


Fall 2004

# Computational approaches to the design and analysis of stability of polypeptide multilayer thin films

Bin Zheng

Follow this and additional works at: <https://digitalcommons.latech.edu/dissertations>

 Part of the [Computer Sciences Commons](#), [Materials Science and Engineering Commons](#), and the [Mathematics Commons](#)

---

COMPUTATIONAL APPROACHES TO THE DESIGN AND  
ANALYSIS OF STABILITY OF POLYPEPTIDE  
MULTILAYER THIN FILMS

by

Bin Zheng, B.S.

A Dissertation Presented in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

COLLEGE OF ENGINEERING AND SCIENCE  
LOUISIANA TECH UNIVERSITY

November 2004

UMI Number: 3148961

### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

**UMI**<sup>®</sup>

---

UMI Microform 3148961

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

LOUISIANA TECH UNIVERSITY

THE GRADUATE SCHOOL

09/14/2004

Date

We hereby recommend that the dissertation prepared under our supervision  
by Bin Zheng

entitled Computational Approaches to the Design and Analysis of Stability of Polypeptides Multi-  
layer Thin Films

be accepted in partial fulfillment of the requirements for the Degree of  
Ph.D in Computational Analysis and Modeling

Donald J. Hynice  
Supervisor of Dissertation Research  
Richard F. Grechis  
Head of Department  
Mathematics Dept.  
Department

Recommendation concurred in:

Raja Nataraj  
Wishnu Dri  
[Signature]

Advisory Committee

Approved: [Signature]  
Director of Graduate Studies

[Signature]  
Dean of the College

Approved: [Signature]  
Dean of the Graduate School

## ABSTRACT

The focus of this research is the development of computational approaches to understanding the physical basis of layer-by-layer assembly (LBL), a key methodology of nanomanufacturing. The results provided detailed information on structure which cannot be obtained directly by experiments.

The model systems chosen for study are polypeptide chains. Reasons for this are that polypeptides are no less polyelectrolytes than the more usual polyions, and one can control the primary structure of a polypeptide on a residue-by-residue basis using modern synthetic methods. Moreover, as peptides constitute one of the four major classes of biological macromolecules, research in this direction is expected to advance development of bionanotechnology. Polypeptide thin films are a type of new material, and there is great potential for applications in biocompatible implants, drug delivery, and other areas.

A key consideration in polypeptide design for LBL is charge properties as a function of pH. This work presents a computational approach to identify structural motifs in amino acid sequence data and to minimize the immune response to polypeptides based on the structural motifs and demonstrate by experiments.

This work also presents innovative molecular dynamics (MD) work on LBL. All-atom models have been used to investigate polypeptide LBL at the sub-molecular level. The peptide structures studied – homopolymers of lysine and of glutamic acid, and

designed cysteine-containing peptides – correspond to ones for which experimental data have been obtained in the Haynie research laboratory. Simulations were carried out to study structural and dynamical properties of peptide models having some combination of parallel and anti-parallel  $\beta$  sheets, as such structures are known to be formed by the indicated peptides in LBL films.

The MD work suggests that hydrophobic interactions too play an important role in polypeptide LBL. Moreover, hydrogen bonding appears to be a consequence of polypeptide LBL instead of a major driving force for stabilizing secondary structures in polypeptide multilayer thin films. Results of simulations of 6-residue and 8-residue peptides further suggest that if the shorter peptides can form a stable superstructure in the vicinity of 350 K, the most likely conformation will be anti-parallel  $\beta$  strands within a layer and parallel  $\beta$  strands between layers.

## APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Dissertation. It is understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Dissertation. Further, any portions of the Dissertation used in books, papers, and other works must be appropriately referenced to this Dissertation.

Finally, the author of this Dissertation reserves the right to publish freely, in the literature, at any time, any or all portions of this Dissertation.

Author Bin Zhang  
Date 11/11/2004

## TABLE OF CONTENTS

<b>LIST OF TABLES.....</b>	<b>viii</b>
<b>LIST OF FIGURES.....</b>	<b>ix</b>
<b>ACKNOWLEDGMENTS.....</b>	<b>xii</b>

### **CHAPTER 1:**

<b>INTRODUCTION.....</b>	<b>1</b>
1. 1 Motivation.....	1
1. 2 Background.....	2
1.2.1 A Brief Review of Layer-by-Layer Assembly.....	2
1.2.2 LBL.....	3
1.2.3 Peptide Design.....	5
1.2.4 Computational Approach.....	7
1.2.4.1 Computer Simulation for Polyelectrolyte Systems.....	7
1.2.4.2 Simulating Internal Forces.....	9
1.2.4.3 Theoretical Description of a Polymer Chain.....	10
1.3 Organization of this Dissertation.....	11

### **CHAPTER 2:**

<b>GOVERNING EQUATIONS OF MD SIMULATION.....</b>	<b>13</b>
2.1 Introduction.....	13
2.2 Integration Methods.....	14
2.3 Forces Fields.....	15
2.4 Solvent Models.....	18
2.4.1 Explicit Solvent Models.....	19
2.4.2 Implicit Solvent Models.....	19

### **CHAPTER 3:**

<b>PEPTIDE MOTIF DATABASE.....</b>	<b>22</b>
3.1 Introduction.....	22
3.2 General Properties of Amino Acids.....	24
3.3 Protein/Peptide Structures.....	25
3.4 Methods.....	28
3.5 Results and Discussions.....	30
3.6 Database Configuration.....	33



3.7 Experimental Results Involving Designed Polypeptides.....	33
<b>CHAPTER 4:</b>	
<b>MD SIMULATIONS OF Y(K)<sub>5</sub> and Y(E)<sub>5</sub> PEPTIDES.....</b>	<b>36</b>
4.1 Introduction.....	36
4.2 Methods.....	37
4.2.1 Simulation Models.....	37
4.2.2 Simulation Details .....	39
4.2.3 Geometrical Parameters .....	40
4.2.4 Simulations of Hexamers in Neutral States.....	40
4.3 Results .....	40
4.3.1 Dimers and Trimers.....	40
4.3.2 Tetramers.....	46
4.3.3 Hexamers.....	46
4.3.4 Comparisons of Hexamers in Charged and Neutral States.....	47
4.4 Discussions.....	52
<b>CHAPTER 5:</b>	
<b>MD SIMULATIONS OF CYSTEINE-CONTAINING PEPTIDES.....</b>	<b>58</b>
5.1 Introduction.....	58
5.2 Methods.....	60
5.2.1 Simulation Models.....	60
5.2.2 Simulation Details.....	61
5.3 Results .....	62
5.3.1 Dimers and Trimers.....	62
5.3.2 Tetramers.....	65
5.3.3 Hexamers.....	67
5.4 Discussions.....	70
<b>CHAPTER 6:</b>	
<b>SUMMARY AND PROSPECTS.....</b>	<b>74</b>
6.1 Summary and Contributions.....	74
6.2 Future Study and Prospects.....	76
<b>APPENDIX:</b>	
<b>SOURCE CODES.....</b>	<b>77</b>
<b>REFERENCES.....</b>	<b>89</b>

## LIST OF TABLES

Table	Page
2.1 Geometrical and energetic properties of various water models .....	21
3.1 Properties of five charged amino acid residues.....	26
3.2 Motif length vs. the number of amino acids .....	28
3.3 Amino acid frequency and propensity to form secondary structure .....	32
3.4 Motifs information stored in the database.....	34
3.5 Four amino acid sequences for experimental work.....	34

## LIST OF FIGURES

Figure	Page
1.1 Schema of oppositely-charged polyelectrolytes in LBL .....	4
1.2 Schema of a polyelectrolyte chain. (a) Continuous chain. (b) Discrete chain.....	11
2.1 Lennard-Jones potential .....	16
2.2 Internal interactions (a) Bond-angle. (b) Torsion angle.....	17
2.3 Explicit water models. (a) Structures of TIP3P, SPC, PPC. (b) Structure of TIP4P.....	20
3.1 Structure of an amino acid. (a)The zwitterionic form, predominant at neutral pH is shown. (b) Structure of a dipeptide. A peptide bond joins the two amino acids. The dihedral bond angles $\phi$ and $\psi$ are shown. ‘R’ represents the side chain.....	24
3.2 Flowchart of the identification sequence motif process .....	27
3.3 The number of non-redundant amino acids as function of sequence length.....	30
3.4 20 amino acids secondary propensity values.....	35
3.5 Adsorption of SN, SP, LN, and LP.....	35
4.1 Schema diagram of Dimers, Trimers, Tetramers and Hexamers. Black and gray colors represented negatively-charged Tyr(Glu) <sub>5</sub> and positively-charged Tyr(Lys) <sub>5</sub> at neutral pH, respectively.....	38
4.2 The geometrical parameters used to evaluate the structural changes.....	41
4.3 The fraction of the number of hydrogen bonds of dimers throughout the simulation.....	42
4.4 The average distance of dimers as a function of time.....	42
4.5 RMSD of trimers relative to the starting structure .....	43

4.6 The fraction of the number of hydrogen bonds of trimers throughout the simulation.....	43
4.7 The average distance of trimers as a function of time.....	43
4.8 RMSD of trimers relative to the starting structure .....	44
4.9 Ramachandran plot for residues 2 to 5 of Dimer 1. The background picture was from <a href="http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html">www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html</a> .....	45
4.10 Ramachandran plot for residues 2 to 5 of Dimer 2. The background Ramachandran plot was from <a href="http://ww.cryst.bbk.ac.uk/PPS2/course/section3/rama.html">ww.cryst.bbk.ac.uk/PPS2/course/section3/rama.html</a> ....	45
4.11 RMSD values of tetramers.....	47
4.12 The fraction of hydrogen bonds of tetramers as a function of time.....	47
4.13 The average distance within the sheets of tetramers as a function of time.....	48
4.14 The average distances between sheets of tetramers as a function of time.....	48
4.15 The fraction of hydrogen bonds of hexamers as a function of time....	48
4.16 RMSD values of hexamers as a function of time.....	49
4.17 Average distances within the sheets of hexamers as a function of time.....	49
4.18 Average distances between $\beta$ sheets of hexamers a a function of time.....	49
4.19 Comparison of remaining hydrogen bonds of hexamers in different charge status. Black and pink stands for the non-charged and charged peptides, respectively.....	50
4.20 Comparison of intra-distances of hexamers as a function of time. Black stands for the non-charged peptides and pink stands for charged peptides.....	51
4.21 Comparison of inter-distances of hexamers as a function of time. Black stands for the non-charged peptides and pink stands for charged peptides.....	52
5.1 CD of multilayers of Polypeptides neutral pH. (a) PLL-PLGA (b) Designed positive and negative polypeptides.....	60
5.2 Atomic simulation models. Black and red colors represented negatively- and positively-charged peptide sequences at neutral pH, respectively.....	61
5.3 RMSD values of dimers .....	63

5.4 The proportion of remaining hydrogen bonds of Dimer 1 and Dimer 2 as a function of time.....	63
5.5 Average distances of Dimer 1 and Dimer 2 as a function of time.....	64
5.6 RMSD of trimers relative to the starting structure.....	64
5.7 The fraction of remaining number of hydrogen bonds of trimers throughout the simulation.....	64
5.8. The average distances of trimers.....	65
5.9 RMSD values of tetramers as a function of time.....	66
5.10 Percentage of the remaining number of hydrogen bonds of tetramers throughout the simulation.....	66
5.11 Intra-strands distances of tetramers.....	67
5.12 Inter-sheets distances of tetramers.....	67
5.13 RMSD values of hexamers as referencing to the initial structures.....	68
5.14 The fraction of hydrogen bonds of hexamers as a function of time.....	69
5.15 Intra-strands distances of hexamers as a function of time.....	69
5.16 Inter-sheets distances of hexamers as a function of time.....	69
5.17 Comparison of the remaining number of hydrogen bonds of tetramers. KE stands for the peptide group lysine and glutamic acid. CC stands for cysteine-containing peptides. The length of each peptide is 8.....	73
5.18 Comparison of the remaining number of hydrogen bonds of hexamers. KE stands for the peptide group lysine and glutamic acid. CC stands for cysteine-containing peptides. The length of each peptide is 8.....	73

## ACKNOWLEDGEMENTS

My sincere gratitude and appreciation goes first to my advisor, Professor Donald T. Haynie, for providing me the unique opportunity to work in the research area of bioinformatics, and for his guidance, encouragement, patience, and support. As a scientist, he amazes me with his comprehensive scientific knowledge and unique insights on scientific problems. As an advisor, he encouraged and supported me in scientific decisions and taught me how to be a good scientist. I appreciated all his time and advice.

I would also like to thank Dr. Raja Nassar, Dr. Weizhong Dai, and Dr. Sumect Dua, not only for their participation as Advisory Committee members but also for the precious knowledge of statistics, mathematics, computer science, and bioinformatics I obtained from them.

I also want to thank all of the professors who have taught me in the past few years. I learned a lot from them. Although I do not mention their names here, I keep them in mind.

I am very grateful for all of the support I obtained from all of them. I know it is only because of such support that I could stand higher and see further.

Finally, I thank the Center for Entrepreneurship and the Information Technology of Louisiana Tech University for research funding.

# CHAPTER 1

## INTRODUCTION

### 1.1 Motivation

LBL is a well-developed methodology for the fabrication of “nano-structural” multilayer films. It has great potential for applications in a variety of areas, including biomedical engineering. Polyelectrolytes, linear polymers containing highly-charged monomers, are widely used for LBL [1]. Proteins and polypeptides constitute one of the major classes of biological macromolecules and also are polyelectrolytes because of their charge properties. Proteins have been used for LBL in recent years [2]. Designed polypeptides have recently been introduced for LBL [3]. They are attractive because they can be designed and synthesized. Also because there will be almost unlimited choices considering there are 20 natural different build blocks for a single polypeptide. Finding suitable peptide sequences, the first step for LBL, will be the key to success with peptide LBL. In Chapter 3, we propose an approach to identify structural motifs in amino acid sequence data that could be useful for polypeptide LBL. This work will be of interest to anyone interested in LBL, and it will be of particular interest for anyone concerned with the biocompatibility of structure that can be formed by LBL, notably films, coating, and microcapsules.

Polypeptides can form various secondary structures in LBL films. Circular dichroism spectroscopy (CD) has shown that multiple layers made of design peptides contain a significant amount of  $\beta$ -sheet structures at neutral pH [3-5]. Little, however, is known about the details of structural properties of these films at atomic level, and this kind of information cannot be obtained directly by experiments. A more detailed determination of the supramolecular architecture of such films has been difficult since they do not easily form single crystals, and solution-phase nuclear magnetic resonance NMR is unsuitable due to their large aggregate size. This thesis described the molecular dynamics simulation approach we have developed to investigate the behavior of designed peptides for LBL at an atomic level.

## 1.2 Background

In this section, related information is presented on electrostatic LBL, general principles of peptide design and molecular dynamics simulation.

### 1.2.1 A Brief Review of Layer-by-Layer Assembly

Making a thin film on a substrate is nothing new. Already 1200 years ago, Japanese artists used Chinese ink to create decorative patterns on a sheet of paper, the so called “spilled ink” technique. In the 19<sup>th</sup> century Agnes Pockel developed this idea into a technique by solving the problem of determining layer thickness [6]. In the early 1900s, based on the previous experimental methods, Langmuir first studied insoluble monolayers at the air-water interface and then went further to deposit many monolayers onto the same substrate [7, 8], making a pile of layers of expected thickness. Since then, methods of depositing nanoparticles to fabricate functional ultrathin films have been greatly developed. Now there are several well-developed methods available, including



classic layer-by-layer techniques such as Langmuir-Blodgett and Langmuir-Schaefer [8], and LBL self-assembly adsorption of polyelectrolytes proposed by Decher in 1991 [9].

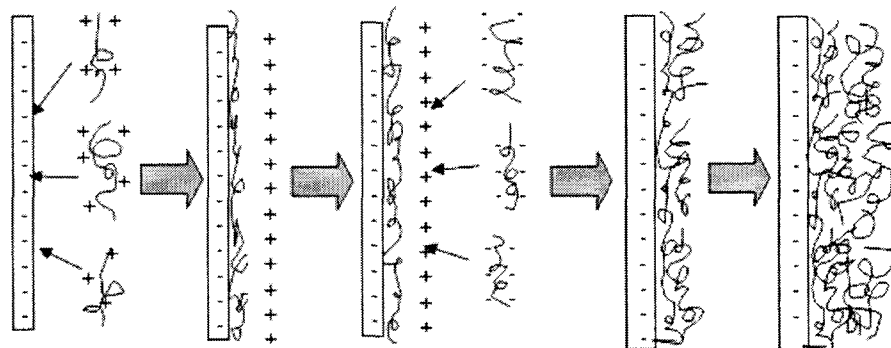
Depending on the materials used, the driving force for LBL can be different, including electrostatic interactions [9-12], hydrogen bonding [13], van der Waals interactions [14, 15], charge-transfer interactions [16, 17], and covalent bonds [18]. Among them, the LBL technique is the most popular method and has been used by the majority of people working in this area, due to the approach's being very straightforward.

### 1.2.2 LBL

Electrostatics LBL is a useful extension of the LBL method [9-11] (Fig 1.1). Different from Langmuir-Blodgett technique, electrophoresis, and other coating techniques, LBL builds up multiple layers by sequential adsorption by oppositely-charged polyelectrolytes. This method was pioneered in the 1960s by Iler [11], based on an earlier work by Langmuir [6, 8, and 19]. But it was Decher and co-workers who demonstrated that repeated adsorption of anionic and cationic polyelectrolytes [20-24] can support the formation of well-ordered "multilayers" [25-29].

By alternating the adsorption of polycations and polyanions on the same support, a multilayer film of nanometer-scale order and defined layer sequence can be obtained. The final structure of a LBL film will depend on the adsorbing species, their concentration, adsorption time, humidity, whether the film was dried after each deposited layer or at some other point, whether the film was agitated during adsorption or rinsing, the ionic, pH, temperature, and solvent of the polyelectrolyte solution [3-5]. Moreover, LBL incorporating many different functional polymers or nanoparticles will have great

potential in a broad range of biomedical-related applications such as drug delivery, membranes, filtration, and electrochromical devices.



**Figure 1.1** Schema of oppositely-charged polyelectrolytes in LBL.

Polyelectrolytes are polymer chains containing a variable proportion of ionizable monomers. A number of synthetic polyelectrolytes have been employed in LBL applications, including sodium poly(styrene sulfonate) (PSS), poly(allylamine hydrochloride) (PAH), *etc.* Such materials, however, are not generally useful for biomedical applications because they are antigenic or toxic [5, 52]. Several biopolymers such as DNA, proteins, and polypeptides have been introduced for LBL in the last several years [2, 30, and 31]. Proteins can be used for LBL; however, there are some difficulties. For example, a protein is usually large and has an irregular surface, and this makes it hard to control the LBL procedure. In contrast, polypeptides, which are short and less complex, have attracted increasing attention as a new material to form multiple thin layers by layer-by-layer self-assembly due to the great potential for applications in biotechnology, medicine, and pharmacology [32].

### 1.2.3 Peptide Design

Peptide design is a case of molecular design. There are 20 natural amino acids; one can easily imagine an almost limitless number of design possibilities, and still exists the possibility of further variation and control of the functionality through the use of non-natural amino acids. Amino acid sequence could be used to control the permeability, thickness, or elasticity of polypeptide thin films. By carefully choosing the sequence, peptide and peptide films could be engineered to have minimal toxicity and immunogenicity. A major goal of any molecular design process is to find motifs with a desired function and to rationally change the segments' structures to make or improve the desired properties. Goals of the design process per se can be achieved *in vivo*, *in vitro*, or even entirely by computer, so-called *in silico*. In this section, some advantages and considerations of the peptide design for LBL will be discussed.

The first synthetic peptide successfully used for self-assembly is a short (16 amino acid residues) sequence EAK16, designed by Suguang Zhang about ten years ago [33]. By alternately linking hydrophobic and hydrophilic amino acids, the 16-mer peptide can self assemble into a very stable insoluble membrane [33]. Since then, extensive effort has been made to understand and make good use of these kinds of peptide systems. These peptides have been reported to form unusual stable  $\beta$  sheets and macroscopic membranes in the presence of salt [47]. Also, the peptides could support mammalian cell attachment [47] and have been used as a scaffold for neurite outgrowth and synapse formation [48]. A recent study showed that these peptides can form fibrillar assemblies [49].

Using peptides as building blocks in LBL has several advantages. One, the number of possible structures is effectively limitless. There are 20 usual amino acid types, including 2 positively- and 3 negatively-charged one, 7 polar amino acid types, and 8 nonpolar types. The number of possible combinations is astronomical. For example, for a peptide of 7 residues, there will be  $20^7 \approx 8 \times 10^6$  combinations. By carefully selecting the sequence, the resulting peptides could have desired biologically relevant characteristics: for instance, minimum toxicity or immunogenicity. Such films have great potential applications in drug encapsulation and as coatings for implants. Two, a peptide is less complicated than a protein. One can easily design or synthesize a short peptide and control an assembly procedure by using short peptides as building blocks. Three, the formation of secondary structures of a peptide can be controlled. Peptides can self-organize into nanostructures with a higher level of complexity as long as the length of a peptide is larger than 4. It has been shown, for example, that the sequential deposition of cationic poly-lysine and the anionic dye congo red (CR) can result in a multilayer film containing  $\alpha$ -helix [36]. Moreover, Boulmedais and colleagues have reported that multilayer films containing  $\beta$ -sheet result from interaction between poly-glutamic acid and poly-lysine several years ago [4]. The role of secondary structure is not clear yet, but they might have some special applications in certain areas. Finally, it is cost less when synthesize a large amount of short peptides instead of synthesize proteins.

Some difficulties associated with peptides are the following. In using peptides as the material, drug delivery system, tissue scaffold, and other medicine-related areas, biocompatibility is mandatory: the resulting product must be minimally toxic to any living cell and minimally immunogenic. A synthetic short peptide will generally have the

potential to elicit an immune response, especially when conjugated to a large molecule, *e.g.* a protein [37]. To be sure of whether a compound is or is not toxic, extensive testing must be done. Prediction of the antigenicity of a peptide can be done to a limited extent using various available computer programs. The most common ones are Kyte and Doolittle [38], Hopp and Woods [39], and Emini [55]. Such approaches can be used to calculate hydrophilicity, hydrophobicity, and surface accessibility of a protein or peptide; a high hydrophilicity combined with a high surface accessibility value is often taken as a marker for a potential antigenic determinant [38-40]. Another concern about peptides is biodegradation. This is more important when designed peptides are used for tissue engineering [41]. A major challenge in peptide engineering is increasing the stability to proteolytic digestion [42]. Various approaches [43], including replacing single amino acids to form disulfide bonds [44], to increase hydrophobicity [45], or to reduce holes in the hydrophobic core [46], have been used to increase the stability of peptides or proteins.

To be able to design peptides to form multilayers with expected properties and predictable functions, it will be necessary to determine the factors that affect peptide self-assembly in certain conditions. The goal is maximum control over the peptide's physical, chemical, and biological properties. As described in Section 2.1.2, environmental factors such as temperature, pH, and ionic strength of solution affect polyelectrolyte adsorption.

#### 1.2.4 Computational Approach

1.2.4.1 Computer Simulation for Polyelectrolyte Systems. Computer simulation is an important complementary tool to experimental research for deepening our understanding of biomolecular systems [50, 54]. This method has attracted increasing interest as a scientific and engineering technique due to greatly increased

computational capability. In the last decade, this technique has been applied to various fields such as physics, chemistry, biology, material science, and drug design. The application of various computational methods to solve physical and chemical problems has accelerated at a prodigious rate. In addition to share promise to predict properties of molecular systems, computational methods have the potential to be an essential engineering tool to design a novel material at the molecular level. So far, many computational methodologies have been developed and validated in various fields of physics and chemistry. For example, rigorous quantum mechanical theory can provide useful information on the electronic properties of materials. To investigate a system with a large number of atoms, however, quantum mechanical computation is too expensive to be used. Another category of computation is the force field method, which allows investigation of the structure and energetics of a larger system, of hundred nanometers, using various statistical ensembles if electrons per se are not of interest.

There are several simulation techniques which fall into this category. To simulate a polyelectrolyte system, the Monte Carlo (MC) method, MD simulations, and the Brownian dynamics (BD) simulations have been applied. A synopsis of each method is given below.

The first computer simulation of a molecular system involved the MC method [54]. Monte Carlo is a statistical method which uses a random number generator and probability to solve problems in physic, chemistry, biology, and even economic. The general procedure is as follows:

1. Assign initial coordinates to all particles in the molecular system.
2. Randomly choose one of the particles  $i$  in the system and displace it randomly.

3. Calculate the energy difference before and after displacement.
4. Choose a random number  $r$  between 0 and 1. If the value  $\exp(-\Delta E/kT)$  is larger than  $r$  ( $T$  is the temperature,  $k$  is the Boltzmann constant), the new position of particle  $i$  is accepted and a new system conformation of the system is created. Otherwise, the displacement is not valid and the particle  $i$  is remaining in its original position.
5. Trajectory data are stored, including energy, velocity, and molecular conformations.

Steps 2-5 are repeated until average values of the trajectory data “converge”. If information on atomic positions of solvent is not need, MC method might be a good choice for simulations a polyelectrolyte.

MD is the most widely used method for studying the dynamic behavior of molecular systems. Like MC, MD also starts with an initial structure and signed velocities. Newton’s second law is applied to the molecular system and solved by numerical methods. Detailed information on time correlation functions can be obtained from the trajectory data.

BD is another simulation approach. The basic principle is similar to that involved in MD simulation. But instead of integrating Newton’s second law, BD integrates the Langevin equations for dynamics and uses an implicit continuum model to simulate the solvent.

1.2.4.2 Simulating Internal Forces. For computer simulations, the complexity of a system must be reduced to some extent, but the model must maintain essential features of the solvent and solute interactions. The solvent environment, for example, will often be

replaced by a mean-field potential or just a dielectric constant. The molecule itself is treated as a linked chain of charged-atoms of regular shape. Then the electrostatic potential between any two charged atoms can be modeled by the Coulomb equation:

$$V_{ij}(r) = \frac{q_i q_j}{4\pi\epsilon_r \epsilon_0 r} \quad (1.1)$$

where  $q$  is the charge measured in Coulomb (C), the unit of electrostatic charge,  $\epsilon_0$  is the permittivity of vacuum ( $8.85 \times 10^{-12}$  Farad/meter),  $\epsilon_r$  is the relative permittivity of the solvent, and  $r$  is the distance between particles  $i$  and  $j$ . The factor  $1/4\pi\epsilon_0$  is also known as the electrostatic constant  $k$ . When simulating non-bonded electrostatic forces, a cutoff distance parameter is introduced to decrease the complexity of the calculation. In general, if  $r$  is larger than the cutoff distance ( $\sigma$ ), the electrostatic force is set to zero by a switching function. The electrostatic force between any pair of charged particles is then given as:

$$V_{ij} = \begin{cases} 0 & r > \sigma_{ij} \\ \frac{q_i q_j}{4\pi\epsilon_r \epsilon_0 r^2} & r < \sigma_{ij} \end{cases} \quad (1.2)$$

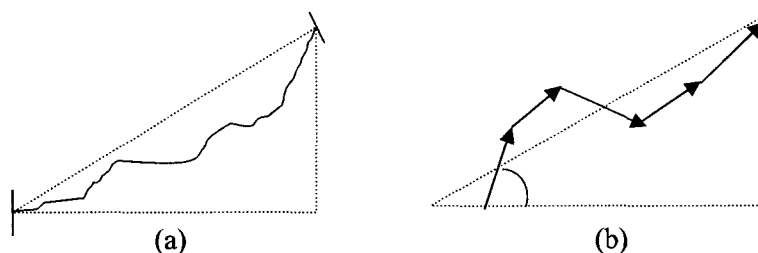
where  $\sigma_{ij}$  is the cutoff distance.

1.2.4.3 Theoretical Description of a Polymer Chain. A polyelectrolyte is a highly-charged polymer in an aqueous solution. A polymer system usually refers to the polyelectrolyte, solvent, counterions, and coions in solution. A number of theoretical approaches have been developed, and experimental studies aimed at describing the relationship between the structure and properties of a polyelectrolyte system have been done. But there are always discrepancies between theory and experiment. One way to



accelerate the closing of the gap between theoretical studies and experimental observations is to develop semi-empirical computer simulation techniques.

A simple way to represent a flexible polyelectrolyte chain is to replace each unit of the chain with a mathematical unit and ignore the true covalent structure. Two such models have been used to describe a polymer chain: a continuous chain and a discrete chain [50]. A continuous chain model, shown in Figure 1.2a can be solved by analytical methods. A discrete chain model, given in Figure 1.2b is used in computer simulations.



**Figure 1.2** Schema of a polyelectrolyte chain. (a) Continuous chain. (b) Discrete chain.

### 1.3 Organization of this Dissertation

This dissertation combines computer science, bioinformatics, polymer science, and knowledge of the immune system to identify “peptide motifs” in human genome data, use the motifs to design polypeptide for LBL, and use computational approaches to deepen our understanding of the structure of polypeptides films. An introduction has been presented in the present chapter.

Chapter 2 is a detailed description of MD.

In Chapter 3 we propose a highly-interdisciplinary computational approach to design peptides LBL.

In Chapter 4 we report results of MD of peptide models based on two synthetic oppositely-charged peptides at the neutral pH.

In Chapter 5 we report results of MD of peptide models based on the peptide sequences designed in Haynie's research group on the basis of results presented in Chapter 3.

Conclusions and further studies are addressed in Chapter 6.

## CHAPTER 2

### GOVERNING EQUATIONS OF MD SIMULATION

#### 2.1 Introduction

The idea of MD simulation was suggested in the late 1950's by Alder and Wainwright [56, 57], who studied the interactions of hard spheres. In 1964, the first simulation of liquid argon using a realistic potential was done by Rahman [58]. The first MD simulation of a realistic system, liquid water, was performed by Rahman and Stillinger in 1974 [59]. The simulation of a protein, bovine pancreatic trypsin inhibitor (BPTI), carried out by McCammon in 1977 [60]. Since then, along with the great increase in the power of computers, MD has become the most widely used method to study the structure, energetics, and thermodynamics of macromolecules at the atomic level.

The basis of a MD simulation is Newton's force equation. It is given as:

$$F_i = m_i a_i \quad (2.1)$$

where  $F_i$  is the force acted on atom  $i$ ,  $m_i$  is the mass of atom  $i$  and  $a_i$  is the acceleration of particle  $i$ . The force can also be expressed as the gradient of the potential energy as given below:

$$F_i = -\nabla_i U \quad (2.2)$$

where  $U$  is the potential energy of the system. Combining these two equations yields

$$\frac{dv_i(t)}{dt} = -m^{-1} \frac{dU}{dr_i} \quad (2.3)$$

where  $v_i(t)$  is the velocity of atom  $i$  at time  $t$ .

For a two-atom system, this equation can be solved exactly, but more than two atoms, it is too complicated to be solved analytically, and therefore it must be solved by numerical methods. Numeric integration is typically done step by step using methods that are called Finite difference methods. These methods are explicit and use the information available at time  $t$  to predict the system's coordinates and velocities at a time  $t + \Delta t$ , where  $\Delta t$  is a short time interval. Two most common used methods, Verlet and leapfrog methods are detailed in the section of Integration Methods.

A general procedure of a MD can be summarized as follows: given the system state  $S(t_0)$ , that is, the position  $r$  and velocity  $v$  of every atom in the system at time  $t_0$ , subsequent states  $S(t_0 + \Delta t)$ ,  $S(t_0 + 2\Delta t)$ , ..., are calculated by numerical integration of Newton's law  $F = ma$ . To calculate  $S(t_0 + (n+1)\Delta t)$  from  $S(t_0 + n\Delta t)$ , first for every particle  $i$ ,  $F_i(t_0 + n\Delta t)$  is calculated.  $F_i(t_0 + n\Delta t)$  is the sum of the forces on  $i$  as exerted by the other particles of the system at time  $t_0 + n\Delta t$ . For every particle  $i$  the force  $F_i(t_0 + n\Delta t)$  is then integrated to get the new velocity  $v_i(t_0 + n\Delta t)$ . Using this velocity, for every particle  $i$  the new position  $r_i(t_0 + (n+1)\Delta t)$  can be calculated. For accurate results small timesteps  $\Delta t$  have to be used.

## 2.2 Integration Methods

All integration methods assume that the positions, velocities, and accelerations can be approximated by a Taylor series expansion. The first and simplest integration

method is Verlet algorithm [61]. Particle positions  $r$  are calculated at time  $t_0 + n\Delta t$ . The formulas for Verlet integration are

$$r_i(t + \Delta t) = r_i(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \quad (2.4)$$

$$r_i(t - \Delta t) = r_i(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \quad (2.5)$$

Adding these two equations leads to the prediction of the position at time  $t + \Delta t$ :

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + a_i(t)\Delta t^2 \quad (2.6)$$

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + \frac{F_i}{m_i}(t)\Delta t^2 \quad (2.7)$$

where  $F_i$  is the force exerted on atom  $i$  with mass  $m_i$ . Note that the velocities do not appear explicitly, but they can be easily computed from the positions. Another basic integration method is Leap-frog algorithm in which the velocity is incorporated. The Leap-frog equation is given in two steps:

$$r_i(t + \Delta t) = r_i(t) + \Delta t v_i(t + \frac{1}{2}\Delta t) \quad (2.8)$$

$$v_i(t + \frac{1}{2}\Delta t) = v_i(t - \frac{1}{2}\Delta t) + \frac{F_i}{m_i}(t)\Delta t \quad (2.9)$$

where  $F_i$  is the force that applied on the atom  $i$  with mass  $m_i$ .

The advantage of this algorithm is that the velocities are explicitly calculated and the numerical error of Verlet algorithm is reduced. However, the disadvantage is that they are not calculated at the same time as the positions.

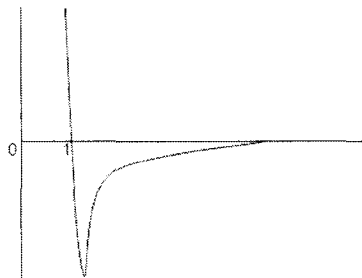
### 2.3 Force Fields

Current force fields (or potential energy functions) provides a reasonably compromise between accuracy and computational efficiency. In an MD simulation there

are two classes of interactions: non-bonded and bonded. Non-bonded interactions model the interactions between the particles are not linked together by a covalent bond, including the van der Waals force and electrostatic interaction. The van der Waals interactions are often represented by an energy potential as a function of distance  $r$  that includes both the attractive force and the repulsion at close range. The most well-known of these is the Lennard-Jones potential [62]. For any two atoms with spherical symmetry at a distance  $r$ , the Lennard-Jones potential is given

$$V_{LJ}(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (2.10)$$

where  $\epsilon$  and  $\sigma$  are the specific Lennard-Jones parameters. The first term gives the repulsions, the second the van der Waals attractions. A plot of the typical Lennard-Jones potential is shown in Figure 2.1. Some important characteristics are illustrated from the plot. At short range ( $r$  is small), the potential energy is very large and positive, indicating that the two atoms are strongly overlapping. As the distance between two atoms is increased, Lennard-Jones force is decreased. When the distance approaches to the sum of the radii of the two atoms, a minimum is reached. The electrostatic force can be modeled by Coulomb's law (Eqn. 1.1) as mentioned in Chapter 1.



**Figure 2.1** Lennard-Jones potential.

The internal interactions model the relatively strong chemical bonds which are not created or broken during a MD simulation. They generally include three parts: bond stretching ( $E_{\text{bond}}$ ), bond-angle bending ( $E_{\text{angles}}$ ), and dihedral torsion angle ( $E_{\text{torsions}}$ ). The bond-stretching potential models the interaction between any two particles linked by a covalent bond, and it is given as follows:

$$V = \frac{1}{2} K_b (r_{ij} - b_0)^2 \quad (2.11)$$

where  $b_0$  is the ideal length of the covalent bond.

The bond-angle bending is bond angles defined by three particles:  $i$ ,  $j$ , and  $k$  (Figure 2.2a). It can be calculated as:

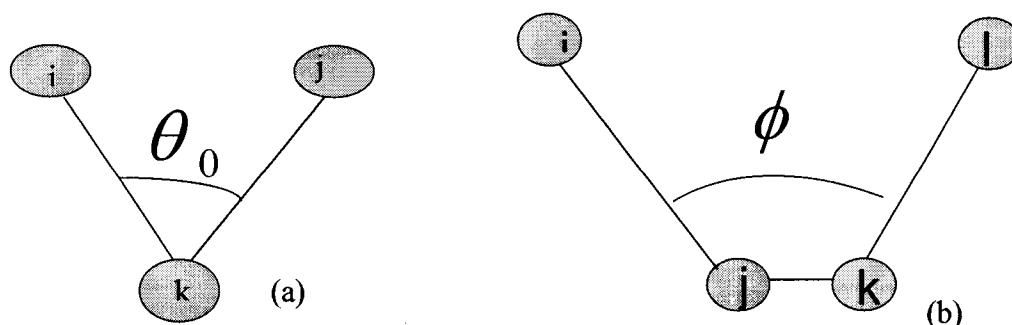
$$V = \frac{1}{2} K_\theta (\theta - \theta_0)^2 \quad (2.12)$$

where  $\theta_0$  is the ideal angle between  $i$ ,  $j$ , and  $k$ .

The dihedral torsion angle  $\phi$ , also known as the improper dihedral angle, is an interaction between four particles,  $i$ ,  $j$ ,  $k$ , and  $l$  (Figure 2.2 b) and is expressed as:

$$V = K_\phi (1 + \cos(n\phi - \phi_0)) \quad (2.13)$$

where  $\phi_0$  is the ideal torsion angle.



**Figure 2.2** Internal interactions (a) Bond-angle. (b) Torsion angle.

The potential energy,  $V$ , is a function of the atomic positions of all the atoms in the system, can be expressed as either the sum of the internal and external interactions, or it can be evaluated using more complex molecular mechanics (MM) force fields [63-65]. The basic components of any forcesfields or energy functions, however, are the same, usually including two parts: internal and external interactions written as below:

$$V(R) = \sum_b K_b (b - b_0)^2 + \sum_\theta K_\theta (\theta - \theta_0)^2 + \sum_\phi K_\phi (1 - \cos(n\phi)) + \sum_{ij} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{ij} \frac{q_i q_j}{r^2} \quad (2.14)$$

where  $K_b$ ,  $K_\theta$ ,  $K_\phi$  are constant values, obtained empirally from the study of small molecules; and  $b_0$  and  $\theta_0$  represent ideal values of bond length and angle, respectively.

There are many different types of forcefield and MD packages, including Charmm, AMBER, MOIL, and GROMOS. The choice for one rather than another depends on the purpose of the simulation and the kind of molecular system used in simulation. The atoms in a system are treated as the smallest particles and the total energy as the sum of all the forces exerted on each one the atoms by other atoms in the system. The goal is that the force field will approximately represent the forces that acturally exist in the system.

#### 2.4 Solvent Models

Solvent plays a key role in determining a biological molecule's structure, function, stability, and intermolecular interactions. A realistic simulation of a molecule must therefore include the effects of water (solvent). There are two ways to incorporate solvent effects in MD: explicit and implicit models. The ideal simulation environment is to use explicit solvent models [66, 67]. The major advantage of an explicit solvation is to



provide detailed information on the interaction between a bimolecular system and its environment. The cost, however, is a great increasing computing complexity and thus the time for simulation. The approach also recognizes a limited number of atoms.

#### 2.4.1 Explicit Solvent Models

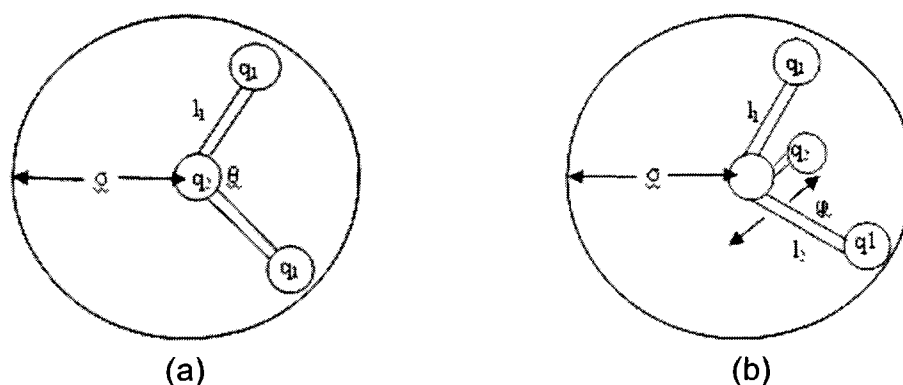
In explicit solvent models, solvent molecules are treated as explicitly. This will provide time dependent information on the motion of all the atoms. Water is the most commonly used solvent model. There are several most popular water models: TIP3P [67], TIP4P [68], SPC [69], PPC [70] (Fig. 2.1). The TIP3P model treats a water molecule as three particles: one oxygen and two hydrogen atoms. There is no internal flexibility. The simple point charge (SPC) model is similar to the TIP3P model. Both have similar atomic partial charges on the oxygen and hydrogen atoms. The TIP4P is a 4-point model with Lennard-Jones interaction between oxygen and three charge sites. Compared with the TIP3P, it adds a negative interaction site in the center of the oxygen atom. Table 2.1 lists the geometry and energy properties of these models.

Overall, all these water models give approximately accurate structure and thermodynamic description of bulk water [68, 70, and 71]. The TIP3P water model and charmm22 force field were chosen for all the simulations we carried out. This combination has been shown to produce an accurate simulation for various peptides systems [72, 73].

#### 2.4.2 Implicit Solvent Models

The major disadvantages of an explicit solvent are time consuming computation and a limited number of atoms. Since a large number of water molecules will slow a simulation, an alternative way to speed it up and obtain a reasonably accurate description

of a solute-solvent interaction is to use an implicit solvent model. Instead of using explicit water molecules, the average solvent forces on a solute are used to represent the solvent effect. There are many different implementations of this method. Among them, the generalized Born (GB) method is a popular one [74]. This method was first introduced by Still and co-workers and then extended by Dominy and Brooks [74, 75, and 76]. In a GB model, the total solvation energy is equivalent to the sum of solvent-solvent cavity energy ( $G_{cav}$ ), solute-solvent van der Waals energy ( $G_{vdw}$ ), and solute-solvent electrostatic polarization energy ( $G_{pol}$ ). The first two terms are approximated by calculating the solute surface accessible area. The last term is calculated by the finite difference solution to the Poisson equation (FDPB).



**Figure 2.3** Explicit water models. (a) Structures of TIP3P, SPC, PPC. (b) Structure of TIP4P.

Table 2.1 Geometrical and energetic properties of various water models.

Properties	SPC <sup>[60]</sup>	TIP3P <sup>[59]</sup>	TIP4P <sup>[59]</sup>	PPC <sup>[61]</sup>
$l_1$ (Å)	1.0	0.9542	0.9542	0.9542
$l_2$ (Å)	-	-	0.15	0.06
$\theta$ (deg)	109.47	104.52	104.5	106.00
$q_1$	0.41	0.417	0.52	0.517
$q_2$	-0.82	-0.843	-1.04	-1.034
$\sigma^a$	3.166	3.15061	3.15365	3.234
$\varphi$	-	-	52.26	127.00
$\varepsilon$ (kJmol/mol)	625.5	636.4	648	600

<sup>a</sup> distance between two same charged atoms.

## CHAPTER 3

### PEPTIDE MOTIF DATABASE

In last few years, designed peptides have aroused extensive interest in bioengineering and nanotechnology [33, 77, 78, 79, 80]. Composed of different amino acids, designed peptides have many applications including coatings, drug delivery, and artificial skin. In our research, we are interested in the design of peptides for LBL and the development of applications that will not stimulate an immune response, making them suitable for medical usage. A new strategy has been conceived and developed to minimize the immune reaction. 87,779 human protein sequences have been extracted from National Center of Biotechnology Information (NCBI) protein database in autumn 2001. Following our criteria, 54117 positively- and 27117 negatively-charged non-redundant sequence motifs have been identified in the human protein sequences. A sequence motif has physical properties advantageous for LBL.

#### 3.1 Introduction

LBL is an established technique in which ultra thin films are assembled by the alternating adsorption of oppositely-charged polyelectrolytes. The process is based on the reversal of the surface charge of the film after the deposition of each layer. This process is repeated until a film of desired thickness is formed. Because of the generality

and relative simplicity of the process, LBL allows for the deposition of many different types of materials onto many different types of surfaces. There are, therefore, a vast number of possible useful combinations of materials and surfaces [1, 2, 4, and 5].

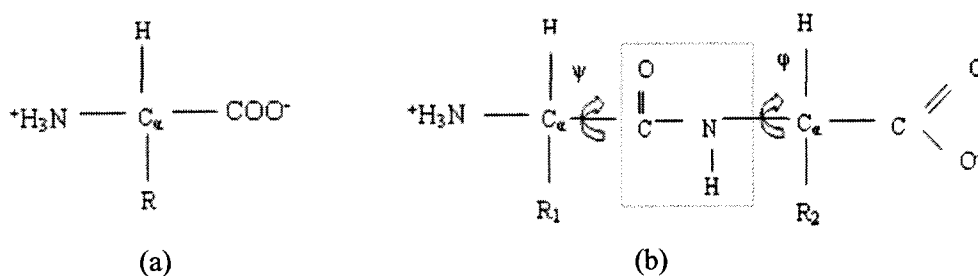
A number of synthetic polyelectrolytes have been employed in LBL applications, including sodium poly (styrene sulfonate) (PSS), poly (allylamine hydrochloride) (PAH), poly (diallyldimethylammonium chloride) (PDDA), poly (acrylamide-co-diallyldimethylammonium chloride), poly (ethyleneimine) (PEI), poly (acrylic acid) (PAA), poly (anetholesulfonic acid), poly (vinyl sulfate) (PVS), and poly (vinylsulfonic acid). Such materials, however, are not generally useful for biomedical applications because they are antigenic or toxic [5].

Proteins, being polymers with side chains having ionizable groups, can be used in LBL for various applications, including biomedical applications. Examples of proteins that have been used in LBL include cytochrome *c*, hen egg white lysozyme, immunoglobulin G, myoglobin, hemoglobin, and serum albumin [2]. There are, however, difficulties with using proteins for this purpose. These include limited control over multilayer structure (because the surface of the protein is highly irregular and proteins will not ordinarily adsorb on a surface in a regular pattern), restrictions on pH due to the pH-dependence of protein solubility and structural stability, lack of biocompatibility when using exogenous proteins, and the cost of scaling up processes if the gene has not been cloned, making the protein effectively unaffordable for large-scale production.

By contrast, polypeptides, which are generally smaller and less complex than proteins, constitute an excellent class of material for LBL assembly, and polypeptide film

structures formed by LBL will be useful in a broad range of applications. In this paper, we present a novel idea to design peptides, which will exhibit several useful properties, including without limitation, completely determined primary structure, minimal secondary structure in the solution, monodispersity, completely controlled net charge per unit length, the ability to form cross-links on demand, the ability to form vastly superior thin films than protein, and relatively inexpensive large scale production cost.

Polypeptides designed using the method has been shown useful for LBL of thin film structures with possible applications in biomedical technology, food technology, and environmental technology. Such polypeptides could be used, for example, to fabricate artificial red blood cells, drug delivery devices, and antimicrobial films.



**Figure 3.1** Structure of an amino acid. (a).The zwitterionic form, predominant at neutral pH is shown. (b) Structure of a dipeptide. A peptide bond joins the two amino acids. The dihedral bond angles  $\phi$  and  $\psi$  are shown. ‘R’ represents the side chain.

### 3.2 General Properties of Amino Acids

The 20 natural amino acids have a general structure: an  $\alpha$ -carbon atom links an amino group to its left side, a carboxyl group to its right side, a hydrogen atom and a different side chain also attach to the  $\alpha$ -carbon atom (Figure 3.1). The only difference of each amino acid is due to its side chain (R group). All amino acids found in live

creatures are of L configuration ( $^+$ NH<sub>3</sub> group on the left). According to the side chain's properties, the 20 amino acids can be divided into charged, polar or nonpolar groups. There are five charged natural amino acids. Arginine, lysine, histidine are basic hydrophilic amino acids and each of them contains an amino group. Aspartic acid and glutamic acid are acidic hydrophilic amino acids and each of them contains a carboxylate group. The properties of the five charged amino acids are summarized in Table 3.1.

### 3.3 Protein/Peptide Structures

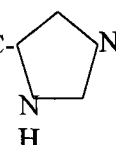
There are nearly 1,800 proteins and peptides deposited in the protein databank (www.pdb.org). The structures of more than 85% of proteins in PDB database are determined by X-ray crystallography, and the others are determined by nuclear magnetic resonance (NMR) or by theoretical models. For most proteins only a single structure is known, but in some cases multiple structures can be found. Proteins are not rigid; they can adopt several, but similar conformations in aqueous environment depending on the pH, the ionic strength, the solvent, *etc.* A protein can have several levels of hierarchical structures, *i. e.* primary, secondary, tertiary, and quaternary structures. A polypeptide structure may have a primary and secondary structure.

Primary structure refers to the linear combination of different amino acids by a covalent peptide bond. A peptide bond forms between the carbon atom (C) of the carboxyl group and the nitrogen atom (N) of the amino group.

Secondary structure refers to the arrangement in space of a polypeptide chain. The conformation of a polypeptide can be described by rotations about the bonds C<sub>α</sub>-C ( $\psi$ ), N-C<sub>α</sub> ( $\phi$ ), and CO=NH ( $\omega$ ) bonds. The symbols of  $\phi_i$ ,  $\psi_i$  refer to the torsion angles of amino acid residue  $i$ , defined as the angle of the atoms C<sub>*i-1*</sub>-N<sub>*i*</sub>-C<sub>*i*</sub><sup>α</sup>-C<sub>*i*</sub> and N<sub>*i*</sub>-C<sub>*i*</sub><sup>α</sup>-C<sub>*i*</sub>-

$N_{i+1}$ , respectively, and  $\omega_i$  refers to the angle between the atoms  $C_i-C_{i+1}$  (Figure 3.1). There are three common types of secondary structure:  $\alpha$  helix,  $\beta$  sheet, and  $\beta$  turn. An ideal  $\alpha$  helix in a protein may contain about 10 amino acid residues (1.5nm in length). This structure is stabilized by hydrogen bonds formed between the CO of residue  $i$  and the NH of residue  $i + 4$ . There are several types of helixes: the common right-handed  $\alpha$ -helix, the rare left-handed  $\alpha$ -helix, the  $3_{10}$  helix, and others. All residues participating in an ideal  $\alpha$ -helix have about the same  $\phi$ ,  $\psi$  angles.

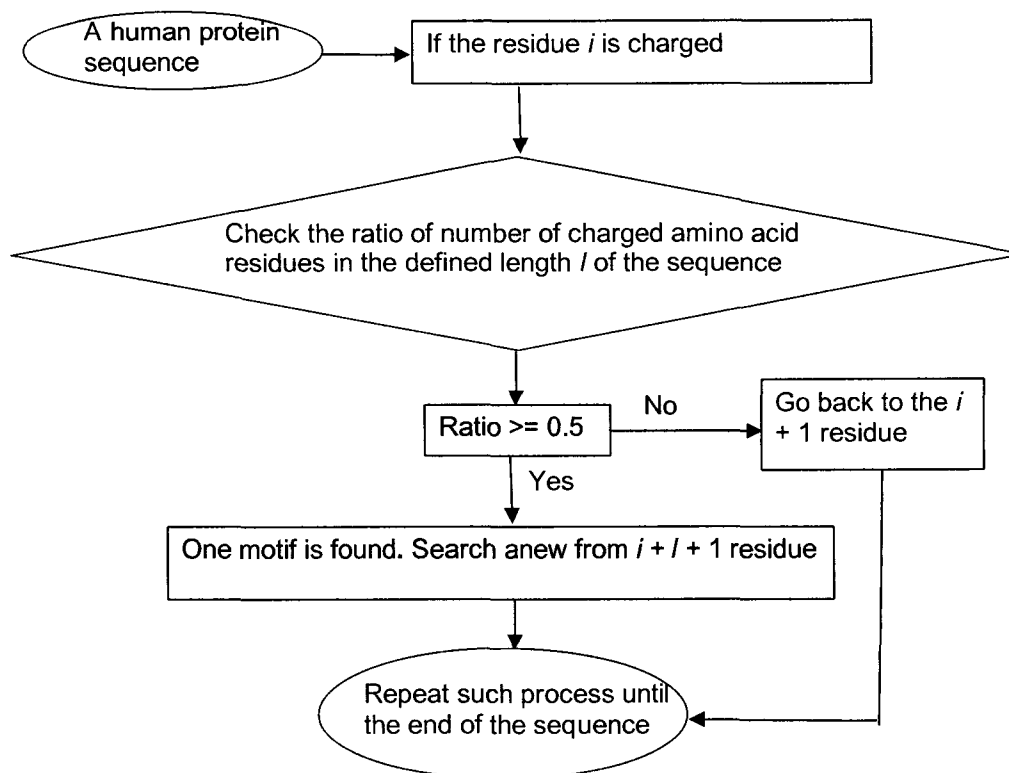
Table 3.1 Properties of five charged amino acid residues.

Name Three-letter code One-letter code	Formula	Side Chain Group	pK <sub>R</sub> *	pI**	pK <sub>a</sub> - NH <sub>3</sub> <sup>+</sup> *	pK <sub>a</sub> - COOH*
Arginine Arg R	C <sub>6</sub> H <sub>14</sub> N <sub>4</sub> O <sub>2</sub>	-NH-C-NH <sub>2</sub>    NH <sub>2</sub> <sup>+</sup>	12.48	10.74	8.99	1.82
Lysine Lys	C <sub>6</sub> H <sub>14</sub> N <sub>2</sub> O <sub>2</sub>	-NH <sub>3</sub> <sup>+</sup>	10.54	9.80	9.06	2.16
Histidine His H	C <sub>6</sub> H <sub>9</sub> N <sub>3</sub> O <sub>2</sub>	-H <sub>2</sub> C- 	6.04	7.49	9.33	1.80
Aspartic acid Asp D	C <sub>4</sub> H <sub>7</sub> NO <sub>4</sub>	R-COOH	3.90	2.95	9.90	1.99
Glutamic acid Glu E	C <sub>5</sub> H <sub>9</sub> NO <sub>4</sub>	R-COOH	4.07	3.09	9.47	2.10

\* Source: Dawson, R.M.C., Elliott, D. C., Elliott, W. H. and Jones, K. M.. 1986. *Data for Biochemical Research (3<sup>rd</sup> ed.)*. 1-31. Oxford Science Publications.

\*\* For positively charged amino acids, the pI value = (pK<sub>a</sub>-COOH + pK<sub>R</sub>) / 2. For negatively charged amino acids, the pI value = (pK<sub>a</sub>-NH<sub>3</sub><sup>+</sup> + pK<sub>R</sub>) / 2. For other amino acids, pI = (pK<sub>a</sub>-NH<sub>3</sub><sup>+</sup> + pK<sub>a</sub>-COOH) / 2.





**Figure 3.2** Flowchart of the identification sequence motif process.

A  $\beta$  sheet is an extended form of polypeptide in which hydrogen bonds are formed between two adjacent segments of polypeptide backbone. The  $\beta$  strands in a  $\beta$  sheet can be either parallel or anti-parallel. ‘Anti-parallel’ refers to the two adjacent strands running in opposite directions; this is the more common form of  $\beta$ -sheet. ‘Parallel’ refers to the two adjacent strands running in the same direction. If the strands in a  $\beta$  sheet have an anti-parallel or parallel orientation only, it is called an anti-parallel or parallel  $\beta$  sheet, respectively. Otherwise, it is called a mix sheet.

A  $\beta$ -turn is a short secondary structure, containing only about 4 residues. It is formed by hydrogen bonds between the CO group of residue *i* and the NH group of residue *i*+3.

### 3.4 Methods

A sequence motif is a design element of a polypeptide intended to optimize the physical, chemical, and biological properties of structures fabricated by LBL. There are several criteria for identifying a qualified motif in a polypeptide sequence. The most important is that a sequence motif must satisfy a certain charge requirement in order to be suitable for LBL. It must have a net charge of at least 0.5, and ordinarily all charged amino acids residues in a motif will be of the same sign. Also, the length of a motif should not be too long, because long peptides are difficult to synthesize and manipulate. The length of a sequence motif was set to 7 for reasons discussed below. Figure 3.2 shows the flowchart of the motif identification procedure. In Table 3.2 the minimum numbers of charged amino acids of motif are given for different lengths, respectively. The numbers of unique sequence motifs in available human protein sequence data was calculated for different lengths from 1 to 15 (Figure 3.3). Based on our calculations, the length of 7 gave the maximum number of motifs.

Table 3.2 Motif length vs. the number of amino acids

AA length	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Minimum Number of charged AA	1	2	2	3	3	4	4	5	5	6	6	7	7	8	9

The pseudocode for identifying a sequence motif given as below:

**Input:** A human protein sequences file

**Output:** Redundant negative or positive sequences length of seven

**Function Extraction**

*n*: length the protein sequence

*l* : the defined length of a peptide motif

count: counting the positively- or negatively-charged peptide amino acid

ratio: count / *l*

*i*: current amino acid

**While** *i* != *n* **do**

**If** *i* is charged amino acid **then**

        Check the following *l* amino acids

**If** ratio > 0.5 **then**

            A peptide motif is found and record

            Check the next *i* + *l* amino acid

**Endif**

**Else**

            Check the *i*+1 amino acid

**Endif**

        Check the next amino acid to *i*

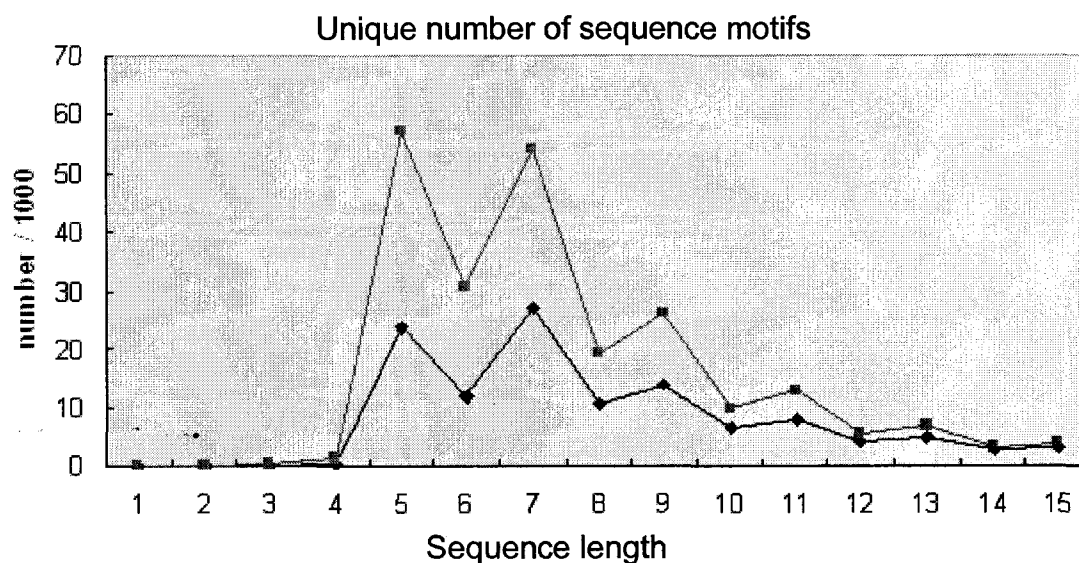
**Endwhile**

**End** extraction

Protein secondary structure prediction has been of interest topic for half a century.

Many computer algorithms have been developed to predict the secondary structure of a polypeptide based on amino acid sequence alone [82, 83, and 85]. Among them, Chou and Fasman's [82, 83] method to calculate  $\alpha$  helix and  $\beta$  sheet propensity values is a simple but efficient one and is still used today some 30 years after its development. Besides secondary structure prediction, other physical properties of the proteins have also been studied widely, such as hydrophobicity, hydrophilicity [38], surface accessibility, and antigenic site prediction [55]. To obtain a set of more accurate secondary structure parameters, we increased the sample space from 15 to about 1,500 proteins. The 1,500 protein structures from the Protein Data Bank were chosen using the following criteria: the resolution of each protein was better than 2.0 Å; the technique for determining the

protein structures was X-ray crystallography only; and proteins of 50% or greater and theoretical models were excluded. The secondary propensity values of the 20 amino acids we calculated are shown in Table 3.3.



**Figure 3.3** The number of non-redundant amino acids as function of sequence length.

### 3.5 Results and Discussions

This work identified a total of 54,117 and 27,115 non-redundant positively- and negatively-charged amino acids motifs with a length of 7 in human protein sequences. These sequence motifs were stored in a relational database and can be accessed by a graphical interface written in Visual Basic [84].

The secondary propensity values of the 20 amino acids were calculated as described here and the results are given in Table 3.3. There is a good agreement between our calculated secondary structure propensities and those of Chou and Fasman. Alanine, glutamic acid, glutamine, and leucine are found most often in  $\alpha$ -helix, while the

hydrophobic amino acids valine, isoleucine, and phenylalanine are the most common ones in  $\beta$ -sheets in proteins (Figure 3.4). There are still some noticeable differences, however, which are highlighted by '\*' in Table 3.3. Comparing the probability of  $\alpha$ -helix propensity values, histidine decreased from 1.24 to 0.96, and arginine increased from 0.79 to 1.17. Comparing the  $\beta$ -sheet propensity values, glutamic acid increased from 0.26 to 0.74, glutamine decreased from 1.23 to 0.79, and methionine decreased from 1.67 to 0.96. Such discrepancies are not surprising: Chou and Fasman's method is a statistical method; the total number of proteins they used to calculate the secondary propensity values was just 15; the database of protein structures was very limited 30 years ago.

The probability of the 20 amino acids appearing at the N terminus (N-end) or the C terminus (C-end) of a helix is given in Table 3.3. Proline has the lowest  $\alpha$  helix,  $\beta$  sheet, and C-end propensity value, but the highest N-end propensity value of (Figure 3.4). This agrees well with experimental research showing that proline destabilizes secondary structure; if there is a proline in a protein sequence, it almost always appears at the beginning of an  $\alpha$  helix as a  $\alpha$  helix breaker [87]. Asp ranks second in the  $P_{N\text{-end}}$  column, consistent with its ability to interact favorably with the helix dipole. Glu is the third common amino acid at the N-end. For the C terminus, the three positively charged amino acids, L, H, and A, are the most common ones, again consistent with their ability to interact favorably with the helix dipole. It should be noted that even if the sample space is 100 times larger than Chou and Fasman's, as suggested by Kyngäs and co-workers, the results still cannot be highly accurate due to the limitation of the method itself [86]. But we should not doubt that Chou and Fasman's method is a useful one for secondary structure properties.

Table 3.3 Amino acid frequency and propensity to form secondary structure

Name	$F_{\alpha}^a$	$P_{\alpha}^a$	$F_{n-end}^b$	$P_{n-end}^b$	$F_{c-end}^b$	$P_{c-end}^b$	$F_{\beta}^a$	$P_{\beta}^a$
Ala	0.49	1.32	0.21	0.84	0.26	1.09	0.18	0.78
Arg	0.44	1.17*	0.20	0.79	0.28	1.16	0.18	0.82
Asn	0.34	0.90	0.24	0.95	0.27	1.12	0.17	0.77
Asp	0.35	0.94	0.34	1.35	0.18	0.77	0.14	0.61
Cys	0.33	0.88	0.20	0.82	0.24	1.02	0.27	1.20
Gln	0.45	1.21	0.24	0.98	0.26	1.07	0.18	0.79*
Glu	0.48	1.27*	0.31	1.23	0.23	0.96	0.16	0.74*
Gly	0.25	0.65	0.28	1.12	0.26	1.07	0.17	0.76
His	0.36	0.96*	0.23	0.93	0.27	1.14	0.21	0.95
Ile	0.37	0.98	0.21	0.86	0.22	0.94	0.35	1.57
Leu	0.46	1.21	0.18	0.73	0.29	1.21	0.25	1.11
Lys	0.42	1.13	0.21	0.82	0.30	1.27	0.18	0.80
Met	0.42	1.12	0.19	0.77	0.26	1.09	0.21	0.96*
Phe	0.37	0.98	0.23	0.93	0.24	1.01	0.29	1.31
Pro	0.22	0.58	0.44	1.75	0.11	0.45	0.11	0.51
Ser	0.35	0.92	0.27	1.09	0.23	0.95	0.20	0.89
Thr	0.32	0.85	0.29	1.18	0.21	0.88	0.27	1.19
Trp	0.39	1.05	0.24	0.95	0.21	0.88	0.28	1.23
Tyr	0.37	0.98	0.23	0.91	0.25	1.04	0.29	1.30
Val	0.33	0.87	0.24	0.98	0.21	0.86	0.38	1.70

<sup>a</sup>  $F$ ,  $P$  refer to the frequency and propensity values of each amino acid occurred in  $\alpha$ -helix or  $\beta$ -helix.

<sup>b</sup>  $F_{end}$ ,  $P_{end}$  refer to the frequency and propensity values of each amino acid occurs in the N-end or C-end of  $\alpha$ -helix.

\* Significant difference from Chou and Fasman value.

Biocompatibility was assessed by computer algorithms. We computed the antigenicity of each peptide motif by several most frequently-used methods: Kyte and Dolittle [38], Hopp and Woods [39], and Emini accessibility [55]. The hydrophobic and hydrophilic surface areas and surface accessibility values based on the three methods were calculated for each motif. A high hydrophilicity value combined with a high accessibility value is assumed to indicate a possible antigenic determinant.

### 3.6 Database Configuration

To store and use the huge amount of reference data, a relational database has been built. Oracle 8.0 (personal edition) was chosen as the backend database engine. Oracle 8 personal edition is designed especially for personal use but has the major features of the commercial Oracle 8.0 database. The advantages of it include: easy to use, stable, powerful, and free.

Two tables designed to store the motif information. The main information stored is summarized in Table 3.4. In the database, identification number is the unique key for locating motifs quickly.

### 3.7 Experimental Results Involving Designed Polypeptides

Motifs for use in experimental work were selected from human blood proteins using the process described in Methods: complement C3 (gi|68766) for the anionic peptides, and lactotransferrin (gi|4505043) for the cationic peptides. The positive and negative motifs were repeated 4 (for short polypeptides) or 6 (for long polypeptides) times for peptide design (Table 3.5). A glycine was introduced between each 7-residue motif to inhibit secondary structure formation, tyrosine was put at the beginning of the polypeptide, the N-terminus was acetylated, and the C-terminus was amidated in each case.

The polypeptides were named SN1, SP2, LN3, and LP4, meaning short negative, short positive, long negative, and long positive, respectively. Experiments were done by colleagues in the Haynie research group using pairs of the designed polypeptides, one negative and one positive polypeptide at a time. Multilayer films consisting of at least 5

bilayers of the peptides were deposited onto QCM resonators using standard LBL techniques (each bilayer consists of one layer of polycation and one layer of polyanion). Figure 3.5 shows resonator frequency versus adsorbed layer for different combinations of SP2, SN1, LP4, and LN3. Each combination involved one negative polypeptide and one positive polypeptide, as required by LBL. The linearity of the data is a likely indicator of relatively regular assembly of the polymer during adsorption and an approximately uniform density of the polypeptides in each adsorbed layer. Linear growth of deposited polypeptide mass indicates repeatability of adsorption steps early in the assembly process; frequently shift indicates the general success of the multilayer fabrication process, as this sensible quantity is proportional to mass deposited by the sauertray equation.

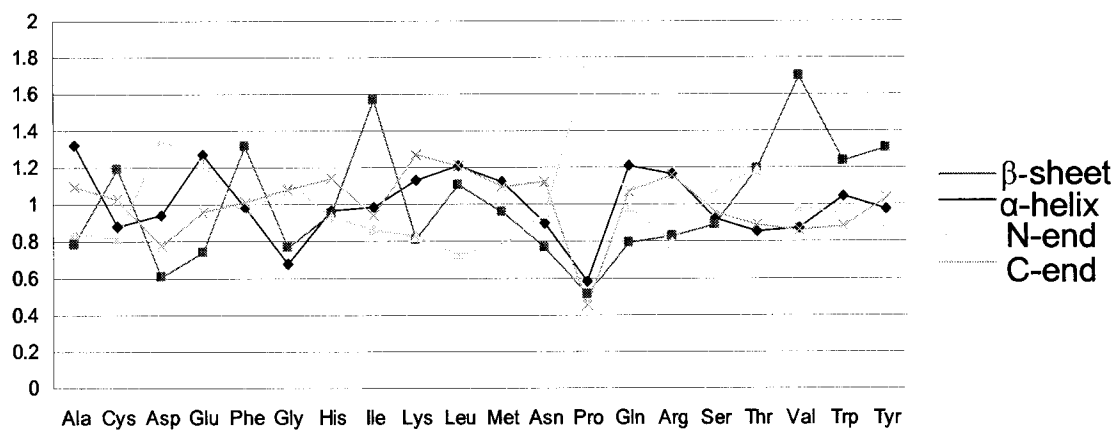
Table 3.4 Motifs information stored in the database

Motif information
Identification numbers
Motif sequences
Secondary structure prediction value
Hydrophobicity/hydrophilicity value
Surface accessibility value

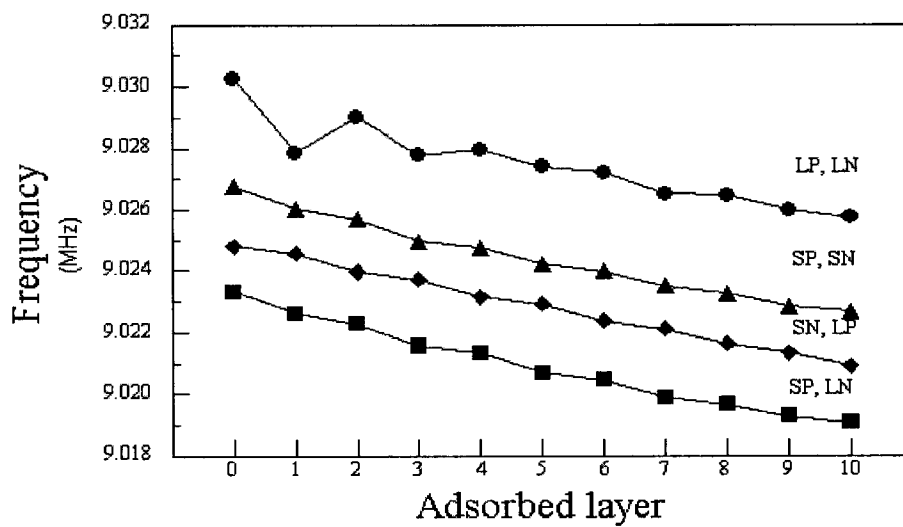
Table 3.5 Four amino acid sequences for experimental work

SEQ ID #1 (SN1)	YRRRRSVQGRRRRSVQGRRRRSVQGRRRRSVQ
SEQ ID #2 (SP2)	YEEDECQDGEDECQDGEDECQDGEDECQD
SEQ ID #3 (LN3)	YRRRRSVQGRRRRSVQGRRRRSVQGRRRRSVQGRRRRSVQ
SEQ ID #4 (LN4)	YEEDECQDGEDECQDGEDECQDGEDECQDGEDECQDGEDECQD





**Figure 3.4** 20 amino acids secondary propensity values.



**Figure 3.5** Adsorption of SN, SP, LN, and LP.

## CHAPTER 4

### MD SIMULATIONS OF Y(K)<sub>5</sub> and Y(E)<sub>5</sub> PEPTIDES

#### 4.1 Introduction

Poly-L-lysine and poly-L-glutamic acid have been used to fabricate multilayer thin films by LBL. Molecular structure in the film has been studied by various experimental tools, including neutron and X-ray reflectivity and X-ray photoelectron spectroscopy, CD [3 and 91], and Fourier transform infrared (FTIR) experiments [4]. The experimental results show that a (PLL-PLGA)<sub>n</sub> multilayer at neutral pH 7.4 contains ~35 %  $\beta$  sheet structure [3-5, and 91]. However, some important information, such as the detailed secondary structure at atomic level still cannot be obtained by experiment. So a computational approach is sought. Each adsorption step of polypeptide LBL usually requires minutes or longer to reach > 90 % completion [51]. It would be unrealistic, therefore, to attempt to simulate  $\beta$  sheet formation from randomly oriented polypeptides; at least before greater capacity for calculations becomes available.

MD simulation has been an important tool for studying protein structure for more than twenty years, and it can provide insights on the nature of PLL and PLGA LBL.

In this work we have adopted the approach of Nussinov and co-workers [73, 88] and built multiple peptide models which initially exhibit ideal  $\beta$  sheet geometry, based on experimental results. We have done MD simulations for each model to obtain atomic-

resolution information on the internal structure of polypeptide LBL films and a better understanding of the relationship between hydrophobic interactions, electrostatic interactions, and hydrogen bonds.

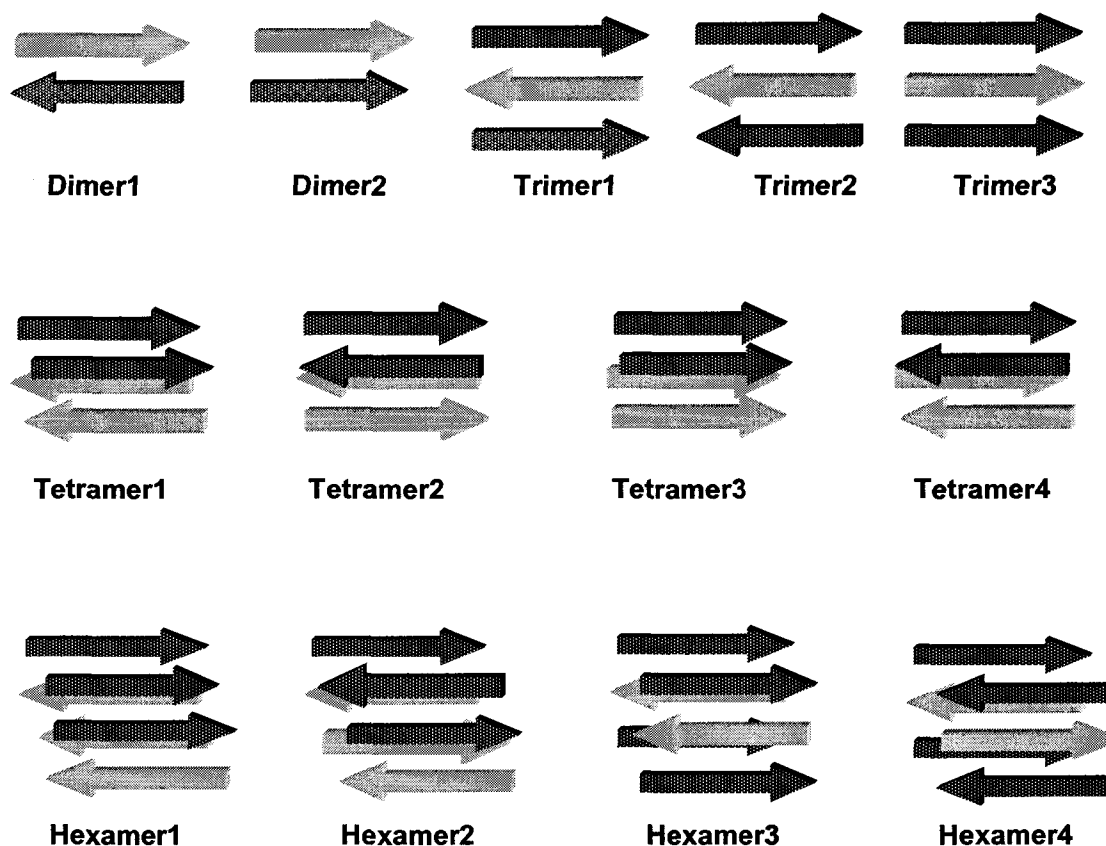
## 4.2 Methods

### 4.2.1 Simulation Models

The sequences of the two peptides for the PLL/PLGA simulations were: Tyr-(Lys)<sub>5</sub> and Tyr-(Glu)<sub>5</sub> where Try = tyrosine, Lys = lysine, and Glu = glutamic acid. A tyrosine appears at the N-terminus for spectroscopic detection of the material peptide; it was included for further comparison with experimental data. The models are displayed in Figure 4.1. Dimer 1 and Dimer 2 consist of one negative and one positive peptide at neutral pH in anti-parallel and parallel conformation. Trimer 1, Trimer 2, and Trimer 3 have two peptides of the same charge and one of opposite charge. The three peptides in Trimer 1 are in an anti-parallel  $\beta$  sheet conformation; in Trimer 3, all are parallel; and in Trimer 2, there is a mixture of anti-parallel and parallel. In the 4-peptide models, there are four combinations: Tetramer 1 and Tetramer 2 contain two anti-parallel  $\beta$ -sheets; but the sheets are parallel in one case (Tetramer 1) and anti-parallel in the other (Tetramer 2). Tetramer 3 and Tetramer 4 consist of two parallel  $\beta$  sheets, either parallel (Tetramer 3) or anti-parallel (Tetramer 4). There are four combinations of Hexamers. Hexamer 1 and Hexamer 2 are two-layer models. Hexamer 1 contains 3 anti-parallel  $\beta$ -sheets, each of which is parallel to the others. In Hexamer 2 the 3 anti-parallel  $\beta$ -sheets are anti-parallel to each other. Hexamer 3 contains 3 layers and two anti-parallel  $\beta$  sheets, each of which is anti-parallel to the others. Hexamer 4 is the same as Hexamer 3 except that the two

anti-parallel  $\beta$  sheets are anti-parallel to each other. In each case, the distance between two peptides in the same sheet is 4.7 Å and the distance between two sheets is 10 Å.

The peptides were built in fully extended conformation using the Biopolymer module integrated with insightII (Accelrys, USA). All simulations were done using CHARMM [63] version 29b1 running on a SGI Origin 2000 with a total of 32 CPUs, 10 TB memory, and 150 GB hard disk. Dimer and trimer required one week of CPU time to finish; for a tetramer, almost half of a month; and for a hexamer, more than twenty days.



**Figure 4.1** Schema of dimers, trimers, tetramers and hexamers. Black and gray colors represented negatively-charged Tyr(Glu)<sub>5</sub> and positively-charged Tyr(Lys)<sub>5</sub> at neutral pH, respectively.

#### 4.2.2 Simulation Details

CHARMM [63] and the all-atom charmm22 force field were used for the MD simulations. The peptides were solvated using TIP3P water molecules [67, 89]. Cubic periodic boundary conditions were applied in all simulations to eliminate boundary effects. Box size was calculated as the sum of the maximal size of a given peptide-system and the cutoff length of the forcefield. For dimer and trimer models, the box size was  $36 \times 36 \times 36 \text{ \AA}^3$ , for tetramer  $40 \times 40 \times 40 \text{ \AA}^3$ , and for hexamers  $46 \times 46 \times 46 \text{ \AA}^3$ .

The cutoff distance for nonbonded interactions was set to  $13.0 \text{ \AA}$ , and a neighbor list was built and updated when necessary using a heuristic test. A switch function was applied for both the electrostatic and Lennard-Jones (LJ) interactions between  $10$  and  $12 \text{ \AA}$  to smooth the change across the cutoff. The SHAKE algorithm [90] was used to constrain all hydrogen atom-heavy atom bond lengths. The pH value for peptides was set to  $7.4$ , where Tyr(glu)<sub>5</sub> is negatively charged and Tyr(Lys)<sub>5</sub> is positively charged.

Next, 200 steps of steepest-descent and 300 steps of Adopted Basis Newton-Raphson (ABNR) were done on initial models yielding a lower energy conformation. The leap-frog algorithm with a time step of  $1 \text{ fs}$  was used in the MD simulation under the conditions of constant number ( $N$ ), volume ( $V$ ), and temperature ( $T$ ). Prior to MD in each case, the whole system was heated from  $240$  to  $350 \text{ K}$  for  $10 \text{ ps}$  and equilibrated for  $10 \text{ ps}$ . Finally, a  $1 \text{ ns}$  simulation was done. The temperature was set higher than room temperature to sample a relatively large conformational space in a limited time period. Trajectory data were saved every  $1 \text{ ps}$  during the final stage. Analysis of trajectory data was done using Decipher (Accelrys, USA) and figures were prepared using PSI-plot.

### 4.2.3 Geometrical Parameters

Three geometrical criteria have been discussed [73] concerning structural changes of the peptide models. They were head-to-head the distance, tail-to-tail distance, and distance between the centers of mass (CM). The two average distances used to assess the structural changes of the peptide models are: intra-strand distance,  $\langle d_{str} \rangle$ , defined as the average of the above three basic distances between any two  $\beta$  strands which belong to the same sheet; inter-sheet distance,  $\langle d_{sh} \rangle$ , defined as for intra-strand distance except that the two  $\beta$  strands belong to different  $\beta$  sheets (see Figure 4.2).

### 4.2.4 Simulations of Hexamers in Neutral States

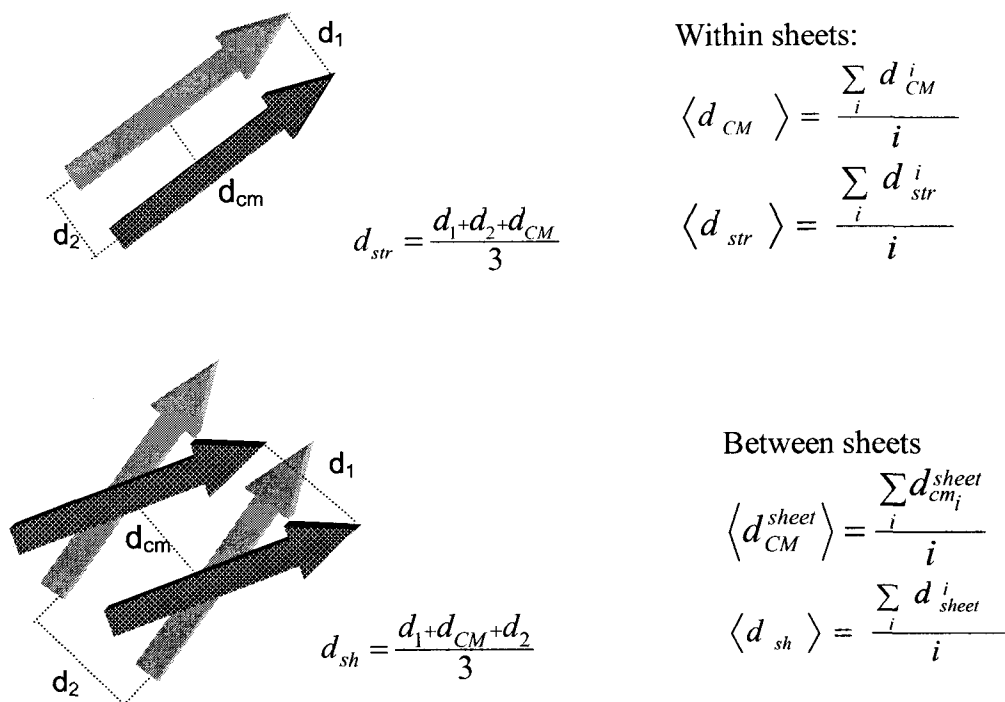
To explore the role of electrostatic interactions in peptides PLL and PLGA LBL films, we also carried out MD simulations for four hexamers with no charge.

## 4.3 Results

### 4.3.1 Dimers and Trimers

Molecular dynamics simulations have been done on two Dimers and three Trimers. Root mean square deviation (RMSD) is defined as the scalar distance between atoms of the same type for two structures. RMSD values of Dimer 1 with the anti-parallel orientation showed that Dimer 1 remained equilibrium during the simulation, and Dimer 2 with parallel orientation reached equilibrium after about 400 ps (Figure 4.5). For the three trimers, the RMSD value of Trimer 1 had a small increase during the first 100 ps, and then reached equilibrium after 300 ps. Trimer 3 behaved like Trimer 1. Trimer 2 had the largest RMSD change during the 1 ns simulation (Figure 4.8). Among these three trimers, only trimer 1 had a fully anti-parallel orientation between the three  $\beta$

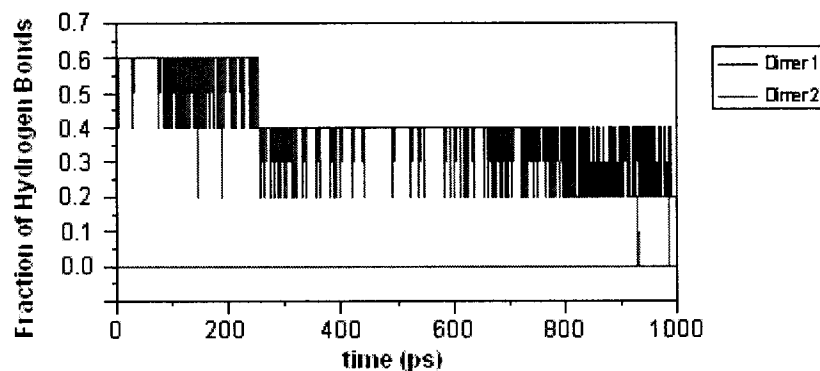
strands, whereas the other two trimers had either a parallel orientation (trimer3) or a mixture parallel and anti-parallel orientation (trimer2), which indicates that anti-parallel is stable than others.



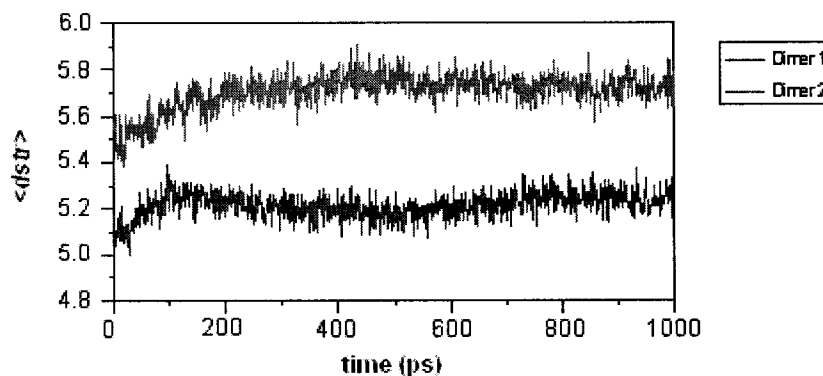
**Figure 4.2** The geometrical measures used to evaluate the structural changes.

The number of remaining hydrogen bonds in the dimer and trimer models is shown in Figure 4.3 and Figure 4.6, respectively. During the 1 ns simulation, Dimer 1 preserved about 40 % of its initial 5 hydrogen bonds over 700 ps, whereas almost none left for Dimer 2 (Figure 4.3). The average distances between the strands of Dimer 1 increased from 5.0 Å to 5.2 Å in the first 100 ps, and then remained at 5.2 Å for the remaining 900 ps. The average distance of Dimer 2 showed the similar tendency, but average distances were 0.5 Å higher than Dimer 1 (Figure 4.4).

Trimer 1 was able to preserve more than 50 % of its initial 10 hydrogen bonds, whereas Trimer 2 and Trimer 3 quickly lost most of theirs (Figure 4.6). Comparing with the number of hydrogen bonds of dimers, only Dimer 1 and Trimer 1 had maintained about half of their original number of hydrogen bonds during the 1 ns simulation. The average distance of Trimer 1 decreased quickly at the beginning of the simulation but was at equilibrium for the rest of the simulation; the fluctuation of Trimer 1 was only about 0.1 Å after 500 ps. The average distance of Trimer 2 and Trimer 3 had similar distance changes as Trimer 1 (Figure 4.7).

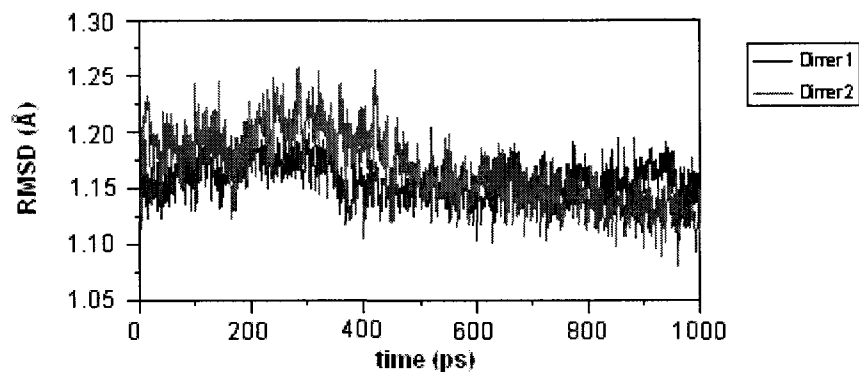


**Figure 4.3** The fraction of the number of hydrogen bonds of dimers throughout the simulation.

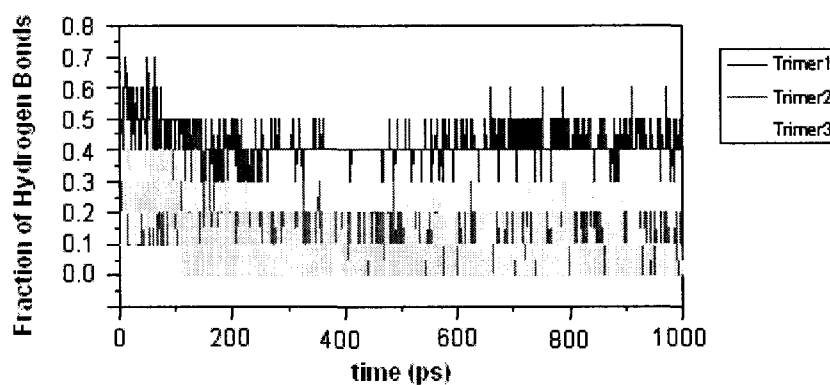


**Figure 4.4** The average distance of dimers as a function of time.

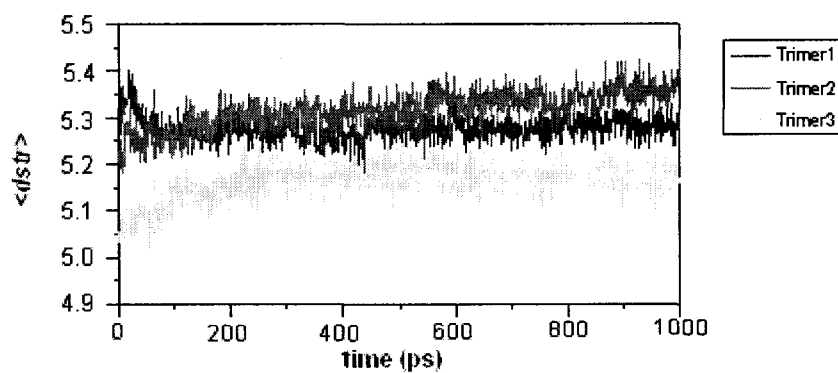




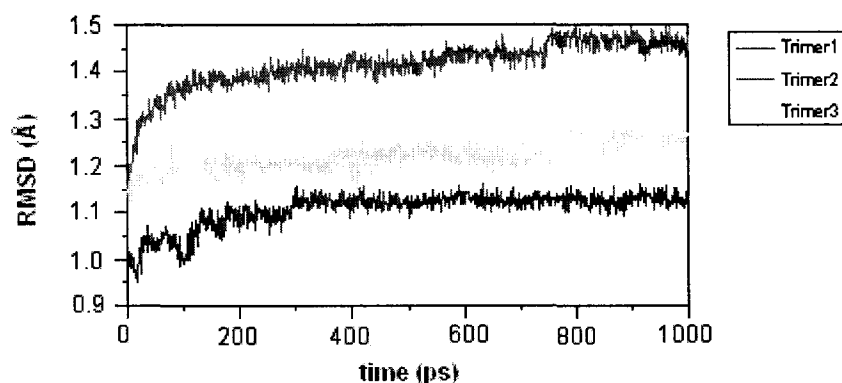
**Figure 4.5** RMSD of dimers relative to the starting structure.



**Figure 4.6** The fraction of the number of hydrogen bonds of trimers throughout the simulation.



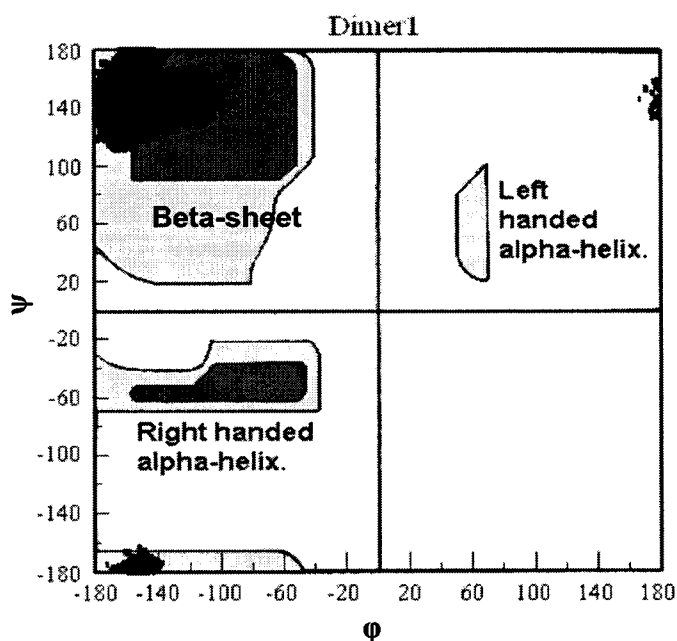
**Figure 4.7** The average distance of trimers as a function of time.



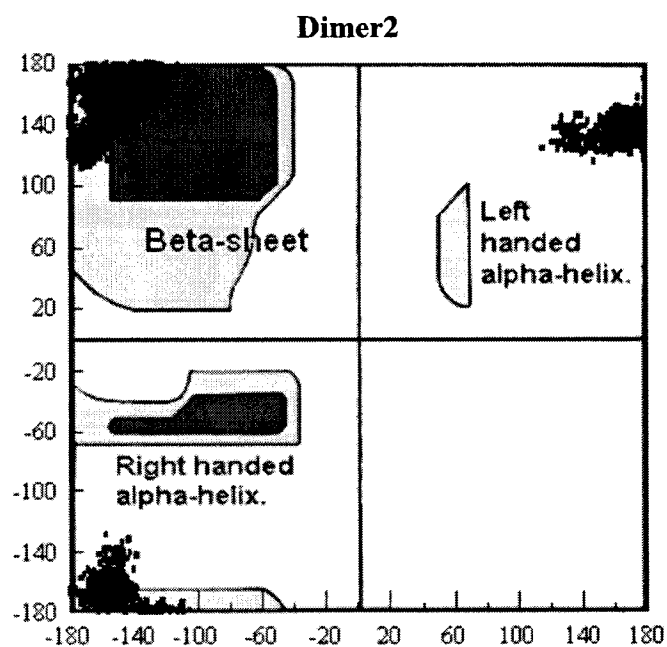
**Figure 4.8** RMSD of trimers relative to the starting structure.

All of the MD simulations began with the peptides a fully extended conformation, in which the  $\phi$  and  $\psi$  angles for each amino acid residue were set at  $180^\circ$ . The Ramachandran plot can be used to study the relationship between backbone torsion angles and type of secondary structure. The bond angles  $\phi$  and  $\psi$  at each residue for an  $\alpha$  helix or a  $\beta$  sheet fall within a relatively restricted range of sterically allowed structures. The  $\phi$ ,  $\psi$  angles for a  $\beta$  sheet are around  $(-90^\circ, 150^\circ)$  [88]. Figures 4.9 and 4.10 show a Ramachandran plot for residues 2 to 5 of Dimer 1 and Dimer 2, respectively. During the 1 ns simulations, the majority of dihedral angles of Dimer 1 remained in a  $\beta$  sheet range, while many of the dihedral angles of Dimer 2 gradually fell out of the  $\beta$  sheet regions; there is no indication of  $\alpha$  helix structure. So far, we have evaluated the stability of a single  $\beta$  sheet in PLL and PLGA films based on several properties: the number of remaining hydrogen bonds during the simulation which is an essential indicator of the stability of a  $\beta$  sheet, the average intra-strands and inter-sheet distances which show the structural integrity of dimer and trimer models, RMSD value which is the overall measurement of the differences between any two structures, and Ramachandran plot. Taken these information together, it showed that Dimer 1 and Trimer 1 are the stable

models for a single sheet and indicated that anti-parallel orientation is the stable  $\beta$  sheet conformation.



**Figure 4.9** Ramachandran plot for residues 2 to 5 of Dimer 1. The background picture was from [www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html](http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html).



**Figure 4.10** Ramachandran plot for residues 2 to 5 of Dimer 2. The background Ramachandran plot was from [www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html](http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html).

### 4.3.2 Tetramers

Four molecular dynamics simulations have been done for the four tetramer models, respectively. The RMSD values of tetramers are shown in Figure 4.11. Tetramer 2 (two  $\beta$ -strands are anti-parallel within each sheet and two sheets are anti-parallel) had the least RMSD changes and Tetramer 3 (two  $\beta$ -strands are parallel within sheets and two sheets are parallel) had the most RMSD changes. The numbers of remaining hydrogen bonds during the 1 ns simulation of tetramers were quite different. The number of hydrogen bonds in the initial structures of tetramers is 10. During the 1 ns simulation, Tetramer 2 was able to keep half of the original hydrogen bonds; the other three tetramers quickly lost the majority of hydrogen bonds (Figure 4.12). The intra- and inter-average distances of tetramers were shown in Figures 4.13 and 4.14.

### 4.3.3 Hexamers

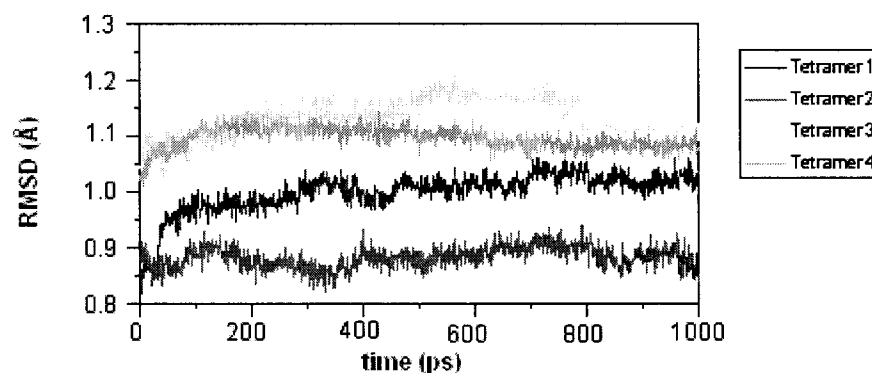
The RMSD results of these four hexamers are displayed in Figure 4.16. Among the four hexamers, Hexamer 2 had the largest fluctuation compared with its initial structure. The RMSD value of Hexamer 2 gradually increased to 1.2 Å during the first 500 ps, then jumped up and down between 1.1 Å to 1.25 Å for about 200 ps, and finally returned to the same level of Hexamer 1 and Hexamer 4. There was a little increase of Hexamer 3 at the first 100 ps, and then it remained in fluctuation as less as 0.05 Å for the rest of simulation. Hexamer 1 and Hexamer 4 were able to keep equilibrium state also although there was a very slow increase.

The remaining numbers of hydrogen bonds of hexamers as a function of time are plotted in Figure 4.15. Hexamer 3 maintained 60~65% of its hydrogen bonds, Hexamer 1 kept about 50~55% of its hydrogen bonds, Hexamer 2 kept 50~60% of its hydrogen

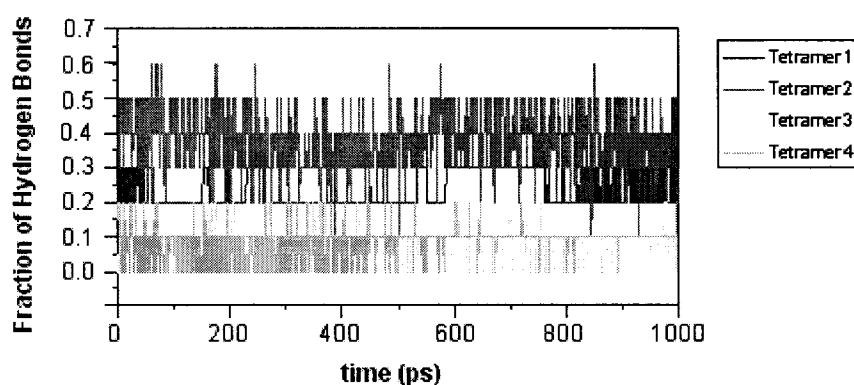
bonds for about 600 ps, then the percentage of it dropped to 50%, and Hexamer 4 maintained about 40~50% of its hydrogen bonds during the simulation. The average intra- and inter-distances of hexamers are shown in Figures 4.17 and 4.18.

#### 4.3.4. Comparisons of Hexamers in Charged and Neutral States

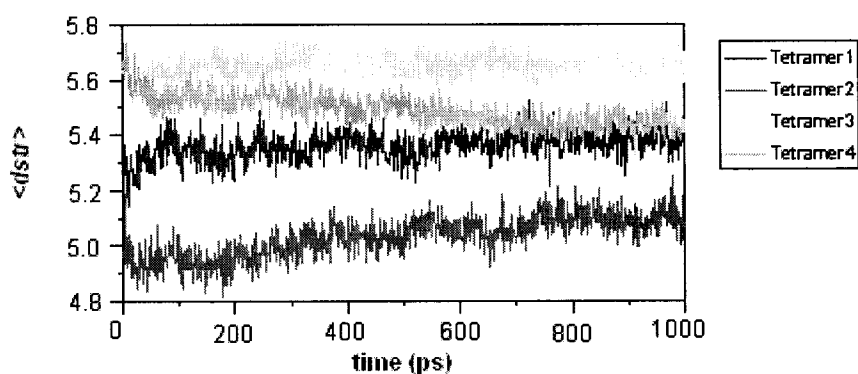
Four MD simulations have been done for the Hexamers without charges. The comparison of the remaining number of hydrogen bonds and the geometrical changes between Hexamers in a charged state and in a neutral state are shown in Figures 4.19, 4.20, and 4.21, respectively.



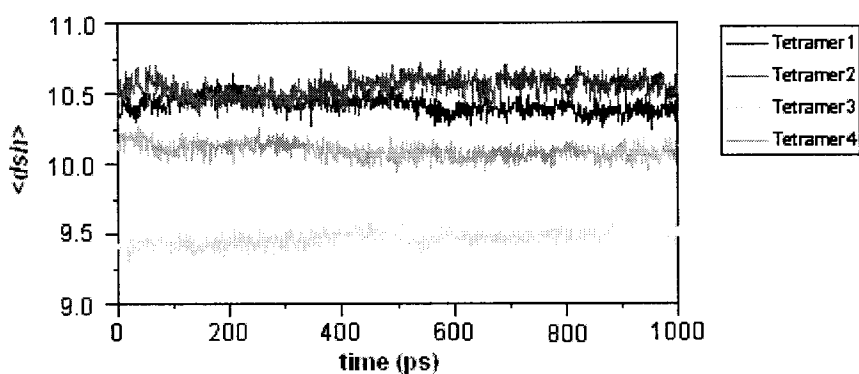
**Figure 4.11** RMSD values of tetramers.



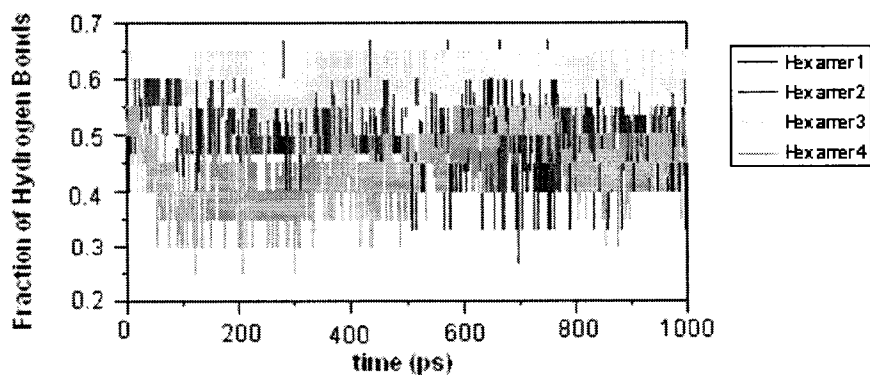
**Figure 4.12** The fraction of hydrogen bonds of tetramers as a function of time.



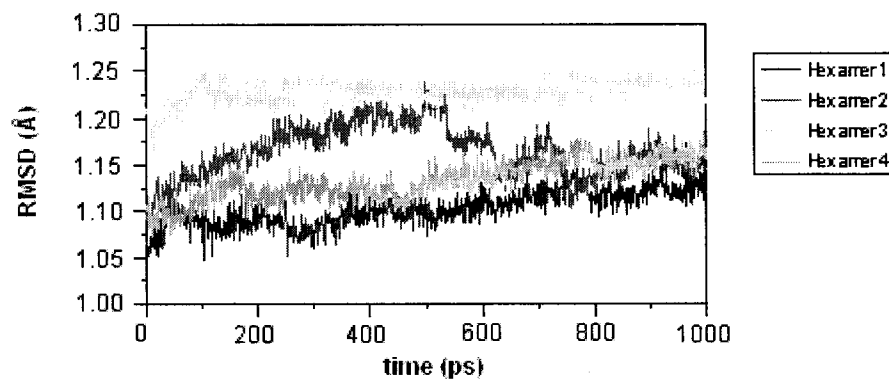
**Figure 4.13.** The average distance within the sheets of tetramers as a function of time.



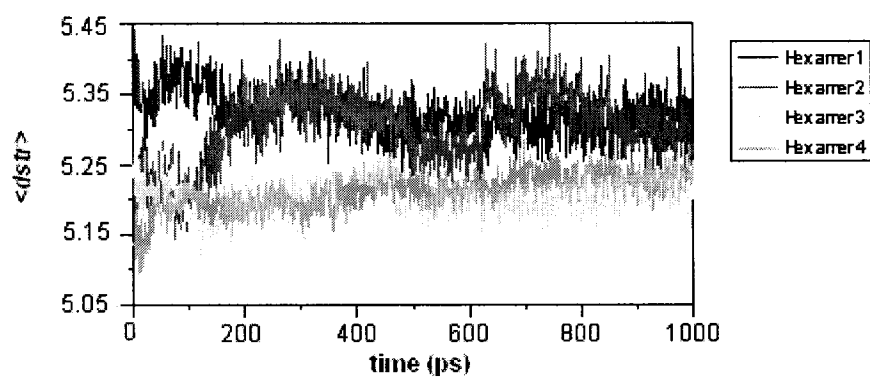
**Figure 4.14** The average distances between sheets of tetramers as a function of time.



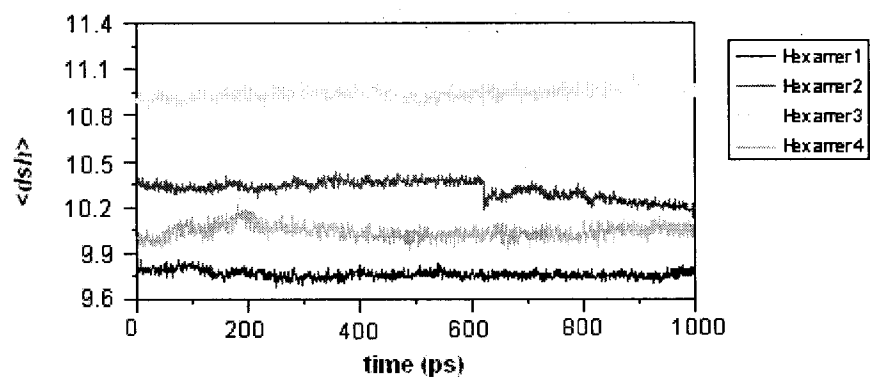
**Figure 4.15** The fraction of hydrogen bonds of hexamers as a function of time.



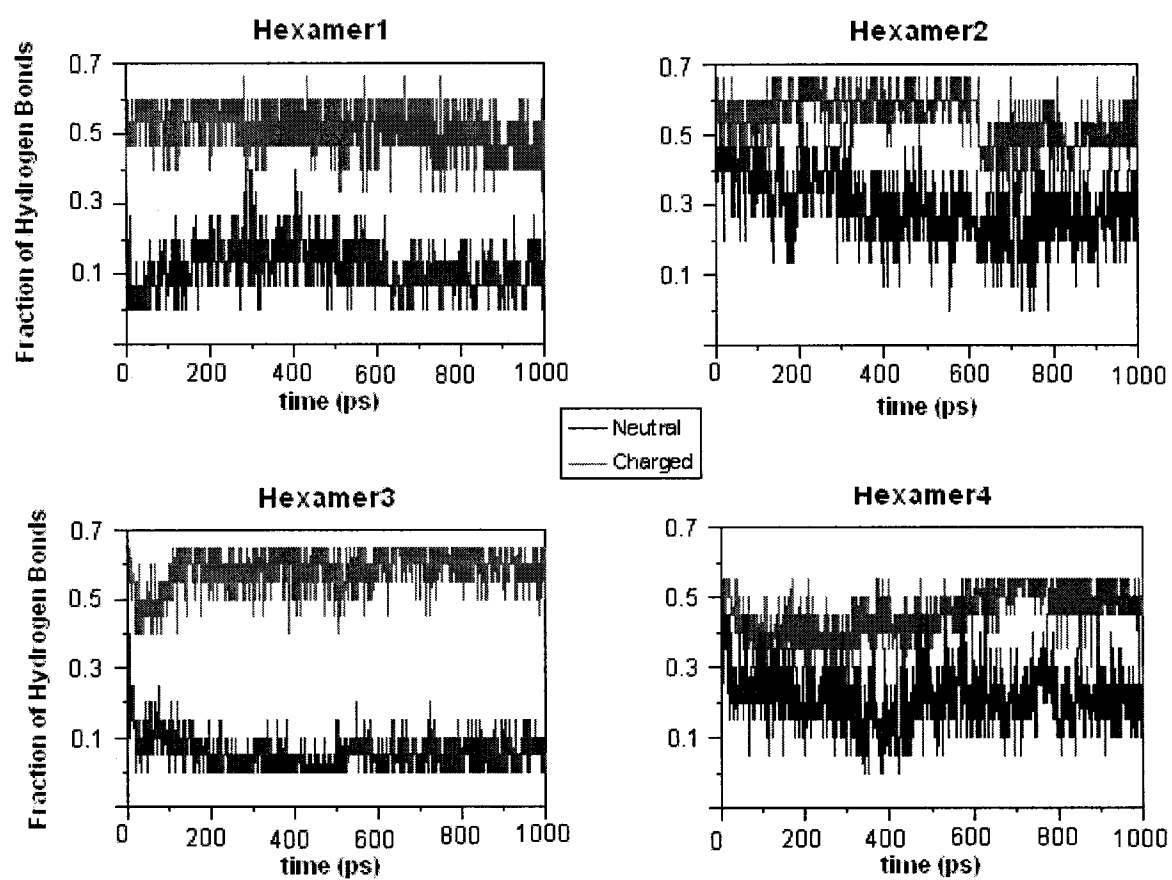
**Figure 4.16** RMSD values of hexamers as a function of time.



**Figure 4.17** Average distances within sheets of hexamers as a function of time.

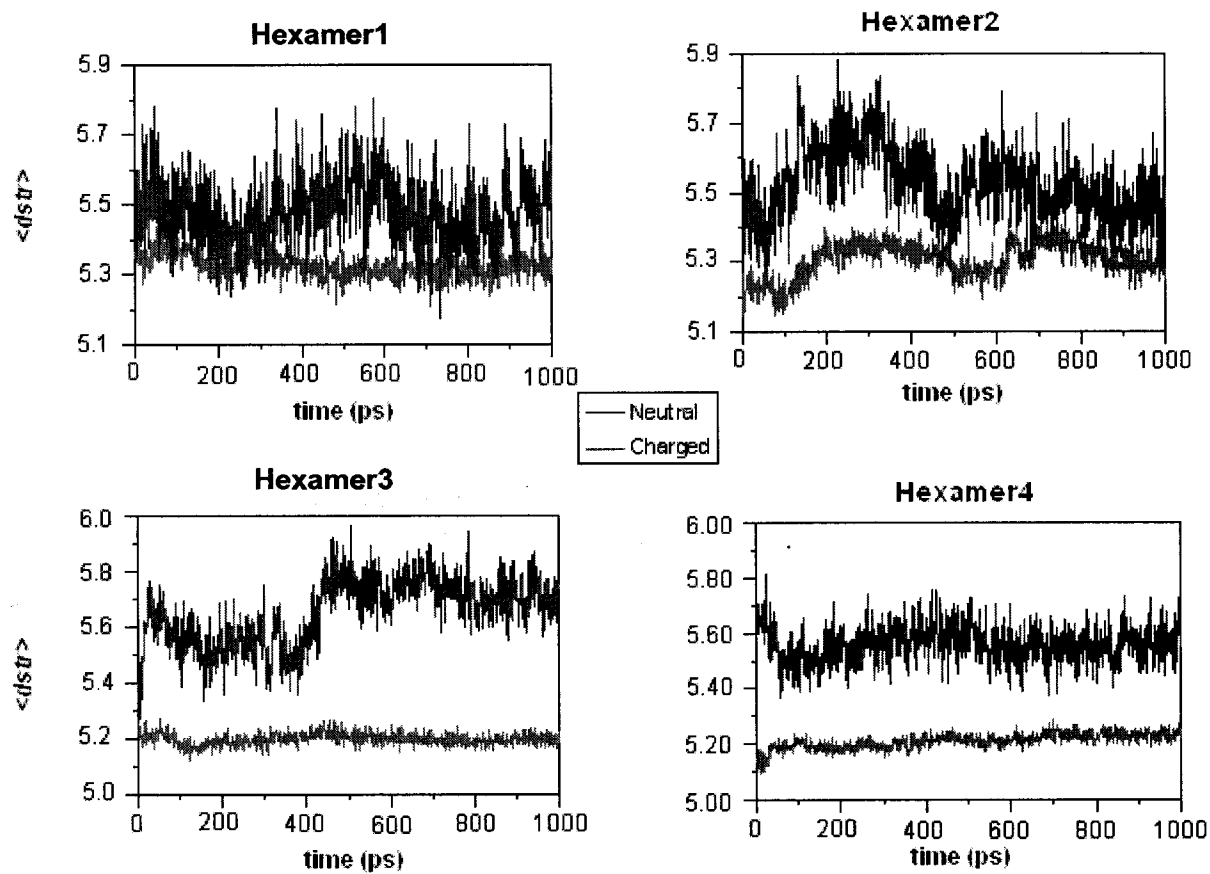


**Figure 4.18** Average distances between  $\beta$  sheets of hexamers as a function of time.

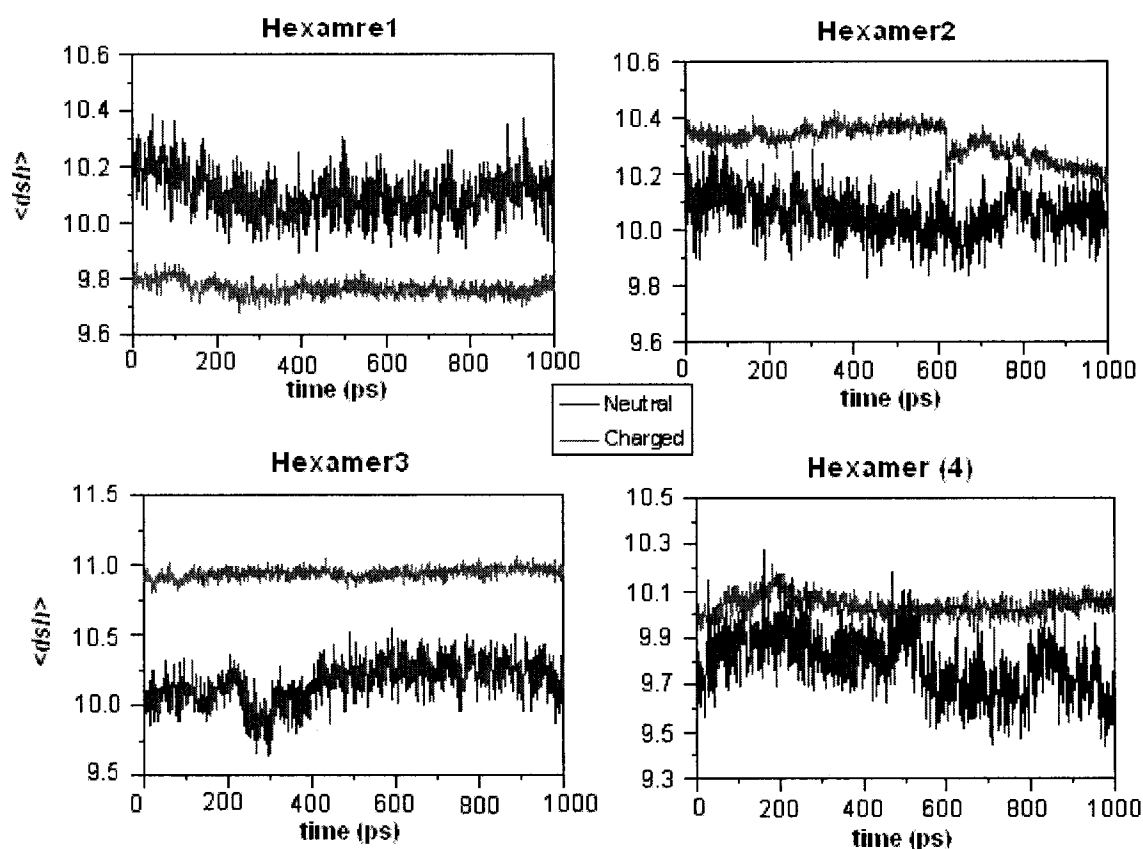


**Figure 4.19** Comparison of remaining hydrogen bonds of hexamers in different charge statuses. Black and pink stands for the non-charged and charged peptides, respectively.





**Figure 4.20** Comparison of intra-distances of hexamers as a function of time. Black stands for the non-charged peptides and pink stands for charged peptides.



**Figure 4.21** Comparison of inter-distances of hexamers as a function of time. Black stands for the non-charged peptides and pink stands for charged peptides.

#### 4.4 Discussions

The peptides PLL/PLGA have been used in LBL due to the great potential applications in a broad range of biotechnology related areas. The amount of adsorbed polymer in LBL and layer structure depends upon the charge density of the polymer, the sign and the density of the surface charge [5]. Among many natural and artificial peptides, the kind of peptides is of special interest due to the potential formation of the salt bridge. They not only share some common features of uncharged peptides, but they also have charge properties that exhibit various forms dependent on different pH and, hence, to be able to allow researchers to control the assembly process. The nanoscaled

multilayer thin films have some distinguished properties. Forming secondary structure property is one of them. Glutamic acid has the second highest  $\alpha$  helix propensity value (1.27, [52]) of the 20 usual amino acid types and lysine also has a high  $\alpha$  helix propensity value (1.13, [52]); however, PLL and PLGA multilayer thin films did not contain a significant amount of  $\alpha$  helix at neutral pH. The reason is because the negatively charged groups repel each other so strongly that they overcome the stabilizing influence of hydrogen bonds on an  $\alpha$  helix. For the same reason, PLL will not form a stable helical structure at neutral pH. This agrees with the experimental work which has shown that individual PLL and PLGA exhibited random coil structure at neutral pH [1, 2, 91], however, their mixture in solution and in the thin films made up contained a large amount of  $\beta$  sheets.

$\beta$  sheet structure and  $\alpha$  helix are two common secondary structures. It is general accept that  $\alpha$ -helix is stabilized by forming a hydrogen bond between  $i$  and  $i+4$  residues along the same polypeptide chain,  $\beta$  sheet by intermolecular hydrogen bonds or hydrogen bonds formed from different parts of a long peptide.

Multiple peptide models have been set up and the structures are mainly parallel or anti-parallel. There is no anti-parallel and parallel mixture in the same sheet in hexamers since this type of structure is the least possible structure in reality [94].

To assess the stability of each peptide model, we mainly monitored three properties: the remaining hydrogen bonds, geometrical parameters, and RMSD with respect to the initial structure. The RMSD values vary in different protein parts. A large RMSD value usually suggests a mobile part of a protein or one lacking secondary structures. When comparison was limited to one  $\beta$  sheet with two  $\beta$ -strands, Dimer 1

(anti-parallel structure between one negatively-charged peptide and one positively-charged peptide) showed greater stability than Dimer 2. Similarly, Trimer 1 was more stable than Trimer 2 or Trimer 3, as it had the greatest number of hydrogen bonds and the smallest RMSD value. When there was only one sheet, the anti-parallel orientation was the most stable one and can therefore be considered the thermodynamically favored arrangement for  $\beta$  strands.

As to tetramers, only Tetramer 2 (anti-parallel between sheets and within sheets) retained half of the original hydrogen bonds; the others lost the majority of their original hydrogen bonds during the simulation (Figure 4.12). RMSD calculations show that Tetramer 2 had the minimal RMSD value (Figure 4.11). In some cases, though, the model with the highest number of hydrogen bonds did not have the lowest RMSD value. Tetramer 2 thus can be considered relatively stable. Other tetramers, although they could not retain hydrogen bonds, still they maintained a degree of structural integrity as seen by the small differences between intra- and inter-sheet distances (Figures 4.13 and 4.14).

There are two groups for hexamers. The first includes Hexamer 1 and Hexamer 2. They were formed by adding another anti-parallel  $\beta$  sheet to Tetramer 1 and Tetramer 2, resulting in an entire sheet being buried. This arrangement allows study of the stability of a  $\beta$  sheet induced by sheet organization. As shown in Figures 4.12 and 4.15, Hexamer 1 had the most dramatic change in terms of number of hydrogen bonds. Hexamer 1 preserved about 50 % of the original hydrogen bonds, whereas Tetramer 1 had just a few of its hydrogen bonds for most of the simulation time. Hexamer 2 also had about an average of 10 % more hydrogen bonds than Tetramer 2. Considering the much stronger hydrophobic interaction introduced by the entire sheet buried inside the peptide system,

these results were not surprising. Average distance calculations show that the four hexamers were able to keep the structural integrity (Figures 4.17 and 4.18).

Two peptide models were used to test the significance of hydrogen bonding: Hexamer 3 and Hexamer 4. The role of hydrogen bonds in protein or polypeptide folding is still unclear. It is said that hydrogen bonding plays a key role in stabilizing a protein's secondary structure. Some recent studies, however, have pointed out that hydrogen bonds are the consequence rather than the reason for protein folding and aggregation [73]. In our second group, Hexamer 3 and Hexamer 4 had five more hydrogen bonds than Hexamer 1 and Hexamer 2. If the first saying is right, Hexamer 3 and Hexamer 4 should be more stable than Hexamer 1 and Hexamer 2. In the 1 ns MD simulations, Hexamer 3 retained about 65% of its original hydrogen bonds and the Hexamer 4 ~40~50%. During the first 600 ps, Hexamer 3 and Hexamer 2 had the same number of hydrogen bonds; Hexamer 4 had the least. After 600 ps, Hexamer 4's hydrogen bond number increased but only at the level of Hexamer 1 and Hexamer 2. With regard to intra-sheet distances, Hexamer 3 and Hexamer 4 were apparently at equilibrium or in a kinetically-Trapped state throughout the simulation. Hexamer 1 and Hexamer 2 fluctuated for the first 500 ps and then gradually came to equilibrium after another 400 ps (Figure 4.17). There was no significant difference in intra-distances for the four hexamers. The inter-sheet distances changes behave the similarly (Figure 4.18). Taken together, the simulations suggest that hydrogen bonding is not a key factor for stabilizing the  $\beta$  sheet structure of PLL-PLGA-based films.

Some important conclusions can be drawn at this point: hydrophobic interactions are undoubtedly essential for peptide LBL; hydrogen bonding plays less of a role in

stabilizing the secondary structure than one might guess. Furthermore, the simulations also suggest that as the number of peptides increase, the stability of the structure they form increases. Also the most stable  $\beta$  structure in a PLL and PLGA multilayer thin film at neutral pH will be one with anti-parallel strands within the sheet.

We had studied the influence of hydrophobic interaction, hydrogen bonds, and the origin of stability of suprastructure formation in PLL-PLGA multilayer thin films. To study the influence of electrostatic interactions involving charged amino acid residues at neutral pH, our approach is simply to use peptides carrying no charge and thus avoid altering the local geometry of peptides. Simulations have been done for hexamer models with no charge. The simulation protocols were the same as for the charged peptides. As shown in Figure 4.19, when hexamers were charged, a greater number of hydrogen bonds remained intact during the simulation than hexamers lacking charge, and intra-sheet distance of charged hexamers in charged status were less than neutral ones, as shown in Figure 4.20. This suggests that electrostatic interactions are important for stabilizing  $\beta$ -sheet structure in PLL and PLGA multilayer thin films. Other researchers have reached a similar conclusion. For example, the MD simulations of Klimov and Thirumalai found that trimers disassembled when the charged amino acids were replaced by non-charged amino acids [92], and Ma and Nassinov have reported that neutral tetramers were the least stable ones [73]. On the other hand, the average distance between sheets in uncharged hexamers was less than the charged counterparts except for the Hexamer 1 (Figure 4.21). A possible explanation is that due to the lack of electrostatic interactions, the peptides can move more freely in response to thermal fluctuations, destabilizing sheet structures; the hydrophobic interaction, however, remain, and these help to hold peptides

together even in the absence of regular structure. This is possible because of hydrophobic interactions are relatively non-specific. By contrast, hydrogen bond formation requires not only specific chemical groups which can serve as donor or acceptor, but also a sufficiently favorable geometrical orientation of donor and acceptor.

In summary, we have studied poly-Lys and poly-Glu for LBL at neutral pH. We have found that hydrophobic interactions are essential for the stability of PLL and PLGA films, even at neutral pH; we know, however, that hydrophobic interactions are non-specific, a sort of 'glue', which sticks peptides together but does not necessarily stabilize a specific type of secondary structures. This agrees with the experimental work: self-assembly is possible at any pH. Electrostatic interactions too are essential for stabilizing  $\beta$  sheet structure in PLL and PLGA films at neutral pH. When PLL and PLGA are fully charged and mixed together, the electrostatic interactions not only reduce the stability of  $\beta$  sheet, but also stabilize them by the interactions between opposite charged groups. We also found that hydrogen bonding is more likely the consequence of hydrophobic interaction and electrostatic interaction than the major factor to stabilize secondary structure. However, once the hydrogen bond formed, it will help to stabilize an  $\alpha$  helix or  $\beta$  sheet. The simulations also suggest that peptides of six residues could form a stable film. This is consistent with the empirical finding that small molecules and ions can be useful for reversing the surface charge of a film and enabling its layer-by-layer assembly.

## CHAPTER 5

### MD SIMULATIONS OF CYSTEINE-CONTAINING

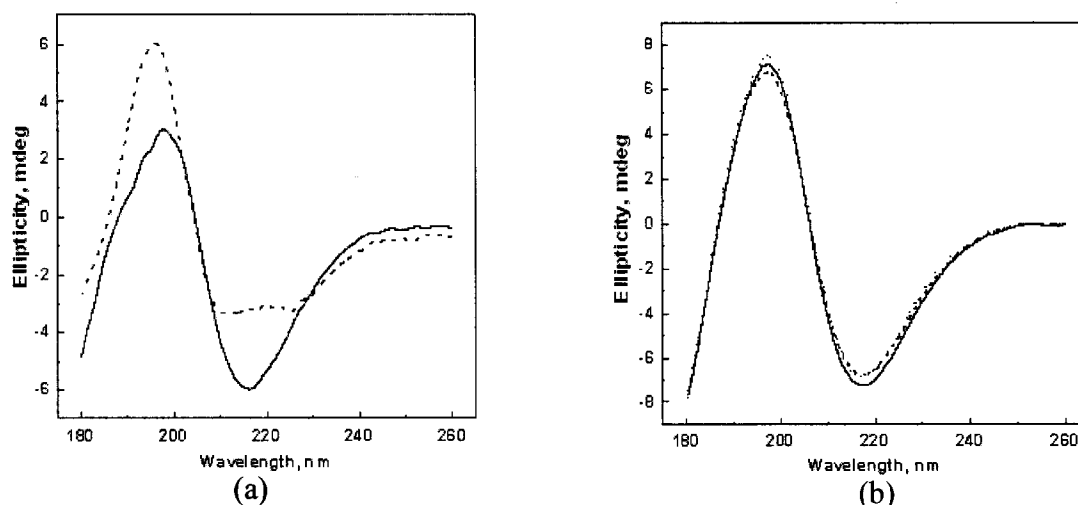
#### PEPTIDES

##### 5.1 Introduction

Polyelectrolyte multilayer thin films have attracted considerable interest for potential applications ranging from polymer electronics to biomaterials. PLL and PLGA have been used in LBL in the past several years and shown to form thin films, coating, and microcapsule [3, 4, 53, and 91]. These polypeptides involve just two natural amino acids: lysine and glutamic acid. Considering the 20 natural amino acids, the number of possible polypeptide sequences is astronomical. There thus is a great range of tremendous possibilities for raw materials for LBL films. Several attempts have been made to achieve this goal. There are generally two ways to stabilize LBL thin films: choose polyelectrolytes of high inherent structural integrity or form cross-links between polyelectrolytes [35]. When biocompatibility is a major concern, the reversible formation of cross-linked of peptides is ideal [34, 35]. Cysteine-containing peptides were first designed in Haynie's research group based on the computational approach described in Chapter 3 [52]. The two peptides each contained 32 residues; one peptide consisted of alternate positively-charged and hydrophobic residues, and another one consisted of



alternate negatively-charged and hydrophobic residues. The key amino acid in these peptides was cysteine (C), which increased the stability of the film by forming S-S bonds between different peptides sequences. Experiments have shown that the multilayer thin films constructed from the two cysteine-containing peptides contain a high amount of  $\beta$  sheets [3, 4] (Figure 5.1). Experimental data, however, do not provide a detailed view of structure. What is the internal  $\beta$  sheet structure? Is there any difference between PLL and PLGA and the designed peptides from the point view of MD simulation? As mentioned in Chapter 4, the self-assembly of  $\beta$ -sheet peptides procedure occurs on a time scale longer than second. A simulation starting from scattered monomers therefore would be too computationally expensive to implement. Here, we followed the approach of Chapter 4 and constructed the final  $\beta$  sheet structure and tested its stability. The simulations were run in explicit solvent, using 8-mers instead of the 32-mers of the experimental work in order to make the simulations doable on a reasonable timescale. The two cysteine-containing 8-mers were: ECEVEVEG, abbreviated by CEV, and KCKVKVKG, abbreviated by CKV. The MD simulation approach did not permit monitoring of S-S bond formation and breakage. In the previous chapter, MD simulations were applied for peptides YE<sub>5</sub> and YK<sub>5</sub>. The results suggested that hexamers of Tyr(glu)<sub>5</sub> and Tyr(Lys)<sub>5</sub> were able to form a stable structure. The hydrophobic interaction plays a key role in stabilizing the supramolecular assembly. At neutral pH, electrostatic force plays an important role in stabilizing  $\beta$  sheet structures in the PLL and PLGA multilayer thin films. The computational results agreed with the experimental observation [4].

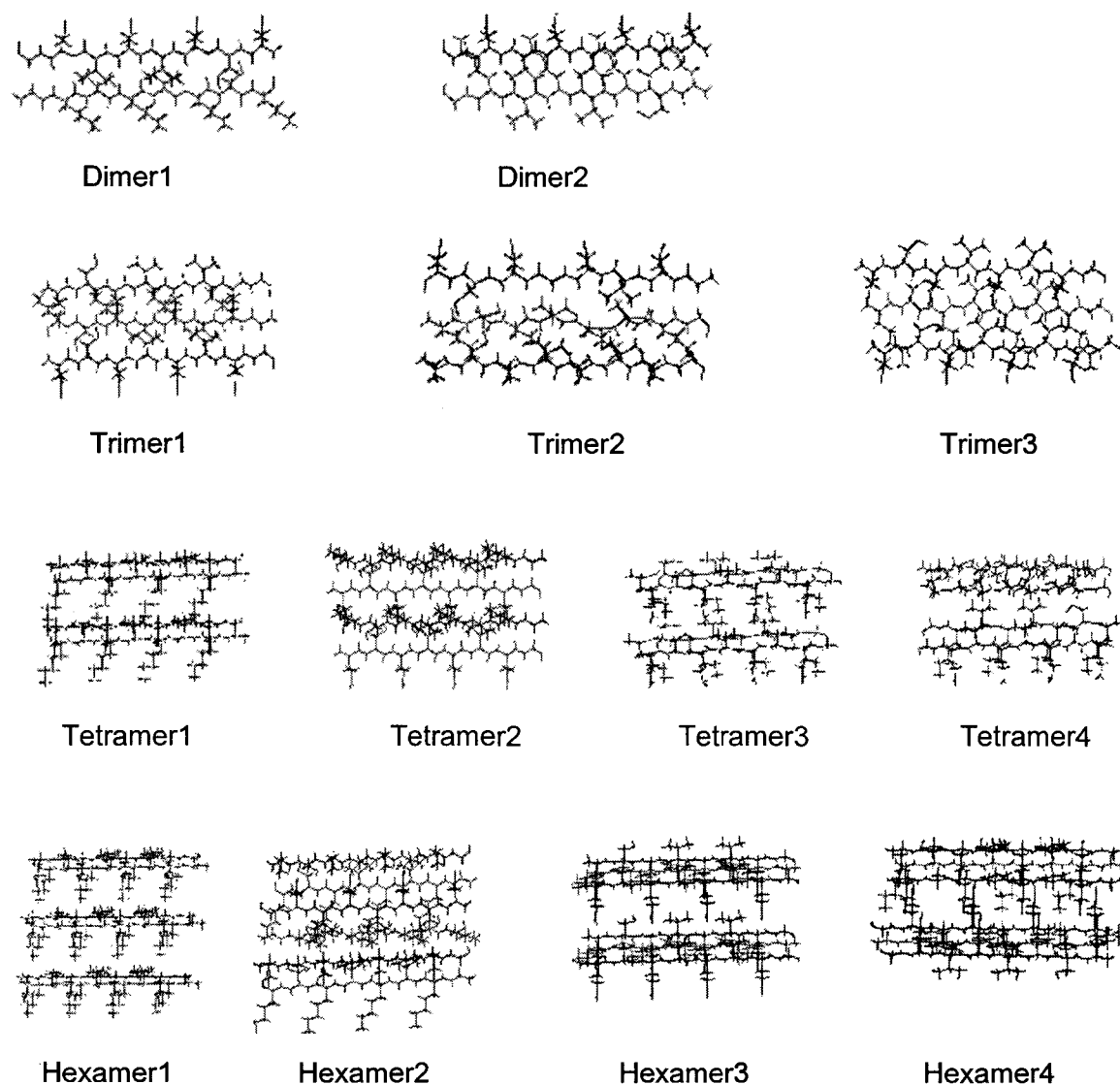


**Figure 5.1** CD of multilayers of Polypeptides neutral pH. (a) PLL-PLGA (b) Designed positive and negative polypeptides (3).

## 5.2 Methods

### 5.2.1 Simulation Models

The sequences of the two peptides are: Lys-Cys-Lys-Val-Lys-Val-Lys-Gly and Glu-Cys-Glu-Val-Glu-Val-Glu-Gly. Thirteen models have been studied as representatives of different combinations of the two peptides CEV and CEK. The models contain 2 to 6 peptides. There are two Dimers, three trimers, four tetramers, and four hexamers as shown in Figure 5.2. In each case, the distance between two peptides in the same sheet is about  $4.7 \text{ \AA}$  and the distance between two sheets is about  $10 \text{ \AA}$ , just as for poly-lys and poly-glu. The peptides were built in fully extended conformation using Biopolymer module integrated with InsightII. All simulations were performed by CHARMM running on the SGI origin 2000, as the simulations done in the Chapter 4.



**Figure 5.2** Atomic simulation models. Black and red colors represented negatively- and positively-charged peptide sequences at neutral pH, respectively.

### 5.2.2 Simulation Details

The peptides were solvated in TIP3P water molecules. CHARMM and the all-atom charmm22 force field were used for the MD simulations. Cubic periodic boundary conditions were applied in all simulations to eliminate the boundary effect. Box size was calculated as the sum of the maximal size of a given peptide-system and the cutoff length

of the forcefield. For the dimers and trimers the box size was  $46 \times 36 \times 36 \text{ \AA}^3$ , for tetramers it was  $46 \times 40 \times 40 \text{ \AA}^3$ , and for hexamers it was  $46 \times 46 \times 46 \text{ \AA}^3$ . The other conditions were the same as described in Chapter 4 section 4.2.2.

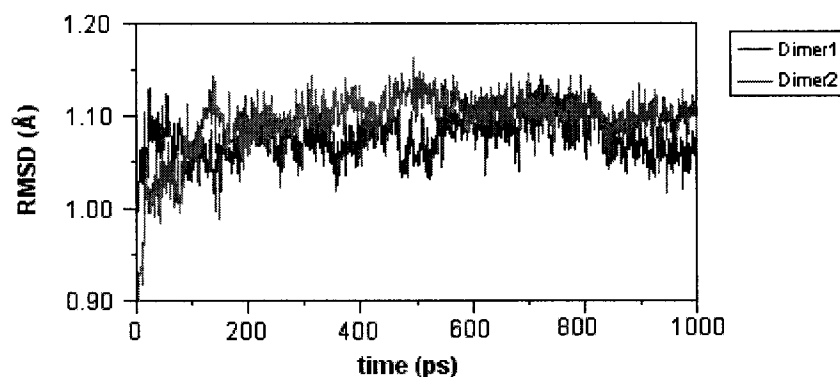
## 5.3 Results

### 5.3.1 Dimers and Trimers

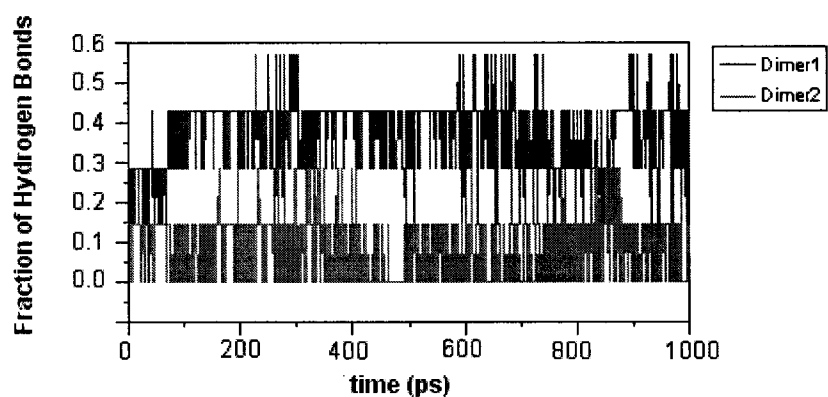
Dimers and trimers are the relatively simple models because they contained only one sheet. Simulations have been done for two Dimers and three trimers. Figure 5.3 shows that during the 1 ns simulation the RMSD for Dimer 1 fluctuated between 1 and 1.1  $\text{\AA}$  and Dimer 2 reached equilibrium at 200 ps and reached an equilibrium status at an average RMSD value about 1.1  $\text{\AA}$ . As to the number of hydrogen bonds, Dimer 1 retained 40-50% of the original ones whereas Dimer 2 lost its structural integrity early on (Figure 5.4). The average distance of strands is shown in Figure 5.5. Although Dimer 2 showed larger fluctuations of distances than Dimer 1, neither had a dramatic average distance change.

The simulations with a single, three-stranded sheet showed similar tendencies. Trimers RMSD values show that the irregularities and the conformational distortions have been gradually disappeared in the first 200 ps and finally reached equilibrium at 1.1  $\text{\AA}$  (Figure 5.6). Trimer 1, the full anti-parallel arrangement, retained ~40-50 % of its initial hydrogen bonds for 1 ns, whereas Trimer 2, with combinations of parallel and anti-parallel orientations, kept only 20-30% of its hydrogen bonds for half the simulation time, and Trimer 3 (full parallel  $\beta$  strands) quickly lost the interactions between the three strands (Figure 5.7). The average distances of trimers are displayed in Figure 5.8. Like Dimers, trimers showed no dramatic change, excluding the small jump at the beginning

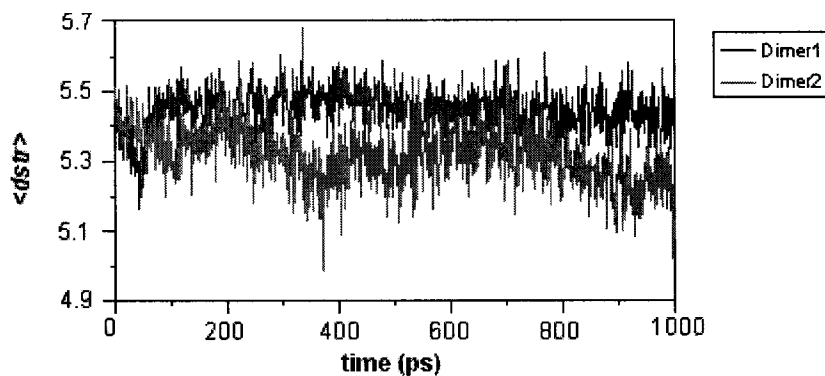
of the Trimer 3 simulation. For a single layer  $\beta$  sheet, the anti-parallel orientation between strands is the most stable conformation.



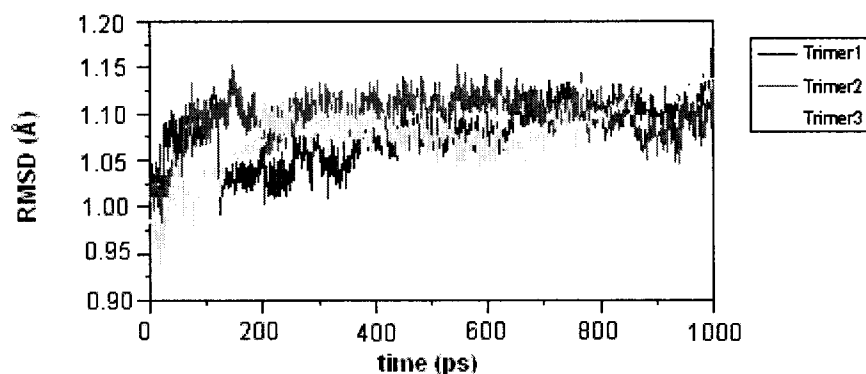
**Figure 5.3** RMSD values of dimers.



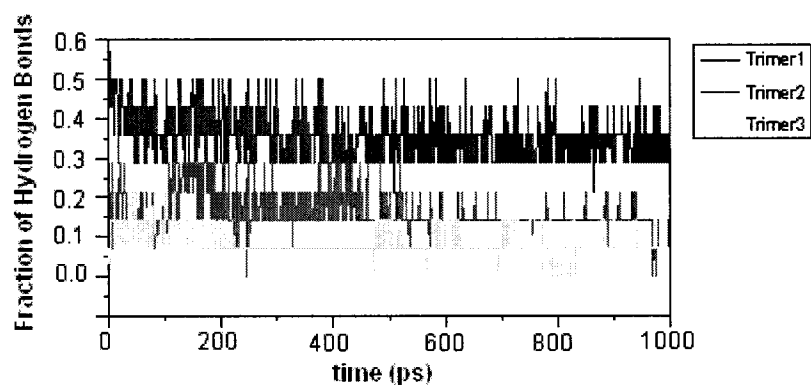
**Figure 5.4** The proportion of remaining hydrogen bonds of Dimer1 and Dimer2 as a function of time.



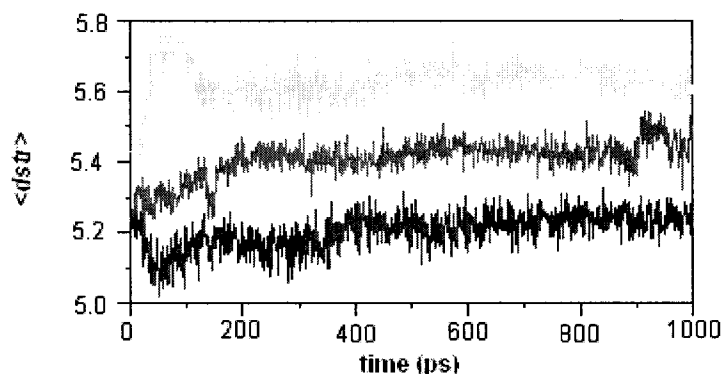
**Figure 5.5** Average distances of Dimer1 and Dimer2 as a function of time.



**Figure 5.6** RMSD of trimers relative to the starting structure.



**Figure 5.7** The fraction of remaining number of hydrogen bonds of trimers throughout the simulation.

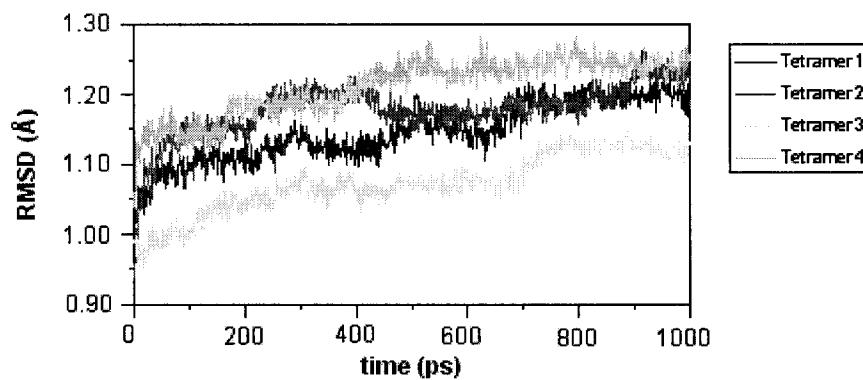


**Figure 5.8** The average distances of trimers.

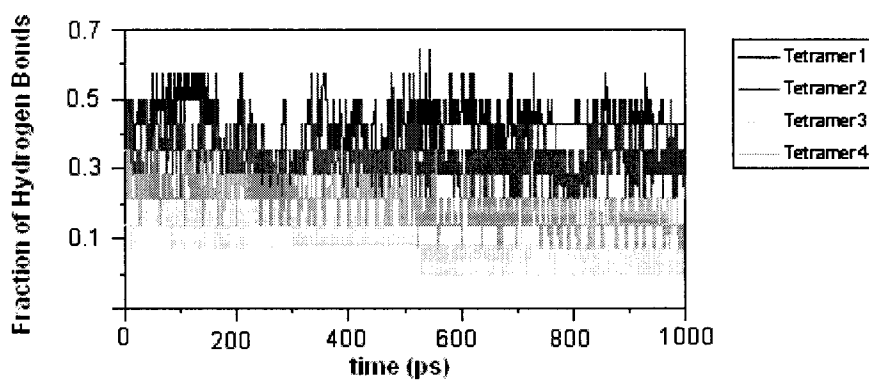
### 5.3.2 Tetramers

A single-layer  $\beta$  sheet is very unstable on its own [79]. In proteins,  $\beta$  sheets often pack onto other elements of secondary structure. We studied multilayered  $\beta$  sheets in the form of tetramers. There are four supramolecular organizations for tetramers. Tetramer 1 and Tetramer 2 consisted of two anti-parallel or parallel  $\beta$  sheets. In each layer, the  $\beta$ -strands are anti-parallel. Tetramer 3 and Tetramer 4 consisted of two anti-parallel or parallel  $\beta$  sheets; in each layer, the  $\beta$ -strands are parallel. Figure 5.9 plots the RMSD values of tetramers with reference to the corresponding initial structures. The overall tendency was for RMSD to increase slowly, but tetramers did pass through local energy minima. The original number of hydrogen bonds of tetramers was 14. Following the hydrogen bonds throughout the simulation, the results showed as before that the anti-parallel orientation within a sheet is the favored one. Tetramer 1 retained above 45 % of its original hydrogen bonds, Tetramer 2 30-45 %, and Tetramer 3, the full parallel structure barely maintain its initial hydrogen bonds (Figure 5.10). Figure 5.11 and 5.12 display the average distances of tetramers. The four tetramers were able to maintain considerable structural integrity for the 1 ns MD simulation, as seen by the small differences between all inter- and intra-sheet distances. In other words, although

Tetramer 3 and Tetramer 4 had lost the majority of hydrogen bonds during the simulation, they still preserved a degree of structural order.

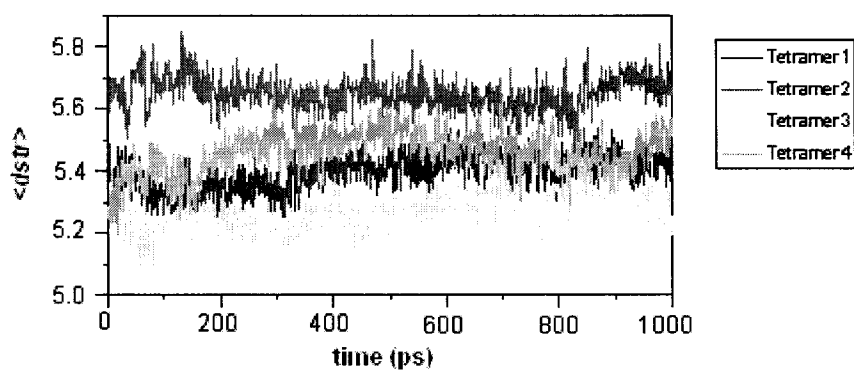


**Figure 5.9** RMSD values of tetramers as a function of time.

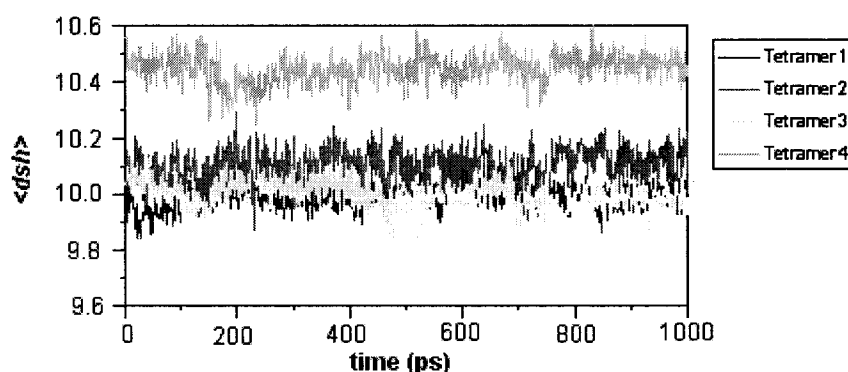


**Figure 5.10** Percentage of remaining number of hydrogen bonds of tetramers throughout the simulation.





**Figure 5.11** Intra-strands distances of tetramers.



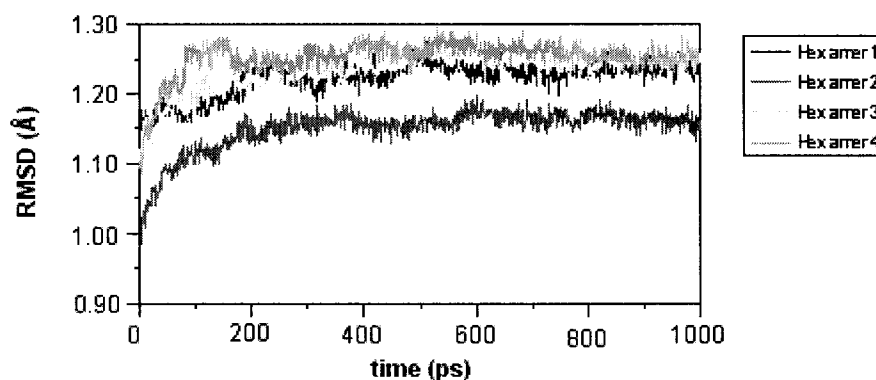
**Figure 5.12** Inter-sheets distances of tetramers

### 5.3.3 Hexamers

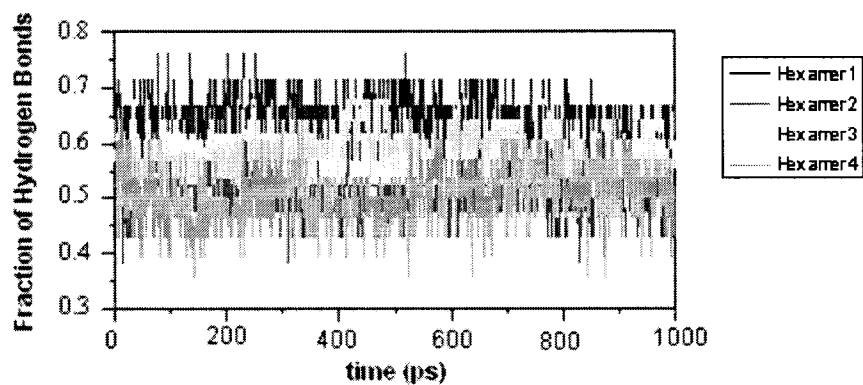
To gain deeper insight, we used Ma and Nussinov's approach and built two hexameric peptide models of three sheets, two strands in each, to check the effect of the stabilization introduced by sheet association. We built another two models of two sheets, three strands in each, to check the stabilization introduced by hydrogen bonds.

Simulation time versus RMSD of the four hexamers is displayed in Figure 5.13. From the comparison of RMSD values, the four hexamers remained at equilibrium throughout the 1 ns simulation time. Hexamer 2 showed the lowest RMSD values

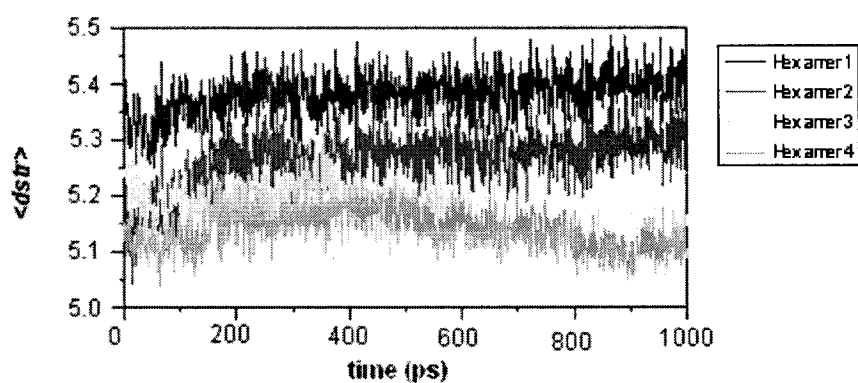
remaining at about 1.15 Å. RMSD values for Hexamer 1, Hexamer 3, and Hexamer 4 were between 1.2 and 1.3 Å with Hexamer 4 a little higher. Regarding the number of original hydrogen bonds, Hexamer 1, Hexamer 2, and Hexamer 3 were able to maintain above 50% of their original interactions, and Hexamer 4 was a little lower but still was able to keep a relatively high percentage (45%) throughout the simulation (Figure 5.14). This is consistent with the behavior of hexamers of YE<sub>5</sub> and YK<sub>5</sub> in Chapter 4, and it agrees with Zanuy and Nussinov's results for amyloid related peptides [73]. With regard to the sheet packing, the four hexamers clearly were able to maintain organized structure, in terms of intra-strands distances (Figure 5.15) and inter-sheets distances (Figure 5.16). Among the four hexamers, Hexamer 1 seems to be the most stable conformation. In this model, two-thirds of the side chains of the three sheets that were buried inside the peptide system contributed to the hydrophobic interactions that played an essential role in stabilizing the  $\beta$  sheet.



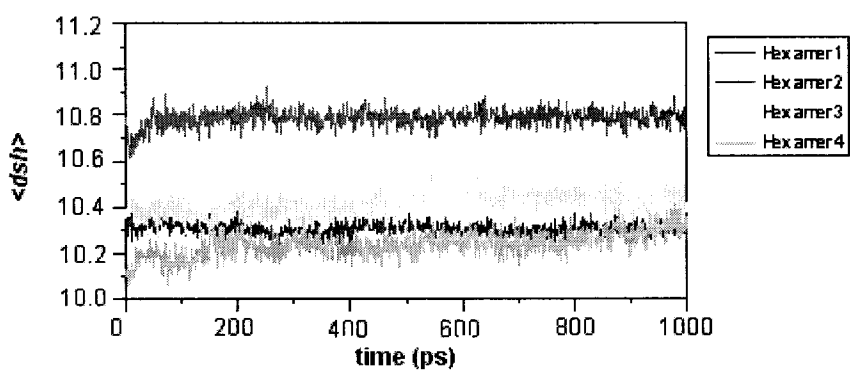
**Figure 5.13** RMSD values of hexamers as referencing to the initial structures.



**Figure 5.14** The fraction of hydrogen bonds of hexamers as a function of time.



**Figure 5.15** Intra-strands distances of hexamers as a function of time.



**Figure 5.16** Inter-sheets distances of hexamers as a function of time.

#### 5.4 Discussions

PLL and PLGA, and cysteine-containing peptides have been shown promising for making thin films, coatings, and microcapsules [3, 4, 5, 34, and 35]. Polypeptide multilayer thin films have been demonstrated to contain various amounts of secondary structure depending on environment. At neutral pH, the majority of regular secondary structure is  $\beta$  sheet [3, 4, and 91]. Detailed dynamical characterization of the internal structure of an LBL film can not be achieved by experimental work.

We have done simulations of multiple copies of peptides YE<sub>5</sub> and YK<sub>5</sub>, from simple structure to supramolecular organization. Following the same simulation approach, we have simulated from 2 to 6 CEV and CKV oligomers. Peptide self-assembly is slow and usually takes minutes to reach its completion. Reproducing the procedure of peptides self-assembly to form  $\beta$  sheet structure on a surface from random coil in solution is beyond our current capabilities. Our focus therefore is on the stability of  $\beta$  sheet structure, and we have probed the physical basis of polypeptide assembly into regular structure in the nano-structured multilayer thin films.

$\beta$  sheet is a common secondary structure of proteins or polypeptides. The first  $\beta$  sheet structure was observed in keratin fibers in 1933 [93]. After almost twenty years, Pauling and Corey proposed the detailed structure of both anti-parallel and parallel sheets and the correct hydrogen bonding patterns for both types [81]. Since we do not have an x-ray or NMR structure for these peptides in an LBL film, following Ma and Nussinov [88], we used a simple planar sheet as a starting conformation to decrease possible bias. We are aware that a  $\beta$  sheet may have various conformations. In fact, most of  $\beta$  sheets observed in proteins are not as planar as Pauling and Corey's models; they all exhibit a

certain degree of twist [14]. Also, there are several configurations for  $\beta$  sheet, including  $\beta$  sandwiches,  $\beta$  barrels, and  $\alpha/\beta$  arrangements [94].

The formation of a  $\beta$  sheet is rather complicated. Unlike an  $\alpha$  helix, the hydrogen bonds are formed either by two parts which belong to the same polypeptide chain but are far away from each other (intramolecular interactions) or two parts of different polypeptides chains (intermolecular interactions). An  $\alpha$  helix is simpler. Another difference between  $\alpha$  helix and  $\beta$  sheet is that a  $\beta$  sheet is less stable than an  $\alpha$  helix when in isolated status. It is not always this case, however, when  $\beta$  sheets appear in a peptide complex. There are several possible contributions to the stability of  $\beta$  sheets, including hydrophobic interactions, electrostatic interactions, hydrogen bonds, and dipole moment interactions. Hydrophobic interactions are related to nonpolar substances minimizing their contacts with water. It has become increasingly clear that hydrophobic interactions are the driving force in protein folding. Our simulations would appear to confirm that hydrophobic interactions are important for stabilizing for peptides in LBL. The charged groups in a peptide or protein contribute to the electrostatic interactions, which are often treated as specific interactions in contrasting to hydrophobic interactions, which are treated as nonspecific. What of hydrogen bonds? Our simulations show that hydrogen bonds have certain impact on stabilizing polypeptide LBL.

Another factor which contributes the stability of  $\beta$  sheets in LBL films is Dipole moment. Dipole moment is believed to have a greater influence in  $\beta$  sheet conformation for short peptides than longer ones [88]. But this also depends on the orientation of  $\beta$  strands because for anti-parallel peptides, the dipole moments will cancel each other,

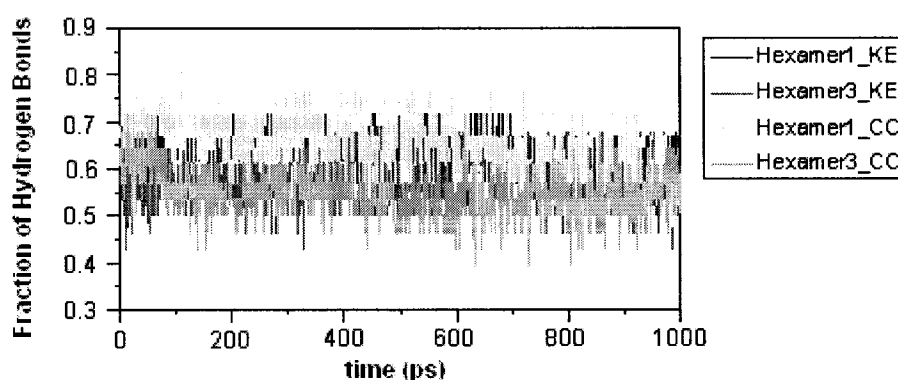
whereas for parallel peptides the dipole moments will be overlapped with the direction of the N-terminus of the  $\beta$  strands [95].

One question of interest is the minimal number of peptides which could form a stable  $\beta$  sheet structure. Our simulations confirmed that oligomer stability increased with the number of the peptides. All hexamers retained about 50 % of their initial hydrogen bonds and maintained internal and external structural integrity throughout the simulation.

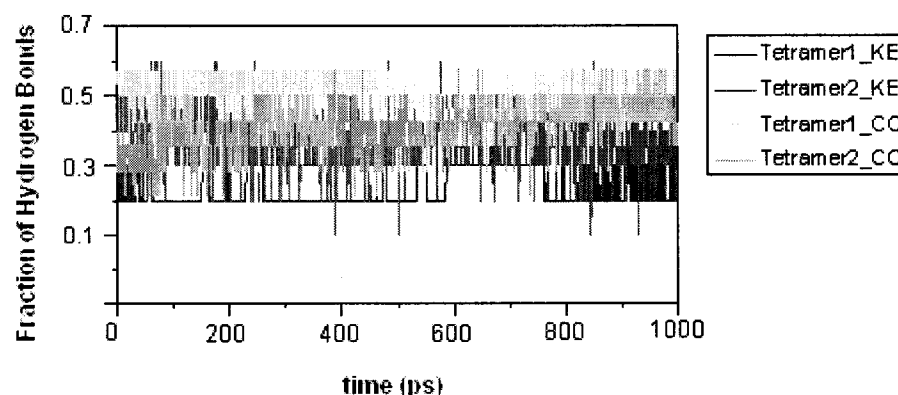
Sequences were designed specially for layer-by-layer assembly. A certain amount of charge in each sequence is an important requirement. The charge density is high in cysteine-containing peptides ( $\sim 50$  %) and it is even higher in YE<sub>5</sub> and YK<sub>5</sub>, respectively, at neutral pH. The results of the simulations resembled studies of natural peptides by other researchers [73, 88]. In our explicit water simulations, the anti-parallel orientation was the most stable one. Longer peptides could possibly be either parallel or anti-parallel with the similar probability [73].

To test the influence of peptide length, we also did simulations of poly-Lys and poly-Glu of length of 8, the same length as the cysteine-containing peptides. Four peptide models were chosen for study, Tetramer 1, Tetramer 2, each of which consists of two sheets anti-parallel or parallel with each other, Hexamer 1, which consists of three parallel sheets, and Hexamer 3, which consists of two sheets anti-parallel with each other. The simulation method was as before. The remaining numbers of hydrogen bonds are displayed in Figures 5.17 and 5.18. For tetramers, peptides K<sub>8</sub> and E<sub>8</sub> had a little higher percentage of hydrogen bonds than CKV and CEV. Tetramer 1 which has the antiparallel  $\beta$  strands and parallel  $\beta$  sheets is more stable than Tetramer 2. All four hexamers maintained at least 50 % percent of initial hydrogen bonds. There was no large

dependence on sequence, as seen by computation of Hexamer 1\_KE and Hexamer 1\_CC, which have similar number of hydrogen bonds, and of Hexamer 3\_KE and Hexamer 3\_CC. The difference was from the conformation itself. The simulations for  $K_8$  and  $E_8$  are consistent with those for CKV and CEV. Taken together, it confirmed the earlier conclusion regarding hydrophobic interactions, electrostatic interactions, and hydrogen bonds in peptide LBL. Also, the simulations show that the most stable and therefore probable  $\beta$  sheet structure is anti-parallel within the sheets and parallel between the sheets.



**Figure 5.17** Comparison of the remaining number of hydrogen bonds of tetramers. KE stands for the peptide group lysine and glutamic acid. CC stands for cysteine-containing peptides. The length of each peptide is 8.



**Figure 5.18** Comparison of the remaining number of hydrogen bonds of hexamers. KE stands for the peptide group lysine and glutamic acid. CC stands for cysteine-containing peptides. The length of each peptide is 8.

## CHAPTER 6

### SUMMARY AND PROSPECTS

#### 6.1 Summary and Contributions

The last several years have witnessed an explosive growth of biological knowledge. The information, however, will be of no use if just keep in a warehouse — sequence databank. Peptide has emerged in recent years as a novel material used for LBL, and they have attracted increasing interest due to potential for bio-related applications. This research constitutes a highly-interdisciplinary approach for creating LBL films.

54,117 and 27,115 unique positively- and negatively-charged sequence motifs were identified by computational approach presented in this dissertaion. Secondary structure prediction is an important aspect of understanding the relationship between polypeptide sequence, structure and function. Based on over 1,000 high-resolution protein structures obtained by X-ray crystallography, the Chou and Fasman secondary structure parameters have been recalculated. The results agree well with available experimental data.

Each sequence motif has been evaluated in terms of antigenicity and secondary structure probability. The results have been stored in a relational database and can be accessed by a user-friendly interface. This will prove invaluable to the polypeptide design



engineering process. One of the key differences between polypeptides and other types of polyelectrolyte is the stability of the former to form higher-order regular structures in the films, namely  $\alpha$  helix and  $\beta$  sheet. Although it is too early to say whether these structures will be advantageous for film applications, we do know they exist and exhibit different secondary structure under various environments. Insight into the internal structure and mobility of the peptides on an atomic level can be sought obtained from MD simulation.

MD simulation of all-atom models has been used for the first time to gain information of the stability of multilayer thin films fabricated from peptides. Simulations have been carried out to study structural and dynamical properties of peptide systems involving the peptide sequences  $Y(L)_5$  and  $Y(K)_5$ , and  $KC(KV)_2KG$ , and  $EC(EV)_2EG$ . Our simulations not only provided a detailed picture of the peptide in LBL film but also shed light on the understanding the physical basis of peptide LBL. We have found that hydrophobic interaction is the most important one for peptide LBL and that the electrostatic interaction plays an essential role in stabilizing  $\beta$  sheet structure at neutral pH. Our simulations also show that hydrogen bonds might be a consequence of forming secondary structure by peptide LBL rather than the cause of it. Moreover, when the number of peptides in a supramolecular structure is relatively large, there is increased likelihood that it will be stable. The most stable suparmolecular structure is anti-parallel within sheets and parallel between sheets.

## 6.2 Future Study and Prospects

This research has explored a way to integrate computer science, biochemistry, biology, bioinformatics, statistics, and materials science. This approach will help to guide the design of experiments and to improve the understanding of experimental results. However, there is still much room for improvement. For example, due to the limitation of current computational ability, we can not simulate the entire process of forming secondary structure by peptides from random conformations in explicit solvent. Implicit solvent is faster for MD simulation and might be a viable alternative at some point, but it is less accurate than explicit solvent. In the future we expect this method will be improved so that the computation time for large molecular systems will be greatly decreased and make it possible to simulate a larger number of molecules with longer durations.

**APPENDIX  
SOURCE CODES**

## 1. Source code for identification of peptide motifs

```
*****
//This program extracts the same charge amino acid sequence with a defined length from human protein
//sequence (supplied by Vinay) and modified to standard format. 7 was used as a motif length. But it can
be //change to any other lengths.
```

```
#include <stdlib.h>
#include <ctype.h>
#include <math.h>
#include <iostream.h>
#include <fstream.h>

ifstream infile( "seq.txt", ios::in );
ofstream outfile( "red_seq1.txt", ios::out );
ofstream outfiled( "red_d.txt", ios::out );
ofstream outfilee( "red_.txt", ios::out );
void function ( char [], int, int &, char [], int );

void main()
{
    int i = 0;
    int s = 0, t = 0, start = 0;
    int sgi = 0, size = 0;
    int countN = 0, countP = 0;
    int tempstart = 0;
    char sgiarray[1000];
    char asc = 'a';
    char samplearray[25000];
    char tempsamplearray[25000];

    while( infile.peek() != EOF )
    {
        asc = infile.get();
        if ( asc == '>' ) {
            asc = infile.get();
            t = infile.tellg();
            infile.seekg( t + 2 );
            do {
                asc = infile.get();
                sgiarray[sgi++] = asc;
            } while( infile.peek() != '|' && sgi <= 10 );
        }
        if( asc == '[' ) {
            asc = infile.get();
            if ( asc >= 65 && asc <= 89 ) {
                s = infile.tellg();
                infile.seekg( s - 1 );
                do {
                    asc = infile.get();
                    samplearray[start++] = asc;
                } while( infile.peek() != '>' );
            }
            outfile << endl;
        }
    }
}
```

```

        outfile << start << endl;
        for (int i = 0; i < start; i++) {
            tempsamplearray[tempstart++] = samplearray[i];
        }

        function (tempsamplearray, tempstart, countN, sgiarray, sgi);
            sgi = 0;
            sgiarray[1000] = '';
            samplearray[25000] = '';
            tempsamplearray[25000] = '';
            tempstart = 0;
            start = 0;
        }
        else
            asc = infile.get();
    }
}
infile.close();
outfile.close();
}

void function ( char ssamplearray[],int ssum, int &cN, char sgia [], int sg )
{
    int sc = 0;
    int saa = 15;
    while( sc<ssum ) {
        if( ssamplearray[sc] == 'D' || ssamplearray[sc] == 'E' ) {
            int counter = 0;
            char temparray[] = "xxxxxxxxxxxxxxxx";
            for(int i=0; i<saa; i++) {
                temparray[i] = ssamplearray[sc];
                sc++;
                if (temparray[i] == 'D' || temparray[i] == 'E')
                    counter++;
            }
            else
                if( temparray[i] == 'K' || temparray[i] == 'R' ||
                    temparray[i] == 'H' || temparray[i] == 'X' )
                    counter = counter-saa + 1;
        }

        sc = sc-saa;
        if( counter >= 8 ) {
            for( int number_gi = 0; number_gi < sg; number_gi++){
                outfile << sgia[number_gi];
            }
            outfile << ",";
            outfile << "n" << ++cN << "," << endl;;
            if( temparray[0] == 'D' ) {
                outfile << ">";
                for( int i = 0; i < saa; i++ ) {
                    cout << temparray[i];
                    outfile << temparray[i];
                }
                outfile << "," << endl;
            }
            else if( temparray[0] == 'E' ) {
                outfile << ">";
            }
        }
    }
}

```

```

        for( int i = 0; i < saa; i++ ) {
            cout << temparray[i];
            outfilee << temparray[i];
        }
        outfilee<<"<<endl;
    }
    sc = sc + saa;
}
else
    sc = sc + 1;
}
sc++;
}
}

```

## 2. Source code for calculation of protein secondary propensity values

```

*****
//This program calculates secondary structure propensity values based on chou and fasman's method.
//The sample of protein structures were obtained Protein data bank (2001).

#include <iostream.h>
#include <fstream.h>
#include <stdlib.h>
#include <ctype.h>
#include <string.h>

void checkPdbId(char [], int &, int &, int &);
int check_seq(char []);
int check_helix(char []);
int check_dbre(char []);

ofstream outfile("aa_seq.txt", ios::out);
ofstream outfile2("pdb_helix.txt", ios::out);
ofstream outfile3("pdb_sheet.txt", ios::out);
ofstream outfile4("seq_helix.txt", ios::out);
ofstream outfile5("seq_sheet.txt", ios::out);
ofstream outfile10("length_seq.txt", ios::out);
ofstream outfile("name_sheet.txt", ios::out);
ofstream outfileh("name_helix.txt", ios::out);

void main()
{
    ifstream infile("v_list.txt", ios::in);
    char asc = 'a'; int sum = 0; int suma=0; int sumb=0;
    char *temp1; temp1 = new char [10];
    char *temp2; temp2 = new char [5];

    int start = 0;
    while(asc != EOF)
    {
        asc = infile.get();
        if(asc == '>')

```

```

        {
            for(int h=0; h<7; h++)
            {
                asc=infile.get();
                temp1[h]=asc;
            }
            checkPdbId(temp1, sum, suma, sumb);
        }
    else;
}
}
void checkPdbId(char aa[], int &sum, int &suma, int &sumb)
{
    char seqName[10]; char amount[100]; char *tamount; tamount = new char[5];
    int startamount=0; int bi = 0; int ta=0;
    char *seqT; seqT = new char[5000]; int sT=0;
    char *seqA; seqA = new char[3500]; int sA=0;
    char *seqB; seqB = new char[2500]; int sB=0;
    char *seqC; seqC = new char[3500]; int sC=0;
    char *seqD; seqD = new char[3500]; int sD=0;
    char *seqE; seqE = new char[3500]; int sE=0;
    char *seqF; seqF = new char[3500]; int sF=0;
    char *seqG; seqG = new char[3500]; int sG=0;
    char final[2500]; int number = 0;
    char finalA[2500]; int numberA = 0;
    char finalB[2500]; int numberB = 0;
    char finalC[2500]; int numberC = 0;
    char finalD[2500]; int numberD = 0;
    char finalE[2500]; int numberE = 0;
    char finalF[2500]; int numberF = 0;
    char finalG[2500]; int numberG = 0;
    int flag_helix = 0; int flag_sheet = 0; int flag_helixA = 0; int flag_sheetA = 0;
    int flag_helixB = 0; int flag_sheetB = 0; int flag_helixC = 0; int flag_sheetC = 0;
    int flag_helixD = 0; int flag_sheetD = 0; int flag_helixE = 0; int flag_sheetE = 0;
    int flag_helixF = 0; int flag_sheetF = 0; int flag_helixG = 0; int flag_sheetG = 0;
    int sstart = 0; int s = 0; int sumh = 0, sums = 0;
    char helixName[10]; int hS=0; char asc = 'a'; char *dbre; dbre = new char [5];
    aa[7] = '\0';
    sum ++;
    outfile<<aa<<" "<<sum<<endl;
    outfile4<<aa<<" "<<endl;
    outfile5<<aa<<" "<<endl;

    ifstream infile (aa, ios::in);

    if(!infile) {
        outfile<<aa<<" does not exist."<<endl;
    }
    else{
        while(asc != EOF) {
            asc = infile.get();
            if(asc == 'D') {
                for(int i=0; i<4; i++) {
                    asc = infile.get();
                    dbre[i] = asc;
                }
            }
        }
    }
}

```

```

        if(check_dbre(dbre) == 1) {
            while(asc != '\n') {
                asc = infile.get();
            }
        }
    }
    else if(asc == 'S') {
        for(int i = 0; i<5; i++) {
            asc = infile.get();
            seqName[i] = asc;
        }
        if(check_seq(seqName) == 1) {
            do {
                asc = infile.get();
                if(asc != ' ' || asc != '\n') {
                    amount[startamount] = asc;
                    startamount++;
                }
            } while(asc != '\n');
        }
    }
    if(startamount > 70) {
        if (amount[5] == 'A') {
            for(int i = 13 ; i<64; i++) {
                if(amount[i]!=' ') {
                    seqA[sA] = amount[i];
                    sA++;
                }
            }
        }
        if (amount[5] == 'B') {
            for(int i = 13 ; i<64; i++) {
                if(amount[i]!=' ') {
                    seqB[sB] = amount[i];
                    sB++;
                }
            }
        }
        if (amount[5] == 'C') {
            for(int i = 13 ; i<64; i++) {
                if(amount[i]!=' ') {
                    seqC[sC] = amount[i];
                    sC++;
                }
            }
        }
        if (amount[5] == 'D') {
            for(int i = 13 ; i<64; i++) {
                if(amount[i]!=' ') {
                    seqD[sD] = amount[i];
                    sD++;
                }
            }
        }
        if (amount[5] == 'E') {
            for(int i = 13 ; i<64; i++) {
                if(amount[i]!=' ') {
                    seqE[sE] = amount[i];
                }
            }
        }
    }
}

```



```

        sE++;
    }
}
}
if (amount[5] == 'F') {
for(int i = 13 ; i<64; i++) {
if(amount[i]!=' ') {
    seqF[sF] = amount[i];
    sF++;
}
}
}
if (amount[5] == 'G') {
    for(int i = 13 ; i<64; i++) {
        if(amount[i]!=' ') {
            seqG[sG] = amount[i];
            sG++;
        }
    }
}
if (amount[5] == ' ') {
    for(int i = 13 ; i<64; i++) {
        if(amount[i]!=' ') {
            seqT[sT] = amount[i];
            sT++;
        }
    }
}
}
}
startamount = 0;
}
}

sT = 0; sB = 0; sA = 0; sC = 0; sD = 0; sE = 0; sF = 0; sG = 0;
seqT[5000] = ''; seqA[3500] = ''; seqB[2500] = ''; seqC[3500] = ''; seqD[3500] = '';
seqE[3500] = ''; seqF[3500] = ''; seqG[3500] = '';
for(int i=0; i<numberA; i++) {
    outfile<<finalA[i];
}
for(int i1=0; i1<numberB; i1++) {
    outfile<<finalB[i1];
}
for(int i2=0; i2<numberC; i2++) {
    cout<<finalC[i2];
    outfile<<finalC[i2];
}
for(int i3=0; i3<numberD; i3++) {
    outfile<<finalD[i3];
}
for(int i4=0; i4<numberE; i4++) {
    outfile<<finalE[i4];
}
for(int i5=0; i5<numberF; i5++) {
    outfile<<finalF[i5];
}

```

```

}
for(int i6=0; i6<numberG; i6++) {
    outfile<<finalG[i6];
}
for(int ii=0; ii<number; ii++) {
    outfile<<final[ii];
}

outfile<<endl;
infile.close();
infile.open(aa);
asc = infile.get();
while(!infile.eof()) {
asc = infile.get(); int head =0; int tail=0; int shead =0; int stail=0;
char hSeq[100]; int sh=0;
    if(asc == 'H') {
        for(int i = 0; i<4; i++) {
            asc = infile.get();
            helixName[i] = asc;
        }
        if(check_helix(helixName)==1) {
            while(asc != '\n') {
                sc = infile.get();
                hSeq[sh] = asc;
                sh++;
            }
            if( sh >70) {
                flag_helix = 1;
                if(hSeq[14]=='A') {
                    if(head<numberA && tail<numberA) {
                        flag_helixA = 1;
                        outfile10<<">"<<(tail-head+1)<<"<<endl;
                        for(int j2=head-1; j2<tail; j2++) {
                            outfile2<<finalA[j2];
                        }
                    }
                    outfile2<<endl;
                }
            }
            else if(asc == 'S') {
                char sSeq[50];
                int ssh = 0; int sbegin = 0; int send = 0;
                for(int i = 0; i<4; i++)
                    asc = infile.get();
                    seqName[i] = asc;
            }
            if(check_seq(seqName) == 2)
                while(asc != '\n') {
                    asc = infile.get();
                    sSeq[ssh] = asc;
                    ssh++;
                }
            if(ssh >70) {
                flag_sheet = 1;
                if(sSeq[16]=='A') {
                    if(sbegin<numberA && send <numberA) {
                        flag_sheetA = 1;
                        for(int j2=sbegin-1; j2<send; j2++) {
                            outfile3<<finalA[j2];
                        }
                    }
                }
            }
        }
    }
}

```

```

        outfile3<<endl;
    }
}
if(sSeq[16]=='B') {
if(sbegin<numberB && send <numberB) {
    flag_sheetB = 1;
    for(int j2=sbegin-1; j2<send; j2++) {
        outfile3<<finalB[j2];
    }
    outfile3<<endl;
}
}
if(sSeq[16]=='C') {
if(sbegin<numberC && send <numberC){
    flag_sheetC = 1;
for(int j2=sbegin-1; j2<send; j2++) {
    outfile3<<finalC[j2];
}
    outfile3<<endl;
}
}
if(sSeq[16]=='D') {
if(sbegin<numberD && send <numberD) {
flag_sheetD = 1;
    for(int j2=sbegin-1; j2<send; j2++) {
        outfile3<<finalD[j2];
    }
    outfile3<<endl;
}
}
if(sSeq[16]=='E') {
if(sbegin<numberE && send <numberE) {
flag_sheetE = 1;
for(int j2=sbegin-1; j2<send; j2++) {
    outfile3<<finalE[j2];
}
    outfile3<<endl;
}
}
if(sSeq[16]=='F') {
if(sbegin<numberF && send <numberF) {
flag_sheetF = 1;
for(int j2=sbegin-1; j2<send; j2++) {
    outfile3<<finalF[j2];
}
    outfile3<<endl;
}
}
if(sSeq[16]=='G') {
if(sbegin<numberG && send <numberG) {
flag_sheetG = 1;
    for(int j2=sbegin-1; j2<send; j2++) {
        outfile3<<finalG[j2];
    }
    outfile3<<endl;
}
}
}

```

```

    }
    if(sSeq[16]==' ') {
    if(sbegin<number && send <number) {
        flag_sheetX = 1;
        for(int j2=sbegin-1; j2<send; j2++) {
            outfile3<<final[j2];
        }
        outfile3<<endl;
    }
    }
}
ssh = 0;
sbegin = 0;
send = 0;
}
}
infile.close();
}
if(flag_helix == 1) {
    if(flag_helixA == 1) {
        for(int i=0; i<numberA; i++) {
            outfile4<<finalA[i];
        }
    }
    if(flag_helixB == 1) {
        for(int i1=0; i1<numberB; i1++) {
            outfile4<<finalB[i1];
        }
    }
    if(flag_helixC == 1) {
        for(int i2=0; i2<numberC; i2++) {
            outfile4<<finalC[i2];
        }
    }
    if(flag_helixD == 1) {
        for(int i3=0; i3<numberD; i3++) {
            outfile4<<finalD[i3];
        }
    }
    if(flag_helixE == 1) {
        for(int i4=0; i4<numberE; i4++) {
            outfile4<<finalE[i4];
        }
    }
    if(flag_helixF == 1) {
        for(int i5=0; i5<numberF; i5++) {
            outfile4<<finalF[i5];
        }
    }
    if(flag_helixG == 1) {
        for(int i6=0; i6<numberG; i6++) {
            outfile4<<finalG[i6];
        }
    }
}
}

```

```

        if(flag_helixX == 1) {
            for(int ii=0; ii<number; ii++) {
                outfile4<<final[ii];
            }
        }
        outfile4<<endl;
    }
    if (flag_sheet == 1) {
        if(flag_sheetA == 1) {
            for(int i=0; i<numberA; i++) {
                outfile5<<finalA[i];
            }
        }
        if(flag_sheetB == 1) {
            for(int i1=0; i1<numberB; i1++) {
                outfile5<<finalB[i1];
            }
        }
        if(flag_sheetC == 1) {
            for(int i2=0; i2<numberC; i2++) {
                outfile5<<finalC[i2];
            }
        }
        if(flag_sheetD == 1) {
            for(int i3=0; i3<numberD; i3++) {
                outfile5<<finalD[i3];
            }
        }
        if(flag_sheetE == 1) {
            for(int i4=0; i4<numberE; i4++) {
                outfile5<<finalE[i4];
            }
        }
        if(flag_sheetF == 1) {
            for(int i5=0; i5<numberF; i5++) {
                outfile5<<finalF[i5];
            }
        }
        if(flag_sheetG == 1) {
            for(int i6=0; i6<numberG; i6++) {
                outfile5<<finalG[i6];
            }
        }
        if(flag_sheetX == 1) {
            for(int ii=0; ii<number; ii++) {
                outfile5<<final[ii];
            }
        }
        outfile5<<endl;
    }
    if(flag_helix ==1 && flag_helixA == 1 || flag_helixB==1 || flag_helixC==1 || flag_helixD==1 ||
    flag_helixE==1 || flag_helixF==1 || flag_helixG==1 || flag_helixX ==1)
        outfileh<<aa<<suma++<<endl;
    if(flag_sheet ==1 && flag_sheetA == 1 || flag_sheetB==1 || flag_sheetC==1 || flag_sheetD==1 ||
    flag_sheetE==1 || flag_sheetF==1 || flag_sheetG==1 || flag_sheetX ==1)
        outfiles<<aa<<sumb++<<endl;

```

```

number = 0; numberA = 0; numberB = 0; numberC = 0; numberD = 0; numberE = 0;
numberF = 0; numberG = 0;
final[2500]=' '; finalA[2500]=' '; finalB[2500]=' '; finalC[2500]=' '; finalD[2500]=' ';
finalE[2500]=' '; finalF[2500]=' '; finalG[2500]=' ';
}
int check_dbre(char db[]) {
    db[4]='\0';
    if(db[0] == 'B' && db[1] == 'R' && db[2] == 'E' && db[3] == 'F' )
        return 1;
    else
        return 0;
}
int check_helix(char b[]){
    if(b[0]=='E' && b[1]=='L' && b[2] == 'I' && b[3] == 'X' )//&& b[4] == 'X')
        return 1;
    else
        return 0;
}
int check_seq(char a[]) {
    if(a[0]=='E' && a[1] == 'Q' && a[2] == 'R' && a[3] == 'E' && a[4] == 'S' ) //&& a[5] == 'S')
        return 1;
    else if(a[0]=='H' && a[1]=='E' && a[2]=='E' && a[3]=='T' )//&&a[4]=='T')
        return 2;
    else
        return 0;
}
}

```

## REFERENCES

1. Decher, G. and Schlenoff J. *Multilayer Thin Films* (ed.), Wiley-VCH, Weinheim, Germany, 2003.
2. Lvov, Y. 2000. "Electrostatic layer-by-layer assembly of proteins and polyions" in *Protein Architecture: Interfacial Molecular Assembly and Immobilization Biotechnology*. (New York: Dekker) pp125-167.
3. Zhang, L.; Li, B.; Zhi, Z.-L. ; Haynie, D.T. Susceptibility of Internal Structure of Polypeptide Multilayer Thin Films to Environmental Perturbation. In press.
4. Boulmedais, F., Bozonnet, M., Schwinté, P., Voegel, J.-C., and Schaaf, P. 2003. Buildup of Exponentially Growing Multilayer Polypeptide Films with Internal Secondary Structure. *Langmuir* 19: 9873-9882.
5. Haynie, D., Balkundi, S., Palath, N., Chakravarthula, K., and Dave, K. 2004. Polypeptide Multilayer Films: Role of Molecular Structure and Charge *Langmuir* 20: 4540-4547.
6. Ulman, U. 1991. *An introduction to ultrathin organic films: from Langmuir-Blodgett to selfassembly*; Academic Press, Inc.: Boston.
7. Blodgett, K.B. 1937. Built-up films of barium stearate and their optical properties *Phys. Rev.* 51: 964-982.
8. Langmuir, I. and Schaefer, V.J. 1938. Activities of Urease and Pepsin Monolayers. *J. Am. Chem. Soc.* 60: 1351-1360.
9. Decher, G. and Hong J.D., 1991. Build-up of ultrathin multilayer films by a self-assembly process, I. Consecutive adsorption of anionic and cationic bipolar amphiphiles on charged surfaces. *Makromol. Chem. Macromol. Symp.* 46, 321-327.
10. Tripathy, S.K., Kumar, J. and Nalwa, H.S.eds. 2002. *Handbook of Polyelectrolytes and their Applications. Vol. 1: Polyelectrolyte-based Multilayers, Self-assemblies and Nanostructures*. Stevenson Ranch, CA: American Scientific.
11. Iler, R.K. 1966. Multilayers of colloidal particles. *J. Colloid Interface Sci.*, 21: 569-594.

12. Caruso, F., Donath, E., Möhwald, H., Georgieva, R. 1998. Fluorescence Studies of the Binding of Anionic Derivatives of Pyrene and Fluorescein to Cationic Polyelectrolytes in Aqueous Solution. *Macromolecules* 31: 7365-7377.
13. Stockton W. B. and Rubner M. F. 1997. Molecular-Level Processing of Conjugated Polymers. 4. Layer-by-Layer Manipulation of Polyaniline via Hydrogen-Bonding Interactions. *Macromolecules*, 30: 2717-2725.
14. Hao, E., Lian, T. 2000. Buildup of Polymer/Au Nanoparticle Multilayer Thin Films Based on Hydrogen Bonding. *Chem. Mater.* 12: 3392-3396
15. Kharlampieva, E., Sukhishvili, S. A. 2003. Polyelectrolyte Multilayers of Weak Polyacid and Cationic Copolymer: Competition of Hydrogen-Bonding and Electrostatic Interactions. *Macromolecules*. 36: 9950-9956.
16. Shimazaki, Y., Mitsuishi, M., Ito, S., Yamamoto, M. 1997. Preparation of the Layer-by-Layer Deposited Ultrathin Film Based on the Charge-Transfer Interaction. *Langmuir* 13: 1385-1387.
17. Zhang, Y., Cao, W. 2001. Stable Self-Assembled Multilayer Films of Diazo Resin and Poly(maleic anhydride-co-styrene) Based on Charge-Transfer Interaction. *Langmuir* 17: 5021-5024.
18. Serizawa, T., Nanameki, K., Yamamoto, K., Akashi, M. 2002. Thermoresponsive Ultrathin Hydrogels Prepared by Sequential Chemical Reactions. *Macromolecules*. 35: 2184-2189.
19. Langmuir, I. 1941. U.S. patent #2,232,539.
20. Decher, G., Hong, J. D. and Schmitt, 1992. Buildup of ultrathin multilayer films by a self-assembly process: III. Consecutively alternating adsorption of anionic and cationic polyelectrolytes on charged surfaces. *Thin Solid Films*. 210/211: 831-835.
21. Klitzing, R. v. and Mohwald, H. 1995. Proton Concentration Profile in Ultrathin Polyelectrolyte Films. *Langmuir* 11: 3554-3559.
22. Klitzing, R. v. and Mohwald, H. 1996. A Realistic Diffusion Model for Ultrathin Polyelectrolyte Films. *Macromolecules*. 29: 6901-6906.
23. Lvov, Y., Haas, H., Decher, G., Möhwald, H., Kalachev, A. 1993. Assembly of polyelectrolyte molecular films onto plasma-treated glass. *J. Phys. Chem.*, 97: 12835-12841.
24. Lvov, Y., Decher, G., Möhwald, H. 1993. Assembly, structural characterization and thermal behavior of layer-by-layer deposited ultrathin films of poly(vinylsulfate) and poly(allylamine). *Langmuir* 9: 481-486.



25. Decher, G. 1997. Fuzzy Nanoassemblies: toward layered polymeric multicomposites. *Science* 277: 1232-1237.
26. Losche, M., Schmitt, J., Decher, G. Bouwman, W. G., Kjaer, K. 1998. Detailed Structure of Molecularly Thin Polyelectrolyte Multilayer Films on Solid Substrates as Revealed by Neutron Reflectometry. *Macromolecules*. 31(25): 8893-8906.
27. Donath, E., Sukhorukov, G.B., Caruso, F., Davis, S.A., Möhwald, H. 1998. Novel hollow polymer shells by colloid-templated assembly of polyelectrolytes. *Angew. Chem. Int. Ed. Engl.* 37: 2202-2205.
28. Sukhorukov, G.B., Donath, E., Davis, S. A., Lichtenfeld, H., Caruso, F., Popov, V.I., Möhwald, H. 1998. Stepwise polyelectrolyte assembly on particle surfaces: a novel approach to colloid design *Polym. Adv. Technol.* 9: 759-767.
29. Caruso, F., Niikura, K., Furlong, D. N., and Okahata, Y. 1997. Ultrathin Multilayer Polyelectrolyte Films on Gold: Construction and Thickness Determination. *Langmuir* 13(13): 3422-3426.
30. Lvov, Y., Ariga, K., Ichinose, I. and Kunitake, T. 1995. Assembly of multicomponent protein films by means of electrostatic Layer-by-Layer Adsorption. *J. Am. Chem. Soc.* 117: 6117-6123.
31. Boulmedais, F., Schwinté, P., Gergely, C., Voegel, J.-C. and Schaaf, P. 2002. Secondary structure of polypeptide multilayer films: an example of locally ordered polyelectrolyte multilayers. *Langmuir* 18: 4523-4525.
32. Elbert, D.L., Herbert, C.B., and Hubbell, J.A. 1999. Thin Polymer Layers Formed by Polyelectrolyte Multilayer Techniques on Biological Surfaces. *Langmuir* 15: 5355-5362.
33. Zhang, S.G., Holmes, T., Lockshin, C., and Rich, A. 1993. Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl. Acad. Sci. USA.* 90: 3334-3338.
34. Li, B.Y. and Haynie, D.T. 2004. Multilayer Biomimetics: Reversible Covalent Stabilization of a Nanostructured Biofilm. *Biomacromolecules*. In press.
35. Li, B.Y., Haynie, D.T., Palath, N., Janisch, D. 2004. Nano-scale Biomimetics: Fabrication and Optimization of Stability of Peptide-based Multilayer Thin Films. *J. Nanosci. Nanotech.* In press.
36. Cooper, T.M., Campbell, A. L. and Crane, R.L. 1995. Formation of polypeptide-dye multilayers by electrostatic self-assembly technique. *Langmuir* 11: 2713-2718.

37. O'Brien-Simpson, N.M., Ede, J.N., Brown, L.E., Swan, J. and Jackson C.D. 1997. Polymerization of Unprotected Synthetic Peptides: A View toward Synthetic Peptide Vaccines. *J. Am. Chem. Soc.* 119: 1183-1188.
38. Kyte, J., Doolittle, R.F. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157: 105-132.
39. Hopp, T.P., Woods, K.R. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA.* 78: 3824-3828.
40. Park, H.J., Choi, J.Y., Seong, Y.M. 2003. Antigenicity of the region encoded by exon8 of the human serine protease, HtrA2/Omi, is associated with its protein solubility. *Biotechnology Letters* 25: 1597-1603.
41. Shakesheff, K., Cannizzaro, S., Langer, R. 1998. Creating biomimetic micro-environments with synthetic polymer-peptide hybrid molecules. *J. Biomater. Sci. Polym. Ed.* 9: 507-518.
42. Kobatake, E., Onoda, K., Yanagida, Y., Aizawa, M. 2000. Design and gene engineering synthesis of an extremely thermostable protein with biological activity. *Biomacromolecules* 1: 382-386.
43. Nosoh, Y., Sekiguchi, T. 1990. Protein engineering for thermostability. *Trends. Biotechnol.* 8: 16-20.
44. Matsumura, M., Becktel, W.J., Levitt, M., Matthews, B.W. 1989. Stabilization of phage T4 lysozyme by engineered disulfide bonds *Proc. Natl. Acad. Sci. U.S.A.* 86: 6252-6566.
45. Yutani, K., Ogasahara, K., Tsuji, T.; Sugino, Y. 1987. Dependence of Conformational Stability on Hydrophobicity of the Amino Acid Residue in a Series of Variant Proteins Substituted at a Unique Position of Tryptophan Synthase  $\alpha$  Subunit *Proc. Natl. Acad. Sci. U.S.A.* 84: 4441-4444.
46. Karpusas, M., Baase, W.A., Matsumura, M., Matthews, B.W. 1989. Hydrophobic Packing in T4 Lysozyme Probed by Cavity-Filling Mutants. *Proc. Natl. Acad. Sci. U.S.A.* 86: 8237-8241.
47. Zhang, S.G., Lockshin, C., Cook, R., Rich, A. 1994. Unusually Stable  $\beta$ -Sheet Formation in an Ionic Self-Complementary Oligopeptide. *Biopolymers* 34: 663-672.
48. Holmes, T.; Delacalle, S.; Su, X.; Rich, A.; Zhang, S. 2000. Proc. Extensive neurite outgrowth and active neuronal synapses on peptide scaffolds. *Natl. Acad. Sci. U.S.A.* 97: 6728-6733.

49. Hong, Y., R. L. Legge, S. Zhang and P. Chen. 2003. Effect of amino acid sequence and pH on nanofiber formation of the self-assembling peptides EAK16-II and EAK16-IV. *Biomacromolecules* 4: 1434–1442.
50. Ullner, M. 2002. “polyelectrolyte models in theory and simulation” in *handbook of polyelectrolytes and their applications*, volume 3. chap. 10 (American scientific publishers) pp275-279.
51. Rozas, J.; Li, B.; Haynie, D.T. Unpublished data.
52. Zheng, B., Haynie, D.T., Zhong, H, Sabnis, K., Surpurija, V., Pargaonkar, N, Sharma, G., and Vistakula, K. 2004. Design of peptides for thin films, coating, and microcapsules for applications in biotechnology. *J. BIOMAT SCI-POLYM E*. In press.
53. Haynie, D.T., Palath, N., Liu, Y., and Parganokar, N. Biomimetic nanotechnology: Disulfide-bonded polypeptide microcapsules are especially stable in harsh environments. Submitted.
54. Frenkel, D. and Smit, B. 2002. *Understanding Molecular Simulations: from Algorithms to Applications*, Academic Press, San Diego, 2nd edition
55. Emini, E.A., Hughes, J., Perlow, D. and Boger. J.1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* 55: 836-839.
56. Alder, B, J. and Wainwright, T.E. 1957. Phase transition for a hard sphere system. *J. Chem. Phys.* 27: 1208-1209.
57. Alder, B.J. and Wainwright, T.E. 1959. Studies in molecular dynamics, i. general method. *J. Chem. Phys.* 31: 459-466.
58. Rahman, A. 1964. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.*136: 405-411.
59. Stillinger, F.H. and Rahman, A.J. 1974. Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.* 60: 1545-1557.
60. McCammon, J.A., Gelin, B.R., and Karplus, M. 1977. Dynamics of folded proteins. *Nature.* 267: 585-590.
61. Verlet, L. 1967. Computer Experiments on Classical Fluids I: Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* 159: 98-103.
62. McQuarrie, D.A. 1976. In *Statistical Mechanics*, pp. 236–37. New York: Harper and Row.

63. Brooks, R. b., Bruccoleri, R. E., Olafson, B. D., Sate, D. J., Swaminathan, S. and Karplus, M. 1983. CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.* 4: 187-217.
64. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., et al. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106: 765–84.
65. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A.. 1986. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comp. Chem.* 7: 230–52.
66. Brooks, C.L., Karplus, M., Pettitt, B.M. 1988. *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics* (John Wiley and Sons).
67. Mahoney, M. W. and W. L. Jorgensen. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* 112: 8910-8922.
68. Robinson, G.W., Zhu, S.B., Singh, S. and Evans, M.W. 1996. *Water in Biology, Chemistry and Physics: Experimental Overviews and Computational Methodologies* (World Science. Singapore).
69. Kusalik, P. G. and I.M. Svishchev. 1994. The spatial structure in liquid water. *Science* 265: 1219-1221.
70. Jorgensen, W.L, Chandrasekhar, J and Madura, J.D. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79: 9606-9616.
71. Levitt. M., Hirshberg, M., and Sharon, R. 1997. Calibration and Testing of a Water Model for Simulation of the Molecular Dynamics of Proteins and Nucleic Acids in Solution. *J. Phys. Chem. B.* 101: 5051-5061.
72. Bursulaya, B.D., and Brooks, C.L. 1999. Folding free energy surface of the three stranded  $\beta$ -sheet protein. *J. Am. Chem. Soc.* 121: 9946-9951.
73. Zanuy, D., Ma, B., Nussinov, R. 2003. Short peptide amyloid organization: stabilities and conformations of the islet amyloid peptide NFGAIL. *Biophysical Journal.* 84: 1884-1894.
74. Still, W. C., Tempczyk, A., Hawley R. C., Hendrickson, T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112: 6127-6129.

75. Qui, D.; Shenkin, P. S., Hollinger, F. P., Still, W. C. 1997. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A*. 101: 3005-3014.
76. Dominy, B. H.; Brooks, C. L., III. 1999. Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. *J. Phys. Chem. B*. 103: 3765-3773.
77. Zhang, S., Holmes, T., DiPersio, M., Hynes, R. O., Su, X., Rich, A. 1995. Self complementary oligopeptide matrices support mammalian cell attachment. *Biomaterials* 16: 1385-1391.
78. Aggeli, A., Boden, N., Zhang, S.G. 2001. *Self-assembling peptide systems in biology, medicine and engineering* (Kluwer academic publishers).
79. Venkatraman, J., Gowda, G.A.N. and Balaram, P. 2002. Design and construction of an open multistranded  $\beta$ -sheet polypeptide stabilized by a disulfide bridge. *J. Am. Chem. Soc.* 124: 4987-4994.
80. Zhang, S.G., Alexander R. 1997. Direct conversion of an oligopeptide from a  $\beta$ -sheet to an  $\alpha$ -helix: A model for amyloid formation. *Proc. Natl. Acad. Sci. USA*. 94: 23-28.
81. Pauling, L., Corey, R. B. 1951. Conformations of polypeptide chains with favored orientations around single bonds two new pleated sheets. *Proc. Natl. Acad. Sci. U.S.A.* 37: 729-740.
82. Chou, P.Y. and Fasman, G.D. 1974. Conformational parameters for amino acids in helical,  $\beta$ -sheet, and random coil regions calculated from proteins. *Biochem.* 13: 211-222.
83. Chou, P.Y. and Fasman, G.D. 1978. Empirical predictions of protein conformation. *Annu Rev Biochem.* 47: 251-276.
84. Zhong, H.; Zheng, B.; and Haynie, D.T. 2002. Development of a Graphical User Interface for Computer-aided Fabrication of Polypeptide Microcapsules. First CERT Biomedical Informatics Symposium
85. Garnier, J., Osguthorpe, D.J.; Robson, B. 1978. Analysis of the accuracy and implications of simple methods for prediction of the secondary structure of globular proteins. *J Mol Biol.* 120: 97-120.
86. Kyngäs, J. and Valjakka, J. 1998. Unreliability of the Chou-Fasman parameters in predicting protein secondary structure, *Protein Eng.* 11: 345-348.
87. Suh, J.Y., Lee, Y.T. and Park, C.B. 1999. Structural and functional implications of a proline residue in the antimicrobial peptide gaegurin. *Eur J Biochem.* 266: 665-674.

88. Ma, B. and Nussinov, R. 2002. Molecular dynamics simulations of alanine rich  $\beta$ -sheet oligomers: Insight into amyloid formation. *Protein* 11: 2335-2350.
89. Jorgensen, W.L., Chandrasekhar, J., Madura, J., Impey, R. W., and Klein, M.L. 1982. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79: 926-935.
90. Ryckaert, J.P., Ciccoti, G., and Berendsen, H.J.C. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327-341.
91. Zhi, Z.-L. and Haynie, D.T. Direct Evidence of Controlled Interlayer to Intralayer Structure reorganization in a Nano-organized polypeptide multilayer thin film. *Macromolecules*. In press.
92. Klimov, D.K. and Thirumalai, D. 2003. Dissecting the assembly of A $\beta_{16-22}$  amyloid peptides into antiparallel beta-sheets. *Structure* 11: 295-307.
93. Chothia, C. 1973. Conformation of twisted beta-pleated sheets in proteins. *J. Mol. Biol.* 75: 295-302.
94. Nesloney, C.L. and Kelly, J.W. 1996. Progress towards understanding beta-sheet structure. *Bioorganic & medical chemistry*. 4: 739-766.
95. Hol, W.G. J., Halie, L.M. and Sander, C. 1981. The dipoles of the  $\alpha$ -helix and the  $\beta$ -sheet: Their role in the folding of proteins. *Nature*. 294: 532-536.