**, 3491 (2010).
[100] J. Batra, K. Xu, and H.-X. Zhou, Proteins: Structure, Function, and Bioinformatics **77**, 133 (2009).
[101] Y. Liu, Y. Shang, H. Liu, Y. Hu, and J. Jiang, Phys. Chem. Chem. Phys. **14**, 15400 (2012).
[102] C. A. Brackley, M. E. Cates, and D. Marenduzzo, Phys. Rev. Lett. **111**, 108101 (2013).
[103] M. M. Waegele and F. Gai, The Journal of Chemical Physics **134**, 095104 (2011).
[104] S. Qin and H.-X. Zhou, Biophysical Journal **97**, 12 (2009).
[105] A. Wildes *et al.*, Phys. Rev. Lett. **106**, 048101 (2011).
[106] O. Krichevsky and G. Bonnet, Reports on Progress in Physics **65**, 251 (2002).
[107] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).
[108] H. M. Sobell, Proceedings of the National Academy of Sciences **82**, 5328 (1985).
[109] A. Spassky and D. Angelov, Journal of Molecular Biology **323**, 9 (2002).
[110] Y. Zeng and G. Zocchi, Biophysical Journal **90**, 4522 (2006).
[111] A. Bonincontro, M. Matzeu, F. Mazzei, A. Minoprio, and F. Pedone, Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1171**, 288 (1993).
[112] G. Bonnet, S. Tyagi, A. Libchaber, and F. R. Kramer, Proceedings of the National Academy of Sciences **96**, 6171 (1999).
[113] N. C. Nelson, P. W. Hammond, E. Matsuda, A. A. Goud, and M. M. Becker, Nucleic Acids Research **24**, 4998 (1996).
[114] C. H. Choi *et al.*, Nucleic Acids Research **32**, 1584 (2004).
[115] J. Fei and T. Ha, Proceedings of the National Academy of Sciences **110**, 17173 (2013).
[116] M. E. Fisher, The Journal of Chemical Physics **45**, 1469 (1966).
[117] C. Richard and A. Guttmann, Journal of Statistical Physics **115**, 925 (2004).
[118] J. Valle-Orero *et al.*, New Journal of Physics **16**, 113017 (2014).
[119] Y. Kafri, D. Mukamel, and L. Peliti, Physica A **306**, 39 (2002).
[120] N. Singh and Y. Singh, Phys. Rev. E **64**, 042901 (2001).
[121] B. S. Alexandrov, L. T. Wille, K. Ø. Rasmussen, A. R. Bishop, and K. B. Blagoev, Phys. Rev. E **74**, 050901 (2006).
[122] S. Ares, N. K. Voulgarakis, K. Ø. Rasmussen, and A. R. Bishop, Phys. Rev. Lett. **94**, 035504 (2005).
[123] T. van Erp, S. Cuesta-López, and M. Peyrard, The European Physical Journal E **20**, 421 (2006).
[124] Z. Rapti *et al.*, EPL (Europhysics Letters) **74**, 540 (2006).
[125] N. Theodorakopoulos, Phys. Rev. E **77**, 031919 (2008).
[126] M. Kenward and K. D. Dorfman, The Journal of Chemical Physics **130**, 095101 (2009).
[127] K. P. N. Murthy and G. M. Schütz, EPL (Europhysics Letters) **96**, 68003 (2011).
[128] M. Zoli, The Journal of Chemical Physics **138**, 205103 (2013).
[129] M. G. Gauthier, J. Herrick, and J. Bechhoefer, Phys. Rev. Lett. **104**, 218104 (2010).
[130] M. McCullagh, I. Franco, M. A. Ratner, and G. C. Schatz, The Journal of Physical

Chemistry Letters **3**, 689 (2012).
[131] S. Kumar and M. S. Li, Physics Reports **486**, 1 (2010).
[132] T. R. Strick *et al.*, Reports on Progress in Physics **66**, 1 (2003).
[133] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Proceedings of the National Academy of Sciences **94**, 11935 (1997).
[134] C. Danilowicz *et al.*, Proceedings of the National Academy of Sciences **100**, 1694 (2003).
[135] K. L. Sebastian, Phys. Rev. E **62**, 1128 (2000).
[136] G. Lee, L. Chrisey, and R. Colton, Science **266**, 771 (1994).
[137] A. D. MacKerell Jr. and G. U. Lee, European Biophysics Journal **28**, 415 (1999).
[138] L. H. Pope *et al.*, European Biophysics Journal **30**, 53 (2001).
[139] A. Noy, D. V. Vezenov, J. F. Kayyem, T. J. Maade, and C. M. Lieber, Chemistry & Biology **4**, 519 (1997).
[140] M. Santosh and P. K. Maiti, Journal of Physics: Condensed Matter **21**, 034113 (2009).
[141] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, Nat Struct Mol Biol **6**, 346 (1999).
[142] R. K. Mishra, T. Modi, D. Giri, and S. Kumar, The Journal of Chemical Physics **142**, 174910 (2015).
[143] A. Neuman, Keir C.and Nagy, Nat Meth **5**, 491 (2008).
[144] G. Binnig, C. F. Quate, and C. Gerber, Phys. Rev. Lett. **56**, 930 (1986).
[145] R. Simmons, J. Finer, S. Chu, and J. Spudich, Biophysical Journal **70**, 1813 (1996).
[146] F. Amblard, B. Yurke, A. Pargellis, and S. Leibler, Review of Scientific Instruments **67**, 818 (1996).
[147] E. Evans, K. Ritchie, and R. Merkel, Biophysical Journal **68**, 2580 (1995).
[148] T. R. Strick *et al.*, Reports on Progress in Physics **66**, 1 (2003).
[149] M.-L. Visnapuu, D. Duzdevich, and E. C. Greene, Mol. BioSyst. **4**, 394 (2008).
[150] A. Singh, B. Mittal, and N. Singh, Phys. Express **3**, 18 (2013).
[151] A. Noy, editor, *Handbook of Molecular Force Spectroscopy*, 2008 ed. (Springer, 2007).
[152] N. Singh and Y. Singh, The European Physical Journal E **17**, 7 (2005).
[153] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. E **58**, 2386 (1998).
[154] S. M. Bhattacharjee, Journal of Physics A: Mathematical and General **33**, L423 (2000).
[155] J. Weeks *et al.*, Biophysical Journal **88**, 2752 (2005).
[156] A. Guttmann, J. Jacobsen, I. Jensen, and S. Kumar, Journal of Mathematical Chemistry **45**, 223 (2009).
[157] P.-G. Gennes, *Scaling Concepts in Polymer Physics*, 1^{st} ed. (Cornell University Press, 1979).
[158] S. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).
[159] J. F. Marko, Phys. Rev. E **57**, 2134 (1998).
[160] C. Bustamante, Z. Bryant, and S. B. Smith, Nature **421**, 423 (2003).
[161] I. Rouzina and V. A. Bloomfield, Biophysical Journal **80**, 882 (2001).
[162] I. Rouzina and V. A. Bloomfield, Biophysical Journal **80**, 894 (2001).
[163] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, Biophysical Journal **82**, 3160 (2002).
[164] L. Shokri, M. J. McCauley, I. Rouzina, and M. C. Williams, Biophysical Journal **95**, 1248 (2008).
[165] M. C. Williams, J. R. Wenner, I. Rouzina, and V. A. Bloomfield, Biophysical Journal **80**,

874 (2001).
[166] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proceedings of the National Academy of Sciences **94**, 6185 (1997).
[167] C. G. Baumann *et al.*, Biophysical Journal **78**, 1965 (2000).
[168] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).
[169] S. Kidoaki and K. Yoshikawa, Biophysical Chemistry **76**, 133 (1999).
[170] F. Ritort, Journal of Physics: Condensed Matter **18**, R531 (2006).
[171] K. Hatch, C. Danilowicz, V. Coljee, and M. Prentiss, Nucleic Acids Research **36**, 294 (2008).
[172] J. M. Huguet *et al.*, Proceedings of the National Academy of Sciences **107**, 15431 (2010).
[173] A. Bosco, J. Camunas-Soler, and F. Ritort, Nucleic Acids Research **42**, 2064 (2014).
[174] U. Bockelmann, Current Opinion in Structural Biology **14**, 368 (2004).
[175] K. R. Chaurasiya, T. Paramanathan, M. J. McCauley, and M. C. Williams, Physics of Life Reviews **7**, 299 (2010).
[176] K. M. Kosikov, A. A. Gorin, V. B. Zhurkin, and W. K. Olson, Journal of Molecular Biology **289**, 1301 (1999).
[177] R. Podgornik, P. L. Hansen, and V. A. Parsegian, The Journal of Chemical Physics **113**, 9343 (2000).
[178] A. Wynveen and C. N. Likos, Soft Matter **6**, 163 (2010).
[179] F. Romano, D. Chakraborty, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, The Journal of Chemical Physics **138**, 085101 (2013).
[180] B. E. K. Snodin *et al.*, The Journal of Chemical Physics **142**, 234901 (2015).
[181] J.-M. Yuan *et al.*, Protein Science **17**, 2156 (2008).
[182] A. R. Singh, D. Giri, and S. Kumar, Phys. Rev. E **79**, 051801 (2009).
[183] H. Li, E. hua Cao, and T. Gisler, Biochemical and Biophysical Research Communications **379**, 70 (2009).
[184] C. Yang, S. Jang, and Y. Pak, The Journal of Chemical Physics **135**, 225104 (2011).
[185] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[186] A. E. Bergues-Pupo, J. R. Arias-Gonzalez, M. C. Morn, A. Fiasconaro, and F. Falo, Nucleic Acids Research (2015), DOI:10.1093/nar/gkv690.
[187] C. Ke, M. Humeniuk, H. S-Gracz, and P. E. Marszalek, Phys. Rev. Lett. **99**, 018302 (2007).
[188] C. Danilowicz *et al.*, Proceedings of the National Academy of Sciences **106**, 13196 (2009).
[189] R. Kapri and S. M. Bhattacharjee, Phys. Rev. Lett. **98**, 098101 (2007).
[190] D. Giri and S. Kumar, Phys. Rev. E **73**, 050903 (2006).
[191] R. Kapri, S. M. Bhattacharjee, and F. Seno, Phys. Rev. Lett. **93**, 248102 (2004).
[192] N. Singh and Y. Singh, The European Physical Journal E **19**, 233 (2006).
[193] S. Srivastava and N. Singh, The Journal of Chemical Physics **134**, 115102 (2011).
[194] A. R. Singh, D. Giri, and S. Kumar, The Journal of Chemical Physics **132**, 235105 (2010).
[195] S. Nath *et al.*, The Journal of Chemical Physics **139**, 165101 (2013).
[196] K. Hatch, C. Danilowicz, V. Coljee, and M. Prentiss, Phys. Rev. E **78**, 011920 (2008).
[197] M. Mosayebi, A. A. Louis, J. P. K. Doye, and T. E. Ouldridge, ACS Nano **0**.
[198] M. Zoli, Phys. Rev. E **79**, 041927 (2009).
[199] M. Zoli, Phys. Rev. E **81**, 051910 (2010).

[200] A. Campa and A. Giansanti, Phys. Rev. E **58**, 3585 (1998).
[201] A. Campa and A. Giansanti, Journal of Biological Physics **24**, 141 (1999).
[202] Y.-l. Zhang, W.-M. Zheng, J.-X. Liu, and Y. Z. Chen, Phys. Rev. E **56**, 7100 (1997).
[203] O. Dahlen and T. S. van Erp, The Journal of Chemical Physics **142**, 235101 (2015).
[204] M. Peyrard, S. Cuesta-López, and D. Angelov, Journal of Physics: Condensed Matter **21**, 034103 (2009).
[205] S. Kumar and D. Giri, The Journal of Chemical Physics **125**, 044905 (2006).
[206] A. Bergues-Pupo, J. Bergues, and F. Falo, Physica A **396**, 99 (2014).
[207] N. K. Voulgarakis, A. Redondo, A. R. Bishop, and K. Ø. Rasmussen, Phys. Rev. Lett. **96**, 248101 (2006).
[208] Z. Rapti, K. Ø. Rasmussen, and A. R. Bishop, Journal of Nonlinear Mathematical Physics **18**, 381 (2011).
[209] M. Barbi, S. Cocco, and M. Peyrard, Physics Letters A **253**, 358 (1999).
[210] M. Zoli, The Journal of Chemical Physics **135**, 115101 (2011).
[211] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
[212] P. M. Morse, Phys. Rev. **34**, 57 (1929).
[213] T. Dauxois, Physics Letters A **159**, 390 (1991).
[214] T. Dauxois and M. Peyrard, *Dynamics of breather modes in a nonlinear helicoidal model of DNA*, Lecture Notes in Physics Vol. 393 (Springer Berlin Heidelberg, 1991).
[215] B. S. Alexandrov *et al.*, Nucleic Acids Research **37**, 2405 (2009).
[216] N. C. Baird, International Journal of Quantum Chemistry **8**, 49 (1974).
[217] M. Peyrard and T. Dauxois, Physics of Life Reviews **11**, 173 (2014).
[218] J. Elschner, ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift fr Angewandte Mathematik und Mechanik **73**, 376 (1993).
[219] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[220] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing, Second Edition* (Cambridge University Press, 1992).
[221] M. Barbi, S. Lepri, M. Peyrard, and N. Theodorakopoulos, Phys. Rev. E **68**, 061909 (2003).
[222] T. S. van Erp, S. Cuesta-Lopez, J.-G. Hagmann, and M. Peyrard, Phys. Rev. Lett. **95**, 218104 (2005).
[223] S. Cocco, R. Monasson, and J. F. Marko, Proceedings of the National Academy of Sciences **98**, 8608 (2001).
[224] S. Cocco, R. Monasson, and J. F. Marko, Phys. Rev. E **65**, 041907 (2002).
[225] G. Kalosakas and S. Ares, The Journal of Chemical Physics **130**, 235104 (2009).
[226] V. Bloomfield, D. Crothers, and J. Tinoco.I., *Nucleic Acids. Structures, Properties and Functions* (University Science Books, Sausalito, CA, 2000).
[227] A. Brunet *et al.*, Macromolecules **48**, 3641 (2015).
[228] S. Buyukdagli and M. Joyeux, Phys. Rev. E **77**, 031903 (2008).
[229] G. S. Manning, Biopolymers **11**, 937 (1972).
[230] K. A. Sharp and B. Honig, The Journal of Physical Chemistry **94**, 7684 (1990).
[231] J. A. Grant, B. T. Pickup, and A. Nicholls, Journal of Computational Chemistry **22**, 608 (2001).
[232] T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, Phys. Rev. Lett. **97**, 128105

(2006).
[233] E. Yeramian, Gene **255**, 139 (2000).
[234] A. Ramachandran, Q. Guo, S. M. Iqbal, and Y. Liu, The Journal of Physical Chemistry B **115**, 6138 (2011).
[235] B. E. K. Snodin *et al.*, The Journal of Chemical Physics **142**, 234901 (2015).
[236] S. Niewieczerza and M. Cieplak, Journal of Physics: Condensed Matter **21**, 474221 (2009).
[237] R. C. DeMille, T. E. Cheatham, and V. Molinero, The Journal of Physical Chemistry B **115**, 132 (2011).
[238] D. A. Potoyan, A. Savelyev, and G. A. Papoian, Wiley Interdisciplinary Reviews: Computational Molecular Science **3**, 69 (2013).
[239] R. D. Blake and S. G. Delcourt, Nucleic Acids Research **26**, 3323 (1998).
[240] R. Dong, X. Yan, and S. Liu, Journal of Physics A: Mathematical and General **37**, 4977 (2004).
[241] M. Zoli, Journal of Physics: Condensed Matter **24**, 195103 (2012).
[242] F. Zhang and M. A. Collins, Phys. Rev. E **52**, 4217 (1995).
[243] K. Drukker, G. Wu, and G. C. Schatz, The Journal of Chemical Physics **114**, 579 (2001).
[244] G. Weber, N. Haslam, J. W. Essex, and C. Neylon, Journal of Physics: Condensed Matter **21**, 034106 (2009).
[245] A. Singh and N. Singh, American Institute of Physics Conference Series **1536**, 1201 (2013).
[246] R. Kapri and S. M. Bhattacharjee, EPL (Europhysics Letters) **83**, 68002 (2008).
[247] D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000).
[248] D. K. Lubensky and D. R. Nelson, Phys. Rev. E **65**, 031917 (2002).
[249] M. Zoli, Journal of Theoretical Biology **354**, 95 (2014).
[250] F. Sanger, A. Coulson, G. Hong, D. Hill, and G. Petersen, Journal of Molecular Biology **162**, 729 (1982).
[251] J. Šponer, K. E. Riley, and P. Hobza, Phys. Chem. Chem. Phys. **10**, 2595 (2008).
[252] R. Kapri, Phys. Rev. E **86**, 041906 (2012).
[253] M. Peyrard, S. Cuesta-López, and G. James, Journal of Biological Physics **35**, 73 (2009).
[254] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Phys. Rev. Lett. **85**, 2400 (2000).
[255] P. Peltomäki, **21**, 1174 (2003).
[256] C. Hensey and J. Gautier, Mechanisms of Development **69**, 183 (1997).
[257] M. M. Vilenchik and A. G. Knudson, Proceedings of the National Academy of Sciences **100**, 12871 (2003).
[258] N. Singh, *Statistical theory of unzipping of DNA molecule*, PhD thesis, Department of Physics, Banaras Hindu University, 2004.
[259] T. Dauxois and M. Peyrard, Phys. Rev. E **51**, 4027 (1995).
[260] A. Singh and N. Singh, Phys. Rev. E **92**, 032703 (2015).
[261] Z. Rapti *et al.*, Phys. Rev. E **73**, 051902 (2006).
[262] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
[263] Y. Sasaki, D. Miyoshi, and N. Sugimoto, Biotechnology Journal **1**, 440 (2006).
[264] D. Miyoshi, S. Matsumura, S. Nakano, and N. Sugimoto, Journal of the American Chemical Society **126**, 165 (2004).
[265] G.-W. Li, O. G. Berg, and J. Elf, Nat Phys **5**, 294 (2009).

[266] J. Shin, A. G. Cherstvy, and R. Metzler, Soft Matter **11**, 472 (2015).
[267] H. C. Fogedby and R. Metzler, Phys. Rev. Lett. **98**, 070601 (2007).
[268] F. Khner, J. Morfill, R. A. Neher, K. Blank, and H. E. Gaub, Biophysical Journal **92**, 2491 (2007).
[269] D. K. Lubensky, AIP Conference Proceedings **665**, 336 (2003).
[270] S. Kumar, EPL (Europhysics Letters) **85**, 38003 (2009).
[271] S. Prakash and Y. Singh, Phys. Rev. E **84**, 031905 (2011).
[272] M. Mosayebi, A. A. Louis, J. P. K. Doye, and T. E. Ouldridge, arXiv:1502.03623 [cond-mat.soft] (2015).
[273] P.-G. de Gennes, Comptes Rendus de l'Acadmie des Sciences - Series {IV} - Physics **2**, 1505 (2001).
[274] B. Chakrabarti and D. R. Nelson, The Journal of Physical Chemistry B **113**, 3831 (2009).
[275] S. Srivastava and Y. Singh, EPL (Europhysics Letters) **85**, 38001 (2009).
[276] C. Danilowicz *et al.*, Journal of Physics: Condensed Matter **22**, 414106 (2010).
[277] A. Singh and N. Singh, Macromolecular Symposia **357**, 64 (2015).
[278] B. Heddi and A. T. Phan, Journal of the American Chemical Society **133**, 9824 (2011).
[279] J. Shin, A. G. Cherstvy, and R. Metzler, New Journal of Physics **16**, 053047 (2014).
[280] A. V. Doghaei, M. Housaindokht, and M. Bozorgmehr, Journal of Theoretical Biology **364**, 103 (2015).
[281] H. Tateishi-Karimata and N. Sugimoto, Nucleic Acids Research **42**, 8831 (2014).
[282] D. Rhodes and H. J. Lipps, Nucleic Acids Research **43**, 8627 (2015).
[283] T. Pal, P. Sadhukhan, and S. M. Bhattacharjee, Phys. Rev. E **91**, 042105 (2015).
[284] S. Kumar and G. Mishra, Phys. Rev. Lett. **110**, 258102 (2013).
[285] G. Mishra, P. Sadhukhan, S. M. Bhattacharjee, and S. Kumar, Phys. Rev. E **87**, 022718 (2013).
[286] R. Kapri, Phys. Rev. E **90**, 062719 (2014).
[287] A. E. Bergues-Pupo, F. Falo, and A. Fiasconaro, EPL (Europhysics Letters) **105**, 68005 (2014).

List of Publications and Presentations

International Journals:

- *"Pulling short DNA molecules having defects on different locations"*
  Amar Singh and Navin Singh, Phy. Rev. E, **92** 032703 (2015)
- *"Effect of salt concentration on the stability of heterogeneous DNA"*
  Amar Singh and Navin Singh, Physica A, **419** 328 (2015)
- *"Phase diagram of mechanically stretched DNA: The salt effect"*
  Amar Singh and Navin Singh, Physica A, **392** 2052 (2013)
- *"Thermal denaturation of dsDNA in presence of molecular crowders"*
  Amar Singh and Navin Singh, (Communicated)
- *"Denaturation of DNA at high salt concentrations"*
  Arghya Maity, **Amar Singh** and Navin Singh, (Communicated)
- *"Critical length in mechanical unzipping of DNA"*
  Amar Singh and Navin Singh, (manuscript under preparation)

Conference proceedings in International Journals:

- *"Molecular Crowding effects on Stability of DNA Double Helix"*
  Amar Singh and Navin Singh, AIP Conf. Proc. (2015, In press)
- *"Pulling DNA: The effect of chain length on the mechanical stability of DNA chain"*
  Amar Singh and Navin Singh, Macromolecular Symposia, **357** 64 (2015)
- *"Role of chain stiffness & end entropy in unzipping of DNA chain"*
  Amar Singh and Navin Singh, AIP Conf. Proc. **1536** 1201 (2013)
- *"Force induced unzipping of dsDNA: The solvent effect"*
  Amar Singh, Bhaskar Mittal and Navin Singh, Phys. Express, **3** 18 (2013)
- *"Stabilization of dsDNA Molecule With Salt Concentration of Solution"*
  Amar Singh and Navin Singh, Int. J of Phy. and Math. Sc., **2** 58 (2012)

List of conferences/schools attended/participated:

- *"International Conference Condensed Matter & Applied Physics (ICC 2015)"*, October 30-31, 2015, GEC, **Bikaner, INDIA.**
- *"International Conference on Soft Materials (ICSM 2014)"*, October 01-06, 2014, MNIT **Jaipur, INDIA.**
- *"4^{th} RRI school on Statistical Physic"*, April 01-13, 2013, RRI, **Bangalore, INDIA.**
- *"38^{th} Conference of the Middle European Cooperation in Statistical Physics MECO38"*, March 25-27, 2013, ICTP, **Trieste, ITALY**
- *"International Conference on Recent Trends in Applied Physics & Material Science"*, February 01-02, 2013, GEC, **Bikaner, INDIA.**
- *"Condensed Matter Days 2012"* August 29-31, 2012, BIT, **Mesra (Ranchi), INDIA.**
- A workshop on *"Scientific/Research Paper Writing"*, June 08-10, 2012, C.I.F.E., **Mumbai, INDIA**.
- *"National Conference on Condensed Matter Physics"*, February 24-25, 2012, **BITS Pilani, INDIA** .
- A workshop on *"Physics of DNA"*, August 11-13, 2010, Banaras Hindu University, **Varanasi, INDIA**.

Brief Biography of the Supervisor

Dr. Navin Singh is Assistant Professor in Physics at Birla Institute of Technology & Science Pilani, Pilani Campus. He received his Ph. D. degree from Department of Physics, Banaras Hindu University in 2004. During his doctoral degree, he worked under the supervision of Professor Yashwant Singh in the research area of soft condensed matter physics. After his doctoral degree, he spent a year at Max Planck Institute for Polymer Research, Mainz, Germany as Post doctoral fellow. He joined Physics Department of BITS Pilani, Pilani Campus in December 2006. He has published several highly cited papers in reputed international journals. His research interests are in the Physics of biological systems like DNA, RNA or Proteins. Currently his research deals with force induced unzipping of dsDNA. He has successfully completed one UGC project and is presently dealing with a DST project.

Brief Biography of the Student

Mr. Amar Singh obtained his Masters degree in Physics from Malaviya National Institute of Technology (MNIT), Jaipur in 2008. He worked more than a year as a project associate at CSIR-CEERI, Pilani. He was involved in designing and fabrication of MEMS pressure sensors. He joined Department of Physics, BITS Pilani as a UGC Project fellow in June 2010. He is currently pursuing Ph.D. in the same Department in the area of soft condensed matter physics. His research interests include the structural changes in DNA helix and force induced unzipping of DNA molecules. He has published several papers in international journals and few are under review. In addition to this, he has participated and presented his work in several national and international conferences of high repute.