CHAPTER 1

Introduction

Polymer is a large macromolecule that consists of repetitive units called monomers [1–3]. Polymers can be classified in various categories: (i) structure-wise, (ii) based on composition (different kinds of monomers), and (iii)source or origin wise. The functionality (f') is a parameter which is defined as the number of reactive end-groups associated with a monomer [2–5]. For $f' < 2$, polymerization does not take place and hence, polymer cannot be formed. For $f' = 2$, a linear polymer chain can grow, while $f' > 2$, a complex structure appears such as branched polymer, cross-linked polymers, etc. In Fig. 1, we show cases for $f' \geq 2$.

Polymers may be classified in terms of its constituent monomers. If monomers are identical, it is termed as homopolymer, whereas if monomers are different, that is known as heteropolymer. Heteropolymers may further be classified as random copolymer, block copolymer. In a random copolymer, no periodicity is maintained between different kinds of monomers [4, 6]. In contrast, in the block copolymer, a block of two or more monomers gets repeated along the polymer's backbone. In Fig. 1 we

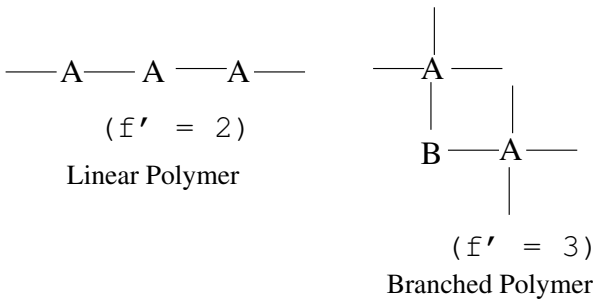

Figure 1.1. Schematic representation: Polymer classification based on functionality.

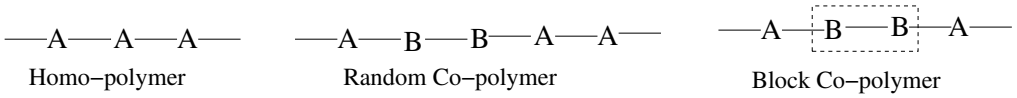

Figure 1.2. Schematic representation: Polymer classification based on different kind of monomer

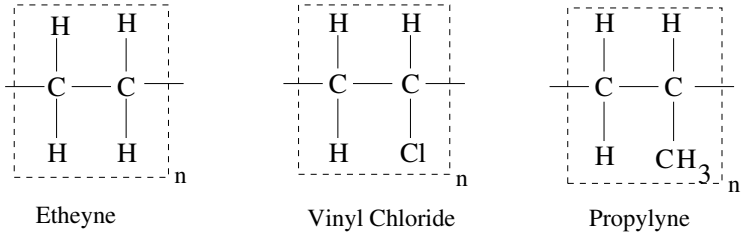

Figure 1.3. Examples of monomers of synthetic polymer

illustrate some examples of homo and heteropolymers.

There are two different types of polymer, namely, (i) synthetic polymers, and (ii) natural polymers. Synthetic polymers are synthesised in the laboratory, such as polyethylene, polyvinyl chloride, polypropylene, etc., where the repetitive units are ethylene, vinyl chloride, propylene respectively (Fig. 1) [7, 8]. On the otherhand, proteins, RNA, DNA, polysaccharides, etc., are natural polymers, also known as biopolymers.

Proteins are made up several smaller units called amino acids that are attached to one other by peptide bonds forming a long chain [9, 10]. Amino acids are composed of amino ($-NH_2$) and Carboxyl ($-COOH$) functional groups along with a side chain (R-group) which is unique for each amino acid. There are 20 different kinds of amino acids linked by peptide bonds to form a unique three-dimensional structure of the protein. There are four stages in protein folding. The linear sequence of amino acids is known as the primary structure. The hydrogen bonding between amino and carbonyl groups often folds into a two-dimensional β-sheet or α-helix, which is known as a secondary structure. The combination of folds and formations in a single linear chain forms polypeptides which are called tertiary structures. Lastly, multiple polypeptide chains

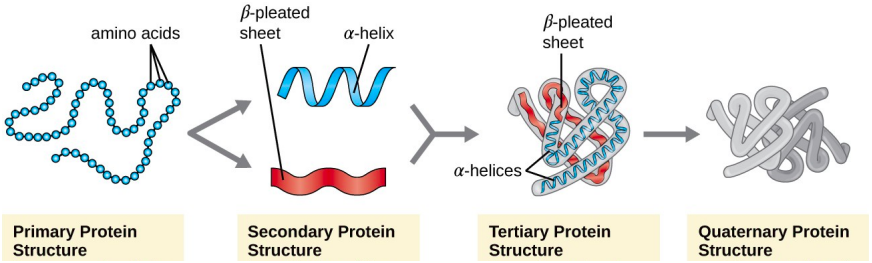

Figure 1.4. Various stages of protein folding. Image is taken from $https$: $//courses.lumenlearning.com/microbiology/chapter/proteins/$

fold into a complex quaternary structure [11]. The occurrence of the structures during folding are shown in Fig. 1.4.

Nucleic acids (DNA and RNA) are another examples of biopolymers. Nucleotides are the repetitive units of DNA and RNA which are composed of a nitrogenous base, sugar, and phosphate [9, 12]. While sugar and phosphate are the same for each nucleotide, there are four nitrogenous bases, namely Adenine (A), Thymine (T), Guanine (G), and Cytosine (C), in the case of DNA (Fig. 1.5). While for the RNA, Thymine (T) is replaced by Uracil (U). There is a Watson-Crick interaction between two single stranded DNA (ssDNA) where A makes base pair with T by two hydrogen bonds, whereas G pairs with C by three hydrogen bonds. This makes G-C base-pairs stronger in compare to AT base-pairs. This is known as complementary base-pairing. In double stranded DNA (dsDNA), the two strands of the DNA are antiparallel to each other, determined by the position of sugar and phosphate classified by $3'$ and $5'$. The $5'$ and $3'$ designations refer to the number of carbon atoms in Deoxyribose molecules to which a phosphate group gets attached.

In the living cell, DNA is usually found in the nucleus. It contains all the genetic information and transfer it to another cell during cell division. During cell division, DNA undergoes through the process of replication where the genetic information of mother DNA is being copied to the daughter DNA [13]. There is another biological process known as transcription, where RNA gets synthesized from the DNA [9, 14]. Both these processes require the unwinding of dsDNA into two single strands

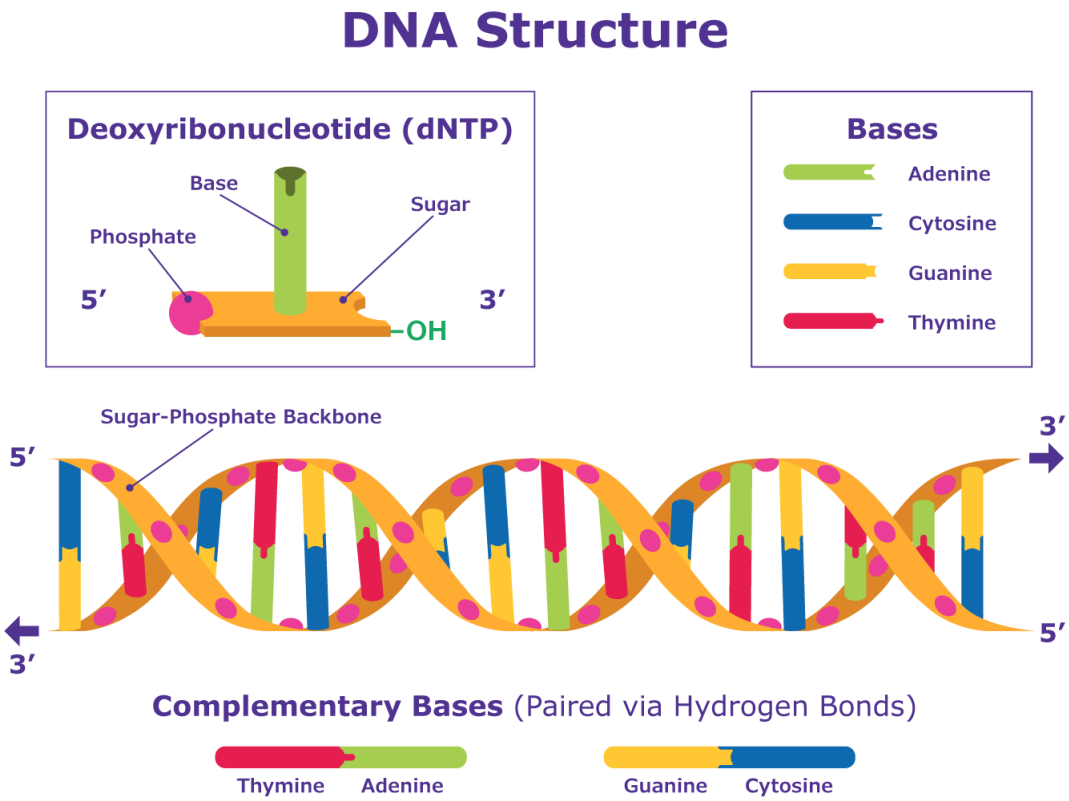

Figure 1.5. Schematic representation of dsDNA structure. Image taken from $https://www.sigmaaldrich.com/IN/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing$

Table 1.1. Range of forces and corresponding displacements for various experimental probes.

Techniques	$f_{max} - min(N)$	$\Delta(x)_{min}$ (m)
Optical tweezers	$10^{-13} - 10^{-10}$	10^{-9}
AFM	$10^{-11} - 10^{-7}$	10^{-10}
Micro-needles	$10^{-12} - 10^{-10}$	10^{-9}

mediated by enzyme action [15]. *In vitro*, the separation is induced either by temperature (DNA melting) or change in solvent conditions (DNA denaturation) [16]. The melting temperature (T_m) is usually defined when half of the base-pairs of the given strand are opened [17]. The melting temperature usually occurs in the range of $\approx 85°C$ to $95°C$ depending upon the sequences. For A-T rich sequence, melting temperature is found to be less compared to G-C rich region. This is because A-T binds with two hydrogen bonds, whereas G-C binds with three hydrogen bonds[18]. In *vivo*, replication and transcription take place at much lower temperature ($36 - 38\ °C$). In fact the separation of dsDNA *in vivo* is mediated by different types of enzymes (DNA helicase, polymerase, etc.) which exerts a force which eventually reduces the melting temperature. Motivated by this, efforts are being made in recent years to measure the forces responsible for processes like DNA unzipping, protein unfolding, etc.

In the following, we discuss about the different types of forces involved in the biological processes [19, 20]. The smallest force ranges between $4-6$ pN arising due to stochastic fluctuations. The molecular forces arising due to vander waals interaction and hydrogen bonding usually lies in the range of $8 - 120$ pN. The largest force is arising due to stretching of covalent bonds which is of the order of 10^3 pN. The single molecule force spectroscopy (SMFS) experiments allow us to measure these forces. For example, optical tweezer measure the force in the range of $0.1 - 100$ pN, whereas Magnetic tweezer can measure $10^{-3} - 10^4$ pN, AFM $10 - 10^4$ can measure force (Table 1.1) [21]. Therefore, it is now possible to study

the force induced transitions *in vitro*, *e.g.* folding-unfolding transition of protein, stretching of DNA, unzipping transition of dsDNA, etc.

New challenges have been seen in the area of protein unfolding, DNA unzipping etc. when removed from an artificial, controlled environment *in vitro* and relocated to the cellular environment [22]. Cells have a crowded environment, because they are composed of various types of biomolecules which occupy a large fraction of the total volume. It is now known that such confinement can influence the stability, dynamics and function of biomolecules. Understanding of equilibrium properties of a biopolymer in confined geometry thus become important because of the shift in the balance between the conformational entropy and the internal energy [17, 23], which leads polymers to modify their shape under external parameter, such as, changes in temperature, or pH of the solvent, or external force. Such confinement may delineate the possible mechanism involved in many biological processes, such as, translocation, transport of protein from the nucleus, ejection of viral DNA from the capsid, etc. [24, 25].

While most experiments and theorists focus on driven system, there is also considerabale interest related to the unforced translocation. An interesting example is when a pore connects two volumes of the solvent of different qualities in case of nascent polypeptide which translocate from cytoplasm of eucaryotic cell to the lumen of the endoplasmic reticulum [26]. It is pertinent to mention here that in some cases the shape of the pore interface (*e.g.* Mycobacterium Smegmatis Porin A (MSPA), HIV capsid, etc.) looks similar to the cone shaped channel [27]. One of the objective of present thesis is to study the effect of asymmetry arising due to cone shaped channel on the DNA melting and compare it with a pore on the flat interface.

In this context confinement arising due to impenetrable surface have been studied quite extensively in the context of adsorption of biopolymers on the surface. This has potential applications in biocompatibility of surgical implants, prosthetics depends on the cellular interaction with

metal /non-metal surface, etc. There is a considerable amount of studies related to adsorption of biopolymers, *e.g.*, proteins in serum and blood which adsorps to the substrate [28, 29]. On the other hand, adsorption of DNA and RNA molecules on gold nanoparticle surface has utmost importance due to RNA sequencing, which is useful for early clinical diagnosis of AIDS virus, hepatitis C, COVID-19 virus, etc. [30, 31]. The dynamics of dsDNA adsorption on the gold surface are different from ssDNA or RNA as flexible ssDNA uncoil its bases easily absorbed by the charged surface. While the rigid dsDNA is itself stable due to double helix geometry [32]. This makes the study of adsorption of dsDNA on attractive surface noteworthy and is worth exploring.

Apart from temperature, solvent and confinement, another property which drives the dynamics of biopolymer is different types of gradients in the living system, such as chemical gradient, thermal gradient, gradient arising due to geometry, etc [33–35]. A chemical gradient (salt gradient) may increase or decrease the motion of biomolecules [36]. Similarly, a thermal gradient may give rise convection flow [37, 38]. Different geometries, such as MSPA, viral capsid induces geometrical gradient arising due to the gradual increase or decrease in configurational entropy of biopolymers as it approaches or go away from the confined geometry.

Manipulation of configurational properties of biomolecules in above mentioned processes (melting, unzipping, adsorption-desorption, etc.) may be understood in the framework of statistical mechanics. The sophistication in computer simulation studies may shed deeper insight in the mechanism involved in such processes. The main advantage of computer simulations relies on explicitly modeling all the degrees of freedom of the biopolymeric system. Despite limitations in length and time scales, the physical properties of the biopolymers in a given constraint can be extracted quite accurately using the simulation.

1.1 Theoretical modelling of polymer

In recent years, a large amount of data has been generated by different SMFS techniques, which provided the various properties of biomolecules [39]. The two state model explains some of the properties as seen in the experiments [40]. However, there are many intriguing issues, which are beyond the two state model. A considerable interest has been generated to understand such issues. Since, the number of monomers in these experiments varies in the range $10^4 - 10^5$, different models of the polymer / biopolymer have been developed in the framework of statistical mechanics. Broadly they are classified in two categories: (i) Continuum model, and (ii) lattice model.

The simplest representation of polymer is a random walk (RW) model which can be realized both in continuum and discrete space. The underlying concept is straightforward, that the walk can intersect itself. In continuum space model, there is no restriction in bond length (distance between two monomers), whereas in lattice RW follows fixed bond length. Next we will discuss the random walk model and illustrate its limitations. These limitations may be overcome by self-avoiding walk (SAW) which is a subset of random walk, but resembles many properties of polymers.

1.1.1 Continuum model

A polymer chain in solution changes its shape continuously, where the instantaneous shape is known as conformation. The Freely Jointed Chain (FJC), Gaussian chain, Worm-like chain (WLC), etc., are few examples of modelling of polymer in the continuum limit. All these models belong to the class of ideal polymer where monomer of the chain can intersect and belong to the same universality of random walk [3, 5].

Freely Jointed Chain (FJC)

The simplest polymer model is the Freely Jointed Chain (FJC) (Fig. 1.6). The bond length (b) is fixed between two adjacent monomers. We may associate a vector $\vec{r}$ with each bond which orients randomly and is in-

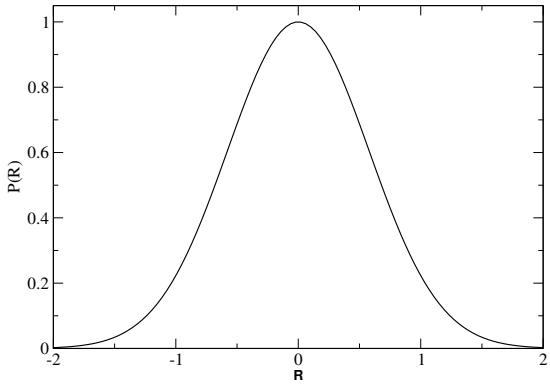

Figure 1.6. Probability distribution $P(\vec{R}, N)$ for FJC model of polymer.

dependent of its previous bonds [3, 5]. For a polymer of N bonds with starting end at origin and the other end with position vector $\vec{R_N}$ may be defined as

$$R(\vec{N}) = \sum_{l=1}^{N} \vec{r_l}, \tag{1.1}$$

The probability distribution of end-to-end vector $\vec{R}$(N) is given by,

$$P(\vec{R}, N) = \int d\vec{r_1} \int d\vec{r_2} .... \int d\vec{r_N} \delta(\vec{R} - \sum_{i=1}^{N} \vec{r_i}) \phi(\vec{r_i}). \tag{1.2}$$

$\phi(\vec{r_i})$ is the bond-vector distribution which is given by,

$$\phi(\vec{r_i}) = \frac{1}{4\pi b^2} \delta(\vec{r} - b) \tag{1.3}$$

which inturn gives

$$P(\vec{R}, N) = \frac{3}{2\pi N b^2} exp(-\frac{3R^2}{2Nb^2}) \tag{1.4}$$

A fixed bond length in physical system is a far-fletched simplification and in general there are fluctuations found in bond lengths which are

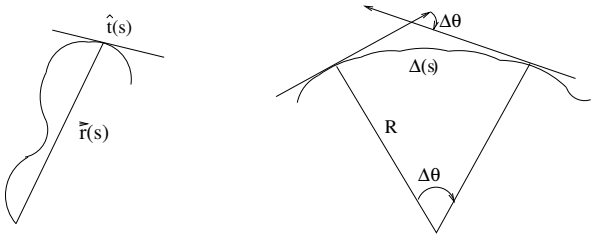

Figure 1.7. Worm-like chain model

independent and given by,

$$\begin{aligned} \langle \vec{r_l} \rangle &= 0 \\ \langle \vec{r_l} \vec{r}_m \rangle &= b^2 \delta_{lm} \end{aligned} \tag{1.5}$$

with the distribution,

$$\phi(\vec{r}) = (\frac{3}{2\pi b^2})^{3/2} exp(-\frac{3r^2}{2b^2}) \tag{1.6}$$

This is the **Gaussian** distribution of the distance between two points along the length of the polymer and may be written as,

$$P(\vec{r_l} - \vec{r}_m, l-m) = (\frac{3}{2\pi b^2 |l-m|})^{3/2} exp(\frac{-3R^2}{2|l-m|b^2}) \tag{1.7}$$

A continuous description of the polymer chain where energy is associated with bending of chain is termed as **Worm-like chain** model (Fig. 1.7). If s denotes the distance of a point on the polymer from one end, then the function $\vec{r}(s)$ denotes the position vector of the particular point with s running from 0 to L. The tangent vector at any point s is given by,

$$t(s) = \frac{\partial \vec{r(s)}}{\partial s} \tag{1.8}$$

where $t(s)$ is a unit vector *i.e.* tangent to the curve at point s. Now, in order to define bending energy, we need curvature parameter ($\kappa(s)$)

which is defined as,

$$\kappa = \frac{1}{R} = \lim_{\Delta s \to 0} \frac{\Delta \theta}{\Delta s} = |\frac{\partial^2 \vec{r}(s)}{\partial s^2}| \tag{1.9}$$

This in turn gives bending energy of WLC model which is quadratic in curvature κ and can be written as,

$$E = \int_0^L ds[\frac{a}{2}(\frac{\partial^2 \vec{r}(s)}{\partial s^2})^2] \tag{1.10}$$

where a is the rigidity parameter which has the dimension of (energy $\times$ length).

1.1.2 Lattice model

Polymer modelling in discrete space is known as the lattice model of polymer. The simplest model of the polymer is a random walk model where it follows the mechanism of the FJC chain with restricted bond angles associated with the lattice. The polymer size characteristics are the same as $\langle \vec{r_i} \rangle = 0$ and $\langle r^2 \rangle = Nb$.

In physical systems, random walk representation of polymers holds less significance as a single point in space (lattice) can be accessed by an infinite number of monomers. In order to avoid this discrepancy in modelling, there are several models which are shown in Fig. **??**. One of them is directed walks (DW) model in which self-interaction is excluded. There are two types of DWs: (i) Fully Directed walk (FDW), and (ii) Partially Directed walk (PDW). In FDW a walker can move only along the preferred direction. For example, in a two-dimensional lattice model, only $+x$ and $+y$ direction. In PDW, a walker can move along $\pm y$ direction, however it can move only in $+x$ direction, *i.e.* walker cannot go backward [4].

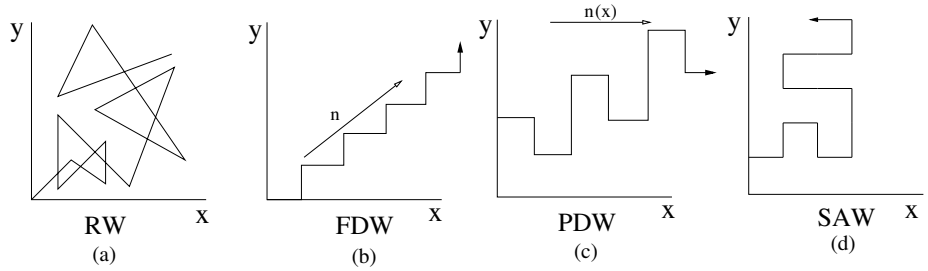

Figure 1.8. Schematic representation of the (a) Random walk (RW), (b) Fully Directed walk (FDW), (c) Partially Directed Walk (PDW), and (d) Self-avoiding Walk (SAW).

1.1.3 Self-avoiding Walk

The oversimplification of DW makes it unrealistic to model the polymer which can acquire the confirmation in any direction ($\pm x$, $\pm y$, and $\pm z$). The occupancy of the same lattice site by the walker arises when the walker moves in $\pm$ direction. In order to avoid self occupancy, the concept of excluded volume is introduced where two monomers cannot occupy the same site. If this concept is introduced in RW, we get a self-avoiding walk (SAW) where the walker is not allowed to visit sites which are already visited by the walker (Fig. **??** (d)). For this model,

$$\langle R^2 \rangle = N^{2\nu}, \tag{1.11}$$

where $\nu = \frac{3}{d+2}$ which is known as the Flory relation. This is the remarkable relation which gives the exact value of $\nu = 1, 0.75, 0.5$ for dimension $d = 1, 2$ and 4 respectively and differs by 0.001 to the best reported value for $d = 3$ [1, 2, 4]. The relation may be obtained by considering the self-avoiding walk of N monomers with mean field interaction energy among the monomers in three dimension as,

$$U \propto \frac{N^2 b^3}{R^3} \tag{1.12}$$

Therefore, the total free energy is given by,

$$\frac{F}{K_B T} = \frac{N^2 b^3}{R^3} + \frac{R^2}{N b^2} \tag{1.13}$$

Now, minimizing the free enrgy for equilibrium condition at $\bar{R}$ reduces to the following equation,

$$\Big(\frac{\partial F/k_B T}{\partial R}\Big)_{R=\bar{R}} = 0 \qquad \Rightarrow \qquad \bar{R} \propto b N^{3/5}, \tag{1.14}$$

This gives $\nu \approx 3/5$, very close to the experimentally observed value $\nu = 0.59$.

de Gennes and des Cloizeaux [2] established a relation between polymer statistics and q-vector spin model of magnetization in the limit $q \to 0$. This correspondence allows polymer science to be benefitted from the vast knowledge of critical phenomena. A polymer chain in good solvent can be described by SAW [3, 4]. If z is step fugacity in the SAW, then associated generating function can be expressed as:

$$C(z) = \lim_{p\to\infty} \frac{1}{p} \sum_{N=1} C_N(p) z^N \tag{1.15}$$

where p is the number of lattice points and C_N is the distinct number of SAWs for a N step walk. We define,

$$\bar{C_N} = \lim_{p\to\infty} \frac{C_N(p)}{p} \tag{1.16}$$

The number of configurations (C_N) increases exponentially. It is reported that the number of configurations,

$$\bar{C_N} \sim \mu^N N^{\gamma-1} \tag{1.17}$$

where, $\mu = log(\frac{C_{N+1}}{C_N}) = 2.638159..$ for square lattice and $\gamma = 43/32$ is critical constant.

Apart from self-avoidance, polymer can have long range interaction in

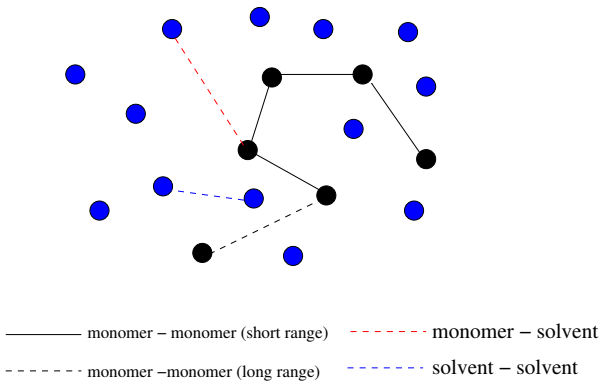

Figure 1.9. Schematic representation of different types of interactions in the solvent.

the solution [2]. These interactions are well descibed in three categories (Fig. 1.9):

- Monomer - Monomer interaction $\sim \frac{T}{2}\chi_{MM}\phi^2$.
- Monomer - Solvent interaction $\sim T\chi_{MS}\phi 1-\phi$
- Solvent - Solvent interaction $\frac{T}{2}\chi_{SS}(1-\phi)^2$

Where T is the temperature and ϕ is the fraction of sites occupied by monomers in solution. The Flory interaction parameter is given by,

$$\chi = \chi_{MS} - \frac{1}{2}(\chi_{MM} + \chi_{SS}) \tag{1.18}$$

Now, if χ is positive, then the solvent is good. If χ is negative, solvent is poor and for $\chi = 0$ solution is athermal (Free energy is only from entropy) [1, 2]. In a good solvent, the polymer is in coil form, whereas in the poor solvent, the self attraction in free energy dominates to have a globular shape of the polymer. Changing the solvent quality or temperature leads polymer to acquire conformation from the coil state to the globule state and the transition is referred to as coil-glouble transition. The temperature at which the coil-glouble transition takes place is known as θ-temperature [4].

1.2 Modelling of DNA

In the following we discuss few models of DNA which successfully describe the equilibrium properties of DNA.

1.2.1 Poland-Scheraga Model

Poland-Scheraga model considers the DNA molecule is composed of alternating region of zipped and bubbles (regions with broken base-pairs) formed by two linear polymer chains (strands) interacting with each other with suitable constraint [41]. The bubble state (loop) is entropically favorable, while the zipped region (bound) is energetically rich. Both of these factors contribute to making the free energy minimum in the equilibrium condition. In this simplified description, details of DNA, *e.g.* stiffness, helicity, chemical composition, excluded volume effect of nucleotides etc. have been ignored. Later the model was generalized by Fisher to take care of the excluded volume effect [42].

1.2.2 Peyrard and Bishop Model

In 1989, Peyrard and Bishop proposed a dsDNA model, which assumed that each strand of dsDNA is a series of point masses (beads), corresponding to each nucleotide [43, 44]. In this model, the vibration of each bead in transverse direction along the strands is represented by harmonic potential. While Morse potential used in the P-B model represents the interactions within complementary strands as well as the solvent effect. These models do not allow formations of hairpin, DNA condensations, non-native interactions between bases. The role of conformational entropy is minimized in these models, as they are mostly $(1+1)$ dimensional models. These shortcomings can be resolved in the lattice representations of DNA.

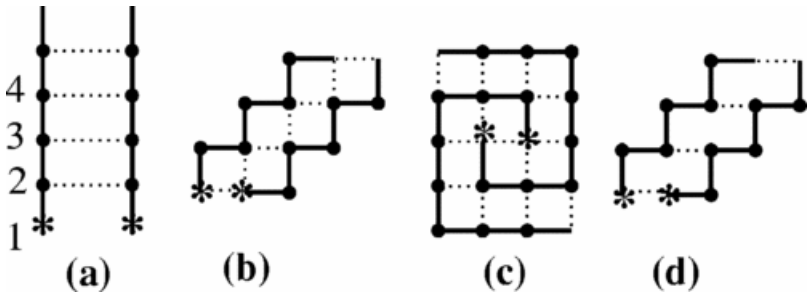

Figure 1.10. Schematic representations of lattice moddels of DNA: (a)-(d) represent the possible models for lattice representations of DNA. For model A, (a) and (b) are two possible states with (c) representing a possible ground state. For model B, (a) represents the ground state and (d) represents a partial bound state. Here, (d) differs from (b) in the nature of interactions represented by the dotted lines. In model B, (c) has no valid interaction and would represent an open state.

1.2.3 Lattice model of DNA

In minimal representation of DNA (lattice model)(Fig. 1.10), we consider two models A and B, which differ only in the interaction type in between the two strands [45]. Let us consider two mutually attracting self avoiding polymer chains that can not cross each-other. Monomers are called 'bases' and interaction among 'bases' are called base-pair. There is an attractive interaction between monomers or bases only if they are of opposite strands and are nearest neighbors on the lattice. This nearest neighbor attraction mimics the short range nature of hydrogen bonds in DNA. In model A, any monomer of one strand can interact with any monomer of the other strand. In model B only native base-pairing is possible such that i^{th} monomer of one strand makes base-pair with i^{th} monomer of the other strand. In both models, base-pairing energy is $\epsilon = -1$ and originating lattice point of each strand is kept fixed with a single lattice parameter distance.

1.3 Methods: Theory and Simulations

1.3.1 Exact Enumeration technique

The thermodynamic properties associated with the collapse of polymer or unzipping transition of DNA are obtained from the partition function which can be written as a sum over all possible configurations of polymer or DNA[4],

$$Z_N(\omega, u) = \sum_{m,y} C(m, y)\omega^m u^y. \tag{1.19}$$

Here N is the chain length of the polymer or of each of the two strands of DNA. $\omega = exp(1/T)$ is the Boltzmann weight associated with each non-bonded nearest neighbor and m is the number of nearest neighbors in case of polymer. Now, for dsDNA representation, m describes the number of bound base-pairs. $u = exp(g/T)$ is Boltzmann weight associated with the force. Here, y is the distance of interest. As an example, for DNA unzipping by pulling two strands apart, y represents the distance between the two ends of the DNA. Analysis of the partition function through series analysis gives greater accuracy by suitable extrapolation technique. To achieve the same accuracy by the Monte Carlo method, a chain of about two orders of magnitude larger than in the exact enumeration method must be considered.

1.3.2 Monte-Carlo simulation

Real polymers comprise a large number of monomers ($10^4 - 10^5$), so a better understanding of polymer behavior on a large length scale is needed sometimes [46]. Various Monte-Carlo techniques, such as Monte Carlo Metropolis, Wang-Landau simulation, Bond Fluctuation model, Pruned Enriched Rosenbluth Method (PERM), are used to study the large system [46]. Next, we describe briefly different moves associated with Monte-Carlo simulation.

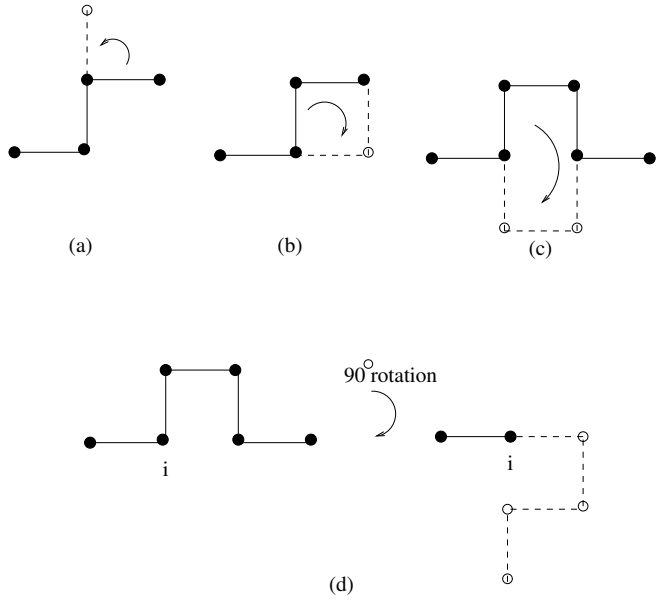

Figure 1.11. Schematic representation of various moves: a) end-point rotation, b) kink, and c) crankshaft moves. These are called local moves. Figure (d) represents pivot move (global move) where the rotation of polymer can be done after i^{th} monomer.

Moves

The accuracy of Monte-Carlo methods depends on the structural moves which try to scan the configurational space accessible to the polymer as much as possible. Moves are broadly categorized in two ways,

Local moves

Single monomer moves such as endpoint rotation (one bond length changed), corner flip (monomer in between two bonds changed) are often used. Whereas crankshaft is a local move where two monomers and three bonds positions are changed (Fig. 1.11). However, these moves suffer from correlated configurations [47, 48].

Global moves

Pivot move in lattice polymer simulation is an efficient move where any monomer(i) along the polymer (length N) is chosen randomly and by

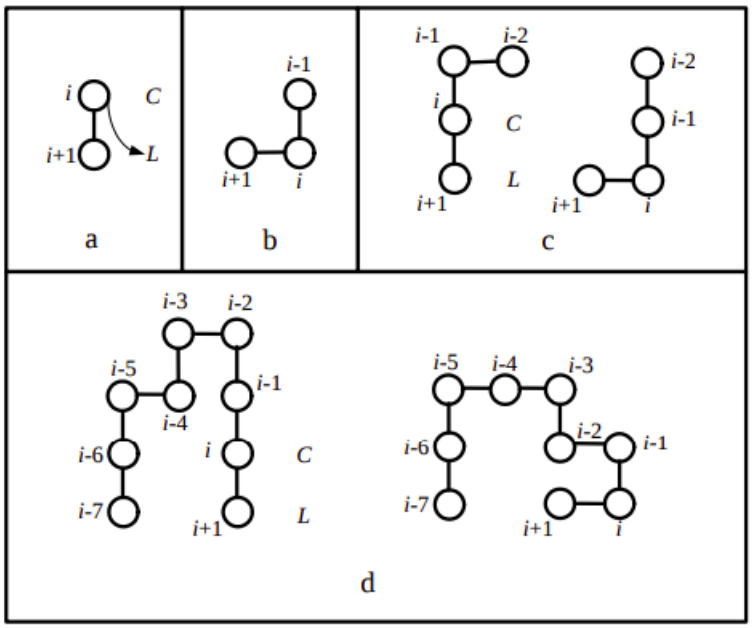

Figure 1.12. Schematic representation of global pull moves. This figure is taken from the ref. [49]

arbitrary rotation (90 or -90 degree rotation) the rest of the polymer ($i+1 \rightarrow N$) orientation position is changed. This works nicely for long polymer chains but lacks efficiency in compact polymer structures or low-temperature cases. To avoid that, another efficient Monte Carlo move, Pull moves, is used, which is very efficient in generating compact self-avoiding walks in lattice settings.

In the pull move (Fig. 1.12) a) vertex i shifts to a free lattice site L. (b) In the case where the fourth corner site C holds vertex $i-1$; the move is complete. (c) Otherwise, vertex $i-1$ is shifted to C. This does not complete the move. Then (d) as it does not complete the move, vertex $i-2$ is moved to the lattice site earlier occupied by vertex i, vertex $i-3$ now is shifted to the site earlier held by vertex $i-1$, and so on until a valid self-avoiding configuration is reached. In this example, $i-3$ is shifted which completes the move.

The Monte Carlo technique allows as much as possible independent

configuration based on a random number generator. Any physical observable can be calculated by using the average of all such configurations. The master equation governs the probability of choosing a new configurational state n over m:

$$\frac{dP_n(t)}{dT} = -\sum_{n \neq m} [\omega_{nm} P_n(t) - \omega_{mn} P_m(t)], \tag{1.20}$$

where $P_n(t)$, $P_m(t)$ is the probability of having the system (polymer) at state n and at state m in time t. ω_{nm} and ω_{mn} are the transition probabilities from state n to m and m to n respectively. For equilibrium condition $\frac{dP_n(t)}{dT} = 0$, leading to

$$\omega_{nm} P_n(t) = \omega_{mn} P_m(t) \tag{1.21}$$

This is known as detailed balance condition and must be satisfied during the simulation run. So, the ratio of transition probabilities,

$$\frac{\omega_{nm}}{\omega_{mn}} = \frac{P_m}{P_n} \tag{1.22}$$

And the selection criterion is given by,

$$\omega_{nm} = min(\frac{P_m}{P_n}, 1) \tag{1.23}$$

This is the basis of the Monte Carlo method. In the next section, we briefly describe two important MC simulation techniques: (i) Metropolis technique (ii) Wang-Landau technique.

1.3.3 Metropolis technique

Starting with an initial configuration of polymer with energy E_n at state n, the probability of the state is given by [46],

$$P(E_n, t) = \frac{exp(-E_n K_B T)}{Z} \tag{1.24}$$

Where Z is the partition function and is given by,

$$Z = \sum_n exp(-\frac{E_n}{K_B T}) \tag{1.25}$$

In general during the simulation E_n is unknown, but this is bypassed by the Markov process as we are interested only on the selection criterion which depends on $\frac{P_m}{P_n}$ or the factor $\exp(-\Delta E/K_B T)$, where $\Delta E = E_n - E_m$. So, the probability of choosing the state m over n eventually depends on,

$$\omega_{nm} = min(exp(-\frac{\Delta E}{K_B T}), 1) \tag{1.26}$$

which depicts that if $E_m < E_n$, the probability of choosing the m state is 1. This is in demand because the system always tends to have lower energy (equilibrium). If $E_m > E_n$ then the exponential factor is compared with random number r $(0 \rightarrow 1)$ and if $r < \exp(-\frac{\Delta E}{K_B T})$, then only the state with energy E_m is considered, otherwise simulation carries on with the previous energy E_n. The process usually goes on for 2 x 10^8 steps where 10^8 steps are for equilibration, and the next 10^8 steps are for calculating average properties. During the simulation (after the equilibrium), various properties of the system is measured, and average values are given by,

$$\langle A \rangle = \frac{1}{S} \sum_i^S A(x_i) \tag{1.27}$$

where S is the total number of Monte Carlo steps. This is very straightforward to work with but comes with difficulties. As the simulation directly depends on the temperature, there are some metastable states at low temperatures where polymer gets trapped, and averaging does not give an accurate picture. Also, the polymer does not behave accurately near the phase transition, which is known as critical slowing down. From systems where transitions from one state to another as a function of time (Monte Carlo moves) need to be studied, the Metropolis method fits the bill. But for the thermal phase diagram where polymer configurational states for

the whole temperature range need to be analyzed, the Metropolis method lacks accuracy.

1.3.4 Wang-Landau technique

The challenges in determining the average thermodynamical properties in low temperatures can be overcome by generalized histogram techniques such as the Wang-Landau simulation method. Wang-Landau simulation deals with the density of states (DOS) $g(E)$, which is the number of configurations for a given energy state E and is independent of the system temperature [46]. In this case, the walker moves randomly on the energy space of the system and visits every distinct energy point corresponding to different conformations of the polymer. If the maximum and minimum energy of the polymer is known prior to the simulation (which may be known by ground-state search of the model), that energy space $E_{min} < E < E_{max}$ can be divided uniformly in energy bins (E_i) with equal interval. The initial DOS attributed to each E is $g(E) = 1$ and $h(E) = 0$ for all E, where $h(E)$ is the quantity that monitors the flatness of the histogram of energy visits of the walker. Along with the modification factor f (which is initially set $\ln(f) = 1$), $h(E)$ controls the simulation run. A complete flow-chart diagram of Wang-Landau technique is shown in Fig. 1.13.

A Monte Carlo trial move from a state A with energy E_A to a state B with energy E_B is accepted on the basis of the transition probability

$$P(A \to B) = \min(\frac{g(E_A)}{g(E_B)}, 1). \tag{1.28}$$

After each successful move, we update DOS in the following way: $g(E_B) = g(E_B) \cdot f$ and $h(E_B) = h(E_B) + 1$, otherwise $g(E_A)$ and $h(E_A)$ are updated in the same fashion. Energy distribution in $h(E)$ is checked in each run, once $h(E) \geq 0.8 < h(E) >$, where $< h(E) >$ is the average histogram, then the modification factor is reduced to $f_{i+1} = \sqrt{f_i}$. The simulation continues until $\ln(f)$ reaches its final value 10^{-6}, which is a standard choice

for WL simulation and then the final $g(E)$s are extracted from the system. We can calculate the final DOSs, and corresponding partition function of the system as,

$$Z(\beta) = \sum_E g(E) \exp(-\beta E) \tag{1.29}$$

Now, to calculate the physical quantities of the system (average nucleotide density, average number of base pairs, end-to-end distance etc.), we perform the simulation once without updating the final DOSs $g(E)$ for all E and following the same acceptance rule (Eq. 5). When all the bins of the energy histogram $h(E)$ have reached a sufficient number of hits ($\sim 10^7$), we stop the sampling and the observable quantity Q can be calculated in the following way:

$$< Q(\beta) >= \frac{1}{Z(\beta)} \sum_E g(E) \exp(-\beta E) \overline{Q(E)}, \tag{1.30}$$

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

$$\overline{Q(E)} = \frac{\sum_Q h(E,Q)Q}{\sum_Q h(E,Q)}, \tag{1.31}$$

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

Here we have briefly surveyed the literature on polymers and biopolymers in confined environment. We have highlighted some significant issues which will be addressed in the following chapters. In order to do this we have discussed few models of biopolymers and techniques required to calculate the physical observable in the framework of statistical mechanics.

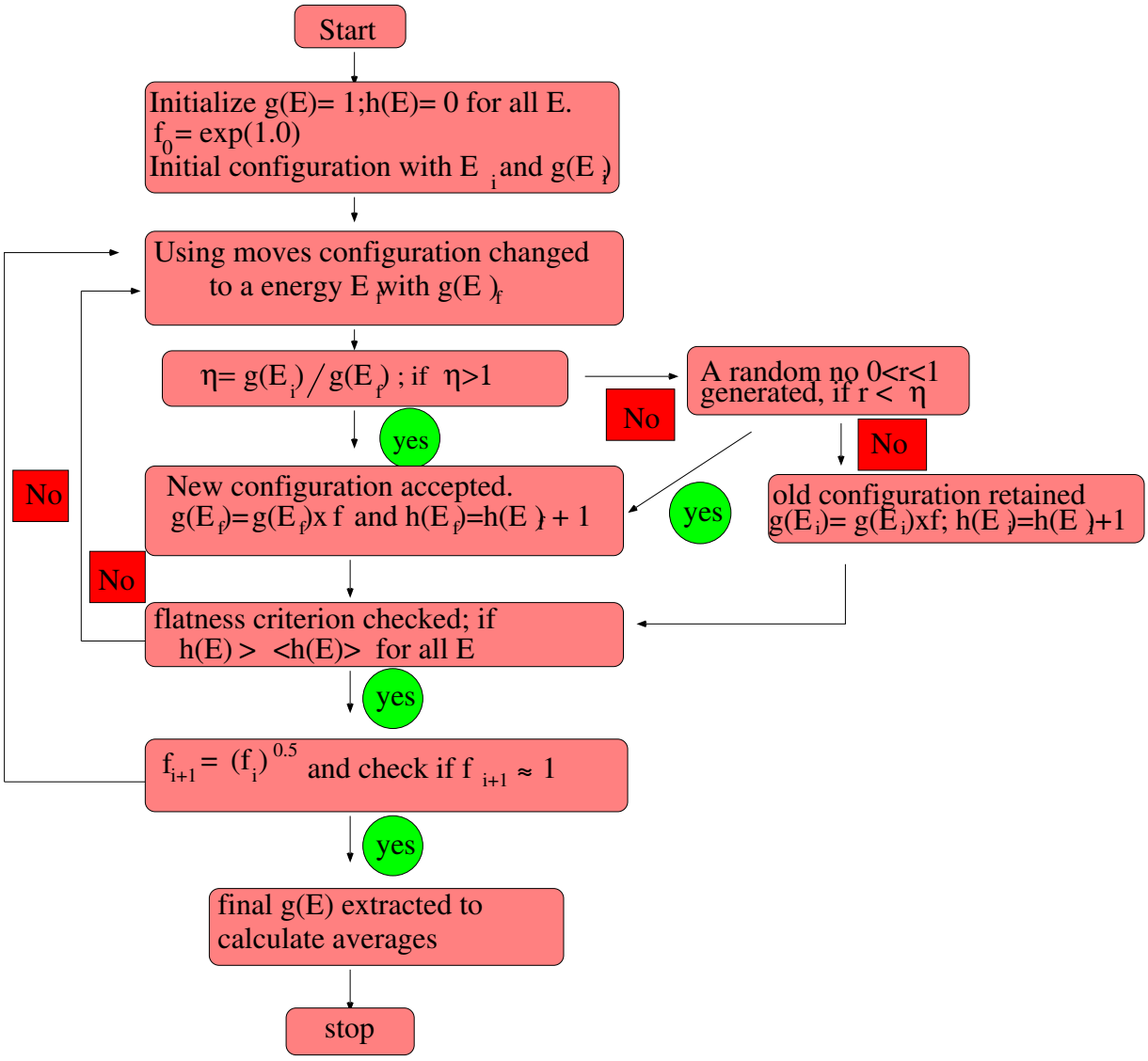

Figure 1.13. Flowchart of Wang-Landau algorithm