2015

Biomolecular Folding Rates As Understood From Single-Reaction-Coordinate Langevin Dynamics And Kramers' Theory

Md Adnan Kabir
University of Mississippi

BIOMOLECULAR FOLDING RATES AS UNDERSTOOD FROM SINGLE-REACTION-COORDINATE LANGEVIN DYNAMICS AND KRAMERS' THEORY

A Thesis
presented in partial fulfillment of requirements
for the degree of Master of Science
in the Department of Physics and Astronomy
The University of Mississippi

Submitted by

MD ADNAN KABIR

December, 2015

ABSTRACT

Langevin dynamics was used to model the folding and unfolding of simple, hairpin-like biomolecules whose ends are attached to laser-trapped beads, as occurs in optical tweezers experiments. The Langevin process was evolved numerically, using parameters motivated by real experimental systems. Folding trajectories were generated and analyzed to extract the folding rate as a function of the force applied to the beads. The observed rate was compared to the analytical predictions of Kramers' theory. Strong discrepancies were noted. The failure of the Kramers' theory was attributed to the slow dynamical response of the beads, which it does not account for. The results of this work highlight the necessity to include in the modeling the experimental systems that mediate force along the length of the biomolecule.

DEDICATION

Dedicated

To

My wife

ACKNOWLEDGEMENT

I express my heartiest regards, profound gratitude to my supervisor, Dr. Kevin Beach, Assistant Professor, Department of Physics and Astronomy, University of Mississippi, for his constant patience, guidance, keen interest and valuable inspiration to carry out my research work.

I am also grateful and expressing my deepest appreciates to my committee members Dr. Luca Bombelli who is also my grad-advisor and Dr. Cecille Alista Labuda for giving their valuable time for me and thankful to my department for providing me the adequate financial support.

Finally, I want to give special thanks to my beloved wife for her continuous support and active encouragement to pursue my graduate study abroad.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Proteins

Proteins are large biological molecules or macromolecules that carry out vital functions within living cells. They are the most important molecules found in living organisms. They do most of the work in cells and are required for the structure and function of the body's tissues and organs, including the support of the skeleton, muscle movements, digestion of food and protection against infection [1]. They come in all shapes and sizes. These are the most abundant components within a cell, and they make up nearly twenty percent of the total body weight.

Proteins consist of one or more long chains of building blocks known as amino acids. Each amino acid consists of an amino group (NH_2), a carboxyl group (COOH), a hydrogen atom (H), and a fourth group, referred to as the R-group (side chain). The basic structure is common to all amino acids; they differ because of the presence of the side chain. Some of these side chains are nonpolar and hydrophobic (water-repelling), others are hydrophilic (water-attracting). This diversity in side-chain properties defines the characteristic of each protein.

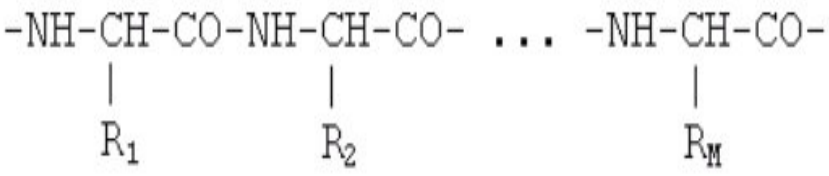


Figure 1: Chemical formula of a protein. Figure reproduced from Ref. 2.

Proteins are linear polymers; they are built up by amino acids that are linked into a peptide (covalent chemical bond formed between two amino acid molecules) chain. The sequence of amino acid residues (a "residue" is the portion of a free amino acid that remains after polymerization) is unique for each protein [2]. There are twenty different forms of amino acid residues found in our body (Figure 2) and they vary dramatically between amino acids. For example, if the R-group consists only of hydrogen, the amino acid is known as glycine while methyl (CH_3) group is known as alanine and a more complex structure found in tryptophan.

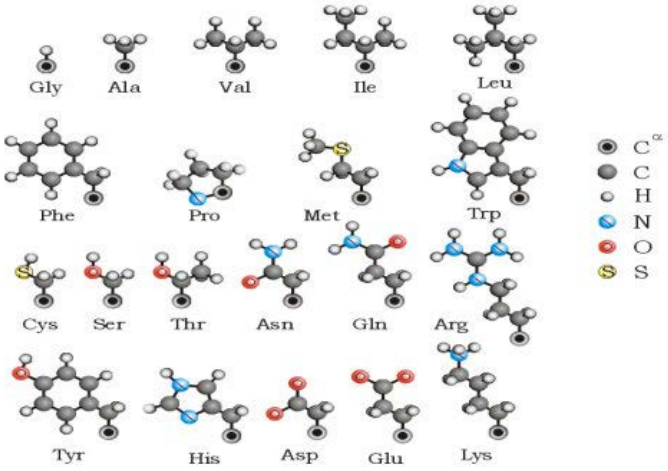


Figure 2: The side chains of 20 standard amino acid residues. Figure reproduced from Ref. 2.

1.2 Protein Folding

Countless efforts have been made to study the properties of proteins and protein folding. Protein folding is a physical process by which a polypeptide chain folds into its functional shape or conformation from a random coil [5, 8]. This conformation is a well-defined three-dimensional structure (Figure 3) and is known as the folded state or the "native state" of the proteins. This resulting three-dimensional structure is determined by the amino acid sequence in the protein [9].

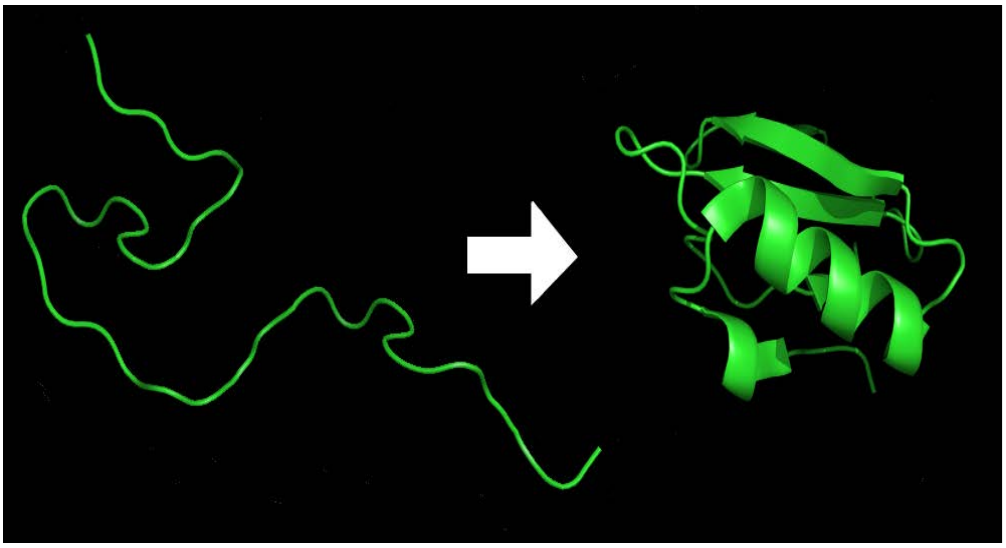


Figure 3: An unfolded protein (left) and its native state (right). Figure reproduced from Ref. 3.

Each amino acid in the chain has different chemical features, for example, hydrophobic, hydrophilic, or electrically charged. These amino acids interact with each other and their surroundings in the cell to produce the folded protein. Protein folding occurs very rapidly, from microseconds to seconds and produces molecules with specific biological functions. For a standard protein, the free energy gap between the folded and the unfolded state at room temperature is of the order of 20 k_BT ~ 10 kcal/mol for a ~ 100 residue domain [5, 6]. Therefore, protein stays folded and functional with a free energy of a fraction of k_BT per amino acid.

1.3 Kinetic and Thermodynamic Hypotheses

The kinetics and thermodynamics are two major issues in protein folding. Historically, two prevalent theories, thermodynamic and kinetic hypothesis have been proposed. The thermodynamic hypothesis was first suggested by Christian Anfinsen by his pioneering work in 1957, and later in 1968 C. Levinthal proposed the kinetic hypothesis which provides an alternative view of the native state of protein [10].

According to the thermodynamic hypothesis the native state is unique and thermodynamically most stable state for proteins. Proteins remain folded under normal physiological conditions and keep the lowest possible Gibbs free energy [11]. The native state depends only on the amino acid sequence and the environment it is in. It does not depend on the kinetic folding routes. Moreover, there are no other configurations that can have a comparable free energy to challenge the free

energy minimum. The thermodynamic hypothesis has been widely used in computer simulations to study the folding dynamics for its better predictions of the native structures. It simplifies calculations, yielding results which are in strong agreement with experimental findings [9].

On the other hand, the kinetic hypothesis predicts that the native state of protein is kinetically stable but not thermodynamically and is at a local energy minimum [12]. There are a large number of kinetically inaccessible states in the conformational space might exist which have lower free energy and thus are more stable than the native state.

CHAPTER 2

FOLDING MECHANISMS AND MODELS

2.1 Hydrophobic Collapse Theory

Many models have been proposed to explain the protein folding problem; the hydrophobic collapse model is one of them. It predicts that the native conformation of proteins forms by rearrangement of a compact collapsed intermediate structure which is also referred to as a molten globule structure, therefore constitutes an early step in the folding pathway. This model is based on the fact that most folded proteins have a hydrophobic core in which side chain packing stabilizes the folded state, leaving most of the polar residues on the solvent exposed protein surface to interact with surrounding water molecules [16, 17, 18]. It is generally accepted that sequestration of hydrophobic side chains exposed to water stabilizes the folding intermediates. The energy of the intermediate states is lower than that of the denatured state but higher than that of the native state. Hydrophobic collapse is an early event in the folding pathway and it occurs before the formation of secondary structures. It is believed that the compact intermediate or "molten globule" was larger than the native form of the protein and has an intact secondary structure. Any polypeptide chain of near-native composition and length will exist either in the molten globule state, or as an ensemble of such states when placed in water [19, 20].

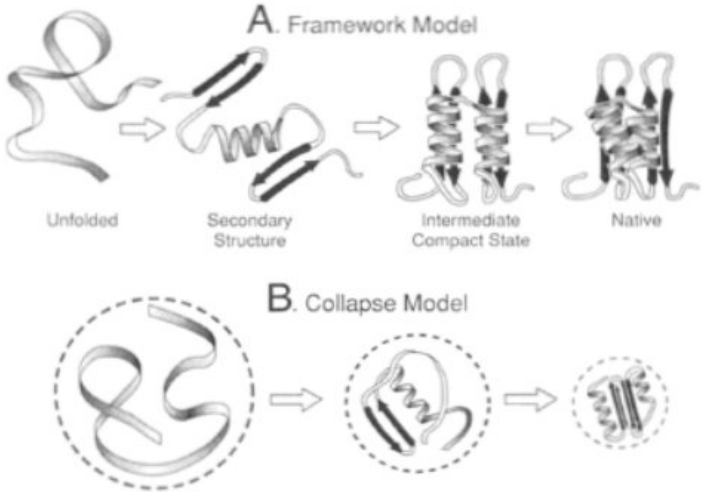


Figure 4: The framework model (A) and the collapse model (B): (A) Describes the model outlined by Ptitsyn et al. which dictates that secondary structure form first, and is followed by tertiary structure formation. (B) Outlines the initial collapse of the molecule into a smaller structure with tertiary spatial organization, and then initiates secondary structure formation. Figure reproduced from Ref. 16.

2.2 Multiple Folding Pathways

Extensive experimental and theoretical efforts have been made to study the kinetics of protein folding. This specific model predicts that there is no specific way that leads proteins from the random state to the native state. According to this theory, proteins may fold to their unique native state along multiple unpredictable kinetic pathways and intermediate conformations instead of a single specific way. This theory is also named as funnel landscape theory because of its funnel shaped energy landscape. This was originally proposed by Dill et al. and it is based on the thermodynamic hypothesis which stated the native state was at an energy minimum [19]. An example of a funnel landscape is given in Figure 5 which pictures that protein must fold in an energetically downhill manner as they go down along the *z* axis.

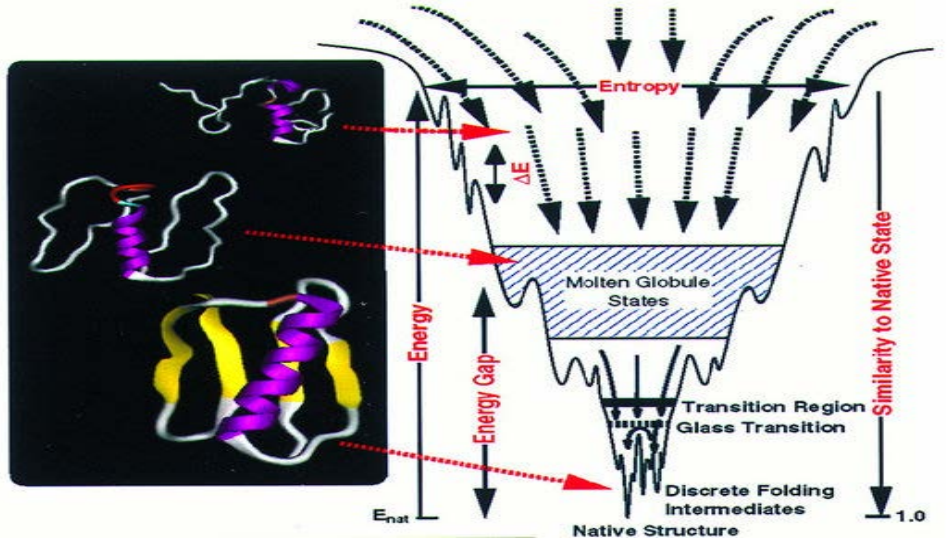


Figure 5: A schematic of funnel landscape of protein folding. Figure reproduced from Ref. 22.

Along this energy landscape the partially folded protein can be trapped at many non-native local minima while the theory assumes that the native state is a free energy minimum. This theory has been used in both computational simulations and experimental studies to improve methods for protein structure prediction and design.

2.3 Two-State Folding

Protein folding is a complex process and therefore it is very difficult to obtain a quantitative understanding of protein folding mechanism. To avoid the complexities of the other models we have discussed so far and to overcome the difficulties in analyzing the folding data, the two-state folding model has been introduced. This is the most popular mechanism currently used to study protein denaturation process.

This is the simplest model where it is assumed the protein folding process is an all-or-none reaction in which protein is in either the fully folded (F) state or fully unfolded state (U) and there is no intermediate state. The protein folding process can be described in a straightforward method with a before and after state as a reaction with an activation barrier. The equilibrium condition can be written as:

$$\frac{[F]}{[U]} = e^{(E_{\mathrm{f}} - E_{\mathrm{u}})/k_{\mathrm{B}}T}, \qquad (1)$$

where, [F] and [U] are the instantaneous (time-dependent) concentrations of the folded and unfolded conformations, E_f and E_u are the energy barriers for folding and unfolding respectively, k_B is Boltzmann's constant, and T is the temperature of the system. Furthermore, the folding and unfolding rates can be written as:

$$R_f = Ae^{-E_f/k_B T}, \quad (2)$$

$$R_u = Ae^{-E_u/k_B T}, \quad (3)$$

where R_f and R_u are the folding and unfolding rates respectively, and A is the normalization constant. Using these kinetic rates, one may obtain a quantitative understanding of the protein folding process and ignore the complexity of the other models for folding.

2.4 Necessity of Protein Folding

Proteins are "nano-machines" and they are responsible for most of the biological functions in living organisms. To carry out these functions, protein must 'fold' first. They cannot do their job without being folded. Therefore, understanding of how proteins fold is very interesting and important topic. Protein folding also prevents wrong interaction between proteins in the intercellular environment where several hundred thousand proteins are present. Misfolded proteins cause serious diseases in animal body. Neurodegenerative and other diseases such as Alzheimer's disease (caused by amyloid β-protein), Parkinson's disease (α-Synuclein) and Creutzfeldt-Jakob disease (Prion protein), bovine spongiform encephalopathy (mad cow disease) are believed to result from the accumulation of misfolded proteins. Therefore, it may be more important to understand exactly why some proteins do not fold.

Protein folding depends on its amino acid sequence and it folds spontaneously during or after synthesis. The folding mechanism also depends equally on the nature of the primary solvent (water or lipid), the concentration of salts, and the temperature of the environment. Cells need a specific temperature range to live in. So temperatures below or above this range will cause thermally

unstable proteins. Moreover, protein stability depends on the free energy change between the folded and unfolded states which is expressed by the following,

$$-RT\ln K = \Delta G = \Delta H - T\Delta S\,, \qquad (4)$$

where R represents the Avogadro number, K is the equilibrium constant, ΔG the free energy change between folded and unfolded states, ΔH the enthalpy change and ΔS the entropy change from folded to unfolded states. Proteins become more stable if the free energy of the unfolded protein (G_u) increases relative to the free energy of the folded or native protein (G_f) or the entropy difference between two states decreases. Increase in binding energy (dispersion forces, electrostatic interactions, van der Waals potentials and hydrogen bonding) also helps the folded protein to become more stable [21, 25].

CHAPTER 3

PROTEIN FOLDING SIMULATION ALGORITHMS

3.1 Monte Carlo Algorithm

Monte Carlo (MC) is a general method in computational calculations by which thermodynamics and structural properties of an arbitrary system can be obtained. The name has come from the famous Monaco casino to emphasize the importance of randomness [24]. MC algorithm is a stochastic approach where a set of parameter values are selected randomly and a function having these parameters is evaluated by numerical integration. In conventional MC simulations only the potential energy is required for calculations [27]. For very small chain configurations the average thermal energy and entropy can be computed by the following formula:

$$\langle A \rangle = \frac{\int dr^N \exp[-\beta U(\boldsymbol{r}^N)] A(\boldsymbol{r}^N)}{\int dr^N \exp[-\beta U(\boldsymbol{r}^N)]} \qquad (5)$$

where $\beta = 1/k_{\mathrm{B}}T,\ U$ is the potential energy, $\boldsymbol{r}^N$ is the configuration of an N particle system (i.e., the positions of all N particles). Since it is not needed to solve Newton's equation of motions in MC simulations, dynamical information cannot be gathered by using these techniques. Moreover, it is not useful in defining the efficient moves for molecules with large chain since it is very difficult to devise the simple structural perturbations in these molecules [27]. Therefore, conventional MC methods are inefficient for exploring the configurational space of large biomolecules when compared to molecular dynamics.

3.2 Molecular Dynamics (MD) Algorithm

This simulation technique was first introduced by Alder and Wainwright in the late 50s. It was used to simulate perfectly elastic collisions between hard spheres [26]. Now MD is frequently

used to refine the three-dimensional structures of proteins by simulating the folding of a polypeptide chain from a random coil. This simulation technique is highly used to study the microscopic properties of a biological system including the atomic positions and velocities of the biomolecule in the system [26, 28]. This method follows Newton's second law in the classical limits. For a system of classical particles Newton's equation of motion are applied and can be written as:

$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{F}_i(t), \qquad i = 1,2,3,\dots,N \ , \tag{6}$$

where $\boldsymbol{r}_i$ is the position vector of the i^{th} particle and $\ddot{\boldsymbol{r}}_i$ is the corresponding acceleration, $\boldsymbol{F}_i$ is the vector of forces acting on the i^{th} particle. The positions at times $(t + \Delta t)$ and $(t - \Delta t)$ can be written using the Taylor expansion around time *t*,

$$\boldsymbol{r}_i(t + \Delta t) = \boldsymbol{r}_i(t) + \dot{\boldsymbol{r}}_i(t)\Delta t + \frac{1}{2}\ddot{\boldsymbol{r}}_i(t)\Delta t^2 + \frac{1}{6}\dddot{\boldsymbol{r}}_i(t)\Delta t^3 + O(\Delta t^4) \ , \tag{7}$$

$$\boldsymbol{r}_i(t - \Delta t) = \boldsymbol{r}_i(t) - \dot{\boldsymbol{r}}_i(t)\Delta t + \frac{1}{2}\ddot{\boldsymbol{r}}_i(t)\Delta t^2 - \frac{1}{6}\dddot{\boldsymbol{r}}_i(t)\Delta t^3 + O(\Delta t^4). \tag{8}$$

The sum of equations (7) and (8) gives,

$$\boldsymbol{r}_i(t + \Delta t) + \boldsymbol{r}_i(t - \Delta t) = 2\boldsymbol{r}_i(t) + \ddot{\boldsymbol{r}}_i(t)\Delta t^2 + \ O(\Delta t^4). \tag{9}$$

Using equation (6) the following equation is obtained:

$$\boldsymbol{r}_i(t + \Delta t) = 2\boldsymbol{r}_i(t) - \boldsymbol{r}_i(t - \Delta t) + \frac{1}{m_i}\boldsymbol{F}_i(t)\Delta t^3 + \ O(\Delta t^4). \tag{10}$$

We should calculate equation (10) iteratively to obtain trajectories of atoms in the system. Moreover, in this deterministic process, one can yield the trajectory from the force on each atom by integrating the equations of motion and, thus get the information about the positions, velocities and accelerations of that particular atom.

The important advantage one can get by using the MD simulation over MC is that it is time dependent and it provides the ability to find the folding path for both small and large biomolecules.

But this simulation can be time consuming and computationally very expensive. However, calculations up to nanosecond time scale can be achieved in simulations of solvated proteins by using this method [28].

3.3 Langevin Dynamics

Now if the object is more complex than a classical particle or when the curvilinear coordinates are used, it is recommended to use generalized Lagrange equations of motions instead [29, 30]. But every biological process takes place at the molecular level where physics is different from that of the macroscopic world and requires a different approach. When the system is treated at the fully atomic level, Langevin dynamics is used to mimic the interactions of the solute molecules with the environment. Langevin dynamics was first introduced by French physicist Paul Langevin in 1908 on the motion of Brownian particles in the fluid [30]. Langevin dynamics mimics the viscous aspect of solvent, which eventually turns equation (6) into a stochastic differential equation expressed by equation (11):

$$\boldsymbol{F}_i = m\dot{\boldsymbol{v}}_i = -\nabla_{r_i} U(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N) - m_i\gamma_i\dot{\boldsymbol{r}}_i + \boldsymbol{R}_i(t), \tag{11}$$

$$\dot{\boldsymbol{r}} = v\,, \tag{12}$$

where $\dot{\boldsymbol{r}}_i$ and γ_i are the velocity and friction coefficient of atom *i*, respectively; U is the potential energy of the system ($-\nabla_{r_i} U$ being the potential force acting in the atom *i*) and $\boldsymbol{R}_i(t)$ are random forces arising from the collision of atom *i* with the molecules of the environment. The random force has zero mean and is completely uncorrelated at different times [31]. Therefore, we can write:

$$\langle \boldsymbol{R}_i(t) \rangle = 0, \tag{13}$$

$$\langle \boldsymbol{R}_i(t).\boldsymbol{R}_i(t') \rangle = 2m_i\gamma_i k_{\mathrm{B}} T\delta(t - t'), \tag{14}$$

where *T* is the absolute temperature of the system, k_{B} is the Boltzmann constant, and $\delta(x)$ is the Dirac δ-function.

Now to solve Newton's equation of motion numerically Verlet-type algorithms have been introduced for its low computation cost and higher accuracy. This method is based on the truncated

Taylor series expansions of a particle with mass m, coordinate $r(t)$, velocity $v(t)$ and force $f(r,t)$ [29, 32]. Therefore, the velocity Verlet can be written as:

$$r^{n+1} = r^n + v^n dt + \frac{dt^2}{2m} f^n, \tag{15}$$

$$v^{n+1} = v^n + \frac{dt}{2m}(f^n + f^{n+1}). \tag{16}$$

Now starting with the continuous Langevin equations (11) – (14), we are integrating equation (11) over a small time interval dt between two times $t_n \ and \ t_{n+1} = t_n + dt$

$$\int_{t_n}^{t_{n+1}} m\dot{v}\, dt' = \int_{t_n}^{t_{n+1}} f\, dt' - \int_{t_n}^{t_{n+1}} m\gamma\dot{r}\, dt' + \int_{t_n}^{t_{n+1}} R(t')\, dt'. \tag{17}$$

With no approximation, this can be written as

$$m(v^{n+1} - v^n) = \int_{t_n}^{t_{n+1}} f\, dt' - m\gamma(r^{n+1} - r^n) + R^{n+1}, \tag{18}$$

where

$$R^{n+1} = \int_{t_n}^{t_{n+1}} R(t')\ dt' \tag{19}$$

is a Gaussian random number with $\langle R^n \rangle = 0$ and $\langle R^n R^l \rangle = 2m_i\gamma_i k_{\mathrm{B}} T dt \delta_{n,l}$

Now integrating equation (12) gives

$$\int_n^{n+1} \dot{r}\, dt' = r^{n+1} - r^n = \int_{t_n}^{t_{n+1}} v\, dt', \tag{20}$$

which can be approximated with [32]

$$r^{n+1} - r^n \approx \frac{dt}{2}(v^{n+1} + v^n). \tag{21}$$

Now working with equation (18) and then replacing v^{n+1} from equation (22) into equation (21) gives us

$$(v^{n+1} - v^n) = \frac{1}{m}\int_{t_n}^{t_{n+1}} f\,dt' - \gamma(r^{n+1} - r^n) + \frac{1}{m}R^{n+1}\,, \tag{22}$$

$$=\gg \quad r^{n+1} - r^n = \frac{dt}{2}\left(v^n + \frac{1}{m}\int_{t_n}^{t_{n+1}} f\,dt' - \gamma(r^{n+1} - r^n) + \frac{1}{m}R^{n+1} + v^n\right)$$

$$=\gg \qquad r^{n+1} - r^n = \frac{dt}{2}\left(2v^n + \frac{1}{m}\int_{t_n}^{t_{n+1}} f\,dt' - \gamma(r^{n+1} - r^n) + \frac{1}{m}R^{n+1}\right)$$

$$=\gg \qquad (r^{n+1} - r^n)\ \ (1 + \frac{\gamma dt}{2}) = dtv^n + \frac{dt}{2m}\int_n^{n+1} f\,dt' + \frac{dt}{2m}R^{n+1}$$

$$r^{n+1} - r^n = bdtv^n + \frac{bdt}{2m}\int_n^{n+1} f\,dt' + \frac{bdt}{2m}R^{n+1}\,, \tag{23}$$

where

$$b = \frac{1}{1 + \frac{\gamma dt}{2}}\,. \tag{24}$$

For any R^n the equation (19) is correct to 2nd order in dt while equation (20) is not. Therefore the integral of the force was approximated in such a way that both equations are correct to 2nd order in dt [32]:

$$r^{n+1} = r^n bdtv^n + \frac{bdt^2}{2m}f^n + \frac{bdt}{2m}R^{n+1}\,, \tag{25}$$

$$v^{n+1} = v^n + \frac{dt}{2m}(f^n + f^{n+1}) - \gamma(r^{n+1} - r^n) + \frac{1}{m}R^{n+1}\,. \tag{26}$$

Now if $\gamma = 0$ the above equations reduce to the standard velocity-Verlet given in equations (15) and (16). Equations (25) and (26) constitute a simple functional Verlet-type scheme for solving stochastic Langevin equations and therefore we will use these two equations in the simulation later on.

3.4 Brownian Dynamics

Brownian dynamics (BD) is used when a molecule is surrounded by the solvent of high effective viscosity. In order to produce a representative diffusional trajectory in BD, a large friction coefficient γ_i has been introduced in the Langevin equation**.** We know the Langevin equation as:

$$m\dot{\boldsymbol{v}}_i = -\nabla_{r_i} U(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N) - m_i\gamma_i\dot{\boldsymbol{r}}_\iota + \boldsymbol{R}_i(t) \ . \qquad (27)$$

In the over-damped limit $\gamma \gg \omega$ (where ω is the characteristic frequency of the system), the dissipative terms ($m_i\gamma_i\dot{\boldsymbol{r}}_\iota$) dominate over the inertial terms ($m_i\ddot{\boldsymbol{r}}_\iota$) and equation (27) can be written as:

$$m_i\gamma_i\dot{\boldsymbol{r}}_\iota = -\nabla_{r_i} U(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N) + \boldsymbol{R}_i(t). \qquad (28)$$

Now after integrating equation (28) and doing some calculations we found:

$$\boldsymbol{r}_i(t + \Delta t) = \boldsymbol{r}_i(t) + \frac{\boldsymbol{F}_i(t)}{\gamma_i}\Delta t + \sqrt{\frac{2k_{\mathrm{B}}T}{\gamma_i}\Delta t}\ \boldsymbol{\omega}_i \ , \qquad (29)$$

where Δt is the time step and $\boldsymbol{\omega}_i$ is the random noise vector obtained from a Gaussian distribution.

Brownian dynamics is a useful and widely applied approach in studying the protein folding dynamics. Proteins are treated as rigid bodies in BD simulations and time scales in the microsecond or millisecond range are readily accessible [27]. It is useful for long time simulations. The main disadvantage of this method is that it cannot provide the correct values of the thermodynamic properties because of the neglected inertia terms [30]. Proper treatment of Langevin dynamics can avoid the problem.

Among all the techniques have been discussed so far, we will be using the Langevin dynamics for our folding simulation.

CHAPTER 4

REACTION COORDINATE & KRAMERS' THEORY

A reaction coordinate is a real coordinate system which represents the progress along a reaction pathway. It can be the extension of a particle, bond length, angle, etc. This abstract one-dimensional coordinate system is often plotted against free energy of a system in order to demonstrate a schematic form of the potential energy profile. Therefore, the kinetics of an *N*-body system can be projected directly onto a single reaction coordinate by integrating the cumulative positions of the system over time and then considering the position of the biomolecule as a variable of the coordinate.

The reaction rate theory was first introduced by Hendrik Kramers in 1940 to model the chemical reaction of classical particles of mass *m* moving in a one-dimensional double well potential $V(x)$ divided by a potential barrier (Figure 6). Here the double well potential represents the folded and unfolded state and the separating barrier along the reaction coordinate.

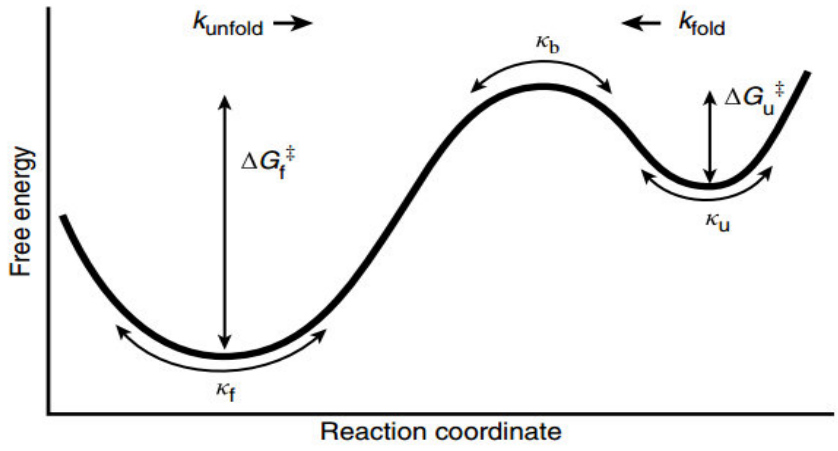


Figure 6: Schematic of the potential energy profile and the transition rates from Kramers' theory. Figure reproduced from Ref. 33.

For intermediate to strong friction Kramers' theory can be written as:

$$k = \frac{1}{\omega_{\mathrm{b}}}\left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \omega^2{}_{\mathrm{b}}} \right)\left\{ \frac{\omega_{\mathrm{R}}}{2\pi} \exp\left[\frac{-\Delta G}{k_{\mathrm{B}}T}\right] \right\}, \tag{30}$$

where, ω_{R}, ω_{b} are the frequencies of the confined potential well and barrier top, respectively, and γ is the friction coefficient. Moreover, for strong (over-damped) friction, $\gamma \gg \omega_{\mathrm{b}}$, the expression for the rate (equation 30) simplifies to

$$k = \frac{\omega_{\mathrm{b}}}{\gamma}\frac{\omega_{\mathrm{R}}}{2\pi} \exp\left[\frac{-\Delta G}{k_{\mathrm{B}}T}\right]. \tag{31}$$

After simple modification, folding and unfolding rates can be written in terms of the stiffness of the potential profile as [33, 34]:

$$k_{\mathrm{unfolded}} = \frac{\sqrt{\kappa_{\mathrm{f}}\kappa_{\mathrm{b}}}}{2\pi k_{\mathrm{B}}T} D\, e^{-\frac{\Delta G_{\mathrm{f}}}{k_{\mathrm{B}}T}}, \tag{32}$$

$$k_{\mathrm{folded}} = \frac{\sqrt{\kappa_{\mathrm{u}}\kappa_{\mathrm{b}}}}{2\pi k_{\mathrm{B}}T} D\, e^{-\frac{\Delta G_{\mathrm{u}}}{k_{\mathrm{B}}T}}, \tag{33}$$

where, $D = k_{\mathrm{B}}T/\gamma$ the diffusion constant over the barrier, κ_{f}, κ_{u} are the stiffness of the potential well in the folded and unfolded states respectively, and κ_{b} is the stiffness of the barrier. This stiffness can be measured by calculating the curvature of the potential wells and the top of the potential barrier. Moreover, $\Delta G_{\mathrm{f}}, \Delta G_{\mathrm{u}}$ are the heights of the energy barriers for the folded and unfolded states.

Now the inverse dependence on the friction coefficient γ in equations (32) and (33) plays an important role in finding the transition rates from the high-friction Kramers' rate theory. Due to the multiple, random re-crossings over the barrier, the transition rates reduce by a factor $\gamma/\sqrt{\kappa_{\mathrm{b}}}$. The transition rates also depend on the shape of the surface of the energy barrier, because smaller values of κ_{b} make the barrier broader, flatter, and allows more re-crossing over the barrier. On the other hand, greater curvature (larger κ_{b}) allows less re-crossing and enhances the reaction rate.

CHAPTER 5

OPTICAL TWEEZERS

The determination of the energy landscape of protein folding is one of the main approaches in this field. For many decades these landscapes have been determined with numerous techniques that measure the protein in traditional bulk methods. In these methods, many different folding routes and non-cumulative intermediate states are often not clearly resolved [36]. Recent progresses in single molecule manipulation techniques provide us a different approach to reinvestigate the protein folding problem.

In these techniques the single molecules can be directly manipulated and investigated with the help of a well-defined reaction coordinate. These methods present a number of advantages over the traditional techniques, allowing us: i) to study the different molecular conformations ii) to measure the kinetics and statistics of these complex biological processes iii) to measure the potential of mean force of a molecule as a function of its extension, iv) to explore the disorder of the folded structure, v) to predict the precise location of the bio molecules in the substrate over time.

Optical tweezers are scientific instruments that use a highly focused laser beam to trap nanometer and micron-sized dielectric molecules in space. It can hold and move the microscopic dielectric molecules by applying an attractive or repulsive force to them. The force acting on the molecule is typically in the order of piconewton. This technique was first introduced by Arthur Ashkin and colleagues in 1986 when they discovered that a tightly focused beam of light capable of holding microscopic particles of high refractive index in three dimensions [37].

In optical tweezers the center of the beam is known as the beam waist and it contains a very strong electric field gradient. Dielectric particles are attracted along this field gradient. The laser light also exerts a force on the particles along the direction of beam propagation. This force is

known as the scattering force and because of it particles are slightly displaced from the beam waist, as seen in Figure 7.

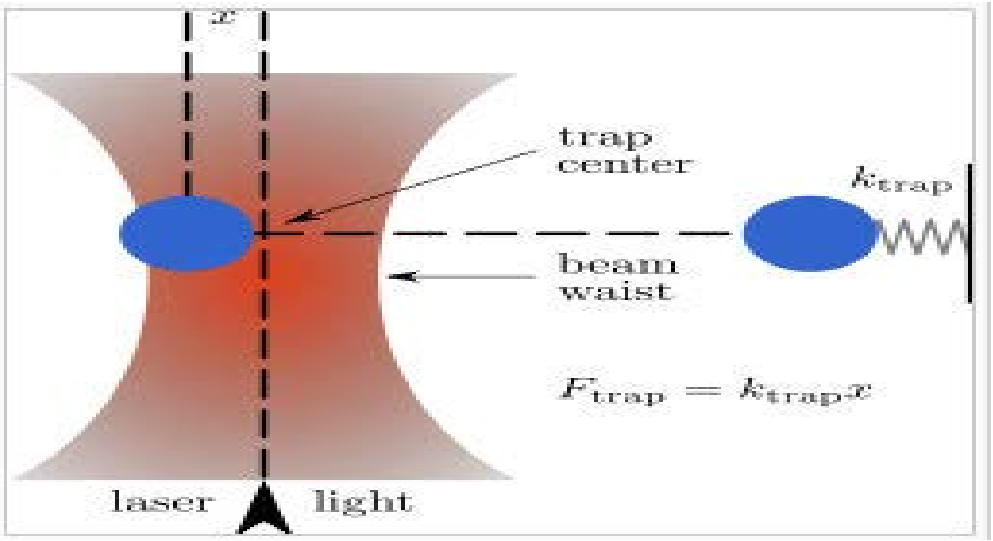


Figure 7: Schematic of optical tweezers. Figure reproduced from Ref. 38.

Moreover, the force applied to the dielectric particle is linear in its displacement from the center of the trap as long as the displacement is small. Therefore, for sufficiently small displacements from equilibrium, the trapping potential is harmonic. In this way, an optical trap can be compared to a simple spring, which follows Hooke's law:

$$F_{\mathrm{trap}} = k_{\mathrm{trap}}x\,, \qquad (34)$$

where F_{trap} is the applied force, k_{trap} the "spring constant" and x is the extension. A force applied to the molecule will push or pull it away from its equilibrium position. The bigger the force the bigger the displacement is from the equilibrium. Therefore, multiplying the measured displacement by the "spring constant" of the tweezers determines the size of the force.

A major task in single-molecule manipulation technique is to manipulate nanometer-sized proteins with micrometer sized beads because a direct attachment of the protein molecule to the beads would require the large tethering surfaces which increase the interaction to a maximum [36]. Thus it compromises the measurements. Therefore, to get rid of this problem, the double strand DNA molecules are used to connect the protein to polystyrene beads in order to keep the interactions between the tethering surfaces negligible [39].

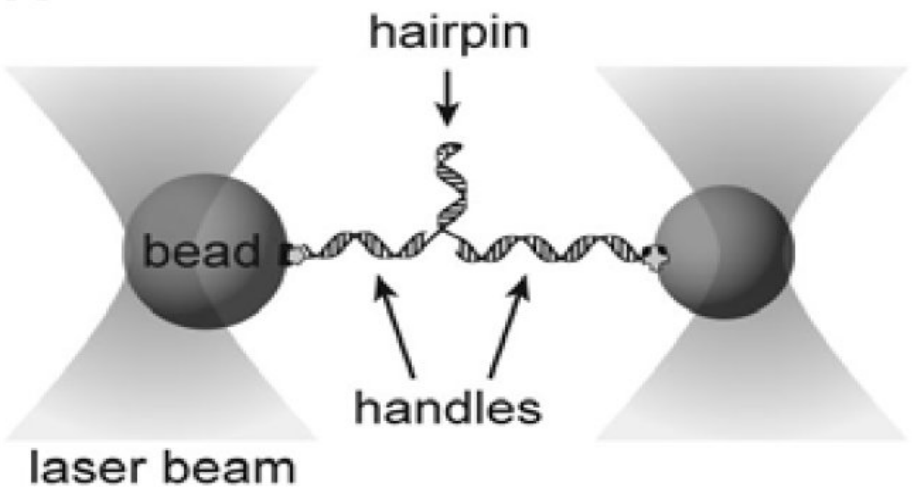


Figure 8: Schematic of a single hairpin held in optical traps. Figure reproduced from Ref. 41.

Single-stranded DNA form molecular structures, like hairpins and a hairpin constructs consisted of a single hairpin connected at each end to a double strand DNA handles of several hundred bp (base pairs) in length, which are attached to polystyrene beads of several hundred nanometers in diameter. The experiment is performed in the lab with two orthogonally polarized laser beams used to exert forces on the beads [40, 42]. Bead displacements in each trap are measured independently with two polarized detectors, and trap stiffness was determined by standard techniques. Folding and unfolding rates are obtained from the distribution of dwell times in the two states [40, 42].

CHAPTER 6

SIMULATIONS

The simulation is done in C++ and it modeled the experiments by assuming the diffusive motion of a biomolecule over one-dimensional potential profile to mimic that of the hairpin in Figure 8 where two potential wells are divided by a barrier [41]. Our main purpose was to obtain the folding and unfolding rates of the biomolecule in a 1D potential energy profile by using the Langevin dynamics and then compares these rates with that of calculated from Kramers' theory to get a better understanding of the reaction rate theory.

Therefore, to do that we have used a symmetric one-dimensional potential energy profile which was $V(x) = 0.33 \cdot (x^2 - 10)^2 - F \cdot x$ (Figure 9) in the simulation and Langevin dynamics have been used to find the extension trajectory of the biomolecule. In the potential two wells are defined as left and right well where they represent the folded and unfolded state of the protein folding respectively.

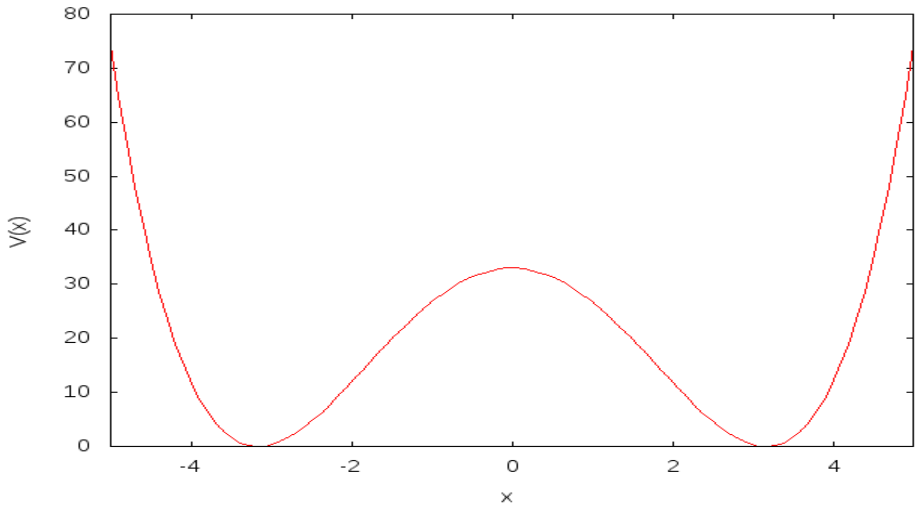


Figure 9: One-dimensional potential profile

In the simulation the biomolecule having mass m = 9.203 × 10^{-6} pg is tethered at one end with a bead of radius r = 410 nm via a spring with stiffness k = 0.4 pN/nm and a constant force F is applied as a tension to the bead. The time step Δt is 10^{-6} s and the thermal energy $k_B T$ is 4.1 pN·nm. The chosen time step is very small to ensure the thermal stability in the system. Time step Δt must be smaller than the characteristic relaxation time, $\tau = m/\gamma$ to get the pictures of the fast folding process. Moreover, there are two types of forces acting on the molecule and the bead and they are stochastic and non-stochastic forces. Stochastic forces on the molecule and bead are drawn from the Gaussian distribution of width $(2\gamma k_B T \Delta t)^{1/2}$ (with diffusion constant $D = 5.8 \times 10^{-12}$ m^2/s) and $\gamma = 6\pi\eta r$ (with viscosity $\eta = 10^{-3}$ Pa·s). $-V'(x_1) - k(x_2 - x_1)$ and $k(x_1 - x_2) - Fx_2$ are the non-stochastic forces acting on the molecule and bead respectively, where x_1 is the extension of the molecule, x_2 is the position of the bead and $V(x_1)$ is the potential landscape for the folding.

In this work x_1 is treated in terms of the Brownian dynamics whereas x_2 is treated according to the Langevin dynamics. The Langevin dynamics simulation was done using the Verlet-style algorithm as we stated before.

CHAPTER 7

RESULTS & DISCUSSIONS

The biomolecule tethered with the polystyrene beads usually folds as a two-state system under a constant force: folded and unfolded states. The extension trajectories measured under three constant forces are shown in Figure 10:

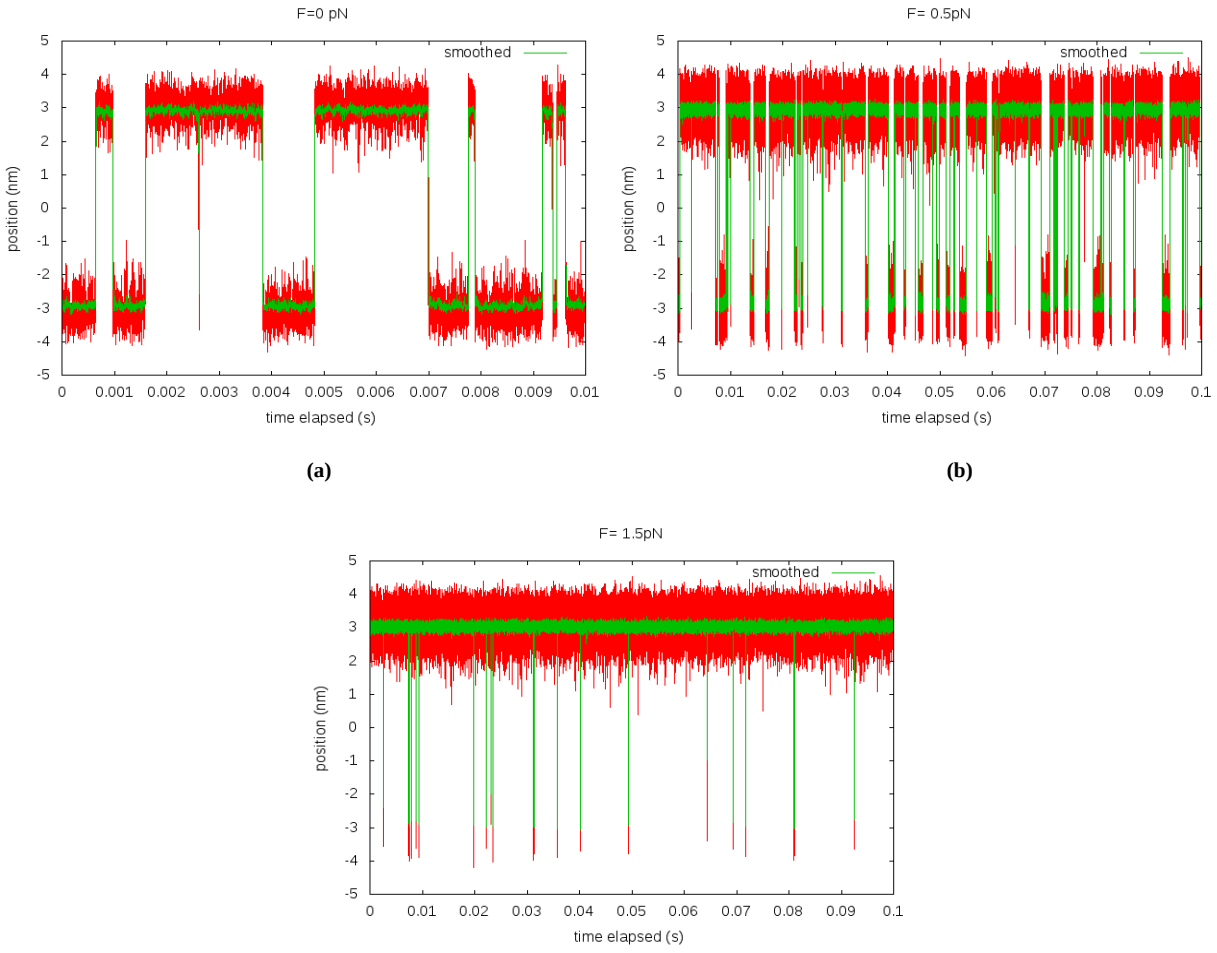
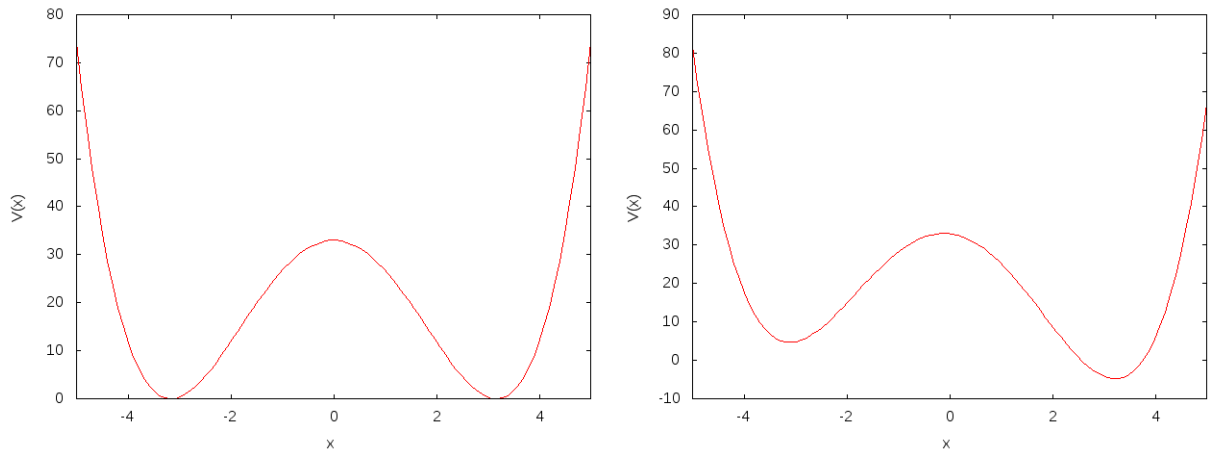


Figure 10: Extension trajectory of the biomolecule held at constant force (a) 0 pN, (b) 0.5 pN, (c) 1.5 pN, showing fluctuations between folded and unfolded state.

The force at which the two states are equally likely, means the biomolecule has 50% probability of being unfolded is donated by $F_{1/2}$. At this force the biomolecule exhibits hopping between the folded and unfolded states [41].

Extension trajectories have been made by calculating the cumulative positions of the biomolecule over time. Here the green data have been filtered to illustrate the two states more explicitly in order to get rid of the thermal noise present in the red data. Binomial filter has been used for the filtration process. We can see from the Figure 10 that as we increase the force the chance of getting the folded state diminishes which is as expected.

In our simulation we have used a symmetric one-dimensional potential energy profile which we stated before, and we let the molecule evolve back and forth in that profile. We applied the tension force F on the bead and kept it constant in the rest of the simulation. The force has been measured in the piconewton scale.

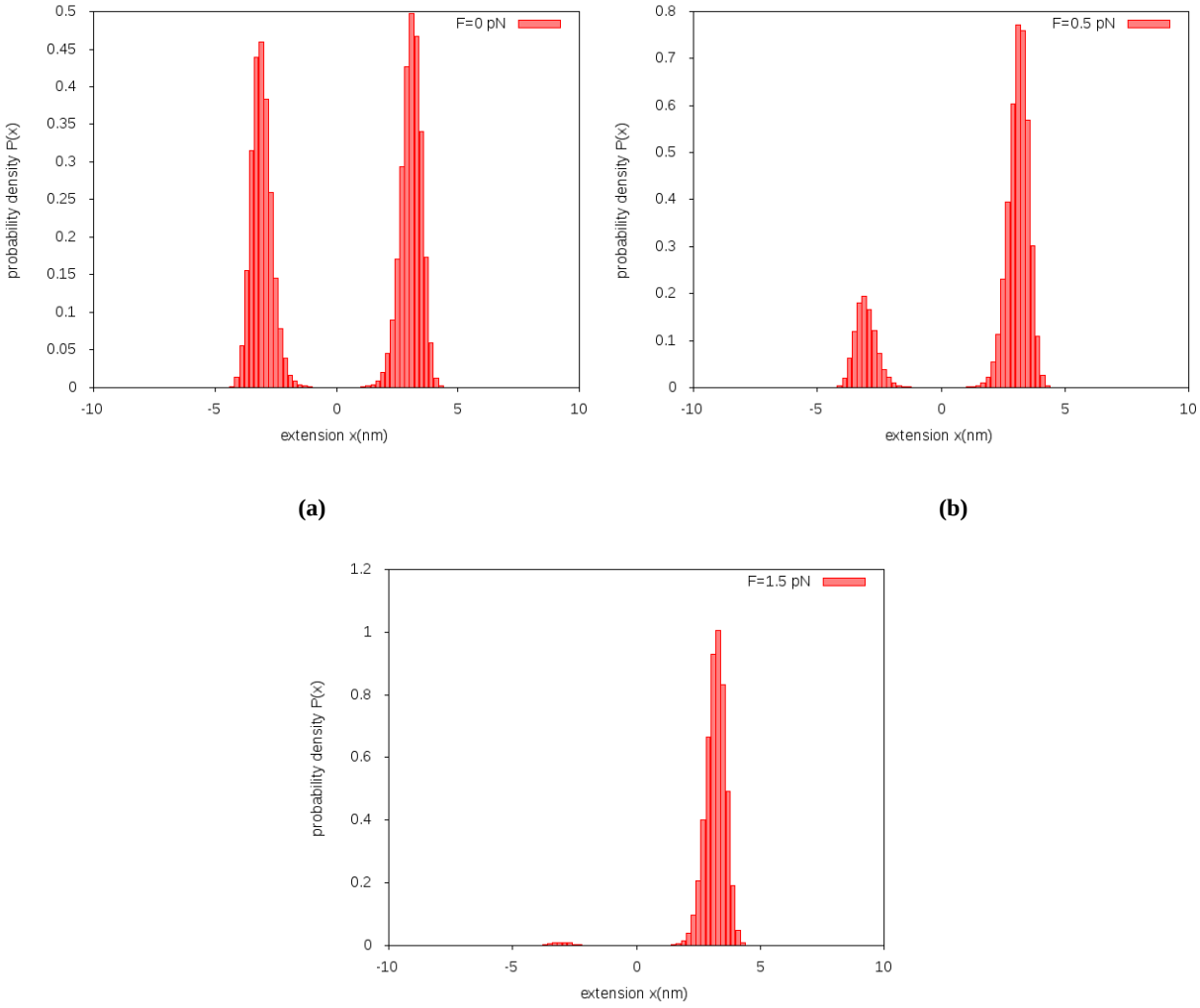


(a) **(b)**

Figure 11: Energy landscape varied with force (a) F = 0 pN, (b) F = 1.5 pN

The presence of two mechanical states indicates that the energy landscape of the biomolecule is dominated by two potential wells separated by a barrier. The applied force tilts the landscape (Figure 11) by changing the depth of the two wells. Here we can see that as we change the force the height of the energy barrier also changes keeping the position of the barrier constant.

We have varied the force from 0 pN to 1.5 pN and found $F_{1/2} \approx 0$ pN in good agreement with the potential energy. Figure 12 illustrates the distribution histogram of the extension trajectory of the biomolecule for three different values of the tension force.



(a) **(b)**

(c)

Figure 12: Distribution of the extension trajectory of the molecule under force (a) 0 pN, (b) 0.5 pN, (c) 1.5 pN

To get a better picture of the distribution of the biomolecule in the folded and unfolded states, we have fitted two Gaussians (Figure 13) to the extension histogram (Figure 12) which gives an extension of 6.25 ± 0.5 nm.

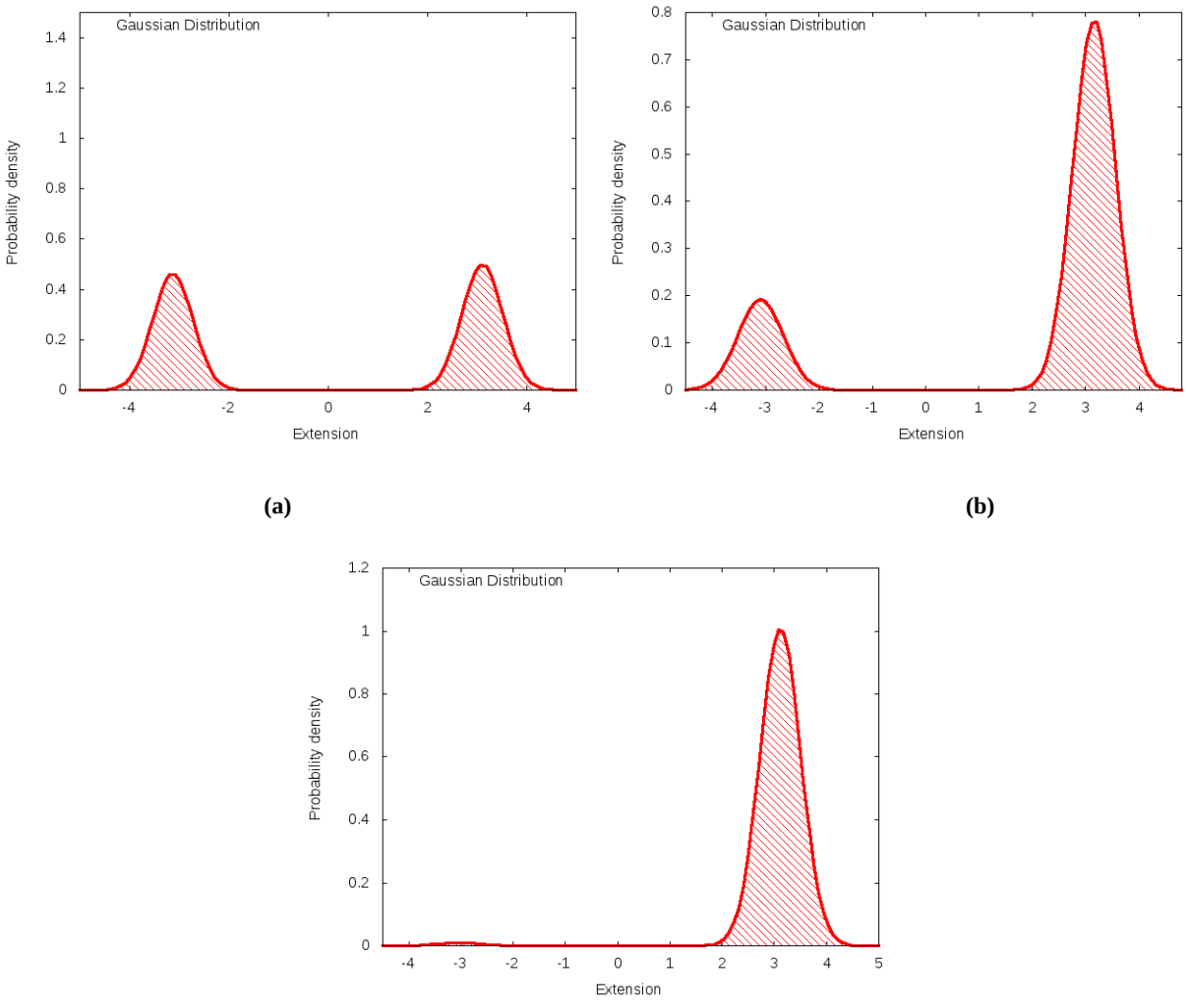
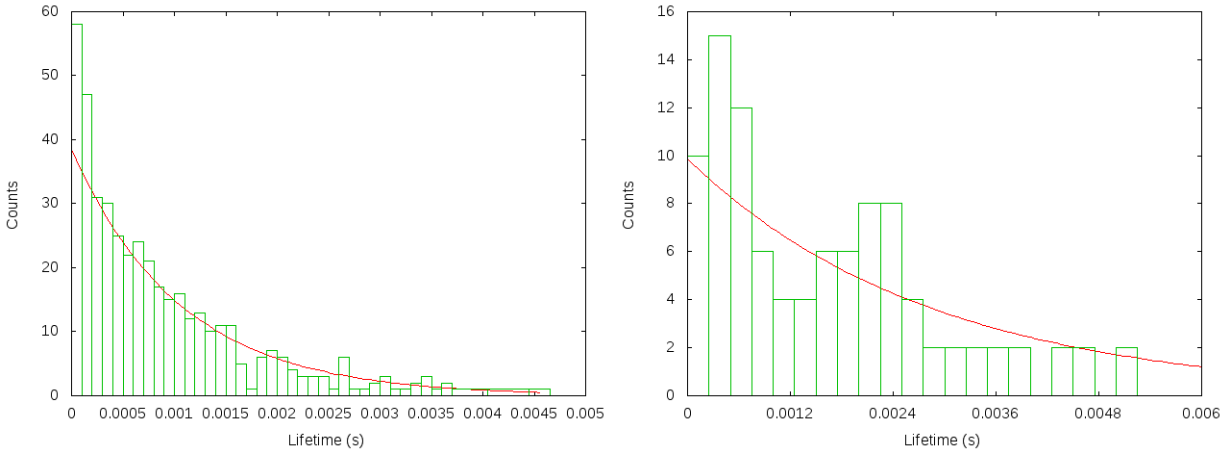


Figure 13: Gaussian distribution of the extension trajectory of the molecule under force (a) 0 pN, (b) 0.5 pN, (c) 1.5 pN

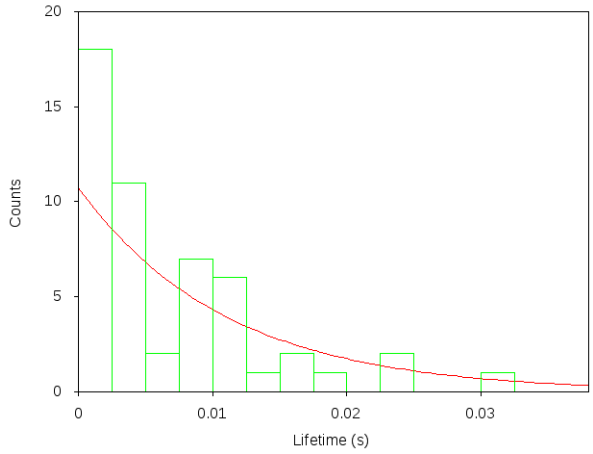
We have set the activation barrier as 2/3 of the actual barrier for both folding and unfolding processes, filtered out the data by dividing the regions as left (L), transition (T), and right (R) and then looked for the events that can be classified as 'LTT…TTR' (for unfolding) or 'RTT…TTL' (for folding) but not, for example, ' LTT...TTL'. The choice of the barrier height is quite arbitrary; this height was chosen because we have seen that a large fraction of re-crossing events happened just before that height thus has increased the accuracy of transition rates. We have run the program for about two days for each force to get better statistics and then plotted the resulting data as 'Counts vs. lifetime' in Figure 14. Lifetimes have been filtered out from the raw data and then plotted as histograms for folded and unfolded states. Figure 14 represents the lifetimes in the unfolded state for three different forces which are distributed exponentially, yielding the folding

rates. Unfolding rates can also be found following the same procedure. Furthermore, there was no restriction on the bin sizes. Different bin sizes have been chosen for different forces to get the best fit from the obtained data.



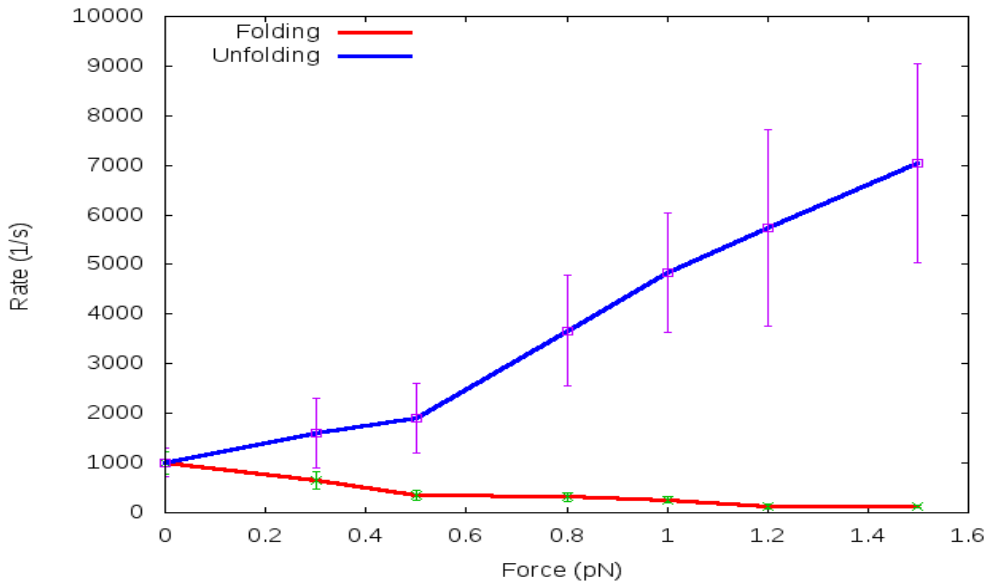
(a)

(b)

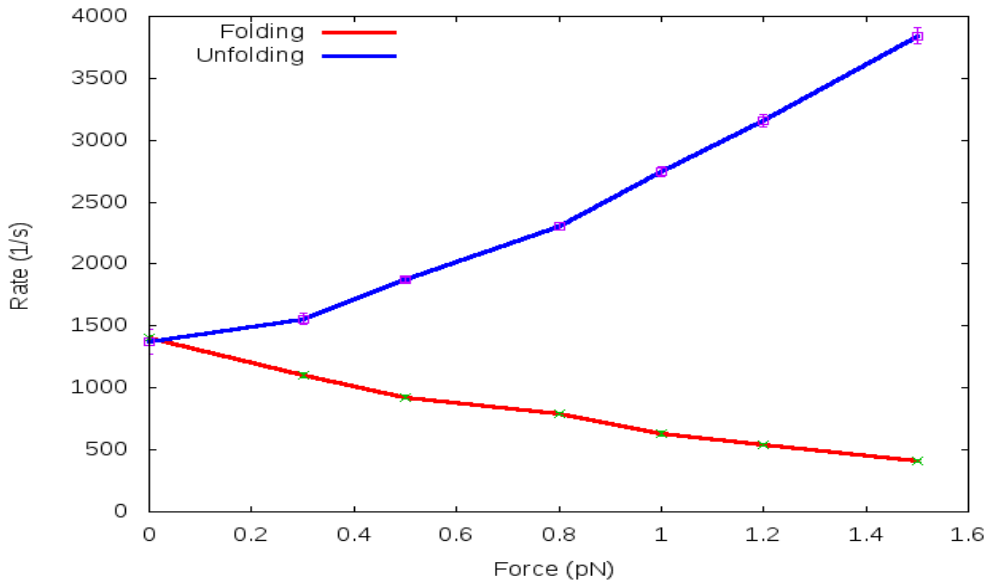


(c)

Figure 14: Exponentially distributed lifetimes in the unfolded state under force (a) 0 pN, (b) 0.5 pN, (c) 1.5 pN



(a)



(b)

Figure 15: Transition rates (with error bars) vary with applied forces (a) plot extracted from the simulation data, (b) plot extracted from Kramers' theory

Our job was to calculate the rates from Kramers' theory and then to compare those with the rates obtained from the simulation data. Because $\gamma \gg \omega_b$ equations (32) and (33) have been used to

calculate the unfolded and folded rates respectively where the stiffness of the wells and barrier top were measured by doing the quadratic fit in each region.

We have plotted the rates against the forces in Figure 15, from the simulation data and Kramers' theory and for both cases we can see that as we increase the forces the chance of being folded decreases thus lowering the folding rates with the increase of the unfolding rates.

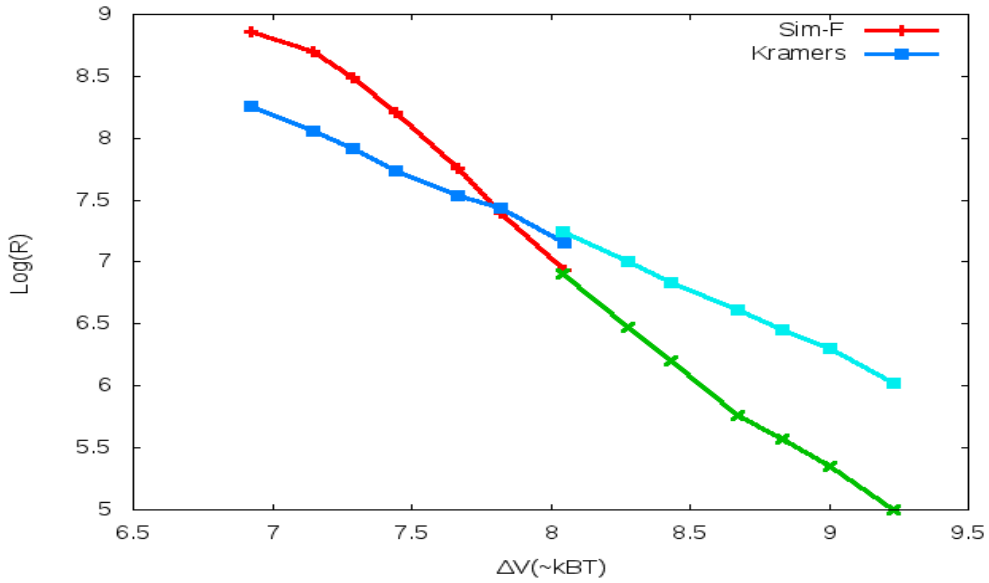


Figure 16: Comparison of the transition rates of the simulation and Kramers', (a) for unfolding and (b) for folding

Moreover, in Figure 16 we have illustrated the variation of the transition rates with the depth of the potential well and compared the values extracted from the simulation data and Kramers' theory. The red and green data represent the unfolding and folding rates obtained from the simulation whereas blue and light blue data have been calculated from Kramers'. We can see that at low barrier heights Kramers' line goes just under the simulation line because the transition rates reduced by a factor $\gamma/\sqrt{\kappa_b}$ which is as expected. But as the barrier height increased the folding rates become higher than the simulation folding rates data which is also reasonable to understand. Because Kramers' theory is just an assumption and it does not provide well prediction about the interaction between the biomolecule and bead molecule which reduces the folding the rates quite significantly when larger forces applied to the bead. Therefore, the prediction of folding rates from Kramers' theory is not well in the high barrier region.

CHAPTER 8

CONCLUSIONS

In this work we have studied the kinetics of the biomolecule with the help of Langevin dynamics and compared the transition rates obtained from the simulations with Kramers' theory. We found that Kramers' theory works well in a certain region of the potential barrier only. It would give better prediction about the transition rates if the interaction between the bead and biomolecule is taken into account. Moreover, a better statistic can be achieved if the program runs for a very long period of time which however requires high computation power and cost as well. Furthermore, we haven't studied kinetics of the bead molecule. It could be more challenging to extract the kinetic data of the bead molecule from the simulation since it's so noisy and it shows spurious transitions over time. Furthermore, we have considered the molecule can stay only in the folded and unfolded states and predicted that there might be a transition state in between. Therefore, finding the kinetics of this transition state could be a further study of my thesis.

Knowledge of protein folding has its own demands. As we already knew, protein makes 20% of our body mass and it helps our body to work. Misfolded protein causes numerous serious diseases. So it is highly important to get an idea on protein folding. Scientists have been trying for decades to get a clear picture of how proteins exactly fold and why? This requires the knowledge of physics, chemistry and biology. Therefore, further work on protein folding can give us a breakthrough in the development of physical science.

REFERENCES

1. Pietzsch, Joachim. The importance of protein folding. Retrieved from http://www.nature .com/ horizon/proteinfolding/background/importance.html
2. Antonie, Octavian. Langevin dynamics of protein folding: influence of confinement. Retrieved from http://www.ens-lyon.fr/DSM/SDMsite/M2/stages_M2/Antonie.pdf
3. Protein. Retrieved from https://en.wikipedia.org/wiki/Protein#cite_note-16
4. Dill, Ken A. and Hue Sun Chan. From Levinthal to pathways to funnels. Nature Structural Biology, 1997; 4(1).
5. Creighton, Michael E. Protein folding. New York Freeman W.H., 1992.
6. Branden, Carl, John Tooze. Introduction to protein structures. New York Garland Publishing, 1999.
7. Sneppen, Kim, Giovanni Zocchi. Physics in molecular biology. Oxford University Press, 2005.
8. Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walters. The shape and structure of proteins. Molecular Biology of the Cell; Fourth Edition. New York and London: Garland Science, 2002.
9. Anfinsen, C. The formation and stabilization of protein structure. Biochem. J., 1972; 128 (4): 737–49.
10. Levinthal, C. Are there pathways for protein folding? Extrait du Journal de Chimie Physique, 1968; 65: 44-45.
11. Voet, D. et al. Fundamentals of Biochemistry. John Wiley & Sons, Inc.
12. Onuchik, J.N. et al. Theory of protein folding. Current Opinion in Structural Biology, 2004; 14: 70-75.
13. Park, S. et al. Folding and misfolding of the collagen triple helix: Markov analysis of molecular dynamics simulations. Biophysical J., 2007; 93: 4108-4115.
14. Bai, Y. Hidden intermediates and levinthal paradox in the folding of small proteins. Biochemical and Biophysical Research Communications, 2003; 305: 785-788.
15. Ptitsyn, O. J. Protein folding: hypothesis and experiments. Protein Chem, 1987; 6: 273-293.
16. The physics of protein folding. Retrieved from http://www.physics.mcmaster.ca/ phys4a03/Example_essay.pdf

17. Dill, K.A. *et al*. Principles of protein folding--a perspective from simple exact models**.** Protein Science, 1995; 4: 561-602.
18. Sali, A. *et al*. How does a protein fold? Nature, 1994; 369: 248-251.
19. Creighton, T.E. Review Article on Protein folding. Biochem J., 1990; 270: 1-16.
20. Richards, F.M. The protein folding problem. Scientific American, 1991: 34-41.
21. Mulcahy, Stephen**.** The thermodynamics of protein folding. Retrieved from http://www.csn.ul.ie/~stephen/reports/unfolding.html
22. Chemical physics of protein folding. Retrieved from http://www.pnas.org/content/95/19/11037/F1.expansion.html
23. Scholtz, J.M., Baldwin R.L. The mechanism of alpha-helix formation by peptides. Annu. Rev. Biophys. Biomol. Struct., 1992; 21: 95-118.
24. Jackson, S.E. Folding and Design, 1998; 3: R82-R91.
25. Nosoh, Y., and Sekiguchi. Protein stability and stabilization through protein engineering. Wiseman, A., Ellis Horwood Limited. England: 124-130.
26. Movahed, Hanif Bayat. Simulation of protein folding. Retrieved from http://www.chem.utoronto.ca/~hbayat/simulation.pdf
27. Adcock, Stewart A. and J. Andrew McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev., 2006; 106(5): 1589–1615.
28. Theory of molecular dynamics simulation. Retrieved from http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html
29. Frenkel, D, Smit B. Understanding molecular simulation: from algorithms to applications. New York: Academic Press 1996.
30. Scheraga, Harold A., Mey Khalili, and Adam Liwo. Protein-folding dynamics: Overview of molecular simulation techniques. Annu. Rev. Phys. Chem., 2007; 58: 57–83.
31. Langevin dynamics simulation. Retrieved from http://cmm.cit.nih.gov/intro_simulation/node24.html
32. Grønbech-Jensen, Niels & Oded Farago. A simple and effective Verlet-type algorithm for simulating Langevin dynamics. Mol. Phys., 111: 983–991.
33. Neupane, Krishna, Dustin B. Ritchie, Hao Yu, Daniel A. N. Foster, Feng Wang, and Michael T. Woodside. Transition path times for nucleic acid folding determined from

energy-landscape analysis of single-molecule trajectories. Phys. Rev. Lett., 2012; 109: 068102.

34. Hanggi, P., P. Talkner, and M. Borkovec. Reaction-rate theory - 50 years after Kramers. Rev. Mod. Phys., 1990; 62: 251-341.
35. Kramers. Retrieved from https://www.uniulm.de/fileadmin/website_uni_ulm/nawi.inst.251/Lehre/TAPC/kramers.pdf
36. Studying protein folding with laser tweezers. Retrieved from http://research.physics.berkeley.edu/bustamante/pubs/Proceedings-Varenna.pdf
37. Ashkin, A, Dziedzic JM, Bjorkholm JE, Chu S. Observation of a single-beam gradient force optical trap for dielectric particles. Opt. Lett., 1986; 11 (5): 288–290.
38. Optical tweezers. Retrieved from Wikipedia.
39. Cecconi, C. *et al*. Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers, in preparation. Biophysical J., 2008 Jul; 37(6): 729–738.
40. Woodside, M.T., W.M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and Steven M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. Proc. Natl. Acad. Sci. U. S. A. 103: 6190–6195.
41. Woodside, M.T., John Lambert, Kevin S. D. Beach. Determining intra-chain diffusion coefficients for biopolymer dynamics from single-molecule force spectroscopy measurements. Biophysical J., 2014; 107(7): 1647–1653.
42. Greenleaf, W. J., M. T. Woodside, E. A. Abbondanzieri and S. M. Block. Passive all-optical force clamp for high-resolution laser trapping. Phys. Rev. Lett., 2005; 95: 208102.

APPENDIX

Codes:

The code was written in C++.

```
#include <cassert>
#include <cstdio>
using std::sprintf;
#include <cstdlib>
using std::atof;
using std::exit;
#include <iostream>
using std::cout;
using std::cerr;
using std::endl;
#include <iomanip>
using std::setw;
#include <fstream>
using std::ofstream;
#include <cmath>
using std::cos;
using std::sqrt;
using std::fmod;
#include <vector>
using std::vector;
#include "particle.hpp" // defining particle class

#include "mtrand.hpp" // random number generator
mtrand R(10);

// units: Force [pN], distance [nm], time [μs], mass [pg]
// [pN] = [pg·nm/μs^2]
// [pN·μs/nm] = [pg/μs]
```

```
// at room temp, 1 [kJ/mol] = 0.4034 kBT = 1.657974 [pN·nm]
// simulation constants

const double dt = 1.0E-6; // [μs] //simulation time step
const unsigned long int skip = 0.1/dt; //  time step for file output
const double kBT = 4.11; // [pN·nm] --thermal energy

// biomolecule characteristics

const double rho = 1.05E-9; //[pg/nm^3] -- density of polystyrene;
const double rad = 410; // [nm] -- bead radius -- small 300 large 410
const double pi = M_PI;

const double eta = 1.002E-3; // [pN·μs /nm^2] -- viscosity of water
//const double gamma2 = 6pi*eta*rad; // [pN·μs /nm]
//const double diffusion2 = kBT/gamma2; // [pN·nm] / [pN·μs /nm] = [nm^2/μs]
//const double tau2 = m_bead/gamma2; // [μs]

// Properties of the biomolecule
const double m_bio = rho*2*pi*2*700; // [pg] //mass of the biomolecule
const double gam1 = 6*pi*eta*sqrt(2*700); // [pN·μs /nm] = [pg/μs]
const double diffusion1 = kBT/gam1; //  [pN·nm] / [pN·μs /nm] = [nm^2/μs]
const double tau1 = m_bio/gam1; // [μs]

// Properties of the bead
const double m_bead = rho*4.0/3.0*pi*rad*rad*rad; // [pg] // mass of the bead
const double gam2 = 6*pi*eta*rad; // [pN·μs /nm] = [pg/μs]
const double diffusion2 = kBT/gam2; //  [pN·nm] / [pN·μs /nm] = [nm^2/μs]
const double tau2 = m_bead/gam2; // [μs]

// measurement characteristics
```

```
double initialClampingForceValue; // [pN]
double loadingRate; // [pN/μs]
const double x0=.1;
const double scale = 0.33;

double TrapForce(const particle &p)
{
	return initialClampingForceValue + loadingRate*p.t;
}

double LandscapeTest(const particle &p)
{
	const double F = initialClampingForceValue + loadingRate*p.t; // tension force applied on
the bead
	// U(x) = (x^2 - 10)^2
	//      = x^4 - 20 x^2 + 100
	const double y = p.x*p.x-10;
	return scale*y*y - F*p.x; // one-dimensional double well potential
}

double ForceTest(const particle &p)
{
	// U'(x) = 4 x^3  - 40 x - F
	//       = 4 x (x^2 - 10) - F
	const double F = initialClampingForceValue + loadingRate*p.t;
	return scale*(-4*p.x*(p.x*p.x-10)) + F; // poteintial force
}

double EffectiveCoupling(const particle &p1,const particle &p2)
{
	const double K=0.4;
```

```
        return K*(p2.x-p1.x);
}
void evolve(particle &p1,particle &p2, double dt, double (&force1)(const particle&),double (&force2)(const particle&),double (&force12)(const particle&, const particle&))
{
        const double xi1 = R.randNorm(0.0,sqrt(2.0*gam1*kBT*dt)); // Gaussian random number
        const double xi2 = R.randNorm(0.0,sqrt(2.0*gam2*kBT*dt)); // Gaussian random number
        const double b1 = 1.0/(1.0 + 0.5*gam1*dt/p1.mass);
        const double b2 = 1.0/(1.0 + 0.5*gam2*dt/p2.mass);
        const double f12=force12(p1,p2); // coupling force
        const double a1 = (force1(p1)+f12)/p1.mass;
        const double a2 = (force2(p2)-f12)/p2.mass;

        const double dx1 = b1*dt*(p1.v + 0.5*dt*a1 + 0.5*xi1/p1.mass);
        const double dx2 = b2*dt*(p2.v + 0.5*dt*a2 + 0.5*xi2/p2.mass);

        p1.x += dx1; // final positions of the biomolecule over successive times
        p2.x += dx2; // final positions of the bead over successive times

        const double ff12=force12(p1,p2);
        const double aa1 = (force1(p1)+ff12)/p1.mass;
        const double aa2 = (force2(p2)-ff12)/p2.mass;

        p1.v += 0.5*dt*(a1+aa1) - gam1/p1.mass*dx1 + xi1/p1.mass; //velocity of the biomolecule
        p2.v += 0.5*dt*(a2+aa2) - gam2/p2.mass*dx2 + xi2/p2.mass; // velocity of the bead molecule
        p1.t += dt; // time of the simulation
        p2.t += dt; // time of the simulation
}

void usage(void)
```

```
{
        cerr << "Usage: langevin [clamping force in pN] [loading rate [pN/s]]" << endl;
        exit(1);
}

int main(int argc, char *argv[])
{
        initialClampingForceValue = argc > 2 ? atof(argv[1]) : 0.0;
        cout<< "initialClampingForceValue ="<<"\t"<<initialClampingForceValue <<endl;
        loadingRate = (argc == 3 ? 1.0E-6*atof(argv[2]) : 0.0); // convert [pN/s] input to [pN/μs]
internal working units
        cout<<"loadingRate ="<<"\t"<<loadingRate<<endl;
        if (argc > 3) usage();

        particle bio(m_bio,0.0,0.0); // calling the particle class
        particle bead(m_bead,0.0,0.0); //  calling the particle class

        cout << "simulating with" << endl;
        cout << " biomolecule --  mass : " << bio.mass << " [pg]" << endl;
        cout << "            -- gamma : " << gam1 << " [pN·μs /nm] = [pg/μs]" << endl;
        cout << "            --   tau : " << tau1 << " [μs]" << endl;
        cout << "            --     D : " << diffusion1 << " [nm^2/μs] ";
        cout << endl;
        bio.t = 0.0;
        bead.t = 0.0;
        //ofstream fout("bio_bead_pos.dat"); // open a file
         ofstream fout1("bio_pos.dat"); // open a file named "bio_pos.dat"
        //fout.setf(std::ios::fixed);
        fout1.setf(std::ios::fixed);
        //fout.precision(16);
        fout1.precision(16);
```

```
        unsigned long int n0 = 1;
        int tempState=2;
        for (unsigned long int n = 0; n < 10000000000; ++n)
        {
                evolve(bio,bead,dt,ForceTest,TrapForce,EffectiveCoupling);

                if (n%skip == 0)
                {
                        //fout << bio << endl;
                        fout1<<bio.t<<"\t"<<bio.x<<"\t"<<bead.x<<endl;  //writing the
respective position of the biomolecule in a file

                }

        }
        //fout.close();
        fout1.close();
        return 0;
}
```

VITA

MD ADNAN KABIR

50-10 Broadway, Apt #3D • Woodside, NY 11377 • (662) 380-0309 • makabir@go.olemiss.edu

EDUCATION:

M.S., Department of Physics and Astronomy, University of Mississippi, December 2015

Concentration: Biophysics

Thesis: Biomolecular Folding Rates as Understood from Single-Reaction-Coordinate Langevin Dynamics and Kramers' Theory

M.S., Department of Physics, University of Dhaka, October 2010

Concentration: Solid State Physics

Thesis: The Study of Electrical and Spectrophotometric Properties of Organic Dyes

B.S., Department of Physics, University of Dhaka, January 2009

TEACHING EXPERIENCE:

Teaching Assistant, 2012-2014

University of Mississippi

Courses: Laboratory Physics I, General Physics I

SKILLS:

Extensive knowledge on Vacuum Evaporation Technique of preparing
Thin films and on FTIR spectroscopy
Operating system: Windows, Mac-OS, Linux
Software & languages: Microsoft Office, C++, C#, Python, Lab VIEW, Matlab
Excellent skills on photography & Adobe Photoshop Elements
Good communication skills-oral, writing, presentation & good team skills