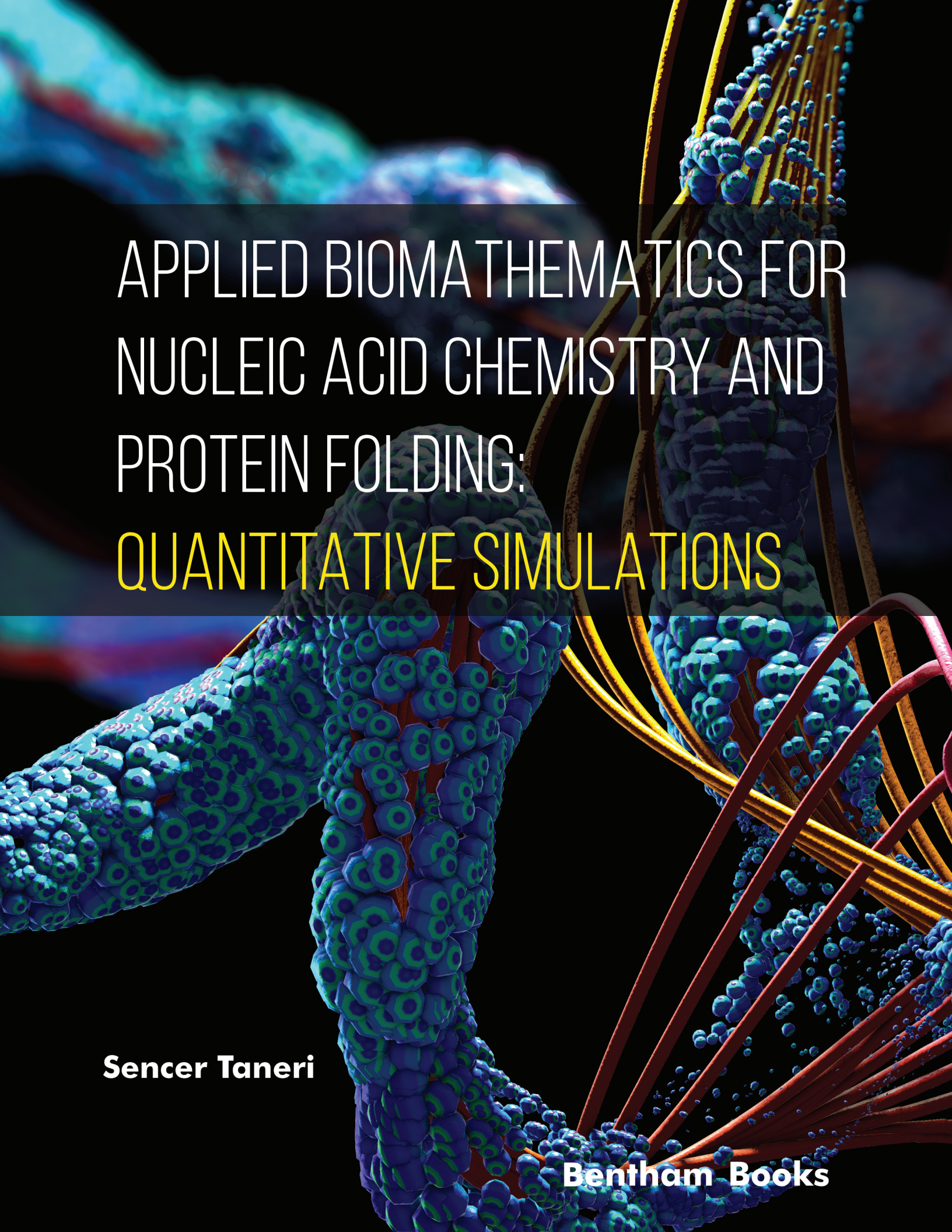

APPLIED BIOMATHEMATICS FOR NUCLEIC ACID CHEMISTRY AND PROTEIN FOLDING: QUANTITATIVE SIMULATIONS

Sencer Taneri

Bentham Books

Applied Biomathematics for Nucleic Acid Chemistry and Protein Folding: Quantitative Simulations

Authored By

Sencer Taneri

University of Southern California
Los Angeles,
USA

Applied Biomathematics for Nucleic Acid Chemistry and Protein Folding: Quantitative Simulations

Author: Sencer Taneri

ISBN (Online): 978-981-5179-96-5

ISBN (Print): 978-981-5179-97-2

ISBN (Paperback): 978-981-5179-98-9

First published in 2023.

need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.

3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Pte. Ltd.
80 Robinson Road #02-00
Singapore 068898
Singapore
Email: subscriptions@benthamscience.net

CONTENTS

FOREWORD

Rigorous description of biological phenomena through physics and mathematics is a daunting task both theoretically and computationally. One has to simplify usually very complex biological systems so as to be able to write equations that can be solved at least via numerical methods. There is always a high level of difficulty in physical understanding of a complicated biological process: how much the system is simplified, which degrees of freedom are chosen and which are discarded, and then how the numerical solution methods are implemented all contribute to the success or failure in understanding the particular biological system through physics and mathematics. Biological systems with their many degrees of freedom thus bring a highest possible challenge to a working physicist. This book contains state of the art theoretical and numerical techniques for analysis of several biological phenomena by a biophysical approach. In it, readers would find many invaluable insights into the biological processes, which can be utilized in diverse applications, including the spread of disease in a pandemic situation.

This book discusses three biological processes utilizing the techniques of biophysics: melting and vitrification of DNA molecules which can be described collectively as the chemistry of nucleic acids, theoretical discussion of percolation model and the description of folding of protein Crambin via two different methods. In chapter 1, Morse potential is used to describe the hydrogen bond interaction between nucleic acid base pairs in DNA molecule, and then a metropolis algorithm is used for the total potential energy as well as including the quantum fluctuation in terms of random displacement of the π electrons. This way the melting temperature of base pairs is calculated. Then in chapter 5, the same principles are used together with the inclusion of effect of longitudinal phonon vibrations to calculate the vitrification temperature of the base pairs. In chapter 2, an analytical analysis for a percolation simulation is presented: in a bit-string model of invading species in a random environment, the Hausdorff dimensions are calculated for the fractals and the conditions on invasion are analyzed analytically. Chapters 3 and 4 are reserved for analysis of the folding dynamics of the plant-seed protein Crambin in a liquid environment. A stochastic approach used to take into account the viscosity results in a 2D-Langevin equation, solution of which is established with a Molecular Dynamics simulation, accompanied by a delicate Monte Carlo technique. The final image of folded protein is found to be in very good geometric agreement with the real shape of the protein chain.

Finally, a much useful discussion of the well known 10-12 potential of hydrogen-hydrogen covalent bond is shared in the appendix.

This book consists of the research articles the author published during his postdoc studies at the Feza Gürsey Institute. Feza Gürsey Institute was a major center for theoretical physics and mathematics in Turkey. During the first decade of this century many top level research was conducted in this institute, and Dr. Taneri's work was up there with the very best. As in any good work, this book has many layers. It can be a valuable tool for the graduate students learning the subject, or it can be equally useful for established researchers.

Cemsinan Deliduman
Physics Department, Mimar Sinan University
İstanbul, Türkiye

PREFACE

The present book is intended to lay out research in Applied Biomathematics carried out by the author at the beginning of this millennium. It involves sample Monte Carlo simulations which have been widely used among scientists to understand stochastic molecular dynamics in biological matter of living systems. The subject material contains research essentially on three aspects of biomolecular structure and dynamics: (i) DNA Melting and DNA Vitrification, (ii) Evolutionary Genetics and Gene Mutation, and (iii) Protein Folding. The book consists of five chapters and an appendix. There have been some extensions to the collection of previous research articles published in International Journal of Modern Physics C and in Modern Physics Letters B which are World Scientific Publishing Journals. The collection of articles benefits from simple computer algorithms for hard mathematical physics problems of the past such as mesoscopic, fractals, percolation, metropolis algorithm and Langevin Dynamics. It can be useful for understanding the pandemic as the consequence of the spread of epidemic, and for understanding the recovery as the result of computer-aided drug discovery and cryopreservation.

Sencer Taneri
University of Southern California
Los Angeles
USA

ACKNOWLEDGEMENTS

The author would like to thank Dr. Ayşe Erzan from Istanbul Technical University ($IT\ddot{U}$), Dr.Cemal Yalabık from Bilkent University, and Dr.Cemsinan Deliduman from Mimar Sinan University (MSÜ) for fruitful discussions. The author would also like to thank Dr.Tonguç Rador from Boğaziçi University (BU) for the technical support. Most of the computer simulations have been carried out on the Intel (R) Celeron (R) CPU 2MHz node and Compaq nx8220 Centrino CPU 2MHz in Fortran 77 at home, and the author would also like to thank Dr. Nazım Ziya Perdahçı from Mimar Sinan University (MSÜ) for the technical support. The author is indebted to his friend Mr.Burak Gökmen for his critical reading of the manuscript.

Finally, I would like to thank my parents for the kind hosting of my writing sessions.

Comments, corrections and suggestions for improvement will be welcome.

CHAPTER 1

A Stochastic Mechanism for DNA Melting

Abstract: In Chapter 1, we have DNA as a kind of nucleic acid consisting of two strands which are made up of two Watson-Crick base pairs: adenine-thymine (AT) and guanine-cytosine (GC). There are three components of the total energy. These are the inharmonic stacking interaction, hydrogen bond interaction and kinetic energy. Morse potential is used to mimic the hydrogen bond interaction between bases on the opposite strands for the overlapping π electrons, when two neighboring bases move out of the stack. The AT pair has 2 hydrogen bonds and the GC pair has 3 of them. The π electrons obey $Bose-Einstein$ (BE) statistics, and the overlapping of them results in quantum fluctuation. It will be shown that this can be simplified into $<\Delta y(t)\Delta y(t)>=2D_q\Delta t$ type fluctuation between the base pairs. Thus, a metropolis algorithm can be developed for the total potential energy by superposing two potential energy terms as well as including the quantum fluctuation in terms of random displacement of the π electrons. So, one can calculate the melting temperature of base pairs.

Keywords: Biological matter, Living systems, Stochastic analysis methods.

INTRODUCTION

Deoxyribonucleic acid (DNA) melting is a widely addressed topic that has been intensively studied both experimentally, and theoretically/computationally. DNA melting has been lately studied from molecular dynamics point of view [1-6]. In these studies mesoscopic models for stretching are developed and Langevin equation is solved numerically for both of with and without solvent cases. Diffusion constant is facilitated during the solution of Langevin equation. Path integral formulation is developed to gain advantage in tedious numerical computations [5].

DNA melting transition has been recently studied by utilizing both Poland-Scheraga (PS) and Peyrard-Dauxois-Bishop (PDB) models of DNA, and theoretical techniques (mean field analysis) as well as numerical ones (Monte Carlo (MC) and Brownian Dynamics (BD) simulations) are elaborated for various types of comparison [6]. Fourth-order Runge-Kutta method is used in numerical solution of equations of motion for PS model. PDB and PS models are simulated using BD and Metropolis MC algorithm respectively with periodic boundary conditions for simplicity. Instead, Metropolis MC algorithm for PDB model of DNA will be utilized in this article to investigate how DNA behaves and melts. Instead of calculating the mean separation of strand we will calculate the separation of the single nucleotide in the strand.

Sencer Taneri

As it is well known, DNA has helical structure made up of two strands each of which consists of deoxyribonucleic acids in Fig. (**1**).

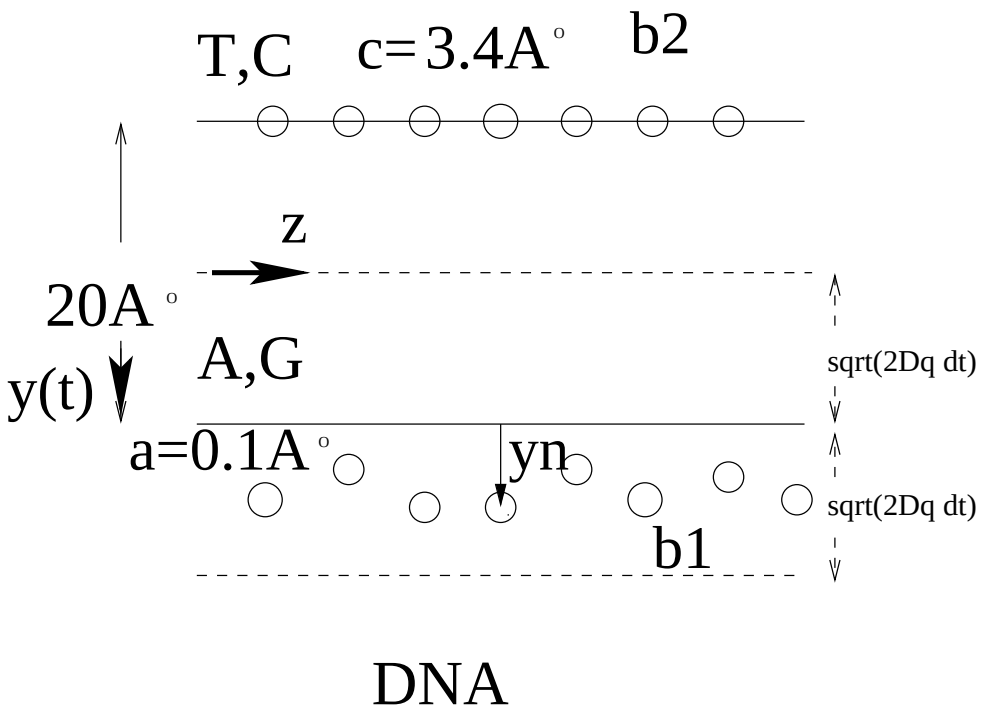

Fig. (1). DNA strand, opened helix as a planar ladder diagram of either uniform homopolymer AT or homopolymer GC. Here, the upper strand is assumed to do a similar motion during the motion of lower one (y_n is the vertical separation of the n^{th} monomer, the total stretch is $b = b_1 + b_2 = 2A^o$, a is the radius of monomer, c is the horizontal separation of monomers) [21].

These acids being on opposite strands are paired against one another as adenine-thymine (AT) and guanine-cytosine (GC) pairs. The AT pair has 2 hydrogen bonds and the GC pair has 3 of them [7]. There are two types of interactions going on between those base pairs which give rise to Peyrard-Bishop Model [8, 9]. The first one is the inharmonic stacking interaction between the neighboring base pairs, due to (UV) absorption process where the energy for each neighboring base pair can be expressed as, when one of the base pairs has a transverse stretching of amount y_n and the neighboring base pair has y_{n+1} [1, 10]. Here, n is the pair index, K is the backbone harmonic coupling constant, ρ and α are non-linear positive parameters independent of the type of the base pair. The harmonic portion denotes the effective potential for the optical branch, and the inharmonic coupling is due to the coupling

of the probability of bound states. This term, in general, represents several effects including the stacking, backbone flexibility, hydrophilic/hydrophobic interactions and any long range interactions [6]. The second one is the hydrogen bond energy which is, due to probability of the bound electron shared between the hydrogen atom pairs when one of the base pairs has a vertical separation of y_n. Here, D is the dissociation energy of the base pair and a is the inverse length defining the spatial scale of the potential. This term, in general, is an on-site Morse potential describing the net attraction between strands due to a combination of hydrogen bonding, solvent interactions, and the repulsion of negatively charged phosphate groups [6].

$$W_1(y_n, y_{n+1}) = \frac{K}{2}(y_n - y_{n+1})^2(1 + \rho exp(-\alpha(y_n + y_{n+1}))) \quad \textbf{(1)}$$

$$W_2(y_n) = D(1 - e^{-ay_n})^2 \quad \textbf{(2)}$$

Another relevant quantity to be calculated is the spatial fluctuation which is closely related to the diffusion constant utilized in solution of the Langevin equation mentioned in the beginning of the article [2-6]. That is, in BD, the system evolves *via* the Langevin equation,

$$m\ddot{y}_n = -\widehat{\nabla}V(y_n) - \gamma\dot{y}_n + \xi(t) \quad \textbf{(3)}$$

where the noise $\xi(t)$ is random Gaussian white noise with $< \xi(t) >= 0$ and $< \xi(t)\xi(t') >= 2D_c\delta(t - t') = 2\gamma k_B T\delta(t - t')$, and m mass of the n^{th} nucleotide, V is the potential $(V(y_n) = W_1(y_{n-1}, y_n) + W_1(y_n, y_{n+1}) + W_2(y_n))$, γ is the damping constant, D_c is the classical diffusion coefficient, k_B is Boltzmann constant and T is temperature. Note that, due to the definition of the diffusion coefficient, the spatial fluctuation obeys $< \Delta y_n \Delta y_n >= 2D_c\Delta t$.

On the other hand in Fig. (**2**), the spatial fluctuation of π electrons obeys,

$$I_{\Delta t \to \tau_c} =< \Delta y \Delta y > \sim \frac{24\hbar^4}{k^2 m_e^2 e^4} \frac{1}{exp(\frac{2\varepsilon_{2e}}{k_B T}) - 1} \quad \textbf{(4)}$$

where τ_c is the characteristic time, k is the dielectric constant of free space, m_e is the mass of the electron, and $2\varepsilon_{2e}$ is the total energy of the electron doublet which are boson, with respect to chemical potential level. Then, one ends up with quantum diffusion constant,

CHAPTER 2

A Theoretical Analysis of a Percolation Model

Abstract: In Chapter 2, we demonstrate an analytical analysis of a previously published research for a percolation simulation. In that research the effect of mutations on adaptability was investigated in a bit-string model of invading species in a random environment. However, analytical analysis was missing which will be the topic here. The Hausdorff dimensions are calculated for the fractals and the conditions on invasion are analyzed analytically by manipulation of partial differential equations. Thus, various conclusions may be reached without having to run long simulations.

Keywords: Boundary value problems, Fractals, Hausdorff dimension, Monte carlo methods, Ordinary and partial differential equations, Percolation.

INTRODUCTION

Much progress has been made since $1950's$ when percolation was a child and many open problems of the last decade have been solved. With such solutions we have seen the evolution of new techniques and questions and the consequent knowledge has shifted the ground under percolation. The mathematics of percolation is now fairly mature although there are mature questions which remain largely unanswered. Percolation technology has emerged as a cornerstone of the theory of disordered physical systems [1].

Percolation of a bit-string model [2] was developed by the author long ago in which evolution of invading species was simulated [3]. There, it was observed that mutations enabled the invasion of species even in environments with a high fitness threshold value. However, if the decay of species is introduced, the invasion sets still in a more simple manner. Huge amount of computer time is spent to get simple but as well as accurate results in those simulations. Lots of results can be attained by mathematical analysis which will be our task in this paper.

CALCULATION

In the bit-string model different mutants and the local ideal type are represented by random bit-strings consisting of 1's and 0's , which can be written as $\vec{v}_i(0,1,\dots.)$ (length description of genes in genome of the organism) and $\vec{h}_i(0,1,\dots.)$(length description of genes in genome of the ideal type) at site $\vec{\imath}$ respectively [2]. The

fitness function is defined as $f(d_{\vec{i}\vec{j}}) = 1 - d_{\vec{i}\vec{j}}$ where $\vec{j}$ is the site as the nearest neighbor and $d_{\vec{i}\vec{j}}$ is the Hamming distance introduced as,

$$d_{\vec{i}\vec{j}} = \frac{1}{l}\sum_{\alpha=1}^{l} \left| v_{\vec{i}\alpha} - h_{\vec{j}\alpha} \right| \quad \textbf{(1)}$$

with $l = 16$. The fitness function and the barrier height have the relation $f(d_{\vec{i}\vec{j}}) \geq r$ as there is invasion to the nearest neighbor. When this condition is not satisfied, it is set vacant. In the presence of mutations, there is a chance to bring the neighboring organism to the target site. Two variants as with decay (WD) and without decay (WOD) are simulated, when $g_2(r) = 1 - g_1(r)$ is the probability to decay, and $g_1(r)$ is the probability to invade the nearest neighbor. If we have the fitness threshold $r = \frac{m}{l}$ where m is the number of matching alleles and the number of mismatching alleles is n, then the probability of the case will be,

$$P(n) = \frac{1}{2^l}\frac{l!}{(l-n)!n!}. \quad \textbf{(2)}$$

So, the probability that the fitness function is larger than the fitness threshold r, namely $l - n \geq m$ is,

$$g_1(r) = \frac{1}{2^l}\sum_{0}^{l-m-1} \frac{l!}{(l-n)!n!} \quad for \quad \frac{m}{l} \leq r < \frac{m+1}{l}. \quad \textbf{(3)}$$

Various values for $g_1(r)$ are calculated in Table **1** for the following Monte Carlo Simulation (MCS).

RESULTS

The organisms are allowed to start from the center of 513×513 lattice by various different threshold values r. As for the inclusion of mutations, organisms occupying the sites are mutated with a probability μ per bit at each time step according to $mod(x + 1,2)$ where x is 0 or 1. The results are run 1000 times for different initial conditions of $\vec{v}_i$ for WD and WOD cases and averaged, see Tables **1**-**2**. In the absence of mutations there is percolation for $r \leq 0.5$ in our model. This is always the case (*e.g* no restrictions on r) when there is mutation for the WOD. Whatever mutation rate is used, there is no percolation when $r > 0.5625$ for the WD case, which is a more humble result in comparison to the WOD case. These

results in Table **1-2** can be obtained by analytical methods as will be discussed in the next section.

Table 1. Percolation times for the method without decay (*WOD*) case for various fitness threshold and mutation rates [2, 10].

r	$g_1(r)$	μ	$MCS\ Time$
0.0625	1.000	0.00	256 ± 0
0.125	0.998	0.00	256 ± 0
0.1875	0.989	0.00	256.0 ± 0.2
0.25	0.962	0.00	256.9 ± 0.7
0.3125	0.895	0.00	260.4 ± 1.1
0.375	0.773	0.00	270.9 ± 1.7
0.4375	0.598	0.00	298.2 ± 3.0
0.5	0.402	0.00	493.7 ± 63.7 $(g_{1eff}(r) = 0.5)$
0.5	0.402	0.01	324.3 ± 2.7
0.5625	0.227	0.01	408.8 ± 4.3
0.625	0.105	0.01	609.8 ± 8.8
0.6875	0.038	0.01	1152.0 ± 17.9

Table 2. Percolation times for the model with decay (*WD*) for various fitness treshold and mutation rates [2, 10].

r	μ	$MCS\ Time$
0.5	0.01	330.2 ± 3.1
0.5	0.5	326.4 ± 2.9
0.5625	0.01	465.7 ± 7.8

CHAPTER 3

A Monte Carlo Assisted Simulation of Stochastic Molecular Dynamics for Folding of the Protein Crambin in a Viscous Environment

Abstract: In Chapter 3, we investigate the folding dynamics of the plant-seed protein Crambin in a liquid environment, that usually happens to be water with some certain viscosity. To take into account the viscosity, necessitates a stochastic approach. This can be summarized by a 2D-Langevin equation, even though the simulation is still carried out in 3D. Solution of the Langevin equation will be the basic task in order to proceed with a Molecular Dynamics simulation, which will accompany a delicate Monte Carlo technique. The potential wells, used to engineer the energy space assuming the interaction of monomers constituting the protein-chain, are simply modeled by a combination of two parabola. This combination will approximate the real physical interactions, that are given by the well known Lennard-Jones potential. Contributions to the total potential from torsion, bending and distance dependent potentials are good to the fourth nearest neighbor. The final image is in very good geometric agreement with the real shape of the protein chain, which can be obtained from the protein data bank. The quantitative measure of this agreement is the similarity parameter with the native structure, which is found to be $0.91 < 1$ for the best sample. The folding time can be determined from Debye-relaxation process. We apply two regimes and calculate the folding time, corresponding to the elastic domain mode, which yields $5.2ps$ for the same sample.

Keywords: Computer simulation, Diffusion, Theory and modeling.

INTRODUCTION

We propose a molecular dynamic simulation (MDS) of the plant seed protein $Crambin$ which is made up of 46 amino acids. Experiments are done in liquid environments, usually in water with viscosity $0.89cp$ at $25C$. The protein is made up of monomers where the C_α's are centered. These monomers constitute the backbone of the protein chain in continuum space [1], not on a cubic lattice, which is the case mostly studied in literature [2-6]. In these studies, the monomers of the chain are considered to lie on the points forming the cubic lattice, corresponding to a self-avoiding walk, where the Metropolis Algorithm [7] is used in Monte Carlo Simulation (MCS).

This is how our MDS will progress: Certain physical interactions are present as the Van der Waals interaction and the repulsive interaction which can be summoned with the Lennard-Jones $6-12$ potential. There are also Hydrogen bonds that are represented by $10-12$ potential in the absence of solvent [8]. We also include

Sencer Taneri

elastic collisions with the surrounding water molecules, giving rise to a Gaussian white noise [9-12]. The viscosity of water causes the velocity proportional frictional force during the motion of the backbone, in response to thermal excitations. All these effects are to be summarized under one equation known as the Langevin equation that will demand a delicate MCS technique to solve. Since the backbone length is assumed to stay constant as a constraint, the angular variables will be presented in spherical coordinates and the Langevin equation will be two dimensional per monomer.

One can attain the time dependence of internal energy as the result of MDS and make a Debye relaxation fit to get the time constants of the folding process. The internal energy stems from the springs that are used to model the physical interactions. It is recognized as the elastic domain modes, which are known to take $100fs$ to several ps [13]. The final folded sequence consists of two α helices and a β sheet. All of the final images are not in good geometric agreement with the native shape of the protein chain. Nevertheless, we hope that our results regarding the dynamics, probed by means of our hybrid Monte Carlo and Molecular Dynamics approach, will be found interesting.

METHODS

The protein chain is modeled as a connected sequence of rods. A monomer is located at each edge of the rod, which can be oriented arbitrarily within the constraint of self-avoidance. The orientation of the polymer can be defined by the bending angles θ_i and the relative azimuthal angels ϕ_i as shown in Fig. (**1**). Then, we consider a potential, which depends on these angles and the distances between the rod centers up to the fourth neighbors. The kind of monomers in the neighborhood determines the parameters of this potential. The bending potential and the torsion potential depend on θ_i and ϕ_i and the arguments of the distance dependent potential are simply the distances between the centers of the neighboring rods that join the different types of monomers up to the fourth neighbors. There is also the constraint of self avoidance. Then, the total potential can be written as:

$$V_{chain} = \sum_{i=2}^{N} V_{\mathrm{B}}(t_{i-1}, t_i, \theta_i) + \sum_{i=3}^{N} V_{\mathrm{T}}(t_{i-2}, t_{i-1}, t_i, \phi_i) + \sum_{n=1}^{4} \sum_{ij}{}^{(d)} V_{\mathrm{N}}^{(d)}(t_i, t_j, d_{ij}) \quad \textbf{(1)}$$

where V_{B} is the bending, V_{T} is the torsion potential and V_{N} is the distance dependent potential.

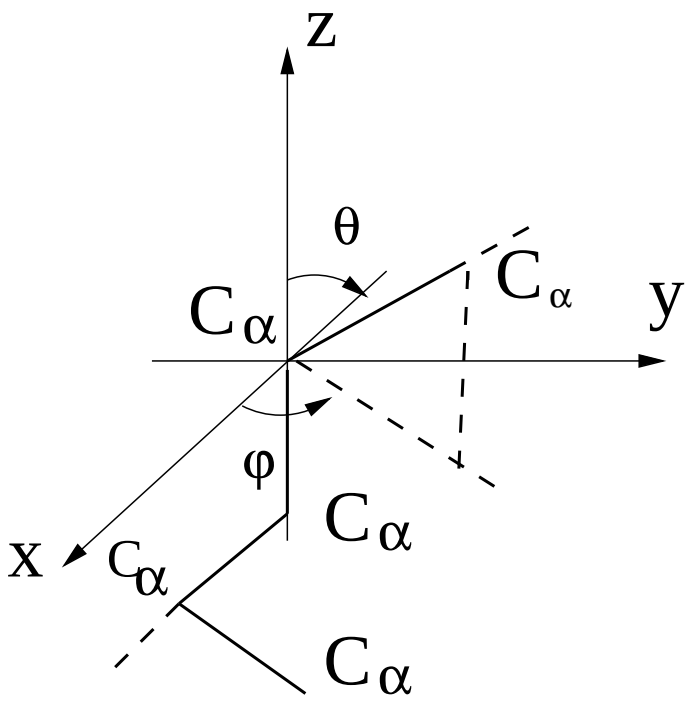

Fig. (1). Definitions of the relative angles used in our model. The bending angle θ_i is the angle between the $i-1$ st and i th monomer. The torsion angle ϕ_i is the azimuthal angle between the projection of i th monomer onto the x-y plane and $i-2$ nd monomer when the $i-1$ st monomer lies along the -z axis and the $i-2$ nd monomer lies in the x-z plane [29].

Here t_i is the type of the monomer, d is the neighbor type, d_{ij} is the distance between centers of the rods joining the monomer centers (C_α). The potentials are weighted by 0.8692 for distance dependent potentials and by 1.6085 for the angle dependent potentials so that the distance dependent spring constant is 10kcal/mol/A^{o2} and angular dependent spring constant is 30kcal/mol/rad^2 [14,15]. The superscript (d) on the summation sign shows that the summation runs on that type of neighbors. The first and the second term depicts the emphasized short range, and the third term depicts interactions between close by pairs of monomers depending long range effects arising from adjustment of angular variables down the chain. I should still point out that 44 bending,43 torsion(the first and the second terms), and 3 center of mass degrees of freedom, are good enough to span the configuration space, so that $3(46-1)$ degrees of freedom, minus $(46-1)$ constraints which yields 90 dimensional space again.

CHAPTER 4

Continuum Space Model for Folding of the Protein Crambin

Abstract: In Chapter 4, we have studied the chain length dependence of folding time for proteins by implementing a novel Monte Carlo (MC) method. The physical parameters in our model are derived from the statistics for bending and torsion angles and distances between the centers of the monomers up to the fourth neighborhood. By assigning potential wells to each of the physical parameters, we are able to use a modified Metropolis algorithm to efficiently trace the later conformations of the proteins as time evolves. Our prescription for microscopic dynamics for the protein "Crambin" results in an increase in folding times with increasing chain length. The folding times are determined *via* Debye relaxation process.

Keywords: Crambin, Debye, Folding, Metropolis.

INTRODUCTION

Protein folding problem continues to be interesting for the computationalists because of the unique conformation of the native state. One may take a sequence of amino acids and may expect to get the native conformation by a simulation. However, problems arise because, a minimization of the energy will lead to a final state which may be very different from the native state due to kinetics of the folding process [1]. Thus, the final state may very well be a trap in the folding pathway to the native conformation [2] and the protein is said to have made a collapse transition. This displays the importance of kinetics of folding. Because of this practical problem and due to fundamental interest, the relevant kinetics have been analyzed extensively and in particular on a cubic lattice [2-6]. In these studies, the monomers of the polymer are assumed to lie on the points forming the cubic lattice, corresponding to a self-avoiding walk. The degrees of freedom for the kinetics are then taken as the corner flip and the crankshaft moves [3-5]. The Metropolis algorithm [7] is used for the Monte Carlo MC simulation.

A natural extension of the MC procedure to continuum space involves small local modifications in the configurations of monomers in two dimensional space [8]. Because of the large amount of time necessary for the accumulation of small changes to result in a major change in the structure of the polymer, the computer time necessary to analyze such processes grows prohibitively large as the size of the polymer increases.

Sencer Taneri

We hope to illuminate the kinetics of folding in the early stages of folding process by analyzing the Debye relaxation process. We are introducing a different type of method, in which parameters modified in the MC procedure are relative, and therefore modifications propagate to all parts of the system rapidly. Although this approach would model the physical folding process only under limited circumstances due to the allowed deviation used in the statistics, it does produce a very efficient method and folding time does increase with the chain length by a power law.

In our present work we have developed a MC simulation based on a simple continuum space model. In this model, the monomers that make up the protein *Crambin* are used. *Crambin* is a plant seed protein that consists of 46 amino acids. Empirical potentials are derived from the statistics obtained from *Crambin* itself which is used to define the kinetics. These statistics include implicitly the information about several kinds of interaction such as the Van der Waals interaction and the hydrogen bonds [2]. Current methods often aim to fold *Crambin*, but most real proteins of interest have hundreds of amino acids.

The Model

The polymer is modeled as a joined sequence of rods of lengths corresponding to particular monomers, which can be oriented arbitrarily within the constraint of self avoidance. The orientational state of the polymer can be defined by the bending angles $\{\theta_i\}$ between the monomers and the relative azimuthal angles $\{\phi_i\}$ as shown in Fig. (**1**).

We then assume a potential which depends on these angles and the distances between the monomer centers up to the fourth neighbors. The parameters of this potential depend on the types of monomers in the neighborhood: The parameters of the bending potential and torsion potential which are functions of θ_i and ϕ_i depend on the types of the monomers i, $i-1$ and $i-2$ and the parameters of the distance dependent potential depend on the two types of monomers that it relates. There are no longer range interactions in the model apart from the constraint of self avoidance. The total potential describing the chain may then be written as:

$$V_{chain} = \sum_{i=2}^{N} V_{\mathrm{B}}(t_{i-1}, t_i, \theta_i) + \sum_{i=3}^{N} V_{\mathrm{T}}(t_{i-2}, t_{i-1}, t_i, \phi_i) + \sum_{n=1}^{4} \sum_{ij}{}^{(n)} V_{\mathrm{N}}^{(n)}(t_i, t_j, d_{ij}) \qquad \textbf{(1)}$$

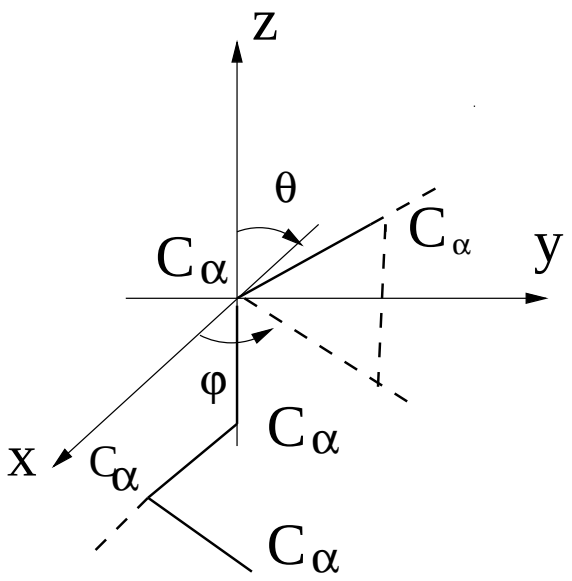

Fig. (1). Definitions of the relative angles used in our model. The bending angle θ_i is the angle between the $i-1$ st and i th monomer. The torsion angle ϕ_i is the azimuthal angle between the projection of i th monomer onto the x-y plane and $i-2$ nd monomer when the $i-1$ st monomer lies along the -z axis and the $i-2$ nd monomer lies in the x-z plane [15].

where V_{B} is the bending, V_{T} is the torsion potential and V_{N} is the distance dependent potential. Here t_i is the type of the monomer, n is the neighbor type (1 for the nearest neighbor, (2) for the second nearest neighbor, *etc.*), d_{ij} is the distance between centers of the rods joining the monomers (C_α) $i, i+1$ and $j-1, j$. The potentials for $n=1$ and $n=2$ are weighted by 0.5 and the rest by unity. The weighting numbers are determined in an empirical way. Superscript (n) on the summation sign indicates that the summation runs on that type of neighbors.

This potential covers the approximation to the hydrogen bond and the collapse energy terms of the total Hamiltonian around the equilibrium point of the total energy surface which is a good approximation for the α- helices [9,10]. Bending or torsion terms mostly represent the hydrogen bonds in the absence of solvent ($10-12$ potential, see Appendix) while distance dependent potentials mostly represent the collapse energy ($6-12$ Lennard-Jones potential, see Appendix). That the distance dependent potential assumes interactions up to the fourth neighbor is justified by the fact that the collapse energy including the hydrophobic interactions

CHAPTER 5

A Stochastic Mechanism for DNA Vitrification

Abstract: In Chapter 5, DNA is a kind of nucleic acid consisting of two strands which are made up of two Watson-Crick base pairs: adenine-thymine (AT) and guanine-cytosine (GC). Vitrification (from Latin *vitreum, "glass"*) on the other hand is the transformation of a substance into a glass. DNA vitrification is achieved by rapidly cooling DNA in a liquid state through the glass transition. The quantum fluctuation in terms of random displacement and specific heat capacity of the π electrons in hydrogen bonds was studied earlier to calculate the DNA melting temperature. Same principles along with the inclusion of longitudinal phonon vibrations will be used here in order to calculate the vitrification temperature (glass transition temperature) of base pairs. This has an important application in cryonics and cryopreservation.

Keywords: Biological physics, Statistical physics.

INTRODUCTION

As it is well known, Deoxyribonucleic acid (DNA) is an organic nanomaterial with a helical structure of twin strands of nucleic acids, see Fig. (**1**) [1]. DNA melting has been lately studied both experimentally and computationally [2-13]. A stochastic mechanism for DNA melting has also been studied by the author before [14]. There, inharmonic stacking interaction, hydrogen bond interaction and kinetic energy components of the total energy were used to implement a metropolis algorithm. Morse potential was used to mimic the hydrogen bond interaction between the basis on the opposite strands for the overlapping π electrons. The AT pair had 2 bonds and the GC pair had 3 of them. The π electrons obeyed $Bose-Einstein$ (BE) statistics, and the overlapping of them resulted in quantum fluctuation. It was shown that this could be simplified into $<\Delta y(t)\Delta y(t)>=2D_q\Delta t$ type fluctuation between the base pairs. Thus, a metropolis algorithm was developed for the total potential energy by superposing two potential energy terms as well as including the quantum fluctuation in terms of random displacement of the π electrons, see Fig. (**2**). Here, the quantum diffusion constant D_q, is given by,

$$D_q \sim \frac{12\hbar^4}{k^2 m_e^2 e^4 \tau_c} \frac{1}{exp(\frac{2\varepsilon_{2e}}{k_B T})-1}, \quad \textbf{(1)}$$

and τ_c is the characteristic time. The previous parameters along with potential parameters can be found in literature.

Sencer Taneri

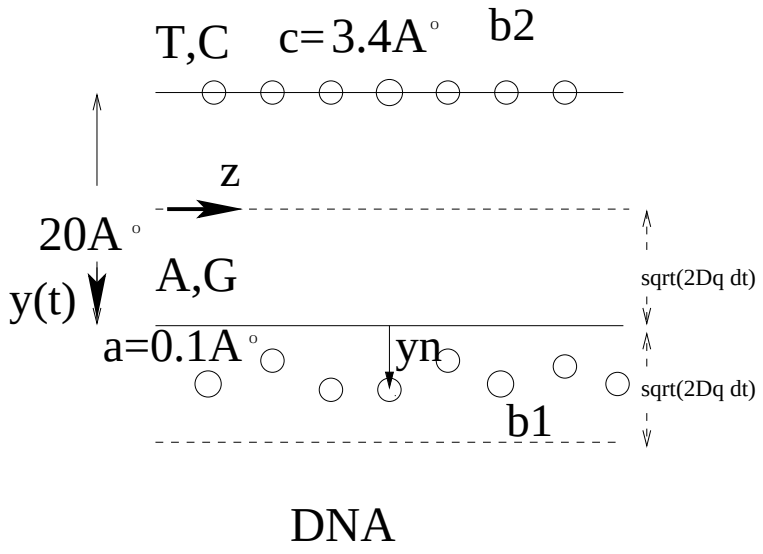

Fig. (1). DNA strand, opened helix as a planar ladder diagram of either uniform homopolymer AT or homopolymer GC. Here, the upper strand is assumed to do a similar motion during the motion of lower one (y_n is the vertical separation of the n^{th} monomer, the total stretch is $b = b_1 + b_2 = 2A^o$, a is the radius of monomer, c is the horizontal separation of monomers) [14].

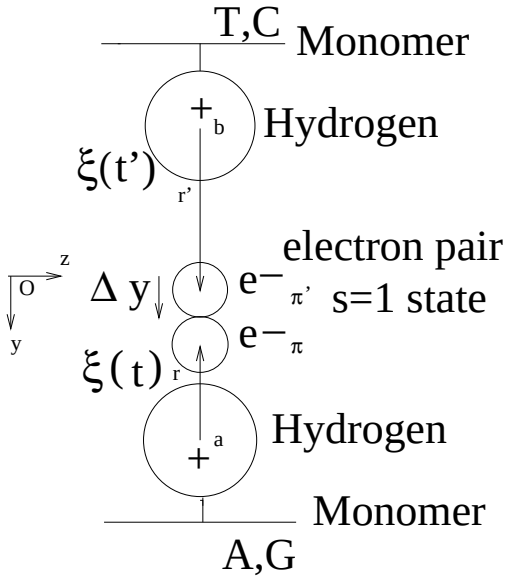

Fig. (2). $\Delta y = \pm\sqrt{2D_q \Delta t}$ is the quantum fluctuation in the mesoscopic scale, ξ is the random force exerted to hydrogen atom and thus to monomers by the electron pair of hydrogen bond. r and r' are coordinate labels of electrons π and π' with respect to protons a and b [14].

As DNA melts, we confront second order Arrhenius equation. The height of the energy barrier ΔE_m that DNA base pairs have to fall down to get melted can be regarded as approximately same as the height of the energy barrier ΔE_g that DNA base pairs aiming the liquid state have to climb up for glass transition, see Fig. (**3**).

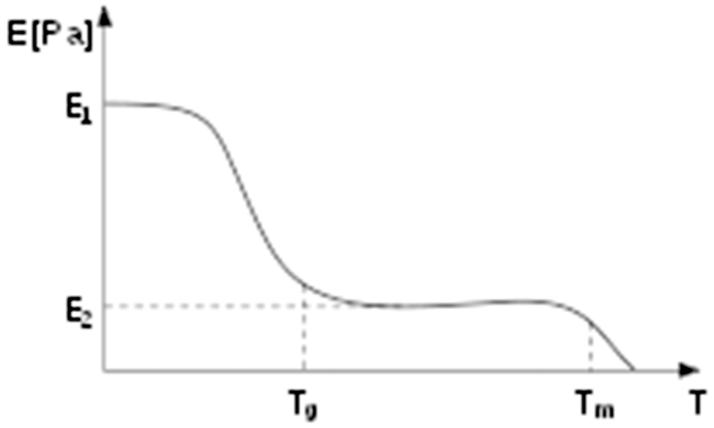

Fig. (3). Stiffness *versus* temperature curve. Stiffness $E\alpha N_p$ is proportional to population of paired DNA molecules N_p. Note the difference between the glass transition temperature and melting temperature. Also, the energy difference in glass transition and in melting processes are assumed to be approximately equal ($\Delta E_g \sim \Delta E_m$) [16].

A glass has the random structure of the liquid from which it is derived by cooling below the freezing point, without crystallization [15]. So, vitrification temperature (glass transition temperature) of the base pairs can be calculated by superposing the specific heat capacity of electrons in hydrogen bonds and longitudinal vibrations of DNA strands (acoustic phonons). The rationale here is that the thermal motion in liquids can be decomposed into elementary longitudinal vibrations (or acoustic phonons) while transverse vibrations (or shear waves) were originally described only in elastic solids exhibiting the highly ordered crystalline state of matter [16]. At the end of the day, this crude scheme is still good enough to offer us kilo base-pair limitation for the model.

THEORY

The heat capacity or the thermal kinetic energy of the electrons in the Fermi gas of the hydrogen bonds is given by,

CHAPTER 6

A Theoretical Investigation on 10-12 Potential of Hydrogen-Hydrogen Covalent Bond

Abstract: In Appendix, we have an analytical investigation of the well-known 10-12 potential of hydrogen-hydrogen covalent bond. In this research, we will make an elaboration of the well-known 6-12 Lennard-Jones potential in case of this type of bond. Though the results are illustrated in many text books and literature, an analytical analysis of these potentials is missing almost everywhere. The power laws are valid for small radial distances, which are calculated to some extent. The internuclear separation as well as the binding energy of the hydrogen molecule are evaluated with success.

Keywords: Electronic structure and bonding characteristics, Materials, Molecular hydrogen and isotopes, Solid state.

INTRODUCTION

The fact that H_2 is explained quantum mechanically by the behavior of the electronic eigenfunction is described by the charge distribution of the system, as two hydrogen atoms approach one another. The resulting charge distribution does lead to electrostatic attraction, but it is a charge distribution that can be interpreted as a sharing of electrons by both atoms. The binding is called covalent, [1]. One can write the total potential energy of two hydrogen atoms at separation R as the $10-12$ potential for the hydrogen-hydrogen covalent bond that differs from the non-covalent hydrogen bond which is important in protein interactions, See Refs. [2-6]. An example of such non-covalent hydrogen bonds is illustrated in literature, [7].

Likewise, one may also consider two identical inert gas atoms at separation R large in comparison with the radii of the atoms. The atoms induce oscillating dipole moments in each other, and the induced moments cause an attractive interaction between the atoms. The attractive interaction varies as the minus sixth power of the separation of the two oscillators which is called the van der Waals interaction, known also as the London interaction. An analytical proof for this interaction which will be our guide for evaluation of $10-12$ potentials is given in some solid state books, [8].

As two atoms are brought together, their charge distributions gradually overlap, thereby changing the electrostatic energy of the system. At sufficiently close separations the overlap energy is repulsive, mostly because of the Pauli exclusion

Sencer Taneri

principle. The elementary statement of the principle is that two electrons can not have all their quantum numbers equal, [8]. This repulsive interaction varies as the minus twelfth power of the separation of the two electrons. So, the attractive and the repulsive potentials can be summarized in one equation as the Lennard-Jones potential, which is valid for inert gases. Here, ε and σ are some parameters. A detailed discussion about such interatomic potential models can be found in literature, [9].

$$V(R) = 4\varepsilon\left(\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6}\right) \quad \textbf{(1)}$$

THE THEORY

One can refer to Fig. (**1**) to calculate the "10" attractive interaction part of $10-12$ potential by taking into account the dipole-dipole interaction. The total potential due to this dipole-dipole interaction may be written as,

$$V_{dd} = \frac{1}{2}Cx_1^2 + \frac{1}{2}Cx_2^2 - kq^2\left(\frac{1}{R-\frac{x_1+x_2}{2}} - \frac{1}{R+\frac{x_2-x_1}{2}} - \frac{1}{R-\frac{x_2-x_1}{2}} + \frac{1}{R+\frac{x_1+x_2}{2}}\right) \quad \textbf{(2)}$$

where C is the force constant for the hydrogen atom, x_1 and x_2 are separation of dipoles, k is the electromagnetic force constant, $q = -(-e)$ is minus the electronic charge and R is the inter-dipole coordinate and by making Taylor series expansions for the radial part when $x_1, x_2 \ll R$, one finds leading to fourth order in x,

$$V_{dd} = \frac{1}{2}Cx_1^2 + \frac{1}{2}Cx_2^2 - \frac{2kq^2x_1x_2}{R^3} - \frac{kq^2x_1x_2\left(x_1^2+x_2^2\right)}{R^5}. \quad \textbf{(3)}$$

Here, the force constant can be evaluated by writing the effective interatomic potentials for each dipole as,

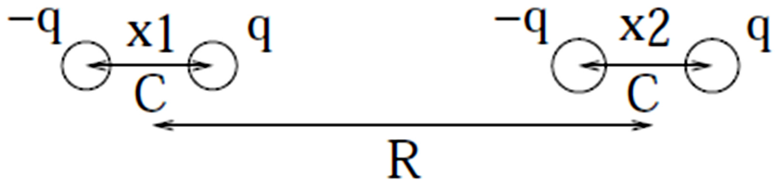

Fig. (1). Coordinates of dipole oscillators modeling the electromagnetic interaction between the charge clouds of hydrogen molecule [12].

$$V_0 = \frac{\hbar^2 J(J+1)}{2m_e x^2} - \frac{kq^2}{x} \quad \textbf{(4)}$$

where $\hbar$ is the Planck constant, J is the total angular momentum of the electron, m_e is the electron mass for the rotational kinetic energy, [10]. This effective potential may also be written as, around equilibrium value of x_{eq} and by equating the first derivative of the potential to zero the equilibrium value yields.

$$V_0 = V(x_{eq}) + \frac{dV(x)}{dx}|_{x_{eq}}(x - x_{eq}) + \frac{d^2V(x)}{2dx^2}|_{x_{eq}}(x - x_{eq})^2. \quad \textbf{(5)}$$

$$x_{eq} = \frac{\hbar^2 J(J+1)}{kq^2 m_e}. \quad \textbf{(6)}$$

Plugging this into the second derivative of the potential,

$$C = \frac{k^4 q^8 m_e^3}{\hbar^6 J^3 (J+1)^3}. \quad \textbf{(7)}$$

On the other hand, one should remember that $electron - electron$ wavefunctions overlap to give the resultant wavefunction. Fig. (**2**) depicts this generation, [10, 11]. So, one can superpose $electron - electron$ pairs in place of the spatial coordinates where the probability distribution peaks. Thus, one can evaluate the electromagnetic interaction of electron pairs with nuclei to model the correlation because of overlapping of two electron wavefunctions. Fig. (**3**) shows how this is handled. Now, one can write for the pair-pair interaction,

$$V_{pp} = \frac{1}{2}C' x_1'^2 + \frac{1}{2}C x_2'^2 - kq^2\left(\frac{2}{R' - \frac{x'_1 + x'_2}{2}} - \frac{4}{R' + \frac{x'_2 - x'_1}{2}} - \frac{1}{R' - \frac{x'_2 - x'_1}{2}} + \frac{2}{R' + \frac{x'_1 + x'_2}{2}}\right) \quad \textbf{(8)}$$

where C' is the force constant for $electron\ pair - nucleus$, x'_1 and x'_2 are separation of dipoles and R' is the inter-atomic coordinate and by making Taylor series expansions for the radial part when $x'_1, x'_2 \ll R'$ one finds leading to second order in x',

$$V_{pp} = \frac{1}{2}C' x_1'^2 + \frac{1}{2}C' x_2'^2 + \frac{kq^2}{R'} - \frac{3kq^2(x'_2 - x'_1)}{2R'^2} - \frac{4kq^2 x'_1 x'_2}{R'^3} + \frac{kq^2(x'_2 - x'_1)^2}{4R'^3}. \quad \textbf{(9)}$$

SUBJECT INDEX

F

G

H

I

K

L

M

N

O

P

Q

S

T

V

W

X

Sencer Taneri

Prof. Sencer Taneri, Ph.D. is an associate professor in physics and a private researcher in mathematics in Kartal, Istanbul, Turkey. He did Ph.D. in muscle contraction in physics from Bilkent University in Ankara, Turkey in 2002. He studied the topic ''Deformation and Finite Size Effects in Cooperative Molecular Motors'' and completed his thesis in July 2002. He also studied protein folding , evolutionary genetics and random walk problems after he entered Feza Gürsey Institute to pursue his career with a postdoctoral position in Tübitak-Boğaziçi in Istanbul, Turkey in 2003. Then, he continued his studies on DNA melting , DNA vitrification and quarks as a private researcher since 2008. He has many research papers published in international journals of reputation in the International Science Citation Index. Presently, he is publishing an applied biomathematics book which is about nucleic acid chemistry and protein folding. It includes quantitative simulations programmed in Fortran 77 Language.