Fall 2007

Gene synthesis, cloning, expression, purification and biophysical characterization of the C2 domain of human tensin

Kiran Sukumar Gajula
Louisiana Tech University

GENE SYNTHESIS, CLONING, EXPRESSION, PURIFICATION AND

BIOPHYSICAL CHARACTERIZATION OF THE

C2 DOMAIN OF HUMAN TENSIN

by

Kiran Sukumar Gajula, BE

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

COLLEGE OF ENGINEERING AND SCIENCE
LOUISIANA TECH UNIVERSITY

November, 2007

UMI Number: 3283299

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®

UMI Microform 3283299

LOUISIANA TECH UNIVERSITY

THE GRADUATE SCHOOL

08-23-07
Date

We hereby recommend that the dissertation prepared under our supervision

by ____KIRAN SUKUMAR GAJULA_____

entitled_GENE SYNTHESIS, CLONING, EXPRESSION, PURIFICATION AND BIOPHYSICAL_____

_____CHARACTERIZATION OF THE C2 DOMAIN OF HUMAN TENSIN_____

_____

_____

_____

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

_____DOCTOR OF PHILOSOPHY, BIOMEDICAL ENGINEERING_____

_____
Supervisor of Dissertation Research

_____
Head of Department

____BIOMEDICAL ENGINEERING_____
Department

Recommendation concurred in:

_____

_____

_____ Advisory Committee

_____

Approved:

_____
Director of Graduate Studies

Approved:

_____
Dean of the Graduate School
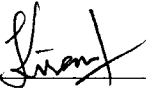
_____
Dean of the College

GS Form 13a
(6/07)

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 _____

Date _____08-29-2007_____

ABSTRACT

Tensin is a large "docking" protein found in the adhesive junctions of animal cells and recruited early in the development of cell-substrate contacts. There it binds to the cytoplasmic domain of integrin β_1 and caps the barbed ends of filamentous actin. This forms a rational basis for its implication in a direct role in the mechanics of membrane-cytoskeleton interactions. Tensin provides a physical link between the actin cytoskeleton, integrins, and other proteins at the cell-substrate contacts. Its overall biochemical properties are a function of its domain composition and architecture, i.e., the domains that are present and their relative positions in the molecule, and specific details of amino acid sequence and post-translational modifications. Tensin can be used as an investigative tool to help explain the physiology of cell-substrate contacts at the molecular scale. The C2 domain indicated to be present at the N-terminus of tensin, along with the phosphatase domain, bears a close homology to PTEN (Phosphatase and Tensin Homolog), a well known tumor suppressor. The main research objective was to study the structural features of the C2 domain of human tensin as well as to investigate its thermal stability.

For this purpose, the C2 domain gene, after being synthesized by Polymerase Chain Reaction, was cloned into the engineered pET-14b vector that would also encode a hexa-Histidine tag. The C2 domain gene was then overexpressed in *E.coli* and the inclusion bodies thus formed were solubilized and lysed. Ni-NTA metal affinity chromatography was performed to obtain the purified C2 domain. Circular dichroism was

used for the initial study of the structural features of the purified C2 domain. After determining that the purified C2 domain was unstable, refolding attempts were done. Later studies of the domain by circular dichroism in the far-UV and near-UV wavelengths indicated that the domain unfolded gradually with increasing amounts of the denaturant GuHCl and also that it retained some amount of tertiary structure prior to denaturation.

The absence of an endothermic peak in the Differential Scanning Calorimetry experiments only suggests the need for further extensive refolding methods not attempted in this work and might also indicate the need for the presence of a phosphatase domain at its N-terminus for thermodynamic stability. The results of this work suggest the presence of a C2 domain in human tensin, which has not been documented previously. Determination of all the structural and functional characteristics of the C2 domain in the long run will contribute to the understanding of the role of tensin in tumor suppression and cell signaling.

DEDICATION

This dissertation is dedicated to that moment on 29th August 2005 at Louisiana Tech, when I came to understand and accept the message of salvation of my Lord and Savior Jesus Christ.

TABLE OF CONTENTS

LIST OF TABLES

ix

LIST OF FIGURES

x

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Donald T. Haynie for helping me to work in this rather challenging and interesting project. I am always thankful to him for his guidance and patience with me through the duration of my PhD course. I am also indebted to Dr. Steven A. Jones for his kind help in advising me in the absence of Dr. Haynie. I also appreciate the time and support of my other committee members Dr. Walter Besio, Dr. Mark A. DeCoster and Dr. William J. Campbell.

I also thank my team members Yogesh Kulkarni, Wanhua Zhao, Naveen Palwai, Sonsy Zachariah, Vinney George and Sunitha Velicharla, for their helpful discussions.

I also thank Dr. John Ward, Jeevan Vemagiri, Regin Koshy, Deeba Balachandran, Prathima Kapa, Anita Alfred and Preethi Kukreja from the International Bible Study group for their prayers and friendship. I also convey my thanks to my friends Kanvasri Jonnalagadda, Kranthi Vistakula, Sivaganesh Mullapudi and Poornachander Yepuru for their support and encouragement.

I also thank Ms. Deborah Wood, Mr. Scott Williams, and Mr. Donald Tatum for their assistance in my experiments at the Biotechnology laboratory in the IFM.

Finally I thank my parents, Daya Rao Gajula and Karuna Gajula; my brothers, Praveen Gajula and Naveen Gajula; my sisters-in-law, Padmaja Gajula and Anne Gajula for their constant prayers, support and encouragement.

CHAPTER 1

INTRODUCTION

1.1. The Human Genome

Cells are the fundamental working units of every living system. The DNA sequence within any living organism spells out the exact instructions required to create that particular organism with its own unique characteristics. The complete set of DNA within an organism is termed "genome", and it varies widely in size from organism to organism [1]. "Genomics involves using reagents, tools and technologies for the high throughput sequencing of DNA and the subsequent storage and annotation of data" [2]. This process focuses on the information of one target molecule, DNA, in the nucleus of cells. Consequently, there is one genome for each organism. The human genome comprises approximately of 3 billion chemical bases which includes 20,000-25,000 protein-coding genes. The U.S. Human Genome Project, begun in 1990, had the goals of identifying all the 20,000-25,000 protein coding genes, determining the sequences of the 3 billion bases that make up the human DNA and hence make them accessible for further biological studies. Though initially planned to last for 15 years, rapid technological advances have accelerated the project to be completed within 13 years in the year 2003 [1].

1

1.2. <u>What is Proteomics?</u>

Proteomics refers to the large-scale study of proteins, particularly their structures and functions. The term 'proteome' of an organism refers to the entirety of proteins in existence in that organism throughout its life cycle [3]. Each of the cells present in an organism contains all the information necessary to make that organism complete. However, not all the genes are expressed in all the cells. Genes that code for enzymes essential to basic cellular functions are expressed in virtually all the cells, whereas those with highly specialized functions are expressed only in specific cell types. Thus, all cells express genes whose protein products provide essential functions and also those genes whose protein products provide unique cell-specific functions. Thus, every organism has one genome, but many proteomes [4]. While the genome is a rather constant entity, the proteome differs from cell to cell and is dynamic in nature through its biochemical interactions with the genome and the environment. As such, an organism has radically different protein expression in different parts of its body, in different stages of its life cycle and in different environmental conditions. Hence proteomics is considered much more complicated than genomics [3]. The biochemical context of genomics and proteomics is depicted in Figure 1.1.

The Human Genome Project also revealed far fewer protein-coding genes in the human genome than there are proteins in the human proteome. This large increase in protein diversity is believed due to alternative splicing and post-translational modification of proteins. The quantity and complexity of the data derived from the sequencing and mapping of the human proteome are estimated to be at least three times greater than that

involved in the Human Genome Project [3]. The field of proteomics is particularly important because most diseases are manifested at the level of protein activity. As such, it seeks to correlate directly the involvement of specific proteins, protein complexes and their modification status in a given disease state. Such knowledge will provide a fast track to commercialization and will speed up the identification of new targets that can be used to diagnose and treat diseases [2].

DNA ➡ *Genome*
"Genomics"

⬇

mRNA

⬇

proteins ➡ *Proteome*
"Proteomics"

⬇

Cell functions
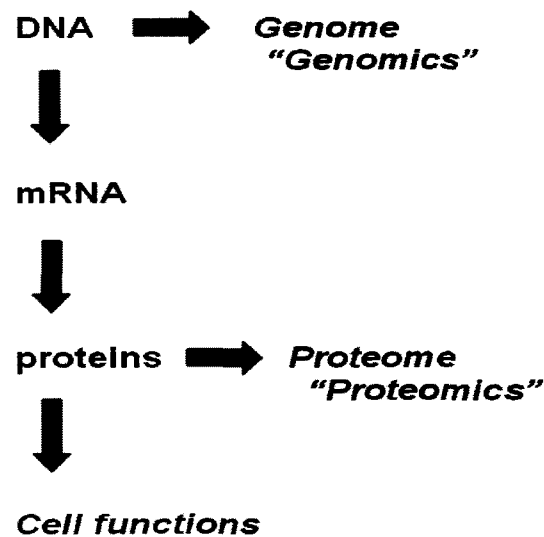
Figure 1.1. Biochemical context of genomics and proteomics [4].

1.3. Tensin in the Context of a Proteome

Tensin is one of the several thousands of proteins that are expressed in animal cells including *H.sapiens*. It is a component of cell-substrate contacts with a close homology to a well established tumor suppressor protein PTEN [38]. Since the exact functions of tensin "modules" or "domains" in humans have not been established, this current research on the individual domains of tensin will attempt to contribute to the understanding of the structure and functionality of the protein and might explain its

predicted role in tumor suppression. Overall the work performed in this project involving various tools of proteomics will contribute to the understanding of the proteome.

1.4. <u>Overview of the Dissertation</u>

Chapter Two discusses the background information underlying the tensin research. It talks about cell adhesion in general and introduces tensin as a key component of focal contacts. The various domains of tensin and their functions are briefly discussed. The motivation for tensin research is highlighted.

Chapter Three discusses the theory and background information relating to the procedures involved in the design and synthesis of the C2 domain gene. All the experimental methods involved in the above procedures including their results and discussions are also mentioned.

Chapter Four discusses the theory and background information relating to the procedures involved in the expression, detection and purification of the C2 domain are briefly discussed. All the experimental methods involved in the above procedures including their results and discussions are also mentioned.

Chapter Five discusses the theory and background information relating to the procedures involved in the refolding and biophysical characterization of the C2 domain. All the experimental methods involved in the above procedures including their results and discussions are also mentioned.

Chapter Six lists the conclusions derived from the results of the experimental work performed on the C2 domain of human tensin. It also discusses the future work.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

2.1.1. Cell Adhesion

Cell adhesion plays an important role in the genesis and maintenance of both three-dimensional structure and normal function in tissues. The functional units of cell adhesion are multiprotein complexes of three general classes of proteins: the cell adhesion molecules/adhesion receptors, the Extra Cellular Matrix (ECM) proteins, and the cytoplasmic plaque/peripheral membrane proteins. Cell adhesion receptors are typically transmembrane glycoproteins that mediate binding to ECM molecules or to counter-receptors on other cells. These molecules also determine the specificity of cell-cell or cell-ECM interaction. The ECM proteins are fibrillar providing a complex structural and functional network that can interact simultaneously with multiple cell surface receptors. The cytoplasmic plaque/peripheral membrane proteins provide structural and functional linkages between adhesion receptors and the actin microfilaments, microtubules, and intermediate filaments of the cytoskeleton [5, 6]. The cell adhesion complexes are not static, but rather dynamic and can capture and integrate signals from the extracellular environment. The functions of these complexes are regulated precisely by biochemical events within the cell. Adhesion-mediated signaling

5

influences the cellular processes of gene expression, cell cycle and programmed cell death [5].

2.1.2. Focal Adhesions

Integrins are a family of transmembrane proteins that link the cytoskeletal elements at the cytoplasmic face with ECM components. Integrins are heterodimers of α and β subunits that contain a large extracellular domain responsible for ligand binding, a single membrane-spanning region and a short cytoplasmic domain [7]. Integrin-mediated cell-ECM adhesion sites are complex specialized structures known as focal contacts or focal adhesions [8]. At focal adhesions, clusters of integrins bind externally to ECM proteins and internally to several cytoplasmic proteins that in turn bind to actin filaments [9, 10] as shown in Figure 2.1.
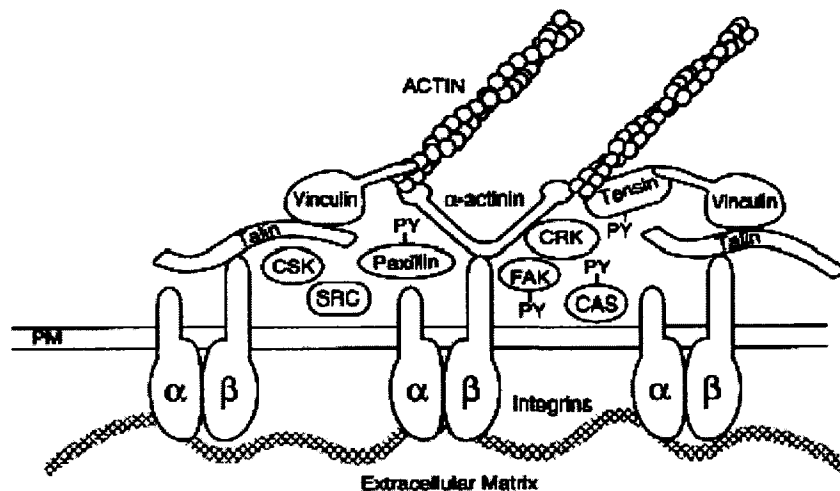
Figure 2.1. Schematic diagram of the molecular architecture of a focal adhesion [10].

Focal adhesions are dynamic and they change in their size and composition as the cell adhesion progresses. Initially, small freckled structures are formed at the sites of cell-to-substratum contact and as the focal contact matures, the actin filaments extend and

bundle to form prominent stress fibers. At focal adhesions, there is a complex network of interactions among the specialized cytoplasmic proteins [5]. The cytoskeletal proteins that directly bind to integrins include talin and α-actinin, which in turn bind to other proteins including vinculin, paxilin and tensin, ultimately leading to the recruitment of actin filaments. Some of the key structural proteins involved in focal contacts are described in Table 2.1.

Table 2.1. Key structural proteins involved in focal contacts.

Protein	Localization	Structure/ Orientation	Binding/ Function
α-Actinin	It is found in Z-discs in skeletal muscles, membrane-associated dense plaques, cytoplasmic dense bodies of smooth muscle, and adherens-type junctions in non-muscle cells [11].	It is a rod-shaped anti-parallel homodimer with a subunit molecular mass of 104 kDa [11].	It binds to vinculin [8], zyxin [12], and the cytoplasmic domains of integrin $\beta1$, $\beta2$, and $\beta3$ subunits [5]. It associates actin molecules to focal adhesions [13].
Talin	It is localized to focal adhesions [14]	It is a homodimer of two 270 kDa polypeptides [15] and is arranged in an anti-parallel orientation.	It binds to actin, vinculin [16], focal adhesion kinase (FAK) [17] and integrin $\beta1$, $\beta3$ subunits [8]. It connects actin cytoskeleton to focal adhesions and is also crucial for the initial formation of focal adhesions [14, 15].

Vinculin	It is localized to focal contacts and cell-cell adherens type junctions [8].	It is a 116 kDa polypeptide with globular head and a rod-like tail region [20]	Its head region binds to talin and α-actinin, while its tail to paxilin and actin [21, 22]. It binds to phosphatidylinositol 4, 5-biphosphate (PIP2), which regulates its focal contact formation and stability [23].
Paxillin	It is localized to the membrane at focal adhesions [24, 25]	It is a 68 KDa protein. The N-terminus has a Src homology region 3 (SH3) domain and multiple SH2 binding sites, while its C-terminus has four LIM domains that are involved in protein-protein interactions [26].	It binds to FAK [27], vinculin [28], p47^{gag-crk} [29], C-terminal Src Kinase (CSK) [26] and pp60^{src} [30]. As such it acts as an adapter protein involved in forming multiprotein complexes.
Filamin	It is localized into the stress fibers and dense bodies in non-muscle cells and Z bands in both skeletal and cardiac muscle cells [32].	It is a homodimer with two polypeptides associating only at the C-terminus, while the N-terminal ends link with actin [33].	It associates with the β2 integrin cytoplasmic domain with is important in spreading and extension of lamellopodia during cell movement and phagocytosis [34].
Tensin	It is localized to various adherens junctions and Z-lines of skeletal muscle [35].	It is a 220 kD protein containing an actin binding region [36], PTP [37], C2 [38], SH2 [39] and PTB [40] domains.	It links cytoskeleton to the plasma membrane, coordinates cell attachment and suppresses cellular transformation [41]. It also plays a role in signal transduction [39].

2.1.3. Tensin

Tensin is located at the transmembrane junctions between the ECM and the cytoskeleton [41-43]. In *H. sapiens,* tensin contains 1735 amino acid residues [44] and is encoded by a gene on chromosome 2 at locus q35-36. Tensin is present in at least two different cell pools, either in the cytoplasm or the focal adhesion complex. The cytoplasmic tensin can suppress cell migration by inactivating downstream regulators, and localization of tensin at focal adhesions is insufficient to promote cell motility [45]. Tensin has three signal transduction-related domains: a protein tyrosine phosphatase (PTP) domain [37], a Src homology 2 (SH2) domain [39], and a phosphotyrosine binding (PTB) domain [40] as shown in Figure 2.2. The PTP, SH2 and PTB domains are well-known to be involved in phosphorylation-mediated interactions in related signaling proteins such as the tumor suppressor PTEN (phosphatase and tensin homolog), Src family kinases, and Shc. A C2 domain (similar to the second conserved domain from protein kinase C) has also been indicated to be present in the protein tensin [38]. Recently, the C2 domains have also been shown to bind phosphorylated peptides increasing the odds of their involvement in signaling pathways [46, 47]. As such it very much seems likely that the participation of tensin in membrane-cytoskeleton interactions will be modulated by signaling involving phosphorylation.
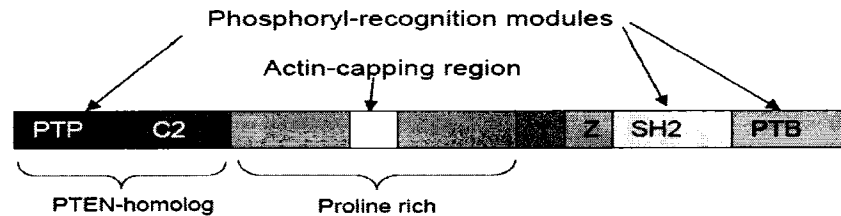
Figure 2.2. Domain architecture of tensin. "Y" and "Z" are identified putative domains (Haynie, D. T, unpublished)

There are three regions in the full-length tensin polypeptide (Haynie, unpublished): N-terminus, central region, and C-terminus (Figure 2.2). The N-terminus consists of the PTP and C2 domains. PTP domain is probably inactive because the essential nucleophilic cysteine is mutated to asparagine [37]. This N-terminal region of tensin is homologous to auxilin [37], cyclin G-associated serine/threonine kinase (GAK) [48], and PTEN [38], also known as multiply-mutated in advanced cancers (MMAC) [49] as shown in Figure 2.3. Torgler *et al.* have shown that the phosphatase domain of tensin in *Drosophila* is required for the localization at muscle attachment sites and Z lines, and also that the deletion of the phosphatase and C2 domains completely inactivates the protein [50]. This region of tensin is also thought responsible for the wing blister phenotype of certain *Drosophila* mutants, which also exhibit a reduced hatching rate of laid eggs [51]. The C2 region of the protein PTEN regulates the tumor suppressor activity of PTEN [53].

Tensin-like proteins
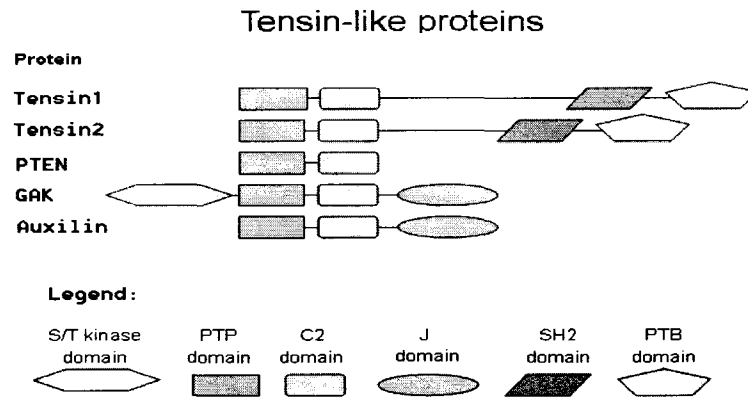
Figure 2.3. Domain architecture of tensin-like proteins (Haynie, D. T, unpublished).

By *in vitro* assays, it was determined that the central region of tensin binds to the barbed ends of filamentous actin (F-actin) (Figure 2.4) and retards the rate of actin polymerization [53]. Tensin can also crosslink actin filaments [36].
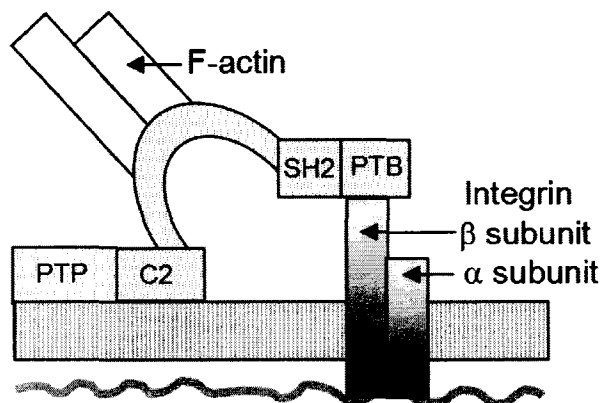
Figure 2.4. Schematic of the hypothetical model of interaction of tensin with other focal adhesion components, including the plasma membrane. Integrin interacts with FN in the ECM and with tensin inside the cell. The F-actin filament bundle will extend into the cytoplasm (Haynie, D. T, unpublished).

The C-terminal region of tensin contains an SH2 and a PTB domain, both of which are phosphotyrosine interaction motifs involved in various signal transduction pathways. Binding of phosphotyrosine by the SH2 domain may be important for tensin-

promoted cell migration [45]. The PTB domain of human tensin binds the cytoplasmic domain of various β integrins [53]. So far, three isoforms have been identified for tensin. They are tensin1, tensin2, and tensin3; all of them are homologous in their N- and C-terminal domains but divergent in their central regions [54, 55].

Tensin plays a role in wound healing and cell migration as suggested by the delay in the skeletal muscle regeneration of tensin knock out mice [56]. A study of tensin (-/-) mice also showed that cellular development was normal for several months, but later multiple cysts were formed in kidney cells that led to kidney degeneration and caused renal failure [57]. Tensin is also a substrate of the focal adhesion protease, calpain II, which is involved in the assembly and disassembly of focal contacts [44]. Tensin is phosphorylated on serine, threonine, and tyrosine residues [41]. Tyrosine phosphorylation of tensin is induced by growth factors such as platelet derived growth factor (PDGF), thrombin, angiotensin, and oncogenes like v-src or brc/abl. [39, 58].

2.1.3.1. PTP (Protein Tyrosine Phosphatase) domain

Tyrosine phosphorylation plays a role in communication between and within cells, maintaining the shape and motility of cells, cellular processes like regulation of gene transcription, mRNA processing, and transport of molecules in or out of cells. It also plays an important role in the coordination of these processes among neighboring cells in embryogenesis, organ development, tissue homeostasis, and the immune system. Abnormalities in tyrosine phosphorylation play a role in the pathogenesis of numerous inherited or acquired human diseases from cancer to immune deficiencies [59]. Reversible protein phosphorylation is a common element of intracellular signal transduction pathways which regulate metabolism, gene expression, cell division and

differentiation, development, transport, locomotion and learning and memory [60]. The structural and functional consequences of incorporating a phosphate group into a protein are diverse, and probably underlie the fundamental role of protein phosphorylation in cellular regulation [61].

Protein Tyrosine Phosphatases (PTPs) are a family of signal transduction enzymes that catalyze the removal of phosphate groups from phosphotyrosine containing proteins. By doing so, they dynamically oppose the action of protein tyrosine kinases (PTKs) which are key to the phosphorylation of proteins [60, 62]. PTPs are classified as receptor-like or intracellular based on cellular localization. Receptor-like PTPs contain a transmembrane domain, an extracellular receptor-like domain, and two intra-cellular PTP domains. The first intracellular domain generally accounts for most of the catalytic activity; the function of the second domain is unknown in most cases. Dual specificity phosphatases (DSPs) constitute a sub-family of intracellular PTPs. While PTPs can dephosphorylate only the tyrosine residues in phosphorylated proteins, DSPs can dephosphorylate serine, threonine as well as tyrosine residues [56, 58].

A key feature of the members of the PTP family is the active site sequence known as the PTP signature motif: $(H/V)C(X)_5R(S/T)$ [56, 58]. 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. The signature motifs of PTP in tensin, PTEN and VHR are shown in Figure 2.5.

```
Signature motif : WPD...H█XXGXXRS
Tensin:            WPD...HNKGNRGRI
PTEN:              FED...H█KAGKGRS
VHR:               AND...H█REGYSRS
```

Figure 2.5. Signature motif of PTP in tensin, PTEN, and VHR. Conserved residues are highlighted in yellow, while the essential nucleophilic cysteine residues are highlighted in green.

The active site of the PTP domain of tensin is different from the usual signature motif of PTPs and DSPs. The essential nucleophilic cysteine is mutated to asparagine. The probable effect of this might be a loss of catalytic activity in human tensin. However, an important aspartic acid, which functions as a general acid or base in catalysis, is conserved in tensin on a flexible loop approximately 30-40 residues towards the N-terminus of the signature motif. The arginine and glycine residues of the signature motif are also conserved in tensin. Glycine is important for the formation of the turn in the P loop. This suggests that the basic shape of the binding pocket will not be substantially different from that in a known PTP and that the PTP domain in tensin will recognize a phosphorylated chemical group.

The principal structural difference between the tyrosine specific PTPs and the dual specific PTP VHR is the depth of the active site pocket. The relatively shallow pocket of VHR can accommodate phosphotyrosine or phosphothreonine; the deeper pocket of PTP1B can only accommodate phosphotyrosine. The PTP domain of tensin is distinguished from other PTPs by its lack of catalytic activity [37]. Also, in the *C.elegans* variant of tensin, the histidine preceding the cysteine is mutated to phenylalanine. Both the above mutations are likely to eliminate all phosphatase activity. The PTP domain in

tensin will presumably protect a phosphorylated chemical group on a binding partner from the catalytic activity of phosphatase.

The phosphatase domain of PTEN, a known tumor suppressor, can also dephosphorylate phosphoinositide substrates invitro besides tyrosine-, serine-, and threonine- phosphorylated peptides. It has an enlarged active site, and it specifically dephosphorylates phosphatidylinositol (3, 4, 5) - triphosphate (PI(3,4,5)P₃), a lipid second messenger produced by the phosphoinositide 3-kinase (PI3K) and activates downstream effectors. These effectors include the Akt/PKB kinase that has potent antiapototic and growth stimulatory effects [38].

2.1.3.2. SH2 (Src Homology 2) domain

SH2 domain is the prototype for protein-protein interaction modules that mediate the formation of multiprotein complexes during signaling [63, 64]. SH2 domains specifically function in protein tyrosine kinase (PTK) pathways, due to the dependence of their binding on tyrosine phosphorylation [65]. SH2 domains represent the largest class of known p-Tyr (phosphotyrosine) recognition domains [64]. SH2 domains are usually found in proteins together with other modules, either catalytic domains like phosphatase domain, or other protein binding modules like SH3, PH, or PTB domains. Hence, depending on the nature of those other modules, proteins containing SH2 domains can participate in a wide variety of intracellular functions [66].

SH2 domains carry out their function by binding with high affinity to phosphotyrosine containing protein targets in a sequence-specific and largely phosphorylation-dependent manner. In addition to the p-Tyr residue, each SH2 domain recognizes several additional flanking residues, usually three to five amino acids C-

terminal to p-Tyr, thereby acquiring selectivity for specific phosphorylated sites [67]. The peptide-binding specificity of large number SH2 domains has been investigated with libraries of peptides phosphorylated on a tyrosine residue and randomized at the +1, +2 and +3 positions C-terminal to the phosphotyrosine residue. Individual SH2 domains bind to unique sequences, while many SH2 domains bind to the sequence pYEEI as shown in Table 2.2.

Table 2.2. SH2 family proteins and their binding partners.

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. Such binding could be a property of the SH2 domain in tensin. Alternatively, the SH2 domain of tensin could possibly interact with a phosphotyrosine residue in the same tensin molecule (intramolecular association as in Src) or in another tensin molecule (intermolecular association as in dimerization).

An SH2 domain consists of ~100 amino acids with two α-helices and seven β-strands, ordered $\beta\alpha\beta\beta\beta\beta\alpha\beta$, with a central anti-parallel β-sheet flanked by α-helices [68]. The folded structure of the SH2 domain is of α+β class. Each SH2 domain-containing protein has distinct sites for recognizing p-Tyr and particular amino acid chains. Recognition also depends on the specific amino acid sequences flanking the phosphotyrosine in a specific target; indeed, the variability in SH2 domain sequences provides a structural basis for target selectivity. In many cases when an SH2 domain is

present, so too is an SH3 domain. The domains are frequently found as repeats in a single protein sequence, which suggests that the functions of the SH2 and SH3 domains are inter-related.

SH2 domain covers the region from 1463 to 1557 in tensin at the C-terminus. SH2 domain of tensin binds to Src, paxillin, p130, and other tyrosine-phosphorylated proteins. The tensin SH2 domain can also interact with certain tyrosine-phosphorylated proteins, notably PI3K and p130^{CAS} [69, 70].

2.1.3.3. PTB (Phospho Tyrosine Binding) domain

The PTB domain plays an important role in protein-protein interactions and signal transduction [71-73]. The PTB domain was first identified in the signaling proteins Shc (Src homolog-collagen homolog) and IRS-1 (Insulin receptor substrate-1) as an alternative to the SH2 domain for phosphotyrosine recognition [72, 74, 75]. In contrast to SH2 domains which bind exclusively to peptides containing phosphotyrosine, PTB domains bind to phosphotyrosine sites on activated tyrosine kinases and are involved in diverse cellular functions. Shc binds to pTyr sites on growth-factor activated tyrosine kinases and oncogene products. The PTB domains of Shc and IRS-1 bind to other proteins or ligands containing NPXpY motif, where 'pY' is phosphorylated tyrosine and 'X' any amino acid [76].

The PTB domain of tensin seems to prefer non-phosphorylated substrate. Tensin is co-localized with integrins in the absence of extracellular ligand binding, which suggests that it associates with integrins before it bonds to ligands. Moreover, tensin is phosphorylated in response to the integrin adhesion, suggesting that it is a part of the down stream-signaling pathway. The PTB domain of tensin binds strongly to the

cytoplasmic tails of integrin $\beta3$, $\beta5$ and $\beta7$ but binds weakly to $\beta1$A, which implicates the tendency of the domain to bind in a phosphorylation independent manner. The potential binding partners of the PTB domain are PI3K, p130^{CAS} [69].

In other proteins, the PTB domains have been found to participate in phosphotyrosine-independent interactions. For example, members of the X11 family of proteins contain a PTB domain that binds peptides in a phosphotyrosine-independent fashion (Table 2.3).

Table 2.3. PTB family proteins and their ligand-binding partners. Residues in bold italics are known to be involved in binding or are conserved among PTB domain binding targets [77].

Proteins with PTB domain	Binding Partner	Amino acid sequence of the peptide binding site
Shc	EGFR	SLD*NP*D*Y*QQDF
IRS-1	IR	*LYA*ASS*NPEYL*AS
NUMB	LNX	GLD*NP*A*YT*SSV
	NAX	GFS*NM*S*F*EDFP
	Peptide Screen	GFS*NM*S*F*EDFP
X11	APP	G*YENPT YK*F*F*E

Sequence analysis and experimental studies on integrin and tensin show that the interaction of integrin with tensin takes place through the PTB domain of tensin. The PTB domain of tensin thus plays a key role in the function of 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.

2.1.4. C2 Domain

C2 domains were first identified in classical protein kinases C (PKCs) [78] as calcium-dependent lipid binding modules. They are intracellular and are present in

numerous proteins with functions ranging from signal transduction to vesicular trafficking [79]. C2 domains are categorized into calcium-dependent or calcium-independent forms by the way they interact with phospholipids. In PLCδ1, PKC, and cPLA$_2$, the C2-domains are of type II topology, and play a regulatory role by mediating the enzymes to bind to phospholipids in a calcium dependent manner. On the other hand, C2 domains in synaptotagmin I, rabphilin-3, RIM, and Munc13, are of type I topology, and bind Ca^{+2} through three loops implicated in mediating association with the plasma membrane. Proteins exhibiting the type II topology are involved in the generation of lipid second messengers (e.g. cPLA2, PLCs, and PI3K), protein phosphorylation, activation of GTPases (e.g. Ras-GAP), and ubiquitin ligation (e.g. Nedd4) [79].

The C2 domain of PTEN folds into a β-sandwich structure consisting of two antiparallel β sheets with two short α helices intervening between the strands (as in Figure 2. 6). This β sandwich is identical to type II topology of the C2 fold as in PLCδ1, PKCδ, and cPLA$_2$. The C2 domain of PTEN also lacks canonical Ca^{+2} ligands and is therefore unlikely to bind Ca^{+2}.
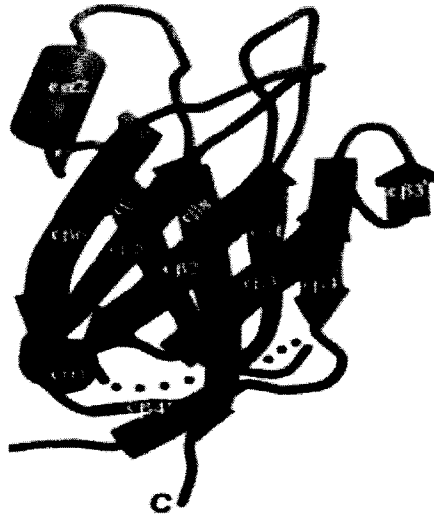
Figure 2.6. Ribbon diagram of the structure of the C2 domain in PTEN [38].

PTEN has a calcium-binding region 3 (CBR3) loop that has the same structure as in the Ca^{2+}-dependent C2 domains [38]. This loop may be important for membrane binding. The CBR3 loop of PTEN has a positive charge at neutral pH, resulting from Lys-260, Lys-263, Lys-266, Lys-267, Lys-269, and the partially buried His-259-Asp-268 pair. This situation resembles the positive charge on the Ca^{2+}-dependent C2 domains that results from the binding of Ca^{2+} to the CBR3 loops. The positive charge on the C2 domains has been implicated in their affinity for phosphoryl groups of phospholipids.

The PTEN C2 domain also contains two solvent-exposed hydrophobic residues, Met-264 and Leu-265, both at the tip of the CBR3 loop. About half of the known C2 domains contain at least one hydrophobic residue at this position. Experimental results suggest that these hydrophobic residues may enable synaptotagmin to insert into a lipid bilayer. The PTEN C2 domain has an additional patch of basic residues on the adjacent Cα2 helix, resulting from the solvent-exposed Lys-327, Lys-330, Lys-332, and Arg-335. This is a position similar to helix of $cPLA_2$ that has been implicated in membrane

binding. The PTEN C2 domain may not only help recruit PTEN to the membrane, but it may also serve to position and orient the catalytic domain optimally with respect to the membrane-bound substrate.

A molecular model of PTP-C2 domain interface in PTEN is shown in Figure 2.7. The conserved Arg and the two Tyr residues of the PSXXRY(L/V)XY motif shown in ball-and-stick representation, point into the domain interface. Also, the conserved Asp residue in the (L/V)XGD(I/V)X(L/V) motif is immediately in front of the left most ball-and-stick Tyr within the domain interface. Since these hydrophobic residues pack together within the domain interface, additional stability is induced in the C2 domain of PTEN.

Figure 2.7. PTP-C2 domain interface in PTEN. Part of the PTP domain is at the top of the figure, part of the C2 domain is at the bottom. Secondary structures are shown as ribbons, α helices in red, β strands in cyan, and turns in green. (Haynie, D. T, unpublished)

The homology of the PTEN phosphatase domain has with tensin and auxilin, both

of which lack key phosphatase active site residues, maps primarily to the hydrophobic

core and to residues on the surface of the pα6 helix as mentioned earlier. Since these

helix residues pack onto the C2 domain, tensin, auxilin, and GAK may also contain a C2

domain [38]. The sequence alignment of various C2-like regions in different proteins is

shown in Figure 2.7. 1D5RA and NP_072174 are the C2 domain sequences of human

PTEN and tensin respectively. Though the sequence heterogeneity is generally high,

relatedness is discernable.
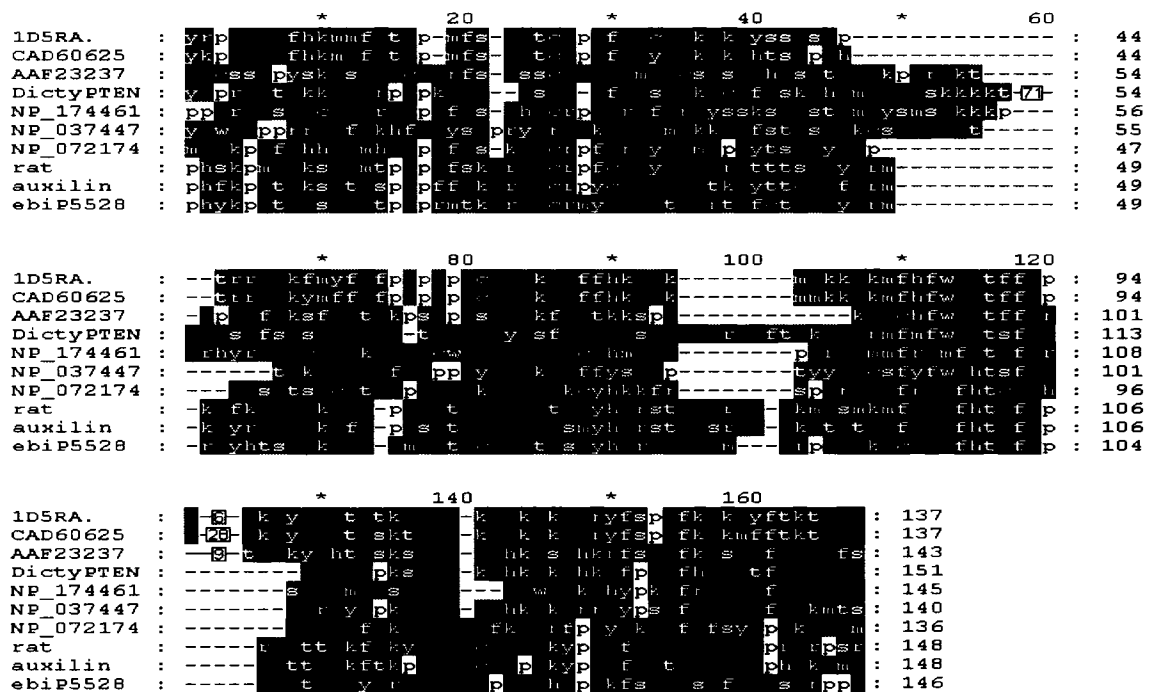
Figure 2.8. Sequence alignments of various C2-like regions. 1D5RA=human PTEN, CAD60625=zebra fish PTEN-like protein, AAF23237=fruit fly PTEN, DictyPTEN= slime mold PTEN, NP_174461=thale cress (plant) actin binding protein, NP_037447= human transmembrane PTEN isoform alpha, NP_072174=human tensin, rat= rat GAK (Cyclin G-associated serine/threonine kinase), auxilin= cow auxilin and ebiP5528=thale cress (plant) auxilin-like protein. (Haynie, D. T, unpublished)

Figure 2.9 shows the ribbon diagram of the C2 domain of PTEN and tensin colored according to the secondary structure. Overall, the structures are quite similar.
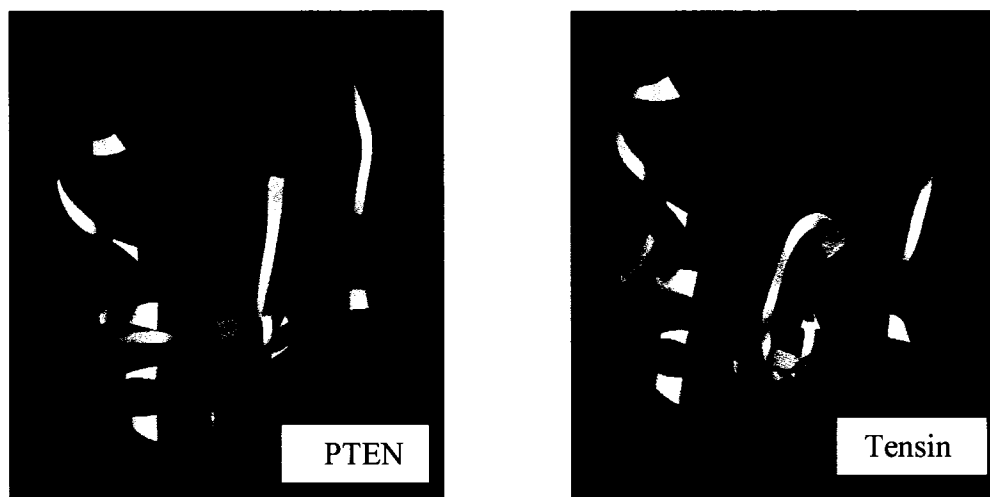
Figure 2.9. Ribbon diagrams of the C2 domain of PTEN and tensin. Red: α-helix; blue: β sheets; green: turns and white: irregular structure. (Haynie, D. T, unpublished)

A comparison of structural features is made in Table 2.4. The backbone RMSD is as low as 0.5 Å. Ratio of SAS/TMS is similar in both cases, *c.* 40%. The number of hydrogen bonds per amino acid is close to 1. All the analyses are close to the average protein structure characterization data.

Table 2.4. Structural comparison of C2 domain of PTEN and tensin models (Haynie, D. T, unpublished).

	PTEN	TENSIN
H-bonds/amino acid	0.98	0.82
Solvent accessible surface (SAS)	8,804 Å^2	10,052 Å^2
Total molecular surface (TMS)	24, 153 Å^2	25, 067 Å^2
SAS/TMS	36.4%	40.1%
RMSD (backbone)	~0.5 Å	

The computational modeling results as shown in Table 2.4, strongly lead to the hypothesis that a C2 domain similar to that of a PTEN is present in tensin.

2.2. Research Objective

To prove the hypothesis that a C2 domain similar to that of PTEN is present in tensin, it is important to determine the structural features and thermodynamic stability of the C2 domain of tensin. Hence the main research objective of the current project was to determine the structural features of the C2 domain of human tensin as well as to investigate its thermal stability.

CHAPTER 3

DESIGN, SYNTHESIS AND CLONING
OF THE C2 DOMAIN GENE

3.1. Design and Synthesis of the C2 Domain Gene

3.1.1. Introduction

This section consists of various subtopics covering the theory and background relating to the procedures involved in the design and synthesis of the C2 domain gene.

3.1.1.1. The genetic code

Genetic information is specified by the order of nitrogenous bases *viz* adenine (A), cytosine (C), guanine (G) and thymine (T) that make up the deoxyribonucleic acid (DNA) molecule. A DNA molecule consists of two long polynucleotide chains composed of four types of nucleotide subunits. These nucleotides are composed of five-carbon deoxyribose to which is attached one or more phosphate groups and a nitrogenous base (A, G, C or T). The nucleotides are covalently linked together in a chain through the sugars and phosphates, which forms a sugar-phosphate backbone. The three-dimensional structure of DNA is the double helix, which is formed by the chemical and structural features of its two polynucleotide chains [80]. Because these two chains are held together by hydrogen bonding between the bases on the different strands, all the bases are on the inside of the double helix, and the sugar-phosphate backbones are on the outside as shown in Figure 3.1A. A always pairs with T, and G with C as shown in Figure 3.1B.

25

RNA is also a linear polymer like DNA except that it has the base thymine (T) substituted by uracil (U) and also deoxyribose by ribose [81].

(A)   (B)
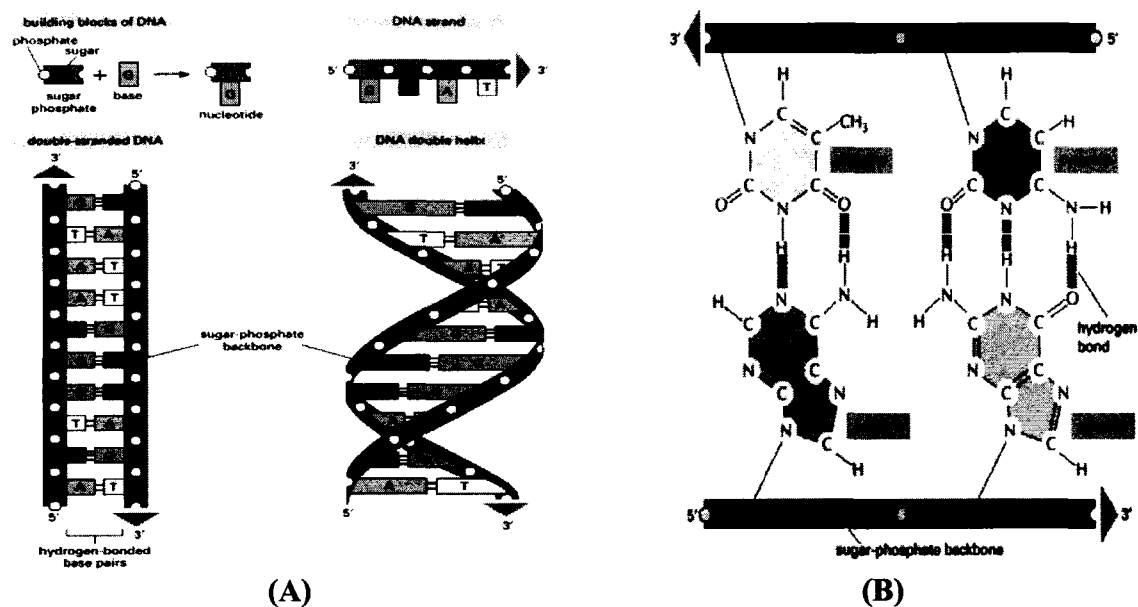
Figure 3.1. Structure of DNA. **(A)** DNA and its building blocks. **(B)** Complementary base-pairing in the DNA double-helix [81].

In a living cell, during the process of transcription, the genetic information contained within its double-stranded DNA is transferred onto the messenger RNA (mRNA). While in translation, the mRNA is decoded to produce a specific polypeptide at the ribosome. The rules for the synthesis of a polypeptide chain are specified by the genetic code, which defines a mapping between tri-nucleotide sequences called 'codons' and amino acids. Each codon specifies a single amino acid [82]. Most organisms use a universal codon usage referred to as standard genetic code such as the one for *Escherichia coli (E.coli) K12* strain as shown in Appendix C.

3.1.1.2. Codon bias

The frequencies with which different codons appear in *E.coli* genes are different from those in human. The frequency of any codon reflects the amount of tRNA specific to that codon. As such any tRNA that recognizes a rarely used codon is present in low amounts. Therefore, human genes that contain rare codons in *E.coli* may be inefficiently expressed. Hence, this problem can be solved either by exchanging codons in the target gene for codons more frequently used in *E.coli*, or, alternately, by co-production of the rare tRNAs [83].

3.1.1.3. Polymerase chain reaction

Polymerase chain reaction (PCR), developed by Kary Mullis in 1983, is a molecular biology technique which allows a small amount of DNA molecule to be amplified many times, in an exponential manner. A PCR requires several basic components such as DNA template or cDNA, which contains the region of the DNA fragment to be amplified; two primers that determine the beginning and end of the region to be amplified; DNA polymerase such as *Taq* or *Pfu*, which copies the region to be amplified; deoxynucleotide triphosphates (dNTPs), from which the DNA polymerase builds the new DNA and a buffer which provides a suitable chemical environment for the DNA polymerase [84].

A PCR process usually consists of a series of twenty to thirty-five cycles. Each of these cycles consists of the following three steps. The first step is called *denaturing*, where the double-stranded DNA (dsDNA) template is heated to 94-96° C to break the hydrogen bonds connecting the two DNA strands. In the second step of *annealing*, the temperature of the reaction mixture is lowered to around 55°C at which the primers are

hybridized onto the template. The final step called *extension,* involves raising the temperature of the reaction mixture to 72°C at which the DNA polymerase starts at each of the annealed primers and copies the DNA strands by extending the primers. A schematic of the PCR process is as shown in Figure 3.2. At the end of first PCR cycle, there would be two copies of the template DNA as shown. As such at the end of 30 cycles, 2^{30} copies of the template DNA would be present.
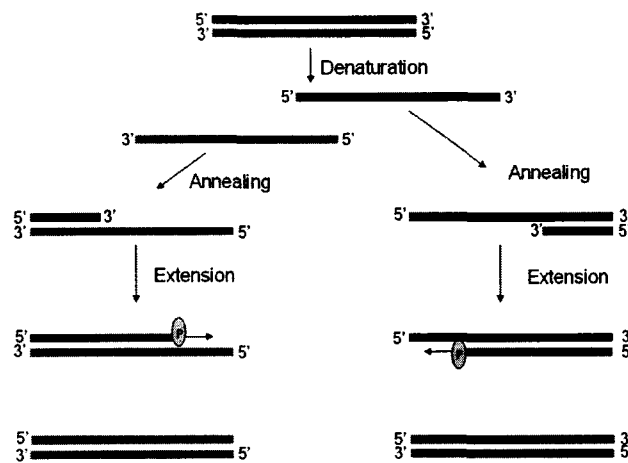
Figure 3.2. Schematic of the PCR process. All the three steps involved in the first cycle of PCR are shown here. 'P' indicates DNA polymerase.

In this current project, PCR has been employed in multi-stage to synthesize the C2 domain gene as described in Section 3.1.2.1.

3.1.1.4. DNA electrophoresis

It is an analytical technique used to separate DNA fragments based on their size. This technique employs agarose, a polysaccharide polymer material extracted from seaweed. Because agarose contains large number of hydroxyl groups, it readily forms low-melting point gels with aqueous buffers and creates an immobile meshwork of microscopic fibers. The DNA fragments to be separated are loaded onto different wells in

an agarose gel and are made to migrate through it by the application of an electric field. Due to the net negative charge of the phosphate backbone of a DNA chain, DNA molecules are normally made to migrate from negative to positive potential. Smaller DNA fragments migrate faster through the meshwork, whereas the larger ones migrate slower. To visualize the DNA, ethidium bromide, a fluorescent dye that intercalates the double stranded DNA during the migration is generally used. The sizes of unknown DNA bands are determined by comparing them to a set of DNA markers [85].

3.1.2. Methods, Results and Discussion

This section consists of various subtopics covering the experimental methods involved in the design and synthesis of the C2 domain gene including their results and discussion.

3.1.2.1. Design of the C2 domain gene

As the first basic step, the amino acid sequence of human tensin under accession number NP_072174 was acquired from the protein database of National Center for Biotechnology Information (NCBI). Boundaries for each of the domains of tensin have been identified by their sequence homology alignment with the respective domains in other known proteins. Once the domain boundaries were identified for the C2 domain, the genes that would encode for its amino acids were built. Also, this was done by picking up only the high frequency codons from the codon usage table of *Escherichia coli K12* strain as shown in Appendix C.

After the gene for the C2 domain has been built, restriction enzyme sites *viz Nde I* and *Xho I* were incorporated at its 5' and 3' ends respectively. A stop codon was also added in the gene design just preceding the *Xho I* site to terminate the amino acid chain

during translation. At the 5'-end of the C2 domain gene, a *Bcl I* site was incorporated a

few bases following the *Nde I* site so that it could be linked to the C-terminal region of

the PTP gene. Care was also taken to include sufficient bases on either side of the

restriction enzyme sites for efficient cleavage. The C2 domain gene designed after

incorporating the above features is shown in Figure 3.3.

Nde I site
↓

```
AGGGAATTC CATATG TCGGGCTCCATTAAAATGAATAATAAACCACTGTTT
  R   E   F   H   M
TCCCTTAAG GTATAC AGCCCGAGGTAATTTTACTTATTATTTGGTGACAAA
```

Afl II site
↓

```
CTGCATCACG TGATCA TGCATGGTATTCCGAATTTTGAAAGCAAAGGTGGC

GACGTAGTGC ACTAGT ACGTACCATAAGGCTTAAAACTTTCGTTTCCACCG
```

```
TGTCGTCCTTTTTTACGTATCTATCAGGCGATGCAGCCTGTATACACCAGC
  C   R   P   F   L   R   I   Y   Q   A   M   Q   P   V   Y   T   S
ACAGCAGGAAAAAATGCATAGATAGTCCGCTACGTCGGACATATGTGGTCG
```

```
GGCATTTATAACATTCCGGGCGATTCCCAGACGAGCGTATGCATAACCATC
  G   I   Y   N   I   P   G   D   S   Q   T   S   V   C   I   T   I
CCGTAAATATTGTAAGGCCCGCTAAGGGTCTGCTCGCATACGTATTGGTAG
```

```
GAACCGGGCCTCCTTCTGAAAGGGGACATTCTGCTTAAATGCTACCATAAA
  E   P   G   L   L   L   K   G   D   I   L   L   K   C   Y   H   K
CTTGGCCCGGAGGAAGACTTTCCCCTGTAAGACGAATTTACGATGGTATTT
```

```
AAATTCCGCAGCCCGGCTCGCGACGTCATCTTCCGCGTGCAATTTCATACG
  K   F   R   S   P   A   R   D   V   I   F   R   V   Q   F   H   T
TTTAAGGCGTCGGGCCGAGCGCTGCAGTAGAAGGCGCACGTTAAAGTATGC
```

```
TGCGCGATTCATGATCTGGGCGTAGTGTTCGGTAAGGAAGACCTAGATGAT
  C   A   I   H   D   L   G   V   V   F   G   K   E   D   L   D   D
ACGCGCTAAGTACTAGACCCGCATCACAAGCCATTCCTTCTGGATCTACTA
```

GCCTTCAAAGATGATCGCTTTCCAGAGTATGGGAAAGTAGAATTCGTCTTT

A   F   K   D   D   R   F   P   E   Y   G   K   V   E   F   V   F

CGGAAGTTTCTACTAGCGAAAGGTCTCATACCCTTTCATCTTAAGCAGAAA

AGTTACGGCCCTGAGAAAATCCAAGGCATGGAACATCTGGAAAACGGCCCG

S   Y   G   P   E   K   I   Q   G   M   E   H   L   E   N   G

TCAATGCCGGGACTCTTTTAGGTTCCGTACCTTGTAGACCTTTTGCCGGGC

TCTGTAAGCGTGGACTATAATACTAGTGATCCGCTTATTCGTTGGGATTCC

S   V   S   V   D   Y   N   T   S   D   P   L   I   R   W   D   S

AGACATTCGCACCTGATATTATGATCACTAGGCGAATAAGCAACCCTAAGG

TACGATAACTTCTCCGGCCACCGTGATGACGGCATGGAAGAAGTAGTTGGT

Y   D   N   F   S   G   H   R   D   D   G   M   E   E   V   V   G

ATGCTATTGAAGAGGCCGGTGGCACTACTGCCGTACCTTCTTCATCAACCA

Xho I site

↓

TAA|CTCGAG|CGG

*   L   E   R

ATT|GAGCTC|GCC

Figure 3.3. Design of the C2 domain gene. The highlighted text in bold indicates the amino acid sequence that would be encoded by the C2 domain gene. The sequences of the corresponding coding and non-coding strands are indicated just above and below it respectively. '*' indicates a stop codon.

3.1.2.2. Synthesis of the C2 domain gene by PCR

Ten oligonucleotides, five each from the coding and the non-coding strands (sets of bases that are underlined in Figure 3.3) were chosen from the gene design so that these oligonucleotides would serve as templates for the synthesis of full length double-stranded C2 domain gene. The schematic of the PCR strategy used is as shown in Figure 3.4.

PCR-I   PCR-II   PCR-III

US1 (74)   US2 (79)   US3 (76)   US4 (76)   US5 (78)

LS1 (74)   LS2 (71)   LS3 (75)   LS4 (80)   LS5 (75)

PCR-IV

DS2 (246)

DS3 (132)

PCR-V

DS1 (235)

DS4 (358)

C2 (573)
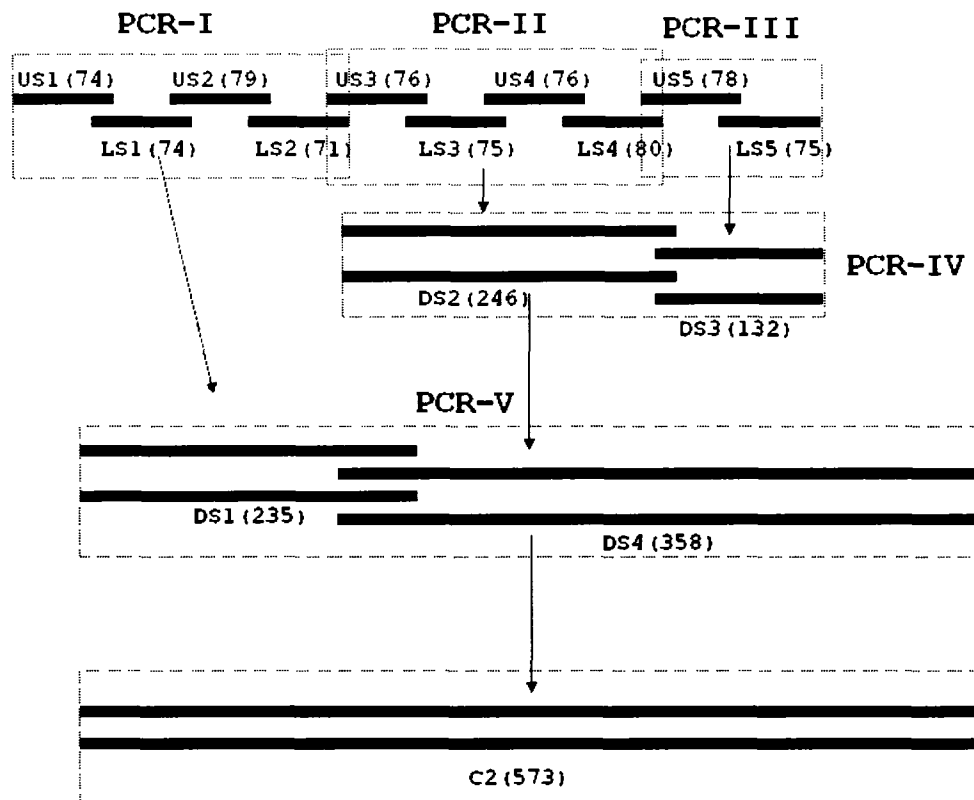
Figure 3.4. Schematic of the PCR strategy to synthesize the C2 domain gene.

The five oligonucleotides from the coding strand are indicated US1, US2, US3, US4 and US5, while the five from non-coding strand are LS1, LS2, LS3, LS4 and LS5. The sequences of each of these oligonucleotides from 5' to 3' ends are shown below with their lengths indicated in parentheses.

US1- AGGGAATTCCATATGTCGGGCTCCATTAAAATGAATAATAAACCACTGT
TTCTGCATCACGTGATCATGCATGG (74)

US2- GTCCTTTTTTACGTATCTATCAGGCGATGCAGCCTGTATACACCAGCGG
CATTTATAACATTCCGGGCGATTCCCAGAC (79)

US3- CCTTCTGAAAGGGGACATTCTGCTTAAATGCTACCATAAAAAATTCCGC
AGCCCGGCTCGCGACGTCATCTTCCGC (76)

US4- CGTAGTGTTCGGTAAGGAAGACCTAGATGATGCCTTCAAAGATGATCGC
TTTCCAGAGTATGGGAAAGTAGAATTC (76)

US5- CATCTGGAAAACGGCCCGTCTGTAAGCGTGGACTATAATACTAGTGATC
CGCTTATTCGTTGGGATTCCTACGATAAC (78)

LS1- GATAGATACGTAAAAAAGGACGACAGCCACCTTTGCTTTCAAAATTCGG
AATACCATGCATGATCACGTGATGC (74)

LS2- GAATGTCCCCTTTCAGAAGGAGGCCCGGTTCGATGGTTATGCATACGCT
CGTCTGGGAATCGCCCGGAATG (71)

LS3- CTTCCTTACCGAACACTACGCCCAGATCATGAATCGCGCACGTATGAAA
TTGCACGCGGAAGATGACGTCGCGAG (75)

LS4- GACGGGCCGTTTTCCAGATGTTCCATGCCTTGGATTTTCTCAGGGCCGTA
ACTAAAGACGAATTCTACTTTCCCATACTC (80)

LS5-CCGCTCGAGTTAACCAACTACTTCTTCCATGCCGTCATCACGGTGGCCGG
AGAAGTTATCGTAGGAATCCCAACG (75)

The strategy to synthesize the full length C2 gene involved several steps. Initially the ten oligonucleotides above were grouped into 3 sets of four, four and two respectively. The next step involved performing an overlap-extension PCR using the four oligonucleotides US1, LS1, US2 and LS2 to yield the double-stranded DNA fragment DS1 as shown in Figure 3.4. Similarly, performing an overlap-extension PCR of oligos US3, LS3, US4 and LS4 would yield the double-stranded DNA fragment DS2, while the overlap-extension PCR of oligos US5 and LS5 would yield the fragment DS3. To amplify each of the fragments DS1, DS2 and DS3, two different end primers were designed, the sequences of which are as shown below:

P1-AGGGAATTCCATATGTCGG (19)

P2-GAATGTCCCCTTTCAGAA (18)

P3-CCTTCTGAAAGGGGACATTCTGCT (24)

P4-GACGGGCCGTTTTCCAGATGTTCCA (25)

P5-CATCTGGAAAACGGCCC (17)

P6-CCGCTCGAGTTAACCAACTACTTCT (25)

The lengths of each of the starting oligos were kept in the range of 70-80 bases. The overlapping regions of the successive oligos were also kept to be around 20 bases in length. These regions would start and end with GC pairs for better annealing.

In the next step, overlap-extension PCR on DS2 and DS3 would yield another double-stranded DNA fragment DS4. In the final step, overlap-extension PCR on the fragments DS1 and DS4 would yield the entire C2 domain gene.

All the oligonucleotides and primers were received from a commercial source in lyophilized form. The optical density of the oligonucleotides was 15 at 260 nm, while that of the primers was 10 at 260 nm. Stock solutions were prepared by initially adding 500µl of DI water to each of the oligonucleotide vials. Amounts of 90 µl, 100 µl, 76 µl, 75 µl, 105 µl and 76 µl of DI water were added to the primer vials P1, P2, P3, P4, P5 and P6 respectively. The resulting oligonucleotide calculations and their concentrations after initial dilution are as shown in Table 3.1. The formulae used are shown in Appendix D.

Table 3.1. Oligonucleotide calculations and their concentrations after initial dilution.

Oligo/Primer	A	G	C	T	E_{260} $(M^{-1}cm^{-1})$	Mol.wt (Da)	pmol	µg	Stock Conc (µg/µl)
US1	23	15	15	21	737267.1	22805.88	20345.4	463.99	0.92798
LS1	26	15	16	17	753625.3	22818.06	19903.79	454.16	0.90832
US2	18	16	22	23	740328.7	24201.93	20261.27	490.36	0.98072
LS2	14	22	18	17	674628.9	21942.37	22234.44	487.87	0.97574
US3	19	15	24	18	717322.2	23243.24	20911.10	486.00	0.97200
LS3	21	18	22	14	731704.6	23062.09	20500.70	472.77	0.94554
US4	23	21	12	20	772591.2	24497.14	19415.18	458.38	0.91676
LS4	17	16	22	25	742304.5	24025.84	20207.34	495.02	0.99000
US5	20	18	18	22	754809.0	24025.84	19872.57	477.45	0.95000
LS5	18	17	23	17	710068.7	22995.06	21124.71	485.76	0.97152

P1	5	6	3	5	190664.7	5867.85	52448.09	307.75	3.41900
P2	5	3	5	5	172330.7	5468.63	58027.96	317.33	3.17330
P3	7	6	6	7	226015.5	7343.86	44244.70	324.92	4.27000
P4	7	7	7	7	229237.3	7649.06	43622.91	333.6	4.44000
P5	2	4	6	2	165700.2	5164.39	60349.95	311.67	2.96000
P6	7	3	9	7	227991.3	7536.9	43861.32	330.57	4.34960

To 10 µl of each of US1, LS1, US2, LS2, US3, LS3, US4, LS4, US5 and LS5,

82.8 µl, 80.8 µl, 88 µl, 87.6 µl, 87.2 µl, 84.5 µl, 81.7 µl, 89 µl, 85 µl and 87.15 µl of DI

water was added respectively to get a final oligo concentration of 100 ng/µl. Also to 3 µl

of each of P1, P2, P3, P4, P5 and P6, 99.5 µl, 92.2 µl, 125.1 µl, 130.2 µl, 85.8 µl and

127.5 µl of DI water was added respectively to get a final primer concentration of 100

ng/µl.

Initially, to synthesize the double stranded piece DS1, the following components

were taken into a PCR tube:

Oligo-US1---------------------1.0 µl (100 ng/µl)

Oligo-LS1---------------------1.0 µl (100 ng/µl)

Oligo-US2--------------------1.0 µl (100 ng/µl)

Oligo-LS2--------------------1.0 µl (100 ng/µl)

Pfu polymerase----------------1.0 µl (2.5 U/µl)

10x *Pfu* buffer-----------------5.0 µl

dNTP mix (25mM each) -----0.8 µl

DI water-----------------------39.2 µl

The components were mixed well using a pipette, after which the PCR tube was

placed in one of the sample wells of the thermal cycler. Then its components were

subjected to the following thermal cycles:

```
94°C------------------1 min  ⌐
                             |
25°C------------------10 min ⟩─ 1 cycle
                             |
65°C------------------15 min ⌐

94°C------------------1 min  ⌐
                             |
55°C------------------1 min  ⟩─ 1 cycle
                             |
72°C------------------1 min  ⌐
```
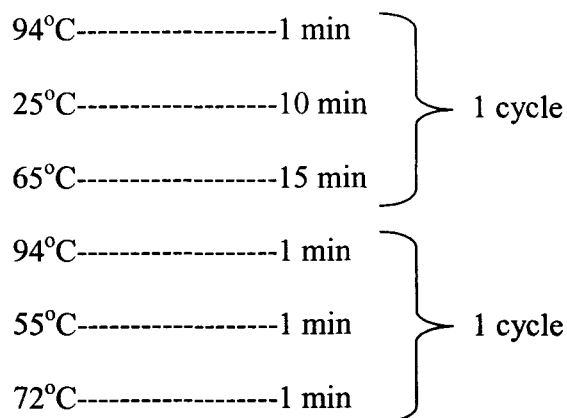
Thereafter, 1.25 μl of each of the primers P1 and P2 at 100 ng/μl concentration were added to the above PCR tube, mixed well and was further subjected to the following thermal cycles:

```
94°C------------------1 min  ⊐─1 cycle

94°C------------------1 min  ⌐
                             |
55°C------------------1 min  ⟩─35 cycles
                             |
72°C------------------1 min  ⌐

72°C------------------5 min  ⌐
                             ⟩─1 cycle
10°C------------------Hold   ⌐
```
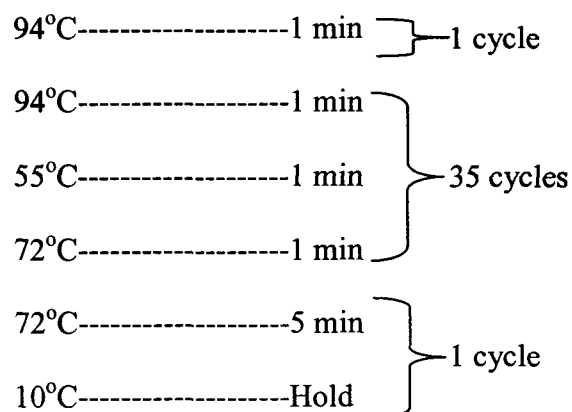
After all the cycles were completed, the resulting sample was taken out, mixed with 10 μl of 6x loading dye and then loaded onto two lanes of a 1.5% (w/v) agarose gel that was prepared as per the protocol B.1 in Appendix B. The gel was run at 60 V, 400 mA for 1 hour. After that, the gel was visualized on a UV transilluminator and was photographed. The gel image thus obtained is as shown in Figure 3.5. The gel clearly showed DNA bands at around 235 bp size, indicating the synthesis of the double-stranded DS1.

Figure 3.5. Agarose gel image showing the synthesis of the double-stranded DNA fragment, DS1.

Later the bands from the agarose gel were excised and the DNA within was extracted by gel extraction procedure as per the protocol B.1 described in Appendix B.

Next to synthesize the double-stranded DS2, the following components were taken into two PCR tubes separately.

Oligo-US3-------------------1.0 μl (100 ng/μl)

Oligo-LS3-------------------1.0 μl (100 ng/μl)

Oligo-US4-------------------1.0 μl (100 ng/μl)

Oligo-LS4-------------------1.0 μl (100 ng/μl)

Pfu polymerase---------------1.0 μl (2.5 U/μl)

10x *Pfu* buffer----------------5.0 μl

dNTP mix (25mM each) -----0.8 μl

DI water----------------------39.2 μl

After being mixed well, the components in the PCR tubes were subjected to the following thermal cycles:

```
94°C------------------1 min  ⎤
                            ⎬ 1 cycle
25°C------------------10 min ⎮
                            ⎮
65°C------------------15 min ⎦

94°C------------------1 min  ⎤
                            ⎮
55°C------------------1 min  ⎬ 1 cycle
                            ⎮
72°C------------------1 min  ⎦
```
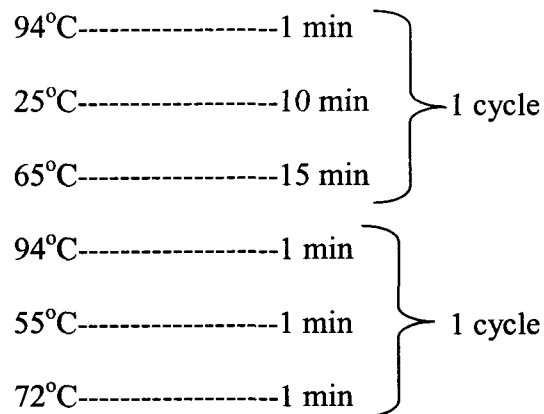
Thereafter, 1.25 µl of each of the primers P3 and P4 at 100 ng/µl concentration were added to the above PCR tubes and were further subjected to the following thermal cycles:

```
94°C------------------1 min  ⎤⎯1 cycle

94°C------------------1 min  ⎤
                            ⎮
55°C------------------1 min  ⎬ 35 cycles
                            ⎮
72°C------------------1 min  ⎦

72°C------------------5 min  ⎤
                            ⎬ 1 cycle
10°C------------------Hold   ⎦
```
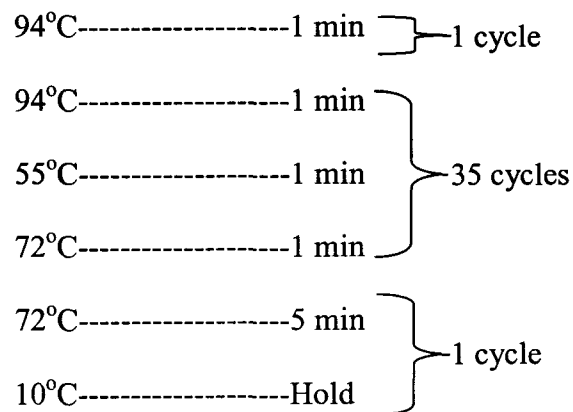
After all the cycles were completed, the resulting tubes were taken out and the components loaded onto a 1.5% (w/v) agarose gel. The gel after it was run was visualized on a UV transilluminator and is as shown in Figure 3.6. The gel clearly showed DNA bands at around 246 bp size, indicating the synthesis of the double-stranded DS2. As before, the bands were excised and the DNA within was extracted and stored.
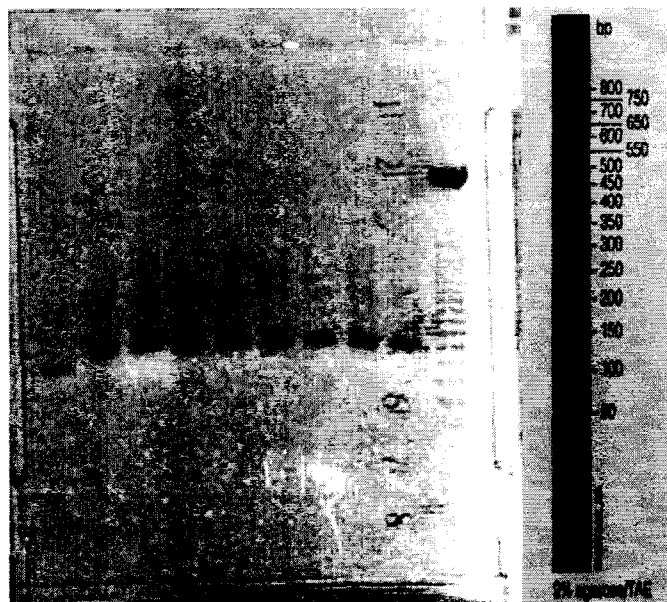
Figure 3.6. Agarose gel picture showing the synthesis of the double-stranded DNA fragment, DS2.

Next, to synthesize the double stranded fragment DS3, the following components were taken into two PCR tubes separately:

Oligo-US5--------------------1.0 µl (100 ng/µl)

Oligo-LS5--------------------1.0 µl (100 ng/µl)

Pfu polymerase---------------1.0 µl (2.5 U/µl)

10x *Pfu* buffer----------------5.0 µl

dNTP mix (25mM each) -----0.8 µl

DI water-----------------------41.2 µl

After being mixed well, the components in the PCR tubes were subjected to the following thermal cycles:

94°C------------------1 min ⎫
25°C-----------------10 min ⎬ 1 cycle
65°C------------------15 min ⎭

94°C------------------1 min ⎤
55°C------------------1 min ⎬—1 cycle
72°C------------------1 min ⎦

Thereafter, 1.25 µl of each of the primers P5 and P6 at 100 ng/µl concentration were added to the above PCR tubes and were further subjected to the following thermal cycles:

94°C------------------1 min ⎤—1 cycle

94°C------------------1 min ⎤
55°C------------------1 min ⎬—35 cycles
72°C------------------1 min ⎦

72°C------------------5 min ⎤
⎬— 1 cycle
10°C------------------Hold ⎦
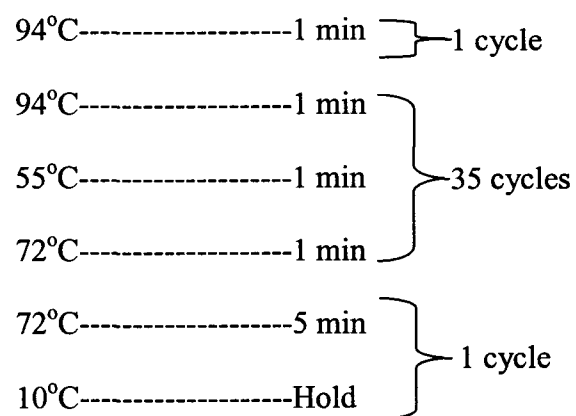
After all the cycles were completed, the resulting tubes were taken out and the components loaded onto a 1.5% (w/v) agarose gel. The gel after it was run as visualized on a UV transilluminator is shown in Figure 3.7. The gel clearly showed DNA bands at around 132 bp size, indicating the synthesis of the double-stranded DS3.
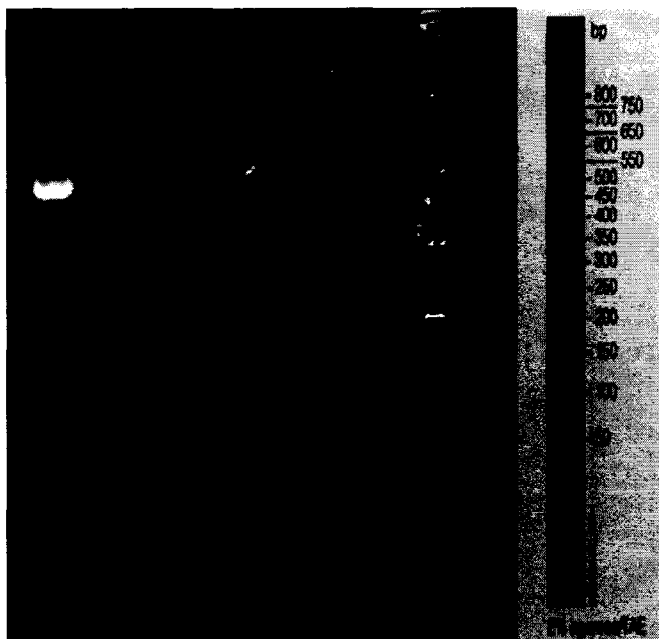
Figure 3.7. Agarose gel image showing the synthesis of the double-stranded DNA fragment, DS3.

Next, each of the gel extracted double-stranded fragments DS1, DS2 and DS3, were amplified by their corresponding primers in three separate PCR tubes as follows. In addition to 1 µl (100 ng/µl) of the each of the template and 1.25 µl (100 ng/µl) of the corresponding primers, the following components were also taken into each of those tubes.

Pfu polymerase----------------1.0 µl (2.5 U/µl)

10x *Pfu* buffer----------------5.0 µl

dNTP mix (25mM each) -----0.8 µl

DI water----------------------41.2 µl

After being mixed well, the components were subjected to the following thermal

cycles:

94°C--------------------1 min ⌐⌐— 1 cycle

94°C-------------------1 min ⌐

55°C------------------1 min }—35 cycles

72°C------------------1 min ⌐

72°C-----------------5 min ⌐

}—1 cycle

10°C------------------Hold ⌐
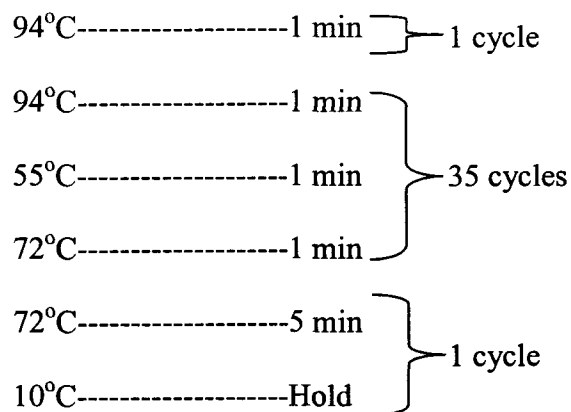
Thereafter the contents of each of those tubes were loaded onto a 1.5% (w/v)

agarose gel. The gel after it was run was visualized and photographed. The gel image as

in Figure 3.8 clearly shows that all the three double-stranded pieces were successfully

amplified.
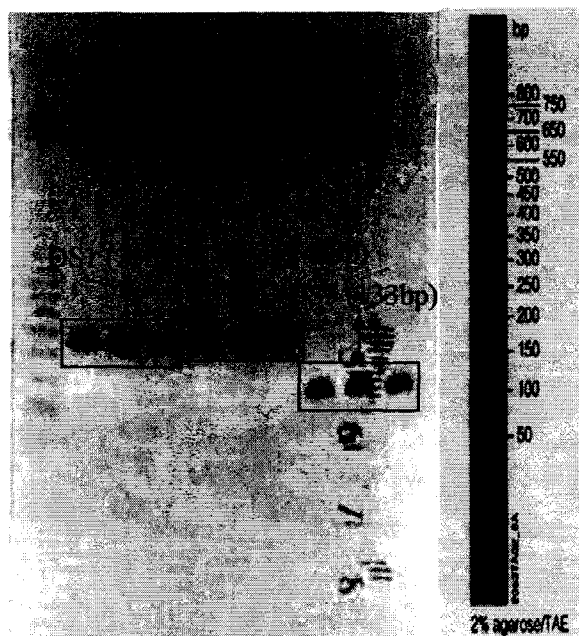
Figure 3.8. Agarose gel image showing the amplification of the double-stranded DNA fragments DS1, DS2 and DS3.

Next to synthesize the double-stranded fragment DS4 from DS2 and DS3, the following reaction components were taken into two separate PCR tubes:

DS2--------------------1.0 µl (100 ng/µl)

DS3--------------------1.0 µl (100 ng/µl)

DI water----------------39.6 µl

The above tubes were subjected to denaturation at 94 °C for 10 minutes. Then the tubes were taken out of the thermal cycler and vortexed. Thereafter the following components were added:

Pfu polymerase----------------0.5 µl (2.5 U/µl)

10x *Pfu* buffer----------------5.0 µl

dNTP mix (25mM each) -----0.4 µl

Then the components in the PCR tubes were subjected to the following thermal cycles:

94°C------------------1 min ⎫
25°C------------------10 min ⎬ 1 cycle
65°C------------------15 min ⎭

94°C------------------1 min ⎫
55°C------------------1 min ⎬ 1 cycle
72°C------------------1 min ⎭

After that, 1.25 µl of each of the primers P3 and P6 at 100 ng/µl concentration were added to the above PCR tubes and were further subjected to the following thermal cycles:

```
94°C------------------1 min  ⊐}— 1 cycle

94°C------------------1 min  ⊓
                              │
55°C------------------1 min   }— 35 cycles
                              │
72°C------------------1 min  ⊔

72°C------------------5 min  ⊤
                             }— 1 cycle
10°C------------------Hold   ⊥
```
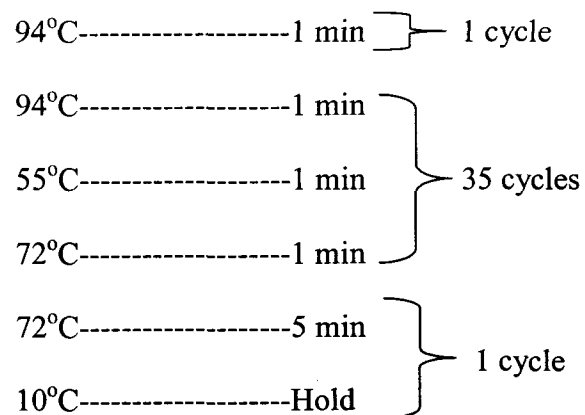
After all the cycles were completed, the components were loaded onto a 1.5%

(w/v) agarose gel. The gel thus obtained showed DNA bands at around 359 bp size,

indicating the synthesis of the double-stranded DS4 as in Figure 3.9.
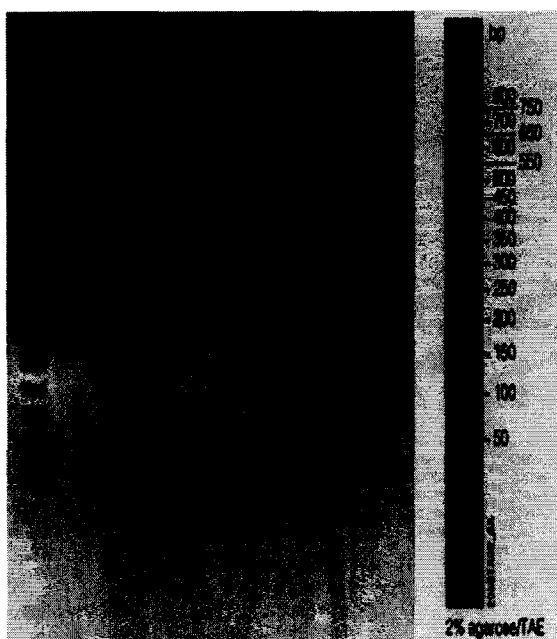
Figure 3.9. Agarose gel image showing the synthesis of the double-stranded DNA fragment, DS4.

The bands on the gel thus obtained were excised and the DNA within, was

extracted as before and stored. After verifying the amplification of DS4 with its end

primers, synthesis of DS5 was attempted using DS1 and DS4. The following components

were taken into two separate PCR tubes:

DS1--------------------1.0 μl (100 ng/μl)

DS4--------------------1.0 μl (100 ng/μl)

Pfu polymerase----------------0.5 μl (2.5 U/μl)

10x *Pfu* buffer-----------------5.0 μl

dNTP mix (25mM each) -----0.4 μl

DI water-----------------------41.2 μl

The components in the two tubes were then subjected to the following thermal

cycles:

94°C------------------10 min ⎤
35°C-----------------3 min ⎥
                                    ⎬ 5 cycles
60°C-----------------3 min ⎥
72 °C------------------1 min ⎦

Thereafter, 1.25 μl of each of the primers P1 and P6 at 100 ng/μl concentration

were added to the above PCR tubes along with 0.5 μl of *Pfu* polymerase and 0.3 μl of

dNTP mix and were further subjected to the following thermal cycles:

94°C-------------------1 min ⎤⎦— 1 cycle

94°C------------------1 min ⎤
55°C------------------1 min ⎬ 35 cycles
72°C------------------1 min ⎦

72°C------------------10 min ⎤
                                       ⎬ 1 cycle
10°C------------------Hold ⎦

Thereafter, the contents of both the tubes were loaded onto a 1.5% (w/v) agarose gel and were run. The gel thus obtained is as shown in Figure 3.10. The gel clearly showed bands at 573 bp size.
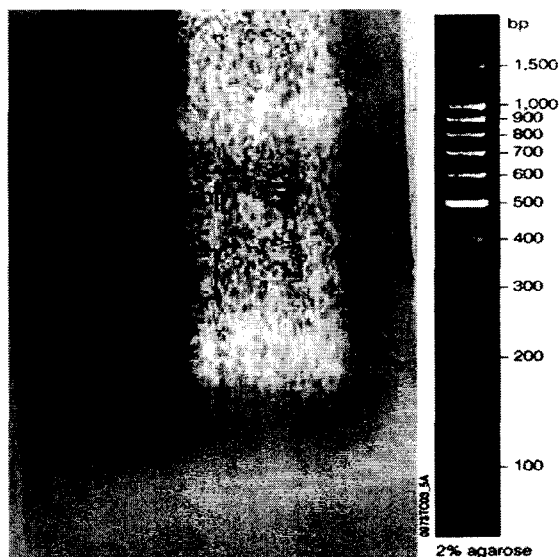
Figure 3.10. Agarose gel picture showing the synthesis of the double-stranded, DS5.

The bands from the gel were cut and the DNA extracted and stored. Thereafter, the double-stranded DS5 was attempted to be amplified using the end primers, P1 and P6. In spite of many attempts, the gene could not be amplified. As such an alternative method to synthesize the C2 domain gene was sought as described in Section 3.1.2.3.

**3.1.2.3. Alternate method for the C2 domain
  gene synthesis**

A 552 bp gene corresponding to the C2 domain and consisting of only the highest frequency codons was ordered from a commercial source to be synthesized and cloned into a pDrive cloning vector. The 552 bp gene ordered is as shown in Figure 3.11.

```
ATGACGGGCTCCATTAAAATGAACAACAAACCGCTGTTTCTGCATCATGTG
 M  T  G  S  I  K  M  N  N  K  P  L  F  L  H  H  V
TACTGCCCGAGGTAATTTTACTTGTTGTTTGGCGACAAAGACGTAGTACAC

ATCATGCATGGCATTCCGAACTTTGAAAGCAAAGGCGGCTGCCGCCCGTTT
 I  M  H  G  I  P  N  F  E  S  K  G  G  C  R  P  F
TAGTCCGTACCGTAAGGCTTGAAACTTTCGTTTCCGCCGACGGCGGGCAAA

CTGCGCATTTATCAGGCGATGCAGCCGGTGTATACCAGCGGCATTTATAAC
 L  R  I  Y  Q  A  M  Q
GACGCGTAAATAGTCCGCTACGTCGGCCACATATGGTCGCCGTAAATATTG

ATTCCGGGCGATAGCCAGACCAGCGTGTGCATTACCATTGAACCGGGCCTG
 I  P
TAAGGCCCGCTATCGGTCTGGTCGCACACGTAATGGTAACTTGGCCCGGAC

CTGCTGAAAGGCGATATTCTGCTGAAATGCTATCATAAAAAATTTCGCAGC
 L
GACGACTTTCCGCTATAAGACGACTTTACGATAGTATTTTTTAAAGCGTCG

CCGGCGCGCGATGTGATTTTTCGCGTGCAGTTTCATACCTGCGCGATTCAT
 P  A  R  D  V  I  F  R
GGCCGCGCGCTACACTAAAAAGCGCACGTCAAAGTATGGACGCGCTAAGTA

GATCTGGGCGTGGTGTTTGGCAAAGAAGATCTGGATGATGCGTTTAAAGAT
 D  L  G  V  V  F  G  K  E  D  L  D  D  A  F  K  D
CTAGACCCGCACCACAAACCGTTTCTTCTAGACCTACTACGCAAATTTCTA

GATCGCTTTCCGGAATATGGCAAAGTGGAATTTGTGTTTAGCTATGGCCCG
 D  R  F  P  E  Y  G  K  V  E  F  V  F  S  Y  G  P
CTAGCGAAAGGCCTTATACCGTTTCACCTTAAACACAAATCGATACCGGGC

GAAAAAATTCAGGGCATGGAACATCTGGAAAACGGCCCGAGCGTGAGCGTG
 E  K  I  Q  G  M  E  H  L  E  N  G  P  S  V  S  V
CTTTTTTAAGTCCCGTACCTTGTAGACCTTTTGCCGGGCTCGCACTCGCAC

GATTATAACACCAGCGATCCGCTGATTCGCTGGGATAGCTATGATAACTTT
 D  Y  N  T  S  D  P  L  I  R  W  D  S  Y  D  N  F
CTAATATTGTGGTCGCTAGGCGACTAAGCGACCCTATCGATACTATTGAAA
```

AGCGGCCATCGCGATGATGGCATGGAAGAAGTGGTGGGCTAA

S G H R D D G M E E V V G *

TCGCCGGTAGCGCTACTACCGTACCTTCTTCACCACCCGATT

Figure 3.11. The 552 bp gene fragment used in the alternate method for the C2 domain gene synthesis. '*' indicates a stop codon.

Two end primers were also designed for the above gene such that a PCR performed on the pDrive cloning vector using them would yield the 573 bp C2 gene, which can later be cloned into a pET14b vector. The strategy to synthesize the C2 gene by PCR using the pDrive cloning vector as the template and then cloning it into the pET14b vector is illustrated in Figure 3.12.
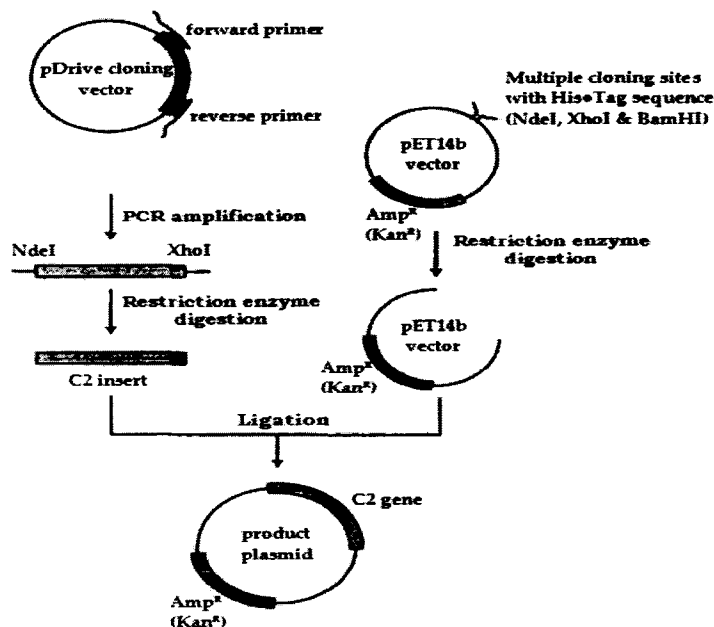
Figure 3.12. The PCR and cloning strategy used for the C2 gene to be expressed in *E.coli*.

The sequences of the forward primer, PF and the reverse primer, PR from 5' to 3' are given by AGGGAATTCCATATGACGGGCTCCATTAAAATG (33 bases) and

CCGCTCGAGTTAGCCCACCACTTCTTCCA (29 bases) respectively. The two primers

were supplied in lyophilized form with an O.D of 5 at 260 nm. The oligonucleotide

calculations for the primers 'PF' and 'PR' are shown in Table 3.2.

Table 3.2. Primer calculations and their concentrations.

Primer	A	G	C	T	Mol.Wt (Da)	E at 260nm $(M^{-1}cm^{-1})$	pmol	µg
PF	11	8	6	8	10185.67	337488	14815.34	150.904
PR	5	4	13	7	8709.63	250810.9	19935.33	173.629

About 251.5 µl of DI water was added to the vial of forward primer, PF to get a

stock concentration of 600 ng/µl. Similarly, 289.4 µl of DI water was added to the vial of

reverse primer, PR to get a stock concentration of 600 ng/µl.

Next, a hot start PCR was performed by assembling the following components

into a PCR tube:

C2 containing pDrive cloning vector--------1.00 µl (100 ng/µl)

Forward primer, PF--------------------------- 1.25 µl (600 ng/µl)

Reverse primer, PR-------------------------- 1.25 µl (600 ng/µl)

10x *Pfu* polymerase buffer--------------------5.00 µl

dNTP mix---------------------------------------2.00 µl (5 mM each)

DI water---------------------------------------38.80 µl

The above components were subjected to 94°C for 3 minutes in the thermal

cycler. After 3 minutes, 0.7 µl of *Pfu* polymerase was added to it and was further

subjected to the following thermal cycles:

```
94°C-----------------------5 min  ┐
                                   ├─ 1 cycle
94°C-----------------------1 min  ┐
                                   │
55°C-----------------------1 min  ├─ 35 cycles
                                   │
72°C-----------------------2 min  ┘

72°C-----------------------10 min ┐
                                   ├─ 1 cycle
10°C-----------------------Hold   ┘
```
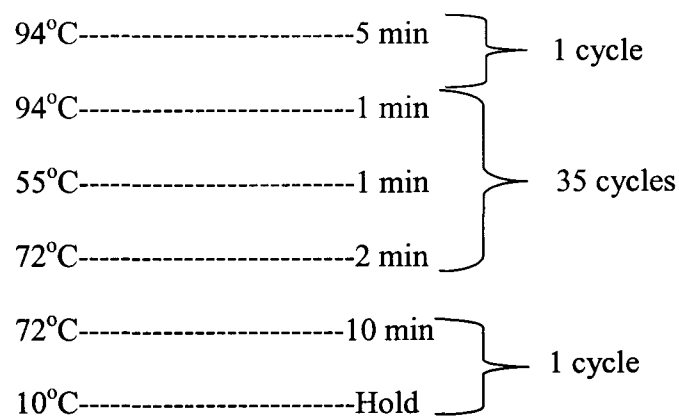
After the reaction was complete, the components were loaded onto a 0.7% (w/v)

agarose gel. The gel thus obtained is as shown in Figure 3.13. The bands obtained at 573

bp from the gel were cut and the DNA within extracted and stored.
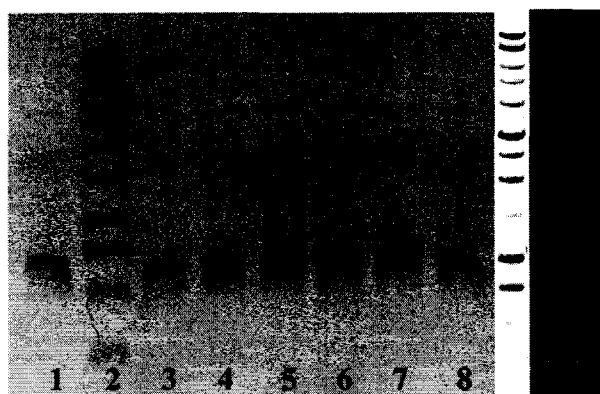
Figure 3.13. Agarose gel image showing the synthesis of C2 gene by PCR from pDrive cloning vector. Lane 2 shows the DNA marker, while all other lanes show the C2 domain gene PCR product.

Thereafter, the gel extracted C2 gene was successfully amplified by PCR using

the end primers, PF and PR.

3.2. Cloning of the C2 Domain Gene

3.2.1. Introduction

This section consists of various subtopics covering the theory and background relating to the procedures involved in the cloning of the C2 domain gene.

3.2.1.1. Recombinant DNA technology

Recombinant DNA refers to the insertion of DNA from one type of organism to another and allows one to construct any protein, simply by changing the genetic plans that are used to build it. For this purpose, naturally occurring enzymes, called restriction enzymes isolated from bacteria are used. These enzymes cut DNA at specific palindromic sequences and create sticky or blunt ends [84]. Examples of various restriction enzymes creating sticky and blunt ends in DNA are shown in Figure 3.14.
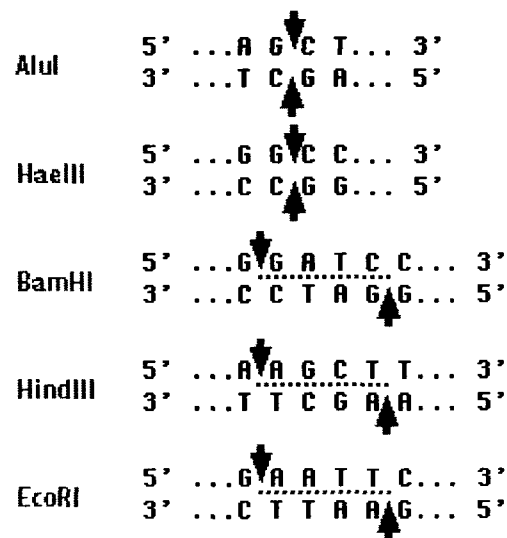
Figure 3.14. Restriction enzymes producing sticky and blunt ends. AluI and HaeIII produce blunt ends, while BamHI, HindIII and EcoRI produce sticky ends.

The sticky ends readily associate with other sticky ends of complementary sequence such as that from a bacterial DNA and are stuck together using another enzyme

called DNA ligase. Once the ends are ligated, the resulting recombinant plasmid is used

to transform bacterial cells and is multiplied within it as the bacterium divides. This

process is known as DNA cloning. Any protein of interest is generally expressed in

engineered cells, using expression vectors. These vectors are plasmids that contain the

gene specifying the protein along with a highly active promoter sequence. The promoter

then aids in the production of messenger RNA in large quantities based on the plasmid

DNA in the vector. Finally at the ribosomes, the cell synthesizes protein of interest in

large quantities based on this messenger RNA.

3.2.1.2. pET14b vector

pET14b was the vector of choice for the cloning and expression of the C2 domain

gene. It has an N-terminal hexa-histidine tag followed by a thrombin site and three

cloning sites as can be seen from its cloning/expression region in Figure 3.15.
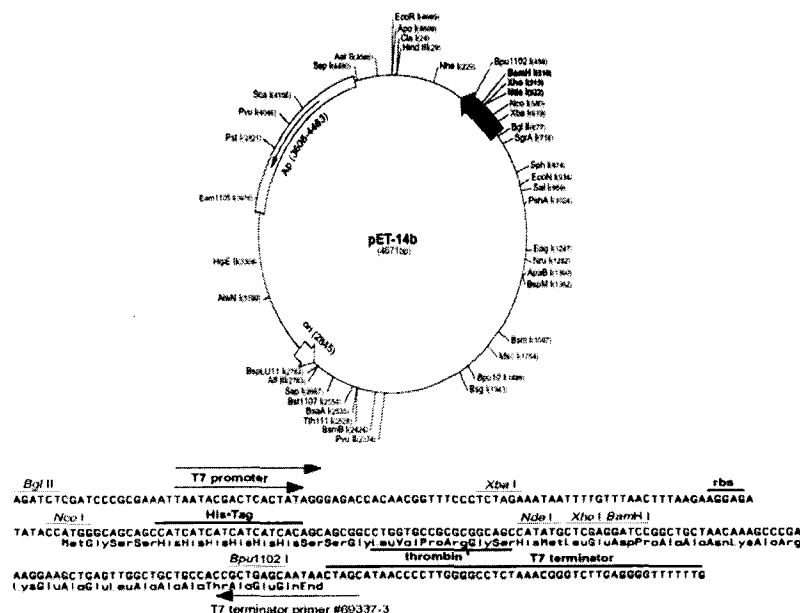
Figure 3.15. Vector map of pET14b with its cloning/expression region [87].

NdeI and *XhoI* are the two sites chosen to clone the C2 domain gene. More about the features of the pET system are discussed in Section 4.1.1.3.

3.2.1.3. DNA sequencing

It is usually done by employing the dideoxy method in which DNA is synthesized by PCR in the presence of chain-terminating dideoxyribonucleoside triphosphates. To determine the complete sequence of a DNA fragment, the double-stranded DNA is first separated into its single strands, and one of the strands is used as the template for sequencing, which consists of four separate DNA synthesis reactions. In each of these reactions, along with the template, there exists DNA polymerase, a short primer, the four deoxyribonucleoside triphosphates (dATP, dGTP, dCTP, dTTP) and one of the four different dideoxyribonucleoside triphosphates (ddATP, ddGTP, ddCTP, ddTTP). The dideoxyribonucleoside triphosphates lack the 3' hydroxyl group in comparison to the normal deoxyribonucleoside triphosphates as shown in Figure 3.16. Hence during the above reactions, when the growing DNA chain encounters the dideoxyribonucleotide present, the addition of the next nucleotide is blocked and the DNA chain is terminated. Each of the above reactions produces a set of DNA copies that terminate at different points in the sequence. The products of these four reactions are then separated by electrophoresis in four parallel lanes of a polyacrylamide gel. The newly synthesized fragments are detected by a radioactive or a fluorescent label that had been incorporated into the primer or into one of the deoxyribonucleoside triphosphates. In each lane, the bands represent fragments that have terminated at a given nucleotide but at different positions in the DNA. By reading off the bands in order, starting at the bottom of the gel

and working across all lanes, the DNA sequence of the newly synthesized strand can be

determined [81].

deoxyribonucleoside triphosphate   dideoxyribonucleoside triphosphate

allows strand extension at 3' end   prevents strand extension at 3' end
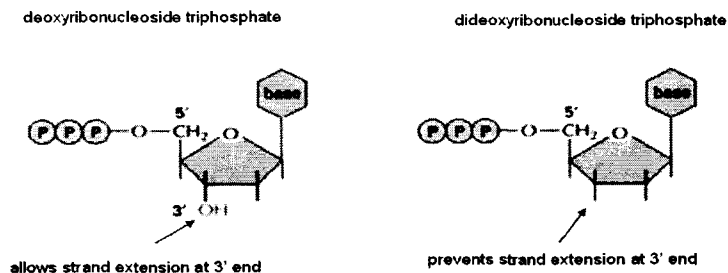
Figure 3.16. Chemical structures of deoxy- and dideoxyribonucleoside triphosphates [81].

3.2.2. Methods, Results and Discussion

The subtopics within this section deal with the experimental methods used in the

cloning of the C2 domain gene including their results and discussion.

3.2.2.1. Cloning of the C2 domain gene

Initially, restriction enzyme digestion of the pET14b vector was performed by

arranging the following components in a PCR tube:

RE 10x buffer---------------------- 4.00 μl

Acetylated BSA--------------------- 0.40 μl (10 μg/μl)

pET 14b vector----------------------1.00 μl (100 ng/μl)

Nde I restriction enzyme----------- 1.00 μl (10 U/μl)

DI water-----------------------------32.10 μl

The PCR tube was then incubated at room temperature for 12 hours. After that 1.5

μl of *Xho I* (10 U/μl) was added to it and the restriction digestion was continued for

another 5 hours. Thereafter, the product was PCR purified.

Next, a restriction enzyme digestion for the C2 domain gene product was setup in a PCR tube with the following components:

RE 10x buffer---------------------- 4.00 μl

Acetylated BSA--------------------- 0.40 μl (10 μg/μl)

C2 gene product----------------------1.00 μl (100 ng/μl)

Nde I restriction enzyme------------ 1.00 μl (10 U/μl)

Xho I restriction enzyme-------------1.00 μl (10 U/μl)

DI water-----------------------------32.60 μl

The PCR tube was then incubated at room temperature for 8 hours. Thereafter, the product was PCR purified. After the above two restriction digestion procedures, a ligation reaction was setup in a PCR tube with the following components:

RE digested pET 14b---------------- 0.5 µl

RE digested C2 gene---------------- 3.0 µl

10x T4 DNA ligase buffer-----------1.50 µl

T4 DNA ligase------------------------0.75 μl (0.5 U)

DI water-------------------------------9.25 µl

The above reaction sample was kept at 15°C overnight. After the reaction, the ligation mixture was stored at -20°C. LB agar plates containing ampicillin at 5 μg/ml final concentration were prepared as per the protocol B.4 in Appendix B. Three micro centrifuge tubes, each containing 50 µl of XL2 blue ultra-competent cells were transformed with 1 µl of uncut pET14b (1 μg/μl), 5 μl of the above ligation mixture, 5 μl of the restriction digested pET14b respectively as per the protocol B.6 in Appendix B. After the transformation, the three samples were plated onto the LB agar plates and then

incubated at 37°C overnight. Next morning, upon the inspection of plates, it was found that the positive control plate (uncut pET14b) had a lot of colonies as expected. The negative control plate (cut pET14b) had five colonies, while the plate with the ligation mixture had just one colony. This single colony formed was picked up and used to inoculate 100 ml of freshly prepared LB broth containing 5 μg/ml of ampicillin. LB broth was prepared as per the protocol B.5 in Appendix B. The culture was then kept shaking at 37°C overnight. The next morning, minipreps kit was used to extract the plasmid from the transformed cell culture as per the protocol B.7 in Appendix B. After the plasmid extraction, 10 μl of it was taken and was run on a 0.7% agarose gel along side 10 μl of pET14b (50 ng/μl). The gel obtained thereafter is as shown in Figure 3.17. From the figure, pET14b vector having a size of 4671 bp can be seen in Lane1, while the recombinant C2 plasmid having an expected size of 5244 bp (4671+573) can be seen in Lane 3. The difference in the sizes of both the plasmids does indicate the presence of an insert within the recombinant plasmid.
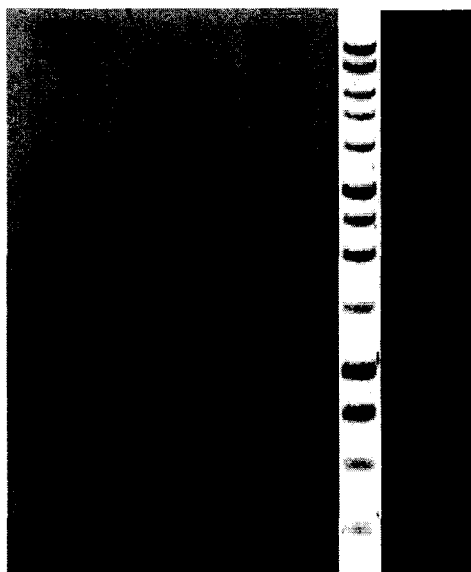
Figure 3.17. Agarose gel image showing the restriction digestion of pET14b and C2 containing pET14b. Lane 1: Pet14b; Lane 2: DNA marker; Lane 3: C2 containing pET14b.

Again to check the presence of insert within the picked-up colony, the following restriction enzyme tests were done. Four different restriction enzyme reactions were setup in tubes A, B, C and D with components as shown in Table 3.3.

Table 3.3. Restriction enzyme reaction components for the C2 domain gene cloning.

	Tube-A	Tube-B	Tube-C	Tube-D
10x Restriction enzyme buffer	4.0 µl	4.0µl	4.0µl	4.0µl
Acetylated BSA	0.4 µl (10 µg/µl)	0.4 µl (10 µg/µl)	0.4 µl (10 µg/µl)	0.4 µl (10 µg/µl)
C2-pET14b	4.0 µl	4.0 µl	4.0 µl	4.0 µl
XhoI	-	1.5 µl (10 U/µl)	1.5 µl (10 U/µl)	1.5 µl (10 U/µl)
NdeI	1.0 µl (10 U/µl)	-	-	1.0 µl (10 U/µl)
DI water	32.1µl	32.1µl	30.6µl	29.1 µl

All the tubes were then incubated at 37°C for 4 hours and later run on a 0.7% agarose gel. The gel thus obtained is as shown in Figure 3.18. Lanes 1 and 2 that were loaded with samples from tubes A and B respectively, show a single band at 4671 bp, which represents the linearized pET14b. Lane 4 loaded with the sample from tube C shows a single band at 5244 bp, which represents the linearized recombinant C2 plasmid. Lane 5 loaded with the sample from tube D show a band at 4671 bp as expected. However, the band corresponding to the insert could not be visible at 573 bp on the gel.
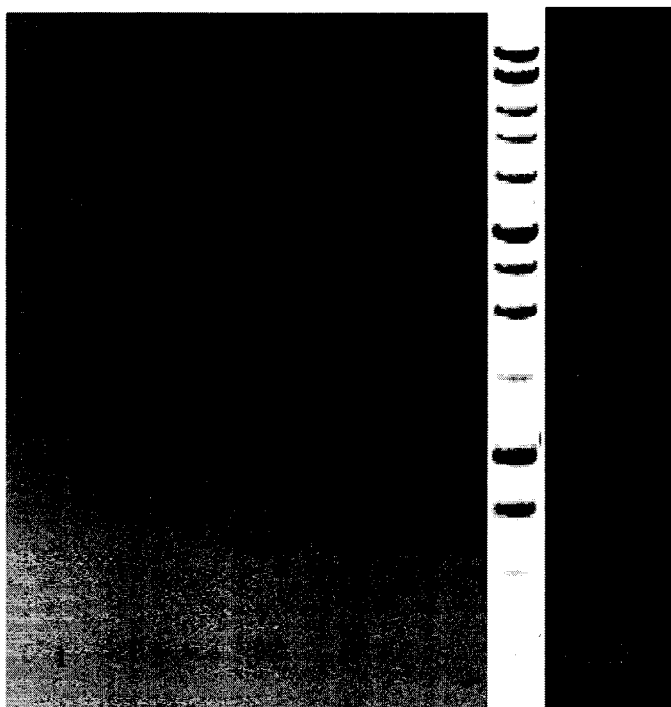
Figure 3.18. Agarose gel image showing the single and double restriction digestion analysis of the recombinant C2 plasmid. Lanes 1 and 2: linearized pET14b vector; Lane 3: DNA marker; Lane 4: linearized recombinant C2 plasmid with single digestion and Lane 5: linearized recombinant C2 plasmid with double digestion.

To confirm the presence of the C2 insert within the recombinant plasmid by PCR, the following components were set up in a PCR tube, T1:

C2 containing pET14b------------------------1.00 μl

Forward primer, PF----------------------------0.62 μl (100 ng/μl)

Reverse primer, PR----------------------------0.62 μl (100 ng/μl)

10x *Pfu* polymerase buffer--------------------2.50 μl

dNTP mix--1.00 μl (5 mM each)

DI water--18.76 μl

Another PCR tube, T2 was used to accommodate with the same volumes of the above components, except for the concentrations of both the primers, which was 600 ng/μl. Both the tubes are subjected to PCR with the following thermal cycles:

94°C------------------------5 min ⎤⎰ 1 cycle

94°C------------------------1 min ⎤
55°C------------------------1 min ⎬ 30 cycles
72°C------------------------2 min ⎦

72°C------------------------10 min ⎤⎰ 1 cycle
10°C------------------------Hold ⎦

After the reaction, both the PCR products were run on a 0.7% agarose gel, along side the restriction enzyme cut samples from tubes A and B. The resulting gel picture is as shown in Figure 3.19. Both Lanes 1 and 2 show the linearized pET14b at 4671 bp. Lanes 4 and 5 that were loaded with PCR products from tubes T1 and T2 respectively show the evidence of the C2 domain gene insert within the recombinant C2 plasmid.
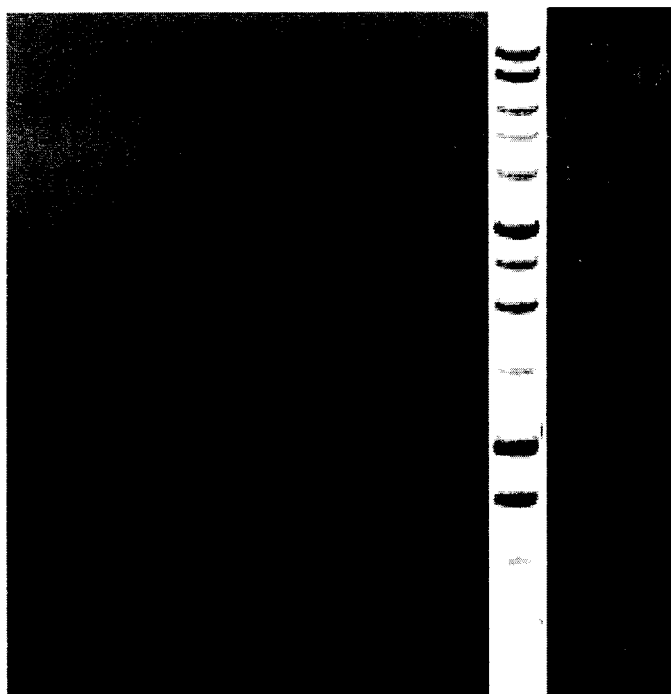
Figure 3.19. Agarose gel image confirming the presence of the C2 gene insert within the recombinant C2 plasmid. Lanes 1 and 2: linearized pET14b vector; Lane 3: DNA marker; Lane 4: PCR product from tube T1 and Lane 5: PCR product from tube T2.

3.2.2.2. DNA sequencing of the C2 gene

After making sure that the insert present within the recombinant plasmid was at the expected size of 573 bp, the recombinant C2 plasmid (C2 containing pET14b) was sent to Retrogen, Inc (San Diego, CA) for DNA sequencing. Dideoxynucleotide method of sequencing was used for the sequence analysis of the recombinant C2 plasmid. Both forward and reverse sequencing reactions were done using the T7 promoter and T7 terminator primers respectively. The sequencing results thus obtained were analyzed by Chromas software. The chromatogram of the forward sequencing reaction of the recombinant C2 plasmid using T7 promoter primer is as shown in Figure 3.20.
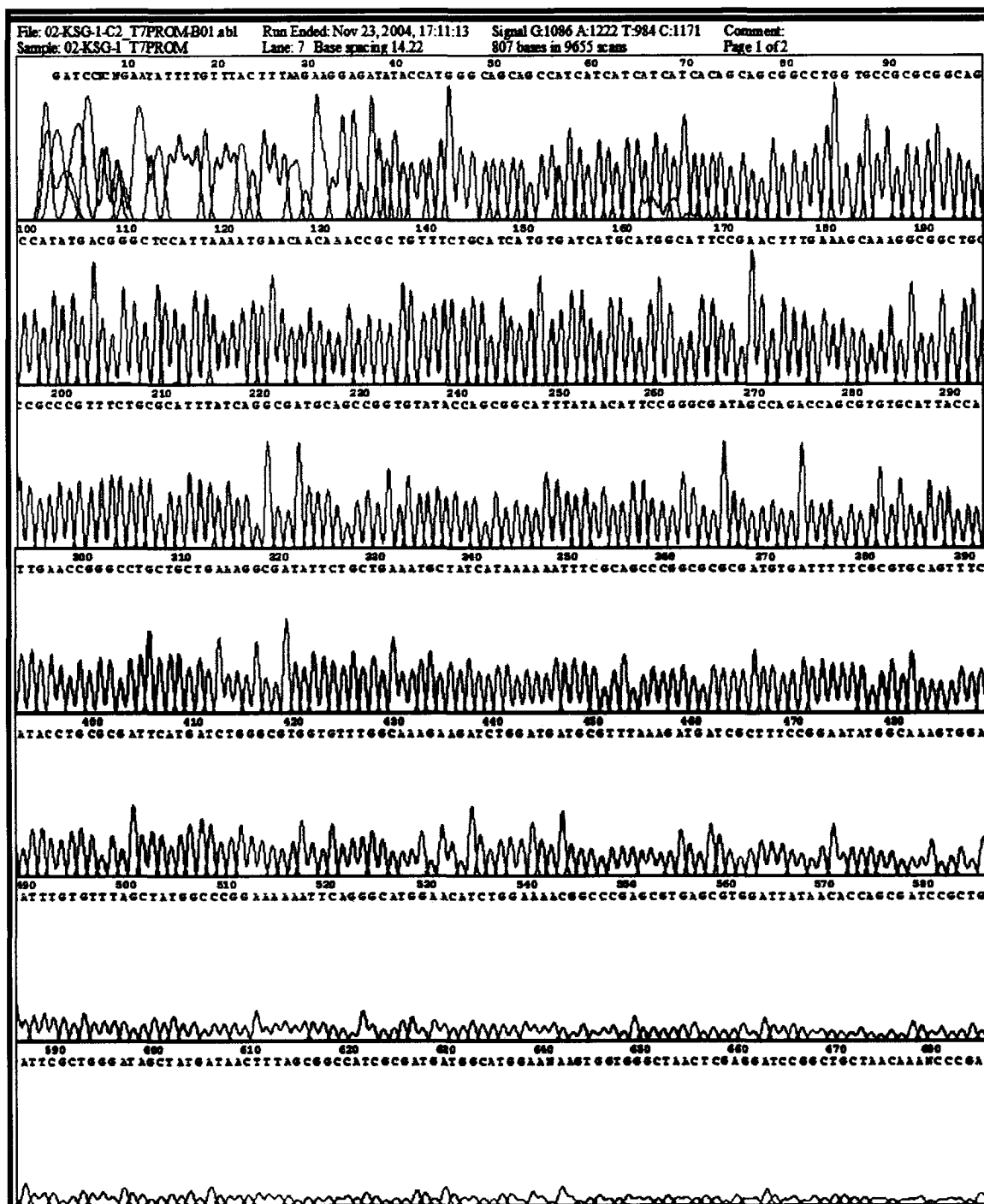
Figure 3.20. Chromatogram of the forward sequencing reaction of the recombinant C2 plasmid using T7 promoter primer.

The text sequence exported from the corresponding chromas file is as follows.

GATCCTCNGAATATTTTGTTTACTTTAAGAAGGAGATATACCATGGGCAGCA

GCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATAT

GACGGGCTCCATTAAAATGAACAACAAACCGCTGTTTCTGCATCATGTGATC

ATGCATGGCATTCCGAACTTTGAAAGCAAAGGCGGCTGCCGCCCGTTTCTGC

GCATTTATCAGGCGATGCAGCCGGTGTATACCAGCGGCATTTATAACATTCCG

GGCGATAGCCAGACCAGCGTGTGCATTACCATTGAACCGGGCCTGCTGCTGA

AAGGCGATATTCTGCTGAAATGCTATCATAAAAAATTTCGCAGCCCGGCGCG

CGATGTGATTTTTCGCGTGCAGTTTCATACCTGCGCGATTCATGATCTGGGCG

TGGTGTTTGGCAAAGAAGATCTGGATGATGCGTTTAAAGATGATCGCTTTCCG

GAATATGGCAAAGTGGAATTTGTGTTTAGCTATGGCCCGGAAAAAATTCAGG

GCATGGAACATCTGGAAAACGGCCCGAGCGTGAGCGTGGATTATAACACCAG

CGATCCGCTGATTCGCTGGGATAGCTATGATAACTTTAGCGGCCATCGCGATG

ATGGCATGGAA█AAGTGGTGGGCTAACTCGAGGATCCGGCTGCTAACAAANC

CCGAAAGGAANCTGANTTGGCTGCTGCCACCGCTGANCAATAACTANCATAA

CCCCTTGGGGCCTCTAAACNGGTCTTGAGGGGTTTTTTGCTNAANGANGAACT

ATNTCCNGATNTNCNNNNGANN

The text highlighted in yellow corresponds well to the actual sequence of the C2 gene at every base except at the base numbered 640 (highlighted in green). The chromatogram of the reverse complement of the reverse sequencing reaction of the recombinant C2 plasmid using the T7 terminator primer is as shown in Figure 3.21.

Figure 3.21. Chromatogram of the reverse complement of the reverse sequencing reaction of the recombinant C2 plasmid using T7 terminator primer.

The text sequence exported from the corresponding chromas file is as follows.

NNCCCNAANNGNNNANCCCNATNTTNCCCCNTNGNNGANGTNGGCNATATA

GGNGCCAGCANCCGCACCTGTGGNGCCGGTGANNCCGGCCNCGATGCGTCC

GGCGTAGAGGATCGAGATCTCGATCCCCGCGAAATTAATACGACTCACTATA

GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGG

AGATATACCATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGG

TGCCGCGCGGCAGCCATATGACGGGCTCCATTAAAATGAACAACAAACCGCT

GTTTCTGCATCATGTGATCATGCATGGCATTCCGAACTTTGAAAGCAAAGGCG

GCTGCCGCCCGTTTCTGCGCATTTATCAGGCGATGCAGCCGGTGTATACCAGC

GGCATTTATAACATTCCGGGCGATAGCCAGACCAGCGTGTGCATTACCATTG

AACCGGGCCTGCTGCTGAAAGGCGATATTCTGCTGAAATGCTATCATAAAAA

ATTTCGCAGCCCGGCGCGCGATGTGATTTTTCGCGTGCAGTTTCATACCTGCG

CGATTCATGATCTGGGCGTGGTGTTTGGCAAAGAAGATCTGGATGATGCGTTT

AAAGATGATCGCTTTCCGGAATATGGCAAAGTGGAATTTGTGTTTAGCTATG

GCCCGGAAAAAATTCAGGGCATGGAACATCTGGAAAACGGCCCGAGCGTGA

GCGTGGATTATAACACCAGCGATCCGCTGATTCGCTGGGATAGCTATGATAA

CTTTAGCGGCCATCGCGATGATGGCATGGAA█AAGTGGTGGGCTAACTCGAG

GATCCGGCTGCTAACAAGCCCGAAGAAGNAT

The text highlighted in yellow again corresponded well to the actual sequence of

the C2 gene at every base including the one at the base numbered 640 (highlighted in

green) from the forward sequencing chromas file. Thus, it was concluded that the gene

corresponding to the C2 domain of human tensin was successfully synthesized, cloned

and sequence verified.

CHAPTER 4

EXPRESSION, DETECTION AND PURIFICATION
OF THE C2 DOMAIN

4.1. Expression and Detection of the C2 Domain

4.1.1. Introduction

This section contains subtopics covering the theory and background information

relating to the procedures involved in the expression and detection of the C2 domain of

human tensin.

4.1.1.1. Gene expression systems

The different systems used for the heterologous expression of proteins are *E.coli*,

yeast, insect cells, and mammalian cells. Expression of proteins in *E.coli* is easy, quick

and cheap with the availability of many commercial and non-commercial expression

vectors with different N and C terminal tags. It also contains several different strains

which are optimized for special applications. More about *E.coli* is discussed in Section

4.1.1.2. Yeast is a eukaryotic organism that can be grown to very high densities and also

provides most of the eukaryotic post translational modifications such as phosphorylation,

glycosylation, acetylation and acylation. Insect cells are a higher eukaryotic system than

yeast and can carry out more complex post-translational modifications than *E.coli* and

yeast. They also have the best machinery for the folding of mammalian proteins and

65

hence produce highly soluble proteins of mammalian origin. The disadvantages of insect cells are the higher costs and low expression levels [88].

4.1.1.2. Significance and features of Escherichia coli (E.coli)

E.coli is the most frequently used prokaryotic expression system for the production of heterologous proteins. The advantages of *E.coli* are its ease of growth using inexpensive media, rapid growth in comparison to higher cells, well-characterized genetics and the availability of increasingly large number of cloning vectors and mutant host strains [89-91]. Virtually any gene can be expressed to a significant level in *E.coli*. However, the major disadvantages are the frequent accumulation of expressed gene product as insoluble aggregates called 'inclusion bodies' and the lack of extensive post translational processing machinery as in higher cells [89, 90]. A typical *E.coli* expression vector consists apart from the gene of interest, an origin of replication, a gene that confers antibiotic resistance, a promoter and a transcription terminator. The origin of replication determines the number of plasmid copies. Plasmid instability generally occurs with multi-copy plasmids, when certain number of cells inherits no copies of the plasmid upon cell division. As such the use of expression vectors encoding an antibiotic resistance gene coupled with the addition of antibiotics in the growth medium prevents plasmid instability. There are many strong promoters in *E.coli* that give very high rates of transcription initiation. Among those, *tac*, *lacUV5*, and *lpp-lac* are all regulated by the binding of the *lac* repressor and are derepressed by the addition of isopropyl-β-thiogalactoside (IPTG) [90-91].

4.1.1.3. Regulation of protein expression
in pET system

An ideal promoter should be tightly regulated and completely shut off in the absence

of an inducing agent. But with most promoters, this regulation is not possible and there is a

basal level of transcription even in the absence of an inducing agent. As such the

recombinant protein expressed in *E.coli* may interfere in the normal functioning of the cell

and therefore may be toxic to the bacteria. The degree of toxicity varies from protein to

protein. All the native *E.coli* promoters and those that are controlled by the *lac* repressor are

leaky because they always have a basal level of transcription [91, 92].

The pET system, developed by Studier [93] and colleagues, is a powerful protein

expression tool because the protein expression can be tightly controlled. In pET vectors,

genes are expressed from T7 promoters, which are transcribed only by the T7 RNA

polymerase. T7 RNA polymerase is a very specific and active enzyme that carries out mRNA

chain elongation at a rate five times higher than the *E.coli* RNA polymerase. The pET

vectors are transformed in a strain carrying a T7 RNA polymerase gene in the chromosome

under the control of a *lac* promoter which is inducible by IPTG. Since there will be small

amount of T7 RNA polymerase even in the uninduced state, the resulting low level of

transcription is harmful to the cells, if the target gene product is toxic. To avoid this problem,

the pET vector containing the target gene is expressed in host strains such as pLysS or pLysE

(See Figure 4.1A). These strains contain a compatible chloramphenicol-resistant plasmid that

provides a small amount of T7 lysozyme, which in turn is a powerful inhibitor of the T7

RNA polymerase (See Figure 4.1B). The T7 lysozyme thus synthesized by the host cells is

sufficient to inhibit transcription by the basal levels of T7 RNA polymerase [91, 92]. In this

present project, the C2 cloned pET-14b was transformed into BL21 (DE3) pLysS strain for the tight regulation of protein expression.
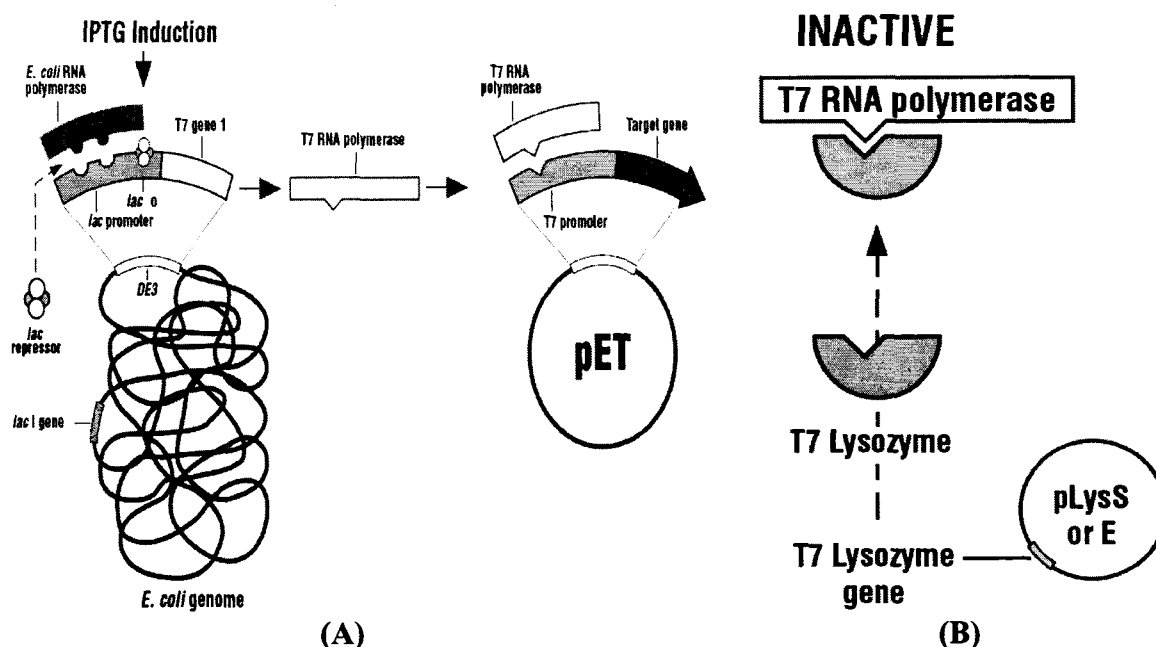
Figure 4.1. Regulation of protein expression in the pET system. (A) Synthesis of T7 RNA polymerase. (B) Inhibition of T7 RNA polymerase by T7 lysozyme in pLysS and pLysE host strains [92].

4.1.1.4. Protein electrophoresis

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) is the most widely used method for analyzing protein mixtures qualitatively. SDS $(CH_3-(CH_2)_{10}-CH_2OSO_3^-Na^+)$ is an anionic detergent. The protein samples to be run on SDS-PAGE are first boiled for 5 minutes in sample buffer containing β-mercaptoethanol and SDS. The mercaptoethanol reduces any disulphide bonds present within the sample holding together the protein tertiary structure, while the SDS binds strongly to the proteins and denatures them. Each protein in the sample is fully denatured by this treatment and opens up into a rod-shaped structure with a series of negatively charged SDS molecules along the polypeptide chain. The original native charge on the molecules is therefore completely saturated by the

negatively charged SDS molecules. Apart from SDS and β-mercaptoethanol, the sample buffer also contains bromophenol blue, an ionizable tracking dye, that allows the electrophoretic run to be monitored, and glycerol, which gives the sample solution density to be settled onto the bottom of the loading well [94]. In the preparation of gels for SDS-PAGE, acrylamide is mixed with bisacrylamide to form a cross-linked polymer network (as in Figure 4.2), when the polymerizing agent ammonium persulphate is added. For the same, ammonium persulphate produces free radicals faster in the presence of TEMED (N, N, N, N'-tertamethylenediamine) [95].
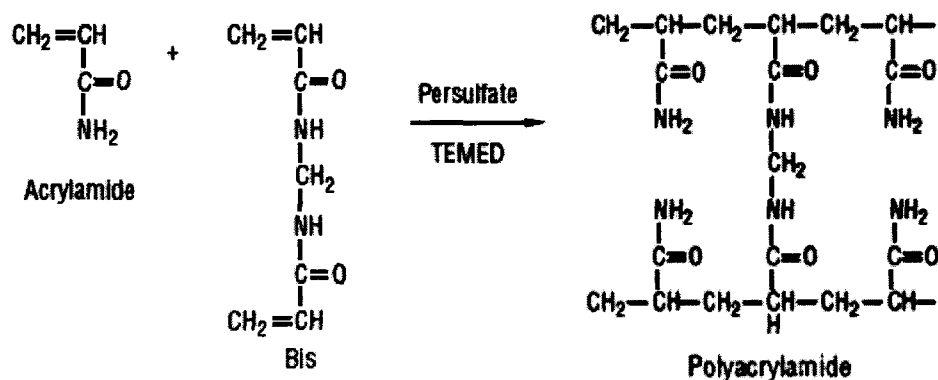
Figure 4.2. Polymerization and cross-linking of acrylamide [95].

During SDS-PAGE, when the current is applied the negatively charged protein-SDS complexes move towards the anode and because they have the same charge per unit length, they travel into the gel under the applied electric field with the same mobility. However, as they pass through the gel, the proteins separate. The smaller proteins migrate more easily within the gel, while the larger proteins are retarded by frictional resistance due to the sieving effect of the gel. The size of the pores created in the gel is inversely related to the amount of acrylamide. Gels with a lower percentage of acrylamide are typically used to resolve large

proteins, while gels with higher percentage of acrylamide are used to resolve smaller proteins [95]. Typically, the resolving gels used are 15% polyacrylamide gels. All the electrophoresis experiments used in the expression and purification of the C2 domain of human tensin were made of 12% polyacrylamide.

4.1.1.5. Protein (Western) blotting

Protein blotting is a method in which a pattern of separated proteins from a gel are transferred onto a sheet of nitrocellulose paper. This transfer is done by passing a current at right angles to the gel, which causes the separated proteins to electrophorese out of gel and into the nitrocellulose sheet. The nitrocellulose with its transferred protein is called as a blot. Once transferred onto the nitrocellulose, the blot is probed by using an antibody to detect a specific protein. The blot is first incubated in a protein solution usually 10% (w/v) bovine serum albumin (BSA) or 5% (w/v) non-fat dried milk, which will block all the remaining hydrophobic binding sites on the nitrocellulose sheet. The blot is then incubated in the diluted solution of primary antibody (IgG) directed against the protein of interest. This IgG molecule will bind to the blot if it detects its antigen, thus identifying the protein of interest. To visualize this interaction, the blot is incubated further in a solution of a secondary antibody. This secondary antibody is appropriately labeled so that its interaction with primary antibody can be visualized on the blot. One of the most common detection methods uses an enzyme-linked secondary antibody. In this method, following treatment with enzyme-linked secondary antibody, the blot is incubated in enzyme-substrate solution. Then, the enzyme coverts the substrate into an insoluble colored product precipitated onto the nitrocellulose. The presence of colored band therefore indicates the position of the protein of interest. By careful comparisons of the blot with a stained gel of the same sample, the protein

of interest can be identified. One of the enzymes used in enzyme-linked antibodies is horseradish peroxidase, which, with H_2O_2 as substrate, oxidizes 4-chloro-1-napthol into an insoluble blue product [97].

Dot blot is very similar to the western blotting except that the proteins are applied directly onto the nitrocellulose rather than transferred from a gel. Cell lysates containing the over expressed protein can also be applied to the nitrocellulose membrane in dot blot procedures. In the detection of histidine tagged C2 domain, Penta-His HRP conjugate kit was used as described in Section 4.1.2.3.

4.1.2. Methods, Results and Discussion

The subtopics within this section deal with the experimental methods related to the overexpression and detection of the C2 domain, including their results and discussion.

4.1.2.1. Transformation of the C2 plasmid into E.coli BL21 (DE3) pLysS cells

After verifying the sequencing results for the cloned C2 gene, C2 plasmid and pET14b (positive control) were used separately to transform BL21 (DE3) pLysS cells as described in the following protocol.

(1) Two aliquots each of 100 μl of BL21 (DE3) pLysS competent cells were taken out from a -80 C freezer and thawed on ice.

(2) The cells were then gently mixed and were taken into two 14-ml falcon polypropylene round-bottom tubes.

(3) 1.7 μl of the β-mercaptoethanol was added to each of the polypropylene tubes to its final concentration of 25 mM. Then the tubes were gently swirled.

(4) The tubes were incubated on ice for 10 min swirling gently every two min.

(5) Next, about 10 μl of C2 plasmid and 1 μl of pET14b (50 ng/μl) were added to the two tubes separately and gently swirled.

(6) Then both the tubes were incubated on ice for 30 minutes. During this time, the SOC medium supplied with the kit was preheated in a water bath at 42°C.

(7) After that, each of the two transformation reactions was heat-pulsed in the water bath at 42°C for 45 seconds. Then they were incubated on ice for 2 minutes.

(8) Then, 0.9 ml of the preheated SOC medium from step 6 was added to each of the transformation reactions and then both the tubes were kept shaking at 37°C for 1 hr at 225 rpm.

(9) After that 200 μl of cells from each of those transformation reactions were spread onto two separate LB agar plates containing ampicillin (50 μg/ml) and chloramphenicol (34 μg/ml).

(10) The plates were then incubated at 37°C overnight for 15 hours and were inspected for the colonies thereafter.

About 30 colonies were formed on the plate that was plated with C2 transformed cells, while there were about 50 colonies on the positive control plate that was plated with pET14b transformed cells. Six random colonies were picked from the C2 transformed plate and were used to inoculate six different flasks containing fresh LB broth cultures. A single random colony was also picked from the pET14b transformed plate and was used to inoculate a flask containing fresh LB broth culture. Each of those flasks contained 50 ml of LB broth culture including ampicillin and chloramphenicol at concentrations of 50 μg/ml and 34 μg/ml respectively. Then those flasks were kept shaking at 37°C for 15 hours at 225 rpm. The next

day, all the seven flasks were observed for bacterial cell growth. As all had bacterial growth in them, 50% glycerol stocks were made for each of those seven cultures and were stored at -80°C in a freezer.

4.1.2.2. Overexpression of the C2 domain gene

From one of the six C2 glycerol stocks, a small quantity of C2 transformed cells was used to inoculate a fresh 100 ml culture of LB broth containing ampicillin and chloramphenicol at 50 μg/ml and 34 μg/ml respectively in a 250 ml Erlenmeyer flask. The flask was then kept shaking at 37°C for 10 hours at 225 rpm. Thereafter, 15 ml of this culture was used to inoculate a fresh 135 ml culture with required antibiotics contained in another 500 ml Erlenmeyer flask. The resulting 150 ml of culture was kept shaking at 37 °C and its growth was monitored by periodically measuring its absorbance at 600 nm. When the absorbance at 600 nm reached 0.5, overexpression of C2 was induced by adding IPTG to a final concentration of 0.4 mM to the culture. The shaking was then continued, and the bacteria were allowed to grow for another 3.5 hours. Samples of 1.5 ml volume from the culture were taken at 0 hr, 1.5 hr and 3.5 hr post induction times. The cell pellets of the above samples after centrifugation at 14,000 rpm were stored at -20°C for SDS-PAGE analysis.

A 12% SDS-PAGE was prepared as per the protocol B.8 in Appendix B. To each of the above post-induction samples, 50 μl of 2x Laemmli sample buffer (62.5 mM Tris-HCl, pH 6.8, 2% SDS, 25% glycerol, 0.01% Bromophenol Blue, 5% β-mercaptoethanol) was added and briefly vortexed. The samples were then heated on a hot plate at 95°C for 5 minutes. After that the supernatant from 0 hr post induction sample was loaded onto Lanes 1 and 2, while the supernatants from 1.5 hr and 3.5 hr post induction samples were loaded onto Lanes 4-6 and 8-10 respectively. 5 μl of protein mass ladder was also loaded onto each of the

Lanes 3 and 7. After loading, the gel plate assembly was clamped onto the main

electrophoresis apparatus and the gel tank filled with SDS run buffer (25 mM Tris, 192 mM

glycine, 0.1% SDS). The safety cover was then kept in place and the electrodes were

connected to a power supply. The gel was run at 200 V, 50 mA for about 50 minutes until the

dye front reached the bottom of the gel. Then the gel was taken out and stained in about 50

ml of Coomassie Blue R-250 staining solution (0.1% Coomassie Blue R-250, 40% methanol,

10% acetic acid) for about 10 hours. It was then followed by 3 hours of destaining in the

destaining solution (40% methanol, 10% acetic acid). The gel thus obtained was

photographed and was as shown in Figure 4.3.

Figure 4.3. SDS-PAGE indicating the overexpression of the C2 domain gene. Lanes 1 and 2:
0 hr post-induction; Lanes 3 and 7: mass ladder; Lanes 4, 5 and 6: 1.5 hr post-induction;
Lanes 8, 9 and 10: 3.5 hr post-induction.

Lanes 4-6 and 8-10 clearly show thick overexpressed protein bands slightly below the

25 kDa mark. This result clearly indicated that the 6xHis-tagged C2 domain was over

expressed. However, more tests were needed at this point to prove that it was expressed at the right size of 22.92 kDa.

4.1.2.3. Detection of the Histidine tag by dot blotting

Two 250 ml Erlenmeyer flasks each containing 100 ml of fresh LB broth with required antibiotics were inoculated separately with glycerol stocks of C2 and pET14b transformed cells and were kept shaking at 37°C for 12 hours. Then, 15 ml of each of the resulting cultures was used to inoculate two separate fresh LB broth cultures of 135 ml each with required antibiotics. Both the flasks were then kept shaking at 37°C and their cell growth was monitored as before by measuring the absorbance at 600 nm. When the absorbance reached 0.5, IPTG to a final concentration of 0.4 mM was added to each of the two flasks and the cells were allowed to grow for another 3.5 hours. Thereafter the cells for both C2 and pET14b in batches of 50 ml volumes were harvested by centrifugation at 10,000 rpm and the resulting cell pellets stored at -80°C.

To one of the cell pellets resulting from 50 ml of overexpressed C2, 250 μl of dilution buffer for denaturing conditions (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris, pH 8.0) was added and incubated at room temperature for 10 minutes with occasional stirring. The same was done for one of the cell pellets resulting from 50 ml of overexpressed pET14b. 2 μl each of the resulting pET14b solution (negative control) was applied at the four corners of a nitrocellulose membrane, while 2 μl each of the resulting C2 solution was applied in 12 dots towards its center. The membrane was then dried at room temperature for 10 minutes. Thereafter, the following protocol was used for the detection of 6xHis-tag.

(1) The membrane was washed twice for 10 min each time with TBS buffer (10 mM Tris-HCl, 150 mM NaCl, pH 7.5) at room temperature.

(2) Then the membrane was incubated for 1 hr in anti-His HRP conjugate blocking buffer (3% BSA in TBS buffer) at room temperature.

(3) The membrane was then washed twice for 10 min each time in TBS-Tween/ Triton buffer (20 mM Tris-HCl, 500 mM NaCl, 0.05% Tween20, 0.2% Triton x-100, pH 7.5) at room temperature.

(4) The membrane was again washed for 10 min with TBS buffer at room temperature.

(5) Then the membrane was incubated in Anti-His HRP conjugate solution (1/1000-1/2000 dilution of conjugate stock solution in blocking buffer) at room temperature for 1 hr.

(6) The membrane was then washed twice for 10 min each time in TBS-Tween/Triton buffer at room temperature.

(7) The membrane was washed again for 10 min in TBS buffer at room temperature.

(8) The membrane was then stained with HRP staining solution until the signal was clearly visible. HRP staining solution was prepared by dissolving 18 mg of 4-chloro-1-naphthol in 6 ml of methanol and then adding it to 24 ml of 1x Tris saline (10x Tris saline: 9% (w/v) NaCl in 1M Tris-HCl, pH 8.0) followed by further addition of 60 µl of H_2O_2.

(9) The chromogenic reaction was then stopped by rinsing the membrane twice with water.

(10) The membrane was then dried and photographed soon after.

The photograph of the membrane taken thereafter is as shown in Figure 4.4. Only the dots corresponding to C2 have been induced a color change. This result indicated the presence of 6xHis-tag in the over expressed protein.
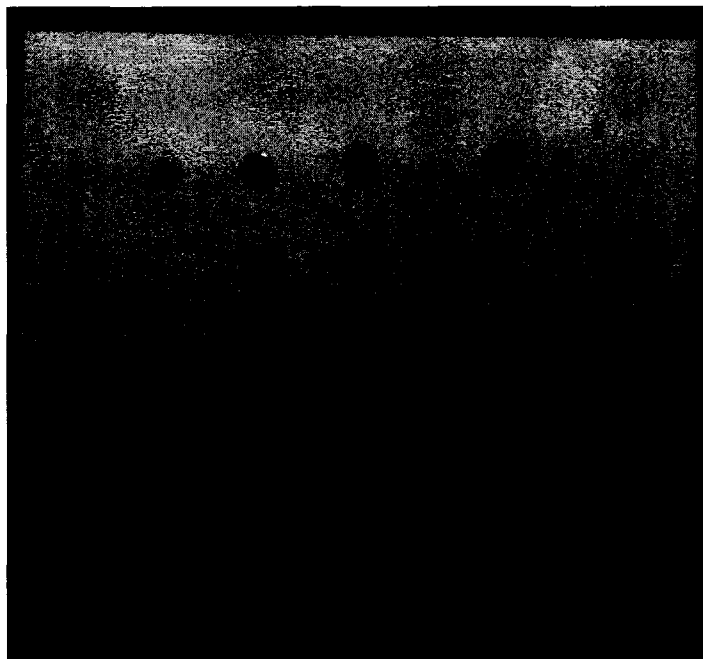
Figure 4.4. Nitrocellulose membrane after the dot blot procedure. The 12 dots corresponding to the over expressed histidine-tagged C2 domain yielded a color change. The 4 dots corresponding to the overexpressed pET14b (negative control) did not yield any color change.

4.2. Purification of the C2 Domain

4.2.1. Introduction

The subtopics within this section cover the theory and background information relating to the procedures involved in the purification of the C2 domain of human tensin.

4.2.1.1. Inclusion bodies

Recombinant proteins expressed in *E.coli* are often deposited in the form of insoluble aggregates called inclusion bodies. High levels of protein expression lead to an increase in the concentration of the nascent polypeptide chains which in turn induces the formation of inactive aggregates. Reducing the level of gene induction and fermentation of transformed cells under reduced growth temperature are generally done to increase the yield of native, soluble and active protein. The formation of inclusion bodies on the whole is determined by a

combination of factors such as rates of folding, aggregation, and protein synthesis, the solubility and thermodynamic stability of folding intermediate and the native states, the susceptibility to proteolytic degradation and the interaction with chaperones [97]. As such it is difficult to predict whether a protein would be expressed in soluble fraction or isolated in inclusion bodies. Though considered as a major obstacle in the production of active soluble proteins, inclusion body formation has certain advantages. First, the formation of inclusion bodies normally protects the gene product from host-cell proteases. Also, the product is inactive and cannot harm the host cell, often giving high expression levels. Furthermore, the dense inclusion bodies can be readily recovered by centrifugation and a relatively high purity and degree of concentration of the gene product are obtained after solubilization.

The process of solubilization aims at the formation of no or minimum non-native intra- or inter-chain interactions, a schematic of which is shown in Figure 4.5. The choice of solubilizing agents--- e.g., urea, guanidine hydrochloride, or detergents--- plays a key role in solubilization efficiency, in the structure of the protein in denatured state, and in subsequent refolding. Urea and guanidine hydrochloride lead to a flexible and disordered structure of proteins. Typically, 6-8 M urea and 6-7 M guanidine hydrochloride are required to completely unfold and solubilize the proteins. Even at the highest concentration of the denaturant, intra- and inter-molecular interactions can occur. Such a non-native structure can lead to aggregation or misfolding upon the removal of denaturant.
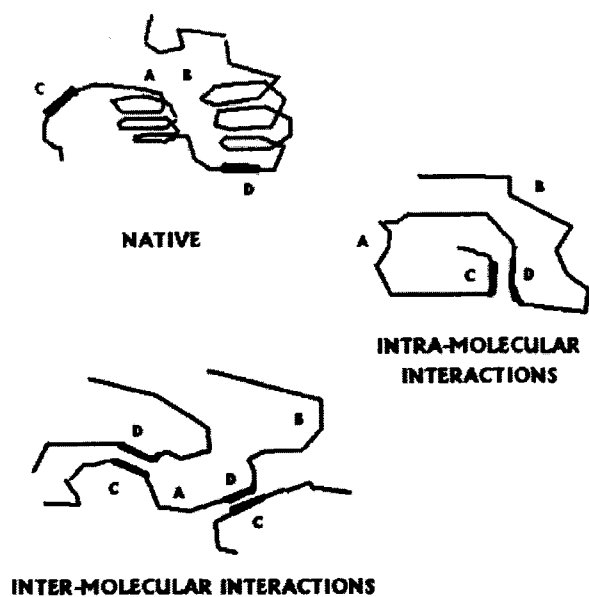
Figure 4.5. Non-native intra- and inter-molecular interactions [97].

4.2.1.2. Protein purification pathway

A typical protein purification process follows the pathway as depicted in Figure 4.6. The pathway taken for the purification process of the 6xHis-tagged C2 domain from *E.coli* is shown in the same figure involving only those processes marked with dots.
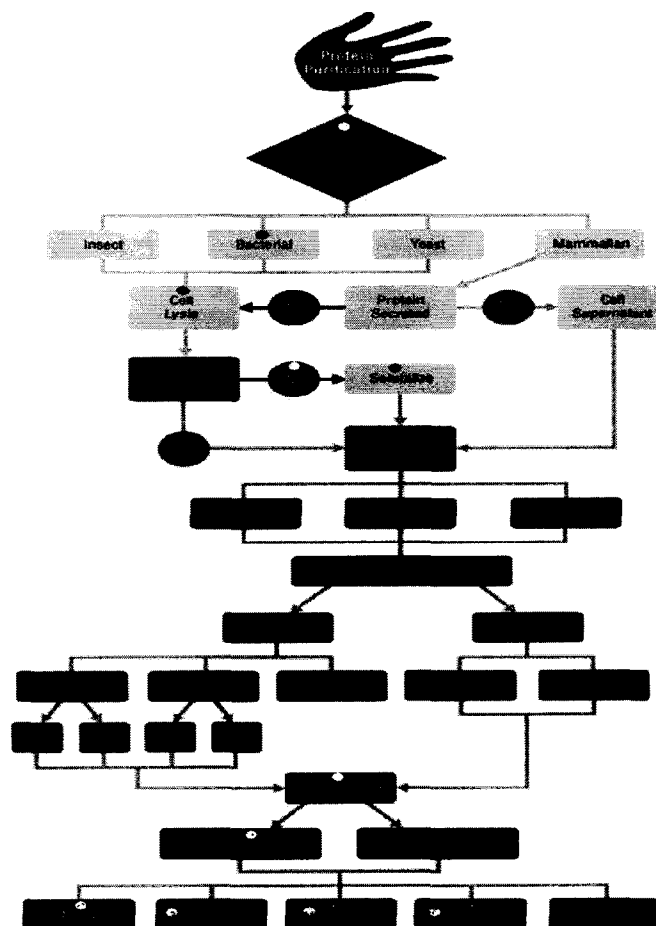
Figure 4.6. Protein purification pathway. Copyright: Pierce Biotechnology, Inc.

4.2.1.3. Immobilized metal affinity chromatography (IMAC)

IMAC is one of several different chromatographic techniques used in the purification of proteins. In this type of chromatography, immobilized ions such as Cu^{+2}, Zn^{+2}, Hg^{+2} or Cd^{+2} or transition metal ions such as Co^{+2}, Ni^{+2}, or Mn^{+2} are used to bind proteins selectively by reaction with the imidazole groups of histidine residues, thiol groups of cysteine residues and indole groups of tryptophan residues. The immobilization of protein involves the formation of coordinate bond that must be sufficiently stable to allow protein attachment and retention during the elution of non-binding contaminating material. The subsequent release of

the protein can be achieved either by simply lowering the pH, thereby destabilizing the protein-metal complex, or by the use of complexing agents such as EDTA [95].

For the purification of the 6xHis-tagged C2 domain, Ni-NTA technology has been used. In this, nitriloacetic acid (NTA) bound to the chromatographic matrix acts as a chelating ligand. It occupies four of the six ligand binding sites in the coordination sphere of the nickel ion. The other two sites are free to interact with the 6xHis tag as seen in Figure 4.7. The 6xHis tag binds strongly to the immobilized nickel ion. For optimal performance, both the 6xHis tag-nickel and the nickel-NTA interactions are important [98].
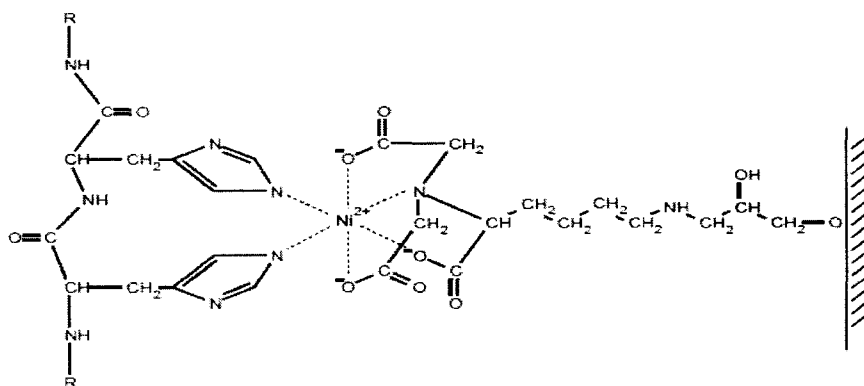
Figure 4.7. Interaction between neighboring residues in the 6xHis tag and Ni-NTA matrix [98].

Before an IMAC procedure is attempted, it is important to determine whether the recombinant protein is expressed in soluble form or in insoluble aggregates called 'inclusion bodies'. The interaction between Ni-NTA and the 6xHis-tag of the recombinant protein does not depend on tertiary structure, and hence proteins can be purified either under native or denaturing conditions. If the protein is found soluble, then there is a higher potential for binding of background contaminants than under denaturing conditions. The imidazole ring is a part of the structure of histidine as shown in Figure 4.8, and the imidazole rings in histidine

residues of the 6xHis-tag bind to the nickel ions immobilized by the NTA groups on the matrix. Imidazole itself can bind to the nickel ions and disrupt the binding of dispersed histidine residues in nontagged background proteins. At low imidazole concentrations, nonspecific, low-affinity binding of background proteins is prevented, while 6xHis-tagged proteins still bind strongly to the Ni-NTA matrix. Therefore, adding imidazole at low concentrations of 10-20 mM to the lysis and wash buffers lead to greater purity in fewer steps [98].
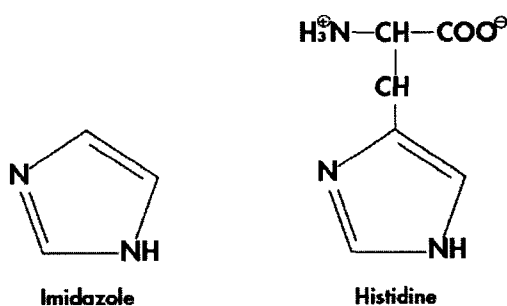
Figure 4.8. Chemical structures of imidazole and histidine [98].

Copurification of host proteins that may have formed disulphide bonds with the protein of interest during cell lysis can be prevented by adding low levels of β-mercaptoethanol (up to 20 mM) in the lysis buffer. Detergents such as Triton X-100 and Tween20 (up to 2%), or high salt concentrations (up to 2 M NaCl) can also be added to reduce nonspecific binding to the matrix caused by hydrophobic or ionic interactions. If the protein is found insoluble, then strong denaturants such as 6 M GuHCl or 8 M urea are added to completely solubilize inclusion bodies and 6xHis-tagged proteins. Under these denaturing conditions, the 6xHis-tag on the protein will be fully exposed so that it's binding to the Ni-NTA matrix will improve, and thereby the efficiency of the purification procedure will be

maximized by reducing the potential for nonspecific binding. The histidine residues in the 6xHis-tag have a pK_a of approximately 6.0 and will become protonated, if the pH is reduced to the range 4.5-5.3. Under these conditions, the 6xHis-tagged protein can no longer bind to the nickel ions and will dissociate from the Ni-NTA resin. Similarly, if the imidazole concentration is increased to 500 mM, the 6xHis-tagged proteins will also dissociate because they can no longer compete for the binding sites on the Ni-NTA resin. Also, reagents such as EDTA, chelate nickel ions and remove them from the NTA groups. This causes the 6xHis-tagged protein to elute as a protein-metal complex [98].

4.2.1.4. Cleavage of the Histidine tag

Cleavage of the fusion tag from recombinant protein is critical if the produced protein is intended to be used as a therapeutic product or if the protein is aimed to be used for structural determination by X-ray crystallography. In most cases, there is hardly any need for the affinity tag to be removed. Thrombin is an endoprotease which converts fibrinogen to fibrin and hence functions as a blood clotting factor. It is also one of the most active site-specific proteases, and a very low mass ratio of enzyme-to-target protein is needed for efficient cleavage. The cleavage frequently occurs after the proline-arginine residue pair when properly exposed in a three-dimensional structure [98]. For the cleavage of 6xHis-tag from the C2 domain, restriction grade thrombin from Novagen was used. It was functionally tested for the specific cleavage of the sequence LeuValProArg↓GlySer found in other recombinant proteins.

4.2.2. Methods, Results and Discussion

The subtopics within this section deal with the experimental methods related to the purification of the C2 domain including their results and discussion.

4.2.2.1. Inclusion body test

The 6xHis-tagged C2 domain was over expressed in a 500 ml Erlenmeyer flask to a final volume of 250 ml similar to the procedure as described in Section 4.1.2.2. The resulting culture was taken into five 50 ml centrifuge tubes and was centrifuged at 10,000 rpm for 15 minutes. To the cell pellets resulting from two of the tubes, 4 ml of cold isolation buffer was added to each of them and vortexed. Then they were sonicated on ice for 2 minutes. After that both the tubes were centrifuged at 10,000 rpm for 15 minutes. The resulting supernatant and cell pellet were analyzed on a 12% SDS-PAGE gel as did before. The gel thus obtained is as shown in Figure 4.9. From the gel picture, high intensity bands in Lane 2 confirm that the 6xHis-tagged C2 domain was expressed mostly in inclusion bodies.
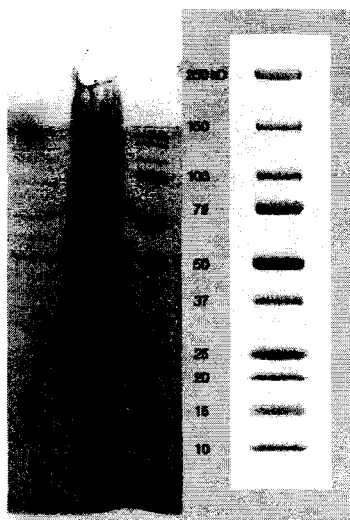
Figure 4.9. Polyacrylamide gel indicating the C2 inclusion body test results. Lane 1: supernatant, Lane 2: cell pellet and Lane 3: mass ladder.

4.2.2.2. Affinity chromatography
of the C2 domain

After identifying that the 6xHis-tagged C2 domain was being expressed in the *E.coli* inclusion bodies, a suitable protocol involving their solubilization and purification was adopted as mentioned below. Initially 6xHis-tagged C2 was overexpressed in an 1 liter Erlenmeyer flask to a final volume of 500ml as before. Samples at post induction times of 0 hr, 1 hr, 2.5 hr and 3.5 hr were taken out for SDS-PAGE analysis. After 3.5 hr, cells were harvested by taking the entire culture volume into four 50 ml centrifuge tubes several times and centrifuging them at 7,000 rpm. Next, to each of the cell pellets resulting from 125 ml of culture, the following protocol was adopted.

(1) The cell paste was resuspended in 5 ml of resuspension buffer (20 mM Tris-HCl, pH 8.0) along with 50 µl of protease inhibitor cocktail mix.

(2) The cells were then disrupted with sonication on ice at 4 x 10 sec.

(3) Then they were centrifuged at 12,000 rpm for 10 min at 4°C.

(4) The resulting cell pellet was resuspended in 4 ml of cold isolation buffer (20 mM Tris-HCl, 2 M urea, 0.5 M NaCl, 2% Triton X-100, pH 8.0) and was sonicated as above.

(5) Then they were centrifuged at 12,000 rpm for 10 min at 4°C.

(6) Steps 4 and 5 were repeated.

(7) To the resulting cell pellets from all the tubes, 10 ml of solubilization buffer (20 mM Tris-HCl, 6 M GuHCl, 0.5 M NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol, pH 8.0) was added along with 200 µl of DNaseI (10 µg/ml) and was stirred for 30-60 min. on a rocker platform at room temperature.

(8) Then they were centrifuged for 15 min at 12,000 rpm at 4°C.

(9) The particulate matter was removed by passing the sample through a 0.45 μm filter.

(10) The sample was degassed for 5 minutes.

Next, a 5 ml His-Trap HP column was connected to the AKTA prime instrument and was then equilibrated with 50 ml of solubilization buffer. The sample obtained from step 10 above was then loaded onto the column. After loading the sample, 50 ml of solubilization buffer was passed through the column again to get a steady baseline in the absorbance reading of the instrument at 280 nm. Thereafter, the instrument was auto-zeroed and 50 ml of refolding buffer (20 mM Tris-HCl, 6 M urea, 0.5 M NaCl, 5 mM imidazole and 1 mM β-mercaptoethanol, pH 8.0) was passed through the column. Then, an on-column refolding of the bound C2 domain was attempted by making use of a linear 6-0 M urea gradient, starting with the refolding buffer above and finishing with the wash buffer. A gradient volume of 100 ml and a flow rate of 1 ml/min were used. After the gradient has reached its end point, 30 ml of wash buffer (20 mM Tris-HCl, 0.5 M NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol, pH 8.0) was continued to pass through the column. Next, the elution of the bound C2 domain was attempted by using a 50 ml linear gradient at a flow rate of 2 ml/min starting with wash buffer and ending with elution buffer (20 mM Tris-HCl, 0.5 M NaCl, 0.5 M imidazole, 1 mM β-mercaptoethanol, pH 8.0). At the start of the elution gradient, a sharp peak was seen on the chart recorder paper corresponding to an absorbance of 1.0 at 280 nm. During this time, 20 ml of the eluted protein was collected in a 50 ml centrifuge tube. This eluted protein was then dialyzed twice against 4 liters of 25 mM sodium phosphate, 10 mM NaCl, 1 mM β-mercaptoethanol, pH 7.8. Thereafter 20 μl of this dialyzed protein was analyzed on a 12% SDS-PAGE gel alongside post-induction samples. The gel was then run at 200 V at 50 mA for 45 minutes. After staining and destaining procedures, the gel obtained is as shown in

Figure 4.10. Lanes 6 and 7 of the gel clearly show good intensity bands of purified C2 domain at around 23 kDa size. This result indicates that the purification of the C2 domain has been done successfully.

Figure 4.10. Polyacrylamide gel confirming the purification of the C2 domain. Lanes 1, 2, 3 and 4: 0 hr, 1 hr, 2.5 hr and 3.5 hr post induction respectively; Lanes 5: mass ladder; and Lanes 6 and 7: purified 6xHis-tagged C2 domain.

4.2.2.3. Cleavage of the Histidine tag of the C2 domain

The remaining dialyzed C2 protein from the above experiment was taken into five 5 ml vials and was freeze dried using a lyophilizer. To the freeze dried C2, DI water was added to get a final concentration of 1 µg/µl. Restriction grade thrombin was used to cleavage the 6x His tag. Five different thrombin cleavage reactions each of 50 µl total volume were setup in tubes 1, 2, 3, 4 and 5 as shown in Table 4.1.

Table 4.1. Thrombin cleavage reaction components for the His-tagged C2 domain.

	Tube-1	Tube-2	Tube-3	Tube-4	Tube-5
10x Thrombin cleavage buffer	5 μl	5 μl	5 μl	5 μl	5 μl
1:25 diluted thrombin	1 μl	2 μl	3 μl	4 μl	2 μl
DI water	41 μl	40 μl	39 μl	38 μl	41 μl
C2 domain	3 μl (1 μg/μl)	3 μl (1 μg/μl)	3 μl (1 μg/μl)	3 μl (1 μg/μl)	-
Control protein	-	-	-	-	2 μl (1 μg/μl)

Tube 5 was a positive control in which a 48 kDa control protein was used. Successful cleavage of this protein with thrombin would yield two protein fragments at sizes 35 kDa and 13 kDa. All the above tubes were then incubated at room temperature for 24 hours. Then to each of the tubes 1, 2, 3 and 4, 2 μl of C2 was added and were analyzed by running them on a 12% SDS-PAGE as before. The gel obtained after the staining and destaining procedures is as shown in Figure 4.11.

Figure 4.11. Polyacrylamide gel showing the cleavage of 6xHis-tag from the recombinant C2 protein using variable amounts of thrombin. Lanes 1 & 2, Lanes 3 & 4, Lanes 6 & 7, and Lanes 8 & 9: thrombin cleavage of C2 with 1 μl, 2 μl, 3 μl and 4 μl of 1:25 diluted thrombin; Lane 5: mass ladder and Lane 10: thrombin cleavage of the control protein with 3 μl of 1:25 diluted thrombin.

Lanes 1-4 and 6-9 clearly indicate thrombin cleavage yielding a cleaved C2 at 20.5 kDa along with the uncleaved C2 at 22.9 kDa. One of the cleaved fragments of the control protein at 35kDa can also be seen in the positive control Lane 10. Hence from this experiment, we can conclude that the purified 6xHis-tagged C2 protein has an accessible histidine tag that can be successfully cleaved by thrombin to yield the desired C2 domain.

CHAPTER 5

REFOLDING AND BIOPHYSICAL
CHARACTERIZATION
OF THE C2 DOMAIN

5.1. Introduction

This section contains subtopics covering the theory and background information

relating to the procedures involved in the refolding and biophysical characterization of

the C2 domain of human tensin.

5.1.1. Protein Refolding

Refolding is a process that leads to conformation change from unfolded (inactive)

state to folded (active) state. At high denaturant concentrations, proteins are unfolded,

well solvated, and flexible. But in aqueous buffers, they are folded, rigid and compact.

Generally, transfer of protein molecules from denaturant solution to aqueous solvent will

force them to collapse into a compact structure. But such a drastic process usually leads

to misfolding and aggregation. Once misfolded or aggregated, in the absence of

denaturants, proteins have no flexibility to disaggregate and refold into the native

structure. A key to refolding is in the intermediate concentration of the denaturant of the

denaturant, where denaturant is low enough to force protein molecules to collapse, yet

can allow them to stay in solution and be flexible to reorganize their structures. However,

which intermediate concentration works depends on the proteins and how the denaturant

concentration is reduced. Intermediates have structures that are unstable and less soluble.

90

As such they readily misfold and aggregate. It is therefore important to facilitate folding of the intermediate into the more stable native structure (N), yet maintain the solubility and flexibility of the intermediate (See Figure 5.1).
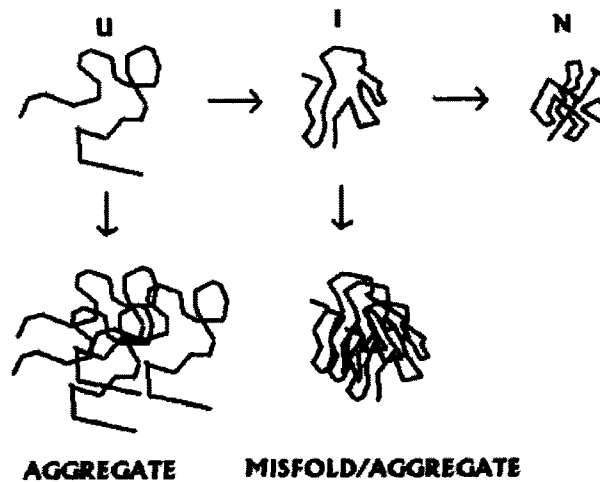
Figure 5.1. Schematic diagram of refolding course. U, I, and N correspond to the unfolded, intermediate, and native state of protein, respectively [97].

Optimal refolding can be achieved by forcing the denatured protein to collapse but still maintain its solubility and flexibility, and this is done by the way by which denaturant concentration is reduced. How fast the denaturant is reduced and how long the protein molecules are exposed to intermediate denaturant concentration determines the rate of folding and the degree of flexibility or solubility of folding intermediates. In addition, the rate of folding vs. misfolding or aggregation can be manipulated by small molecule additives. Certain additives enhance structure formation or collapse, while others increase flexibility or solubility of proteins.

5.1.1.1. Dialysis

In one-step dialysis, the denatured and unfolded protein molecules in the concentrated denaturant solution are dialyzed against a refolding buffer containing much

less or no denaturant. As such, the protein solution is exposed to a descending concentration of the denaturant and the denaturant concentration within the dialysis bag is decreased with time to the concentration of the refolding buffer. As the concentration of denaturant is decreased, the rate of folding into the intermediate and native structure increases. However, the rate of misfolding or aggregation also increases. There is also a high chance of aggregation, if the protein concentration is high in the dialysis tubing. This aggregation is due to the reasons that the protein concentration remains relatively constant within the tubing through the process and also as the denaturant concentration is slowly decreased, the interaction of hydrophobic patches is quite fast due to the high concentration of protein. Hence the initial protein concentration in the denaturant is critical. Also, if the rate of folding is slow, there is a greater chance for aggregation because the moderate to low concentration may not be enough to keep the unfolded or intermediate structures soluble.

In step-wise dialysis, the unfolded protein sample is first brought to equilibrium with high denaturant concentration, then with middle concentration, and with low concentration. In contrast to one-step dialysis, here the equilibrium is established at each denaturant concentration.

5.1.1.2. Dilution

In dilution, protein samples at high denaturant concentration are delivered into a large volume of refolding buffer. This process induces a rapid collapse in the unfolded sample bypassing the intermediate denaturant concentration. Here, both the denaturant and protein concentrations increase as the unfolded protein in concentrated denaturant is delivered into the refolding buffer.

In pulsed dilution, after an aliquot of the denatured protein solution is diluted into the refolding buffer, refolding is allowed to occur for some period before adding the next aliquot. As such this avoids the accumulation of high concentration of folding intermediates that occur in one-step dilution.

5.1.1.3. Role of co-solute assistance

Co-solutes are classified into folding enhancers and aggregation suppressors. Folding enhancers enhance protein-protein interactions, while aggregation suppressors reduce side chain interactions and association of folding intermediates without interfering with refolding process. Polyethylene glycol, cyclodextran, L-arginine, and proline are a few examples of aggregation suppressors. L-arginine is the most frequently used co-solute, but how it reduces aggregation of folding intermediates is unclear. However, studies by Reddy et al. demonstrated that L-arginine seems to increase the equilibrium solubility of unfolded and intermediate states of hen egg white lysozyme during refolding [99]. The use of L-arginine and pulsed dilution has been moderately successful in the refolding experiments of the C2 domain as described in Section 5.2.2.

5.1.2. Circular Dichroism (CD) Spectropolarimetry

Circular dichroism is mainly used for studying changes in the secondary and tertiary structure of proteins. There are two requirements for a molecule or a group of atoms in a molecule to exhibit a circular dichroism (CD) spectrum. The first is the presence of a chromophore i.e., a group that can absorb radiation by virtue of the electronic configuration of its ground state at room temperature. The energy absorbed results in a transition to a higher-energy or excited state, which has a different distribution of electrons around the nucleus. It can therefore interact with its environment

in a way that differs from the ground state. In proteins, Tryptophan, Tyrosine and Phenyl alanine are the main chromophores in the near-UV (240- to 320-nm), while the peptide bond is the main chromophore in the far-UV (180- to 240-nm). The second requirement for a CD is that the chromophore be present in an optically asymmetric environment [100].

Plane polarized radiation consists of two circularly polarized vectors of equal intensity, one right handed and the other left handed as shown in Figure 5.2, which are separately measured in the CD spectrometer by a photoelastic modulator. A chromophore situated in an optically symmetric environment will normally absorb the two components equally so that, when recombined after passing through a solution of the chromophore, they result once again in radiation oscillating in a single plane. A chromophore situated in an optically asymmetric environment, however, will absorb each of the two components to a different extent, the difference being ΔA. The resultant radiation would now be elliptically polarized i.e., the resultant would trace out an ellipse. The dichroism at a given wavelength of radiation is expressed as either the difference in absorbance of the two components ($\Delta A = A_L - A_R$) or as the ellipticity in degrees (θ). ($\theta = \tan^{-1}(b/a)$, where 'b' and 'a' are the minor and major axes of the resultant ellipse.
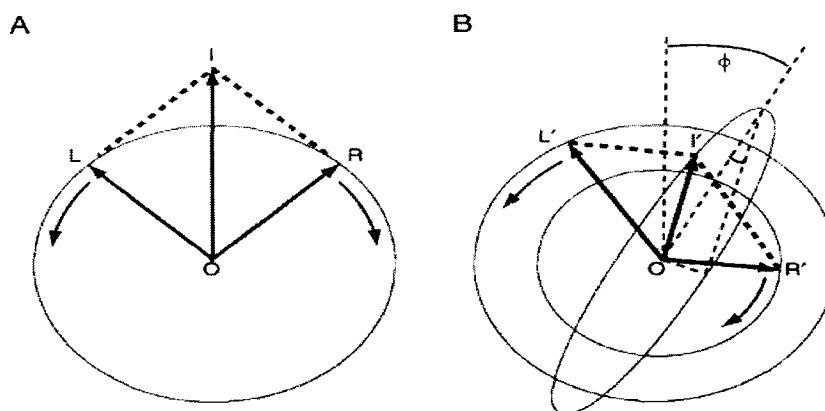
Figure 5.2. The relation of ellipticity to the differential absorption of circularly polarized radiation. (A) Plane-polarized radiation, made up of left- and right-circularly polarized components, **OL** and **OR** respectively (B) Interaction of radiation with a chiral chromophore leads to unequal absorption causing the emerging vectors **OL'** and **OR'** to describe an elliptical path [101].

In the far-UV, the peptide bond is the principal absorbing group and studies in this region can give information on the secondary structure. In the near-UV, the aromatic amino acid side chains (phenyl alanine, tyrosine and tryptophan) absorb in the range of 250-290 nm. The tertiary folding of the polypeptide chain can place these side chains in chiral environments, thus giving rise to CD spectra which can serve as characteristic fingerprints of the native structure.

5.1.3. Differential Scanning Calorimetry (DSC)

A protein in aqueous solution is in equilibrium between the native (folded) conformation and its denatured (unfolded) conformation, therefore, $K_{unf}=$ [U]/[N]. where [U] is the concentration of unfolded protein, [N] is concentration of native protein, and K_{unf} is the equilibrium constant between native and denatured protein. The stability of the native state is based on the magnitude of Gibbs free energy (ΔG) in the thermodynamic relationships $\Delta G=\Delta H-T\Delta S$ and $\Delta G=-RTlnK_{unf}$. Where ΔH is the enthalpy change and ΔS

is the entropy change. A positive ΔG indicates the native state is more stable than the unfolded state. The higher the ΔG, the greater the stability. Protein stability depends on the balance between enthalpic changes (due to changes in hydrophobic interactions and hydrogen bonding) and entropic changes (due to solvation and conformational freedom). For a protein to unfold, stabilizing forces need to be broken. Conformational entropy weakens stabilizing forces allowing the protein to unfold at high temperature where the $-T\Delta S$ term becomes dominant [101].

A DSC measures the ΔH of protein folding due to heat denaturation. In a DSC instrument, a protein solution and a reference solution are in matched cells, and heated at a constant rate. The power input into the sample solution is compared to the power required to raise the temperature of the reference solution. If the protein in the sample cell undergoes thermal unfolding from a low-temperature native state to a high-temperature denatured state, this process is prominently seen in the resulting plot of differential power vs. temperature.

The unfolding transition is recognized as a sharp endothermic peak centered at a characteristic temperature called the transition midpoint (Tm). By integrating the area under the peak, the heat of unfolding (ΔH) can be determined. T_m is an indicator of thermostability and in general, the higher the T_m, the more stable the protein. DSC also measures the change in heat capacity (ΔC_p) of a protein during denaturation. The ΔC_p for protein unfolding is almost positive, because a denatured protein has a higher heat capacity than a native protein. Heat capacity changes associated with protein unfolding are primarily due to changes in hydration of side chains that were buried in the native state, which become solvent-exposed in the denatured state.

5.2. Methods, Results and Discussion

The subtopics within this section deal with the experimental methods related to the refolding procedures of the C2 domain, its biophysical characterization along with their results and discussion.

5.2.1. Protein Preparation for Biophysical Studies

About 20 ml of purified C2 protein fractions obtained from one of the affinity chromatography procedures (described in Section 4.2.2.2) was taken into a dialysis tubing of 10 kDa cutoff. The protein was dialyzed for 4 hours against 4 liters of 25 mM sodium phosphate, 10 mM NaCl, 1 mM β-mercaptoethanol, pH 7.8 contained in a glass beaker. After 4 hours, buffer was changed with a freshly prepared one and the dialysis was continued overnight. After dialysis, the protein was recovered from the dialysis tubing and was concentrated using 10 kDa cutoff centricon filters. The concentration of the C2 protein thus obtained was then measured to be 1.062 mg/ml by measuring the absorbance at 280 nm.

Next, 3.5 ml of the phosphate buffer from the centricon filtrates, resulting from the above procedure, was taken into a 1 cm path length cuvette and its baseline CD spectrum was measured in the near-UV range of 250 nm-360 nm. 120 scans were performed. Thereafter, 3.5 ml of the C2 protein was taken into the cuvette, and its near-UV CD spectrum was measured in the same wavelength range with the same number of scans. The spectrum thus obtained after baseline subtraction is as shown in Figure 5.3.
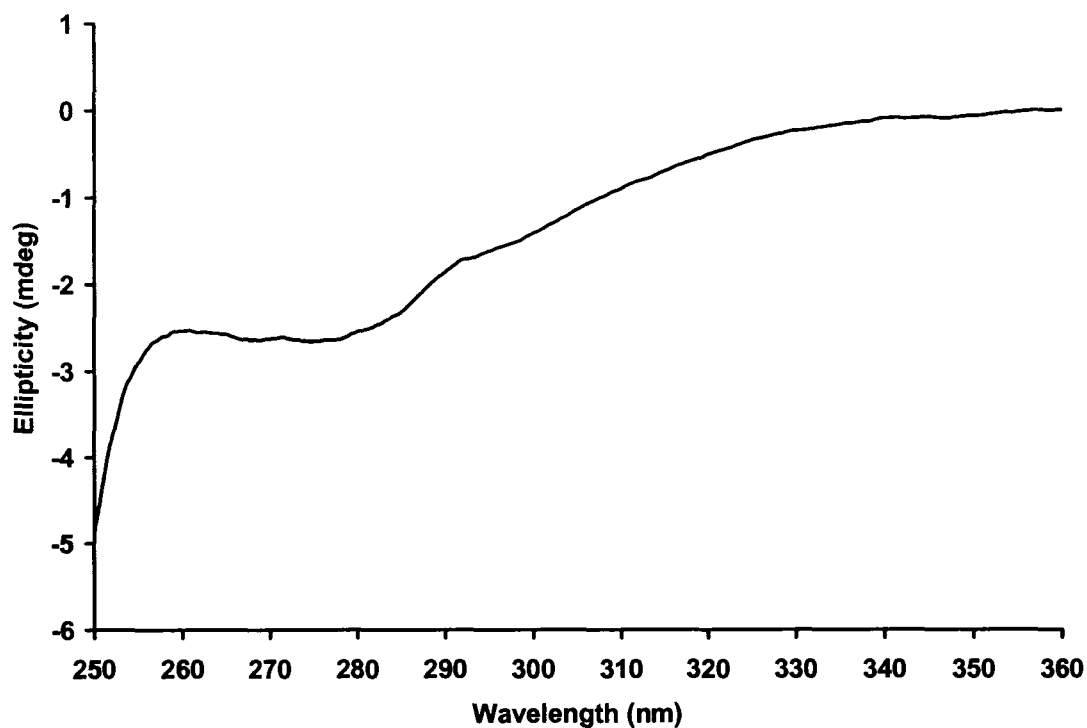
Figure 5.3. Near-UV CD spectrum of the C2 domain in 25mM phosphate buffer, pH 7.8.

The near-UV spectrum of the C2 domain as depicted in Figure 5.3 has no aromatic asymmetry signifying any amount of tertiary structure. However, the protein was used to do some far-UV experiments as follows.

Next 300 ml of 250 mM sodium phosphate, 100 mM NaCl, 10 mM β-mercaptoethanol, pH 7.8 (10x buffer) was prepared to be used for guanidine hydrochloride (GuHCl) CD titration experiments in the far-UV. Samples for the far-UV GuHCl CD titration points of 0 M, 0.5 M and 1 M were prepared by mixing the respective components as depicted in Table 5.1 and were incubated for 1 hour. The corresponding baseline samples were also prepared replacing 30 µl of protein with the buffer.

Table 5.1. Sample compositions for the far-UV GuHCl titration points of 0 M, 0.5 M and 1 M in 25 mM sodium phosphate buffer, pH 7.8.

	0 M	0.5 M	1 M
10x buffer	30 µl	30 µl	30 µl
C2 protein	50 µl	50 µl	50 µl
8 M GuHCl	0 µl	18.75 µl	37.5 µl
DI-H₂O	220 µl	201.25 µl	182.5 µl
Total	300 µl	300 µl	300 µl

It was observed that the C2 protein got precipitated in the 0.5 M and 1 M vials soon after their preparation. However, for each of the above three points, CD spectra were measured, both for the baseline and C2 samples in the far-UV wavelength range of 190 nm-260 nm. A 1-mm path length quartz cuvette was used for the same. The far-UV baseline subtracted CD spectra of each of these points are overlaid and are as shown in Figure 5.4.
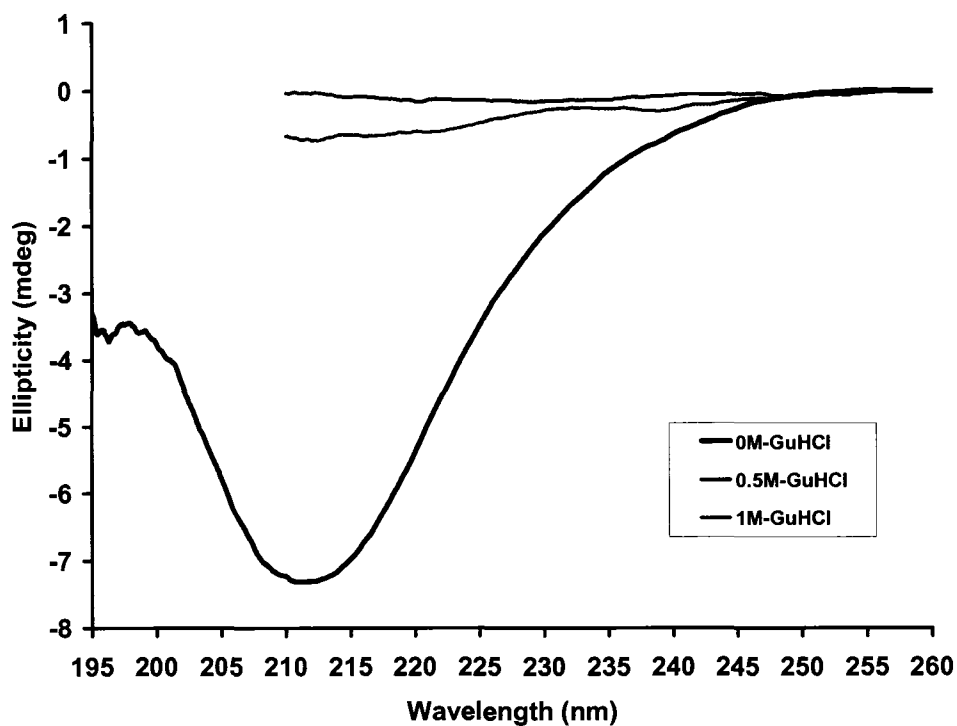
Figure 5.4. Far-UV CD spectra of the C2 domain in 25 mM phosphate buffer, pH 7.8 for 0 M, 0.5 M and 1 M GuHCl titration points.

It can be seen that for the 0.5 M and 1 M titration points, the CD spectra yielded near zero signals; the reason being the C2 protein getting precipitated out of the solution from these samples. These results for the C2 domain along with its near-UV CD spectrum in Figure 5.5 indicate that the protein may be partly folded with very low tertiary structure. Another reason assumed for the protein precipitation might be the experimental pH of 7.8 which is close to the isoelectric point of 6.29. The isoelectric profile for the C2 domain for both the tagged and untagged versions is as shown in Figure 5.5.
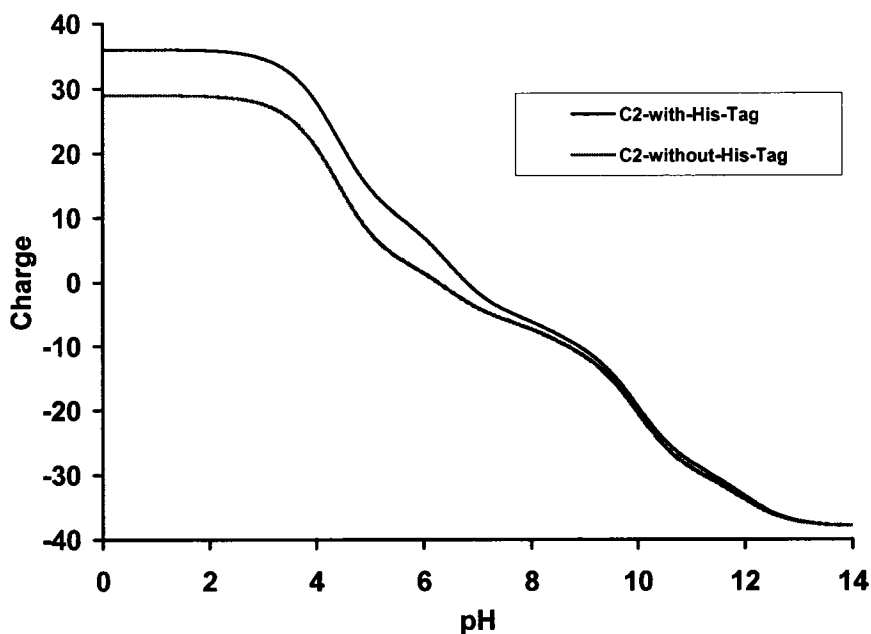
Figure 5.5. Isoelectric profile of the C2 domain with or without histidine tag. The histidine tagged version has a pI of 6.29, while that of the other has 5.81.

In an attempt to prevent the precipitation an alternative buffer, i.e., a carbonate buffer with a pH of 9.6 was sought. 300 ml of 250 mM sodium carbonate, 100 mM NaCl, 10 mM β-mercaptoethanol, pH 9.6 (10x buffer) was prepared to be used for the far-UV GuHCl CD titration experiments. Samples for the far-UV CD GuHCl titration points of 0 M and 0.5 M were prepared just as before as per the Table 5.1 and were incubated for an hour. This time also it was observed that the protein got precipitated in the 0.5 M sample. But still the spectra for the 0 M and 0.5 M samples were measured as before. The resulting spectra are overlaid and are as shown in Figure 5.6.
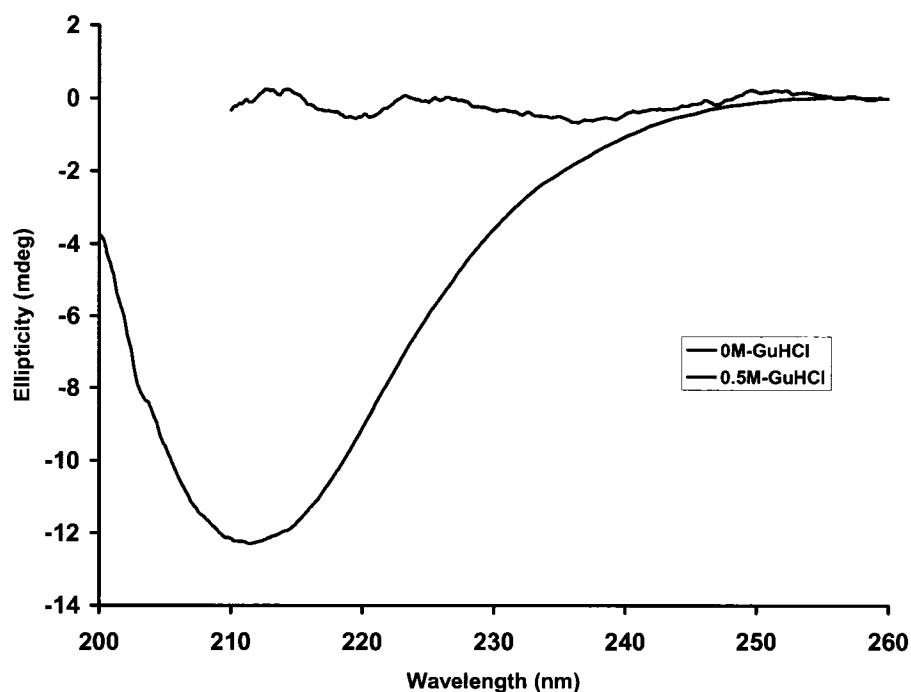
Figure 5.6. Far-UV CD spectra of the C2 domain in 25 mM carbonate buffer, pH 9.6 for 0 M and 0.5 M GuHCl titration points.

From Figure 5.6, it can be seen that the 0.5 M sample yielded a near zero signal. Hence it was concluded that the cause for the precipitation is not the pH of the buffer, but rather the protein itself with regard to its stability and folded state after the on-column refolding procedure. As such various refolding procedures for the C2 domain were sought as described in Section 5.2.2.

5.2.2. Refolding Trials for the C2 Domain

As the first method to refold the C2 domain, one-step dialysis was sought. In this method, 3 ml of the C2 protein was completely denatured in 6 M GuHCl. Then 500 ml of 25 mM sodium carbonate, 10 mM NaCl, 50 mM L-arginine, 1 mM β-mercaptoethanol, pH 9.6 was prepared. The denatured C2 protein was then dialyzed against the above prepared buffer for 4 hours. After 4 hours, it was noticed that the protein got precipitated

in the dialysis bag. However, the buffer was replaced by a freshly prepared one and the dialysis was continued overnight. In the morning, it was noticed that the protein was still in precipitated form. Hence an alternative method of step-wise dialysis was sought as described next.

Here in this method, the C2 protein from the above procedure was denatured completely in 6 M GuHCl. Then 500 ml of 25 mM Tris, 4 M GuHCl, 100 mM L-arginine, 5 mM β-mercaptoethanol, pH 8.0 was prepared. The denatured protein was taken into dialysis tubing and was dialyzed against the above prepared buffer for about 4 hours. After 4 hours, the buffer in the beaker was replaced by 500 ml of 25 mM Tris, 2 M GuHCl, 100 mM L-arginine, 5 mM β-mercaptoethanol, pH 8.0. The dialysis was continued overnight. Next morning, the buffer in the beaker was replaced by 500 ml of 25 mM Tris, 1 M GuHCl, 100 mM L-arginine, 5 mM β-mercaptoethanol, pH 8.0 and the dialysis was continued for 4 hours. After 4 hours, the buffer in the beaker was replaced by 500ml of 25 mM Tris, 100 mM L-arginine, 5 mM β-mercaptoethanol, pH 8.0. The dialysis was then continued. As soon as the dialysis was started, the C2 protein within the dialysis tubing began to precipitate out of the solution. However, the dialysis was continued for another 4 hours. After 4 hours, the precipitated protein from the dialysis tubing was taken out and stored at +4°C. From the above observations, it was evident that though the refolding process was slow when the concentration of GuHCl was decreasing slowly, the interaction between hydrophobic patches among the protein molecules was quite fast, due to the relatively high concentration of the protein, thus leading to aggregation. As such, an alternative refolding method of pulsed dilution was sought as described next.

In this method, the C2 protein from the above procedures was solubilized and denatured in 6 M GuHCl. To this solution, L-arginine to a final concentration of 100 mM was also added. Thereafter, the protein was concentrated by 10 kDa cut-off centricon filters to a final volume of about 1 ml. 25 ml of 25 mM sodium carbonate, 500 mM L-arginine, 10 mM NaCl, 1 mM β-mercaptoethanol, pH 10.0 was prepared in a glass beaker of 120 ml volume. To this buffer, 1 ml of the concentrated C2 protein was added drop wise slowly with rapid stirring. Thereafter, the diluted protein solution was taken into two dialysis tubings and was dialyzed overnight against 3 liters of 25 mM sodium carbonate buffer, 10 mM NaCl, 1 mM β-mercaptoethanol, pH 9.6. Next morning the protein from the dialysis tubings was taken out and was concentrated using 10 kDa cut-off centricon filters. After concentrating the protein to a volume of about 4 ml, its concentration was estimated to be 1.063 mg/ml.

5.2.3. Near-UV CD Measurement of the C2 Domain after Refolding

Next, about 3.5 ml of buffer from the filtrate of the centricon filters above was taken in a quartz cuvette of 1 cm path length and its near-UV baseline CD spectrum was measured in the wavelength range of 250 nm-360 nm. In all, 150 scans were performed. After that, 3.5 ml of the concentrated C2 protein was taken into the cuvette and its near-UV CD spectrum was measured in the same wavelength range with the same number of scans. The spectrum thus obtained after baseline subtraction and normalizing for protein concentration is as shown in Figure 5.7.
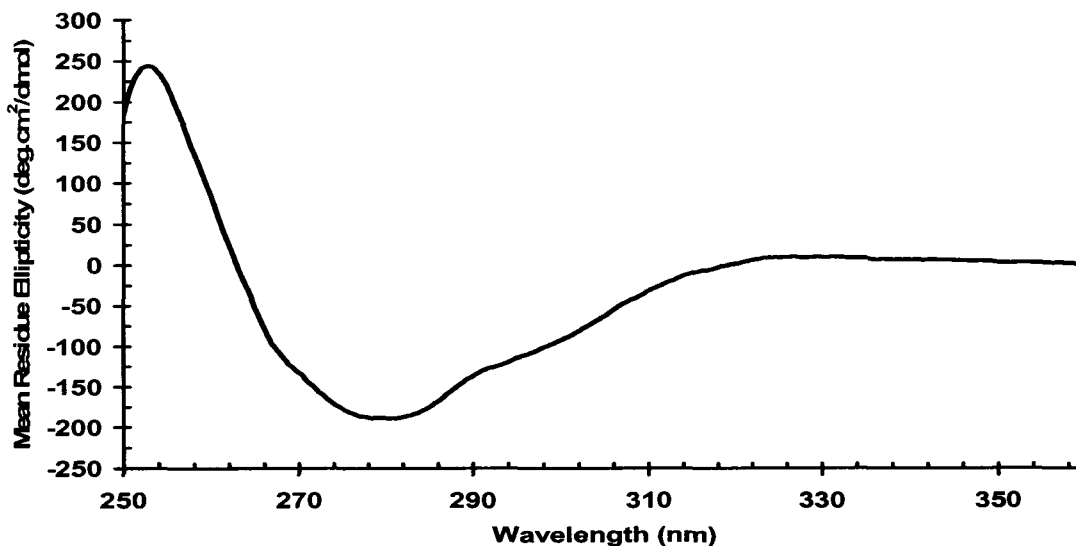
Figure 5.7. Near-UV CD spectrum of the C2 domain in 25 mM carbonate buffer, pH 9.6 after refolding by pulsed-dilution procedure.

From Figure 5.7, the shoulders seen at around 268 nm and 289 nm can be attributed to the 21 aromatic residues (12 Phenyl alanines, 1 Tryptophan and 8 Tyrosines) present in the 6x His-tagged C2 domain. The aromatic asymmetry of the spectrum thus indicates that some of the protein molecules are folded or partly folded.

5.2.4. Far-UV CD Measurements of the C2 Domain after Refolding

After attempting to refold the C2 domain by pulsed dilution, the following CD titration experiments were done in the far-UV range.

5.2.4.1. GuHCl titration experiments

Using the same refolded protein preparation, far-UV GuHCl CD titration experiments were performed for the points from 0 M to 6 M for every 0.25 M increment. Totally 80 scans were performed both for the baseline and actual C2 samples. It was observed that there was some precipitation of the protein in GuHCl titration samples of

0.25 M, 0.5 M, 0.75 M and 1 M. But for other samples there was no precipitation at all. The baseline subtracted and normalized spectra for the 0 M and 6 M titration points are as shown in Figure 5.8.
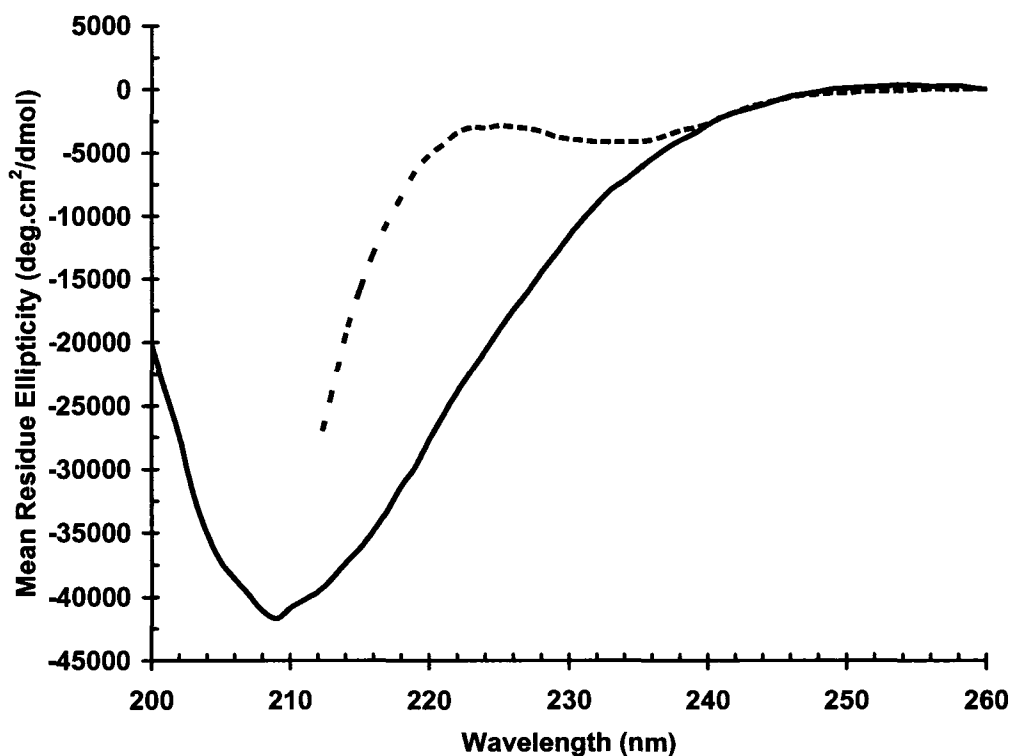
Figure 5.8. Far-UV CD spectra of the C2 domain in 25 mM carbonate buffer, pH 9.6 after the refolding procedure. Solid line: 0 M GuHCl; Dashed line: 6M GuHCl.

C2 domains in general consist of a compact β-sandwich composed of two four-stranded β-sheets [79]. The CD spectra of β-rich proteins are classified as β_I or β_{II} on the basis of the far-UV CD spectrum. The far-UV CD spectra of β_I proteins resemble those of model β sheets, having a significant positive band around 190 nm and a comparable negative band in the 210-220 nm region [102, 103]. The 0 M spectrum from Figure 5.8 shows a negative band near 208 nm, suggesting a high content of β structure as in β_I proteins. Alternately it can be interpreted that the domain is substantially unfolded under

the conditions of data collection. The C2 domain is likely to be completely unfolded in 6

M GuHCl because its spectrum closely resembles that of a typical random coil [104].

Some folded structure is present in the absence of GuHCl according to the near-UV

spectrum and the change in the shape of the far-UV spectrum on the addition of GuHCl.

The baseline subtracted and normalized spectra of all the GuHCl titration points from 0

M to 6 M resulting from the same set of experiments are overlaid in Figure 5.9.
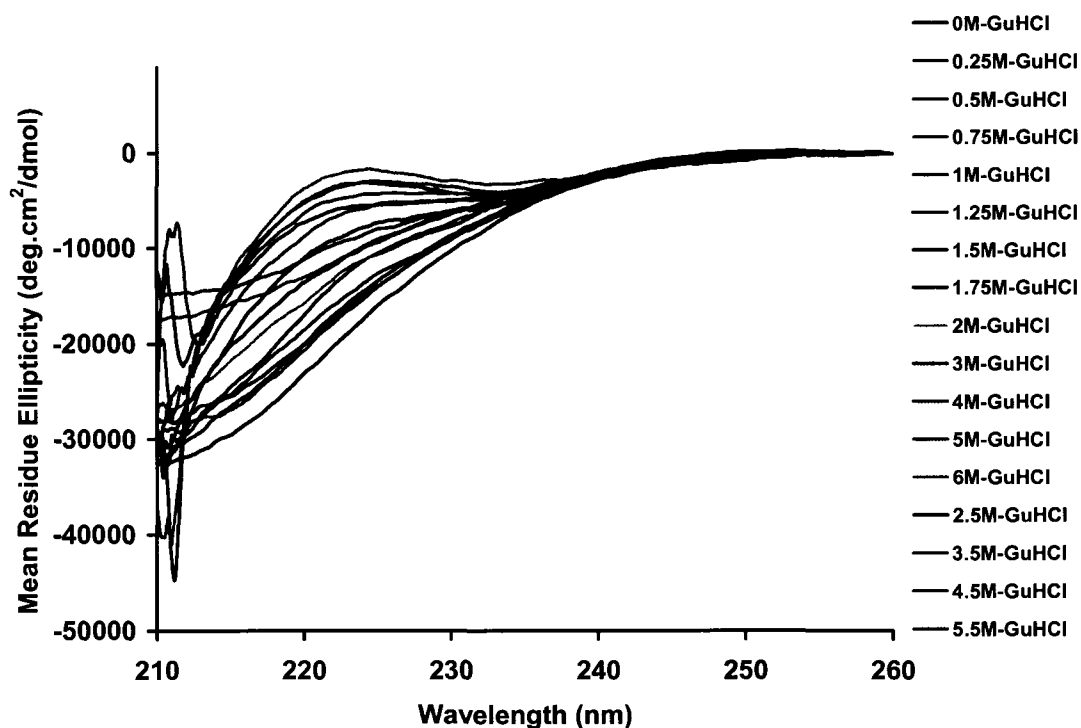
Figure 5.9. Normalized far-UV CD spectra of the C2 domain in GuHCl titration points from 0 M to 6 M. For the sake of clarity, between the points 2 M and 6 M, only spectra at every 0.5 M increment are shown.

Chemical denaturation of the domain was also monitored by the change in

ellipticity at 220 nm at various concentrations of GuHCl. Ellipticity at 220 nm as a

function of GuHCl concentration is as shown in Figure 5.10.
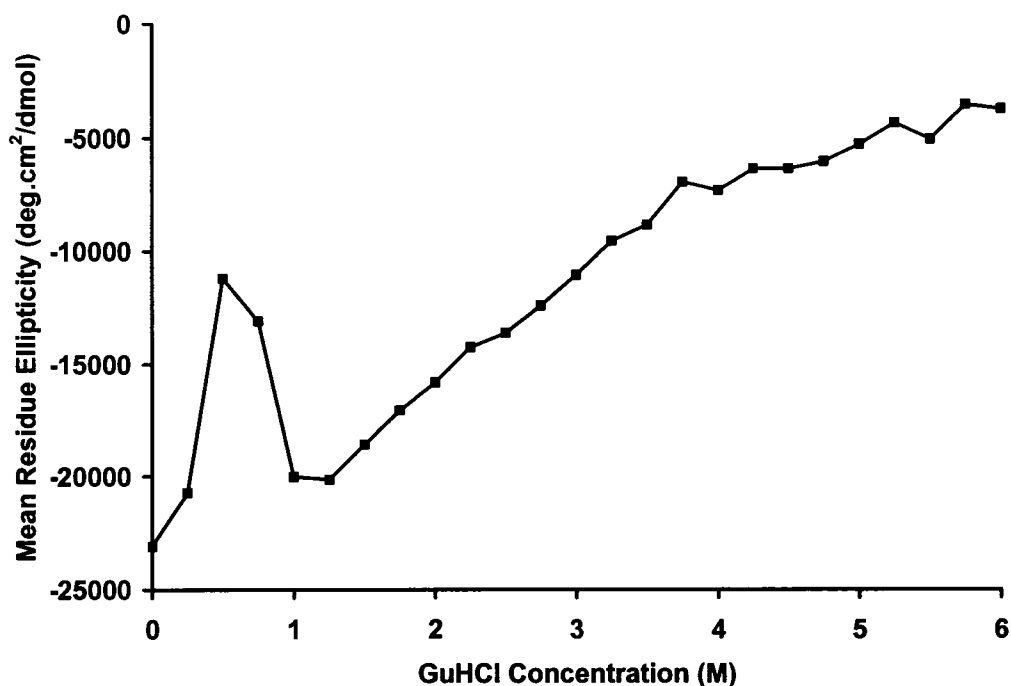
Figure 5.10. Ellipticity at 220 nm as a function of GuHCl concentration.

From the profile in Figure 5.10, it can be seen that the partial precipitation of the C2 protein in 0.25 M, 0.5 M, 0.75 M and 1 M GuHCl samples caused a spike, thus making it hard to interpret the pattern of protein unfolding in GuHCl.

Studies by Edwin and colleagues [105] have shown that for some acidic proteins for example papain, low concentrations of GuHCl can induce aggregation and that it can be prevented by the presence of low concentrations of urea in the buffer. In addition, stabilization of protein can also be induced.

5.2.4.2. Urea titration experiments

Before attempting to do the same for the C2 domain, GuHCl was replaced by urea to perform the titration experiments. As such, all the titration experiments were repeated with urea as before in 25 mM carbonate buffer, pH 9.6. The baseline subtracted and

normalized spectra for all the points from 0 M to 6.5 M at every 0.5 M of urea

concentration are superimposed and are shown in Figure 5.11.
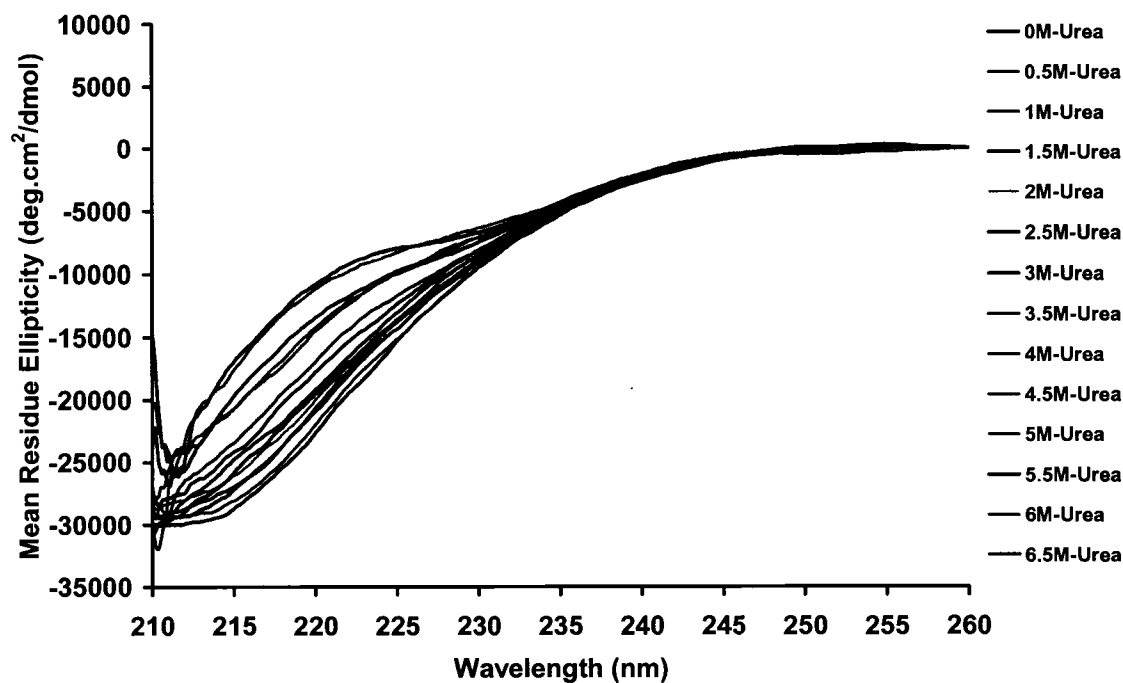
Figure 5.11. Normalized far-UV CD spectra of the C2 domain in urea titration points from 0 M to 6 M.

In this case also, the chemical denaturation was monitored by the change in

ellipticity at 220 nm at various concentrations of urea as shown in Figure 5.12.
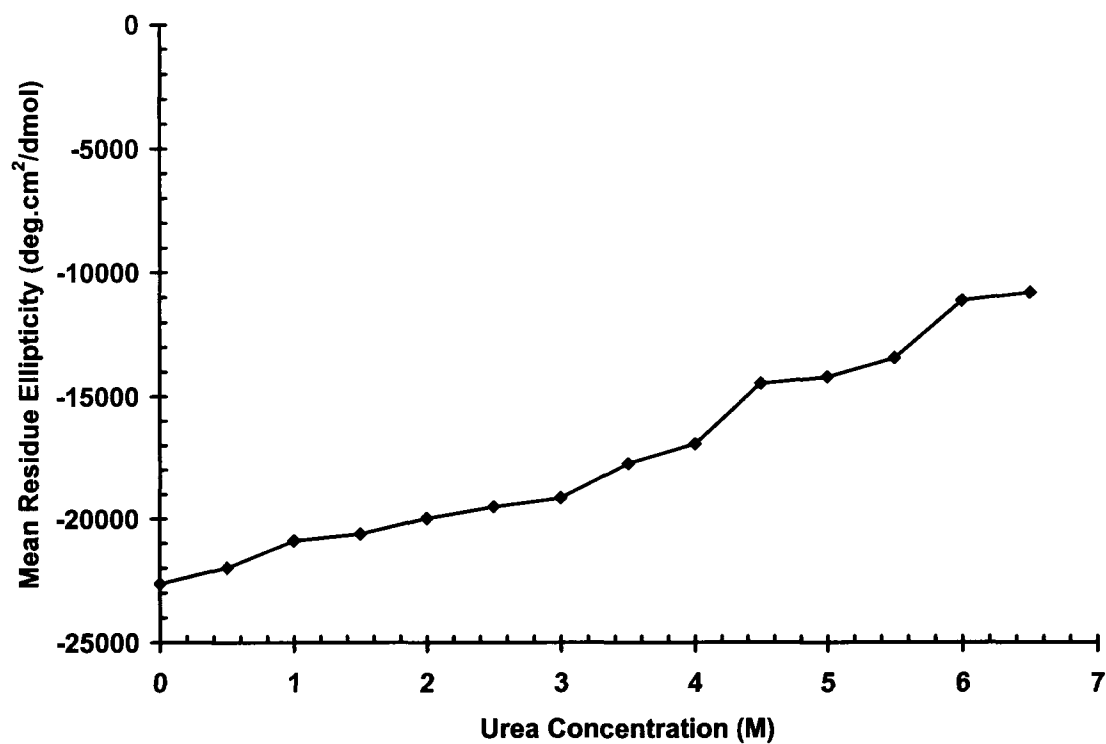
Figure 5.12. Ellipticity curve at 220 nm for the C2 domain as a function of urea concentration.

The urea denaturation data from Figures 5.11 and 5.12 suggest a gradual unfolding of the domain. However, the far-UV CD spectrum in 6.5 M urea shows evidence for some residual structure in the protein. The comparison of the spectra in 6.5 M urea and 6 M GuHCl is as shown in Figure 5.13. The spectrum in 6 M GuHCl very closely resembles that of a random coil, while that of 6.5 M urea does not. Hence, it can be concluded that a mild denaturant like urea would not be appropriate to use to study the thermostability of the C2 domain.
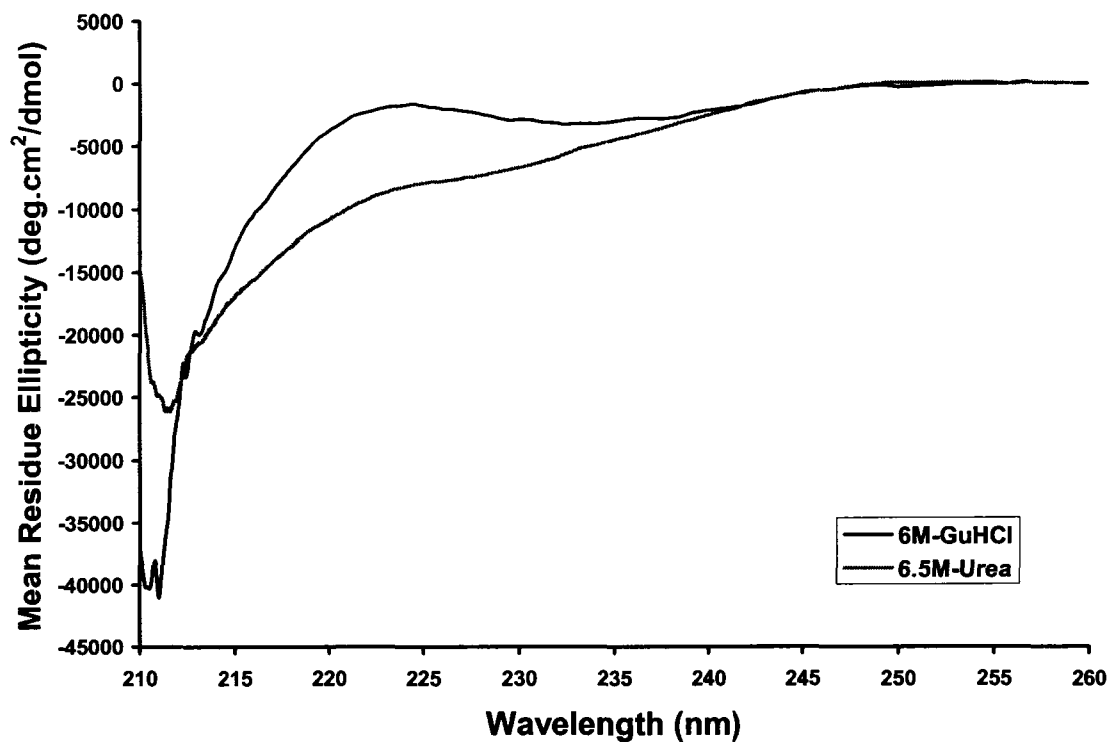
Figure 5.13. Comparison of far-UV CD spectra of C2 domain in 6 M GuHCl and 6.5 M urea.

Therefore, as initially suggested by Edwin et al. [103], urea was used in the GuHCl titration experiments to get around with the problem of protein precipitation.

5.2.4.3. GuHCl titration experiments with low amounts of urea

All the GuHCl titration experiments are repeated with 0.5 M urea added to 0.5 M, 1 M, 1.5 M and 2 M GuHCl samples. When these samples were observed after 1 hour incubation, there was no precipitation of the protein at all. Next, the CD experiments were performed in the far-UV range as before. The baseline subtracted and normalized spectra from 0 M to 6 M GuHCl points are as shown in Figure 5.14.
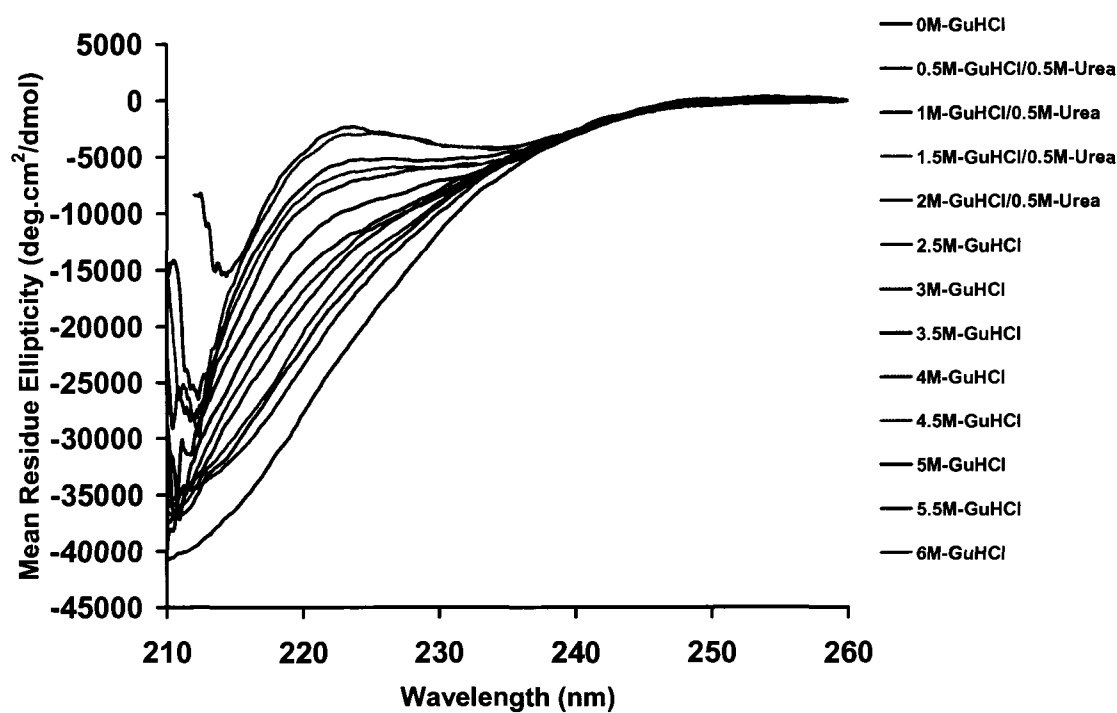
Figure 5.14. Normalized far-UV CD spectra of the C2 domain in GuHCl titration points from 0 M to 6 M with low amounts of urea.

From Figure 5.14, a very gradual transition can be seen in the shape of the spectra from a β-rich pattern to that of a random coil. Also, the ellipticity curve at 220 nm as a function of GuHCl concentration is shown in Figure 5.15. It shows a gradual unfolding of the domain probably signifying the partial folding of the domain in the absence of denaturant.
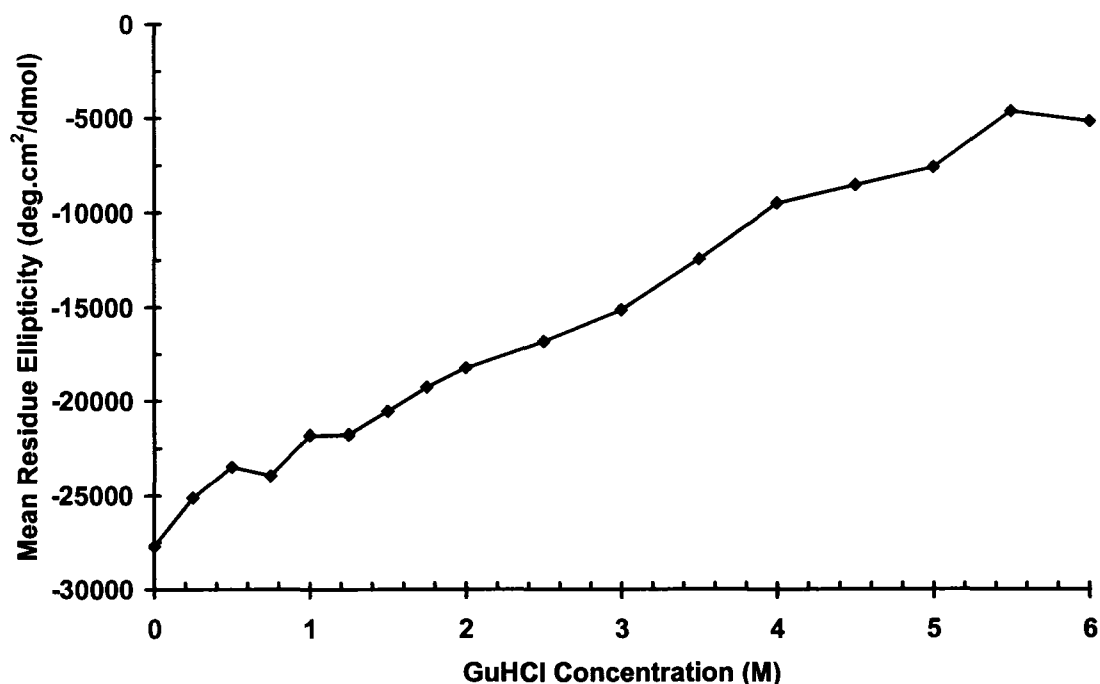
Figure 5.15. Ellipticity curve at 220 nm for the C2 domain as a function of GuHCl concentration (with low amounts of urea).

5.2.5. Differential Scanning Calorimetry Experiments

The C2 domain resulting from the pulsed dilution procedure of refolding was used for a DSC experiment. Initially the buffer alone, i.e., 25 mM sodium carbonate, 10 mM NaCl, 1 mM β-mercaptoethanol, pH 9.6 was loaded onto the reference and sample cells as per the manufacturer's protocols. Four up-scans were performed from 10°C to 110°C at a constant heating rate of 30°C per hour. The baseline scans obtained thereafter are overlaid and are as shown in Figure 5.16.
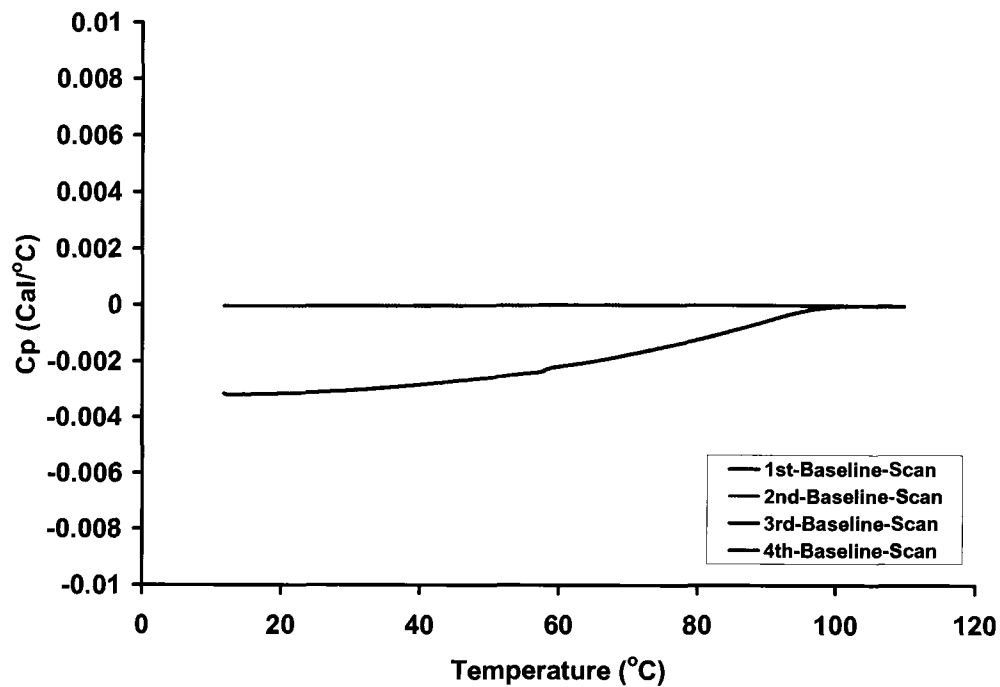
Figure 5.16. Overlay of the DSC baseline scans.

After the baseline scans were done, the reference and sample cells were cleaned. Then the C2 protein in carbonate buffer at 0.8 mg/ml is loaded onto the sample cell, while carbonate buffer is loaded onto the reference cell. Four scans were performed just as before in the same temperature range at the same rate of heating. The profiles of the scans are overlaid and are as shown in Figure 5.17.
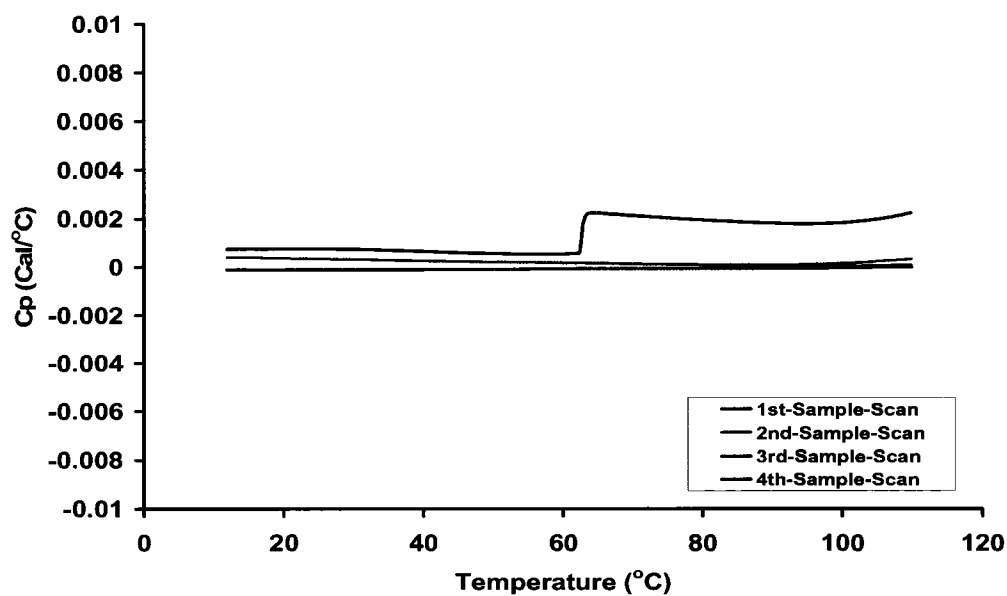
Figure 5.17. Overlay of the DSC C2 scans.

From Figure 5.17, no noticeable endothermic peak could be seen during the heat denaturation process. This result suggests that the C2 domain was partly unfolded at room temperature in the absence of chemical denaturant, or unfolding was gradual on addition of heat. The above possibilities are not mutually exclusive. Indeed, gradual unfolding at room temperature is suggested by the guanidine denaturation data, probably signifying the partial unfolding of the domain in the absence of denaturant.

The partial unfolding of the C2 domain in the absence of denaturant can well be attributed to the reason that the domain naturally tends to get expressed in *E.coli* inclusion bodies and hence would indicate the need for the presence of the PTP domain at its N-terminus in tensin in order to increase its stability just as it is in PTEN.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1. Conclusions

The C2 domain gene of human tensin has been successfully synthesized, cloned

and expressed in *E.coli*. Overexpressed C2 domain in *E.coli* inclusion bodies was

successfully isolated and purified. Initially purified protein was tested for its stability by

circular dichroism. Thereafter, various methods have been employed to refold the C2

domain. After the refolding procedures, the protein was further characterized by circular

dichroism and differential scanning calorimetry.

The biophysical characterization results, based on the near-UV CD spectrum and

the change in shape of the spectra in far-UV GuHCl CD titration experiments indicated

some folded structure in the C2 domain of human tensin. The absence of any endotherm

in the DSC experiments indicates that the domain was partly unfolded at room

temperature in the absence of chemical denaturant, or unfolding was gradual on addition

of heat; the above possibilities being not mutually exclusive. Also, gradual unfolding at

room temperature as suggested by the GuHCl denaturation data probably signifies the

partial folding of the domain in the absence of denaturant. The partial folding of the C2

domain as evident from the experimental results indicates that the C2 domain may be

thermodynamically unstable on its own in solution and hence there is a need for the

presence of the PTP domain at its N-terminus for additional stability. The experimental

116

results also reinforce the existence of a C2 domain in tensin which has not been documented earlier.

6.2. <u>Future Work</u>

Co-expressing the PTP domain gene linked to 5' end of the C2 domain gene might increase the thermostability of the C2 domain and thus make it functional because PTP-C2 seems to be a common supra-domain structural/functional motif and also that both the PTP and C2 domain are required for the focal adhesion localization of the N-terminus of tensin. Both the C2 and PTP domain encoding genes can be linked by doing a PCR on each of the domain genes using primers encoding a *Bcl I* restriction site. This restriction site is compatible with the amino acid sequence of the linker residues between the PTP and C2 domains. Following PCR, the products can be digested with *Bcl I*, purified and ligated. The ligation product can then be amplified by PCR, using end primers. The resulting gene can thus be purified by agarose electrophoresis, digested with *Nde I* and *Xho I*, and cloned into the pET14b vector.

Anti-sera containing antibodies against the PTP, C2 and PTP-C2 domains can be prepared by using animals, e.g., rabbits. Since both these domains are in the N-terminus of tensin, the anti-sera to these domains should be particularly useful for immunoprecipitating proteins that interact with the C-terminus, i.e., with the SH2 domain, PTP domain and, perhaps, the central region of tensin. Such antibodies will have a number of advantages over commercial antibodies. They will be more specific for the corresponding tensin domains than full-length tensin polyclonal antibodies, and they will probably be more suitable for immunoprecipitation than monoclonal antibodies available from a commercial source.

Once the domain or the domain pair is folded, thermostability for the same can be determined by differential scanning calorimetry as was attempted in this work. This approach enables a direct measurement of ΔH of the domain/domain-pair unfolding due to heat denaturation. Heat of unfolding (ΔH) can be determined by integrating the area under the peak (ΔH_{cal}). This will be compared with the computed van't Hoff enthalpy (ΔH_{vH}) to assess the cooperativity of denaturation. Heat capacity (ΔC_p) of the domain/domain-pair can also be measured by DSC, either by non-linear squares fitting of a 2- or multi-state model to baseline subtracted experimental data, or by measurement of the heat of unfolding as a function of pH.

Conditions for growing crystals for the domain/domain-pair can be sought. Initial crystallization conditions can be sought using various crystallization kits, testing a wide range of pH, salt concentrations, and precipitants for the highly concentrated protein. Results from an initial screen will produce crystals or solubility information. When crystals are obtained, information obtained from the screen will be used to optimize crystallization conditions to produce crystals suitable for X-ray diffraction analysis.

More detailed knowledge of the structure and function of tensin resulting from the above kind of experiments for all the domains will lead to a better knowledge of other focal adhesion components, advancing the development of molecular models of cell attachment and migration. It provides a rational basis for drug design pertaining to the cellular processes and can be of commercial value. It will also lead to a better understanding of the nature of molecular assemblies in cells.

APPENDIX A

MATERIALS

MATERIALS

All the oligonucleotides were purchased from Syngen, Inc. (San Carlos, CA). dNTPs, restriction enzymes *viz* NdeI and XhoI, T4 DNA ligase and Wizard plus minipreps DNA purification systems were all purchased from Promega (Madison, WI) while Cloned *pfu* polymerase and XL2-Blue ultracompetent cells were from Stratagene (La Jolla, CA). Both QIAquick gel extraction and PCR purification kits were purchased from Qiagen (Valencia, CA). Agarose and ampicillin were obtained from Bio-world (Dublin, OH), while LB Broth EZ Mix was from Sigma Aldrich (St.Louis, MO).

All PCR reactions were performed on thermal cycler TC-512 from Techne, Inc. (Burlington, NJ). All agarose gels were run on Electrophoresis Subsystem-70 from Labnet International, Inc. (Edison, NJ) and were visualized on DigiDoc-It Bioimaging system from UVP, Inc. (Upland, CA). For the same, TC-312R Fixed-Intensity UV Transilluminator from Spectroline ultraviolet products and an Olympus C-3040 zoom digital camera were used. Spectrafuge 16M benchtop centrifuge and 211DS Incubator/Shaker were purchased from Labnet International, Inc. (Edison, NJ). Petri dishes were purchased from Nalgene Labware (Rochester, NY).

BL21 (DE3) pLysS competent cells for protein expression were purchased from Stratagene (La Jolla, CA). Antibiotics *viz.* ampicillin and chloramphenicol were obtained from Bio-world (Dublin, OH) and Calbiochem (San Diego, CA) respectively. All other chemicals involved in the protein expression, detection and purification were bought from Sigma Aldrich (St. Louis, MO).

Pre-made resolving and stacking buffers for doing SDS-PAGE experiments were purchased from National Diagnostics (Atlanta, GA). All SDS-PAGE experiments

performed were on Dual vertical mini-gel system from Cole-Parmer (Vernon Hills, IL). The resulting gels were visualized on TW-26 White Light Transilluminator and photographed by Olympus C-3040 zoom digital camera. Lab-line 3527 orbital environmental shaker was used for bacterial cell growth. The cell growth was monitored on UV-2100 spectrophotometer from Unico (Dayton, NJ). Penta-His HRP conjugate kit and the nitrocellulose membrane for the dot blotting procedure were purchased from Qiagen (Valencia, CA) and Bio-Rad laboratories (Hercules, CA) respectively.

Cell lysis experiments were performed using Sonicator 3000 from Misonix, Inc (Farmingdale, NY). The HisTrap HP column and the AKTA prime system used for the histidine-tagged purification of the C2 domain were purchased from GE Healthcare (Piscataway, NJ). Dialysis membranes of 6-8 kDa MWCO were purchased from Spectrum laboratories (USA). Centricon filter devices of 10 kDa MWCO were purchased from Millipore (USA). Agarose, Tris, and glycerol were from Bio-world (USA). 8M guanidine hydrochloride (GuHCl) and 8M urea solutions used in CD titration experiments were from Fluka (Switzerland). All other chemicals were from Sigma-Aldrich (USA).

APPENDIX B

EXPERIMENTAL PROCEDURES

122

B.1. <u>Agarose Gel Electrophoresis (X% w/v)</u>

1. Take 'X' g of agarose and dissolve it in 100ml of 1X TBE buffer in an Erlenmeyer flask.

2. Heat the flask in a microwave oven until it begins to boil. Mix the solution well by shaking the flask at this point.

3. Now add $0.5\mu l$ of 10mg/ml ethidium bromide solution to the flask, mix well and continue boiling for another 30 seconds.

4. Take out the flask from the oven and allow the solution to cool for 5 minutes.

5. Now pour the solution into the casting tray and place the well-comb in its position. Wait for 45 minutes for the gel to be cast and then load the samples into different lanes.

6. Pour 1X TBE buffer into the gel tank well enough to cover the surface of the gel.

7. Now connect the terminals to the power supply making sure that the negative terminal is on the side of gel lanes.

8. Run the gel at 80 V for about an hour until the dye front has reached 70% of the length of the gel.

9. Turn off the power supply and visualize the DNA bands present on the gel under a UV Transilluminator.

B.2. <u>PCR Gel Extraction</u>

1. Excise the DNA fragment from the agarose gel with a clean, sharp scalpel.

2. Weigh the gel slice in a colorless tube. Add 3 volumes of Buffer QG to 1 volume of gel (100mg~100μl).

3. Incubate at 50°C for 10 min (or until the gel slice has completely dissolved). To help dissolve gel, mix by vortexing the tube every 2-3 min during the incubation.

4. After the gel slice has dissolved completely, check that the color of the mixture is yellow. If the color is orange or violet, add 10μl of 3M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.

5. Add 1 gel volume of isopropanol to the sample and mix.

6. Place a spin column in a provided 2ml collection tube.

7. To bind DNA, apply the sample to the column, and centrifuge for 1 min.

8. Discard flow-through and place column back in the same collection tube.

9. Add 0.5 ml of Buffer QG to column and centrifuge for 1 min.

10. To was, add 0.75 ml of Buffer PE to column and centrifuge for 1 min.

11. Discard the flow-through and centrifuge the column for an additional 1 min at 13,000 rpm.

12. Place the column into a clean 1.5 ml microcentrifuge tube.

13. To elute DNA, add 50 μl of Buffer EB or H$_2$O to the center of the membrane and centrifuge the column for 1 min. For increased DNA concentration, add 30 μl of elution buffer to the center of the membrane, let the column stand for 1 min, and then centrifuge for 1 min.

B.3. PCR Purification

1. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix.

2. Place a spin column in a provided 2 ml collection tube.

3. To bind DNA, apply the sample to the column and centrifuge for 30-60 sec.

4. Discard flow-through. Place the column back into the same tube.

5. To wash, add 0.75 ml Buffer PE to the column and centrifuge for 30-60 sec.

6. Discard flow-through and place the column back in the same tube. Centrifuge for an additional 1 min.

7. Place the column in a clean 1.5 ml microcentrifuge tube.

8. To elute DNA, add 50 μl Buffer EB or H_2O to the center of the membrane and centrifuge the column for 1 min. For increased DNA concentration, add 30 μl of elution buffer to the center of the membrane, let the column stand for 1 min, and then centrifuge.

B.4. Preparation of LB Agar Plates

1. Add 3.56 g of LB agar to 100ml of DI water in a bottle and shake it well.

2. Autoclave the bottle for 15 minutes.

3. Take the bottle out of the autoclave machine and swirl the bottle gently.

4. Allow the solution to cool to around 55^0C. Then add required amount of antibiotic/ antibiotics to the solution and mix it by shaking.

5. Then pour the solution in small amounts onto sterile plates/dishes. Tilt the plates thereafter so that the solution covers the entire surface of the plates/dishes.

6. Leave the plates at room temperature for about 15 min to solidify. Then, wrap them in aluminum foil and store at $+4^0C$.

B.5. Preparation of LB Broth

1. Dissolve 20.6 g of LB Broth EZmix in 1 liter of DI water in a bottle.

2. Autoclave the bottle for 15 minutes.

3. Take out the bottle and allow it to cool to room temperature.

4. Store the broth at $+4^0$C.

B.6. Transformation of XL2-Blue Ultracompetent Cells

1. Pre-chill two 14-ml falcon polypropylene round-bottom tubes on ice (One tube is for the experimental transformation and the other for the pUC 18 control). Preheat NZY$^+$ broth to 42°C.

2. Thaw the ultracompetent cells on ice. When thawed, gently mix and aliquot 100 μl of cells into each of the two pre-chilled tubes.

3. Add 2 μl of the β-mercaptoethanol provided with this kit to each aliquot of cells.

4. Swirl the tubes gently. Incubate the cells on ice for 10 min, swirling gently every 2 min.

5. Add 0.1-50 ng of the experimental DNA to one aliquot of cells, in a 1-μl volume for maximum efficiency. Dilute the pUC 18 control DNA 1:10 with sterile DI water; then add 1 μl of the diluted pUC 18 DNA to the other aliquot of cells.

6. Swirl the tubes gently, and then incubate the tubes on ice for 30 min.

7. Heat-pulse the tubes in a 42°C water bath for 30 sec. The duration of the heat pulse is critical for maximum efficiency.

8. Incubate the tubes on ice for 2 min.

9. Add 0.9 ml of preheated (42°C) NZY+ broth and incubate the tubes at 37°C for 1 hour with shaking at 225-250 rpm.

10. Plate about 200 μl of the transformation mixture on LB agar plates containing the appropriate antibiotic. For the pUC 18 control transformation, plate 5 μl of the transformation on LB-ampicillin agar plates.

11. Incubate the plates at 37°C overnight and inspect for colonies thereafter.

B.7. Plasmid Extraction by Minipreps Kit

1. Pellet 1-3 ml of cell culture by centrifugation for 1-2 min at 10,000 x g in a microcentrifuge. Pour off the supernatant and blot the tube upside-down on a paper towel to remove excess media.

2. Completely resuspend the cell pellet in 200 μl of cell resuspension solution. Transfer the cells to a 1.5ml microcentrifuge tube if necessary.

3. Add 200 μl of cell lysis solution and mix by inverting the tube 4 times. The cell suspension should clear immediately.

4. Add 200 μl of neutralization solution and mix by inverting the tube 4 times.

5. Centrifuge the lysate at 10,000 x g in a microcentrifuge for 5 minutes. If a pellet has not formed by the end of the centrifugation, centrifuge an additional 15 min.

6. For each minipreps, prepare one minicolumn. Remove the plunger from a 3ml disposable syringe and set it aside. Attach the syringe barrel to the Luer-Lok extension of the minicolumn and pipet 1 ml of the resuspended resin into the barrel. (If crystals or aggregates are present, dissolve by warming the resin to 25-37°C for 10 minutes. Cool to 30°C before use.)

7. Carefully remove all of the cleared lysate from each miniprep (supernatant from step 5) and transfer it to the barrel of the minicolumn/ syringe assembly containing the resin.

8. Carefully insert the syringe plunger and gently push the slurry into the minicolumn.

9. Detach the syringe from the minicolumn and remove the plunger from the syringe barrel. Reattach the syringe barrel to the minicolumn. Pipet 2 ml of column wash solution into the barrel of the minicolumn/ syringe assembly. Insert the plunger into the syringe and gently push the column wash solution through the minicolumn.

10. Remove the syringe and transfer the minicolumn to a 1.5 ml microcentrifuge tube. Centrifuge the minicolumn at 10,000 x g in a microcentrifuge for 2 minutes to dry the resin.

11. Transfer the minicolumn to a new 1.5 ml microcentrifuge tube. Add 50 μl of nuclease-free water to the minicolumn and wait 1 minute. Centrifuge at 10,000 x g in a microcentrifuge for 20 sec to elute the DNA.

B.8. SDS-PAGE (12%)

1. Initially the two glass plates of the dual vertical mini-gel system are thoroughly washed and cleaned.

2. The silicone gasket is secured around one of the plates, and spring clamps are used to hold the two plates against each other, providing a leakproof seal.

3. Now, 5ml of 12% resolving gel is prepared by mixing the following components: 2ml of Protogel solution, 0.325ml of 4x resolving buffer, 50μl of 10% (w/v) APS, 5μl of TEMED and 2.615ml of DI water.

4. Then, the resulting solution is degassed and poured onto the space between the two plates and is allowed to set for about 45 minutes.

5. Now, 1.6ml of 4% stacking gel is prepared by mixing the following components: 0.208 ml of Protogel solution, 0.4 ml of 4x stacking buffer, 16 μl of 10% (w/v) APS, 1.6 μl of TEMED and 0.976 μl of DI water.

6. Then the resulting solution is degassed and poured on the top of the resolving gel. At the same time, a Teflon comb is placed on the top of the gel to make wells in it.

7. Then the gel is allowed to set for about 30 minutes.

8. After 30 minutes, the samples are loaded into the wells of the gel and the gel is run at 200V, 50mA for 45 minutes.

APPENDIX C

CODON USAGE TABLE

Table C.1. Codon usage table for *Escherichia Coli K12* strain

Triplet	Amino acid	Fraction	Frequency: per thousand	Triplet	Amino acid	Fraction	Frequency: per thousand
UUU	F	0.57	22.4	UAU	Y	0.57	16.3
UUC	F	0.43	16.6	UAC	Y	0.43	12.3
UUA	L	0.13	13.9	UAA	*	0.64	2.0
UUG	L	0.13	13.7	UAG	*	0.07	0.2
CUU	L	0.10	11.0	CAU	H	0.57	12.9
CUC	L	0.10	11.0	CAC	H	0.43	9.7
CUA	L	0.04	3.9	CAA	Q	0.35	15.5
CUG	L	0.50	52.8	CAG	Q	0.65	28.8
AUU	I	0.51	30.4	AAU	N	0.45	17.6
AUC	I	0.42	25.0	AAC	N	0.55	21.7
AUA	I	0.07	4.3	AAA	K	0.77	33.6
AUG	M	1.00	27.8	AAG	K	0.23	10.2
GUU	V	0.26	18.4	GAU	D	0.63	32.2
GUC	V	0.22	15.2	GAC	D	0.37	19.1
GUA	V	0.15	10.9	GAA	E	0.69	39.6
GUG	V	0.37	26.2	GAG	E	0.31	17.8
UCU	S	0.15	8.5	UGU	C	0.45	5.2
UCC	S	0.15	8.6	UGC	C	0.55	6.4
UCA	S	0.12	7.1	UGA	*	0.29	0.9
UCG	S	0.15	8.9	UGG	W	1.00	15.3
CCU	P	0.16	7.0	CGU	R	0.38	21.0
CCC	P	0.12	5.5	CGC	R	0.40	22.0
CCA	P	0.19	8.5	CGA	R	0.06	3.5
CCG	P	0.52	23.3	CGG	R	0.10	5.4
ACU	T	0.17	8.9	AGU	S	0.15	8.7
ACC	T	0.44	23.4	AGC	S	0.28	16.0
ACA	T	0.13	7.0	AGA	R	0.04	2.1
ACG	T	0.27	14.4	AGG	R	0.02	1.2
GCU	A	0.16	15.3	GGU	G	0.34	24.9
GCC	A	0.27	25.5	GGC	G	0.40	29.4
GCA	A	0.21	20.3	GGA	G	0.11	7.9
GCG	A	0.36	33.7	GGG	G	0.15	11.0

APPENDIX D

OLIGO CALCULATIONS

OLIGO CALCULATIONS

(1) Molecular weight of an oligonucleotide is given by

M.W= [312.21 X A's] + [289.18 X C's] + [329.21 X G's] + [304.21 X T's] − 62 Da.

(2) Extinction coefficient of an oligonucleotide is given by

E= [16,000 X A's] + [12,000 X G's] + [7,000 X C's] + [9600 X T's]

(3) Quantity of H_2O needed in ml to dissolve an oligo is given by

Q= [O.D of the oligo at 260 nm] / [E X Molar concentration needed for the oligo]

Where 'E' is the extinction coefficient of the oligo.

REFERENCES

[1] Human Genome Project Information
Retrieved July 15, 2007, from
http://www.ornl.gov/sci/techresources/Human_Genome/project/info.shtml

[2] Definition of Proteomics, Human Proteome Organization
Retrieved July 17, 2007 from
http://www.hupo.org/overview/background/proteomics.asp

[3] Definition of Protein, Home page of Transgalactic Ltd.
Retrieved August 14, 2007 from
http://www.bionewsonline.com/5/what_is_protein.htm

[4] Leibler, D.C., "*Introduction to proteomics: Tools for the New Biology*" Humana Press, Ed 2nd, pp.3-4, Totowa, NJ, 2002.

[5] Gumbiner, B.M., "Cell adhesion: The molecular basis of tissue architecture and morphogenesis," *Cell*, vol.84, pp.345-357, 1996.

[6] Aplin, A.E., Howe, A., Alahari, S.K., and Juliano, R.L., "Signal transduction and signal modulation by cell adhesion receptors: The role of integrins, cadherins, immunoglobulin-cell adhesion molecules, and selectins," *Pharmacological Reviews*, vol. 50 (2), pp. 197-263, 1998.

[7] Zamir, E and Geiger, B., "Molecular complexity and dynamics of cell-matrix adhesions," *Journal of Cell Science*, vol. 114 (20), pp. 3583-3590, 2001.

[8] Jockusch, B.M., Bubeck, P., Giehl, K., Kroemker, M., Moschner, J., Rothkegel, M., Rudiger, M., Schluter, K., Stanke, G and Winkler, J., "The molecular architecture of focal adhesions," *Annual Review of Cell and Developmental Biology*, vol. 11, pp. 379-416, 1995.

[9] Burridge, K., Fath, K., Kelly, T., Nuckolls, G and Turner, C., "Focal adhesions: Transmembrane junctions between the extracellular matrix and the cytoskeleton," *Annual Review of Cell and Developmental Biology*, vol.4, pp. 487-525, 1988.

134

[10] Vuori, K., "Integrin signaling: tyrosine phosphorylation events in focal adhesions," *Journal of Membrane Biology*, vol. 165 (3), pp. 191-199, 1998.

[11] Blanchard, A., Ohanian, V and Critchley, D., "The structure and function of alpha-actinin," *Journal of Muscle Research and Cell Motility*. vol. 10(4), pp. 280-289, 1989.

[12] Crawford, A., Michelsen, J.W and Beckerle, M.C., "An interaction between zyxin and a-actinin," *Journal of Cell Biology*, vol. 116, pp. 1381-1393, 1992.

[13] Pavalko, F.M., Otey, C.A., Simon, K.O and Burridge, K., "A direct link between actin and integrins," *Biochemical Society Transactions*, vol. 19, pp. 1065-1069, 1991.

[14] Burridge, K and Connel, A., "A new protein of adhesion plaques and ruffling membranes," *Journal of Cell Biology*, vol. 97, pp. 359-367, 1983.

[15] Rees, D.J., Ades, S.E., Singer, S.J and Hynes, R.O., "Sequence and domain structure of talin," *Nature*, vol. 347, pp. 685-689, 1990.

[16] Burridge, K and Chrzanowska-Wodnicka, M., "Focal adhesions, contractility, and signaling," *Annual Review of Cell and Developmental Biology*, vol. 12, pp. 463-519, 1996.

[17] Chen, H.C., Appeddu, P.A., Parsons, J.T., Hildebrand, J.D., Schaller, M.D and Guan, J.L., "Interaction of focal adhesion kinase with cytoskeletal protein talin," *Journal of Biological Chemistry*, vol. 270, pp. 16995-16999, 1995.

[18] Nuckolls, G.H., Romer, L.H and Burridge, K., "Microinjection of antibodies against talin inhibits the spreading and migration of fibroblasts," *Journal of Cell Science*, vol. 102, pp. 753-762, 1992.

[19] Pavalko, F.M and Burridge, K., "Disruption of the actin cytoskeleton after microinjection of proteolytic fragments of alpha-actinin," *Journal of Cell Biology*, vol. 114, pp. 481-491, 1991.

[20] Molony, L.M and Burridge, K., "Molecular shape and self association of vinculin and metavinculin," *Journal of Cellular Biochemistry*, vol. 29, pp. 31-36, 1985.

[21] Johnson, R.P and Craig, S.W., "An intramolecular association between the head and tail domains of vinculin modulates talin binding," *Journal of Biological Chemistry*, vol. 269, pp. 12611-12619, 1994.

[22] Johnson, R.P and Craig, S.W., "F-actin binding site masked by the intramolecular association of vinculin head and tail domains," *Nature*, vol. 373, pp. 261-264, 1995.

[23] Fukami, K., Endo, T., Imamura, M and Takenawa, T., "α-Actinin and vinculin are PIP₂-binding proteins involved in signaling by tyrosine kinase," *Journal of Biological Chemistry*, vol. 269, pp. 1518-1522, 1994.

[24] Glenney, J.R and Zokas, L., "Novel tyrosine kinase substrates from Rous sarcoma virus-transformed cells are present in the membrane skeleton," *Journal of Cell Biology*, vol. 108, pp. 2401-2408, 1989.

[25] Turner, C.E., "Paxillin: A cytoskeletal target for tyrosine kinases," *Bioessays*, vol. 16(1), pp. 47-52, 1994.

[26] Turner, C.E and Miller, J.T., "Primary sequence of paxillin contains putative SH2 and SH3 domain binding motifs and multiple LIM domains identification of a vinculin and pp125 [fak-] binding region," *Journal of Cell Science*, vol. 107(6), pp. 1583-1591, 1994.

[27] Hildebrand, J.D., Schaller, M.D and Parsons, J.T., "Paxillin, a tyrosine phosphorylated focal adhesion-associated protein binds to the carboxyl terminal domain of focal adhesion kinase," *Molecular Biology of the Cell*, vol. 6(6), pp. 637-647, 1995.

[28] Turner, C.E., Glenney Jr, J.R and Burridge, K., "Paxillin: a new vinculin-binding protein present in focal adhesions," *Journal of Cell Biology*, vol. 111, pp. 1059-1068, 1990.

[29] Birge, R.B., Fajardo, J.E., Reichman, C., Shoelson, S.E., Songyang, Z., Cantley, L.C., and Hanafusa, H., "Identification and characterization of a high-affinity interaction between v-Crk and tyrosine-phosphorylated paxillin in CT10-transformed fibroblasts," *Molecular and Cellular Biology*, vol. 13(8), pp. 4648-4656, 1993.

[30] Sabe, H., Hata, A., Okada, M., Nakagawa, H., and Hanafusa, H., "Analysis of the binding of the Src homology 2 domain of Csk to tyrosine- phosphorylated proteins in the suppression and mitotic activation of c-Src," *Proceedings of National Academy of Sciences*, vol. 91(9), pp. 3984-3988, 1994.

[31] Weng, Z., Taylor, J. A., Turner, C. E., Brugge, J. S., and Seidal-Dugan, C., "Detection of Src homology 3- binding proteins, including paxillin, in normal and v-Src-transformed Balb/c 3T3 cells," *Journal of Biological Chemistry*, vol. 268(20), pp. 14956-14963, 1993.

[32] Mittal, B., Sanger, M., and Sanger, J.W., "Binding and distribution of fluorescently labeled filamin in permeability and living cells," *Cell Motility and the Cytoskeleton*, vol. 8(4), pp. 345:359, 1987.

[33] Gorlin, J.B., Yamin, R., Egan, S., Stewart, M., Stossel, T.P., Kwiatkowski, D. J., and Hartwig, J.H., "Human endothelial actin-binding protein (ABP-280, nonmuscle filamin): a molecular leaf spring," *Journal of Cell Biology*, vol. 111, pp. 1089-1105, 1990.

[34] Sharma, C.P., Ezzell, R.M ., and Arnaout, M.A., "Direct interaction of filamin (ABP-280) with the beta 2- integrin subunit CD18," *Journal of Immunology*, vol. 154(7), pp. 3461-3470, 1995.

[35] Wilkins, J.A., Risinger, M.A., and Lin, S., "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*, vol. 103, pp. 1483-1494, 1986.

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

[37] Haynie, D.T., and Ponting, C.P., "The N-terminal domains of tensin and auxilin are phosphatase homologues," *Protein Science*, vol. 5(12), pp. 2643-2646, 1996.

[38] Lee, J. O., Yang, H., Georgescu, M. M., Di Cristofano, A., Maehama, T., Shi, Y., Dixon, J. E., Pandolfi, P., and Pavletich, N. P., "Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association," *Cell*, vol. 99(3), pp. 323-334, 1999.

[39] Davis, S., Lu, M. L., Lo, S. H., Lin, S., Butler, J. A., Druker, B. J., Roberts, T. M., Q, An., and Chen, L. B., "Presence of an SH2 domain in the actin- binding protein tensin," *Science*, vol. 252(5006), pp. 712-715, 1991.

[40] Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P., "SMART, a simple modular architecture research tool: identification of signaling domains," *Proceedings of National Academy of Sciences*, vol. 95(11), pp. 5857-5864, 1998.

[41] Lo, S. H., Weisberg, E., and Chen, L. B., "Tensin: a potential link between the cytoskeleton and signal transduction," *BioEssays*, vol. 16, pp. 817-823, 1994.

[42] Lo, S. H., An, Q., Bao, S., Wong, W. K., Lui, Y., Janmey, P. A., Hartwig, J. H. , and Chen, L. B., "Molecular cloning of chick cardiac muscle tensin- Full length cDNA sequence, expression and characterization," *Journal of Biological Chemistry*, vol. 269, pp. 22310-22319, 1994.

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

[44] Chen, H., Ishii, A., Wong, W. K., Chen, L. B., and Lo, S. H., "Molecular characterization of human tensin," *Biochemical Journal*, vol.351, pp. 403-411, 2000.

[45] Chen, H., and Lo, S. H., "Regulation of tensin- promoted cell migration by its focal adhesion binding and Src homology domain 2," *Biochemical Journal*, vol. 370(3), pp. 1039-1045, 2003.

[46] Benes, C. H., Wu, N., Elia, A. E., Dharia, T., Cantley, L. C., and Soltoff, S. P., "The C2 domain of PKCdelta is a phosphotyrosine binding domain," *Cell*, vol. 121(2), pp. 271-280, 2005.

[47] Littler, D. R., Walker, J. R., She, Y. M., Finerty, P. J Jr., Newman, E. M., and Dhe-Paganon, S., "Structure of human protein kinase C eta (PKCeta) C2 domain and identification of phosphorylation sites," *Biochemical and Biophysical Research Communications*, vol. 349(4), pp. 1182-1189, 2006.

[48] Kanaoka, Y., Kimura, S. H., Okazaki, I., Ikeda, M., and Nojima, H., "GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain," *Federation of European Biochemical Societies Letters*, vol. 402, pp. 73-80, 1997.

[49] 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., and Tavtigian, S. V., "Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers," *Nature Genetics*, vol. 15, pp. 356-362, 1997.

[50] Torgler, C. N., Narasimha, M., Knox, A. L., Zervas, C. G., Vernon, M. C., and Brown, N. H., "Tensin stabilizes integrin adhesive contacts in *Drosophila*," *Developmental Cell*, vol. 6, pp. 357-369, 2004.

[51] Lee, S. B., Cho, K. S., Kim, E., and Chung, J., "*Blistery* encodes *Drosophila* tensin protein and interacts with integrin and the JNK signaling pathway during wing development," *Development*, vol. 130(17), pp. 4001-4010, 2003.

[52] Georgescu M-M., Kirsch, K, H., Akagi, T., Shishido, T., and Hanafusa, H., "The tumor suppressor activity of PTEN is regulated by its carboxyl-terminal region," *Proceedings of National Academy of Sciences*, vol. 96, pp. 10182-10187, 1999.

[53] Calderwood, D. A., Fujioka, Y., de Pereda, J. M., Garcia-Alvarez, B., Nakamoto, T., Margolis, B., McGlade, C. J., Liddington, R. C., Ginsberg, M. H., "Integrin β cytoplasmic domain interactions with phosphotyrosine- binding domains: A structural prototype for diversity in integrin signaling," *Proceedings of National Academy of Sciences*, vol. 100, pp. 2272-2277, 2003.

[54] Chen, H., Duncan, I. C., Bozorgchami, H., and Lo, S. H., "Tensin1 and a previously undocumented family member, tensin2, positively regulate cell migration," *Proceedings of National Academy of Sciences*. vol. 99, pp. 733-738, 2002.

[55] Cui, Y., Liao, Y. C., and Lo, S. H., "Epidermal growth factor modulates tyrosine phosphorylation of a novel tensin family member, tensin3," *Molecular Cancer Research*, vol. 2, pp. 225-232, 2004.

[56] Ishii, A., and Lo, S. H., "A role of tensin in skeletal-muscle regeneration," *Biochemical Journal*, vol. 356, pp. 737-745, 2001.

[57] Lo, S. H., Yu, Q-C., Degenstein, L., Chen, L. B., and Fuchs, E., "Progressive kidney degeneration in mice lacking tensin," *Journal of Cell Biology*, vol. 136, pp. 1349-1361, 1997.

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

[59] Alonso, A., Sasin, J., Bottini, N., Friedberg, I., Friedberg, I., Osterman, A., Godzik, A., Hunter, T., Dixon, J., and Mustelin, T., "Protein tyrosine phosphatases in the human genome," *Cell*, vol. 117, pp. 699-711, 2004.

[60] Barford, D., Jia, Z., and Tonks, N. K., "Protein tyrosine phosphatases take off," *Nature Structural Biology*, vol. 2(12), pp. 1043-1053, 1995.

[61] Johnson, L. N., and Barford, D., "The effects of phosphorylation on the structure and function of proteins," *Annual Review of Biophysics and Biomolecular Structure*, vol. 22, pp. 199-232, 1993.

[62] Barford, D., Flint, A. J., and Tonks, N. K., "Crystal structure of human protein tyrosine phosphatase 1B," *Science*, vol. 263(5152), pp. 1397-1404, 1994.

[63] Koch, C. A., Anderson, D., Morgan, M. F., Ellis, C., and Pawson, T., "SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins'" *Science*, vol. 252, pp. 668-674, 1991.

[64] Pawson, T., Gish, G. D., and Nash, P., "SH2 domains, interaction modules and cellular wiring," *Trends in Cell Biology*, vol. 11, pp. 504-511, 2001.

[65] Kawata, T., Shevchenko, A., Fukuzawa, M., Jermyn, K. A., Totty, N. F., Zhukovskaya, N. V., Sterling, A. E., Mann, M., and Williams, J. G., "SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in *Dictyostelium*," *Cell*, vol. 89, pp. 909-916, 1997.

[66] Pawson, T., "Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems," *Cell*, vol. 116(2), pp. 191-203, 2004.

[67] Machida, K., and Mayer, B. J., "The SH2 domain: versatile signaling module and pharmaceutical target," *Biochimica et Biophysica Acta*, vol. 1747(1), pp. 1-25, 2005.

[68] Kuriyan, J., and Cowburn, D., "Modular peptide recognition domains in eukaryotic signaling," *Annual Review of Biophysics and Biomolecular Structure*, vol. 26, pp. 259-288, 1997.

[69] Salgia, R., Li, J. L., Lo, S. H., Brunkhorst, B., Kansas, G. S., Sobhany, E. S., Sun, Y., Pisick, E., Hallek, M., and Ernst, T., "Molecular cloning of human paxilin, a focal adhesion protein phosphorylated by P210BCR/ABL," *Journal of Biological Chemistry*, vol. 270(10), pp. 5039-47, 1995.

[70] Auger, K. R., Songyang, Z., Lo, S. H., Roberts, T. M and Chen, L. B., "Platelet-derived growth factor-induced formation of tensin and phosphoinositide 3-kinase complexes," *Journal of Biological Chemistry*, vol. 271(38), pp. 23452-23457, 1996.

[71] Forman-Kay, J. D and Pawson, T., "Diversity in protein recognition by PTB domains" *Current Opinion in Structural Biology*, vol. 9(6), pp. 690-695, 1999.

[72] Van der Geer, P and Pawson, T., "The PTB domain: a new protein module implicated in signal transduction," *Trends in Biochemical Sciences*, vol. 20(7), pp. 277-280, 1995.

[73] Songyang, Z., Margolis, B., Chaudhuri, M., Shoelson, S. E., and Cantley, L. C., "The phosphotyrosine interaction domain of Shc recognizes tyrosine-phosphorylated NPXY motif," *Journal of Biological Chemistry*, vol. 270(25), pp. 14863-14866, 1995.

[74] Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., and Margolis, B., "A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors," *Journal of Biological Chemistry*, vol. 269(51), pp. 32031-32034, 1994.

[75] Kavanaugh, W. M., Turck, C. W., and Williams, L. T., "PTB domain binding to signaling proteins through a sequence motif containing phosphotyrosine," *Science*, vol. 268(5214), pp. 1177-1179, 1995.

[76] Gustafson, T. A., He, W., Craparo, A., Schaub, C. D., and O'Neill, T. J., "Phosphotyrosine-dependent interaction of Shc and insulin receptor substrate 1 with the NPEY motif of the insulin receptor via a novel non-SH2 domain," *Molecular and Cellular Biology*, vol. 15(5), pp. 2500-2508, 1995.

[77] Margolis, B., "PTB domain: The name doesn't say it all," *Trends in Endocrinology and Metabolism*, vol. 10(7), pp. 262-267, 1999.

[78] Nishizuka, Y., "The molecular heterogeneity of protein kinase C and its implications for cellular regulation," *Nature*, vol. 34(6184), pp. 661-665, 1988.

[79] Rizo, J., and Sudhof, T. C., "C2-domains, structure and function of a universal Ca^{2+} binding domain," *J Biological Chemistry*, vol. 273(26), pp. 15879-82, 1998.

[80] Berg J. M., Tymoczko, J., and Stryer, L., "*Biochemistry*" W. H. Freeman, Ed 6th, pp.35-37, New York, NY, 2002.

[81] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., "*Molecular Biology of the Cell*," Garland Science, Ed 4th, pp. 193-195, New York, NY, 2002.

[82] Campbell, P. N., "The biosynthesis of proteins," *Progress in Biophysics and Molecular Biology*, vol. 15, pp. 1-38. 1965.

[83] Jonasson, P., Liljeqvist, S., Nygren, P. A., and Stahl, S., "Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*," *Biotechnology and Applied Biochemistry*, vol. 35, pp. 91-105, 2002.

[84] Sambrook, J and Russell, D. W., "*Molecular Cloning: A Laboratory Manual*," Cold Spring Harbor, Ed 3rd, p. 7, New York, N.Y, 2001.

[85] Westermeier, R., "*Electrophoresis in Practice,*" John Wiley & Sons, Inc, Ed 4th, p.20, Hoboken, NJ, 2005.

[86] Codon Usage Table for *Escherichia coli K12* strain, Kazusa DNA Research Institute.
Retrieved July 15, 2007 from
http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=
Escherichia+coli+K12+[gbbct]

[87] pET-14b Vector Map, Novagen, Inc
Retrieved July 27, 2007 from
http://www.emdbiosciences.com/product/TB044

[88] Comparison of expression systems, Protein Expression and Purification Facility, European Molecular Biology Laboratory.
Retrieved July 16, 2007 from
http://www.embl.de/ExternalInfo/protein_unit/draft_frames/frame_which_expre
ss_ext.htm

[89] Hannig, G. H., and Makrides, S. C., "Strategies for optimizing heterologous protein expression in *Escherichia coli*," *Trends in Biotechnology*, vol. 16, pp. 54-60, 1998.

[90] Baneyx, F., "Recombinant protein expression in *Escherichia coli*," *Current Opinion in Biotechnology*, vol. 10(5), pp. 411-21, 1999.

[91] Cleland, J. L., and Craik, C. S., "*Protein Engineering: Principles and Practice*," Wiley-Liss, Inc, New York, NY, pp.101-105, 1996.

[92] Regulation of Protein Expression, pET system manual, Ed 11th, 2006, Novagen, Inc.
Retrieved July 27, 2007 from
http://www.emdbiosciences.com/product/TB055

[93] Studier, F. W., Rosenberg, A. H., Dunn, J. J., and DuBendorff, J. W., "Use of T7 RNA polymerase to direct expression of cloned genes," *Methods in Enzymology*, vol. 185. pp. 60-89, 1991.

[94] Wilson, K., and Walker, J., "*Principles and techniques of Biochemistry and Molecular Biology*," Cambridge University, Ed 6th, pp.455-72, New York, NY, 2005.

[95] Information about SDS-PAGE, Handbook of Electrophoresis and Staining, Pierce Biotechnology, Inc.
Retrieved July23, 2007 from
http://www.piercenet.com/Objects/View.cfm?type=Page&ID=7D768473-
DF98-40E6-B636-B167EAC4553E

[96] Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T., "Practical considerations in refolding proteins from inclusion bodies, "*Protein Expression and Purification*, vol. 28(1), pp. 1-8, 2003.

[97] Information about Dot Blotting Procedure, Qiaexpress Detection and Assay Handbook, Ed 3rd, 2002, Qiagen, Inc.
Retrieved July 24, 2007 from
http://www1.qiagen.com/HB/QIAexpressDetectionAssay_EN

[98] Information about Thrombin Cleavage, Manual for Thrombin Kits, Novagen, Inc.
Retrieved July 20, 2007 from
http://www.emdbiosciences.com/product/TB188

[99] Reddy, R. C., Lilie, H., Rudolph, R., and Lange, C., "L-arginine increases the solubility of unfolded species of hen egg white lysozyme," *Protein Science*, vol. 14(4), pp. 929-935, 2005.

[100] Coligan, J. E., Dunn, B, M., Ploegh, H. L., Speicher, D. W., and Wingfield, P. T., "*Current Protocols in Protein Science*," John Wiley & Sons, NY.

[101] Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K., "Forces contributing to the conformational stability of proteins," *Journal of Federation of American Societies for Experimental Biology*, vol. 10(1), pp. 75-83, 1996.

[102] Wu, J., Yang, J. T., and Wu, C-S., "β-II conformation of all β-proteins can be distinguished from unordered form by circular dichroism," *Analytical Biochemistry*, vol. 200, pp. 359-364, 1992.

[103] Sreerama, N., and Woody, R. W., "Structural composition of β_I- and β_{II}-proteins," *Protein Science*, vol. 12, pp. 384-388, 2003.

[104] Campbell, I. D., and Dwek, R. A., "*Biological Spectroscopy*" Benjamin-Cummings Publishing Company, 1984.

[105] Edwin, F., Sharma, Y. V., and Jagannadham, M. V., "Stabilization of molten globule state of papain by urea," *Biochemical and Biophysical Research Communications*, vol. 290(5), pp. 1441-1446, 2002.