Summer 2006

Molecular modeling of proteins and peptides related to cell attachment in vivo and in vitro

Wanhua Zhao
Louisiana Tech University

MOLECULAR MODELING OF PROTEINS AND PEPTIDES

RELATED TO CELL ATTACHMENT

IN VIVO AND *IN VITRO*

by

Wanhua Zhao, B.E.

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

COLLEGE OF ENGINEERING AND SCIENCE
LOUISIANA TECH UNIVERSITY

August 2006

UMI Number: 3259731

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®

LOUISIANA TECH UNIVERSITY

THE GRADUATE SCHOOL

10th July 2006
_____
Date

We hereby recommend that the dissertation prepared under our supervision

by___Wanhua Zhao_____

entitled___Molecular Modeling of Proteins and Peptides Related to Cell Attachment___

_____in vivo and in vitro_____

_____

be accepted in partial fulfillment of the requirements for the Degree of

_____Doctor of Philosophy of Biomedical Engineering_____
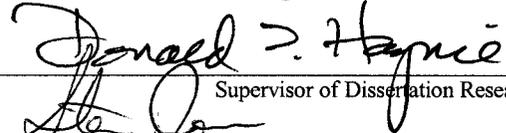
_____
Supervisor of Dissertation Research

_____
Head of Department

Biomedical Engineering
_____
Department

Recommendation concurred in:

_____

_____
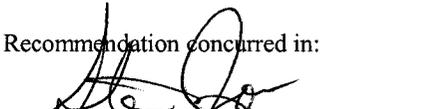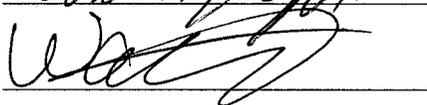
_____   Advisory Committee

_____

Approved:

_____
Director of Graduate Studies

_____
Dean of the College

Approved:

_____
Dean of the Graduate School

ABSTRACT

Polypeptides constitute half of the dry mass of the cell, they form the bulk of the extracellular matrix (ECM), and they are a common element of extra- and intracellular signaling pathways. There is increasing interest in the development of computational methods in polypeptide and protein engineering on all length scales. This research concerns the development of computational methods for study of polypeptide interactions related to cell attachment *in vivo* and *in vitro*.

Polypeptides are inherently biocompatible, and an astronomical range of unique sequences can be designed and realized in massive quantities by modern methods of synthesis and purification. These macromolecules therefore constitute an intriguing class of polyelectrolyte for biomedically-oriented multilayer film engineering (Haynie et al., 2005), Applications of such films include artificial cells, drug delivery systems, and implant device coatings, cell/tissue scaffolds (ECM mimics). The plasma membrane-associated cytoplasmic protein tensin is involved in cell attachment, cell migration, embryogenesis, and wound healing. The tensin polypeptide comprises several modular domains implicated in signal transduction. It has been shown that the N-terminal region of tensin is a close homolog of a tumor suppressor that is highly mutated in glioblastomas, breast cancer, and other cancers.

There are two related areas of development in this work: Polypeptide multilayer films, a type of ECM mimics, and the molecular physiology of tensin. Two studies have

iii

iv

been carried out on polypeptide multilayer films: aggregates of the model polypeptides poly(L-lysine) (PLL) and poly(L-glutamic acid) (PLGA), and interpolyelectrolytes complexes (IPECs) of designed peptides. Molecular models of all known domain of tensin have been developed by homology modeling. The binding properties of the two domain of tensin have been studied.

Molecular dynamics (MD) simulations of PLL/PLGA aggregates suggest that both hydrophobic interactions and electrostatics interactions play a significant role in stabilizing polypeptide multilayer structures. The approach provides a general means to determine how non-covalent interactions contribute to the structure and stability of polypeptide multilayer films. MD simulations of designed polypeptide complexes have been carried out in vacuum and in implicit solvent. The simulation results correlate with experimental data on the same peptides. Energy minimization and MD study of tensin domain-peptide complexes has provided insight on biofunctionality of the tensin molecule and thereby its role in cell adhesion. Such knowledge will be important for determining the molecular basis of cell adhesion in health and disease and engineering treatments of abnormalities involving cell attachment.

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 _____Zhao Wanhua_____

Date _____July 18th 2006_____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

x

ACKNOWLEDGEMENTS

First, I would like to thank Louisiana Tech University for giving me the opportunity to complete my education in the USA.

I would like to express my appreciation and respect for Dr. Donald Haynie for giving me the chance to do my research in computational biology. He has encouraged me to develop a passion for knowledge and to persist in the pursuit of high goals. I thank him for his inspiration and encouragement in guiding me to a deeper understanding of the research subject and his invaluable comments throughout the duration of working on this dissertation.

I also give special thanks to Dr. Steve Jones, the chairman of my dissertation committee, Dr. Walter Besio, Dr. Cheng Luo, Dr. William Campbell and Dr. Andrei Paun for serving as members of my committee and giving me their valuable time.

I am also grateful to everyone in the Bionanosystem Engineering Lab, especially my colleagues in the polypeptide multilayer film group and the tensin group.

I thank my family, especially my mother, for always being supportive of all decisions I have made through my years of education. I thank all the friends who have encouraged me in this journey.

Finally, I would like to thank National Center for Supercomputing Applications and the National Science Foundation for access to computational resources. The work could not have been done without their hardware and software support.

CHAPTER 1

INTRODUCTION

1.1 Biology

1.1.1 Cell Adhesion

Cell adhesion is a fundamental aspect of numerous normal and pathological processes, including cell migration, differentiation, wound healing, inflammation, tumor cell metastasis, and signal transduction [1]. Adhesion involves three broad classes of macromolecules: extracelluar matrix (ECM) molecules, transmembrane adhesion receptors, and intracellular adhesion plaque proteins [2-6]. (See Figure 1.1)

The ECM is a complex structural entity that surrounds and supports cells. It is often referred to as connective tissue. Interactions with the ECM determine the shape, motility, and cytoskeletal rearrangement of a cell *in vivo* [7-10]. Similarly, physical and chemical properties of a tissue culture surface affect the behavior of adherent cells *in vitro*, having a determinative effect on cell development. Hydrophobicity and hydrophilicity [7], surface charge [11-13], roughness [14-16], and topography [17], all influence cell adhesion.

Integrins are adhesion receptor proteins embedded in the plasma membrane of cells. Members of this family of receptor proteins mediate the binding of a cell to the ECM; they play a key role in two-way communication inside and outside the cell.

1

Intracellular adhesion proteins provide structural and functional linkages between adhesion receptors and actin microfilaments and microtubules, i.e., the cytoskeleton. The cytoplasmic domain of integrins bind to cytoskeleton proteins, providing a physical link between the actin cytoskeletal and the ECM [18,19]. Ligation of integrins by the ECM initiates a cascade of intracellular signaling events [3].

Focal adhesions (FAs) are dynamic cellular structures that mediate cell-matix interactions. These structures change in size, shape, and composition during the processes of cell adhesion and migration [20-22]. FAs involve integrin, tensin, and numerous other proteins, for instance, focal adhesion kinase (FAK) [1]. Tensin is translocated to the FA assembly sites by binding to integrins [23]. The foregoing forms a rational basis for the hypothesis that tensin plays a direct role in the mechanics of membrane-cytoskeleton interactions (Figure 1.1).
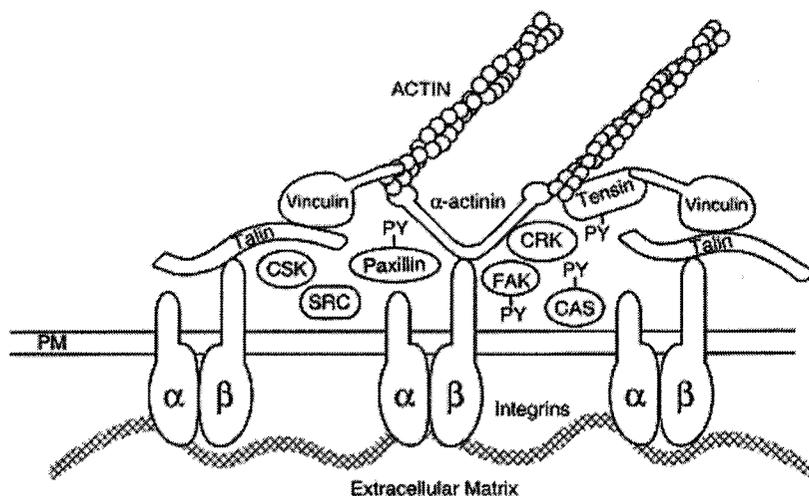
Figure 1.1 Schematic diagram of a focal adhesion. Tensin is shown as binding to vinculin and actin on the basis of biochemical evidence [24]. It is known that the PTB domain of tensin can bind directly to the intracellular domain of β integrin [23]. The figure is from [25].

Two types of adhesion complex are found in the cultured fibroblasts: focal contacts and fibrillar contacts. Focal contacts are sites where a cell attaches to the ECM [2,26]. They are found in the periphery of the cell. Fibrillar contacts are elongated or beaded, mirroring associations with fibronectin fibrils in the center of the cell. In FAs, clusters of integrins bind externally to ECM proteins, and internally to several specialized cytoplasmic proteins, which bind to the "barbed" ends of actin filaments. Integrins interact with numerous proteins at focal contacts and at other sites in the cell. The main apparent functions of an adhesion complex are to transmit tension at the adhesive site and to act as a signaling center. Strong attachment between the cell and the ECM promotes cell growth and gene expression. Abnormal or rapid changes in the morphology, composition, or structure or migration of adhesion complexes during the cell cycle can lead to various diseases including cancer [27].

Integrins regulate the local tension of focal contacts, activate tyrosine phosphorylation, and promote various matrix adhesions in fibroblasts [28]. Phosphorylation of proteins in adherent junctions is critical to the intracellular signaling process [1]. Phosphorylation of proteins can act as a molecular switch to regulate, activate, or suppress a chain of physiological functions inside the cell. Activation of a biochemical process can be achieved by the phosphorylation of a single amino acid, enabling specific binding interactions either with the domains of the same protein or with surrounding proteins at adherent junctions. The regulation and suppression of phosphorylation is achieved by protein tyrosine kinases, protein tyrosine phosphatases, and dual specificity phosphatases.

The ECM comprises a number of different proteins, including fibronectin, vitronectin, and collagen [29]. Cell behavior can be controlled by various bioactive peptides, e.g., growth factors, hormones, and constituents of the ECM, e.g., [30]. A notable example of the last category is fibronectin (FN). (See Figure 1.2) This protein regulates and supports other proteins in the ECM [31-33]. More specifically, FN interacts with integrins in the cell membrane by means of an Arg-Gly-Asp (RGD) signal sequence embedded in the FN polypeptide, promoting cellular spreading and motility. Treatment of cells *in vitro* with RGD-containing peptides can disrupt integrin attachment, e.g., [34].
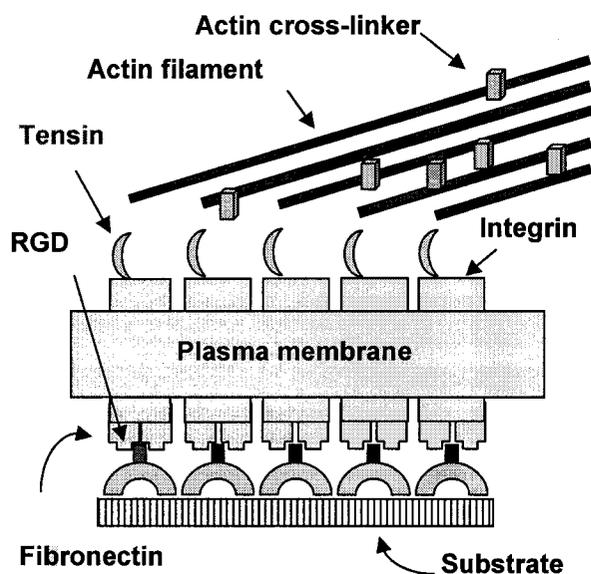
Figure 1.2 Schematic diagram of a cell-ECM contact *in vitro*. The cell secretes various proteins to form the ECM, including FN, binding to the substrate surface. Integrins, proteins embedded in the plasma membrane, bind FN by its RGD sequence. The cytoplasmic part of integrin binds directly to the phosphotyrosine binding domain of tensin and other proteins. Another region of the tensin polypeptide binds to the barbed end of filament-actin. Actin filament boundless are important for the transmission of force generated by myosins striated and smooth in muscle cells and in non-muscle cells. Several actin filaments bundle together to form an actin filament, which is strengthened by cross-linking proteins.

1.1.2 Peptide and Protein Structure

Proteins form one of four classes of biomacromolecule and constitute about half of the dry mass of a living organism [35]. Proteins are the structural building blocks of biomaterials ranging from hair and tendons in mammals to silk produced by insects and spiders. As to size, the well-known globular protein hemoglobin, for example, has a diameter on the order of a few nanometers. The enormous range of possible amino acid side chains, of which the 20 usual ones are but a small subset, makes polypeptides particularly promising for the development of novel biomaterials, for example polypeptide multilayer films (Table 1.1). Degree of polymerization, degree of dispersity, and chemical modification of chain termini or side chains can be controlled, depending on the method of synthesis or purification protocol. Polypeptide chirality is important for biofunctionality and characterization of film structure; it could also play a role in the development of enantioselective films.

Table 1.1 Advantages of polypeptides for multilayer films [36]

Large range of different chemical groups in side chains
Vast number of different combinations of amino acid in a relatively short polymer
Control over synthesis of polymers
Control over contributions of hydrophobicity, hydrophilicity, and hydrogen bonding potential to film structure and stability
Control over secondary structure formation
Control over ability to form "natural" crosslinks
Inherent chirality

Considering the 20 usual amino acids alone, there are $\sim 10^{41}$ distinct chemical structures of unmodified 32-mer peptide. Modern methods of synthesis enable realization in the laboratory of a large proportion of this astronomical range of possibilities.

Important for multilayer film and capsule assembly, some usual amino acid side chains are charged at neutral pH. Other hydrophilic side chains are polar but uncharged at neutral pH, and some side chains are hydrophobic. Inclusion of uncharged amino acids in charged polypeptides influences polymer assembly behavior and film stability by forming hydrogen bonds or hydrophobic interactions [37-39].

A unique feature of polypeptides, one that will certainly influence multiplayer structure and properties, is their ability to form secondary structure. It is known from protein research that various sequences of amino acid show a preference to adopt a type of secondary structure, α helix or β sheet [40]. Both types are stabilized by hydrogen bonds which form between chemical groups in the polymer backbone. The ability of a peptide to fold into a specific structure, the control one can have over peptide sequence, and the range of possible ways of integrating polyelectrolytes with other materials, for example colloidal particles, together provide a remarkable range of opportunities for the design of nanoscale materials. To summarize, hydrophobicity, linear charge density, propensity to form secondary structure at neutral pH, and ability to form chemical crosslinks can be varied according to purpose by design of sequence.

1.1.3 Tensin

1.1.3.1 Introduction

A 220 kD protein, tensin is located at FAs and other transmembrane junctions between the ECM and the cytoskeleton [24,41-43]. Tensin is recruited early in the development of cell-substrate contacts [44], where it binds the cytoplasmic domain of integrin β_1 [23] and caps the barbed end of filamentous F-actin [24]. Tensin bridges the F-actin and integrin components in the supramolecular assembly of the transmembrane

complex. Tensin is present in at least two different pools in a cell, the cytoplasm and FA complexes [45]. Cytoplasmic tensin can suppress cell migration by inactivating downstream regulators. Localization of tensin at FAs is not sufficient to promote cell motility.

Tensin is a substrate of calpain II, an FA protease involved in the assembly and disassembly of the focal contacts [43,46]. Tensin is broadly expressed in mouse embryos and various adult tissues. Cellular development of mice that lack tensin is normal, and the animals are healthy for several months. Later, however, tensin knock-out mice formed multiple cysts in the kidneys. Cyst formation leads to kidney degeneration and renal failure [47]. The role of tensin in wound healing is suggested by the delay in skeletal muscle regeneration of tensin knock-out mice [46,48]. The delay suggests that tensin plays a role in cell migration.

The tensin polypeptide is encoded by a gene located on chromosome 2 in *H. sapiens*, and the intron/exon borders have been defined [46]. There are three regions in the full-length polypeptide: N-terminus, central region, C-terminus. Three signal transduction-related domains have been identified in tensin: a protein tyrosine phosphatase (PTP) domain, a Src homology 2 (SH2) domain and a phosphotyrosin binding (PTB) domain. Protein tyrosine phosphorylation and dephosphorylation is crucial for cell growth, tissue differentiation, inter-cellular communication, and the immune response [49,50]. Tensin also contains a C2 domain, the function of which is not known. The domain architecture of human tensin is shown in Fig. 1.3. The PTP [51-53] were identified computationally by amino acid sequence analysis. The actin-binding region was determined *in vitro* by actin-binding assays [24].
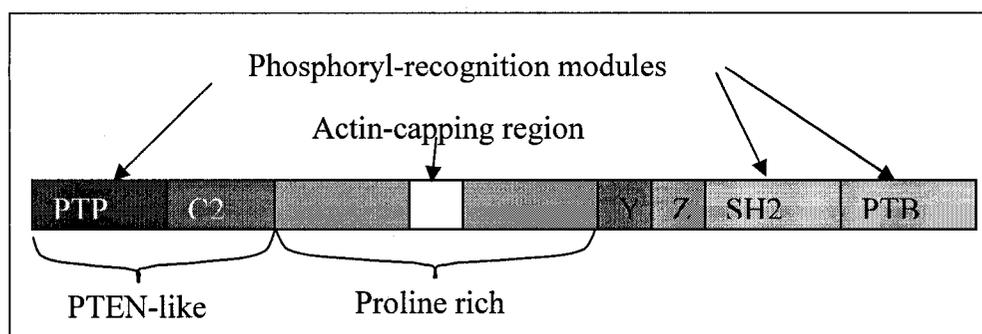
Figure 1.3 Domain architecture of tensin. "Y" and "Z" are putative domains (D.T. Haynie, unpublished results, permitted by D. T Haynie).

The N-terminus of tensin is a homolog of the tumor suppressor phosphatase homolog / tensin homolog (PTEN), also known as multiply-mutated in advanced cancers 1 (MMAC1) [54,55], auxilin [56,57], and cyclin G-associated serine/threonine kinase (GAK) [57,58] (Figure 1.4). The C-terminus of tensin contains two phosphotyrosine binding motifs: an SH2 domain and a PTB domain. Tensin is similar to SH2/collagen homolog (Shc), an adaptor protein that comprises both an SH2 and a PTB domain, but the relative domain organization is different. In view of the relation to Shc, which has no enzymatic activity, tensin may be an "scaffold" protein. This possibility is increased by the hypothetical inactivity of PTP domain [51].

The central region of the tensin polypeptide caps the barbed-ends of F-actin *in vitro* assays [24] (Figure 1.5). Tensin is also able to crosslink actin filaments and is phosphorylated on serine, threonine, and tyrosine residues [42]. In addition, binding of growth factors such as platelet-derived growth factor (PDGF) to cells and activation of oncogenes such as v-src or brc/abl in cells *in vitro* induces tyrosine phosphorylation of tensin [43,52,59].
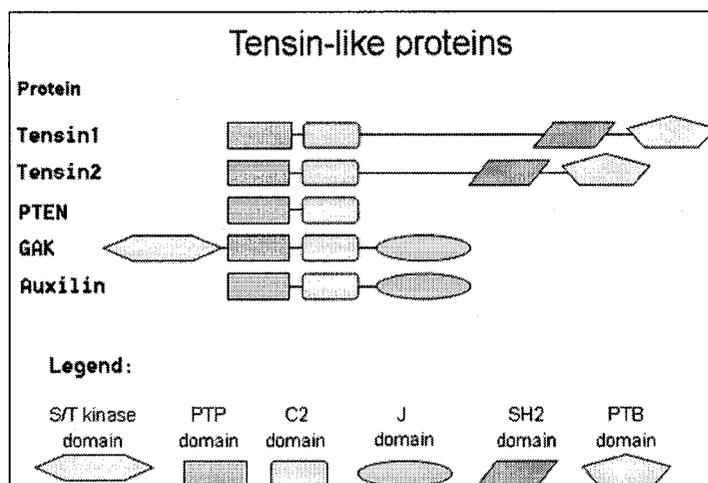
Figure 1.4 Domain architecture of tensin-like proteins (D.T. Haynie, unpublished, permitted by D. T. Haynie).
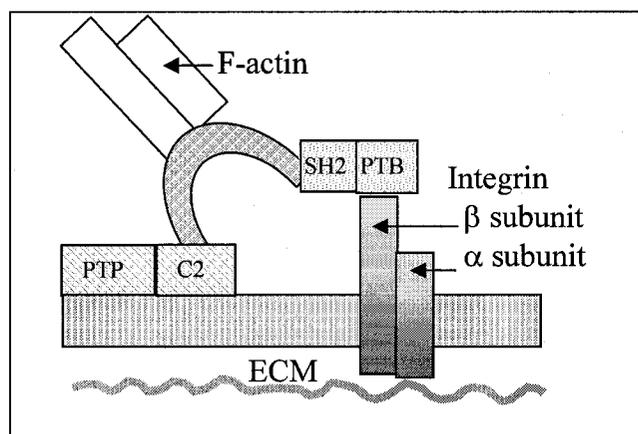
Figure 1.5 Molecular model of interaction of tensin with other focal adhesion components, including the plasma membrane. Integrin interacts with FN in the extracellular matrix and with tensin inside the cell. The F-actin filament bundle will extend into the cytoplasm. (D.T. Haynie unpublished, permitted by D. T. Haynie).

<u>1.1.3.2 PTP domain</u>

PTP domains consist of five central β strands which pack against two α helices on one side and four on the other. PTPs are classified as "receptor-like" or "intracellular" based on cellular localization. Tensin is an intracellular protein. Dual- specific phosphatases (DSPs) are a sub-family of intracellular PTPs. They can dephosphorylate

serine/threonine as well as tyrosine residues. The PTP domain of PTEN, a close homolog

of tensin exhibits DSP activity; that is, it can dephosphorylate both tyrosine- as well as

serine- and threonine- phosphorulated peptides [60]. Moreover, PTEN can

dephosphorylate phosphatidylinositol triphosphate (PtdIns(3,4,5)P$_3$) [61], a lipid involved

in key signaling pathways. Models of phosphotyrosin (pTyr), phosphoserine (pSer),

phosphothreonine (pThr), and inositol(3,4,5)P$_3$ (Ins(3,4,5)P3) are shown in Figure 1.6.
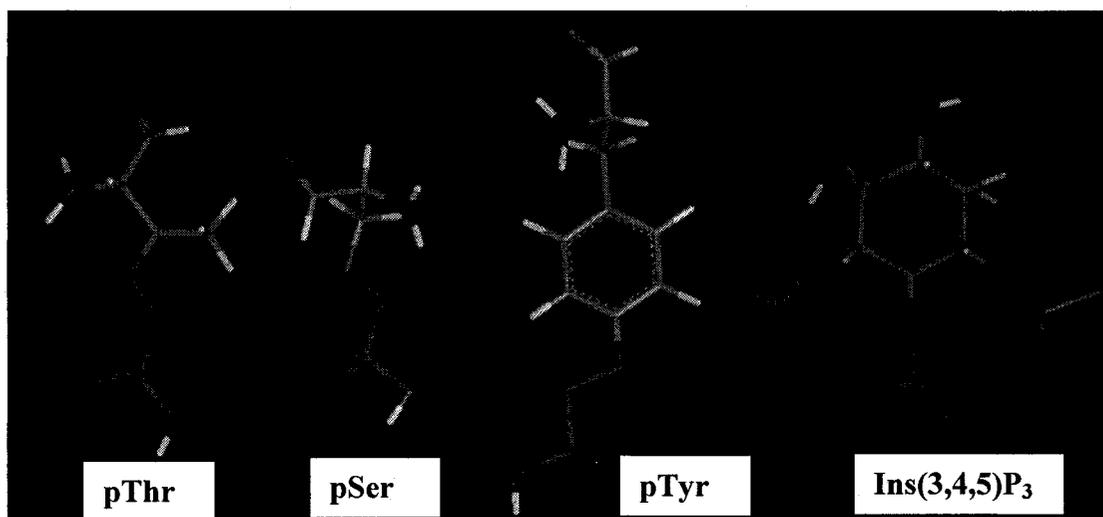
Figure 1.6 Stick model of PTP domain ligands pThr, pSer, pTyr and Ins(3,4,5)P$_3$

Figure 1.7 shows the activie site sequence of PTP tensin, PTEN, and VHR,

known as the PTP signature motif. A key feature of members of PTPs and DSPs is the

sequence:

$$(H/V)C(X)_5R(S/T)$$
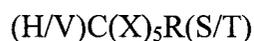
The signature motif forms a loop (called P loop) at the bottom of the active site

pocket. The walls of the pocket are made up of side chain and backbone groups from P

loop residues.

Signature motif	WPD...HCXXGXXRS
Tensin:	WPD...HNKGNRGRI
PTEN:	FED...HCKAGKGRS
VHR:	AND...HCREGYSRS

Figure 1.7 Signature motif of tensin, PTEN, and VHR. Conserved residues are shown in red, the essential nucleophilic cystein residue in blue.

The relative orientation in space of the Cys, Arg, and Ser/Thr side chains are conserved in phosphatases from bacteria and humans. The Cys residue, located at the base of the active site cleft, is essential for formation of a covalent phosphoenzyme intermediate and phosphatase activity. On this basis it has been proposed that the PTP domain of tensin is inactive [51]. Mutation of this residue leads to complete loss of enzyme function. The Arg residue is similarly important for function. The guanidinium group of this residue forms hydrogen bonds to the three non-bridging oxygens in the enzyme-substrate intermediate complex [62,63]. A conserved Ser or Thr residue is located immediately after the Arg residue. The hydroxyl group of this residue is located approximately 3 Å from the active site Cys residue, permitting formation of an S-OH hydrogen bond.

The "active site" of the PTP domain of human tensin is different from the typical signature motif: the essential Cys is mutated to Asn [51]. The probable effect of the mutation is to abrogate activity. Nevertheless, an important aspartic acid, which functions as a general acid or general base in catalysis, is conserved in tensin on a flexible loop approximately 30-40 residues towards the N-terminus of the signature motif. Also conserved in tensin is the signature motif arginine and the signature motif glycine. The latter is important for formation of the turn in the P loop. The view of the foregoing it

seems probable that the basic shape of the binding pocket will not be substantially different from that in PTPs whose structures are known at atomic resolution, and that the PTP domain in tensin will recognize a phosphorylated chemical group.

The principle structural difference between tyrosine-specific PTPs and the vaccinia human-related DSP (VHR) is the depth of the active site pocket. The relatively shallow pocket of VHR can accommodate pTyr or pThr; the deeper pocket of human PTP1B can only accommodate pSer [63]. In the signature motif of PTEN, by contrast, the active site pocket is 0.8 nm deep. The opening is an ellipse 5×11 Å^2. This pocket is wider and deeper than that of 1VHR (a DSP), twice as wide as that of human PTP1B (a PTP), and just as deep as that of human PTP1B. The large width of the pocket is consistent with the relatively large size of $PIns(3,4,5)P_3$, a known substrate. The depth of the active site thus is an important determinant of phopho-amino acid specificity in PTPs.

PTPs are newcomers in the field of drug development. Recent discoveries have revealed that many PTPs play critical roles in a variety of human diseases. Finding the right inhibitors for PTP domain is a new interest in the pharmaceutical industry [64]. Computational studies of binding will complement experimental studies aimed at identifying small molecule inhibitors of PTPs.

1.1.3.3 C2 domain

C2 domains comprise approximately 130 residues and were first identified in protein kinase C [65]. To date, C2 domains have been found in intracellular proteins only. The C2 domain is a member of calcium/lipid-binding domain super family and is involved in binding phospholipids in a calcium-dependent or calcium-independent manner [65]. There are, some C2 domains, however, whose function is not yet known.

The C2 domain fold is a compact β sandwich consisting of two four-stranded β sheets. Despite strong structural similarity, members of the C2 domain family share relatively low sequence identity. Ca^{2+} binding to the C2 domain of synaptogamin induces little conformational change. Instead, ion binding results in change in electrostatic potential, which enhances phospholipid binding [66]. Many C2 domains bind Ca^{2+} through a cluster of aspartic acid residues located in the loop region at one end of the domain. Ca^{2+}-independent C2 domains lack one or more of the calcium coordinating residues.

The C-terminal domain of PTEN has been identified as a C2 domain [67]. The C2 domain in PTEN resembles the C2 domains in phospholipase 1, protein kinase C, and phospholipase A2. These C2 domains are of the type II topology and they are believed to play a regulatory role by mediating the Ca^{2+}-dependent recruitment of the enzymes to phospholipid membranes [65]. The PTEN C2 domain lacks all but one of the Ca^{2+} ligands (Asp268) and thus is hypothesized not to bind Ca^{2+}. The mechanism whereby the PTEN C2 domain binds to the plasma membrane remains unclear.

<u>1.1.3.4 Actin-binding region</u>

This region is P/S/T-rich region. It is much less well conserved between human, bovine, and chicken tensins [46], between various tensin-like proteins, and between human and C. elegans tensins. Some sequences repeat is found in this region, but their function is not known yet.

<u>1.1.3.5 SH2 Domain</u>

SH2 domains are protein modules of about 100 amino acids which are found in a large number of proteins involved in signal transduction [68- 70]. The SH2 domains have so far been found in animals only and apparently do not occur in fungi or plants [71].

SH2 domain has a number of functions, including transducing chemical signals and acting as an adapter between proteins. SH2 domains regulate kinase activity and therby influence a range of cell responses such as proliferation, apoptosis, growth and regulation of enzyme activity [72,73]. SH2 domains have been the targets of extensive drug design efforts [74].

SH2 domains carry out their function by binding with high affinity to pTyr-containing protein targets in a sequence-specific and largely phosphorylation-dependent manner [75,76]. Specificity is conferred by the sequence context of the pTyr within the tyrosine-phosphorylated site, more specifically, by the three residues immediately C-terminal to the pTyr residue [77,78]. The peptide-binding specificity of a large number of SH2 domains has been investigated with libraries of peptides phosphorylated on a Tyr residue and randomized at the +1, +2 and +3 positions C-terminal to the pTyr residue [79,80]. Individual SH2 domains bind to unique sequences (excluding the Src kinase sub-family), while many SH2 domains bind to the sequence pYEEI [81] (Table 1.2).

Table 1.2 SH2 family and their ligand binding partners [81]

Subgroups of SH2 domain	Amino acid sequence of binding partner
Src, Fyn, Lck, Fgr, Abl, Crk, Nck	pY-hydrophilic-hydrophilic-I/P
p8, phospholipase C-gamma, and SHPTP2	pY-hydrophobic-X-hydrophobic
Src sub family (Src, Fyn, Lck and Fgr)	pY-E-E-I

In some cases, an SH2 domain can bind to the target protein or ligand in a non-pTyr dependent manner [60,82]. Such binding could be a property of the SH2 domain in tensin. Alternatively, the SH2 domain of tensin could possibly interact with a pTyr

residue in the same tensin molecule (intramolecular association, as in Src) or in another tensin molecule (intermolecular association, in this case dimerization).

The SH2 domain in tensin is at C-terminus. The optimal phosphopeptide binding specificity of this SH2 domain has been determined to be pY (E or D), N, (I, V, or F) [83]. The localization of tensin requires integrins, talin, and integrin-linked kinase. The N-terminal and C-terminal domain provide essential recruitment signals. The intervening SH2 domain is not localized on its own. Torgler et al. (2004)[84] have suggested a model where tensin is recruited to sites of integrin adhesion via its PTB and N-terminal domains, localizing the SH2 domain so that it can interact with phosphotyrosine-containing proteins and stabilizing the integrin link to the cytoskeleton.

1.1.3.6 PTB domain

PTB domains are also known as phosphotyrosine interaction domains. Each domain consists of 100-150 amino acids. PTB domains are very important in protein-protein interactions and signal transduction [85-88]. The PTB domain was first identified in the signaling proteins Shc and insulin receptor substrate 1 (IRS-1) as an alternative to the SH2 domain for pTyr recognition [88-90]. In proteins like Shc and IRS-1, the PTB domain binds in a pTyr-dependent fashion: The PTB domains of Shc and of IRS-1 bind proteins or ligands containing an NPXpY motif, where pY is the phosphorylated tyrosine residue and X is any amino acid. In other proteins, the PTB domains have been found to participate in pTyr-independent interactions. Member of the X11 family of proteins, for example, contain a PTB domain that binds peptides in a pTyr-independent fashion (Table 1.3). PTB domain binding to pTyr enables the protein to participate in diverse cellular

functions. The PTB domain of tensin binds strongly to the cytoplasmic tails of integrin

β3, β5 and β7 but weakly to β1A [23].

Table 1.3 PTB family and their ligand binding parteners. Residues in bold are known to be involved in binding or are conserved among PTB domain binding targets [91].

Proteins containing PTB domain	Binding Partner	Amino acid sequence of the peptide binding site
Shc	EGFR	SLDNPDYQQDF HIIENPQpYFSDA
IRS-1	IR	LYAASSNPEYLSAS
NUMB	LNX NAX Peptide Screen	GLDNPAYTSSV GLDNPAYTSSV GFSNMSFEDFP
X11	APP	GYENPTYKFFE

The PTB domain of tensin is localized at the FC junction. Sequence analysis and experimental studies on integrin and tensin show that the interaction of integrin and tensin takes place through the PTB domain of tensin [23]. The PTB domain of tensin thus plays a key role in the function of the tensin as an adaptor protein and as a connector between FA molecules and intracellular signaling. Much direct and indirect evidence has shown that tensin acts as an adaptor protein for integrin and is a required component of FAs [92].

1.2 Technology: Layer-by-Layer Assembly

Layer-by-layer assembly (LBL) is a method of making a multilayer thin film from oppositely-charged species [93-97] deposited in succession on a solid support (Figure 1.8). The method has attracted interest because it is both simpler and considerably more versatile than other techniques of thin film preparation, for example, Langmuir-Blodgett deposition. The basic principle of electrostatic LBL, coulombic attraction and repulsion, is far more general than the type of adsorbing species or surface area or shape of support

[96]. Film assembly can be described as the kinetic trapping of charged polymers from solution on a surface [93]. Multilayer film formation is possible because of charge reversal on the film surface after each polyion adsorption step. Film surface charge thus depends on the last adsorbed layer, permitting a degree of control over surface and interface properties. A high density of charge in the adsorbing species will result not only in strong attraction between particles in neighboring layers, but also in strong repulsion between like-charged particles in the same layer. That is, electrostatics both drives film assembly and limits it. Several layers of material applied in succession create a solid, multilayer coating. Each layer of can have a thickness on the order of nanometers, enabling the design and engineering of surfaces and interfaces at the molecular level. Subtle changes in organization and composition can influence film structure and functionality. The layering process is repetitive and can be automated, important for control over the process and commercialization prospects. Constituents of a film could be bioactive or bioresponsive materials.
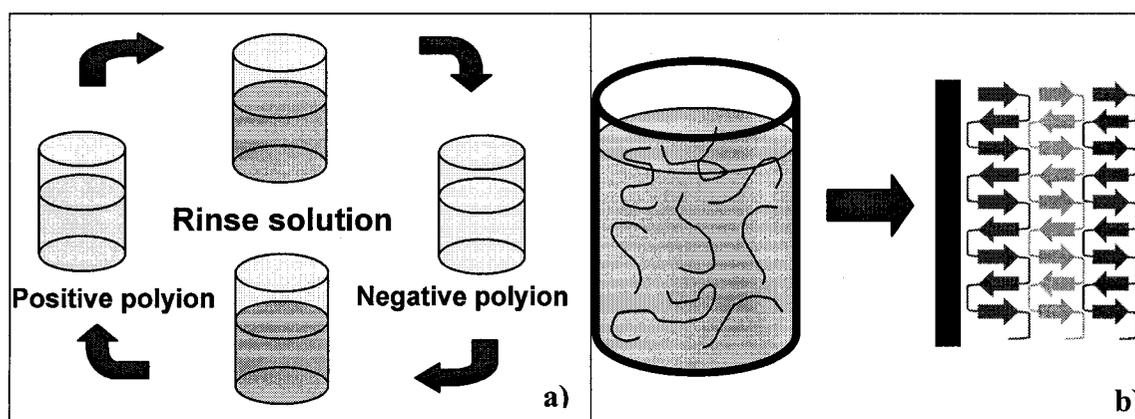
Figure 1.8 Schematic diagram of layer-by-layer assembly. a) LBL procedure b) Rough model of the film made by LBL (D.T Haynie, unpublished, permitted by D. T. Haynie).

The LBL method was pioneered by Iler with rigid colloidal particles [98], based on earlier work by Langmuir [94]. Since the early 1990s there has been considerable interest in exploring multilayer films from linear ionic polymers [93,99]. Such films are being developed for of a variety of applications: for example, contact lens coatings, sustained-release drug delivery systems, biosensors, and functionally-advanced materials with various electrical, magnetic, and optical properties [93-95,100-106]. Many different polyelectrolytes have been studied in this context. Examples are poly-styrene sulfonate, poly-allylamine hydrochloride, poly-acrylic acid, and poly-diallyldimethylammonium chloride. These polymers are called "common" or "conventional" in view of their ready availability from commercial sources and their having been studied extensively. Polyelectrolyte structure, however, would appear to have little effect on whether LBL is possible if the ionic groups are accessible. The polymer chains, once assembled into a multilayer film, tend to become highly interpenetrated [95,107], whether strong polyelectrolytes or weak ones [95]. Besides synthetic polymers, "natural" polyelectrolytes such as nucleic acids, proteins, polysaccharides, and charged nano-objects such as virus particles and membrane fragments have been assembled into multilayer films [97].

Purified proteins, a type of natural polyelectrolyte, have found a role in multilayer films [100]. These "natural" macromolecules present distinct advantages for certain applications of multilayer films, for example, specific functionality. An irregular surface, complex electrostatic properties, pH dependent stability, and relatively large size, however, make proteins less ideal for controlled LBL than one might like. An alternative to proteins is polypeptides designed to meet the electrostatic requirement of LBL but limit

the formation of secondary structure in solution [109]. Films fabricated in this way are envisioned to be useful in a broad range of applications in biomedicine, pharmacology, food science, and other areas [36]. The range of projected applications of polypeptide multilayer films in biotechnology and biomedicine alone encompasses artificial cells, artificial viruses, drug delivery systems, cell/tissue scaffolds, and implantable device coatings.

One of the great challenges of tissue engineering involves attempting to recreate both the cellular and extracellular characteristics of tissue. One approach is to develop artificial extracellular matrices (AECM) from multilayer nanofilms constructed from designed polypeptides [36]. Unlike current ECM protein-based coatings, polypeptide multilayer film-based AECM films are customizable and tunable to specific cell culture functional requirements through control over peptide sequences, type and sequence of layers, type and concentration of embedded molecules, and film surface characteristics. If development of such as AECMs is proven feasible, they will provide researchers and medical practitioners with culture coatings that enable finely-tuned control over the growth and behavior of specific cell types to achieve particular research, diagnostic, and therapeutic objectives. In the present thesis research MD were done on peptides with different sequence and geometrical arrangement to complement experimental work on polypeptide multiplayer films. The related work was described in Chapter 2 and 3. Chapter 2 is related to different sequence used in polypeptide multilayer films. Chapter 3 is about different geometrical arrangements of poly-Lysine and poly-Glutamic acid in polypeptide LBL.

Advantages of polypeptide LBL over other methods of film production are summarized in Table 1.4 and 1.5 [36].

Table 1.4 Why study polypeptide multilayer films

Science	Technology
Physics	**Engineering**
"Unusual" backbone, role of entropy in adsorption	Coatings
Primary structure, role of different interactions	Capsules
Secondary structure, "inherent" nano-scale organization	Self-assembly
Chemistry	Bio-based materials production
"Inherent" covalent crosslinking	
Similarity to protein folding and stability	**Medicine**
Biochemical properties	Tissue engineering
Biology	Artificial cells
"Inherent" Bioactivity	Immunogenicity
Biodegradation	Edibility/biocompatibility
Environmental benignity	

Table 1.5 Properties of polypeptides and of LBL multilayer films

Polypeptides	LbL multilayer films
"Designable"	Nm/A-scale control over thickness
Can be produced en masse in bacteria	Engineered architecture
Susceptible to proteolysis	Arbitrary surface area
Biodegradable	Arbitrary surface shape
Edible	Simple methodology
Environmentally benign	Environmentally-friendly methodology
Sequence-specific immunogenicity	Low-cost methodology
Predictable α helix/β sheet propensity	Can be used to make capsules
Fold into proteins in some cases	Well-suited to an extremely broad range of particles
Specific bioactivity in some cases	Interesting material properties

1.3 Computational Methods

1.3.1 Principles

1.3.1.1 Amino acid sequence alignment

Sequence alignment is usually done computationally. Many programs are available for the purpose. The most popular and probably useful program is the Basic Local Alignment Search Tool (BLAST) [110] and the suites of related programs, e.g., "PSI-BLAST" [111]. All alignment methods aim to achieve the highest similarity in the matched sequence region. Gaps are often introduced to increase alignment. The programs can suggest whether sequences are homologous and what the best alignment of sequences should look like. To establish a functional similarity between sequences, "functional sites" must be identified. Functional sites are usually relatively short, highly conserved sequences. An example is the "signature motif" of PTPs.

Different programs have different estimates of the "cost" of lack of identity between compared residues, and the penalties for gap initiation, elongation, and termination. The "costs" of matches, mismatches, and similarities are derived from the statistics of amino acid substitutions in known homologs; gap penalties are adjusted by various methods and criteria for satisfactory recognition of known homologous proteins.

1.3.1.2 Homology modeling

There is great interest in using computers to predict the 3D structure of proteins when crystallographic or nuclear magnetic resonance spectroscopy (NMR) information is not available. Structural data often greatly enhance the understanding of a protein's function, as well as the search for inhibitors and activators. Experimentally, X-ray diffraction and NMR spectroscopy are the only ways to obtain detailed structural

information. Unfortunately, these techniques involve elaborate technical procedures and many proteins fail to crystallize at all or cannot be obtained or dissolved in large enough quantities for NMR measurements. The size of the protein is also a limiting factor for NMR. At the same time, genome sequencing project produces are providing hundreds of sequences a day. The flow of new 3D structures is smaller by two orders of magnitude; there were only thirty-five thousand structures as of March 2006.

Homology modeling (comparative modeling) is a technique that is used to model protein structure when sequence information only of the target protein is available, but the structure of at least one similar protein is known by X-ray or NMR. Due to the complexity of protein folding, there is currently no known way to go directly from a polypeptide sequence to a protein structure with any real degree of accuracy.

Homology modeling requires at least one sequence with a known 3D structure and with significant similarity to the target sequence. The sequences must be aligned in a way that takes account of secondary structure. The framework of the homology model is computed by averaging the position of each atom in the sequences of template structure, based on the location of the corresponding atoms in the template. Then loops for which no structural information is available in the template structures are assigned either by comparing with loop of the templates or searching for a similar fragment in Protein Data Bank entries. After completing the backbone, side chains are added by searching for the most probable rotamers of each amino acid residue side chain, depending on the backbone conformation. The most favored rotamer is added to the model. Finally, energy minimization with force fields such as CHARMM is done to refine bond geometry and to remove unfavorable non-bonded contacts. It is necessary to keep the number of energy

minimization steps to a minimum to prevent the model drifting away from the control structure. The quality of a model is determined by two criteria: The choice of templates and the quality of the sequence alignment.

Homology modeling is by far the most accurate structure modeling method. Generally, moderate sequence similarity is sufficient for highly similar 3D structure (Figure 1.9). 3D structure is much better conserved than sequence [112]. However, as the number of different amino acids increases, the task quickly becomes more challenging and the results of homology modeling less reliable.
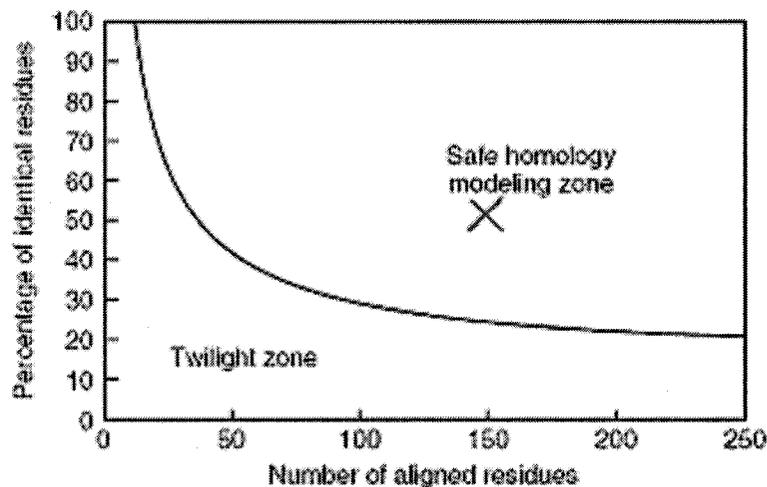
Figure 1.9 Two zones of sequence alignment. The two structures are guaranteed to fold into the same structure if the percentage of sequence identity and number of aligned residues fall into "safe" region. [113].

1.3.1.3 Molecular dynamics

1.3.1.3.1 General view

Molecular dynamics (MD) is one of the most important techniques available for study of a biological system by computer simulation. This method calculates the time-dependent behavior of a molecular system. MD simulations have provided detailed

information on the fluctuations and conformational changes of proteins and nucleic acids. Numerous properties computed from MD simulations can be compared to experimental quantities. The methods are now routinely used to study the structure, dynamics, and thermodynamics of biological molecules and their complexes. They are also used in the determination of structures from X-ray crystallography and from NMR experiments.

The MD simulation method is based on Newton's second law or equation of motion, given by

$$F_i = m_i a_i \qquad \text{(Equation 1.1)}$$

where F is the force exerted on the particle, m is its mass and a is its acceleration. The force can also be expressed as the gradient of the potential energy U,

$$F_i = -\nabla_i U \qquad \text{(Equation 1.2)}$$

Combining these two equations yields

$$\frac{dv_i(t)}{d(t)} = -m^{-1} \frac{dU}{dr_i} \qquad \text{(Equation 1.3)}$$

Newton's equations of motion (Equation 1.1) can then relate the derivatives of the potential energy to the changes in position as a function of time (Equation 1.3). Integration of the equations of motion yields a trajectory that describes the positions, velocities, and accelerations of the particles as they vary with time. Once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. Rather than stepping down an energy gradient, as in energy minimization, MD makes possible an exploration of the energy landscape of the system [114].

MD simulations can be time consuming and computationally expensive, requiring weeks and sometimes months for large, solvated systems even on the fastest computers. As powerful computers become faster and cheaper, however, it has become possible for simulations of solvated proteins to be calculated on the tens of nanosecond time scale with a factor of ~50 less time. The system can be as large as 10^4-10^6 atoms. [115].

MD simulations generate data at the microscopic level, including atomic positions and velocities. The connection between microscopic information and macroscopic observable quantities such as pressure, energy, and heat capacities requires statistical mechanics. Statistical mechanics has rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the molecules and atoms of the system. With MD simulations, one can study both thermodynamic properties (for example, the temperature T, the pressure P, and the number of particles N) and/or kinetic properties (for example, atomic positions Q, momenta P and velocities V).

1.3.1.3.2 Force field

A molecule is considered to be a collection of atoms held together by simple elastic or harmonic forces. The forces are defined in terms of potential energy and are described by mathematical equations as a function of the three-dimensional (3D) structure of the molecule [116]. All empirical force fields have an energy expression similar to a general form (Equation 1.4). Figure 1.10 is the schematic diagram of general format of force field. The exact mathematical expression for each energy term and the parameterizations are the distinguishing features of a force field. The values of the parameters come mainly from experimental structures generated by X-ray crystallography or NMR [114].

Three types of force field were used in this work. CVFF, AMBER and CHARMm. CVFF is a classic force field having a harmonic term and cross-term enhancements [117]. The traditional default force field in the Discover program, CVFF has been used extensively and can be considered well tested and characterized. CVFF was parameterized to reproduce peptide and protein properties. The AMBER (Assisted Model Building and Energy Refinement) force field [118,119] was parameterized with a limited number of organic models. It has been widely used for proteins, DNA, and other classes of molecule. The CHARMm force field (Chemistry at HARvard Macromolecular mechanics) is packaged in a highly flexible molecular mechanics and dynamics engine developed originally in the laboratory of Martin Karplus at Harvard University. It is widely used and can be considered well tested and characterized, e.g., [120].

The general form of force field is

$$E = E_{bond} + E_{angle} + E_{torsion} + E_{oop} + E_{nonbond} + E_{other} \quad \text{(Equation 1.4)}$$

Bond streching

Angle bending

Torsion angle rotations

Out of plane bending
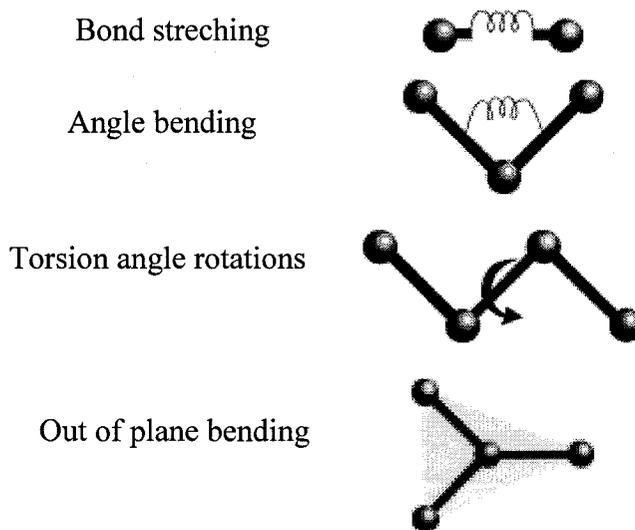
Figure 1.10 Schematic diagram of general format of force field [116].

1.3.1.3.3 Constraints in MD: SHAKE and RATTLE

MD implies numerical integration of the equations of motion. The maximum time step depends on the fastest motions in the system, i.e., the highest frequency motions. These are the bond stretching terms involving hydrogen atoms. These motions determine the largest possible time step. In the SHAKE algorithm, holonomic constraints are placed on the bond lengths: they are made rigid, thereby removing the highest frequency motions from the system. One can either constrain all bonds involving hydrogen atoms or all bonds. SHAKE is an efficient algorithm that incorporates the intrinsic accuracy of the numerical integration algorithm used in MD simulations [121]. It has been shown that application of the SHAKE algorithm to bond stretching degrees of freedom does not affect the overall properties of the system. SHAKE is integrated in CHARMm module of InsightII package. SHAKE was used in the MD simulations discussed in Chapter 3.

RATTLE algorithm is the velocity version of SHAKE during MD runs [122], RATTLE can be used to constrain bonds, angles by two constrained bonds, as well as the distance between any pair of atoms in periodic and non-periodic systems. RATTLE is integrated in Discover module (below) of InsightII package.

1.3.1.3.4 Periodic boundary conditions

Usually, in a simulated system the number of atoms N will be negligible in comparison with the number of atoms in a macroscopic piece of matter (on the order of 10^{23}). Moreover, the ratio of the number of surface atoms to the total number of atoms will be much larger than in reality, causing surface effects to be much more important in the simulations than they should be. Periodic boundary conditions (PBCs) are used to solve this problem. When atoms are enclosed in a box, we imagine that this box is

replicated an infinite number of times by rigid translation in all three cartesian directions, completely filling space. All "ghost" atoms move together, but only one set of them is represented in the computer simulation. Each particle in the box can be thought of as interacting not only with all other particles in the box, but also with their images in nearby boxes [114]. Practically speaking, potentials usually have a relatively short range of interaction. The minimum image criterion can simplify the situation. If the potential energy can take effect only when two particles are within the cutoff distance R_c, then the box size must be larger than $2R_c$. These operating conditions greatly simplify the set up of a MD simulation and therefore are commonly used.

1.3.1.3.5 Solvent

It is very important to consider solvation effects when studying the properties of a biological macromolecule. it has therefore become the norm to study biomolecules in aqueous solution rather than in the gas phase. There is no doubt that a large number of solvent molecules treated explicitly in MD simulations is the most accurate method. Sometimes MD is limited by the amount of real time that can be simulated with current methods and computers. Most of that time is usually spent computing the interactions among water atoms even though it is the solute that is usually the molecule of greatest interest. [114]. This fact has provided a strong reason to replace explicitly represented water molecule with implicit solvent. Implicit solvent models are continuum models that attempt to capture the average effect of the water on a solute. Figure 1.11 presents diagrams of explicit solvent and of implicit solvent.

The continuum "implicit solvent" models have several advantages over explicit water models in MD simulations. For example, the computational cost with implicit

models is considerably lower; the models describe instantaneous solvent dielectric responses, eliminating the need for the lengthy equilibration of water that is typically necessary in explicit water simulations. No viscosity is associated with the implicit water environment; the solute can quickly explore conformation space. Since solvent degrees of freedom are taken into account implicitly, estimating energies of solvated structures is much more straightforward than with explicit water models [123]. The TIP3P model of water and the generalized Born implicit solvent method are the most widely used ones in MD simulations. TIP3P explicit solvent was applied in the work of Chapters 2 and 4. The Generalized Born implicit solvent was used in Chapter 3.
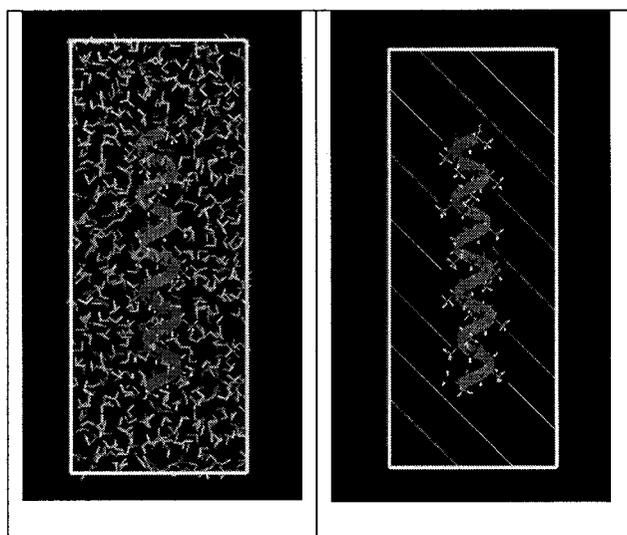
Figure 1.11 Schematic diagram of explicit solvent (left) and implicit solvent (right).

1.3.1.4 Protein-ligand interactions

Protein-ligand interactions are the crucial part of biochemistry. Most physiologic and pharmacological responses such as enzyme catalysis, enzyme transduction, etc., are regulated by specific receptor-ligand interactions. The highly selective receptor-ligand

interactions are known as molecular recognition. Ligand can range from small organic molecules to large biomacromolecules like proteins. The specific protein-ligand interactions are determined by size, shape, and physicochemical property complementarity between the receptor site and ligands, formation of distinctive of specific hydrogen bond. It is of great fundamental and practical interests to understand in atomic detail the nature of these interactions [124].

Molecular modeling techniques have been used to facilitate the biophysical experimental studies. Computational modeling can provide structural and thermodynamic details of ligand-receoptor complexes that are often attainable with experimental techniques. Computational methods are therefore playing an increasingly important role in studying receptor-ligand interactions and in pharmaceutical industry.

1.3.2 Software

1.3.2.1 BLAST

BLAST is a set of similarity search programs designed to search all of the available databases of protein sequences and DNA sequences [110]. The BLAST algorithm detects local as well as global alignments; regions of similarity embedded in otherwise unrelated proteins can be detected. A similarity search by BLAST can provide important clues as to the function of uncharacterized proteins.

There are many different types of BLAST available. Position-specific interated BLAST (PSI-BLAST) is the one used here to search for template sequences for homology modeling of tensin domains [111]. PSI-BLAST produces a position-specific scoring matrix constructed from a multiple sequence alignment of the top-scoring BLAST responses to a given query sequence. This scoring matrix produces a profile

designed to identify the key positions of conserved amino acids within a motif. When a profile is used to search a database, it can often detect subtle relationships between proteins that are distant structural or functional homologues. PSI-BLAST is more sensitive than BLAST; it can detect homologs that are often not detected by a BLAST. PSI-BLAST was used in this work to search more extensively for template sequences for homology modeling of tensin domains.

1.3.2.2 InsightII package

InsightII (Accelrys, Inc., USA) is a comprehensive integrated molecular modeling package for macromolecule building, visualization, simulation, and analysis. The InsightII modules available at Louisiana Tech are Builder, Biopolymer, Homology, Modeler, CHARMm, Discover, Discover_3, Delphi, Docking, Binding Site Analysis and Decipher. Multiple licenses are available for Biopolymer and CHARMm, enabling several simulations to be run simultaneously.

1.   Builder/Biopolymer: Builder and Biopolymer are fully integrated into the InsightII 3-D graphical interface. The two modules are used to build and modify a wide range of biological macromolecules, including proteins, peptides, nucleic acids, and carbohydrates. Molecule construction is the first step in molecular modeling, usually followed by structure refinement and analysis. In the present work, all protein-ligand starting conformations were prepared in Biopolymer. All initial structures of different configurations of designed polypeptide were built by residues in the amino acid library. Atom properties such as atom type, hybridization, potential function parameters, bond order, and geometry were modified in Biopolymer. Builder and Biopolymer were thus used to prepare molecules for subsequent molecular mechanics simulations.

2.   Homology/Modeler: The Homology module can simultaneously optimize both structures and sequence alignments of multiple proteins in the InsightII environment, based on a method developed by Greer [125]. Homology combines automatic model building procedures and user-controlled operations, enabling the user to search protein structure databases for proteins similar to the model being built, find structurally conserved regions, get structures for loops from structure databases or by *de novo* methods, copy coordinates from reference proteins to the model, and refine the new structure by molecular mechanics and dynamics.

Modeler is an automatic homology modeling program developed by Andrej Sali at Rockafeller University [126]. It is integrated into the Homology module in the InsightII package. Before running Modeler, the Homology module allows the sequence alignment and the template sequences to be edited. This is important, as sequence alignment does not guarantee structural similarity. After sequences are aligned, Modeler automatically generates a refined homology model, given only the sequence alignment and a related 3D protein structure. Modeler creates a complete model automatically using a network of empirically-derived spatial restraints and the sequence alignment. Because it optimizes both restraints and stereochemical geometry, Modeler provides an extremely accurate modeling method. Modeler is able to generate excellent structural models given as little as 30 % homology to known structures. The module was used here to model all of the tensin domains.

3.   CHARMm/Discover: CHARMm and Discover can be used to simulate a variety of systems, from isolated small molecules to solvated complexes of large biological macromolecules. The programs carry out energy minimization and molecular dynamics.

It calculates properties such as interaction energy, derivatives, mean square displacements, and vibrational frequencies. It provides tools for doing simulations under various conditions, including constant temperature, constant pressure, constant stress, periodic boundaries, and fixed and restrained atoms. Initial energy refinement of the starting structure complexes of tensin domains and their putative ligands were done with Discover.

1.3.2.3 AMBER

AMBER is the collective name for a suite of programs used to carry out molecular dynamics simulations. The programs important for the present work were LEaP and ANTECHAMBER for molecule preparation; SANDER and PMEMD for simulation runs, and PTRAJ and MM-PBSA for processing trajectories and analyzing results. Amber 8 was used mainly for the polypeptide LBL studies in Chapter 2 and 3.

1.3.2.4 VMD

Visual molecular dynamics (VMD) is a molecular simulation tool for the visualization and analysis of biomacromolecules such as proteins and nucleic acids. VMD can read standard PDB files and display their structure. VMD provides a wide variety of methods for rendering and coloring a molecule: points and lines, CPK spheres and cylinders, licorice bonds, backbone tubes and ribbons, cartoon drawings, etc. VMD can be used to animate and analyze the molecular structural and trajectory files in the Amber format [127]. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign. It is a good companion tool for AMBER.

1.3.3 Hardware

Four computer systems were used locally or remotely to carry out the simulations discussed in this work. They are the following:

1.3.3.1 SGI Origin2000 and 2400

The SiliconGraphics, Inc. (SGI) Origin 2000 and 2400 in Wyly Tower has a 16 microprocessor (MIPRS) 10,000 processor at 250 MHZ and a 16 MIPRS 12,000 processor at 400 MHZ, 10 gigabytes (GB) of random access memory (RAM), and 150 GB of system disk. It was mainly used to simulate the binding of the tensin PTP and PTB domains with putative ligands.

1.3.3.2 SiliconGraphics Fuel R1000

The SGI Fuel V10 Graphics in the Bionanosystems Engineering Laboratory in Engineering Annex has a 600 MHZ processor, R14000A/4MB cache, 512 MB of RAM, an 18 GB system disk, a 21" monitor, 2 GB upgrade of RAM, and a 10,000 RPM internal 3.5" 73 GB hard disk for data storage.

1.3.3.3 SGI Altix 3700 supercluster

The Altix supercluster in the Institute for Microfabrication has 24 Intel Itanium2 microprocessors (1.3 GHz/3.0 MB), 24 GB of RAM, 110 GB of system disk, a 1 Gigabit port Ethernet card, LINUX operating system, a MKL Intel Math Kernel Library 7.0 license, an Intel C++ Compiler with floating license, and an Intel Fortran Compiler with floating license. It supports both the MPI and SHMEM application programming interfaces and global shared memory across partitions or between nodes, and it has parallel processing capabilities. AMBER 8 was installed and run on the Altix system.

1.3.3.4 Xeon Linux cluster

A small allocation grant from the National Supercomputing Applications (NSCA) at University, Illinois at Urbana-Champaign was awarded to do simulation on the Xeon cluster. The system consists of a Dell PowerEdge 1750 server, Intel Xeon microprocessors (3.2 GHz, 533 MHZ system bus with 3 GB ECC DDR SDRAM memory, 512 KB L2 cache, 1MB L3 cache), 70 GB of memory each node, 122 TB of system shared disk, Fortran 77/90/95, and C and C++ compliers. There are 1280 computation nodes. The hardware is primarily intended to run applications of moderate to high levels of parallelism, particularly those codes that perform well in a distributed cluster environment. AMBER was installed and run on this system.

CHAPTER 2

MOLECULAR DYNAMICS MODELING OF DESIGNED

POLYPEPTIDE COMPLEXES

2.1 Introduction

Layer-by-layer assembly (LBL) is a method of making multilayer thin films from oppositely charged species, deposited in succession on a solid support [93-97]. The past 15 years have been witness to considerable growth in research on multilayer films made of linear polyelectrolytes [93,95,97,99]. Such films can be built on planar supports to modify surfaces [93,94,98,100,102-105] or on "sacrificial" core particles to create microcapsules [128-132]. Studies of multilayer films made of polypeptides have recently been initiated for the development of LBL applications in medicine, biotechnology, and other areas [36].

Polypeptides are linear chains of amino acids connected by peptide bonds [35]. The peptide bond is rigid and planar, crucial for formation of "secondary" structures – α helices and β sheets. Polypeptides are "weak" polyelectrolytes: linear charge density can vary significantly with pH. Polypeptides therefore form an interesting class of polyelectrolytes for multilayer film assembly [133]. Important for commercialization prospects of polypeptide multiplayer film technology, a large variety of possible

36

polypeptide sequences can be realized in massive quantities by modern methods of synthesis and purification. In addition, polypeptides are inherently biocompatible, and the multilayer film fabrication process is environmentally benign [36].

Promising biomedical applications of polypeptide multilayer films and microcapsules include coatings for implant devices [134] scaffolds for tissue engineering [135-137], coatings to prevent biofouling [138,139] artificial cells for oxygen therapeutics [140,141] and artificial viruses for vaccine development [36]. The same platform technology could be useful for a variety of non-biomedical applications, for example, biodegradable coatings, coatings of desired surface wettability, and anisotropic coatings. Ref. 36 reviews recent work on polypeptide multilayer films.

In electrostatic polypeptide LBL, film formation is driven primarily by coulombic interactions [134,142-144]. Hydrophobic interactions and hydrogen bonds too, however, can contribute to polymer adsorption and film stability [104,145], the amount depending on polypeptide design [146] and structure of the oppositely charged polymer [136]. At physiological pH most peptides known to be suitable for LBL form films with a comparatively large percentage of β sheet [147-151]. It would appear that β sheet is favored over α helix in this context by the contribution to entropy of the comparatively large variety of ways of forming β structure from a single polypeptide chain [148]. Thickness, surface roughness, density, and structural stability of a polypeptide multilayer film depend substantially on chain length [144], assembly conditions [150,151], and amino acid composition [146,148]. See ref. [133] for a review of the physics of polypeptide multilayer films.

Historically and conceptually, study of polyelectrolyte LBL films is a child of study of IPECs, a special class of polymeric compounds consisting of oppositely charged polyions (see [152] for a review). IPECs form on mixing solutions of oppositely charged macromolecules. LBL multilayer films therefore can be regarded as "layered complexes" of IPECs, the types of interaction being similar on the local level in the two cases. Multilayer films exhibit greater segment density than IPECs and in some cases more extensive stratification of polymer chains [152], but the spatial arrangement of opposing charges in IPECs resembles that in electrostatic LBL films, according to analysis by solid state nuclear magnetic resonance spectrometry, infrared spectroscopy, and other experimental methods [153,154]. Energetics of polyelectrolyte interaction in a multilayer film are expected to show the same trends and magnitudes as IPECs in solution [155]. On this ground we have adopted an IPEC of two particular oppositely charged polypeptide chains as a model of polymer interaction in the corresponding polypeptide multilayer film.

In previous work, six 32mer peptides were designed to probe the relative importance of coulombic interactions, hydrophobic interactions, and hydrogen bonds to multilayer film assembly and stability [146]. Here, we have carried out 52 qualitatively different MD simulations of IPECs of the designed peptides in vacuum and in implicit aqueous solution. The simulations were intended to provide clues on how amino acid sequence might influence peptide complexation, and whether any of the physical observables of simulated IPECs might correlate with the structure, stability, and growth of the corresponding polypeptide multilayer films determined experimentally. We have also developed heuristic models of polypeptide multilayer film assembly in which

peptides are assumed to be in classical β sheet conformation in each layer of the film.

Pertinent experimental findings are summarized in Tables 2.1 and 2.2 [38]. Predictions

based on the models are in remarkably good agreement with experimental results.

Table 2.1 Summary of experimental data. Frequency shift was measured by quartz crystal microbalance for 16-layer films, thickness by ellipsometry for 20 layers, and roughness (granule size) by atomic force microscopy for 20 layers. Data are from [38].

Film structure	Frequency shift (Hz)	Mass deposition	Thickness (nm)	Roughness (nm)	Growth mode
P1/N1	800	Low	4.4	45	Linear
P1/N2	4,700	High	43.9	205	Exponential
P1/N3	1,500	Medium	3.5	105	Linear
P2/N1	6,600	High	72.8	95	Exponential
P2/N2	1,500	Medium	49.9	95	Linear
P2/N3	4,300	High	46.3	150	Exponential
P3/N1	1,400	Medium	14.6	75	Linear
P3/N2	1,000	Low	31.3	160	Linear
P3/N3	500	Low	11.5	140	Linear

Table 2.2 Results of deconvolution of CD spectra after deposition of 15 layers. The data are from [38]. Data not shown for P1/N1 because the film is too thin for sufficient signal -to- noise.

Film structure	Secondary structure content			
	α helix	β sheet	β turn	Random coil
P1/N1	—	—	—	—
P1/N2	0	0.72	0.04	0.24
P1/N3	0.08	0.41	0.21	0.31
P2/N1	0	0.92	0.05	0.03
P2/N2	0	0.19	0.12	0.70
P2/N3	0	0.24	0.19	0.57
P3/N1	0.09	0.35	0.24	0.32
P3/N2	0	0.26	0.23	0.51
P3/N3	0.07	0.37	0.22	0.34

The nine peptide pairs studied here can be divided into three groups on the basis of experimentally determined mass deposited after adsorption of 16 layers: "low" ($\Delta f <$ 1000 Hz), "medium" ($\Delta f \approx$ 1500 Hz) and "high" ($\Delta f >$ 4000 Hz). Frequency shift, Δf, is directly proportional to mass deposited under the conditions of the experiments summarized in Table 2.1.

2.2 Methods

2.2.1 Peptide Design

Basic peptides (P, positively charged at neutral pH) and acidic peptides (N, negatively charged at neutral pH) were designed to study electrostatic interactions (P1 and N1, "Set 1"), hydrophobic interactions (P2 and N2, "Set 2") and side chain hydrogen bonding (P3 and N3, "Set 3") in IPECs [38]. The amino acid sequences of the designed peptides and the IPECs studied here are given in Table 2.3.

Table 2.3 Nomenclature and structure of peptides and IPECs

Peptide[a]	Chemical structure[b]	IPECs[c]
P1	$(KKKK)_7KKKY$	
P2	$(KVKV)_7KVKY$	
P3	$(KVKS)_7KVKY$	P1-N1, P1-N2, P1-N3,
N1	$(EEEE)_7EEEY$	P2-N1, P2-N2, P2-N3,
N2	$(EVEV)_7EVEY$	P3-N1, P3-N2, P3-N3
N3	$(EVEN)_7EVEY$	

[a] P1, P2 and P3 are positively charged, N1, N2 and N3 are negatively charged at neutral pH. [b] K, lysine; V, valine; S, serine; E, glutamic acid; N, asparagine. Tyrosine (Y) is for UV detection of the peptides in the corresponding experimental study. [c] There are nine different combinations of polycation and polyanion.

2.2.2 Model Construction

Thirty-two residue models of the six designed peptides were built with the Biopolymer module of Insight II (Accelrys, USA). Polypeptide multilayer films generally

display a substantial proportion of β structure near neutral pH [147-151]. We therefore

adopted classical β sheet structure as the common starting point for MD simulation of

each peptide complex. Up to four different β sheet configurations of two oppositely

charged peptide sequences were considered in this work: parallel with charged side

chains on the same side of the sheet (PS), parallel on opposite sides (PO), anti-parallel on

the same side (AS), and anti-parallel on opposite sides (AO). Figure 2.1 illustrates these

possibilities. Peptide pairs including either P1 or N1 have but two configurations; there is

no difference between the "same" (S) and the "opposite" (O) configurations in these

cases. Figure 2.2 displays peptide P2 as a β strand with classical dihedral angles.
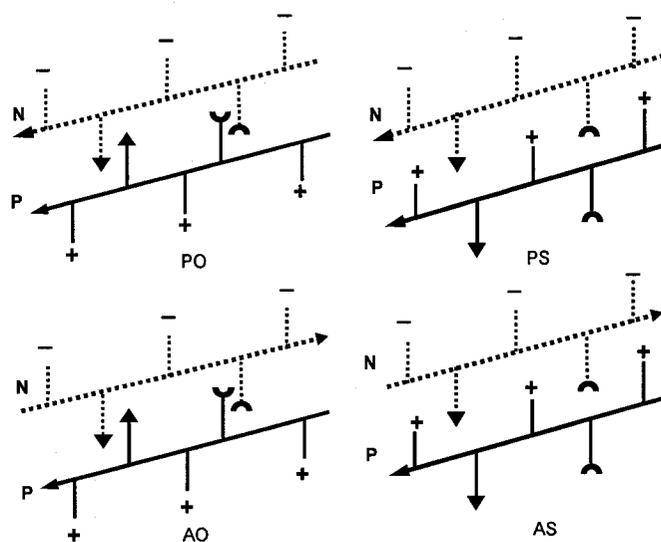
Figure 2.1 Schematic representations of the various configurations of IPEC in sheet conformation. Solid lines, in the plane of the page; dashed lines, below the plane. +, ionized lysine side chain; - , glutamate side chain. Peptides are positively charged (P) or negatively charged (N) as indicated. Relative peptide orientations are abbreviated as follows: P, parallel; A, anti-parallel; S, charged side chains on same side of the sheet; or O, charged side chains on opposite sides of the sheet. ▼ represents the hydrophobic residue Val, ∩ a hydrogen bond donor or acceptor. Hydrogen bonds are formed between donors and acceptors in the polypeptide backbones of each complex (not shown for clarity).
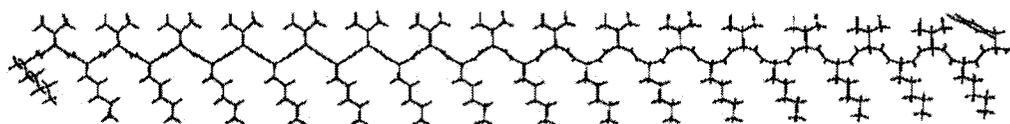
Figure 2.2 Structure of peptide P2. The peptide is shown with classical β sheet bond angles. The terminus is at the left. Lysine side chains point upward, valine side chains downward. The length of the extended chain is *c.* 11 nm. The backbone twists naturally about an axis in the plane of the page.

2.2.3 Simulation Methods

All MD simulations were run with the Amber 8 simulation package [156]. The ff99 force field was employed [157]. Each peptide IPEC configuration was simulated in vacuum and in implicit solvent. Vacuum simulations model a dehydrated polypeptide multiplayer film, implicit solvent simulations a hydrated film. For the latter, the GB solvation model (igb=5) was used [123]. Salt concentration was 15 mM and the pH was 7.4, modeling the corresponding experimental studies [38]. At this pH glutamic acid (Glu, E) is negatively charged and lysine (Lys, K) is positively charged with high probability. The protonation state of Glu and Lys side chains therefore was held ionized. Each IPEC system was energy minimized for 1000 cycles prior to MD simulation, and each residue was constrained to its original position by a harmonic potential with a force constant of 5.0 kcal/mol/Å^2. The system was heated from 240 K to 300 K in 10 ps and weakly coupled to a Berendsen temperature bath at 300 K for 1 ns with a time constant of 2 ps [158]. Lengths of bonds involving a hydrogen atom were constrained with the SHAKE algorithm [121]. The time step for MD was 2 fs.

Fifty-two qualitatively different IPEC simulations have been run, 26 in vacuum and 26 in implicit solvent. Each qualitatively different simulation was repeated at least once. The outcome of a repeat simulation was broadly the same as the original in each

case. Analysis of root mean square deviation (RMSD) of each IPEC (Figure 2.3) and total

potential energy (Figure 2.4) indicated that each IPEC system reached equilibrium well

before 1 ns of MD simulation.
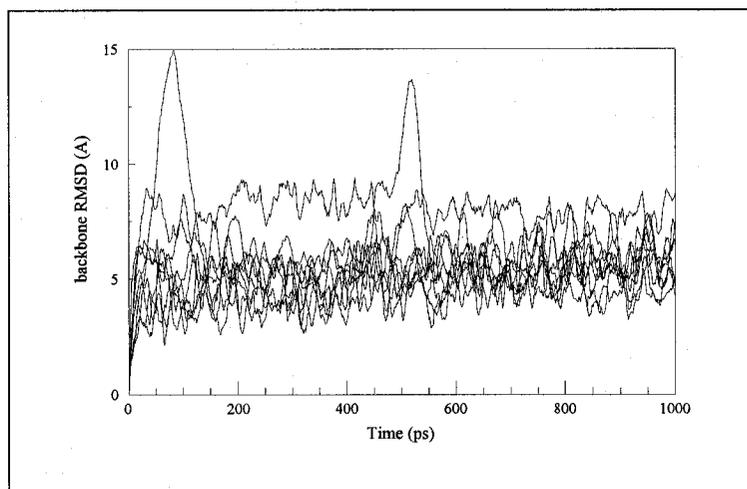
Figure 2.3 Relaxation to equilibrium of the nine PS configurations of designed peptides in implicit solvent. Backbone RMSD with reference to the final structure of the heating step. P2/N3_PS has peaks near 100 ps and 500 ps. Each simulation reaches equilibrium by about 600 ps.
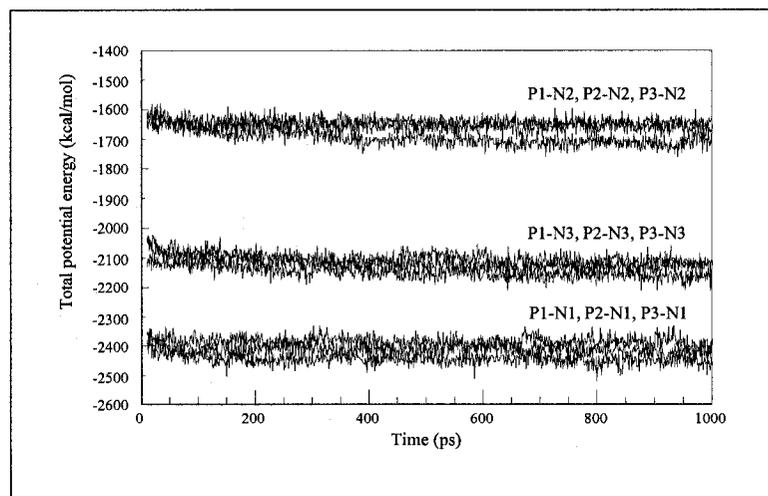
Figure 2.4 Total potential energy of nine PS configurations of designed peptides in implicit solvent during the last 1 ns of simulation. Equilibrium is reached by about 600ps in each case.

2.3 Results and Discussion

2.3.1 MD Results

Figure 2.4 shows the total potential energy of the 9 PS configurations as a function of time. The trajectories form three distinct groups: 1) the anionic polypeptide is N1 and the total potential energy is $c.$ -2400 kcal/mol; 2) the anionic polypeptide is N2 and the potential energy is $c.$ -1600 kcal/mol; 3) the anionic polypeptide is N3 and the potential energy is $c.$ -2100 kcal/mol. Tables 2.4-2.6 show the final structures of the 26 different configurations in implicit solvent and in vacuum. The tables correspond, respectively, to "low," "medium" and "high" mass deposition in multilayer film assembly experiments (see Table 1, of Supplemental Material). The IPECs display a remarkable degree of variety in average conformation at equilibrium.

IPEC conformations at equilibrium can be divided to three categories: "extended" (e.g. P1/N1), "bent" (e.g, P3/N2_AS in vacuum) and "collapsed" (e.g, P3/N2-PO in vacuum). The most extensively "bent" structures and all "collapsed" ones appear in the vacuum simulations. Set 1 combinations, namely, P1/N1 (Table 2.4), P1/N2 and P2/N1 (Table 2.6), P1/N3 and P3/N1 (Table 2.5), exhibit the least substantial changes in structure during MD on the path to equilibrium.

Table 2.4 Average conformation at equilibrium of "low" deposition pairs.[a]

Configuration	Implicit Solvent	Vacuum
P1/N1_P **P1/N1_A**		
P3/N2_AO **P3/N2_AS** **P3/N2_PO** **P3/N2_PS**		
P3/N3_AO **P3/N3_AS** **P3/N3_PO** **P3/N3_PS**		
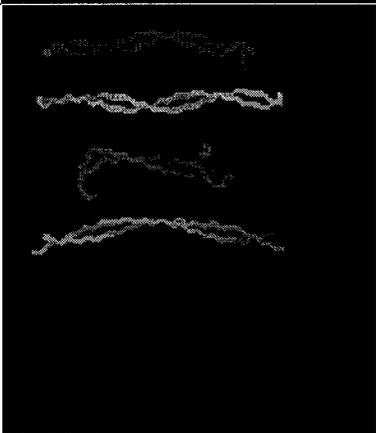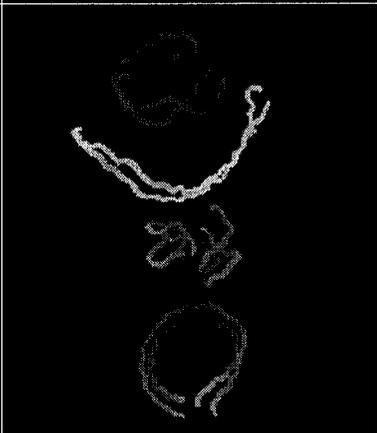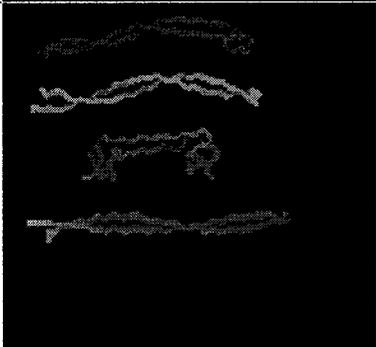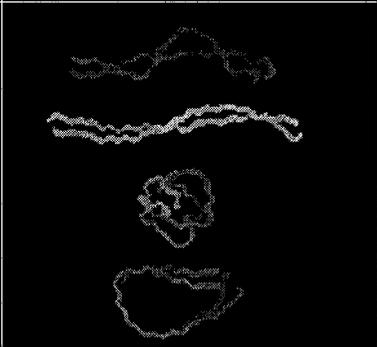
[a]P1/N1, P3/N2 and P3/N3 (see Table 2.1)

Table 2.5 Average conformation at equilibrium of "medium" deposition pairs.[a]

Configuration	Implicit Solvent	Vacuum
P1/N3_P P1/N3_A		
P2/N2_AO P2/N2_AS P2/N2_PO P2/N2_PS		
P3/N1_P P3/N1_A		
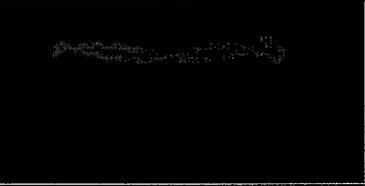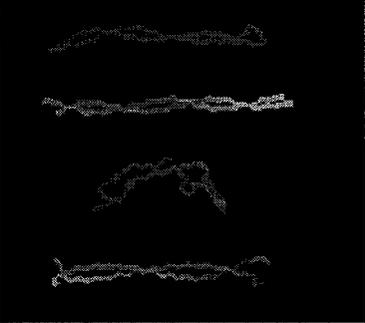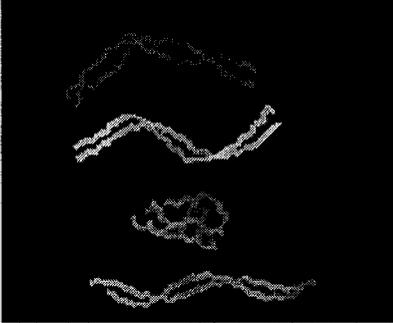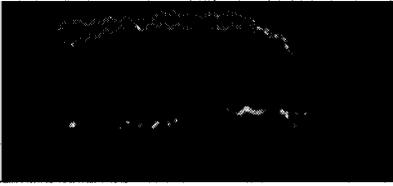
[a]P1/N3, P2/N2, and P2/N2 (see Table 2.1).

Table 2.6 Average conformation at equilibrium of "high" deposition pairs.[a]

Configuration	Implicit Solvent	Vacuum
P1/N2_P P1/N2_A		
P2/N1_P P2/N1_A		
P2/N3_AO P2/N3_AS P2/N3_PO P2/N3_PS		
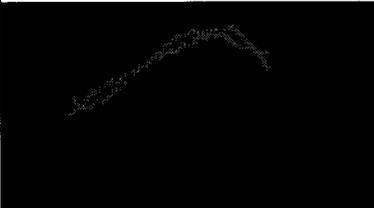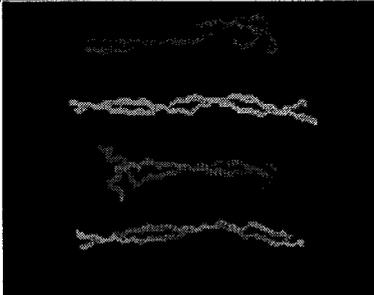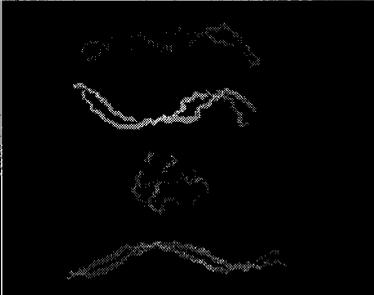
[a]P1/N2, P2/N1, and P2/N3 (see Table 1 of supplemental material).

P1/N1 has the highest charge density per molecule (±1), neutral net charge (0), and the lowest non-bonded potential energy level (Figure 2.5). Strong coulombic repulsion between highly charged side chains makes each molecule of the complex extended, and strong coulombic attraction between highly charged side chains holds the IPEC together. The large negative potential energy indicates that the structure is stable, despite high backbone rigidity; the tightly-bond peptide complex has a relatively low surface area, implying a large increase in entropy on complex formation due to liberation of water molecules. Properties of this IPEC closely resemble the high stability and brittleness of dry polyelectrolyte multilayer films when the linear large density is high.

The simulations would suggest that P1/N2, P1/N3, P2/N1, and P3/N1 – that is, combinations of a member of Set 1 with the oppositely charged member of Set 2 or 3 –

form a distinct group. P2 and N2 have a charge density of 0.5 per residue, and the non-ionized residues are hydrophobic. P3 and N3 are similar to P2 and N2, but half of the hydrophobic residues of the latter are replaced by a hydrogen bond donor or acceptor. P1 and N1 are fully charged at neutral pH. There are differences in "matching" of charge density in the IPECs studied here. The MD snapshots show that P1/N2 and P2/N1 (Table 2.6) are more "twisted" than P1/N3 and P3/N1 (Table 2.5). This suggests that differences in final conformation might be due to mismatch in hydrophobicity between interacting peptides. Peptide pairs in this group have an extended conformation, similar to P1/N1, and the non-bonded potential energy level is in the lower region of the energy spectrum (Figure 2.5a). These peptide pairs also have the lowest solvation energy (Figure 2.5c), implying that they are the most soluble ones in aqueous solution. All these pairs are in the "medium" and "high" deposition groups, suggesting that a low energy value might imply film formation driven primarily by electrostatic interactions. Heuristic models proposed below further help to understand the observed differences in experimental behavior and similarities in MD simulations.

The four pairs without P1 and N1, namely, P2/N2, P2/N3, P3/N2 and P3/N3, show remarkable differences in conformation in vacuum. There is lower conformational diversity in implicit solvent. The differences between IPECs, however, are correlated in the two environments. For example, P2/N2_PO is "bent" and tendency to be "collapsed" at equilibrium in implicit solvent, whereas in vacuum it is fully "collapsed". P2/N2_AO is "bent" in vacuum and implicit solvent, but the extent of bending is greater in vacuum than in implicit solvent. In general, "low" deposition group pairs P3/N2 and P3/N3 (Table 2.4) show a larger degree of conformational change during MD than the "medium" group

pairs P2/N2 (Table 3) and P2/N3 (Table 2.6). This suggests that "collapsed" structure formation might be disadvantageous for film buildup.
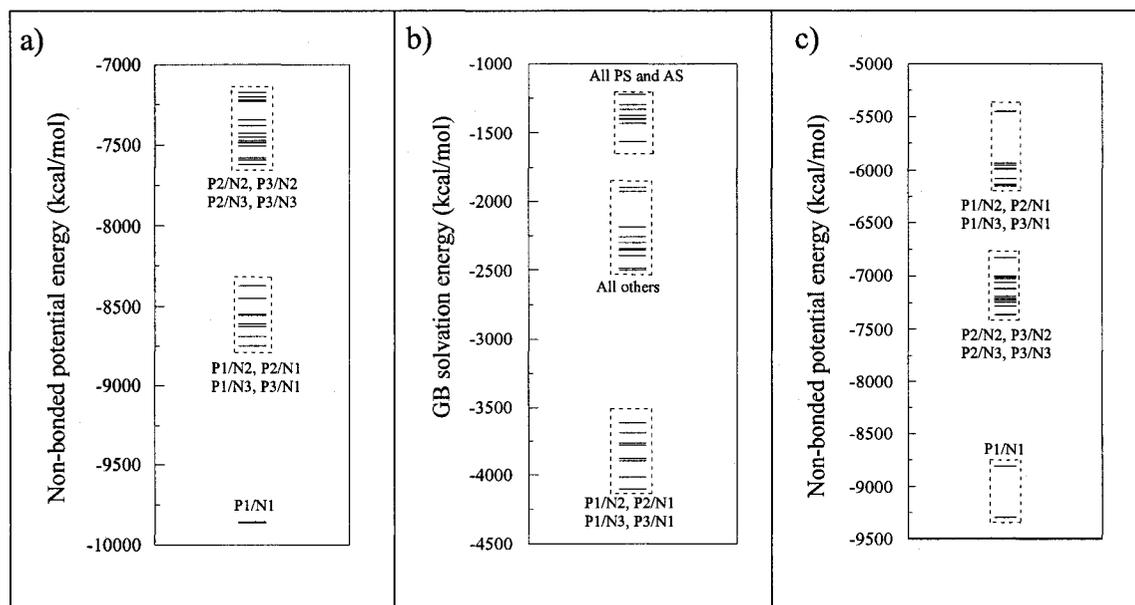
Figure 2.5 Energy spectrum of the 52 qualitatively different simulations. a) Non-bond potential energy in implicit solvent. Three bands are formed: P2/N2, P3/N2, P2/N3 and P3-N3 at c. - 7500 kcal/mol; P1/N2, P2/N1, P1/N3 and P3/N1 at c. -8500 kcal/mol; and P1-N1 at c. − 9850 kcal/mol. Note that every member of the highest energy band is a peptide combination of P2, N2, P3 and N3. b) GB solvation energy in implicit solvent. P1/N2, P2/N1, P1/N3 and P3/N1 form the lowest grouping. c) Non-bond potential energy in vacuum. Three bands are formed in vacuum: P1/N2, P2/N1, P3/N1, P1/N3 at c. − 6000 kcal/mol; P2/N3, P3/N2, P2/N2 and P3/N3 at c. −7000 kcal/mol; and P1/N1 was c. −9000kcal/mol. Non-bonded potential energy is the electrostatic and van der waal part of the total potential energy.

2.3.2 General Discussion

A basic conclusion of mounting data from polypeptide multilayer film studies is that although coulombic interactions provide the main driving force for film assembly, various other types of interaction can participate in the process, notably hydrogen bonding and hydrophobic interactions [35,146,147]. The same types of interaction stabilize native protein structure [35,36]. The participation of multiple types of interaction

in stabilizing a structure resembles the situation with non-polypeptide polyelectrolyte multilayer films [104,145]. Non-electrostatic contributions to film stability can be particularly important with weak polyelectrolytes, especially in a pH range where charge density is low, or with non-ionized but water-soluble polymers. A goal of the present work was to gain insight on how non-covalent interactions might contribute to the physical properties of polypeptide multilayer films. To the extent that such films resemble folded proteins, the results could also inform research on protein folding and stability.

Groundwork in computational study of LBL has recently appeared in the scientific literature. Messina and co-workers, for example, have done Monte Carlo simulations of LBL film assembly from a mixture of oppositely charged polyeletrolytes on a uniformly charged surface or on a spherical charged particle [159-163]. The work is based on the assumption that the final film structure is at equilibrium, though polyelectrolyte adsorption under usual conditions is effectively irreversible [164-167]. MD simulations by Panchagnula et al. concern the sequential adsorption of oppositely charged polyelectrolytes on a spherical charged particle [168,169]. The results appear to confirm that layer build-up in LBL proceeds by surface overcharging during each deposition step [96], and that a multilayer polyelectrolyte system reaches a steady-state regime after a few deposition steps [95]. These simulations are consistent with the view that multilayer formation is based on electrostatic interactions, attraction driving adsorption and repulsion limiting it [96]. Patel et al. have done MD simulations to study influence of degree of polymerization of the polymer chain and fraction of charged monomers on the structure, stability, and mechanism of multilayer formation from dilute

polyeletrolyte solutions at a charged planar surface with discrete charge distribution [170]. Related experiments are reported in ref. [144]

Here, nine peptide combinations in IPECs have been studied by MD simulations in vacuum and in implicit solvent. Observable quantities include potential energy, backbone root mean square deviation, and average conformation during the final 100 ps of MD trajectory. Physical properties of the complexes are compared with each other and with corresponding experimental data on polypeptide multilayer films (Tables 2.1 and 2.2). All films in the experimental studies were dried with nitrogen gas after each adsorption step. The amount of water in the films was minimal. The physical properties of a polypeptide multilayer film can depend significantly on water content. The vacuum and implicit solvent simulations reported here can be viewed as representing two extreme film conditions: dry and wet.

The average potential energy of an anti-parallel configuration (e.g., P2/N2_AS, P1/N1_A) is about the same as that of the corresponding parallel configurations (e.g., P2/N2_PS, P1/N1_P). This suggests that in general the two configurations are approximately equally probable in a dry multilayer film, assuming a constant surrounding environment. The result contrasts to some degree with a study on much shorter peptides by Nussinov and co-workers, in which an energetic preference for the anti-parallel configuration was found [171]. Anti-parallel β sheets have more uniform and apparently more favorable hydrogen bond geometry than parallel β sheets, they can withstand greater distortion (twist and β bulge), they are more likely to be found exposed to solvent in the native state of a protein, and they are more probable if the β sheet contains relatively few residues [172]. These observations, based on crystallographic study of

peptides and proteins, could be supposed to imply that anti-parallel β sheets are more stable than parallel ones, if not in all cases then at least in folded proteins. An extensive analysis of three-dimensional protein structures, by contrast, found no significant difference in the geometry of hydrogen bonds in parallel and anti-parallel β sheets [173]. In the more macroscopic environment of a dehydrated polypeptide multilayer film or even in an IPEC formed from relatively long polypeptides, it may be that other contributions to energetics, for example strong electrostatic interactions between side chains, supersede the presumably greater intrinsic stability of anti-parallel over parallel β sheets.

PO-type configurations (P2/N2_PO, P2/N3_PO, P3/N2_PO, and P3/N3_PO) show the largest degree of conformational change during MD simulation. They are "collapsed" in vacuum without exception. In implicit solvent, the chains both "twist" around each other and are "bent" relative to the starting conformation. By contrast, in vacuum AO configurations (P2/N2_AO, P2/N3_AO, P3/N2_AO and P3/N3_AO) form super-helical structure, the two peptide chains twisting about each other and becoming "bent" to different degrees. There are several possible determinants of the observed differences in conformation. One is side chain. In AS and PS configurations, the charged side chains are on the same side of the β sheet, and strong electrostatic attraction between oppositely charged chemical groups lock the polymers in place, limiting backbone bending (Figure 2.6). In PO and AO configurations, the distance between positively and negatively charged side chains is comparatively large; chain bending could be driven by electrostatic repulsion between like-charged side chains on one side of the IPEC and by hydrophobic attraction between valine side chains on the other side.
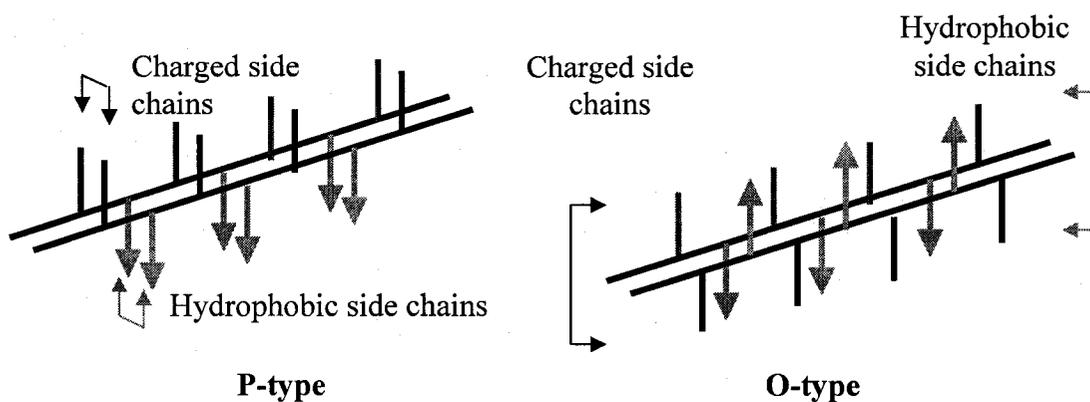
Figure 2.6 Side chain orientations in S-type and O-type configurations lead to different extents of conformation change.

2.3.3 MD with Experiments

Figure 2.7 shows analysis of calculated potential energy values related to average per-residue hydropathy values based on experiments with model compounds.
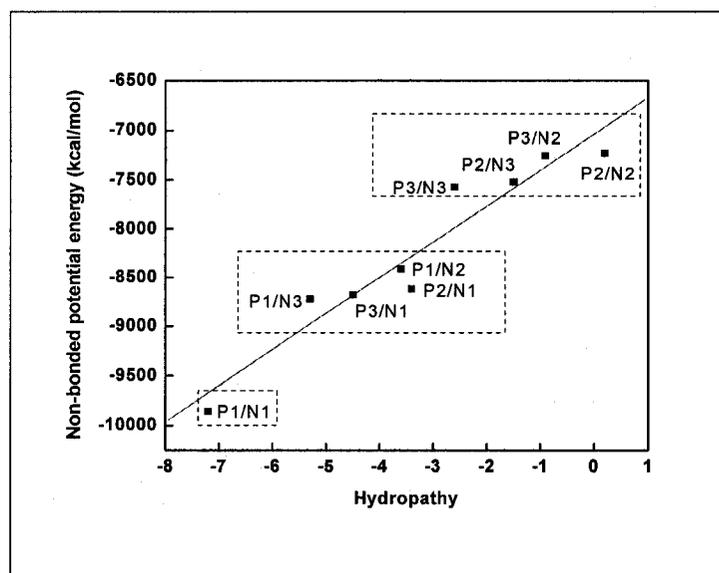
Figure 2.7 Potential energy versus average hydropathy for 9 IPECs. Average hydropathy was calculated with data from ref. 73. N2 has the highest average hydropathy. Generally, the greater the hydropathy, the greater the potential energy. Correlation coefficient $r = 0.96$. Note the clustering of P1/(N2,N3) with (P2,P3)/N1, and of P2/(N2,N3) with (P2,P3)/N2.

Figure 2.7 shows the higher the charge density of a polyelectrolyte, the greater the electrostatic contribution to potential energy of the IPEC, the lower the hydropathy, and the deeper the potential well of the IPEC relative to free peptides in an aqueous medium.

The large conformational changes seen in the MD simulations, assuming they represent the behavior of peptides in experiments, could play a role in the packing of molecules during multilayer film formation. Conformational preferences could also influence film surface roughness.

Figure 2.8 plots calculated total potential energy against surface roughness determined by atomic force microscopy (Table 2.1). IPEC potential energies are clustered in three groups determined by the negative peptide. The order with respect to energy is N1 < N3 < N2. The negative peptides have a more distinctive effect on the electrostatic potential than the positive peptides. All the acidic residues in the negative peptides are glutamic acid residues; all the basic ones in the positive peptides are lysine residues. The lysine side chain, like that of glutamic acid, has a large electrostatic potential at neutral pH, but the hydropathy of the former is higher due to the larger member of methylene groups. The three electrostatic potential groups in Figure 2.8 are also groups in terms of experimental surface roughness. The N1 group has the lowest surface roughness; the N2 group, the highest. In the N1 and N3 groups, surface roughness follows the same pattern seen in total potential energy, namely, P1 (N1 or N3) < P3 (N1 or N3) < P2 (N1 or N3). The order is the same from the point of view of potential energy, namely, P1 < P3 < P2.
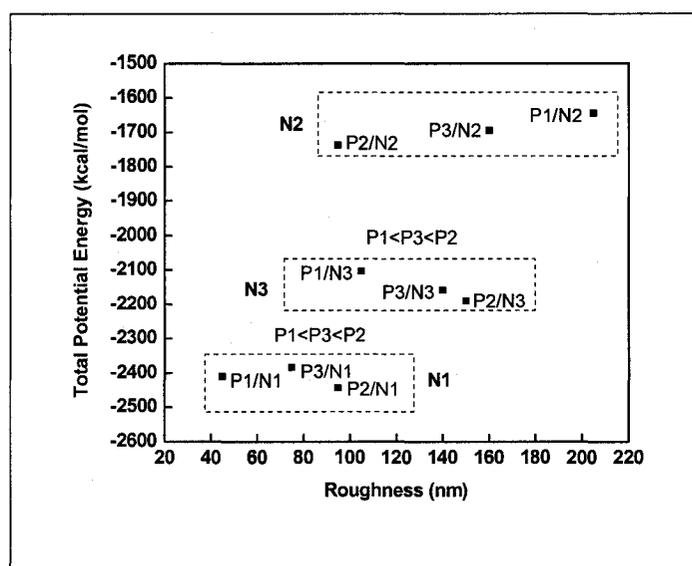
Figure 2.8 Total potential energy in implicit solvent versus surface roughness. With regard to surface roughness, P1 < P3 < P2 in the N1 and N3 groups. With regard to total potential energy, N1 < N3 < N2. Roughly speaking, potential energy correlates with film surface roughness.

We have also studied whether film secondary structure content as revealed by CD analysis might correlate with MD simulation data. Patterns are evident. P1/N2 and P2/N1, for example, have the largest β sheet content of all 9 films: 92 % for P2/N1 and 72 % for P1/N2 (Table 2.2). The MD simulations of these pairs show no collapse of structure at any point along the trajectory. The calculation is consistent with the experimental evidence that these peptide pairs form a large percentage of ordered structure in multilayer nanofilms. Similarly, 41% for P1/N3 and 35% for P3/N1, the middle level of β sheet content, also showed no collapsed structure in simulations (Table 2.5). For lowest β sheet content and highest random coil group: P2/N2 (19%), P2/N3 (24%), P3/N2 (26%) and P3/N3 (37%) (Table 2.2). The final conformation of at least one configuration of each of these four pairs is "collapsed" (see Tables 2.5 and 2.6).

There is also a relationship between calculated solvation energy of a peptide pair and percentage of β sheet determined experimentally. The data suggest that the larger the solvation energy, the more probable the formation of β sheet in the film (Figure 2.9). Similarly, certain peptide structures can form aggregates known as amyloid fibrils. These fibrils are known to be pertinent to several different diseases, for example, Alzheimer's [174]. Fibrillar deposits are organized in cross-β-sheet structures [175]. Modeling studies of the amyloid peptide of Alzheimer's diesease have been done by Nussinov and coworkers [171,176,177].
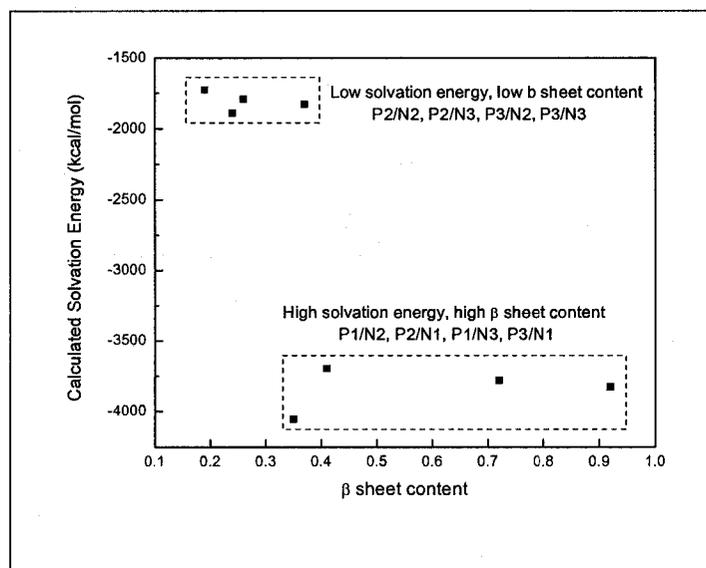
Figure 2.9 Relationship between calculated solvation energy of peptide pairs in implicit solvent and proportion of β sheet in the film. P1-N1 is not shown. The clustering of P2/(N2,N3) with (P2,P3)/N2, and of P1/(N2,N3) with (P2,P3)/N1. The former group of IPECs have a net charge of 0, whereas IPECs of the latter group are only partially charge-neutralized.

In summary, MD simulations have resulted in several relatively distinct types of energetically preferred of peptide IPEC. It is likely that a range of conformations of a complex will be present in a multilayer film for any particular combination of

polypeptides. Some of the likely possible film conformations presumably are suggested by the MD simulations. The simulations have also informed the development of models of designed peptide multilayer films.

2.3.4 Heuristic Models

Experimental study of the designed peptides discussed here (Table 2.1) has revealed that β sheet is the dominant secondary structure comparing to helix in multilayer films at neutral pH (Table 2.2). Figure 10 presents simple models of how the peptides could be organized on a substrate and how the films might grow layer-by-layer. The models, despite their simplicity, help to rationalize observed differences in experimental assembly behavior by providing a microscopic view of the physical basis of multilayer film.

In P1/N1 (Figure 2.10a), the attraction between layers is predominantly electrostatic in character, as in common polyeletrolyte LBL [96]. Each adsorption cycle results in the deposition of one layer of positive chains (P1) or one layer of negative chains (N1). Experimentally determined mass deposition is limited and the film is thin for a given number of layers (Table 2.1), consistent with many non-polypeptide polyelectrolyte multilayer film studies in which the charge density of both polyelectrolytes is high (see [95] for a review). The model is in good agreement with the corresponding experimental data. The high charge density of these peptides results in an extended conformation in the simulated P1-N1 IPEC (Table 2.4) and in a large negative potential energy (Figure 2.5), low hydropathy (Figure 2.6), and small surface roughness (Figure 2.8).

Films P2/N1 and P1/N2 are comparatively thick, and both grow supralinearly with regard to mass deposition (Table 2.1). Similar models are proposed in each case: 3 molecule-thick layers for the two polymer deposition steps of a complete adsorption cycle (Figures 2.10b and 2.10d). Electrostatic between oppositely charged chains and hydrophobic interactions between identical molecules stabilize film structure. The plausibility of the model is increased by knowledge that P2 oligomerizes in solution [176]. Bonds between layers form with relative ease and efficiency, yielding rapid film growth with adsorption step and ordered film structure. The growth curves obtained by quartz crystal microbalance analysis are consistent with the proposed models [38]. Differences in mass deposition are attributable to differences in mass increment per adsorption step. Each of the peptide pairs has but one film growth mechanism.

As to P1/N3 and P3/N1, P3 or N3 forms a double layer on adsorption and P1 or N1 forms a single layer (Figures 2.10c and 2.10k). Side chains with hydrogen bonding potential are present in P3 and in N3, in addition to hydrophobic side chains. The stability of films involving these peptides, however, is expected to be lower at neutral PH than films involving P2 or N2, because a donor and an acceptor must be matched in space in order to form a hydrogen bond, and hydrophobic and hydrophilic residues are close together in P3 and N3, decreasing film density. Experiments are in good agreement with this interpretation (Table 2.1).
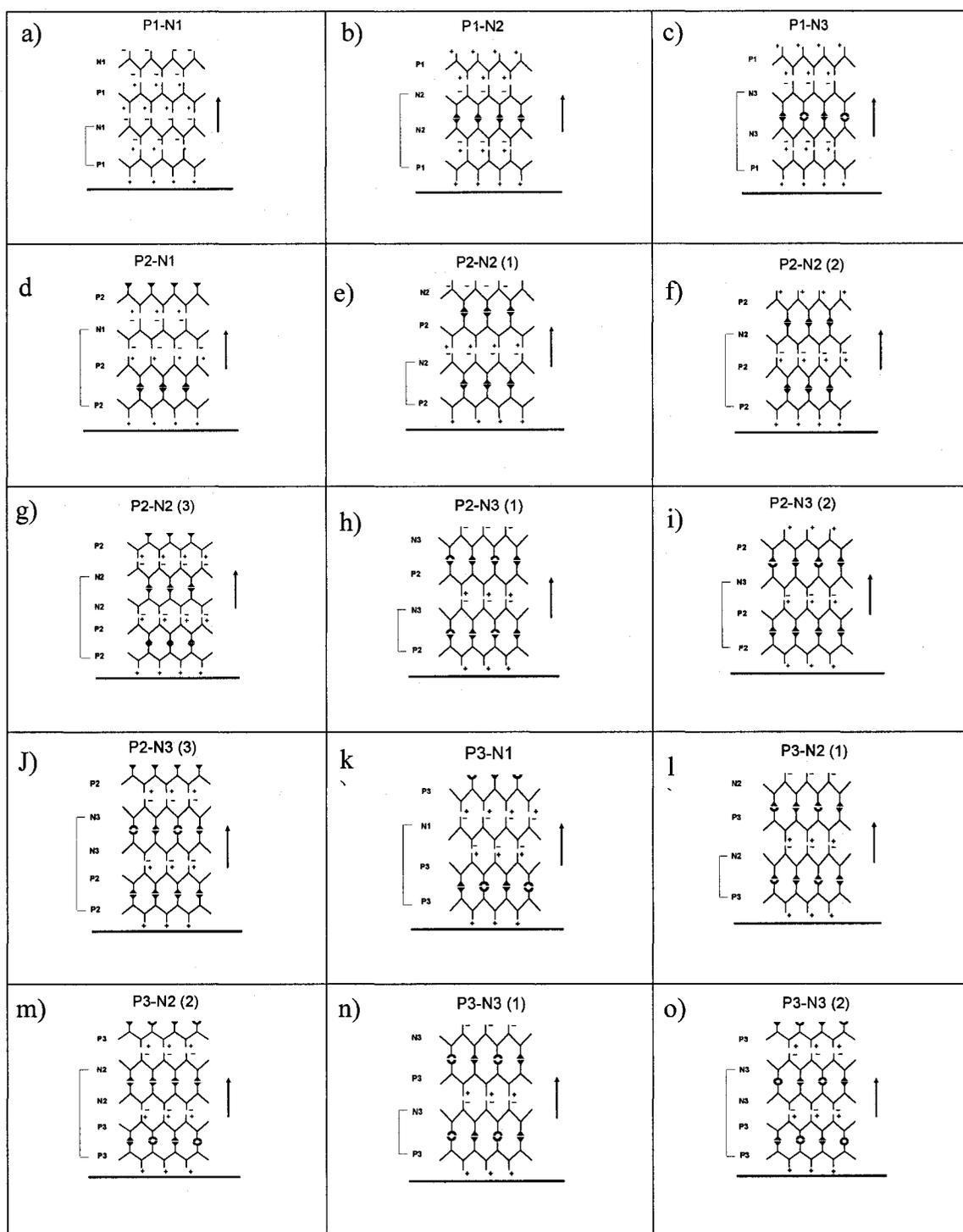
Figure 2.10 Heuristic models of multilayer film assembly of different combinations of designed polypeptide on a negatively charged substrate. +, charged lysine side chain; ⁻, glutamate side chain. P, positively charged chain; N, negatively charged chain. ▼, side chain of the hydrophobic residue Val, ∩, side chain with a hydrogen bond donor or acceptor. ↑, direction of film growth. [, one "bilayer" in the film assembly process.

The various possible adsorption modes for P2/N2 suggest extensive but not necessarily ordered deposition, as described above. MD simulation of the possible P2/N2 β sheet complexes has shown that two out of four configurations are "bent." To the extent that the simulations reflect the structures that actually form on peptide complexation, such bending will influence the strength of electrostatic and hydrophobic interactions between molecules in a layer and between layers in a multilayer film.

Three models are proposed for P2/N3. There are two, three, or four layers for two adsorption steps. The driving force for peptide assembly in each case is a combination of coulombic and hydrophobic interactions. The actual process is likely to be a combination of the three models and other modes of adsorption. As P2 is likely to form oligomers in solution, models 2 and 3 are the more probable ones. Experiments show high mass deposition and exponential growth (Table 2.1), consistent with models 2 and 3. The pattern is similar to that of P1/N2 and P2/N1. P2/N2 and P2/N3 are the only pairs with as many as three obvious possible mechanisms of film growth (Figures 2.10e-2.10g and Figures 2.10h-2.10j, respectively).

The analysis of P2-N2 and P2-N3 would suggest that the degree of order in a polypeptide film is related to number of plausible film growth mechanisms – This would be a source of entropic stabilization of a film. Experiments show more α helix in cases of several plausible film growth mechanisms than in cases of one mechanism (Table 2.2 and Figure 2.10). In any case, the relative amount of β sheet structure in a film at neutral pH can be taken as an indicator of how well ordered the film is.

There are two models for P3/N2. Besides electrostatic interactions, hydrophobic interactions form between N2 and N2, P3 and N2, and P3 and P3; hydrogen bonds form

between P3 and P3, and P3 and N2. Hydrogen bonds are relatively weak electrostatic interactions, and side chain-side chain hydrogen bonding will depend on a donor finding an acceptor in the film. Closeness in space of hydrophobic valine and polar serine side chains would make the film loosely packed and unstable. This might help to explain why so little mass is deposited for this peptide pair, even though multiple layers can be formed in a single adsorption step.

Two models are proposed for P3/N3. This pair shows the least material deposited in the corresponding experiments (Table 2.1). The model helps to illustrate inefficiencies of interaction between non-charged side chains of P3/N3, P3/P3, and N3/N3; the hydrophobic groups and hydrogen bond donors and acceptors all must be specifically positioned for good surface complementarity in the film. Specificity entails a large reduction in entropy during film formation and is therefore unfavorable.

The models presented here help to rationalize experimentally determined properties of the designed peptides [38]. The models are obviously simplistic. Nevertheless, they form a remarkably sound basis for accurate prediction of film growth. For example, the models predict an integral and determinable number of layers of polymer deposited per adsorption cycle for the various peptide systems (manuscript in preparation). Moreover, the models are consistent with the corresponding MD simulations.

2.4 Conclusion

This work constitutes an introductory computational analysis of polyelectrolyte multilayer films. We have simulated the interaction of oppositely charged polypeptide complexes in vacuum and in implicit solvent. Varying peptide sequence and

configuration in the starting structure enabled analysis of the dependence of simulation trajectory on these variables. Little difference was found between parallel and anti-parallel β sheet conformations from the point of view of potential energy and number of hydrogen bonds. This suggests that anti-parallel and parallel β sheets are approximately equally probable in the corresponding polypeptide multilayer films. Hydrophobic groups and polar residues influence final conformations of IPECs in MD simulations and multilayer film assembly in experiments. Heuristic models of film structure help to connect experiments and simulations. Experimental film properties such as surface roughness and β sheet content correlate with computational observable quantities such as potential energy and solvation energy. This work will form the ground of future research on polypeptide multiplayer nanofilms. The results will also be of interest to the broader field of peptide and science.

CHAPTER 3

MOLECULAR DYNAMICS MODELING OF

POLYPEPTIDE MULTILAYER FILMS

3.1 Introduction

Polypeptides in solution can form ordered structures known as α helices and β sheets [178]. In proteins these structures are stabilized by hydrogen bonds and by hydrophobic interactions with surrounding atoms [35]; an individual secondary structure will usually be too small to persist structure in aqueous solution, except under unusual conditions, even if intramolecular hydrogen bond formation is energetically favorable [172]. Nevertheless, secondary structures form the base of the three-dimensional structure of a protein.

Recently, PLL and PLGA have been studied as model peptides in multilayer film fabrication at neutral pH (see [36] for a review). The average charge per monomer under these conditions is high. Analysis by circular dichroism spectroscopy has shown that a (PLL/PLGA)n multilayer film at pH 7.4 contains ~35 % β sheet structure [144,147]. The susceptibility of polypeptide multilayer film structure to environmental perturbation has been probed [148]. It had been found that the adsorption of polypeptides in LBL occurs

63

on a timescale of minutes [146,149]. It therefore would be unrealistic to attempt a full-blown simulation of repeated layer formation from randomly oriented polypeptides in solution; at least for now. On this basis we supposed that initial MD simulations of polypeptide LBL thin films might usefully probe the stability of aggregates of low molecular weight PLL and PLGA in β sheets conformation.

MD studies of aggregates of short identical polymers representing a fragment of the β amyloid peptide have appeared in scientific literature [171,176]. The amyloid peptide self-aggregates in aqueous solution into fibrils containing a large percentage of β sheet. Here, we have adopted the approach of Zauny *et al* [171], and studied multiple peptide models of PLL/PLGA aggregates by MD simulation. The initial structures in our simulations are ideal β sheets, based on the cited experimental evidence for PLL/PLGA multilayer film structure. The results provide insight on the internal structure of a polypeptide multilayer film at atomic resolution; the relationship between hydrophobic interactions, electrostatic interactions, and hydrogen bonds in stabilizing such films; and the relative stability of different possible arrangements of strands in a film containing a large percentage of β sheet.

3.2 Methods

3.2.1 Model Construction

Two peptide sequences, $(Lys)_5Tyr$ and $(Glu)_5Tyr$, have been tested, each six amino acids long (Figure 3.1). Tyr was included to model peptides studied experimentally in our laboratory [144,146]; this amino acid is useful for spepctroscopic detection near 280 nm [148]. Each of the peptide aggregate models tested is a combination of $(Lys)_5Tyr$ and $(Glu)_5Tyr$ (Figure 3.1). Dimer 1 and Dimer 2 consist of

one negative peptide and one positive peptide at neutral pH in an anti-parallel β sheet and a parallel β sheet, repsectively. Trimer 1, Trimer 2, and Trimer 3 have two peptides of the same charge and one of the opposite charge. All three peptides in Trimer 1 are in anti-parallel β sheet configurations; in Trimer 3 all are parallel; in Trimer 2 there is a mixture of anti- parallel and parallel. There are four different of 4-peptide models: Tetramer 1 and Tetramer 2 contain two "stacked" anti-parallel β 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 β strands within a sheet, but the sheets are parallel or anti-parallel relative to each other, respectively. We have also studied four different hexamers. Hexamer 1 and Hexamer 2 are "3-layer" models. The former contain three 2-standed anti-parallel β sheets, each layer parallel to the others, and the latter three 2-stranded anti-parallel β sheets, each layer anti-parallel to each other. Hexamer 3 and 4 are "2-layer" models. Hexamer 3 comprises two 3-stranded anti-parallel β sheets, one layer is anti-parallel to the other. Hexamer 4 is the same as Hexamer 3, except that the two layers are parallel to each other. Hexamers 1 and 4 on the one hand, and Hexamers 3 and 2 on the other hand, represent the same elementary structural concept and comprise the same number of β strands but differ with regard to number of β sheets. The distance between any two peptides in a β sheet was always 4.7 Å, and the distance between two sheets was always 10 Å, close to the average distances for β sheets in crystallographic structures [171,179].

	1	2	3	4
Dimer	4.7		N/A	N/A
Trimer	4.7 4.7			N/A
Tetramer	10 4.7			
Hexamer	10 10 4.7			

Figure 3.1 Schematic representation of dimers, trimers, tetramers, and hexamers. Black and white represent positively charged and negatively charged peptides, respectively. Distances are given in Ångstroms. Two views are shown for Dimer 1 and Dimer 2: ribbon diagrams and the representation introduced here. In keeping with the usual convention, arrows point from the N terminus to the C terminus. A dot (•) signifies the chain coming out of the plane of the page; a star (✦), going into the plane. The distance between two peptides within a β sheet is 4.7 Å, and the distance between two sheets is 10 Å. Hexamer 1 and Hexamer 2 have three layers; Hexamer 3 and Hexamer 4 have two layers. In each extended peptide backbone, the corresponding dihedral angels are $\varphi = -139°$, $\psi = 135°$ for anti-parallel β sheet, and $\varphi = -119°$, $\psi = 113°$ for parallel β sheet. These values are based on crystallographic structures.

3.2.2 Simulation Methods

AMBER 8 and the ff99 force field were used for all MD simulations [156,157].

Peptides were solvated in a period box of TIP3P water molecules, extended in all

directions so that the minimum distance between any atom of the solute and the box wall was at least 15 Å [180]. The cutoff distance for non-bonded interactions was set to 15.0 Å. The particle mesh Ewald approach was used to treat long range electrostatic effects [181]. The pH value for peptides was 7.4: $(Glu)_5Tyr$ was negatively charged and $(Lys)_5Tyr$ was positively charged.

Each fully solvated system was energy minimized by 200 steps of steepest descent followed by 800 steps of conjugate gradient method. The time step for MD simulation was 1 fs. Periodic boundary conditions were applied. Bonds involving hydrogen atoms were constrained to equilibrium lengths by the SHAKE method [121]. The temperature of each system was heated gradually from 240 K to 350 K and then held at 350 K for 200 ps at constant temperature and volume. An additional 1 ns of constant pressure and temperature simulation was done at 350 K for data collection at an interval of 1 ps. The thermal energy was higher than at room temperature to check the stability of the peptide aggregates in a limited time period under "thermal stress" [171,176].

Simulation results presented here have been evaluated by spatial observable quantities defined by Ma and Nussinov [176]. Structural changes in the peptide models were characterized by average head-to-head distance, tail-to-tail distance, and distance between centers of mass (Figure 3.2). Two types of spatial observable were measured: within a sheet (local structure) and between sheets (non-local structure). Hydrogen bonds were calculated and counted. These measures were used to judge the structural integrity and stability of the various peptide aggregates. Final average snapshots are presented to give a more intuitive view.
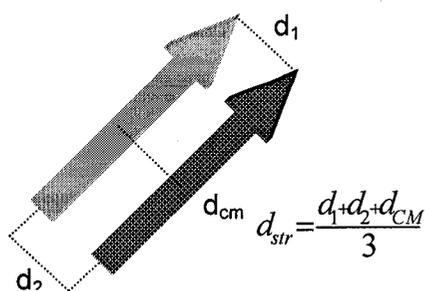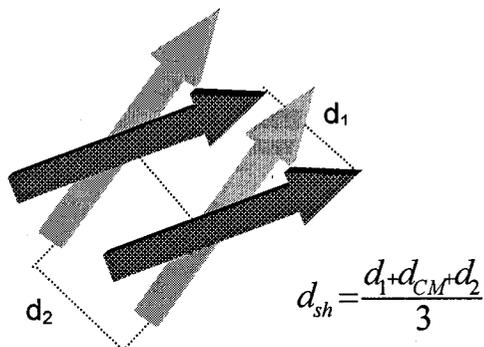
Figure 3.2 Spatial observable quantities used to evaluate structural change during simulation relative to starting structure.

3.3 Results

3.3.1 Dimers

Anti-parallel Dimer 1 is more stable than parallel Dimer 2. The distance within the sheet of Dimer 1 (~7 Å) during MD is much closer to that of the staring structure than that of Dimer 2 (~13 Å) (Figure 3.3a). More hydrogen bonds are present in Dimer 1 than Dimer 2 at equilibrium (Figure 3.3b).
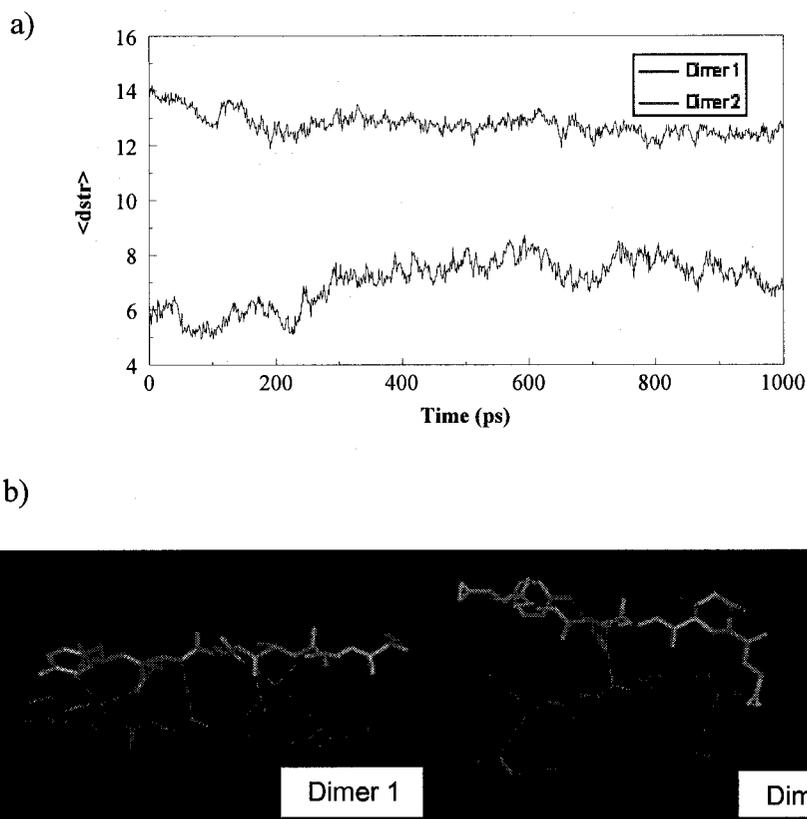
Figure 3.3 Dimer simulations. a) Average distance as a function of time. b) Stick models of average structure during last 10 ps. Dashed green lines represent hydrogen bonds.

3.3.2 Trimers

Trimer behavior resembles that of the dimers: Trimer 1 (fully anti-parallel) has the lowest potential energy of the group (Figure 3.4a); hydrogen bonds are evenly distributed between both pairs of anti-parallel strands at equilibrium (Figure 3.4c). Trimer 2 (partly anti-parallel) has about as many hydrogen bonds between the anti-parallel strands as Trimer 1, but substantially fewer hydrogen bonds are found between the parallel strands (Figure 3.4c). Trimer 3 (fully parallel) has the highest potential energy of trimers and fewer overall hydrogen bonds than Trimer 1 or Trimer 2. Trimer 1 and Trimer 2 behave

very similarly with regard to distance within a sheet, c. 5 Å, shorter than Trimer 3. All

evidence suggests that fully anti-parallel Trimer 1 is the most stable trimer of the three.

a)

Configuration	PE (kcal/mol)
Trimer1	-35675 ± 102
Trimer2	-35312 ± 104
Trimer3	-35430 ± 91

b)

c)
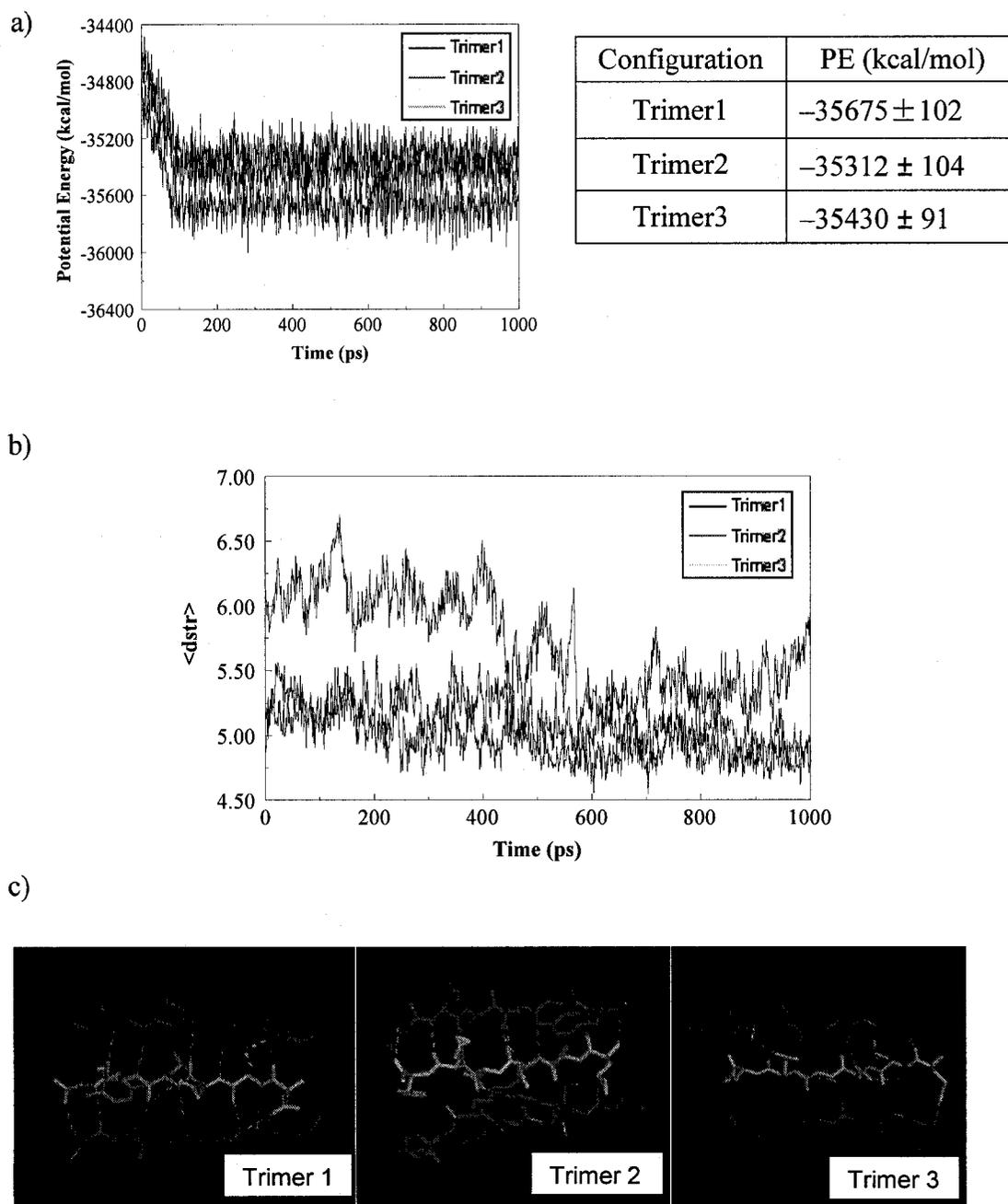
Figure 3.4 Trimer simulations. a) Potential energy during 1 ns simulation. b) Average distance within sheets. c) Stick models of average structure during last 10 ps. Dashed green lines represent hydrogen bonds.

3.3.3 Tetramers

Among the four 4-strand models, each having two layers, Tetramer 1 (parallel between sheets, anti-parallel within sheets) has the lowest potential energy at equilibrium. By contrast, the fully-parallel configuration (Tetramer 3) has the highest potential energy among tetramers (Figure 3.5a). Tetramer 2 has relatively many of the starting-structure hydrogen bonds at equilibrium; the other tetramers have fewer hydrogen bonds (Figure 3.5d). Average distance within and between sheets of tetramers in the final stage of MD simulation is shown in Figures 3.5b and 3.5c, respectively. Tetramer 3 and Tetramer 4 show the largest conformational change relative to the initial structure; these aggregates do not maintain the integrity of the initial structure as well as Tetramer 1 or Tetramer 2. The inter-sheet distance changes of Tetramer 3 (c. 15 Å) and intra-sheet distance changes of Tetramer 4 (c. 8 Å) are comparatively large. Tetramer 1 and Tetramer 2 (anti-parallel within sheets) have a less perturbed overall conformation and more hydrogen bonds than Tetramer 3 or Tetramer 4 (parallel within sheets). A common theme of the data for dimers, trimers and tetramers is that the antiparallel orientation is more stable than the corresponding parallel one. This is in agreement with related studies [171,176].

a)

Configuration	PE (kcal/mol)
Tetramer1	-42906 ± 106
Tetramer2	-40659 ± 107
Tetramer3	-38010 ± 103
Tetramer4	-40547 ± 111

b)

c)
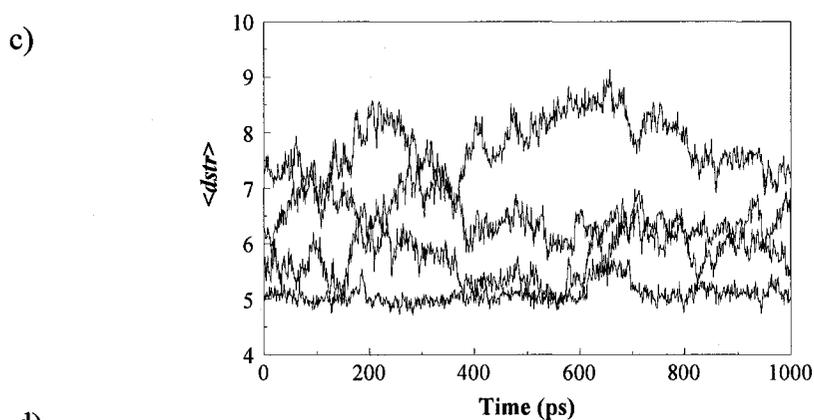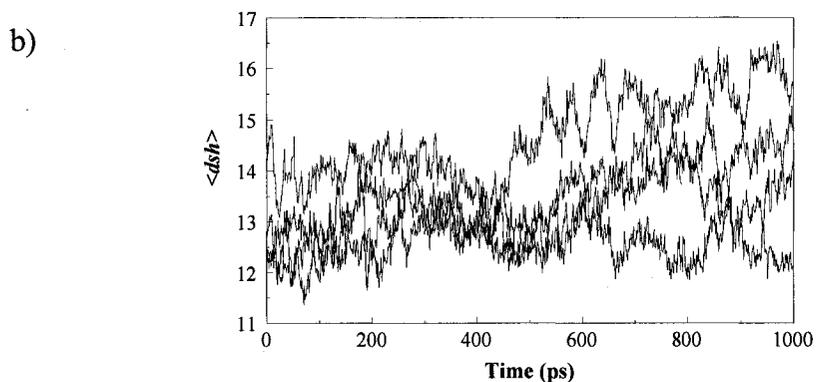
d)
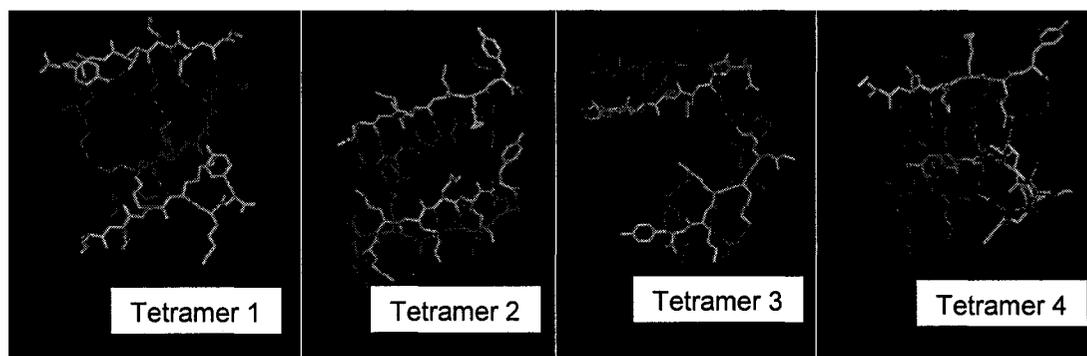
Figure 3.5 Tetramer simulations. a) Potential energy. b) Average distance within sheets. c) Average distance between sheets. Navy, Tetramer 1; magenta, Tetramer 2; yellow, Tetramer 3; cyan, Tetramer 4.d) Stick models of average structure during last 10 ps.

3.3.4 Hexamers

On the grounds of the foregoing results, the basic unit of hexamer simulations was limited to anti-parallel structures. Four configurations have been tested. Hexamers 1 and 4, parallel between sheets and anti-parallel within sheets, correspond to Tetramer 1. Hexamers 2 and 3 are anti-parallel both between and within sheets; they thus correspond to Tetramer 2. Hexamers 1 and 4 on the one hand, and Hexamers 2 and 3 on the other hand, are representations of the same motif but differ with regard to number of sheets (Figure 3.1). Figure 3.6a shows that Hexamers 1 and 3 have a lower potential energy than Hexamers 2 and 4. As to structural rearrangement, Hexamer 3 shows the largest change in intra-sheet distance during MD, from 4.7 Å to 11 Å. Hexamer 2 shows the largest increase in inter-sheet distance, from 10.5 Å to *c.* 15.5 Å. Hexamers 1 and 4 maintain a relatively high degree of structural integrity within sheets throughout the trajectory, as seen by the small differences between the initial and final intra- and inter-sheet distances (Figures 3.6b and 3.6c, and snapshot in left column of Figure 3.7). Hexamers 1 and 4 retain overall structural organization, although one sheet moves away from the others in Hexamer 1. Hexamers 2 and 3 appear disordered at equilibrium. Hexamer 2 maintains structure within sheets to a certain extent, but the location of sheets relative to each other changes substantially. Hexamer 3 is collapsed at equilibrium; it does not have a layered structure.

a)

Configurations	PE charged (kcal/mol)	PE neutral (kcal/mol)
Hexamer1	-47407 ±116	-46276±116
Hexamer2	-43563 ±114	-42150±115
Hexamer3	-46397 ±115	-46218±110
Hexamer4	-43521 ±115	-43925±114

b)
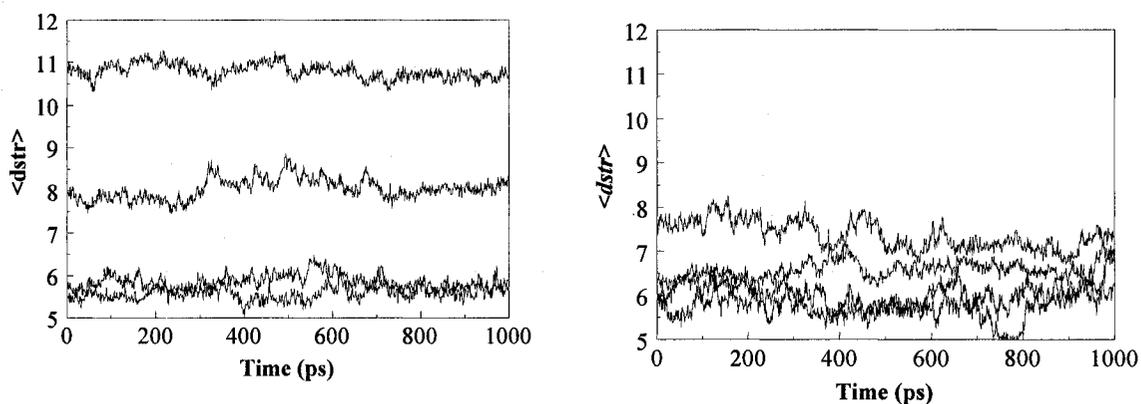
c)
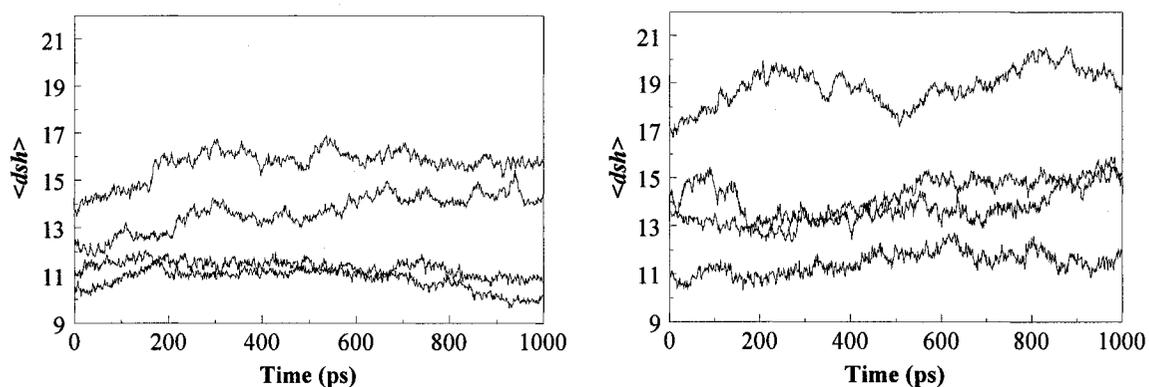
Figure 3.6 Hexamer simulations. a) Potential energy of charged hexamers. b) Average distances within sheets. c) Average distances between sheets. Left, charged. Right, neutral. Navy, Hexamer 1; magenta, Hexamer 2; yellow, Hexamer 3; cyan, Hexamer 4.
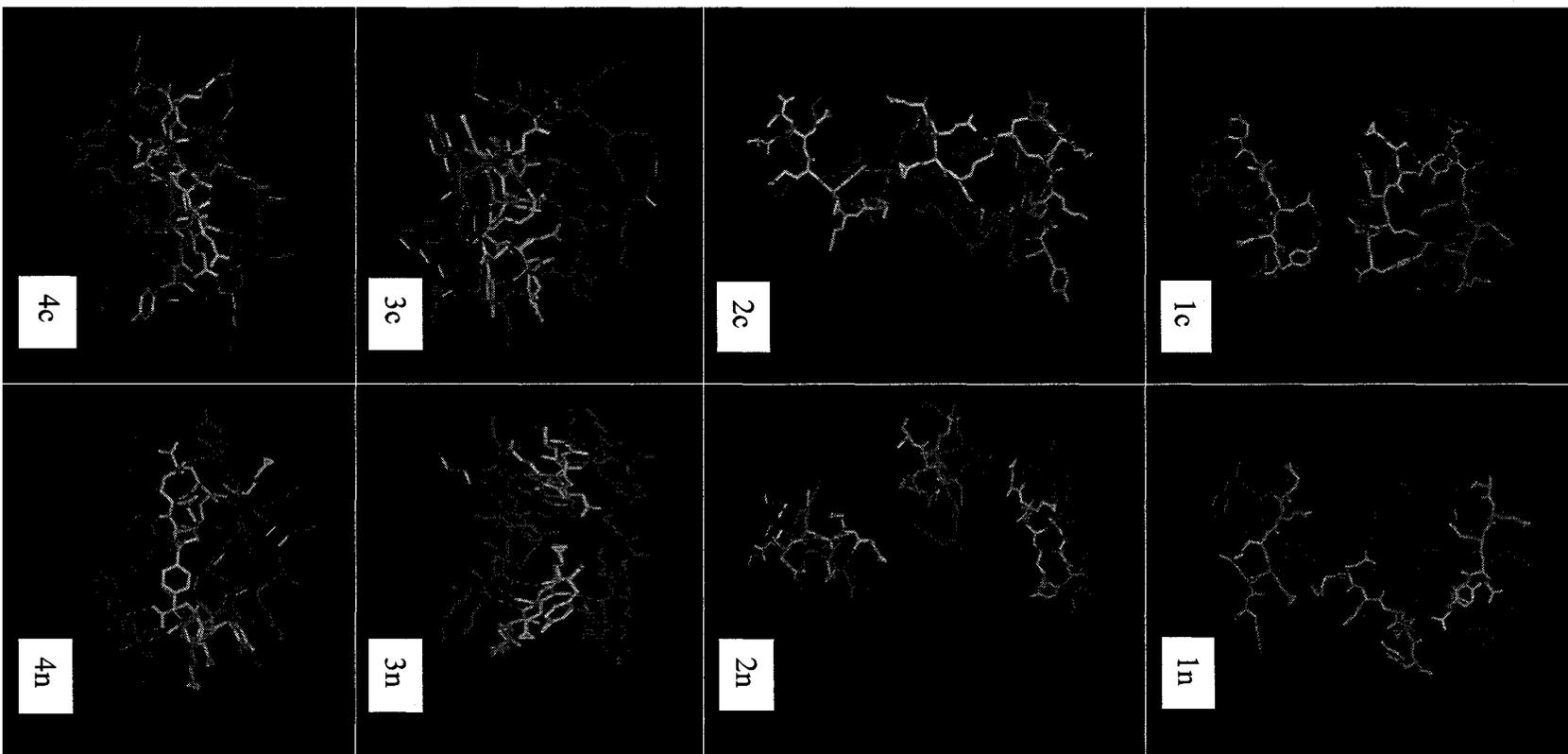
Figure 3.7 Stick models of average structure during last 100 ps of 1 ns simulation. The indicated hexamer aggregate is either charged (c) or neutral (n).

To elucidate further the role of coulombic interactions in a peptide multilayer thin film, we also simulated the four hexamer models in the absence of side chain charges. The structures were otherwise identical to those in the charged hexamer simulations discussed above. Charged and neutral peptide aggregates have been compared on the basis of snapshots and spatial metrics. Figure 3.8 and Figure 3.9 show the average distance within and between sheets of Hexamers 1 and 4. The results for the neutral aggregates are similar to those of the corresponding charged ones. Hexamers 3 and 4 show larger differences for these distances, especially Hexamer 3. Snapshots of neutral and of charged aggregate simulations correlate with the impression formed by comparison of spatial metrics. There is little difference between the charged and the neutral forms of Hexamer 1 and Hexamer 4. For Hexamer 3, the neutral peptides maintain structural integrity relatively well, whereas the charged peptides collapse during simulation. The distance between sheets in neutral Hexamer 2 is larger than in charged Hexamer 2, but the overall structure is basically the same in both cases. In general, the distance between peptides fluctuates more when the aggregate is neutral than charged. Hydrogen bonding in Hexamer 3 differs dramatically from that in Hexamer 4.
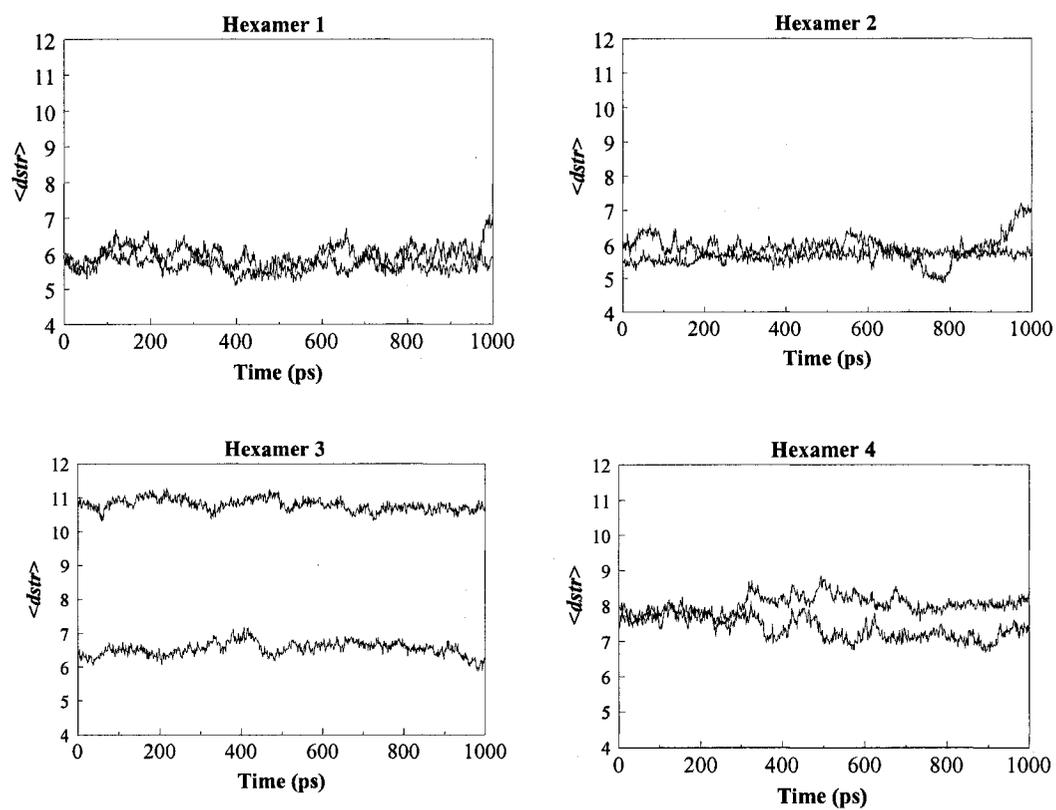
Figure 3.8 Average distances within sheets for hexamers. Black, neutral peptides; gray, charged peptides.

78
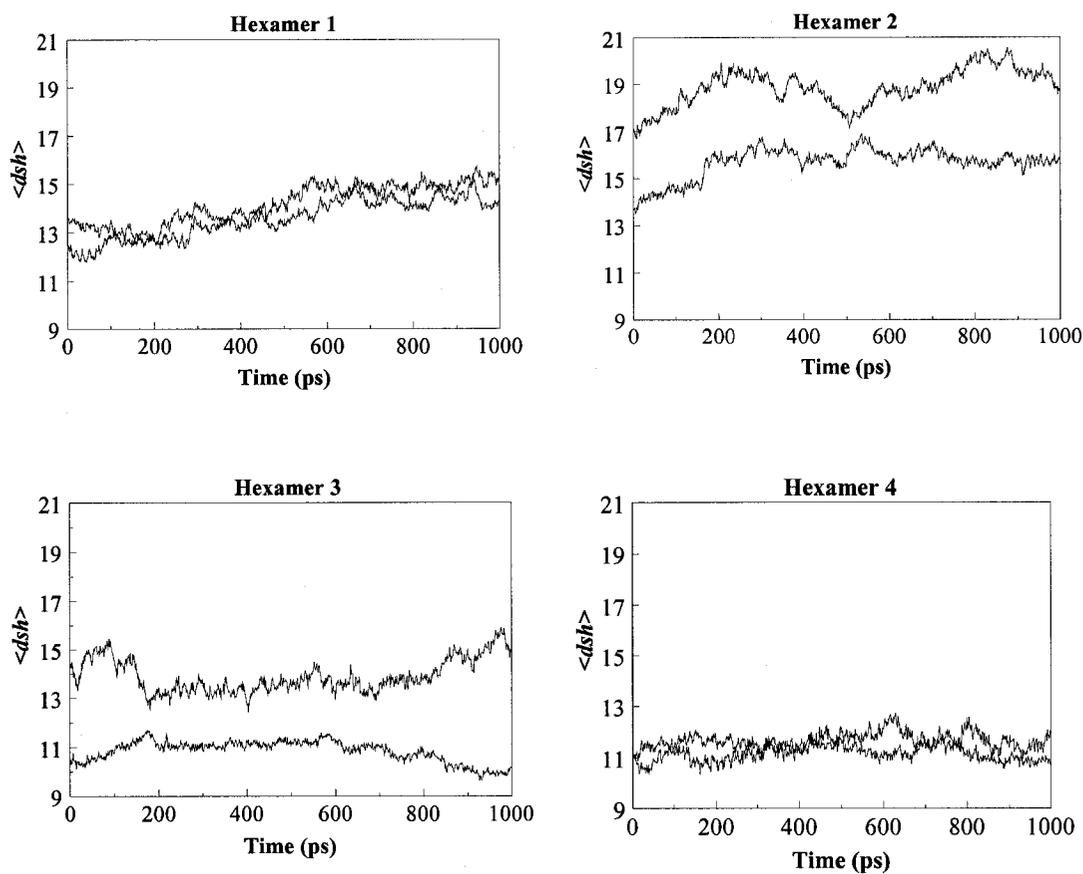
Figure 3.9 Average distance between sheets for hexamers. Black, neutral peptides; gray, charged peptides.

Figure 3.10 shows the hydrogen bond occupancy of hexamers in charged and neutral state. Hydrogen bond occupancy is more than three times greater in neutral Hexamer 3 than charged Hexamer 3. In Hexamer 4, the number of hydrogen bonds is lower by a factor of two in the neutral aggregate as compared to the charged one. This difference, however, is not as much as in Hexamer 1 or Hexamer 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
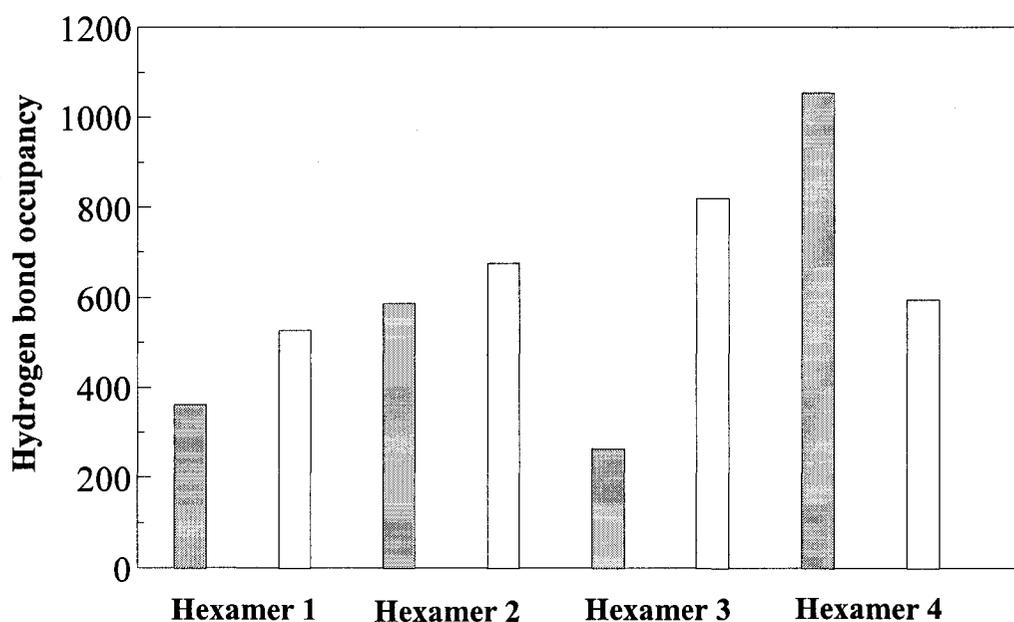
Figure 3.10 Hydrogen bonds in hexamers. Gray, charged aggregate. White, neutral aggregate. The vertical axis is the summed percentage occupancy of individual hydrogen bonds of above 20 % occupancy in the last 500 ps of simulation. In words, 4-5-fold more hydrogen bonds are present in charged Hexamer 4 than in charged Hexamer 3.

3.4 Discussions

Homopolymers of Lys and Glu have been studied extensively in an LBL context in recent years (reviewed in [36]). These polymers are simple, commercially available, relatively inexpensive, and biocompatible. Near neutral pH both PLL and PLGA have an absolute charge per monomer of about one, the maximum value; and nearly all side chains are ionized. The charged polymers will have a large number of degrees of freedom in solution. The backbone degrees of freedom, however, will be smaller in PLL or PLGA than in PSS, PAH, or PAA, due to structural constraints imposed by the peptide bond. This will influence polymer assembly behavior [144] and film internal structure [146].

The amount of polymer adsorbed and layer structure in a multilayer film will be governed by the charge density of polyions sign [182] and density of surface charge of the support [183], and ionic strength of medium from which adsorption occurs [184-186]. At neutral pH and low-to-moderate ionic strength, multilayer films fabricated from 32mer polypeptides contain a significant amount of β sheet structure [147-149]. Even so, the peptide self-assembly process is complicated, and modeling the transition from a random coil in solution to a β strand in a multilayer film will be difficult at best. It seems practical at present to focus on stability of secondary structure in simple film models.

Glu has the second highest α helix propensity value of the 20 usual amino acid types (1.27); Lys too has a high α helix propensity value (1.13) [40,109], Nevertheless, PLL/PLGA films do not contain a significant amount of an α helix at neutral pH [147-149]. The apparent reason is that the negatively-charged groups repel each other so strongly that they (together with thermal fluctuations) overcome the stabilizing influence of hydrogen bonds on helical structures. This will be especially true for short polypeptides [187].

It is generally accepted that there are several types of non-covalent interaction relevant to polyelectrolyte multilayer film assembly. These are electrostatic interactions, hydrogen bonds, and hydrophobic interactions [104]. The simulations discussed here provide insight on the relationship between these forces and the role they play in stabilizing ordered structure in polypeptide multilayer films. The initial, layered peptide structure adopted here was maintained throughout some of the MD simulations. This suggests that electrostatic interactions play a significant role not only in holding polymers together, as in "conventional" polyelectrolyte multilayer films (e.g., [95]), but also in

stabilizing β sheet structure, at least in PLL/PLGA films. It would appear that electrostatic interactions favor secondary structure stability and may be essential to it.

It is well known that hydrophobic interactions play a key role in protein folding and stability [35]. Hydrophobic interactions could also be significant in polypeptide multilayer films. This is suggested by a comparison of MD simulations of charged and of neutral hexamers. Some configurations of neutral hexamer (e.g., Hexamer 3) maintain β sheet conformation better than the corresponding charged configuration. Hydrophobic interactions have a greater effect on Hexamer 3 than the other hexamers. Hydrophobic side chains therefore might act as a general "glue" to hold peptides together in the same region of space. Such interactions, however, will not be specific enough to maintain the geometrical requirements of secondary structure in some cases. The precise role of hydrophobic interactions will depend on the specific location of apolar surface. The hexamer simulations correlate with corresponding multilayer film experimental results. Experiments have tested the pH dependence of assembly of PLL and PLGA [144,188]. pH determines the linear charge density of polypeptides. More material is deposited when both polymers are partially charged than fully charged [144]. Although PLL and PLGA are helical when partially charged, and the difference in polymer deposition could be related to conformation, it is more likely that assembly is governed primarily by polypeptide charge density. Experimental and computational work on the amyloid sequence NFGIL has reached a similar conclusion, that is, that the driving force for amyloid formation is hydrophobic interactions and that sequence matching is critical for amyloid peptide organization, e.g., [24] and [42]. Moreover, Klimov and Thirumalai have found that both interpeptide hydrophobic and electrostatic interactions are critical to

the formation of anti-parallel β sheet structure in Aβ$_{16-22}$ amyloid oligomers [189]. Mutations of either hydrophobic or charged residues destabilize the overall aggregate structure.

The simulation results presented here also suggest that the anti-parallel orientation within sheets is energetically preferred over the parallel orientation. Anti-parallel β sheets have a more uniform and favorable hydrogen bond geometry and can withstand greater distortion than parallel β sheets [75]. The contrast between anti-parallel and parallel orientation is greater for short peptides than longer ones. This is less difference between relative orientation of sheets. The inter-sheet distance is c. 10 Å, generally out of the range of hydrogen bonding formation. The simulations suggest that hydrogen bonds are important within sheets but do relatively little to stabilize overall structure in polypeptide multilayer nanofilms. Instead, overall structure is more likely stabilized by electrostatic interactions and, in some cases, hydrophobic interactions.

3.5 Conclusions

MD simulations of peptides in a charged or a neutral state have been done to clarify some aspects of the physical basis of LBL. Results suggest that electrostatics interactions play a key role in stabilizing β sheet structure in PLL/PLGA. Hydrophobic interactions, however, of substantial importance to protein folding, also play a significant role in polypeptide multilayer films. In PLL/PLGA, hydrogen bond formation between sheets is perhaps better described as a consequence of peptide association than the cause. The simulations also suggest that the anti-parallel orientation is favored over the parallel one in β sheets PLL/PLGA films.

CHAPTER 4

MOLECULAR MODELING OF

TENSIN LIGAND BINDING

4.1 Introduction

This interdisciplinary work sought to clarify the physiology of cell-substrate contacts at the molecular scale. Such contacts play a key role in cell migration, cell differentiation, and wound healing. Tensin is a large cytoskeleton-associated "scaffold" protein which is found in adhesive junctions of animal cells. The structure of tensin has not been solved at atomic resolution because the molecular weight is large and the protein is difficult to purify. Most studies of tensin thus far have proceeded largely from the perspective of the biologist; tensin is closely related to the tumor suppressor PTEN. The approach taken here is to view the scientific problem of tensin physiology from the perspective of the physicist. Near-term aims concern determination of the structure and function of the four known modular domains of tensin. The chief longer-term aim is to develop a molecular-scale model of the architecture and dynamics of cell-substrate contacts. This work will provide insight on the role of tensin and the molecules with which it interacts in the biological processes which involve such contacts. Inspiration for this undertaking comes from extensive work over decades on the molecular mechanics of striated muscle.

83

Tensin provides a physical link between the actin cytoskeleton, integrins, and other proteins at cell-substrate contacts. The overall biochemical properties of tensin must be a function of its domain composition and architecture, *i.e.,* the domains that are present and their relative positions in the protein molecule, and specific details of amino acid sequence and post-translational modification, *e.g.,* phosphorylation. More detailed knowledge of the structure and function of tensin will accelerate acquisition of more detailed knowledge of other focal adhesion components, advancing the development of molecular models of cell attachment and migration. Such knowledge is of interest to basic science, medicine, and biomedical engineering. Moreover, it could also provide a model for nanotechnology development, inspiration the design of novel types of molecular recognition and functionality, and materials design and fabrication.

4.2 Methods

4.2.1 Homology Modeling Protocol

Molecular modeling was done with the InsightII 2000 software package (Accelrys Inc. USA) and the sequence alignment tool GeneDoc 1.0 [190]. Sequences were aligned on a personal computer in the Bionanosystems Engineering Laboratory; homology modeling was done on the SiliconGraphics Fuel.

4.2.1.1 Template searching and structure-based sequence alignment

The primary structure of human tensin was obtained from GenBank, National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). Human tensin has database locus NP_072174 and GI number 66529407.

The full-length tensin peptide is 1735 amino acid residue long. Residues 1-176 correspond to the PTP domain, 177-310 to the C2 domain, 1458-1574 to the SH2 domain, and 1580-1731 to the PTB domain.

The amino acid sequence of each tensin domain was used to do PSI-BLAST [111] search of the PDB for homologous sequences: The default parameters were used (BLOSUM62; Expect = 10; E-value threshold = 0.002; composition based statistics; existence gap penalty = 11; λ = 0.319; and gap λ = 0.270). BLOSUM62 is a general purpose matrix used in BLAST. The BLOSUM matrix assigns a probability score for each position in a sequence alignment, based on the frequency with which that substitution is known to occur among consensus blocks within related proteins. An E-value for alignment score S describes the number of "hits" one can "expect" to find when searching a database of a particular size. The E-value is a convenient way to create a threshold for reporting sequence comparison results. λ is a statistical parameter used in calculating BLAST scores that can be used to convert a raw score to a bit score.

Domain sequences with a PSI-BLAST score of 50 or higher and an E-value of 5×10^{-6} or lower were induced in the alignments presented here. The corresponding crystal structures of the homologs, if available, were downloaded from the PDB.

4.2.1.2 Tensin homology modeling

Three models of each domain of human tensin were generated with the restraint-based comparative modeling program, Modeler (Accelrys). The models differed primarily in the conformation of loops and turns. For each domain, the model with the best loop conformations and stereochemical parameters, as determined by visual inspection and analysis by Procheck [191,192], was selected for further analysis and

binding studies. A summary of the quality of the structure and the Ramanchandran plot are reported here. Bad non-bond interactions (contacts per 100 residues), C_α tetrahedral distortion (deviation of the zeta torsion angle), standard deviation of the main chain hydrogen bond energies, and stereochemical G factors were measured [192].

4.2.2 Binding Affinity Protocol

4.2.2.1 Binding site analysis

Probe-accessible cavity volumes (probe radius = 1.4 Å) were calculated with Binding_Site Analysis (Accelrys Inc.). In this approach, contacts between the probe-sphere and the van der Waals protein surface delimit the probe-occupied cavity, which is similar to the Connolly surface [193]. Usually, more than one possible binding site is found by this method. As all four domains considered here are all well studied, however, it was possible to compare the possible binding sites from computational analysis with the known binding sites. Residues within each binding site were defined in InsightII as a special subset. These residues were given maximum flexibility in later binding studies, whereas residues outside of the binding site were treated as described in the protocol of the Appendix.

4.2.2.2 Geometry optimization of ligands

Initial ligand structures of ligands were either obtained from the PDB or generated within the Biopolymer module of InsightII. Subsequent minimization of the structures was done with the CVFF force field in Discover_3 as implemented in Insight II 2000. Low-energy conformers of each structure were calculated.

4.2.2.3 Binding affinity modeling and analysis

1. pTyr, pSer and pTyr

Ligands were manually inserted into the binding site of the corresponding tensin domain. Examples of PTP domains which bind to pTyr, pSer or pTyr are known from X-ray and NMR studies, providing first-approximation structures for superimposition of the various ligands into the tensin domain structures. This approach was taken whenever possible to maximize the odds of calculating of a meaningful complex structure. Distance constraints were applied to the interactions between ligand and receptor. Optimization of the geometry of the ligand fitted into the binding cavity was done with the CVFF force field as implemented in InsightII. Aspartic acid, glutamic acid, lysine, and arginine residues were taken as charged state, as all binding studies were carried out at pH 7.4 (physiological pH). The complexes were energy minimized with the CVFF force field and studied by MD using the molecule relaxation protocol (see Appendix).

2. Ins(3,4,5) ligand

Binding of the PTP domain to $PIns(3,4,5)P_3$ was studied with the Docking module. The ligand was confined within a 3 Å radius of its starting position (center of mass confinement). The procedure followed a protocol provided by Accelrys Inc. This protocol is suitable for cases in which the protein-ligand structure is not known. The docking calculation was run in two steps. The first one used a Monte Carlo (MC) algorithm to do the initial search of possible orientations of the ligand. The second one used simulated annealing to find the final model of the binding complex.

Step 1: The MC algorithm in the Docking module was used to generate 20 candidate structures with no acceptance filter. The non-bonded method was used with the

coulombic interactions turned off and the quartic van der Waals (VDW) interactions scaled down to 0.1. Ligands were flexible with an energy range of 200 kcal/mol and a tolerance of 10^6 kcal/mol, and 100 steps of minimization were done. All other parameters were set to the default value. Energy tolerance defines the maximum allowable change for succeeding acceptable structures. If the energy of a structure resulting from random moves of the ligand is higher than the energy of the last accepted structureby 10^6 kcal/mol, the resulting structure will not be accepted for the further minimization. A small value for VDW interactions (0.1) was used to avoid large energy fluctuations and adverse effects on the simulation. Coulombic interactions were turned off because interaction energies between ligand and receptor will be well-behaved numerically even if some undesirable contacts are present. This minimizes the chance that Discover will terminate due to numerical instability. The default parameters are well suited to initial searches for approximate placement of the ligand within the binding pocket.

Step 2: The MC structures were put through a separate simulated-annealing search with the program, Cell_Multipole non-bonded interactions, but without the distance-dependent dielectric option. The VDW and coulombic scales, initially 0.1, were brought up to 1.0. The initial and final temperatures of the simulated annealing step were 480 K and 280 K instead of the default 500 K and 300 K degrees to avoid a final minimization step. 10 final structures were filtered for further analysis. The second phase of docking involved a more refined approach. Specifically, a more realistic non-bond setting in Cell_Multipole was used. Since the starting structures for step 2 had reasonable non-bond contacts, scale factors were turned back to their full strength (1.0).

4.3 Results and Discussion

4.3.1 Pros and Cons of Study of the Four Domain of Tensin

The difficulties of homology modeling varied from domain to domain. For example, the biochemical properties of SH2 domains have been studied extensively, many SH2 sequences are known, and a large number of atomic resolution SH2 domain models are available from X-ray or NMR studies. The SH2 domain of tensin, however, has not been studied at high resolution. Moreover, the ligands of the tensin SH2 domain are not known. By contrast, a physiological ligand of the PTB domain of human tensin, namely integrin β, is known. PTB domains, however, are highly diverged, making sequence alignment very difficult. Table 4.1 presents a summary of the main pros and cons for computational study of the four domain structures of tensin.

Table 4.1 Pros and cons of for study of the four domains of tensin

Pros	Cons
PTP	
A high-resolution structure of PTEN is available	Sequences of available PTP structures are generally very different from each other
The tensin domain will bind a phosphorylated molecule, based on structure analysis	Physiological ligands of the tensin PTP domain are not known, increasing the difficulty of evaluating simulations complex
Some ligands and binding complexes are known for PTEN	Specificity of binding of the PTEN ligands to tensin is difficult to assess
C2	
A high-resolution structure of PTEN is available	The high-resolution PTEN crystallographic structure has a disordered loop
Crystallographic details of calcium binding sites and other features are known in distantly-related C2 domains	The teinsin C2 domain would appear to represent a distinct sub-class, differing substantially from the C2 domains for which a 3D structure is available
The tensin C2 doamin can be aligned with the PTEN homolog	The function of the tensin C2 domain is not known
SH2	
Numerous SH2 structures are available at high resolution	There is no closely-related SH2 structure to tensin at high resolution
Many ligands of other SH2 domains are known	Physiological ligands of the tensin SH2 domain are not known
Sequence alignment with the tensin SH2 domain is not very difficult	The tensin SH2 domain sequence has a large insertion, making the entire module difficult to model
PTB	
Physiological ligands of tensin are known	All high resolution PTB structures, which are relatively few in number, are very different from tensin with regard to sequence
Many PTB sequences are known	Known sequences are in general highly diverged
An NMR structure of the Shc PTB domain bound to an NPXpY containing ligand is avaliable	No structure of an integrin binding to a PTB domain is available at this time

4.3.2 PTP Domain

4.3.2.1 Domain alignment

Figure 4.1 shows the PTP domain of tensin aligned with several homologs, including PTEN. The active site is marked. Sequence alignment is not very difficult for

Figure 4.1 PTP domain alignment. The template proteins for homology modeling are PTEN (1D5R), human VH1-related dual-specificity phosphatase (1VHR), *Yersinia* protein tyrosine phosphatase (1YTS), an active site mutant of Pyst1 (1MKP), and cell division control 4b phosphatase (1OHE). 1YTS have two large insertions. Cys is mutated to Asn in tensin. Red, charged; orange, polar uncharged; green, hydrophobic and

most secondary structures, which correspond to the more highly conserved regions of the sequence. Cys is mutated to Asn in tensin and Ser in three other homologs. 1YTS has two large loop insertions, and there are short deletions in some of the other homologs.

4.3.2.2 Domain structure

The coordinates of PTEN (1D5R), human VH1-related dual-specificity phosphatase (1VHR), *Yersinia* protein tyrosine phosphatase (1YTS), an active site mutant of Pyst1 (1MKP), and cell division control 4b phosphatase (1OHE) were downloaded from the PDB (http://www.rcsb.org/). The structure files were edited so that only the PTP domain coordinates and water molecules within the binding pocket were retained. The signature motifs of the PTPs were identified by amino acid sequence alignment (Figure 3). They are residues 123-130 of 1D5R (HCKAGKG); 123-130 of 1VHR (HCREGYS); 292-299 of 1MKP (HSLAGIS); 402-409 of 1YTS (HSRAGVG); A313-A320 of 1OHE (HSKAGLG) and 112-119 of tensin (HNKGNRG) (see Figure 4.1). The highly conserved Cys in the signature motif is the essential nucleophile for phosphatase activity. In tensin this Cys is replace by Asn, rendering the domain non-enzymatic [51]. The tartrate group co-crystalized with PTEN forms hydrogen bonds in the active site with the backbone of R120, K125, D92 and H93, and with the side chains of Q171 and R130 [67].

Models of the human tensin PTP domain were generated with the restraint-based comparative modeling program Modeler and characterized in terms of hydrogen bond per amino acid, solvent accessible surface, and root mean square deviation (RMSD) from the PTEN crystallographic structure. This was done to judge the model quality. Figure 4.2 shows the ribbon diagram of the PTP domain of PTEN and the best model of tensin.
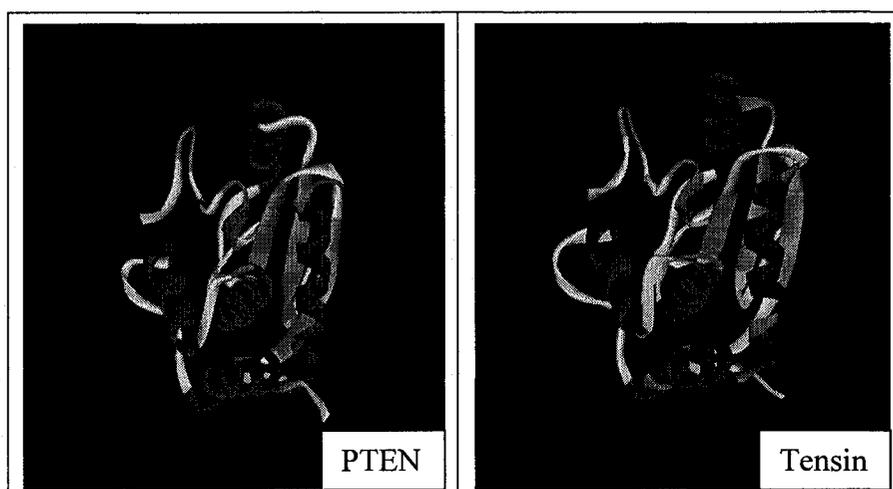
Figure 4.2 Ribbon diagram of PTP domain of PTEN and tensin. Red, α helix; blue, β sheets; green, turns; white, irregular structure.

Table 4.2 shows further details of the structure comparison: number of hydrogen bonds, solvent accessible surface (SAS), total molecular surface (TMS), and root mean-square deviation (RMSD) of the backbone atoms [195]. Hydrogen bonds were calculated by with an angle threshold of 110° and a distance threshold of 3.5 Å. Hydrogen bonds and SAS were calculated with the program described in ref. 196. TMS was calculated based on values for Gly-X-Gly tripeptides in extended conformation, where X represents one of 20 usual amino acid types. For comparison, the average number of hydrogen bonds per amino acid for small globular proteins is about 0.9, and the ratio of SAS to TMS is around 40 % [197].

Table 4.2 Structure characterization of PTP domain of PTEN and tensin.

	PTEN	**Tensin**
H-bonds/amino acid	0.96	0.70
Solvent accessible surface (SAS) Å^2	10045	9153
Total molecular surface (TMS) Å^2	30659	30584
SAS/TMS	32.7 %	30.0 %
RMSD (backbone)	~ 0.35 Å	

The data in Table 4.2 provide a high degree of confidence in the tensin model.

The backbone RMSD is just 0.35. SAS/TMS is similar in the two cases. The PTEN model

is a high resolution structure; its sequence similarity to tensin made construction of the

tensin model by homology modeling relately straight-forward.

Figure 4.3 and Figure 4.4 show the Ramanchandran plot of the PTEN PTP

domain and of the tensin model, respectively. There are no disallowed angles in the

PTEN structure, and only 1.9 % in the computer-generated tensin model.

PTEN PTP Domain
Ramachandran plot: 78.1 % core   18.1 % allow   3.9 % gener   0.0 % disall
All Ramachandrans: 12 labelled residues (out of 171)
Chi1-chi2 plots: 3 labelled residues (out of 123)
Main-chain params: 6 better   0 inside   0 worse
Side-chain params: 5 better   0 inside   0 worse
Residue properties: Max.deviation: 3.4   Bad contacts: 7
Bond len/angle: 5.0   Morris et al class: 1 1 2
G-factors Dihedrals: 0.06   Covalent: 0.46   Overall: 0.22
Main chain bond lengths:100.0 % within limits   0.0 % highlighted
Main chain bond angles: 97.9 % within limits   2.1 % highlighted
Planar groups: 95.2 % within limits   4.8 % highlighted

Figure 4.3 Structure summary and Ramanchandran plot by Procheck of the PTP domain of PTEN.

Tensin PTP Domain
Ramachandran plot: 81.0 % core   13.9 % allow   3.2 % gener   1.9 % disall
All Ramachandrans: 16 labelled residues (out of 174)
Chi1-chi2 plots: 2 labelled residues (out of 111)
Main-chain params: 6 better   0 inside   0 worse
Side-chain params: 5 better   0 inside   0 worse
Residue properties: Max.deviation: 4.1   Bad contacts: 11
Bond len/angle: 9.6   Morris et al class: 1 1 2
G-factors Dihedrals: – 0.24 Covalent: – 0.48   Overall: – 0.31
Main chain bond lengths: 97.7 % within limits   2.3 % highlighted
Main cain bond angles: 89.0 % within limits   11.0 % highlighted   4 off graph
Planar groups: 100.0 % within limits   0.0 % highlighted

Figure 4.4 Structure summary and Ramanchandran plot by Procheck of the PTP domain of tensin model.

The core region (the most favorable phi-psi angles) comprises around 80 % of all residues in both cases. The size of the core region is one of the best guides to judge stereochemical quality of a protein structure. Some of the G-factors for tensin are negative, indicating the presence of some unusual stereochemistry (in the "disallowed" region of the Ramachandran plot). All the other parameters are comparable. Comparison of the structure of PTEN (Figure 4.3) and tensin (Figure 4.4), and the low backbone RMSD of between these models, together suggest that the quality of the tensin PTP

model is high. On these grounds it was decided that the tensin model was likely to be usedful for further study.

4.3.2.3 Domain binding

Analysis of the active site of PTEN [67] and the tensin model has been carried out with the InsightII package. Figure 4.5 is a geometrical comparison of the PTEN and tensin binding pockets. The signature motif is HCKAGKG for PTEN, and HNKGNRG for tensin. The superimposed histidine residues are shown in the lower right-hand corner of the figure.
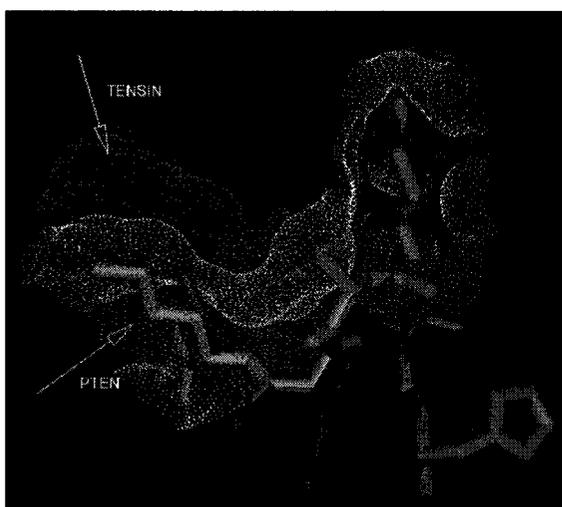
Figure 4.5 Binding pocket of PTP domain of PTEN and tensin. PTEN has a slightly wider pocket than tensin. Depth is about the same in both cases.

The catalytic site in PTEN is surrounded by three positively-charged amino acids (Lys 125, Lys 128, and His 93). This accounts for the known preference of PTEN for highly acidic substrates [194]. The corresponding residues in human tensin are Lys, Arg, and Leu, respectively. Arg, like Lys, is positively charged at neutral pH. Leu, by contrast, is hydrophobic. This suggests that tensin will bind highly acidic substrates with lower affinity than does PTEN.

The choice of ligands for the computational study was based on closely related experimental data (Table 4.3). One pTyr ligand, one pSer and one pThr were chosen for simulation. The binding affinities to PTEN are similar in magnitude, making comparision relatively straightforward. The ligands were built with Biopolymer (Accelrys).

Table 4.3 Activity of PTEN measured with phosphorylated substrates [60]. The peptides in red were chosen for computational study.

Substrate	Activity, pmol/min per mg
Phosphotyrosyl substrates	
1RCM (PDB entry)	88 ± 6.6
EDNDYINASL	51 ± 4.0
Myelin basic protein	21 ± 1.2
Phosphoseryl and -threonyl substrates	
RRRDDDSDDD	210 ± 4.2
RRREEETEEE	161 ± 1.8
Myelin basic protein	11.3 ± 0.6

4.3.2.3.1 Phospho-tyrosine

Simulations were done to model binding of the PTP domains of tensin and of PTEN to the peptides DNDpYINAS and ApYA. DNDpYINAS is known to bind PTEN from experimental data. The tripeptide ApYA has no of flanking residues, Alanine, with its small side chain, limits the contribution of side chain moieties to the overall binding affinity and focuses attention on the phosphorylated chemical group. Figure 4.6 shows the DNDpYINAS ligand in the binding pocket.
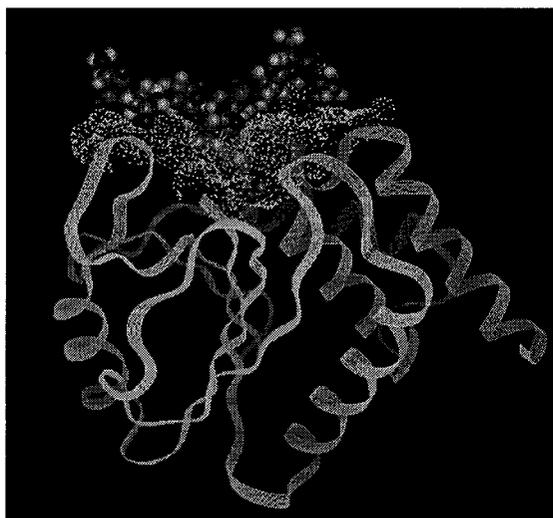
Figure 4.6 Diagram of the PTP domain of tensin complexed with pTyr peptide. The PTP domain is shown in yellow ribbon, the peptide ligand in ball and stick representation, and the binding site in blue.

Results of simulated binding of the PTP domains to the pTyr peptides are shown in Table 4.4. The total interaction energy includes both the VDW and the electrostatic components. The DNDpYINAS peptide binds to PTEN with much higher affinity than to tensin. These results are in good agreement with experimental data on PTEN, with binding to tensin serving as a sort of negative control. The result also suggests that the difference in binding affinity between PTEN and tensin comes primarily from electrostatic interactions – the key to binding specificity for complementary geometrical surfaces. The ApYA peptide also shows a difference in binding affinity to the two proteins, but not as much as for the other peptide. The computational analysis suggests that tensin might not be able to bind the tripeptide at all, given that the electrical component of the binding energy is positive; the VDW interactions are likely to be too weak to compete with thermal energy. In any case, the data underscore the importance of the flanking residues to pY binding. These residues help to determine binding specificity and thermostability of the complex to a substantial extent.

Table 4.4 Interaction energy between the pY ligand and the PTP domain of PTEN and tensin. VDW, van der Waals interaction energy. ELE, electrostatic interaction energy; TOT, total interaction energy.

Ligand	PTEN (kcal/mol)	Tensin (kcal/mol)
DNDpYINAS	VDW = − 48.3 ELE = − 342.3 TOT = − 390.6	VDW = − 63.1 ELE = − 15.5 Tot = − 78.6
ApYA	VDW = − 35.7 ELE = − 27.1 TOT = − 62.8	VDW = − 42.5 ELE = 3.2 TOT = − 39.3

4.3.2.3.2 Phospho-serine

Results of computational analysis of binding of the PTP domain to a pSer peptide are shown in Table 4.5. DDDpSDDD binds to PTEN with about the same affinity as to tensin. The calculated total interaction energy has about the same magnitude as that of binding of DNDpYINAS to PTEN, again consistent with experimental data. ApSA shows higher binding affinity to tensin than to PTEN. As before, the tripeptide data clearly indicate the importance of flanking residues in peptide ligand binding to a molecular protein domain, given the relatively low total energy of ApSA binding. It must be mentioned that the DDDpSDDD peptide is highly negatively charged at neutral pH, making it likely to bind to any structure of opposite polarity; binding to tensin and probably to PTEN is non-specific. Table 4.5 shows that the electrostatic attaction of this peptide to PTEN and tensin is $c. - 430$ kcal/mol − the main contribution to the binding energy. The result contrasts with that of the pY ligand DNDpYINAS, which binds specifically to PTEN. S and N are polar but uncharged, I and A are hydrophobic. There is a large difference between the binding of the pY ligand (Table 4.4) and the of the pS ligand to PTEN and tensin (Table 4.5). The side chain of Y is much larger than that of S.

Table 4.5 Interaction energy between the pS ligand and the PTP domain of PTEN and tensin. VDW, van der Waals interaction energy. ELE, electrostatic interaction energy. TOT total interaction energy.

Ligand	PTEN (kcal/mol)	Tensin (kcal/mol)
DDDpSDDD	VDW = − 29.8 ELE = − 438.2 TOT = − 468.0	VDW = − 37.9 ELE = − 425.8 TOT = − 463.7
ApSA	VDW = − 16.1 ELE = 2.0 TOT = − 14.1	VDW = − 21.3 ELE = − 3.2 TOT = − 24.5

4.3.2.3.3 Phospho-threonine

Results of binding of the PTP domain to a pThr peptide are shown in Table 4.6. EEEpTEEE binds to PTEN with about the same affinity as to tensin: Binding is relatively non-specific. The calculated total interaction energy has about the same magnitude as the binding of DNDpYINAS or DDDpSDDD to PTEN, consistent with experimental data. The ApTA peptide shows a slightly higher binding affinity to tensin than to PTEN. Again, the data indicate the importance of flanking residues in peptide ligand binding to a protein domain, despite the high negative charge of the phosphoryl group and the high positive charge of the coordinating ligands in the binding pocket.

Table 4.6 Interaction energy between pT ligand and PTP domain of PTEN and tensin. VDW, van der Waals interaction energy. ELE, electrostatic interaction energy. TOT, total interaction energy.

Ligand	PTEN (kcal/mol)	Tensin (kcal/mol)
EEEpTEEE	VDW = − 28.8 ELE = − 350.0 TOT = − 378.8	VDW = − 42.6 ELE = − 344.3 TOT = − 386.9
ApTA	VDW = − 21.0 ELE = − 2.7 TOT = − 23.7	VDW = − 23.4 ELE = − 2.9 TOT = − 26.3

4.3.2.3.4 Phospho-inositol

PTEN *in vitro* has a k_{cat}/K_M of 17.3 $(min-mM)^{-1}$ against phosphatidylinositol (3,4,5)-trisphosphate ($PI(3,4,5)P_3$) [67], a lipid second messenger produced by phosphoinositide 3-kinase (PI3K). $PI(3,4,5)P_3$ is a physiological substrate of the PTP domain of PTEN. $PI(3,4,5)P_3$ activates downstream effectors. PTEN dephosphorylates $PI(3,4,5)P_3$ with specificity for the phosphate group at the D3 position of the inositol ring [61]. The $Ins(3,4,5)P_3$ molecule studied here is the head group of $PI(3,4,5)P_3$; the lipid tail was omitted; $Ins(3,4,5)$ is the key part for molecular recognition. The Cys in the bottom of the active site dephosphorylates the molecule.

Table 4.7 and Table 4.8 present details of the conformation analysis by Docking of 10 structures of the $Ins(3,4,5)P_3$ complexed with PTEN and tensin, respectively. Interaction energy between ligand and macromolecule as calculated by Evaluate/Intermolecular (Docking), Ludi_3 score (Ludi), number of hydrogen bonds between ligand and receptor, and orientation of the ligand in the binding pocket are recorded for each structure. A candidate structure was considered "productive" if the P_3 of $Ins(3,4,5)P_3$ was at the bottom of the binding pocket, making dephosphorylation of $PI(3,4,5)P_3$ possible with specificity for the phosphate group at the D3 position of the inositol ring [61]. All models having a productive orientation have relatively high scores in the interaction energy, Ludi_3 score, and hydrogen bond categories.

Analysis would suggest that two candidate strucutures for the PTEN PTP domain-Ins(3,4,5)P3 complex (Table 4.7) and four candidates for the tensin PTP domain-Ins(3,4,5)P3 complex (Table 4.8) are suitable for a more sophisticated MD study. In any case, the initial work presented here would suggest that the binding affinity of Ins (3,4,5)

to the PTEN PTP domain to is higher than to the tensin PTP domain, according to Ludi

scores and interaction potential energy values.

Table 4.7 10 final conformation analysis of the PTEN PTP domain and Ins (3,4,5)P₃.

Model	Interaction Energy (kcal/mol)	Ludi_3 Score	Hydrogen bonds	Productive orientation
1	– 62.0	298	5	
2	– 61.4	144	3	
3	– 52.6	209	3	
4	– 61.4	165	6	
5	– 49.9	157	2	
6	– 86.5	327	6	√
7	– 65.4	421	7	
8	– 79.6	328	7	√
9	– 48.2	261	7	
10	– 60.0	338	6	

Table 4.8 10 final conformation analysis of the tensin PTP domain and Ins (3,4,5)P₃.

Model	Interaction Energy (kcal/mol)	Ludi_3 Score	Hydrogen bonds	Productive orientation
1	– 60.8	125	3	
2	– 42.8	187	3	
3	– 64.5	218	6	√
4	– 49.7	177	3	
5	– 56.8	247	7	√
6	–54.3	243	5	√
7	– 50.2	222	5	
8	– 48.7	195	3	√
9	– 42.9	179	4	
10	– 62.0	239	6	

Figure 4.7 shows an example of a productive orientation of Ins(3,4,5)P3 bound to

PTEN. It is one of the ten models listed in Table 4.7. The yellow ribbon represents the

backbone of the PTP domain. The Ins(3,4,5)P3 ligand, shown in ball-and-stick

representation, is in a productive orientation in the binding pocket; that is, P3 is located at

the bottom of the active site, chosen to the nucleophilic side chain of Cys. The other two

phosphorylated sites in the ligand are P1 and P2. The phosphoryl groups are colored

purple. This is a possible input structure for further MD study.
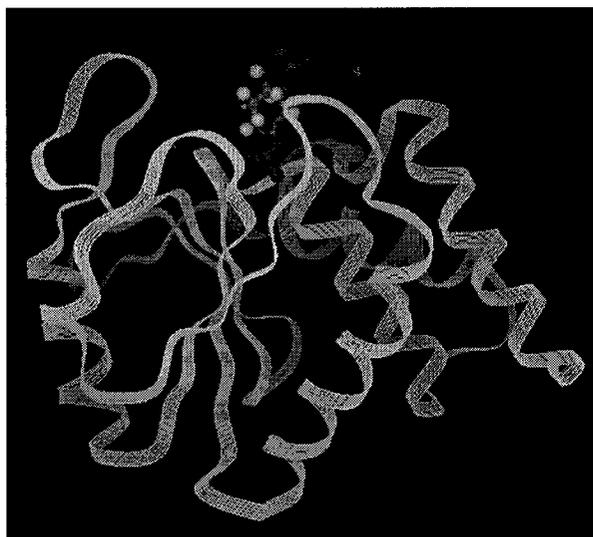
Figure 4.7 Example of a productive orientation of Ins(3,4,5)P3 in the PTP domain of PTEN (model 6 in Table 4.7). The P3 group is located at the bottom of the binding pocket.

4.3.3 C2 domain

4.3.3.1 Domain alignment

Figure 4.8 shows the C2 domain of tensin aligned with several homologs.

Sequence heterogeneity is high. Nevertheless, relatedness is discernable. All the

sequences are presumably descended from the same protogene. Both tensin and PTEN

have a loop insertion close to the C-terminus of the domain. Asp268 in PTEN is conserved throughout the alignment.
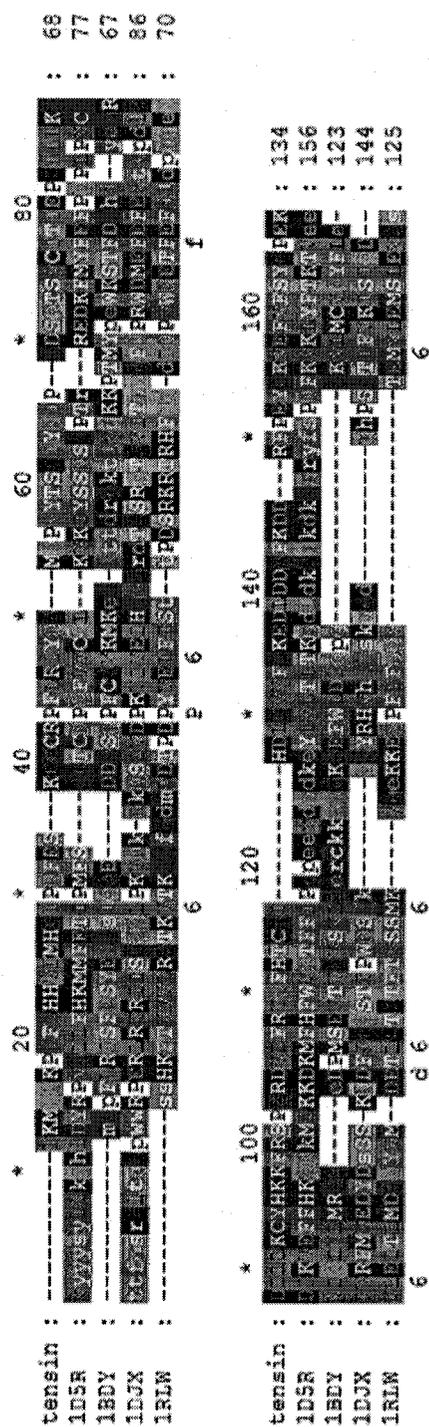
Figure 4.8 C2 domain alignment. Template structures are the C2 domain of PTEN (1D5R), protein kinase C-delta (1BDY), mammalian phosphoinositide-specific phospholipase C-delta 1 (1DJX), and cytosolic phospholipase A2 (1RLW).

4.3.3.2 Domain structure

Figure 4.9 shows the C2 domain of the crystallographic structure of PTEN and the homology model of tensin. The structures are very similar overall. Details of structure characterization are shown in Table 4.9. The backbone RMSD is only 0.5 Å. The SAS/TMS ratio is about the same in both cases, c. 40%. The number of hydrogen bonds per amino acid is close to 1. All values are close to what is found for well-studied small globular proteins.
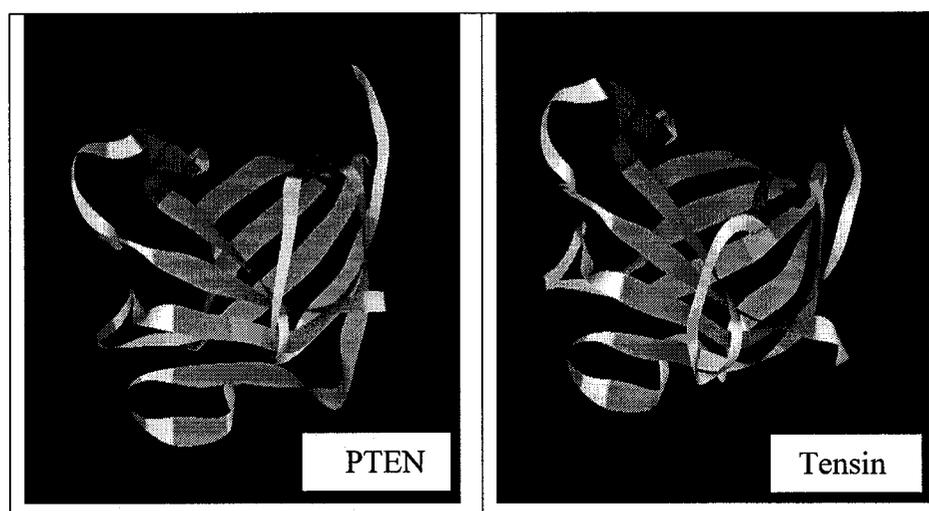
Figure 4.9 Ribbon diagram of C2 domain of PTEN and tensin. Red, α helix; blue, β sheets; green, turns; white, irregular structure. Note that the N- and C- termini are close in space.

Table 4.9 Structure characterization of C2 domain of PTEN and tensin.

	PTEN	Tensin
H-bonds/amino acid	0.98	0.82
Solvent accessible surface (SAS)	8804	10052
Total molecular surface (TMS)	24153	25067
SAS/TMS	36.4 %	40.1 %
RMSD (backbone)	~ 0.5 Å	

Figure 4.10 and Figure 4.11 show the Ramachandran plot of the PTEN C2 domain

crystallographic structure and the tensin model developed here. Only 2.5 % of the angles

in the tensin model fall in disallowed regions. G-factors for tensin are negative, indicating

some unusual stereochemical properties. In general, however, the model seems a good

first approximation of tensin structure.

PTEN C2 Domain
Ramachandran plot: 84.0 % core 12.6 % allow 3.4 % gener 0.0 % disall
All Ramachandrans: 4 labelled residues (out of 131)
Chi1-chi2 plots: 2 labelled residues (out of 93)
Main-chain params: 6 better 0 inside 0 worse
Side-chain params: 5 better 0 inside 0 worse
Residue properties: Max.deviation: 2.9 Bad contacts: 5
Bond len/angle: 5.1 Morris et al class: 1 1 2
G-factors Dihedrals: 0.01 Covalent: 0.40 Overall: 0.17
Main chain bond lengths:100.0 % within limits 0.0 % highlighted
Main chain bond angles: 97.7 % within limits 2.3 % highlighted
Planar groups: 98.1 % within limits 1.9 % highlighted

Figure 4.10 Structure summary and Ramanchandran plot by Procheck of the C2 domain of PTEN model.

Tensin C2 Domain
Ramachandran plot: 78.5 % core   16.5 % allow   2.5 % gener   2.5 % disall
All Ramachandrans: 9 labelled residues (out of 144)
Chi1-chi2 plots: 0 labelled residues (out of 94)
Main-chain params: 6 better   0 inside   0 worse
Side-chain params: 5 better   0 inside   0 worse
Residue properties: Max.deviation: 4.0   Bad contacts: 6
Bond len/angle: 4.7 Morris et al class: 1 1 2
G-factors Dihedrals: − 0.29 Covalent: − 0.48   Overall: − 0.36
Main chain bond lengths: 97.1 % within limits   2.9 % highlighted
Main chain bond angles: 86.6 % within limits 13.4 % highlighted
Planar groups: 100.0 % within limits   0.0 % highlighted

Figure 4.11 Structure summary and Ramanchandran plot by Procheck of the C2 domain of tensin model.

4.3.4 SH2 Domain

4.3.4.1 Domain alignment

Figure 4.12 shows an alignement of the SH2 domain sequence of human tensin with several homologs. Overall, the sequence is highly conserved. Identical residues are in the binding site or core, and the few absolutely conserved residues across all known SH2 domains are involved in coordinating the phosphoryl moiety of phosphotyrosine [71,72]. Sequence differences give some indication of the plasticity of domain structure

and tolerance of mutations in some regions of the sequence. The sequences shown in

Figure 4.12 are relatively easy to align; confidence in the model is high.
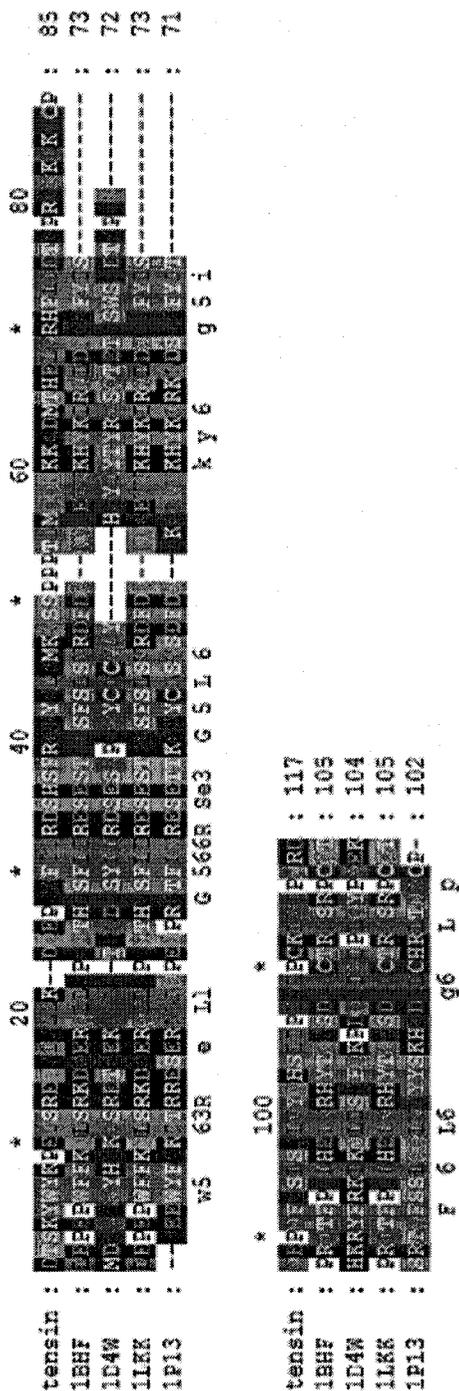
Figure 4.12 SH2 domain alignment of human tensin. Template structures are SH2 domain of human p56-lck (1BHF), XLP protein SAP (1D4W), human p56-lck in complex with peptide pYEEI (1LKK), Src SH2 Domain Complexed with Peptide SDpYANFK (1P13).

4.3.4.2 Domain structure

Figure 4.13 shows the structure of two SH2 domain templates and the tensin model. In all

cases a central β sheet is flanked by an α helix on each side.
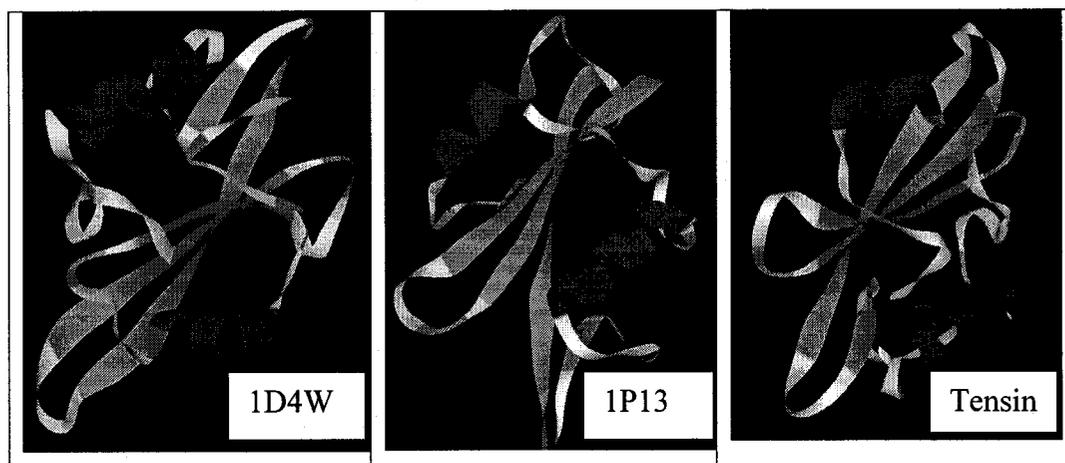
Figure 4.13 Ribbon diagram of SH2 domain of 1D4W, 1P13 and tensin. Red, α helix; blue, β sheets; green, turns; white, irregular structure. Note that the N- and C- termini are close in space.

Table 4.10 and Table 4.11 give quantitive measures of the structures. RMSD of

the backbone between the various structures are shown in Table 4.10. Note that the

highest value (1.64 Å), which is relatively low, is between two templates (1D4W and

1P13). In all cases, the number of hydrogen bonds per amino acid is around 0.8, and

SAS/TMS is between 35% and 48%, close to average in small globular proteins.

Table 4.10 RMSD of backbone in Å between reference proteins and tensin.

	1D4W	1LKK	1P13
1D4W	0.0	0.85	1.64
1LKK	0.85	0.0	1.42
1P13	1.64	1.42	0.0
Tensin	0.95	0.56	0.55

Table 4.11 Structure characterization of SH2 domain of 1P13, 1LKK and tensin. The molecular surface area is larger in the case of tensin due to the insertion around position 80 in the sequence (Figure 4.12).

	1P13	**1LKK**	**Tensin**
H-bonds/amino acid	0.88	0.81	0.71
Solvent accessible surface (SAS)	6136	6591	9520
Total molecular surface (TMS)	17502	17206	19634
SAS/TMS	35.1 %	38.3 %	48.4 %

Figure 4.14 and Figure 4.15 show structure analysis data for the SH2 domains of 1LKK and tensin, respectively. Both structures have at least 85 % of the phi-psi angles in "core" regions of the Ramachandran plot. We can therefore have a high degree of confidence in the tensin model, and the model can be used for computational ligand binding study.
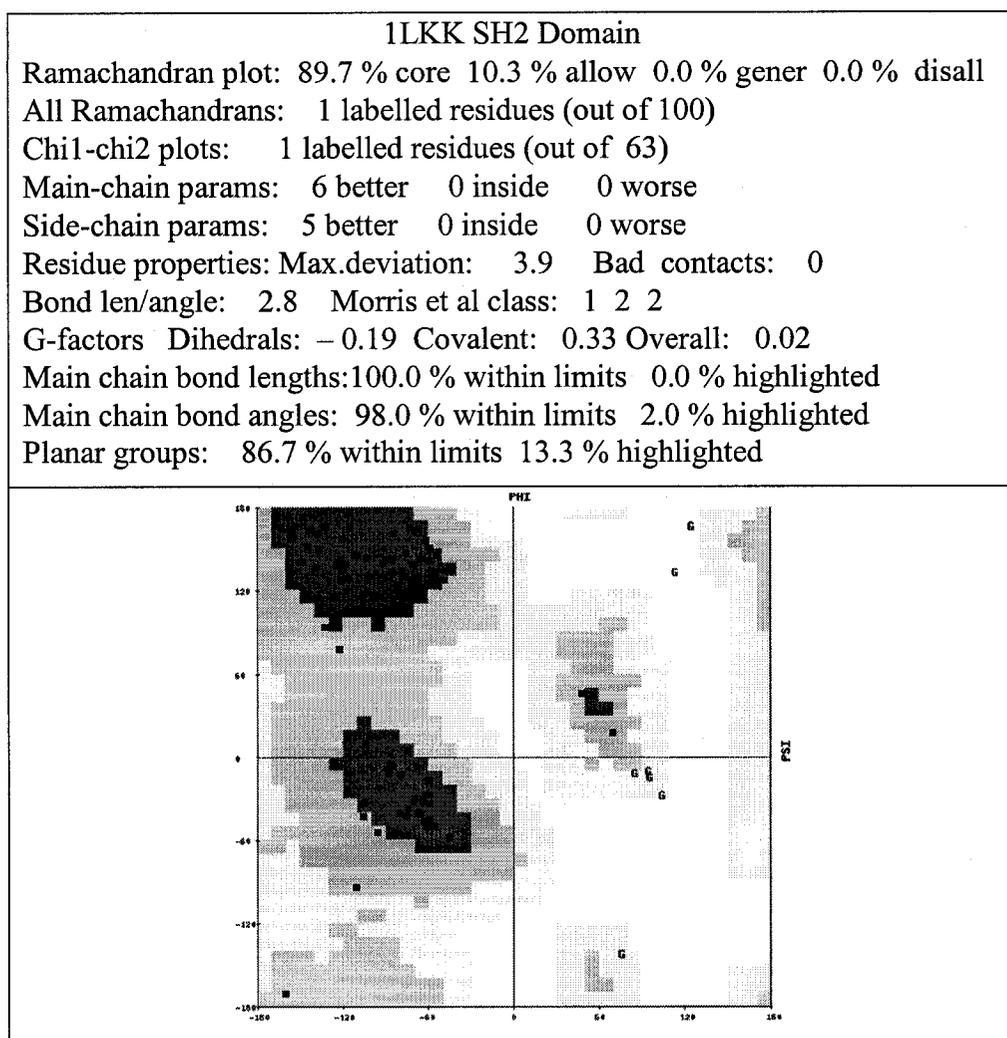
Figure 4.14 Structure summary and Ramanchandran plot by Procheck of the SH2 domain of 1LKK.

Tensin SH2 Domain
Ramachandran plot: 85.2 % core 11.1 % allow 1.9 % gener 1.9 % disall
All Ramachandrans: 12 labelled residues (out of 128)
Chi1-chi2 plots: 3 labelled residues (out of 78)
Main-chain params: 6 better 0 inside 0 worse
Side-chain params: 5 better 0 inside 0 worse
Residue properties: Max.deviation: 4.0 Bad contacts: 4
Bond len/angle: 7.0 Morris et al class: 1 1 2 1 cis-peptides
G-factors Dihedrals: − 0.22 Covalent: − 0.36 Overall: − 0.27
Main chain bond lengths: 98.1 % within limits 1.9 % highlighted
Main chain bond angles: 88.7 % within limits 11.3 % highlighted
Planar groups: 100.0 % within limits 0.0 % highlighted

Figure 4.15 Structure summary and Ramanchandran plot by Procheck of the SH2 domain of tensin model.

4.3.5 PTB domain

4.3.5.1 Domain alignment

Figure 4.16 shows the sequence alignment of human tensin with several homologs. The sequences are highly diverged, reflecting functional specificity and making alignment difficult. The conserved blocks of structure correspond to α helices and β sheets; insertions generally occur between secondary structures.
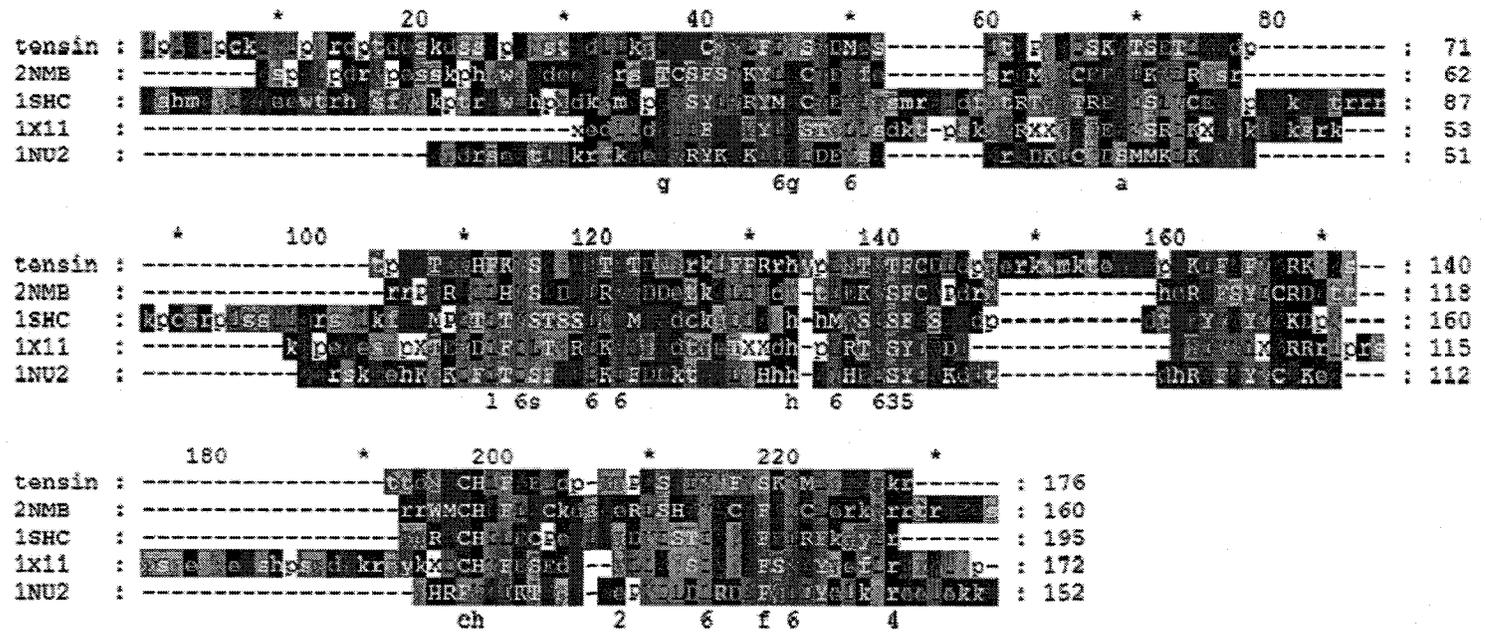
Figure 4.16 PTB domain alignment. Template structures are PTB domain of Numb (2NMB), Shc (1SHC), member 1 of X11 (1X11), and murine Disabled-1 (Dab1) (1NU2) protein. Several insertions and deletions are shown in the sequence of templates and tensin.

4.3.5.2 Domain structure

Figure 4.17 shows the structure of PTB domain template 1SHC and the tensin homology model. Structurally, Shc (1SHC) consists of 3 α helices (red) and β sheets (blue), in which a β sandwich is capped by an α helix. The tensin PTB model conserved the overall secondary and tertiary structure, but the significant differences are consistent with the differences in sequence.
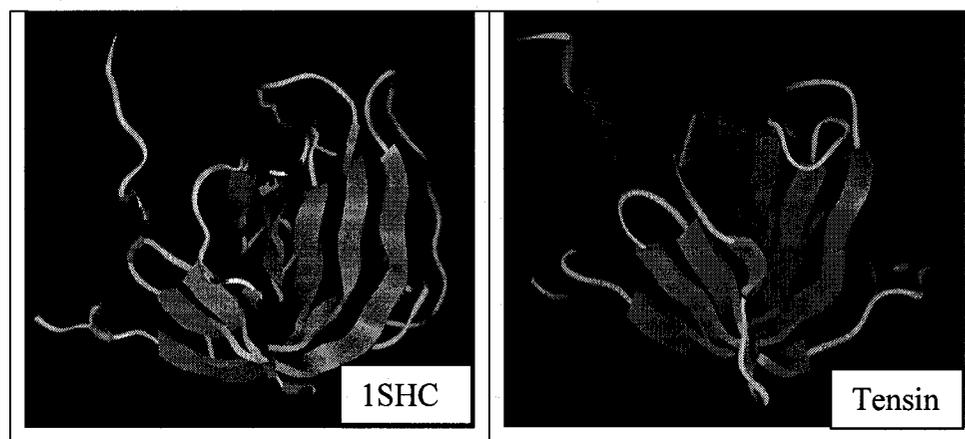
Figure 4.17 Secondary structure diagram of PTB domain of 1SHC and tensin. Red, α helix; blue, β sheets; green, turns; white, irregular structure. Note that the N- and C-termini are close in space.

Table 4.12 and Table 4.13 give quantitive measures of the structures. RMSD of the backbone between the various structures are calculated and shown in Table 4.12. Note that it is difficult to superimpose 1SHC on the other two templates, 1X11 and 2NMB. The lowest RMSD of backbone between structures, 4.5 Å, is between tensin and 2NMB (it is 5.7 Å between 1X11 and 2NMB). The other RMSD measure are over 10 Å. Table 4.13 presents the number of hydrogen bonds per amino acid (close to 1 for 1SHC and tensin), and the SAS/TMS ratio (around 40 %, close to average in small globular

proteins). The molecular surface area is large for 1SHC and tensin due to the several segments of insertions in the sequence.

Table 4.12 RMSD of backbone in Å between template proteins 2NMB, 1SHC, 1X11 and tensin model. An asterisk is shown for cases where the two structures are hard to superimpose.

	2NMB	1SHC	1X11
2NMB	0.0	*	5.7
1SHC	*	0.0	*
1X11	5.7	*	0.0
Tensin	4.5	11.5	10.3

Table 4.13 Structure characterization of PTB domain of 2NMB,1SHC and tensin.

	2NMB	1SHC	Tensin
H-bonds/amino acid	0.70	0.97	0.95
Solvent accessible surface (SAS)	10362	13632	11303
Total molecular surface (TMS)	24840	31975	28420
SAS/TMS	41.7 %	42.6 %	39.7 %

Figure 4.18 and Figure 4.19 show the results of structure analysis by Procheck of the PTB domains of Numb (2NMB) and human tensin, respectively. The tensin model has better better geometrical quality, judging by the Procheck results. In Numb only 52.7% "core" region of Ramachandran plot, comparing to 84.2% for tensin. Large-scale differences between the two structures are evident from comparision of the Ramachandran plots. 2NMB is an NMR structure, and the resolution of an NMR structure cannot be as high as that of an X-ray crystallographic structure. The Numb PTB domain was chosen as one of the template structures because it has been studied.

2NMB PTB Domain

Ramachandran plot: 52.7 % core   39.7 % allow   5.3 % gener   2.3 % disall
All Ramachandrans: 21 labelled residues (out of 145)
Chi1-chi2 plots: 11 labelled residues (out of 81)
Main-chain params: 5 better   0 inside   1 worse
Side-chain params: 2 better   0 inside   3 worse
Residue properties: Max. eviation: 4.0   Bad contacts: 7
Bond len/angle: 1.7   Morris et al class: 4 4 3
G-factors Dihedrals: − 0.44 Covalent: 0.69   Overall: − 0.01
Main chain bond lengths:100.0 % within limits   0.0 % highlighted
Main chain bond angles: 100.0 % within limits   0.0 % highlighted
Planar groups: 100.0 % within limits   0.0% highlighted

Figure 4.18 Structure summary and Ramanchandran plot by Procheck of the PTB domain of Numb model.
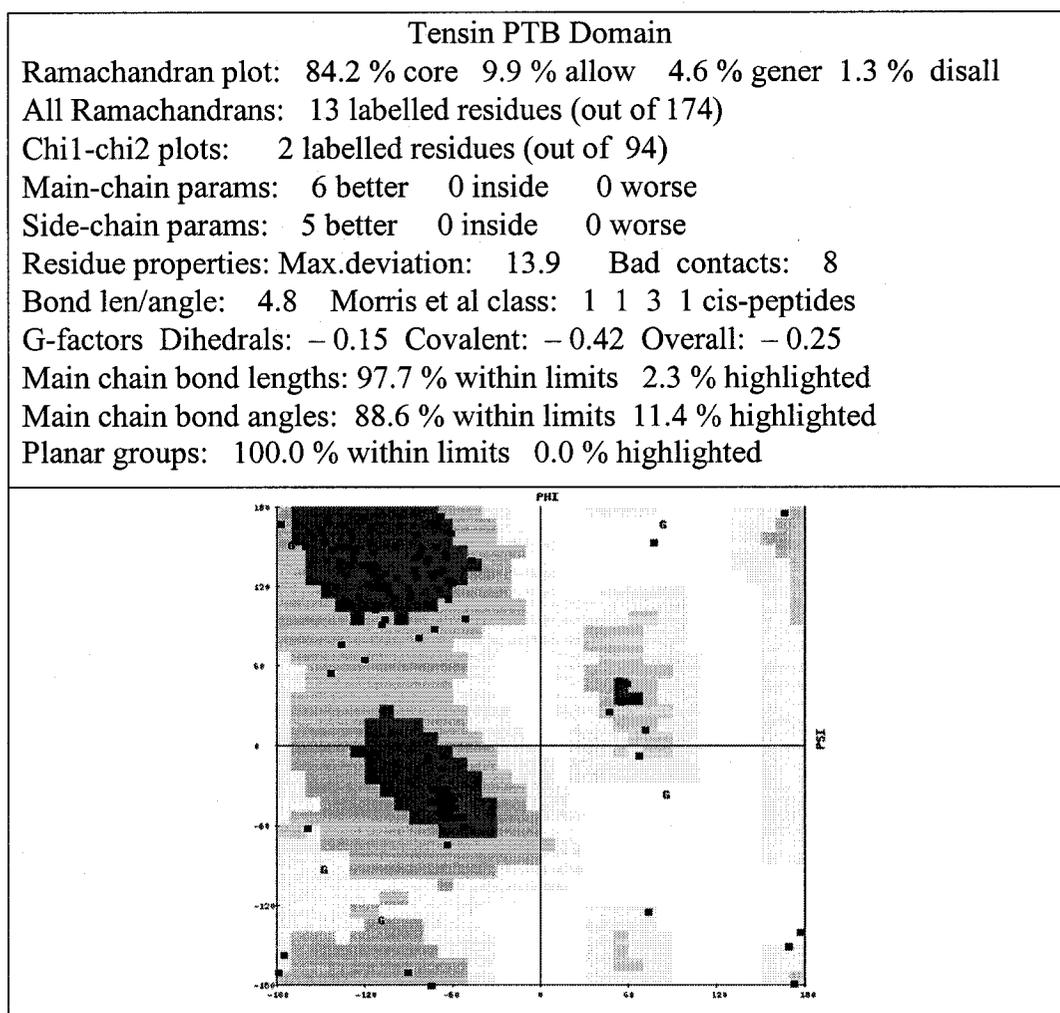
Figure 4.19 Structure summary and Ramanchandran plot by Procheck of the PTB domain of tensin model.

4.3.5.3 Ligand binding

As discussed above, integrins are transmembrane glycoprotein receptors involved in cell-cell interactions and adhesion to the ECM. Most integrin β tails (except β4 and β8) are short and contain one or two NPXY/F motifs that can form β turns. These motifs are part of a canonical recognition sequence for PTB domains, modular structures that are present in a wide variety of signaling and cytoskeletal proteins. Binding of different PTB domains to integrins β1A, β2, β3 β5 and β7 has been studied experimentally in vitro [23].

The generality of the interaction was examined by the binding of a series of recombinant PTB domains to the indicated panel of short integrin β tails. The PTB domains of a number of integrin-related functional proteins such as talin, Dab, tensin, and integrin cytoplasmic tails were cloned, expressed and purified. Specific β tail mutations were introduced by site-directed mutagenesis. Bound proteins were fractionated by SDS-PAGE and analyzed by western blotting. Binding protein levels were evaluated by densitometry. The concentration of bound protein was proportional to the intensity of bands on western blots.

Figure 4.20A shows an alignment of the cytoplasmic portion of integrin β sequences. Integrins β3 and β5, which bind well to tensin, have an Ala residue in the −5 position from Tyr of the NPXY motif; integrins β2 and β7 contain a polar or charged side chain at this position. In β3 and β5, position +2 is occupied by residues with charged groups, whereas in the integrins that do not bind tensin it is Ser. This analysis suggests that residues in positions −5 and +2 with respect to the Tyr of the NPXY/F motif contribute to PTB domain interactions and may play a role in the selectivity of tensin PTB domain recognition by integrins. Integrins β2 and β3 were chosen to simulate binding, the experimental data providing a basis for evaluating the significance of modeling results.
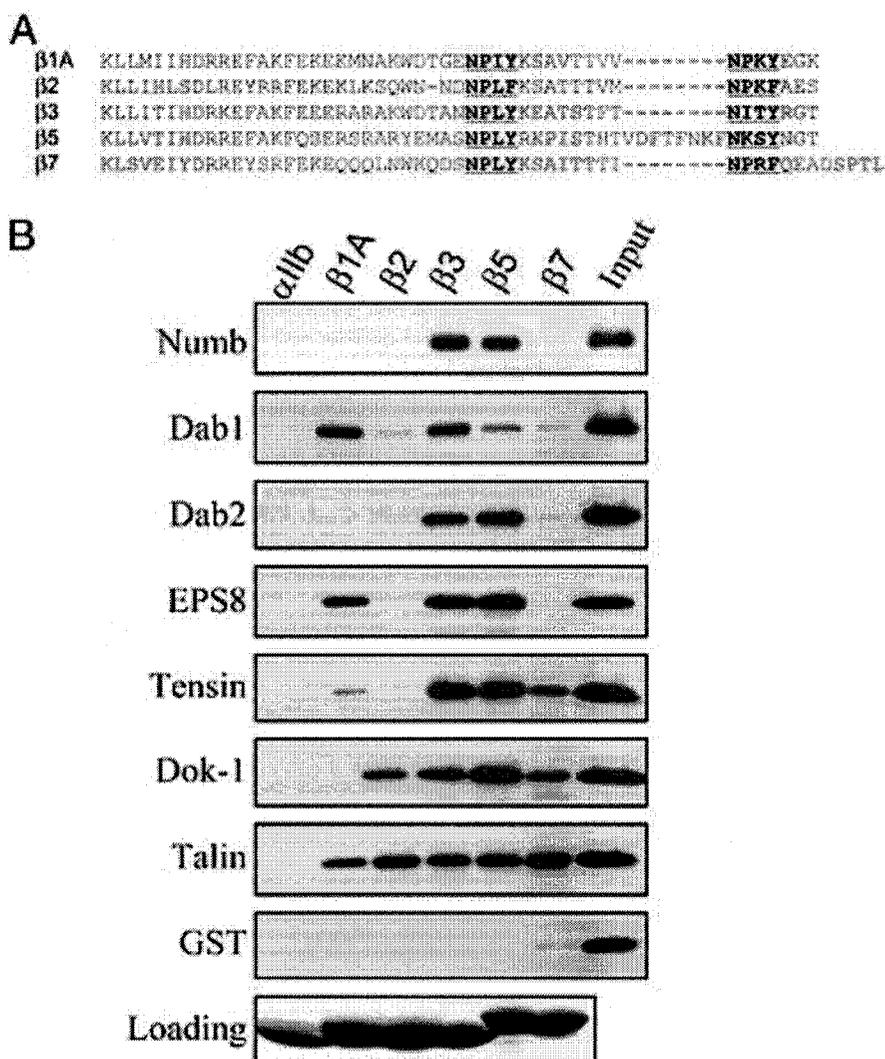
Figure 4.20 PTB domains bind to integrin cytoplasmic tails. (A) An alignment of the amino acid sequences of β integrin cytoplasmic tails. The NPXY or NPXY-like motifs are in bold and underlined. (B) PTB domains were incubated with beads coated with recombinant αIIb, β1A, β2, β3, β5, and β7 cytoplasmic tails. Figure from ref [23].

Table 4.14 shows the calculated interaction energy between two NPXY/F motifs in integrin β2 and β3 tail ligands and PTB domain of tensin. The starting structure of the tensin complex was based on the model of the X11 domain binding to unphosphorylated peptides which contain an NPXY motif; the X11 complex resembles that of phosphorylated peptides bound to the Shc and IRS-1 PTB domains. The

unphosphorylated peptide comprises 12 amino acids. The two NPXY/F motifs in integrin

β2 and β3 were superimposed to this peptide. The ligands were not phosphorylated, as in

the corresponding experimental data [23].

Table 4.14 Interaction energy between the integrin β2 and β3 NPXY/F ligands and the PTB domain of tensin. VDW, van der Waals interaction energy. ELE, electrostatic interaction energy; TOT, total interaction energy.

Ligand	First NPXY/F motif	Second NPXY/F motif
Integrin β2	(NNDNPLFKSA) VDW = − 55.8 ELE = − 49.2 TOT = −105.0	(TVMNPKFAES) VDW = −88.9 ELE = − 40.5 TOT = − 129.4
Integrin β3	(TANNPLYKEA) VDW = − 55.3 ELE = − 107.6 TOT = − 162.8	(TFTNITYGRT) VDW = − 62.7 ELE = − 139.5 TOT = −202.2

The data in Table 4.14 suggest that integrin β3 will bind tensin with higher

affinity than β2. The difference of the binding affinity comes mainly from electrostatic

interactions. Van der Waals interactions are on the same order of magnitude in the two

cases. The two NPXY/F motifs bind to the PTB domain of tensin with somewhat

different affinity: The second one binds more tightly to the tensin PTB domain than the

first one. This suggests that one of the motifs of integrin is more responsible for

interaction with tensin in cells than the other motif. It would be interesting to determine

the binding affinity of integin to PTB domain by experiemental methods, with one

NPXY/F motif as the control. It would also be interesting to determine if

phosphorylation of tyrosine of NPXY motif in integrin β3 affects binding affinity.

4.4 Conclusion

Realistic models of the four known modular domains of human tensin (PTP, C2, SH2 and PTB) have been generated by structure-based alignment and homology modeling. Pros and cons of study of each domain have been adduced. Each domain structure has been characterized qualitively and quantitively by Procheck and other commonly used criteria such as hydrogen bonds per amino acid and surface area. The best PTP and PTB domain models were selected for functional studies. Several ligands of the PTP domain of PTEN have been used to study the binding properties of tensin. Both PTEN and tensin have been studied by the same protocol as a control. The results are reasonable on the grounds of comparion to the experiemenal data for PTEN and predictions based on sequence and structural analysis. The NPXY/F motif of integrin $\beta 2$ and $\beta 3$ tails has been studied as ligands of the tensin PTB domain on the basis of credible experiemtal data. The results suggest the differences in binding affinity between the two integins, which reproduce the corresponding experimental data. The computational work has raised interesting questions and helped to provide a foundation for further study of the molecular basis of cell adhesion.

CHAPTER 5

SUMMARY

This dissertation focused on simulation of the structure and function of peptides and proteins related to cell adhesion *in vivo* and *in vitro*, specifically, polypeptide multilayer nanocoatings and the various molecular domains of the focal adhesion protein tensin. Both objects of the simulation work are closely related to experimental studies in the Haynie laboratory, and both have applications in medicine, biotechnology, nanotechnology and other areas.

In Chapters 2 and 3, MD studies were used to as a general means to understand better how various types of non-covalent interaction contribute to the structure and stability of polypeptide LBL films.

Chapter 2 covered the physical basis of stability of designed polypeptide multiplayer nanofilms in vacuum and in implicit solvent. Three pairs of peptides were designed to study contributions to film formation and stability made by electrostatic interactions (Pair 1, $K_{31}Y$ and $E_{31}Y$), hydrophobic interactions (Pair 2, $(KV)_{15}KY$ and $(EV)_{15}EY$), and hydrogen bonding (Pair 3, $(KVKN)_7KVKY$ and $(EVEN)_7EVEY$). MD simulations of IPECs of the designed peptides pairs have been carried out in vacuum and in implicit solvent, and heuristic models of peptide assembly behavior have been developed. The IPECs model peptide interactions in the corresponding multilayer films

122

and the models help to explain film growth, particularly mass deposited per adsorption step. Three distinct classes of IPEC conformation have been found at equilibrium, as determined by starting configuration and amino acid sequence. Simulation results showed remarkable correlation with experimental data on the same peptide designs. In particular, there is a relationship between IPEC potential energy and surface roughness of the corresponding polypeptide multilayer film. The results of this study constitute a step toward predicting the multilayer film assembly behavior of polypeptides and other polyelectrolytes of arbitrary structure. This simulation work has broken new ground in LBL assembly research. As the peptides are 32-amino acid long, it would be computationally unaffordable at this time to carry out the simulation in explicit solvent. The research reported here has prepared the way for further work. Future study can examine how larger numbers of the designed 32 mer polypeptides interact in aqueous and in vacuum.

In Chapter 3, MD simulations of all-atom models have been used for the first time to gain information on the stability of multilayer thin films in explicit solvent. The simulations were carried out to study structural and dynamical properties of peptide systems involving the sequences EEEEEY and KKKKKY. The sequences were so short in order to make the simulations computationally tractable. The results constitue a detailed picture of new peptides in LBL film but also shed light on the understanding the physical basis of peptide LBL. The simulations suggest that hydrophobic interactions play a key role in peptide multilayer formation, in addition to electrostatic interactions. The simulations also suggest that hydrogen bonds are more a consequence of multilayers formation than the cause of it. As the number of peptides in a supramolecular structure

goes up, there is increased likelihood that it will be stable. The most stable suparmolecular structure was parallel between sheets and anti-parallel within sheets. The importance of hydrophobic interactions is supported by experimental results involving poly(L-lysine) and poly (L-glutamic acid). Due to the limitations of current computational ability, it is not possible to simulate the entire process of forming secondary structure by peptides from random conformations in explicit solvent. This research has studied peptide aggregates up to hexamers. The same method can be used for larger molecular systems with longer durations when the necessary computational resourse become available.

In Chapter 4, the protein tensin, a component of cell-substrate contacts with close connections to cancer, has been used as an investigation tool, to elucidate the physiology of cell-substrate contacts at the molecular scale. Such contacts play a key role in cell migration, cell differentiation, and wound healing. The near-term aim concerned determination of the structure and function of tensin. This aim has been achieved by homology modeling and MD. The structure and function of all four known domains – PTP, C2, SH2 and PTB have been modeled. Binding studies involving the PTP and PTB domains have been initiated, and there sults provide a basis for further work by more sophisticated protocols. The phosphorylation of tyrosine in the integrin β ligand could be studied by the same protocol used here. The longer-term aim of this research is to use the protein tensin as the tool to more fully elucidate the physiology of cell-substrate contacts at the molecular scale and develop a molecular-scale.

APPENDIX

Protocol for Room Temperature Relaxation

125

! Room temperature relaxation protocol

* overlap = 0.01

* cutoff = 50.0

* cutdis = 14.0

* swtdis = 1.5

* dielectric = 1.0

1. Hold the protein fixed and the crystallographic water oxygen atoms

! Solvent water molecules are free to move.

Fixed atom list generation

* add all

* ligand all residues

* add heavy

* protein receptor all residues

! Do preliminary minimization to remove close atom contacts

Minimize

* no cross terms

* no morse

* for 1000 iterations

* using steep descents

* until the maximum derivative is less than 20.0 kcal/Å

2. Relax the system more completely

Minimize

* no cross terms

* no morse

* for 3000 iterations

* using conjugate gradient

* until the maximum derivative is less than 0.250 kcal/Å

3. Do minimal dynamics to allow the solvent molecules to orient

Initialize dynamics

* for 1000 iterations at 300.0 K steps of 1.0

* no cross terms no morse

* write averages every 100 steps

* write history file every 1000 steps

Resume dynamics

* for 5000 iterations

* no cross terms no morse

4. Release all atoms

Fix nothing

! Hold the protein backbone atoms, ligand and water to their initial positions

! Solvent layer can move freely

* clear tethered list

* add heavy

* protein receptor all residues

* add heavy

* ligand all residues

! Start with a large force constant.

5. Do preliminary minimization with Steepest Decendents to remove close atom contacts

Template force with a force constant of 1000.0 kcal/ ^{2}

* no cross terms

* no morse

* for 1000 iterations

* using steep descents

* until the maximum derivative is less than 20.0 kcal/

6. Switch to Conjugate Gradient minimization

Template force with a force constant of 1000.0 kcal/ ^{2}

* no cross terms

* no morse

* for 5000 iterations

* using conjugate gradient

* until the maximum derivative is less than 1.00 kcal/

7. Reduce the force constant

Template force with a force constant of 100.0 kcal/ ^{2}

* no cross terms

* no morse

* for 5000 iterations

* using conjugate gradient

* until the maximum derivative is less than 1.00 kcal/

8. Reduce the force constant further and add cross terms

Template force with a force constant of 50.0 kcal/ ^{2}

* no morse

* for 5000 iterations

* using conjugate gradient

* until the maximum derivative is less than 1.00 kcal/

9. Reduce the force constant further

Template force with a force constant of 15.0 kcal/ ^{2}

* no morse

* for 5000 iterations

* using conjugate gradient

* until the maximum derivative is less than 0.50 kcal/

10. Reduce the force constant further and add Morse potentials

Template force with a force constant of 2.0 kcal/ ^{2}

* for 5000 iterations

* using conjugate gradient

* until the maximum derivative is less than 0.50 kcal/

11. Finally remove tethering restraints and fully relax at 0 K

Clear tethered list

* Minimize for 5000 iterations

* using conjugate gradient

* until the maximum derivative is less than 0.10 kcal/

end

REFERENCES

1. Sastry, S. K., Burridge, K. 2000. Minireview: Focal adhesions: A nexus for intracellular signaling and cytoskeletal Dynamics. *Experimental cell research* **261**: 25-36

2. Jockush, B. M., Bubeck, P., Giehl, C., Kroemker, M., Moschner, J., Rothkegel, M., Rudiger, M., Schluter, K., Stanke., winkler J. 1995. The molecular architecture of focal adhesions. *Annual Review of Cell Developmental Biology* **11**: 379-416

3. Geiger, B., Yehuda-Levenberg, S., Bershadsky, A. D. 1995. Interactions in the membrane-associated plaque of cell-cell and cell-matrix adhesions. *ACTA Anatomica* **154**:46-62

4. Gumbiner B. M. 1996. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* **84**: 345-357

5. Yamada K. M., Geiger B 1997 Molecular interactions in cell adhesion complexes. *Current Opinion in Cell Biology* **9**: 76-85

6. Critchley, D. R. 2000 Focal adhesions; the cytoskeletal connection. *Current Opinion in Cell Biology* **12**: 133-139

7. Grinnell, F., Feld, M.K. 1982. Fibronectin adsorption on hydrophilic and hydrophobic surfaces detected by antibody binding and analyzed during cell adhesion in serum-containing medium. *The Journal Biological Chemistry* **257**: 4888-4893

8. Dunn, G. A., Brown, A.F. 1986. Alignment of fibroblasts on grooved surfaces described by a simple geometric transformation. *Journal of Cell Science* **83**:313-340

9. Chen, C. S., Mrksich, M., Huang, S., Whitesides, G.M., Ingber, D.E. 1998. Micropatterned surfaces for control of cell shape, position, and function. *Biotechnology Progress* **14**: 356-363

10. Chen, C.S., Alonso, J.L., Ostuni, E., Whitesides, G.M., Ingber, D.E. 2003 Cell shape provides global control of focal adhesion assembly. *Biochemical and Biophysical Research Communications* **307**: 355-361

11. Pethica, B.A. 1961. The physical chemistry of cell adhesion *Experimental Cell Research* **8** (Suppl.): 123- 140

12. Curtis, A.S. 1962. Cell contact and adhesion. *Biol. Rev. Camb. Philos. Soc.* **37**: 82-129

13. Qiu, Q., Sayer, M., Kawaja, M., Shen, X., Davies, J.E. 1998. Attachment, morphology, and protein expression of rat marrow stromal cells cultured on charged substrate surfaces *Journal of Biomedical Matererials Research* **42**: 117-127

14. Dufrêne, Y. F., Marchal, T.G.., Rouxhet, P.G. 1999. Influence of substratum surface properties on the organization of adsorbed collagen films: in situ characterization by atomic force microscopy *Langmuir* **15**: 2871- 2878

15. Absolom, D.R., Zingg, W., Neumann, A.W. 1987. Protein adsorption to polymer particles: role of surface properties. *Journal of Biomedical Materials Research* **21**:161-171

16. Kapur, R., Rudolph, A.S. 1998. Cellular and cytoskeleton morphology and strength of adhesion of cells on self-assembled monolayers of organosilanes. *Experimental Cell Research* **244**: 275-285

17. Curtis, A.S., Varde, M. 1964. Control of cell behavior: topological factors. *National Cancer Institute* **33**: 15-26

18. Cowin, P., Burke B. 1996. Cytoskeleton-membrane interactions. *Current Opinionin Cell Biology* **8**: 56-65

19. Mangeat P., Roy C., Martin M. 1999. ERM proteins in cell adhesion and membrane dynamics. *Trends in Cell Biology* **9**:187-192

20. Izzard C.S., Lochner L.R. 1980. Formation of cell-to-substrate contacts during fibroblast motility: an interference-reflexion study. *Journal of Cell Science* **42**: 81-116

21. Bershadsky A.D., Tint I. S., Neyfakh A., Vasiliev J.M. 1985. Focal contacts of normal and RSV-transformed Quail cells, Hypothesis of the transformation-induced deficient maturation of focal contacts. *Experimental Cell Research* **158**: 433-444

22. Izzard, C.S. 1988. A precursor of the focal contact in cultured fibroblasts. *Cell Motility and Cytoskeleton* **10**: 137-142

23. Calderwood, D.A., Fujioka, Y., de Pereda, J.M., García-Alvarez B., Nakamoto T., Margolis B., McGlade C.J., Liddington R.C., Ginsberg M.H. 2003. Integrin β cytoplasmic domain interactions with phosphotyrosine-binding domains: A structural prototype for diversity in integrin signaling. *Proceedings of National Acadamy Science* **100**: 2272-2277

24. Chuang, J.-Z., Lin, D.C., Lin, S. 1995. Molecular cloning, expression, and mapping of the high affinity actin-capping domain of chicken cardiac tensin. *Journal of Cell Biology* **128**: 1095-1109

25. Vuori, K. 1998 Integrin signaling: tyrosine phosphorylation events in focal adhesions. *Journal of Membrane Biology* **165**: 191-199

26. Burridge, K., Chrzanowska-Wodnicka M. 1996. Focal adhesions, contractility, and signaling. *Annual Review of Cell Developmental Biology* **12**: 463-518

27. Volberg, T., Geiger, B., Kam, Z., Pankov, R., Simcha, I., Sabanay, H., Coll, J. L., Adamson, E., Ben-Ze'ev, A. 1995. Focal adhesion formation by F9 embryonal carcinoma cells after vinculin gene disruption. *Journal of Cell Science* **108**: 2253-2260

28. Aplin, A.E., Howe, A., Alahari, S.K., Juliano, R. L. 1998. Signal transduction and signal modulation by cell adhesion receptors: the role of integrins, cadherins, immunoglobulin-cell adhesion molecules, and selectins. *Pharmacological Review* **50**: 197-263

29. Lodish, H., Baltimore, D., Berk, A., Zipursky, S.L., Matsudaira, P., Darnell, J. 1995. *Molecular Cell Biology*, (3rd ed. Scientific American Books, New York)

30. Gilbert, S.F. 2003. *Developmental Biology*, (7th ed. Sunderland, MA: Sinauer).

31. Sasaki, T., Wiedemann, H., Matzner, M., Chu, M.L., Timpl, R. 1996. Expression of fibulin-2 by fibroblasts and deposition with fibronectin into a fibrillar matrix. *Journal of Cell Science* **109**:2895-2904

32. Pereira, M., Rybarczyk, B.J., Odrljin, T.M., Hocking, D.C., Sottile, J., Simpson-Haidaris, P. J. 2002. The incorporation of fibrinogen into extracellular matrix is dependent on active assembly of a fibronectin matrix. *Journal of Cell Science* **115**: 609-617

33. Sottile, J., Hocking, D.C. 2002. Fibronectin polymerization regulates the composition and stability of extracellular matrix fibrils and cell-matrix adhesions. *Molecular Biology of the Cell* **13**:3456-3559

34. Perks, C.M., Newcomb, P.V., Norman, M.R., Holly, J.M. 1999. Effect of insulin-like growth factor binding protein-1 on integrin signalling and the induction of apoptosis in human breast cancer cells. *Journal of Molecular of Endocrinology* **22**:141-150

35. Creighton, T.E. 1993 *Proteins: Structures and Molecular Properties*, (2nd ed. New York: Freeman)

36. Haynie, D.T., Zhang, L., Zhao, W., Rudra, J., Zhong, Y. and Palath, N. 2005. Polypeptide multilayer films, *Biomacromolecules* **6**: 2895-913

37. Haynie D.T., Zhang, L., Zhao, W., and Smith, J.M. 2006. Quantal Self-Assembly of Polymer Layers in Polypeptide Multilayer Nanofilms. Submitted to *Biomacromolecules*

38. Zhang, L., Zhao, W. and Haynie, D.T., 2006. Peptide design for polypeptide multilayer films of desired structure and properties, in preparation to submit to *Biomacromolecules*

39. Zhang, L., Zhao, W., Rudra, J. and Haynie, D.T., 2006. "Green" multifunctional materials engineering at the nanoscale: polypeptide multilayer films in preparation to submit to *Nature Materials*

40. Chou, P.Y.; Fasman, G.D. 1974. Conformational parameters for amino acids in helical, β Sheet, and random coil regions calculated from proteins *Biochemistry*. **13**: 211-222

41. Wilkins, J.A., Risinger, M.A., Lin, S. 1986. Studies on proteins that co-purify with smooth muscle vinculin: identification of immunologically related species in focal adhesions of nonmuscle and Z-lines of muscle cells. *Journal of Cell Biology* **103**: 1483-1494

42. Lo, S.H., Janmey, P.A., Hartwig, J. H., Chen, L. B. 1994. Interactions of tensin with actin and identification of its three distinct actin-binding domains. *Journal of Cell Biology* **125**: 1067-1075

43. Chen, H., Ishii, A., Wong, W. K., Chen, L.B., Lo, S.H. 2000. Molecular characterization of human tensin. *The Biochemical Journal* **351**: 403-411

44. Miyamoto, S., Akiyama, S.K., Yamada, K.M. 1995. Synergistic roles for receptor occupancy and aggregation in integrin transmembrane function. *Science* **267**: 883-885

45. Chen, H., Lo, S.H. 2003. Regulation of tensin-promoted cell migration by its focal adhesion binding and Src homology domain 2. *The Biochemical Journal* **370**: 1039–1045

46. Chen, H, Duncan, I. C., Bozorgchami, H., Lo, S. H. 2002. Tensin 1 and a previously undocumented family member, tensin2, positively regulate cell migration. *Proceedings of National Acadamy of Science* **99**: 733-738

47. Lo, S. H., Yu, Q. C., Degenstein, L., Chen, L. B., Fuchs, E. 1997. Progressive kidney degeneration in mice lacking tensin. *Journal of Cell Biololgy* **136**: 1349–1361

48. Ishii, A., Lo, S. H. 2001. A role of tensin in skeletal-muscle regeneration. *The Biochemical Journal* **356**: 737-745

49. Walton, K. M., Dixon, J. E. 1993. Protein tyrosine phosphatases. *Annunal Review of Biochemistry* **62**:101-120

50. Alonso, A., Sasin J., Bottini N., Friedberg I., Friedberg I., Osterman A., Godzik A., Hunter T. Dixon J., and Mustelin T. 2004. Protein Tyrosine Phosphatases in the Human Genome. *Cell* 117: 699–711

51. Haynie, D. T., Ponting, C. P. 1996. The N-terminal domains of tensin and auxilin are phosphatase homologues. *Protein Science* 5: 2643-2646

52. Davis, S., Lu M. L., Lo S. H., Lin S., Butler J.A., Druker B. J., Roberts, T. M., An Q., Chen L. B. 1991. Presence of an SH2 domain in the actin-binding protein tensin. *Science* 252: 712-715

53. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of National Academy of Science USA* 95: 5857–5864

54. Steck, P.A., Pershouse, M.A., Jasser, S. A., Yung, W. K., Lin, H, Ligon, A. H., Langford, L. A., Baumgard, M. L., Hattier, T., Davis T., Frye C., Hu R., Swedlund B., Teng D.H., Tavtigian S.V. 1997. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature Genetics* 15: 356-362

55. Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S. I., Puc, J., Miliaresis, C., Rodgers, L. McCombie, R., Bigner, S. H., Giovanella, B. C., Ittmann, M., Tycko B., Hibshoosh H., Wigler M.H., Parsons R. 1997. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* 275: 1943-1946

56. Schroeder, S., Morris, S., Knorr, R., Plessmann, U., Weber, K., Nguyen, G. V., Ungewickell, E. 1995. Primary structure of the neuronal clathrin-associated protein auxilin and its expression in bacteria. *European Journal of Biochemistry* 228: 297-304

57. Pankov, R., Cukierman, E., Katz, B.Z., Matsumoto, K., Lin, D.C., Lin, S., Hahn, C., Yamada, K.M. 2000. Integrin dynamics and matrix assembly: tensin-dependent translocation of $\alpha_5\beta_1$ integrins promotes early fibronectin fibrillogenesis. *Journal of Cell Biology* 148:1075-1090

58. Kanaoka, Y., Kimura, S.H., Okazaki, I., Ikeda, M., Nojima, H. 1997. GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain. *FEBS Letters* 402: 73-80

59 Salgia, R., Brunkhorst, B., Pisick, E., Li, J.-L., Lo S.H., Chen L.B., Griffin J. D. 1995. Increased tyrosine phosphorylation of focal adhesion proteins in myeloid cell lines expressing p210BCR/ABL. *Oncogene* 11: 1149-1155

60. Myers, M. P., Stolarov, J. P., Eng, C., Li, J., Wang, S. I., Wigler, M. H., Parsons, R., Tonks, N. K. 1997. P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. *Proceedings of National Academy of Science USA* 94: 9052-9057

61. Maehama, T., Dixon, J. E. 1998. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *Journal of Biological Chemistry* **273**: 13375–13378

62. Barford, D, Flint A. J., Tonks N. K. 1994. Crystal structure of human protein tyrosine phosphatase 1B. *Science* **263**:1397-1404

63. Yuvaniyama, J., Denu J.M., Dixon J. E., Saper M. A. 1996. Crystal structure of the dual specificity protein phosphatase VHR. *Science* **272**: 1328-1331

64. Alonso, A., Sasin, J., Bottini, N., Friedberg I., Friedberg I., Osterman A., Godzik A., Hunter T. Dixon J., and Mustelin T. 2004. Protein tyrosine phosphatases in the human genome. *Cell* **117**: 699–711

65. Rizo, J., Sudhof, T. C. C2-domains, Structure and function of a universal Ca21-binding domain. 1998. *Journal of boilogical Chemistry* **273**: 15879–15882

66. Hurley, J. H., Misra, S. 2000. Signaling and subcellular targeting by membrane-binding domains. *Annunal Review of Biophysics and Biomolecular Structure* **29**:49-79

67. Lee, J. O., Yang, H., Georgescu M.M., Di Cristofano. A., Maehama. T., Shi. Y., Dixon. J. E., Pandolfi. P., Pavletich. N. P. 1999. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell* **99**: 323-334

68. Sadowski, I., Stone, J.C., Pawson. T. 1986. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of fujinami sarcoma virus p130GAG-FPS. *Molecular Cell Biology* **6**: 4396-4408

69. Koch, C.A., Anderson, D., Moran, M.F., Ellis, C., Pawson, T. 1991. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* **252**: 668-674

70. Pawson, T., Scott J.D. 2005. Protein phosphorylation in signaling--50 years and counting. *Trends in Biochemical Science* **30**:286-290

71. Kuriyan, J., Darnell, J. E. Jr 1999. An SH2 domain in disguise. *Nature* **398**: 22-23, 25

72. Sawyer, T. K. 1998. Src homology-2 domains: structure, mechanisms, and drug discovery. *Biopolymers* **47**: 243–261

73. Grucza, R.A., Bradshaw, J.M., Fu¨tterer, K. and Waksman, G. 1999. SH2 domains: from structure to energetics, a dual approach to the study of structure–function relationships. *Medical Research Reviews* **19**: 273–293

74. Smithgall. T.E. 1995. SH2 and SH3 domains: potential targets for anti-cancer drug design. *J. Pharmacol. Toxicol. Methods*. **34**:125-132

75. Branden, C.-I.; Tooze, J. 1999. *Introduction to Protein Structure*, (2nd ed. New York: Garland)

76. Wandless, T.J. 1996. SH2 domains: a question of independence. *Current Biology* **6**: 125-127

77. Kashishian, A., A. Kazlauskas, J. A. Cooper. 1992. Phosphorylation sites in the PDGF receptor with different specificities for binding GAP and PI3 kinase in vivo. *EMBO Journal* **11**: 1373-1381

78. Fantl, W.J., J.A. Escobedo, G.A. Martin, C.W. Turck, M. Rosario, F. McCormick, L.T. Williams. 1992. Distinct phosphotyrosines on a growth factor receptor bind to specific molecules that mediate different signaling pathways. *Cell* **69**: 413-423

79. Songyang, Z., S.E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W.G. Haser, F. King, T. Roberts, S. Ratnofski, R.J. Lechleider, B.G. Neel, R.B. Birge, J.E. Fajardo, M.M. Chou, H. Hanafusa, B. Schaffhausen, L.C. Cantley. 1993. SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**: 767-778

80. Songyang, Z., S.E. Shoelson, J. McGlade, P. Olivier, T. Pawson, X.R. Bustelo, M. Barbacid, H. Sabe, H. Hanafusa, T. Yi, R. Ren, D. Baltimore, S. Ratnofski, R.A. Feldman, L.C. Cantley. 1994. Specific motifs recognized by the SH2 domains of CSK, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, Vav. *Molecular Cell Biology* **14**: 2777-2785

81. Charifson, P.S. Shewchuck, L. M., Rocque, W., Hummel, C.W., Jordan, S.R., Mohr, C., Pacofsky, G.J., Peel, M.R., Rodriguez, M., Sternbach, D.D., Consler, T.G. 1997. Peptide ligands of pp60c-src SH2 domains: a thermodynamic and structural study. *Biochemistry* **36**: 6283-6293

82. Zhao, W., Cavallaro, S., Gusev, P., Alkon, D. L. 2000. PP60C-SRC as a synapse-specific regulatory molecule for spatial learning: specific changes in its gene expression, tyrosine phosphorylation, and protein-protein interaction. *Proceedings of National Academy of Science USA* **97**: 8098-8103

83. Kuriyan, J, Cowburn, D. 1997. Modular peptide recognition domains in eukaryotic signaling. *Annual Review of Biophysics and Biomolecular Structure* **26**: 259-288

84. Torgler, C. N., Narasimha M., Knox A. L., Zervas C. G., Vernon M. C., Brown N. H. 2004. Tensin stabilizes integrin adhesive contacts in *Drosophila. Developmental Cell* **6**: 357-369

85. Forman-Kay, J. D., Pawson, T. 1999. Diversity in protein recognition by PTB domains *Current Opinion in Structural Biology* **9**:690-695

86. Zhou, M.-M., Ravichandran, K. S., Olejniczak E. T., Petros A. M., Meadows R. P., Sattler M., Harlan J. E., Wade W. S., Burakoff S. J., Fesik S. W. 1995 Structure and ligand recognition of the phosophotyrosine binding domain of Shc. *Nature* **378**: 584-592

87. Shoelson, S.E. 1997. SH2 and PTB domain interactions in tyrosine kinase signal transduction. *Current Opinion of Chemical Biology* **1**:227-234

88. Van der Geer, P., Pawson, T. 1995. The PTB domain: a new protein module implicated in signal transduction. *Trends in Biochemical Science* **20**: 277-280

89. Kavanaugh, W. M., Williams L. T. 1994. An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science* **266**:1862-1865

90. Blaikie, P., Immanuel D., Wu J., Li N., Yajnik V., Margolis B. 1994 A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. *The Journal of Biological Chemistry* **269**: 32031-32034

91. Margolis, B., Borg J.-P., Straight S., Meyer D. 1999. The function of PTB domain proteins. *Kidney International*, **56**: 1230-1237

92. Lee, S. B., Cho, K. S., Kim, E., Chung, J. 2003. blistery encodes Drosophila tensin protein and interacts with integrin and the JNK signaling pathway during wing development. *Development* **130**: 4001-4010

93. Decher, G. 1997. Fuzzy nanoassemblies toward layered polymeric multicomposites. *Science* **277**: 1232-1237

94. Tripathy, S.K., Kumar, J., Nalwa, H.S. 2002. *Handbook of Polyelectrolytes and their Applications. Vol.1. Polyelectrolyte-based Multilayers, Self-assemblies and Nanostructures* (Stevenson Ranch, California: American Scientific Publishers)

95. Decher, G., Schlenoff, J.B. 2003. *Multilayer Thin Films: Sequential Assembly of Nanocomposite Materials* (Weinheim: Wiley-VCH)

96. Iler, R.K. 1966. Multilayers of colloidal particles *Journal of Colloid Interface Science* **21**: 569-594

97. Lvov, Y., Möhwald, H. 2000. *Protein Architecture: Interfacing Molecular Assemblies and Immobilization Biotechnology* (New York: Marcel Dekker)

98. Anzai, J., 2001. Development of polyelectrolyte multilayer films and their applications to analytical chemistry. *Bunseki Kagaku* **50**: 585-594

99. Bertrand, P., Jonas, A., Laschewsky, A., Legras, R. 2000. Ultrathin polymer coatings by. complexation of polyelectrolytes at. interfaces: suitable materials, structure and. properties. *Macromolecule Rapid Communications* **21**: 319-348

100. Chen, W., McCarthy, T.J. 1997, Layer by layer deposition: A tool for polymer surface modification. *Macromolecules* **30**: 78-86

101. 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

102. Hammond, P. T. 2000. Recent explorations in electrostatic multilayer thin film assembly. *Current Opinion in Colloid Interface Science* **4**: 430-442

103. Schönhoff, M. 2003. Self-assembled polyelectrolyte multilayers. *Current Opinion in Colloid Interface Science* **8**: 86-95

104. Hammond, P.T. 2004. Form and Function in Multilayer Assembly: New applications at the Nanoscale *Advanced Matererial* **16**: 1271- 1293

105. Müller, M., Meier-Haack, J., Schwarz, S., Buchhammer, H.M., Eichhorn, K. J., Janke, A., Keßler, B., Nagel, J., Oelmann, M., Reihs, T., Lunkwitz, K. 2004. Polyelectrolyte Multilayers and their interactions. *Journal of Adhesion* **80**: 521-548

106. Rubner, M.F. In Decher, G. and Schlenoff, J.B. 2003. (Eds) *Multilayer Thin Films: Sequential Assembly of Nanocomposite Materials* (Weinheim: Wiley-VCH) p. 133

107. Schmitt, J., Grünewald, T., Decher, G., Pershan, P.S., Kjaer, K., Lösche, M. 1993. Internal structure of layer-by-layer adsorped polyelectrolyte films: A neutron and x-ray reflectivity study *Macromolecules*. **26**: 7058-7063

108. Decher, G., Lvov, Y., Schmitt, J. 1994. Proof of multilayer structural organization in self-assembled polyanion-polycation molecular films *Thin Solid Films* **244**: 772-777

109. Zheng, B., Haynie, D. T., Zhong, H.; Sabnis, K., Surpuriya, V. 2005. Design of peptides for thin films, coatings and microcapsules for applications in biotechnology Journal of *Biomaterial Science Polymer Edition* **16**: 285-300

110. Altschul, S.F., Gish W., Miller, W., Myers E.W., Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Boilogy* **215**: 403-410

111. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402

112. Holm, L., Sander, C. 1996. Mapping the protein universe. *Science* **273**: 595–603

113. Elmar, Krieger, Sander B. Nabuurs, and Gert Vriend 2003. *Structural Bioinformatics* Edited by Philip E. Bourne and Helge Weissig (Wiley-Liss, Inc. 507-521)

114. Becker, O. M., MacKerell A. D Jr., Roux, B. 2001. *Computational Biochemistry and Biophysics.* (Marcel Dekker Inc.)

115. Karplus, M., and McCammon J. A. 2002. Molecular dynamics simulations of biomolecules. *Natural Structural Biology* 9:646-652.

116. Jurs, P.C. 1986. *Computer Software Applications in Chemistry.* (John Wiley & Sons Inc.)

117. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., Hagler, A. T. 1988. Structure and energetics of ligand binding to proteins: E. coli dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins* 4: 31-47

118. Weiner, S. J., Kollman, P. A., Case, D. A.; Singh, U. C.; Ghio, C., Alagona, G., Profeta, S. Jr., Weiner, P. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of American Chemical Society* 106: 765–784

119. Weiner, S. J., Kollman, P. A., Nguyen, D. T., Case, D. A. 1986. An all atom force field for simulations of proteins and nucleic acid. *Journal of Computational Chemistry* 7: 230–252

120. Brooks, B. R.Bruccoleri, R.E., Olafson, B. D., States, D. J., Swaminathan S. and Karplus M. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* 4: 187-217

121. Ryckaert, J.P., Ciccoti, G., Berendsen, H.J.C. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes *Journal of Computational Physics* 23: 327-341

122. Andersen, H. 1983. Rattle: a velocity version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* 52: 24–34

123. Onufriev, A., Bashford D., and Case D.A. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* 55: 383-394

124. Lybrand, T. P. 1997. *Protein-ligand interactions. Computational Approaches to Biochemical Reacitivity* (Kluwer Academic Publishers)

125. Greer, J. 1991. Comparative modeling of homologous proteins. *Methods in Enzymology* 202: 239–252

126. Šali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. 1995 Evaluation of comparative protein modeling by MODELLER. *Proteins* 23: 318-326

127. Humphrey, W., Dalke, A. and Schulten, 1996. K., VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics and Modelling* **14**: 33-38

128. Caruso, F., Caruso, R. A., Mvhwald, H. 1998. Nanoengineering of inorganic and hybrid hollow spheres by colloidal templating. *Science* **282**:1111-1114

129. Sukhorukov, G. B., Brumen, M., Donath, E., Möhwald, H. 1999. Hollow polyelectrolyte shells: Exclusion of polymers and donnan equilibrium. *Journal of Physical Chemistry B* **103**: 6434-6440

130. Lvov, Y., Antipov, A.A., Mamedov, A.; Möhwald, H.; Sukhorukov, G.B. 2001. Urease encapsulation in nanoorganized microshells. *Nano Letters* **1**: 125-128

131. Peyratout, C. S., Dähne, L. 2004. Tailor-made polyelectrolyte microcapsules: from multilayers to smart containers. *Angewandte Chemie International Edition* **43**: 3762-3783

132. Haynie, D. T., Palath, N., Liu, Y., Li, B. and Pargaonkar, N. 2005. Biomimetic nanostructured materials: inherent reversible stabilization of polypeptide microcapsules. *Langmuir* **21**: 1136-1138

133. Haynie, D. T. Physics of polypeptide multilayer films. *Journal of Biomedical Material Research B. Applied Biomaterials* DOI: 10.1002/jbm.b.30479

134. Halthur, T.J., Claesson, P.M., Elofsson, U.M. 2004. Stability of polypeptide multilayers as studied by in situ ellipsometry: Effects of drying and post-buildup changes in temperature and pH. *Journal of American Chemical Society* **126**: 17009-17015

135. Richert, L., Arntz, Y., Schaaf, P., Voegel, J.-C., Picart, C. 2004. PH dependent growth and cell adhesion properties of poly(L-Lysine)/poly(L-glutamic) acid multilayer films. *Surface Science* **570**: 13-29

136. Jessel, N.B., Lavalle, P.; Meyer, F., Audouin, F.; Frisch, B., Schaaf, P., Ogier, J., Decher, G., Voegel, J.-C. 2004. Control of Monocyte Morphology on and Response to Model surfaces for implants equipped with anti-inflammatory. Agent *Advanced Materials* **16**: 1507-1511

137. Salloum, D.S., Olenych, S.G., Keller, T.C.S., Schlenoff, J.B. 2005.Vascular smooth muscle cells on polyelectrolyte multilayers:Hydrophobicity-directed adhesion and growth. *Biomacromolecules* **6**: 161-167

138. Etienne, O., Picart, C., Taddei, C., Haikel, Y., Dimarcq, J.L., Schaaf, P., Voegel, J.C., Ogier, J.A., Egles, C. 2004. Multilayer polyelectrolyte films functionalized by insertion of defensin: a new approach to protection of implants from bacterial colonization. *Antimicrobial Agents and Chemotherapy* **48**: 3662-3669

139. Etienne, O., Gasnier, C., Taddei, C., Voegel, J.-C., Aunis, D., Schaaf, P., Metz-Boutigue, H.-M.; Bolcato-Bellemin, A.-L.; Egles, C. 2005. Antifungal coating by biofunctionalized polyelectrolyte multilayered films. *Biomaterials* **26**: 6704-6712

140. Zhi, Z.-l., Haynie, D. T. 2006. High-capacity functional protein encapsulation in nanoengineered polypeptide microcapsules. *Chemical Communication* **2**: 147-149

141. Zhi, Z.-l., Haynie, D. T. 2006. Straightforward and effective protein encapsulation in polypeptide-based artificial Cells. *Artificial Cell Blood Substitutes and immobilization Biotechnology* **34**:189-203

142. Cooper, T. M., Campbell, A. L., Noffsinger, C., Gunther-Greer, J., Crane, R. L., Adams, W. W. 1994. *Material Research Society Symposium Proceedings* **351**: 239-244

143. Boulmedais, F., Ball, V., Schwinte, P., Frisch, B., Schaaf, P., Voegel, J.-C. 2003. Buildup of exponentially growing multilayer polypeptide films with internal secondary structure *Langmuir* **19**: 440-445

144. Haynie, D. T., Balkundi, S., Palath, N., Chakravarthula, K., Dave, K. 2004. Polypeptide multilayer films: role of molecular structure and charge *Langmuir* **20**: 4540-4547

145. Kotov, N. 1999. Layer-by-layer self-assembly: the contribution of hydrophobic interactions *Nanostructured Materials* **B12**: 789-796

146. Haynie, D. T., Zhang, L., Zhao, W. 2005. Polypeptide multilayer films: experiments, simulations, implications *Polymeric Materials Science and Engineering* **93**: 94-97

147. Zhi, Z.-l., Haynie, D. T. 2004. Direct evidence of controlled structure reorganization in a nanoorganized polypeptide multilayer thin film *Macromolecules* **37**: 8668-8675

148. Zhang, L., Li, B., Zhi, Z.-l., Haynie, D. T. 2005. Perturbation of nanoscale structure of polypeptide multilayer thin films *Langmuir* **21**: 5439-5445

149. Li, B., Haynie, D. T., Palath, N., Janisch, D. 2005. Nanoscale biomimetics: fabrication and optimization of stability of peptide-based thin films *Journal of Nanoscience and Nanotechnology* **5**: 2042-2049

150. Li, B., Rozas, J., Haynie, D.T. 2006. Fine tuning of physical properties of designed polypeptide multilayer films by control of pH *Biotechnology Progress* **22**: 126-132

151. Zhong, Y., Li, B., Haynie, D. T. 2006. Structural stability of polypeptide nanofilms under extreme conditions *Biotechnology Progress* **22**: 111-117

152. Kabanov, V. In Decher, G., Schlenoff, J. B. 2003 (Eds.) *Multilayer Thin Films: Sequential Assembly of Nanocomposite Materials* (Weinheim: Wiley-VCH) 47-86

153. Rodriguez, L. N. J., De Paul, S. M., Barrett, C. J., Reven L., Spiess, H. W. 2000. Fast MAS and double quantum ^1H solid-state NMR spectroscopy of polyelectrolyte multilayers *Advanced Materials* 12:1934-1938

154. Farhat, T., Yassin, G., Dubas, S. T., Schlenoff, J. B. 1999. Water and ion pairing in polyelectrolyte multilayers. *Langmuir* 15: 6621-6623

155. Schlenoff, J. B. In Decher, G., Schlenoff, J. B. 2003. Eds. *Multilayer Thin Films: Sequential Assembly of Nanocomposite Materials.* (Published by Weinheim: Wiley-VCH) p 99-132

156. Case, D. A., Darden, T., Cheatham, T. E.,III, Simmerling C., Wang J., Merz K. M., Wang B., Pearlman D. A., Duke R. E., Crowley M., Brozell S., Luo R., Tsui V., Gohlke H., Mongan J., Hornak V., Caldwell J. W., Ross W. S., Kollman P. A. 2004. Amber 8 Manual (University of California, San Francisco)

157. Wang, J., Cieplak P., Kollman, P. A., 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *Journal of Computational Chemistry* 21:1049-1074

158. Berendsen, H.J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., Haak J. R. 1984. Molecular Dynamics with coupling to an external bath. *Journal of Chemical Physics* 81: 3684-3690

159. Messina, R. 2002. Image charges in spherical geometry : Application to colloidal systems. *Journal of Chemical Physics* 117: 11062-11074

160. Messina, R. 2003. Adsorption of Oppositely Charged Polyelectrolytes Onto a Charged rod. *Journal of Chemical Physics* 119: 8133-8139

161. Messina, R., Holm, C., Kremer, K. 2003. Polyelectrolyte multilayering on a charged sphere. *Langmuir* 19: 4473-4482

162. Messina, R., Holm, C.; Kremer, K. 2004. Polyelectrolyte adsorption and multilayering on charged colloidal particles. *Journal of Polymer Science B* 42: 3557-3570

163. Messina, R. 2004. Polyelectrolyte Multilayering on a Charged Planar Surface. *Macromolecule* 37: 621-628

164. Lowack, K., Helm, C. A. 1995. Polyelectrolyte monolayers at the mica/air interface: mechanically induced rearrangements and monolayer annealing. *Macromolecules* 28: 2912-2921

165. Hoogeveen, N. G., Cohen Stuart, M. A., Fleer, G. J., Bohmer, M. R. 1996. Formation and stability of multilayers of polyelectrolytes. *Langmuir* 12: 3675-3681

166. Hoogeveen, N. G., Cohen Stuart, M. A., Fleer, G. J. 1996. Polyelectrolyte Adsorption on Oxides: I. Kinetics and Adsorbed Amounts. *Journal of Colloid Interface Science* **182**:133-145

167. Dubas, S.T., Schlenoff, J.B. 1999. Factors Controlling the Growth of Polyelectrolyte Multilayers. *Macromolecules* **32**: 8153-8160

168. Panchagnula, V., Jeon, J., Dobrynin, A.V. 2004. Molecular dynamics simulations of electrostatic layer-by-layer self-assembly *Physical Review Letters* **93**: 037801(1-4)

169. Panchagnula, V., Jeon, J., Rusling, J.F., Dobrynin, A.V. 2004. Molecular dynamics simulations of polyelectrolyte multilayering on a charged particle. *Langmuir* **21**: 1118-1125

170. Patel, P.A., Jeon, J.; Mather, P.T.; Dobrynin, A.V. 2005. Molecular dynamics simulations of layer-by-layer assembly of polyelectrolytes at charged surfaces: effects of chain degree of polymerization and fraction of charged monomers. *Langmuir* **21**: 6113-6122

171. Zauny, D., Ma, B., Nussinov, R. 2003. Short peptide amyloid organization: stabilities and conformations of the islet amyloid peptide NFGAIL.*Biophysical Journal* **84**: 1884-1894

172. Branden, C., Tooze, J.1991. *Introduction to Protein Structure* (Garland Publishing: New York)

173. Baker, E. N., Hubbard, R. E. 1984. Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology* **44**: 97–179.

174. Kisilevsky, R. 2000. Review: amyloidogenesis-unquestioned answers and unanswered questions *Journal of Structural Biology* **130**: 99-108

175. Sunde, M., Blake, C.C. 1998. From the globular to the fibrous state: protein structure and structural conversion in amyloid formation *Q. Rev. Biophys.* **31**: 1-39

176. Ma, B., Nussinov, R. 2002. Molecular dynamics simulations of alanine rich beta-sheet oligomers: Insight into amyloid formation. *Protein Science* **11**: 2335-2350

177. Ma, B., Nussinove, R. 2002. Stabilities and conformations of Alzheimer's beta -amyloid peptide oligomers (Abeta 16-22, Abeta 16-35, and Abeta 10-35): Sequence effects. *Proceedings of National Academy of Science USA* **99**: 14126-14131

178. Finkelstein, A.V.; Ptitsyn, O. B. 2002. *Protein Physics* (Academic Press)

179. Nesloney, C.L. and Kelly, J.W. 1996. Progress towards understanding beta-sheet structure. *Bioorganic and Medical Chemistry* **4**: 739-766

180. Jorgensen, W.L., Chandrasekhar, J., Madura, J., Impey, R. W., Klein, M.L. 1982. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **79**: 926-935

181. Darden, T.; York, D.; Pedersen, L. 1993. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems *Journal of Chemical Physics* **98**: 10089-10092

182. Van de Steeg, H.G. M., Cohen Stuart, M.A., de Keizer, A., Bijsterbosch, B.H. 1992. Polyelectrolyte adsorption: a subtle balance of forces. *Langmuir* **8**: 2538-2546.

183. Cohen-Stuart, M.A. 1988. *Journal Physics (France)* Polyelectrolyte adsorption. **49**: 1001-1008

184. Cohen-Stuart, M.A.; Fleer, G.J.; Lyklema, J.; Norde, W.; and Scheutjens, J.M. 1991. Adsorption of ions, polyelectrolytes and proteins. *Advances in Colloid Interface Science* **34**: 477-535

185. Papenhuijzen, J., van der Schee, H.A.; Fleer, G.J. 1985. Polyelectrolyte adsorption. I: A new lattice theory. *Journal of Colloid Interface Science* **104**: 540-552

186. Marra, J., van der Schee, H.A., Fleer, G.J., Lyklema, J. 1983. *In Adsorption from solution.* (Ottiwel, R., Rochester, C. H., Smith, A. L., Ed. Academic Press: San Diego, CA)

187. Poland, D., Scheraga, H. A. 1970. *Theory of Helix-Coil Transitions in Biopolymers*; (Academic: New York)

188. Boulmedais, F., Ball, V.; Schwinte, P., Frisch, B., Schaaf, P. and Voegel J.-C. 2003. Buildup of exponentially growing multilayer polypeptide films with internal secondary structure. *Langmuir* **19**: 440-445

189. Klimov, D.K.; Thirumalai, D. 2003. Dissecting the assembly of Abeta16-22 amyloid peptides into antiparallel beta sheets. *Structure* **11**: 295-307

190. Nicholas, K. B., Nicholas, H. B., Jr., Deerfield, D. W., II 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNEW News* **4**:14-15

191. Laskowski, R. A., MacArthur, M. W., Moss D. S., Thornton J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**: 283-291

192. Morris, A. L., MacArthur, M. W., Hutchinson E. G., Thornton J. M. 1992. Stereochemical quality of protein structure coordinates. *Proteins*, **12**: 345-364

193. Connolly, M.L. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**: 709-713

194. Walker, S.M., Leslie, N.R., Perera, N.M., Batty, I.H., Downes, C.P. 2004. The tumour-suppressor function of PTEN requires an N-terminal lipid-binding motif. *Biochemical Journal* **15**: 301-307

195. Miller, S., Janin, J., Lesk, A.M., and Chothia, C. 1987. Interior and surface of monomeric proteins. *Journal of Molecular Biology* **196**: 641–656

196. Gerstein, M. 1992. A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. *Acta Cryst.* **A48**: 271-276

197. Chothia, C.1975. Structural invariants in protein folding. *Nature* **254**:304-308