# Structural Characterization of *The SAC Domain of the Par-4 Protein* by NMR spectroscopy

A thesis presented in partial fulfilment of the requirements for the degree of

**Master of Philosophy**
**in**
**Physics**

at Massey University, Turitea, Palmerston North, New Zealand

**Sachin Raghunath Kate**
**2010**

# ABSTRACT

Prostate Apoptosis Response-4 protein (Par-4): Par-4 gene, first identified in prostate cancer cells undergoing apoptosis, encodes a pro-apoptotic protein. Par-4 selectively induces apoptosis in cancer cells causing regression of tumors in animal models.
Par-4 induces apoptosis in androgen independent prostate cancer cells and Ras-transformed cells but not in androgen-dependent normal cells. Apoptosis induction by the Par-4 involves a complex mechanism that requires activation of the Fas death receptor signalling pathway and coparallel inhibition of NF-kB transcription activity. Par-4 expression is elevated in various neurodegenerative diseases such as Alzheimer's, Parkinson's, Huntington's diseases and stroke.

Rat Par-4 MW 36kDa, full length aa332 protein that shows high homology to human Par-4. Par-4 has two putative nuclear localization sequences NLS1 (aa20-25) and NLS2 (aa 137-153) localized to the N-terminal half of the molecule. SAC domain: Selectivity for apoptosis induction in cancer cells 59-aa (aa137-195) is necessary for apoptosis. Leucine zipper(LZ) domain of 41-aa (aa 292-332) at C terminus which binds to zinc finger domain of aPKC isoforms, Wilms tumor(WT1) protein, p62 and Dlk. Par-4 Δleucine zip aa1-265, zipperless domain.

Different types of expression vectors have been tried for expression of the Par-4 SAC domain in E. coli cells. The first expression vector was H-MBP-3C was tried with *E.coli* rosetta cells. The fusion protein expressed with this vector was not stable. The Par-4 SAC domain protein gene sequence was cloned in different trial vectors. In the beginning we decided to use a vector without any fusion tag, the pPROEX-HTb which shows low level of expression in *E.coli* BL21 (DE3) CP cells. The pETTEV expression vector with a thioredoxin tag shows good expression level and fusion protein stability. The protein sample after purification was characterized by proton NMR spectroscopy, CD and DLS which does not show any presence of protein folding. pH titration by using NMR spectroscopy to simultaneously observe the protonation state of different ionizing functional groups in peptides and proteins with both acidic and basic condition was also done. There was no evidence of any protein folding in the acidic as well as basic pH range. This suggests that protein is natively unfolded.

**ACKNOWLEDGEMENTS**

I thank you my supervisor Steven Pascal for all the help and guidance and for encouraging me to pursue postgraduate studies. He provided me with a great environment to learn and introduced me to some great people along the way.

I thank you to my co-supervisor Gill Norris for helping with my work both in this project and throughout my time at Massey University. I wish I had some exciting crystal for you to help me work with, but thanks for the help during discussion on using different vectors. Thank you also to Mark Patchett for generously supplying us with GroESL plasmid.

I thanks to all the team members in BioNMR research group members Jo, Martin, Hariprasad and David at IFS, Massey University. It has been great to be able to work with a vibrant group of research students. I would like to thank Pat Edwards for helpful discussions on NMR spectroscopy. I would like to thank Trevor Loo for discussions on molecular biology and help for protein purification by HPLC.

I would like to thank to our postgraduate coordinator Prof. Andrew Brodie for encouragement and constant support during my postgraduation studies.

Thanks to my family for supporting me during this study. Thanks especially to my wife Archana who encouraged me to get the degree.

## TABLE OF CONTENTS

## List of Figures

## INDEX OF TABLES

## ABBREVIATIONS

| | |
|---|---|
| **aa** | Amino acid |
| **CD** | Circular Dichroism |
| **DNA** | Deoxyribo nucleic acid |
| **Da** | Dalton |
| **dNTP** | Deoxyribonucleotide triphosphate |
| **DTT** | Dithiothretol |
| **DLS** | Dynamic Light Scattering |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **FPLC** | Fast Protein Liquid Chromatography |
| **HSQC** | Hetero-nuclear single quantum correlation |
| **HPLC** | High Performance Liquid Chromatography |
| **LB broth** | Luria Bertani media |
| **IDP's** | Intrinsically disordered proteins |
| **IPTG** | Isopropyl-β-D-thiogalactopyranoside |
| **IMAC** | Immobilzied ion metal affinity chromatography |
| **LZ** | Leucine zipper domain |
| **MBP** | Maltose binding protein |
| **NMR** | Nuclear Magnetic Resonance spectroscopy |
| **OD** | Optical density |
| **PCR** | Polymerase chain reaction |
| **Par-4** | Prostate apoptosis response-4 |
| **PBS** | Phosphate buffered saline |
| **PKCζ** | atypical protein kinase C zeta |
| **Psi** | Pounds per square inch |
| **pI** | Isoelectric point |
| **SAC** | Selective for apoptosis induction in cancer cells |
| **SEC** | Size exclusion chromatography |
| **SDS-PAGE** | Sodium dodecyl sulphate-polyacrylamide gel electrophoresis |
| **TCEP** | Tris (2-carboxyethyl) phosphine |
| **TEMED** | N, N, N', N'-tetramethylethylenediamine |
| **Tris-HCl** | 2 amino-2-hydroxymethyl propane-1,3 diol |

| | |
|---|---|
| **Tricine** | N-(-2-hydroxy-1, 1-bis (hydroxymethyl) ethyl) glycine |
| **Thesit** | Polyethyleneglycol dodecyl ether |
| **UV** | Ultra violet |
| **WT1** | Wilms' tumor protein 1 |
| **ZF** | Zinc finger domain |

# CHAPTER 1

# INTRODUCTION

## 1.0 Apoptosis in Cancer and neurons:

A cell becomes cancerous, or is transformed, when it taps into its potential for uncontrolled cell division. Tissue homeostasis is regulated by an intricate balance between the rate of cell proliferation and apoptosis. Discordance between these processes leads either to increased growth that culminates in tumorigenesis or increased apoptotic death that culminates in involution of the tissue. Deregulated growth results from the expression of oncogenes or the lack of expression of tumour-suppressor genes[1,2,3]. The physiological cell death in animals, especially in development and in the immune system occurs by the process of apoptosis. The typical features of apoptosis namely the loss of apoptotic mechanisms often result in abridged response to cancer therapy, and therefore, alternative or combinatorial approaches to kill the cancer cells and induce tumour regression are actively pursued[4]. The accumulation of cells harbouring serious mutation or unwanted traits by inhibition of apoptosis leads to cancer and autoimmune disease. An essential feature of anticancer strategies is selective action against cancer cells, with little or no damage inflicted to normal cells. Proteins play important role in signal transduction pathways that explain the propensity of tumor cells to survive and grow. Therefore proteins in these pathways were validated as targets and new drugs were developed, some of which had impressive clinical activity which generated the excitement of this new era of cancer therapeutics.

Apoptosis is a critical process that evolved to regulate development and immunity and to protect multicellular organisms from the accumulation of damaged cells. Apoptosis is achieved through complex mechanisms that should be tightly regulated because defects in the suppression of programmed cell death can result in an uncontrolled loss of essential cells, giving rise to diverse diseases. Therefore, it is not surprising that the improper induction or malfunction of apoptotic cell death is manifested in a variety of diseases, including Alzheimer's disease[5], autoimmune diseases, Parkinson's disease[6],

and cancer. It is believed that increased apoptosis in one or more populations of neurons is behind the development of neurodegenerative diseases such as Alzheimer's[5], Parkinson's[6], Huntington's[7] diseases and stroke. Neurons are longlived cells that do not undergo active regeneration. While apoptosis is a natural process that is required for normal development of the nervous system, its reactivation later in life is pathological. In the nervous system, different neurological disorders arise from degeneration and death of neurons. It is suggested that even when apoptosis is activated in neurons, it is counteracted by responses that slow or reverse this process.

Prostate apoptosis response-4, Par-4, a pro-apoptotic protein was found to play a critical role in number of cancer and neurodegenerative diseases. Par-4 protein was initially identified as the product of a gene upregulated in prostate tumour cells undergoing apoptosis[8, 9]. Par-4 is now being linked to a rapidly growing number of neurodegenerative disorders. The first link between the Par-4 and a neurodegenerative disorder came from studies of Alzheimer's disease[10]. A rapid increase in the Par-4 levels occurs in neurons undergoing apoptosis in a variety of paradigms, including trophic factor withdrawal, and exposure to oxidative and metabolic insults. Par-4 which can be induced at the translational level, acts at an early stage of the apoptotic cascade prior to caspase a activation and mitochondrial dysfunction. The studies of post-mortem tissues from patients and animal models of neurodegenerative disorders, including Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis (ALS), and HIV encephalitis, have documented increased levels of the Par-4 in vulnerable neurons[4]. While the pro-apoptotic role of the Par-4 in cancer should be enhanced, approaches to inhibit the Par-4 expression or function in neurons must be explored. The structural characterization of the Par-4 can help for the molecular therapeutics or targeted therapeutics.

**1.1 Prostate Apoptosis Response (Par-4):**

The prostate apoptosis response-4 (*par*-4) gene was identified by differential screening for genes that are upregulated when prostate cancer cells are induced to undergo apoptosis. Apoptosis is a critical process that evolved to regulate development, immunity and to protect multicellular organisms from accumulation of damaged cells. Apoptosis is achieved through complex mechanisms that should be

tightly regulated because defects in the suppression of programmed cell death can result in an uncontrolled loss of essential cells giving rise to diverse diseases. The Par-4 is a pro-apoptotic protein that is remarkably effective in inducing cancer cell apoptosis and tumour regression in animal models. A unique ability of the Par-4 protein is that it selectively induces apoptosis in androgen-independent prostate cancer cells and in Ras-transformed mouse fibroblasts cells but not in androgen-dependent prostate cancer cells, immortalized cells, or primary cultures of normal prostate epithelial and stromal cells[11,12]. Par-4 was later rediscovered in human cells as a partner of Wilms' tumour 1 (WT1)[13] and atypical protein kinase C (aPKC)[14]. Human Par-4 was found to localize to chromosome 12q21 which is often deleted in pancreatic cancer, this region undergoes reorganization in Wilm's tumorigenesis. Par-4 is detected in almost all rat tissues with varying degrees of expression. It is detected in cells originating from endoderm, mesoderm, or ectoderm. Par-4 was also found to be ubiquitously expressed in all tissues of humans, mice, horses, pigs and cows. Par-4 levels are generally higher in dying cells, for example, in prostate ductal cells of castrated rats and degenerating neurons.

Aligned with the fact that Par-4 is an apoptotic protein, its expression is increased during development in actively dying cells, such as the interdigitating web cells of the mouse embryo and the involuting tadpole tail. Par-4 also plays key role in neuronal development, where an increase in the Par-4 levels results in apoptosis and serves to maintain the number of neurons in the nervous system. While apoptosis is a natural process that is required for normal development of the nervous system, its reactivation later in life is pathological. Neurotrophic factors have been identified that can protect neurons by activation of survival proteins such as NF-κB[15]. Par-4, a pro-apoptotic protein, was found to play a critical role in a number of cancer and neurodegenerative disease paradigms. In addition the high conservation of the Par-4 sequence in the rat, human, and mouse implies a significant biological role for the Par-4 common to many species.

**1.2 Functional domains of Par-4:**

The Par-4 gene produces a 38 kDa protein consisting of 332 amino acids which shows high homology to *human* Par-4[16]. DNA sequence analysis of the Par-4 in all three species of mouse, rat and human showed very high sequence homology. In addition to the general high homology found between three known Par-4 sequences, a more detailed examination revealed significant similarities between different domains. The high conservation of the Par-4 sequence in rat, human, and mouse implies a significant biological role for the Par-4 common to many species, making the search for its physiological functions even more important**.** Sequence analysis of the Par-4 revealed a number of interesting sites and domains. The sequence numbering is made with reference to rat Par-4, to reflect the recombinant rat (rrPr-4) constructs. The sequence identity expressed relative to rat Par-4 of mouse and human is 92% and 76% respectively, whereas African clawed frog and zebra fish share 52% and 47% sequence identity with rat, respectively (Fig.1.1 and Fig.2.1). Two putative nuclear localization sequences (NLS) are present in the amino terminus, the first NLS1 (aa 20-25) and NLS2 (aa 137-153), NLS2 is a functional nuclear localization sequence[17]. To delineate the relationship between Par-4 entry into the nucleus and induction of apoptosis, study has been done by generating untagged and GFP-tagged derivatives of the Par-4, ΔNLS1 and ΔNLS2 which lacked an intact NLS sequences. The PC-3 cells were transiently transfected with these constructs and examined for expression, nuclear entry, inhibition of NF-κB transcriptional activity, and apoptosis. ΔNLS1 showed nuclear entry, whereas ΔNLS2, which lacked intact NLS2 domain, failed to translocate to the nucleus, to inhibit NF-κB and also unable to induce apoptosis[18]. This suggests that a nuclear function of the Par-4 that was further confirmed by the discovery of Par-4/WT1 complex binds to the *bcl*-2 promoter, inhibits transcription, and thereby promotes apoptosis.

The 59 amino acids 137-195 comprise the so-called SAC-domain (selective for apoptosis induction in cancer cells). This domain alone was able to directly induce apoptosis in all studied cancer cell types by inhibition of NF-κB in the nucleus and activation of the Fas pathway. A mutant comprising the region spanning amino acid was identified as the core domain that orchestrates the apoptotic functions of the Par-

4. Upon further deletion up to the amino acid 187, the apoptotic function was lost. It was also observed that small 137-195 domain was constitutively nuclear and caused selective apoptosis of cancer cells upon overexpression. Therefore, this 137-195 mutant is now called SAC, which localized to the nucleus both in normal and cancer cells, caused apoptosis only in cancer cells when overexpressed[19]. This led to the conclusion that in addition to nuclear entry, an activating event that is present only in cancer cells is necessary for the Par-4 to attain its full apoptotic potential. In addition, the Par-4 possesses a number of conserved consensus sites for phosphorylation by kinases, such as PKA (protein kinase A) and PKC, as illustrated in (Fig.1.1 and Fig.2.1)[19]. The SAC region contains a few phosphorylation sites, and so the possibility of phosphorylation as a potential regulator was studied. Two amino acids, a serine at 154 and a threonine 155 are consensus sites for PKA. PKA is a universal kinase that phosphorylates a wide range of substrates involved in the metabolism regulation, cell growth, and differentiation.

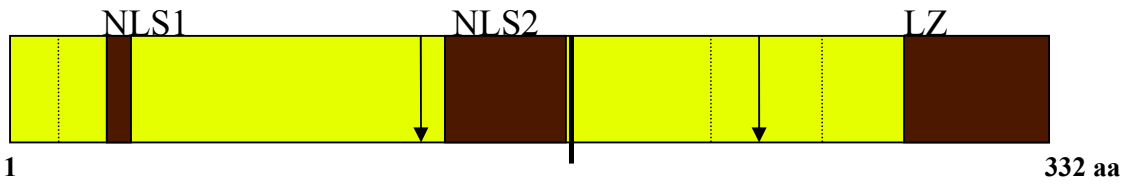


**Figure-1.1 Schematic Structure of the Par-4 showing the nuclear localization sequences NLS1 and NLS2, the apoptosis induction domain (SAC) and highlighted are the leucine zipper domain (LZ). Putative phosphorylation sites conserved in all three species are shown as vertical lines. (⋮ ▼ ▌) CKII, PKC and PKA phosphorylation sites.**

Phosphorylation of T155 is necessary for activating the Par-4. Mutation of T155 abrogated its pro-apoptotic function[20]**.** The Par-4 LZ is a main interaction motif. It contains a 41 amino acid long leucine zipper domain at its C-terminus (aa 292-332). So far all known interacting proteins of the Par-4 interact with the LZ. Collectively presence of these motifs suggests that the function of the Par-4 may be tightly regulated by post-translational modification, localization, and dimerization with partners of biological consequences. It was concluded that this domain is likely to

form a parallel coiled coil. Coiled coils are amphipathic oligomerisation motifs of α-Helices with a seven residue heptad repeat, denoted $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g\text{-})_n$ with the nonpolar residue predominating at position *a*, leucine predominating at position *d*, and polar amino acids dominating the other positions of the repeat. The C-terminal 47 residues of the Par-4 were experimentally established[21]. This structure is packed together in a parallel α-helical coiled coil with the conserved leucine residues playing a key role at position *d* are named leucine zippers. The driving forces for coiled coil formation are interhelical hydrophobic interactions between residues at position *a* and *d*. The stability of the coiled coil is fine tuned by interhelical and intrahelical interactions. Especially repulsive interactions between residues in position *e* and *g* proved to be critical for the supercoiling of the Par-4[21]. Par-4 has at least five confirmed leucine repeats in the C-terminal region[22].

The LZ facilitates homodimerisation with itself or heterodimerisation with all so far known interacting proteins. Deletion of the LZ abolishes the apoptosis sensitizing function of the Par-4[14]. As a result the NF-κB-pathway is inhibited. NF-κB is a transcription factor necessary for the induction of proteins that promotes cell survival and protects cell from apoptosis[18]. Par-4 localizes to the cytoplasm and the nucleus. Once activated it promotes apoptotic pathways and shuts down cell survival pathways. It does so by interacting with other proteins via its LZ. The best-known interacting protein of the Par-4 is the atypical protein kinase C zeta (PKCζ)[23]. Interaction of these proteins takes place between the leucine zipper domain (LZ) of the Par-4 and the zinc finger domain (ZF) of PKCζ. Binding of the LZ to the ZF within the regulatory domain of PKCζ leads to the inactivation the catalytic activity of PKCζ. As an indication of the physiological relevance of the Par-4/aPKC binding, Par-4 was shown to inhibit the expression of reporter genes. PKCζ seems to fit well with suggested functions of the Par-4, both have specific but opposing effects on apoptosis, NF-κB activity, and Fas function and both of them are affected by ceramide.

The Par-4 also interacts with other zinc finger domain of Wilms' Tumor protein 1 (WT1) isoforms bind specifically to the C-terminal end of the Par-4 but not to the N-terminus[23]. Wilms' tumour 1 gene encodes a tumour suppressor whose loss is

associated with the etiology of Wilms' tumour. One main event of the binding of LZ to WT1 is the inhibition of WT1 controlled transcription of Bcl-2. Bcl-2 is an anti-apoptotic protein that protects cells from the release of cytochrome c into the cytosol. Par-4 also binds and enhances the apoptotic function of DAP-like kinase (Dlk/Zip kinase). As shown by studies done the LZ is necessary to sensitize cells for apoptosis but dispensable for the direct induction of apoptosis[24]. However it was shown that over expression of the LZ in cells protects them from apoptosis. In this way the LZ can act as a dominant negative mutant.

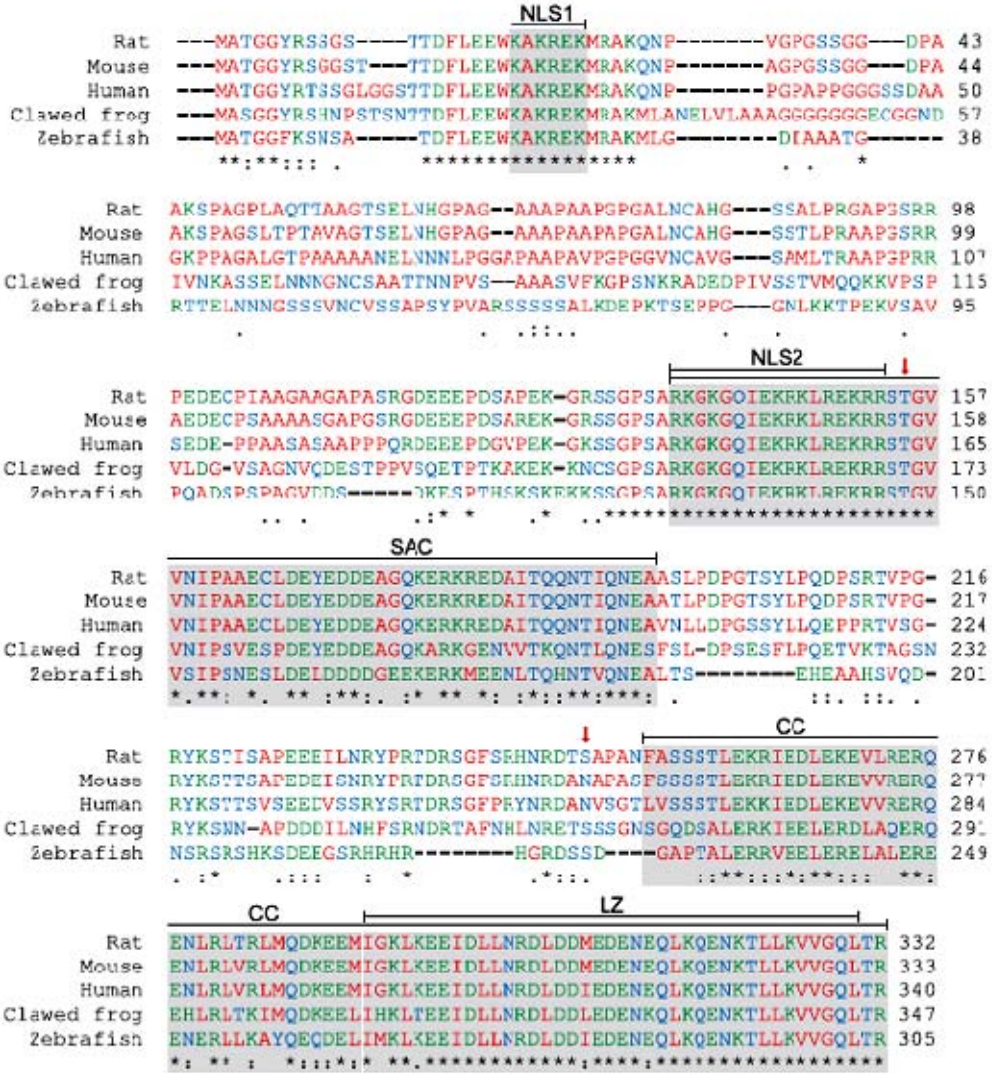


**Figure-1.2 Sequence alignment of the prostate apoptosis response factor 4 (Par-4). A BLASTP/CLUSTALW alignment of sequences of the Par-4 from various**

**species: rat (Rattus norvegicus), mouse (Mus musculus), human (Homo sapiens), African clawed frog (Xenopus laevis) and zebra fish (Danio rero). The amino acids are coloured: red (nonpolar side chains: G, A, V, L, I, M, P, F and W), blue (polar side chains: S, T, N, Q, Y and C) and green (polar, charged side chains: K, R, H, D and E). Symbols: residues in that column are identical in all sequences (*); substitutions are conservative (:); and substitutions are semi-conservative (.). The high degree of sequence conservation of the Par-4 suggests functional significance and thus resistance to evolutionary pressure. With reference to the numbering of rat Par-4, several segments are of notable interest: two nuclear localization sequences (NLS1 (20-25) and 2 (137-153), which are completely conserved among all known Par-4s, and the SAC domain (137-195), which is defined by being the absolute minimum fragment required for apoptosis and includes NLS2. The C-terminal domain (254-332) is a coiled-coil (CC) motif that encompasses a LZ (292-330) as a subset. Two important phosphorylation sites, T155 and S249, are denoted by red arrows.**

All sequence numbering is made with reference to the rat Par-4, to reflect the recombinant rat (rrPar-4) constructs used in the study.

**1.3 Mechanism of Apoptosis by Par-4:**

The role of the Par-4 in apoptosis has been well studied in cell culture models. The pro-apoptotic role of the Par-4 is apparent from its effect in cancer and neurodegenerative disease. Consistently multiple mechanisms are involved in its ability to induce apoptosis (Figure-3). The key defining features of apoptosis are the activation of a series of cysteine aspartyl protease or caspases, the central engines of apoptosis that orchestrate cell death by cleaving a variety of intracellular substrates and triggering cell demise. Caspase activation is followed by chromatin condensation and the display of phosphotidylserine on the cell surface that marks the cell for phagocytosis by specialized macrophage, thus avoiding an inflammatory response. Two general pathways are thought to be responsible for activation of caspase cascades. Apoptosis or programmed cell death usually occurs through one of the two major signalling pathways, known as extrinsic and intrinsic pathways. The extrinsic pathway proceeds through death receptors while intrinsic pathway is mediated by the

mitochondria. The extrinsic pathway is initiated at the cell surface via the activation of transmembrane death receptors of the CD95 (Apo-1 or Fas)/TRAIL/tumour-necrosis factor (TNF) receptor 1 family. Engagement of these receptors with their cognate ligand triggers the assembly of a multiprotein complex that activates the upstream caspase 8[19,20]. Caspases are synthesized as inactive zymogens and activated by proteolytic cleavage by upstream caspases[19,20]. The internal pathway is activated by the release of cytochrome c out of the mitochondria into the cytoplasm thereby activating the upstream caspase 9. Caspases 8 and 9 both activate downstream effector caspases that carry out the final steps of apoptosis.

Par-4 appears to act through extrinsic pathway by enabling the trafficking of Fas and Fas ligand to the plasma membrane. Fas (CD95) is a member of the TNF-R family and is activated by binding to Fas ligand, such binding leads to the formation of a complex consisting of trimeric Fas, the adaptor protein FADD (Fas-associated death domain) and pro-caspase-8. This complex is called the death-inducing signalling complex, or DISC. The formation of DISC in turn leads to the cleavage and activation of the zymogen, pro-caspase 8, to its active caspase form. Par-4 is able to translocate Fas and FasL to the membrane in androgen-independent prostate cancer cell lines as part of the mechanism by which it induces apoptosis in these cells. Direct induction of apoptosis by the Par-4 requires both Fas/FasL translocation and NF-κB inhibition. NF-κB is a anti-apoptotic transcription factor; Par-4 plays a very important role in antagonizing the anti-apoptotic factors which present a major obstacle to the apoptosis cascade. NF-κB is a heterodimer consisting of p65 (RelA) and p50 subunits.

This dimer is held inactive in the cytoplasm as a result of being bound to the inhibitory protein IκB, which on the appropriate signal is phosphorylated by IKKβ, ubiquitinated, and ultimately degraded to release an active NF-κB. The latter is then free to translocate to the nucleus, where it binds to specific sequences in the DNA that activates transcription of specific genes[25]. Par-4 inhibits NF-κB activity and this Par-4 mediated inhibition of NF-κB is carried out without disrupting the DNA binding capacity of the NF-κB complex. Par-4 acts in the nucleus to inhibit phosphorylation of NF-κB by IKKβ or PKCζ which would otherwise confer maximal activity to NF-κB. Also interference with IKKβ phosphorylation of the IκB complex in the cytoplasm acts as another molecular safeguard against the activation of NF-κB. In the cytoplasm,

Par-4 increases Fas and FasL translocation to the membrane. Moreover, it binds and blocks ζPKC functions such as activation of IKKβ and inhibition of Fas signalling. In the nucleus, Par-4 inhibits NF-κB transcriptional activity by mechanism that may, in part, involve ζPKC. In addition, Par-4 inhibits expression of the Bcl-2 promoter by interacting with WT1. (Figure taken directly from N. El-Guendy et.al. 2003)[19].

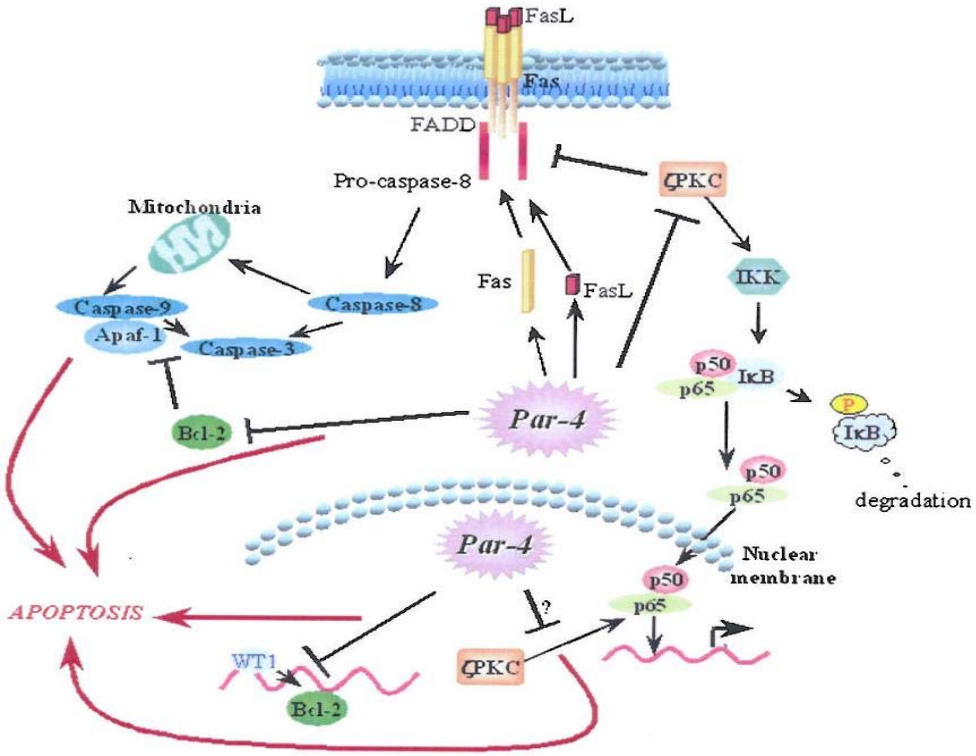


**Figure-1.3 Mechanism of action of the Par-4 showing multiple pathways, regulating cell survival both in the cytoplasm and in the nucleus.**

It is well accepted that NF-κB plays an important role in oncogenesis and in the resistance of tumour cells to chemotherapy[25]. This ability of the Par-4 to inhibit pro-survival factors is essential to its apoptosis sensitizing function. Also the Par-4 inhibits the anti-apoptotic protein Bcl-2 gene expression by binding to its promoter via WT1. In response to the vitamin A derivative ATRA (all trans-retinoic acid), Par-4 translocates to the nucleus in prostate cancer cell lines and the Par-4/WT1 complex binds to the *bcl-2* promoter and inhibits its expression. As a result of Bcl-2 repression,

the prostate cancer cells undergo cell arrest and apoptosis. The ability of the Par-4 to down-regulate and inhibit Bcl-2 can be used as a tool to target cancers resulting from Bcl-2 overexpression.

**1.4 Par-4 in Molecular Therapeutics:**

Par-4 is a pro-apoptotic protein that is highly conserved in mammals and in other vertebrates. It is involved in the critical regulation network that controls cell survival and death. Par-4 was discovered in prostate cancer cells that underwent apoptosis. This induction was specific for apoptotic stimuli but not for stress, necrosis or growth arresting stimuli[2]. It was found that Par-4 is down-regulated in different cancer cell types and that restoration of the Par-4 expression levels or overexpression leads to increased sensitivity to apoptotic stimuli or direct induction of apoptosis[19,22]. Thus evidence of the Par-4 as a candidate for gene therapy has been accumulating. Its ability to inhibit transformation in Ras-expressing cells was a major step in this direction. The mouse experiments wherein a single injection of an adenovirus expressing Par-4 resulted in an amazing regression in solid tumours arising from PC3 cell implant in nude mice and in RM-1 cell derived orthotopic prostatic tumours in immunocompetent mice[26].

In contrast to the essential critical role played by apoptosis in neuronal development, the adverse effect of apoptosis are observed in neuronal disorders, like Alzheimer's, Huntington's and Parkinson's diseases, and stroke. Par-4 was found to be up-regulated in these neuronal cells. Apoptosis in the adult nervous system has serious implications for the whole organism, since the nervous system is not classically known to undergo active regeneration. In contrast to the rapid turnover of cells in proliferative tissues, neurons commonly survive for the entire lifetime of the organism. Studying molecular mechanisms controlling neuronal apoptosis is an essential requirement for identification of targets for drugs against neurodegenerative diseases and Par-4 can be classified as such a target. The expression of endogenous Par-4 is up-regulated in different neurodegenerative diseases. Studies have found that both Par-4 mRNA and protein increase in brain regions that are sensitive to neurodegeneration in Alzheimer's disease but not in those regions that are resistant. Par-4 involvement in sensitization of neurons to apoptosis was noted as an early

event. Subsequently it was demonstrated that the Par-4 dominant-negative mutants or anti-sense oligodeoxynucleotides were able to abrogate apoptosis *in vivo*[23,19].

**1.5 NMR spectroscopy of Proteins:**

NMR (Nuclear Magnetic Resonance) spectroscopy and X-ray crystallography are currently the only techniques capable of determining the structures of biological macromolecules like proteins and nucleic acids at atomic resolution. In addition, it is possible to study time dependent phenomena with NMR, such as intramolecular dynamics in macromolecules, reaction kinetics, molecular recognition or protein folding[27]. Also NMR is the only technique for structure determination of biomacromolecules in aqueous solution and near physiological conditions. NMR solution structure is helpful to ascertain whether structures derived from single crystals represent well and precisely the functionally relevant solution conformation[28]. When an NMR signal of a given nucleus has been assigned in the NMR spectrum, the signal represents a specific environment of the nucleus in the protein structure. The chemical shift is a very sensitive parameter, which might report even subtle conformation changes near the nucleus. Also the important ability of NMR is to characterize the dynamic properties of macromolecules and describe the properties of partially or unfolded proteins. The scalar coupling, represented by the coupling constant may report on conformational changes in dihedral angles. The binding of ligand or the pH dependent change of a charge may result in conformational changes and a change in the NMR parameters for the nuclei in the environment of the chemical modification. Since NMR is specific to the atomic level NMR is by far superior to any other spectroscopic technique to register the changes, which occur as a result of the protein binding to another molecule. In addition, it is possible to study time dependent phenomena with NMR, such as intramolecular dynamics in macromolecules, reaction kinetics, molecular recognition or protein folding. The one dimensional NMR techniques, which yield extremely useful information in small molecules, are of limited applicability to the complex, highly overlapped spectra of biological macromolecules. Also the limitations of NMR spectroscopy result from the low inherent sensitivity of the technique and from the high complexity and information content of NMR spectra. Technological advances enhancing sensitivity,

such as the construction of new high-field magnets are of keen interest. The sensitivity of the acquired NMR data depends critically on the performance of the NMR probe, a sophisticated electronic device used to detect NMR signals. These problems are partially alleviated by new developments. The introduction of cryogenic probes has made a significant impact. Radiofrequency (RF) coils constitute the heart of these probes, and their sensitivity scales with the thermal noise associated with the coil's temperature. Cryogenic probes utilize RF-coils cooled to ~25 K, and the resulting sensitivity enhancement reduces instrument time requirements by factors that range from 4 to 16. Another key advance involves partial deuteration, providing samples that can be studied with improved signal-to-noise ratios that result from their sharper linewidths and longer transverse relaxation times. The combination of partial deuteration and cryogenic probes can provide a factor of 10 or more reduction in the requisite data collection times. These technologies provide the basis for high throughput NMR, and are particularly valuable for samples exhibiting limited stabilities and/or low solubilities. The sensitivity and resolution of NMR are increased by progress in spectrometer technology with high field magnets, devices for stable radio frequency generation and performing double, triple resonance experiments. Progress in the theoretical and practical capabilities of NMR lead to an increasingly efficient utilization of the information content of NMR spectra. Parallel developments in the biochemical methods (recombinant protein expression) allow the simple and fast preparation of protein samples.

NMR structure determinations rely on the nearly complete assignment of chemical shifts, which are obtained using multidimensional $^{13}C$, $^{15}N$, $^{1}H$-triple resonance NMR methods[27,28]. However, a complete set of these experiments often requires far more instrument time than the minimum dictated by signal-to-noise (S/N) requirements. A particular challenge for structural genomics is the development of NMR experiments that allow matching of instrument time investments to the minimum time required for measuring the chemical shift data. For many samples, most of the instrument time is needed not to detect signal, but to ensure appropriate resolution and/or information content of the spectra.

Traditional NMR structure determination relies on measurement of nuclear Overhauser effects (NOEs; through-space dipolar interactions between protons) and

scalar couplings (through-bond interactions between nuclei mediated by nuclear-electron interactions) for deriving distance and torsion angle constraints, respectively. NOE constraints will continue to be key for high-throughput structure determination, but the arsenal of techniques that have recently been developed to recruit additional experimental parameters for structure refinement will play a valuable role in structural genomics. Dipolar coupling constraints can establish the spatial relationship of remote segments of a biological macromolecule and can complement sparse NOE networks for obtaining high-quality structures [29,30]. Current limitations for use in structural genomics are the efficient identification of suitable orienting media in which the protein sample remains soluble. Second, chemical shifts (the NMR resonance frequencies) have long been recognized as a potential source for structural refinement. In particular, 13C$\alpha$ and 13C$\beta$ shifts offer a robust means to map the secondary structure and to derive backbone dihedral angle constraints at an early stage of the structure determination[31]. They are obtained during the resonance assignment process, and are thus of outstanding value for efficient high-throughput efforts. Third, detection of through-hydrogen bond scalar couplings[32] affords valuable unambiguous constraints for characterizing hydrogen-bonded networks, although the small size of these couplings may restrict this to smaller proteins.

To effectively utilize the information available from NMR spectroscopy of biological macromolecules the multidimensional NMR spectroscopy is used. It combines data from several kinds of spectra to establish the mappings for resonance assignments. In general, there are two types of multidimensional spectra, homo-nuclear spin, in which each axis represents the same type of atom, and hetero-nuclear spectra, in which each axis represents a different type of atom. NMR was greatly advanced by new methods for isotope labelling proteins with $^{15}N$, $^{13}C$ and $^{2}H$ using bacterial expression systems. Spectra from these samples can be drastically simplified. The development of triple resonance experiments $^{1}H$-$^{13}C$-$^{15}N$ useful for conformation independent sequential assignments.

The aim of the analysis of NMR spectra is to extract all available information about interatomic distances and torsion angles. In the initial stage of investigation by NMR spectroscopy each resonance must be associated with a specific nucleus in the investigated molecule. This process is called assignment. The strategies employed for

the assignment procedure depend on whether only homonuclear 2D spectra are available (unlabelled proteins), whether $^{15}N$ heteronuclear spectra are available ($^{15}N$ labelled proteins) or whether triple resonance spectra ($^{15}N/^{13}C$ doubly labelled proteins) are available. But in general the assignment can be divided in two parts: The sequential assignment of the amino acids in the protein sequence and the assignment of the amino acid side chains.

Protein structure determination by NMR has three steps which involves spin system assignment. The first stage is the assignment of backbone resonances to sequential residues. Identify resonances for each amino acid then put the assigned resonances in order according to their amino acid sequence i.e. sequential assignment and do the sequence-specific assignment. Initially a hetero-nuclear single quantum correlation (HSQC) spectrum is recorded. This correlates the backbone amide protons and nitrogen. NMR spectrum each resonance is associated with a specific nucleus in the molecule under investigation.

**1.6 Aims of this project:**

The main aim of this project is structural characterization of the Par-4 SAC domain protein by NMR spectroscopy. To express the Par-4 SAC domain protein in suitable cloning vector, isotope labelling and purification for further characterization by different techniques. Since prostate apoptosis response-4 (Par-4), is a human gene coding for a tumour-suppressor protein that induces apoptosis in cancer cells, but not in normal cells. The ability of the Par-4 antisense or dominant-negative mutant to abrogate apoptosis in neurodegenerative disease paradigms makes it an appealing candidate for molecular therapy of cancer and neuronal diseases.

Par-4 selectively induces apoptosis in cancer cells and can shrink solid tumours in animals; it is being developed for molecular therapy of cancer. The structural study of the Par-4 protein is important which will be helpful to design the drugs used for medical treatment. The techniques used for protein characterization are NMR spectroscopy, dynamic light scattering (DLS) and circular dichroism (CD) spectropolarimetry.

# CHAPTER 2

# CLONING OF THE Par-4 SAC DOMAIN

## 2.1 Experimental: General Cloning Methods:

### *Cloning, Ligation and Plasmid Purification*

Synthetic oligonucleotide primers were used to amplify the template Par-4 SAC domain. DNA amplification by PCR reaction in 0.5ml microfuge tube: Add 2µl of template Par-4SAC (1ng/µl), 0.5µl BamHI forward primer, 0.5µl HindIII reverse primer, 2.5µl 10x buffer, 0.5µl dNTP (deoxyribonucleoside triphosphates), 0.20µl *Pyrococcus woesei* (*PWO*) DNA polymerase (add after 5 min of first step) and make up to 25µl with $H_2O$. PCR reaction was carried out using Eppendorf thermal cycler.

Double digestion reaction ~1 µg of DNA plasmid (40ng/µl) was double digested with 2 units of BamHI and HindIII, 3 µl of cut buffer in a total volume of 30 µl and kept at $37^0$C for overnight. ~100ng of DNA plasmid was single digested with 2 units of BamHI and HindIII, 0.5 µl cut buffer in a total volume of 5 µl kept at $37^0$C for overnight. The plasmid cut was determined by running agarose gel electrophoresis as described below. The double reaction was purified with QIAGEN gel/PCR kit. 90 µl of QG buffer and 30 µl of isopropyl alcohol was added to the double digest reaction mix and spindown. The mixture was added to a QIAquick column assembly, centrifuge at 14000 rpm for 1 minute and discard the flow through. 250 µl of PE buffer was added to the column incubated for 3 minute at room temperature. The column tube was centrifuged at 14000 rpm for 1 minute and flow through was discarded. The plasmid DNA was eluted with 30 µl of elution buffer in sterile 1.5mL Eppendorf tube at 14000 rpm for 1 minute. Pure cut plasmid DNA was stored at-$20^0$C for further use.

### *Ligation reaction of DNA Plasmid and insert*

The ligation reaction explained here is used for all other plasmids. The double digested pET-TEV plasmid as described in general methods. The ligation reaction consisted of ~30ng of digested pET-TEV plasmid 1.0 µl, 5ng of insert 1.0 µl, 1 unit of

T4 DNA ligase (Roche) 1 µl, 3 µl of ligase buffer made up to a total volume of 30 µl with sterile water. The reaction was incubated overnight at $16^0$C using an Eppendorf thermal Cycle instrument.

***Agarose gel electrophoresis:***

DNA fragments were separated by size using agarose gel electrophoresis. 0.8%(w/v) agarose gels were prepared by adding 0.24g agarose in 27ml $H_2O$, and 3ml 10x TAE(40mM Tris base, 20mM acetic acid and 2mM EDTA at pH 8.0). The agarose solution was dissolved by heating in microwave. The clear solution of agarose was poured in flask containing 3.6µl ethidium bromide then whole solution was poured on a gel casting slab kept ready with a ten slot comb, air bubbles were removed by tip of micropipette. The gel was kept for settling for half an hour. DNA sample were premixed with 6x DNA loading buffer (0.2% (w/v) bromophenol blue in 50% (v/v) glycerol) and loaded into the wells. Standard marker 1kbp 4µl was loaded. Electrophoresis was carried out using Agarose gel Electrophoresis system (Bio-Rad) in 1x TAE buffer at 150V until the dye front had migrated ¾ of the length of the gel. The DNA fragments were visualized on the gels by exposure to ultraviolet light (302nm). Pictures of the gels were taken using an Alpha Imager gel documentation system (Alpha Innotech Corporation, USA).

***Purification of DNA from agarose gel***

DNA fragments were purified from agarose gel using QIAGEN gel extraction kit. This consisted of cutting out the DNA band from the gel and dissolving it in 3x of buffer QG to 1 volume of gel (100mg~100µl). The solution incubated at $50^0$C for 10 minutes and mixed by vortexing the tube every 2-3 minute during the incubation. After gel slice has dissolved completely the color of the mixture has changed to orange or violet add 10µl of 3M sodium acetate pH 5.0 and mix. Add one gel volume of isopropyl alcohol to the sample and mix. Take a QIAquick spin column with a 2ml collection tube and add 700µl of the sample mixture to the column centrifuge for 1 min with table top centrifuge to bind DNA. To wash the DNA sample add 750µl of PE buffer incubate for 3 minute at room temperature and centrifuge for 1 minute discard the flow through, spin the column again for drying. The flow through was

discarded and the DNA was eluted from the filter tube by adding 30µl of elution buffer (10mM Tris-Cl, pH 8.5) incubate for 1 minute and centrifuge at 14000rpm for 60 seconds. The DNA was collected into a sterile 1.5ml Eppendorf tube and stored at $-20^0$C.

***Transformation of DNA fragments into E. coli XL1-blue cells:***

Competent XL-1 blue cells from the $-80^0$C stock were used. Thaw the 500µl of the competent XL1-blue cells on ice for 30 min. 10 µl of the relevant plasmid DNA was mixed with the XL1- blue cells and incubated on ice for 30 min. The cells were heat shocked for 90 seconds at $42^0$C in a water bath and then incubated on ice for 2 minutes. Then 1ml of SOC (20g Tryptone, 0.5g yeast extract, 0.5g NaCl, 1M KCl 2.5ml in one litre distilled water pH=7.0 was added autoclaved and the sample 20ml sterile 1M glucose was added just before use) medium was added and incubated at $37^0$C shaker for one hour. The cells were plated on LB agar plates containing 100 $mg.mL^{-1}$ ampicillin, 34 $mg.mL^{-1}$ chloramphenicol and incubated for 12-16 hours at $37^0$C. Glycerol stocks of cells were prepared by growing colony from the plate in 25 mL LB (25 µl amp and camp) at $37^0$C till an optical density at 600 nm ($OD_{600}$) of ~0.6 was achieved. The 10ml of inoculated LB was taken out and centrifuged at 4000 rpm for 15 minutes, supernatant from is thrown out and the cell pellet was resuspended in 250 µl of Glycerol and 750 µl LB. The cells were stored in $-80^0$C for future use.

***Preparation of plasmid DNA***

Plasmid DNA was prepared using a High Pure Plasmid Isolation Kit (Roche). This consisted of spinning down 4 ml of cells at 4000 rpm for 60 seconds; the cells had been grown for 12-16 hours at $37^0$C in 5ml LB media containing 5 µl amp and camp. The cells were resuspended in 250 µl of suspension buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, 0.1 mg $mL^{-1}$ RNase A) and 250 µl of lysis buffer (0.2 N NaOH, 1% (w/v) SDS) was added. The mixture was incubated at room temperature for 5 minutes, then 350 µl of chilled binding buffer (4 M guanidine hydrochloride, 0.5 M potassium acetate pH 4.2) was added and the mixture was incubated on ice for 10 minutes. The precipitate was spun down for 10 minutes at 14000 rpm and the supernatant was

added to a filter tube. The tube was centrifuged at 14000 rpm for 60 seconds. 700 µl of washing buffer (2 mM Tris-HCl pH 7.5, 20mM NaCl, 80%(v/v) ethanol) was added and the tube was centrifuged at 14000 rpm for 60 seconds. The flowthrough was discarded and the plasmid DNA was eluted from the filter tube by adding 50 µl of elution buffer and centrifuged into a sterile 1.5mL Eppendorf tube at 14000 rpm for 1 minute. Pure plasmid DNA was stored at -20$^0$C.

***Quantification and Size determination of DNA fragments***

The concentration of a DNA sample was estimated by using nanodrop device. The approximate size of DNA band was estimated by comparing its migration through the agarose gel against that of DNA standards with a known size (1 Kb plus DNA ladder, Invitrogen) which were run alongside the DNA sample of interest.

***DNA sequencing***

DNA sequencing services were provided by the Allan Wilson Centre for Molecular Evolution and Ecology Genome Service at Massey University. DNA sequencing was carried out on either an ABI prism 377-64 sequencer or an ABI Prism 3730 capillary sequencer using BIGDYE labelled dideoxy chain termination chemistries (Applied Biosystems). Sequencing reactions contained at least 300ng DNA and 3.2 pmol of primer (lac operator forward, T7 terminator reverse or forward and reverse primer of corresponding plasmid) in total of 15 µl in sterile Milli-Q water.

**2.2 Cloning of the Par-4 SAC in different vectors:**

**2.2a Cloning with H-MBP-3C vector:**

The novel H-MBP-3C expression vector created by modifying the pMAL-c2X vector[33]. The modification include the insertion of His tag at the N-terminus of MBP and a 3C protease site between the EcoRI and BamHI sites of multiple cloning sites (MCS). This vector has got maltose-binding-protein (MBP) fusion which keeps protein in soluble form and His tag for affinity purification (see Figure 2.1). A cleavage site from human rhinovirus protease [3C protease from HRV-14][34,35]

provides highly specific fusion cleavage even in the case of unfolded or partially folded proteins. The protein sequence code for HRV-14 3C protease site LEVLFQGP. The Par-4 SAC domain 180 nucleotides long which encodes protein product of 59 amino-acids. The Par-4 SAC gene sequence was inserted between BamHI and HindIII restriction enzyme sites.

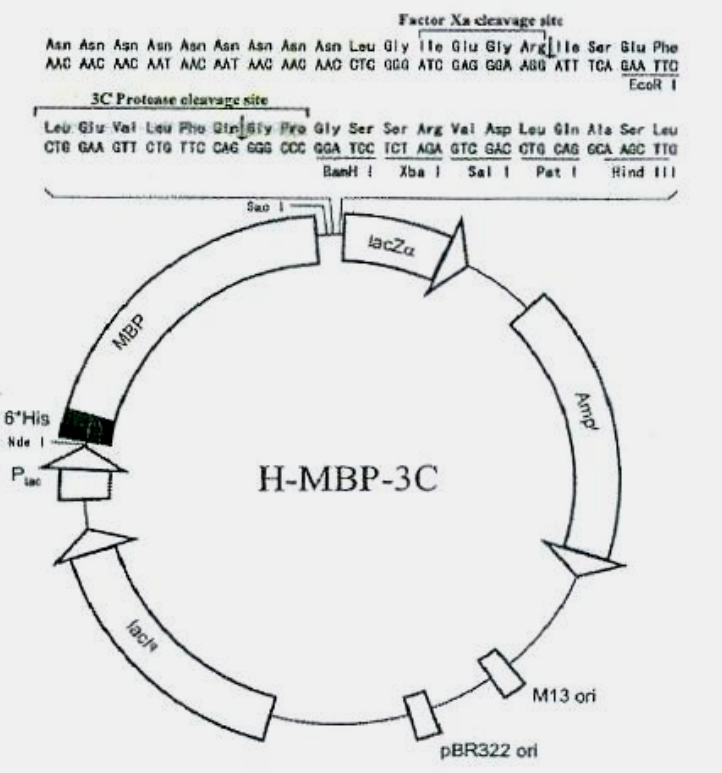


**Figure-2.1: H-MBP-3C vector map 6800bp Important modifications include insertion of a His tag at the N-terminus of MBP and a 3C protease site between *Eco*RI and *Bam*HI sites of the MCS. The MBP fusion increase solubility, and the His tag and 3C site optimize purification and fusion cleavage.[33]**

The rrPar-4SAC (residues 137–195) construct representing the SAC domain was cloned using the primers 5′GAGGATCCAGGAAAGGCAAAGGGCAGATCG-3′ and 5′GCAAGCTTTTATGCTTCATTCTGGATGGTG-3′. The rrPar-4 expression vectors were used to transform *E. Coli* Rosetta (DE3) cells (Novagen). The more details of expression vector construction Par-4 is covered in chapter-6 experimental procedures.

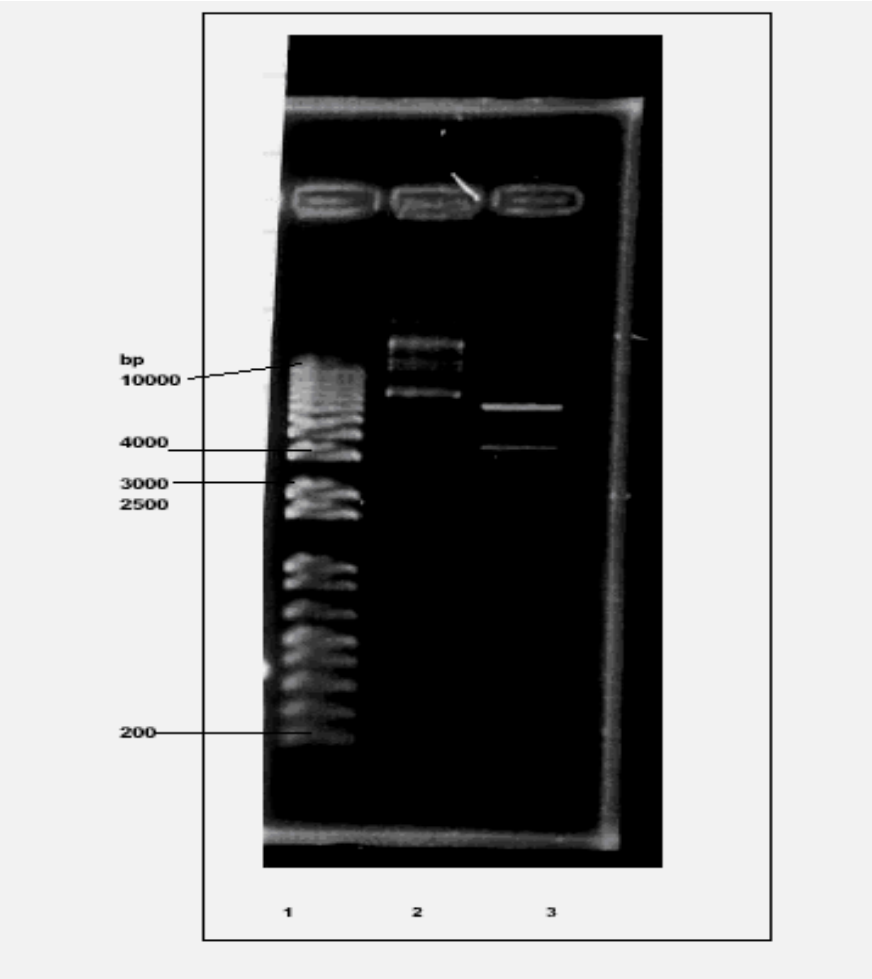


**Figure-2.2: Agarose gel showing cloning of the Par-4 SAC domain in H-MBP-3C vector. Lane 1 2 3 shows marker MassRuler DNA ladder, HindIII digested vector, BamHI digested vector.**

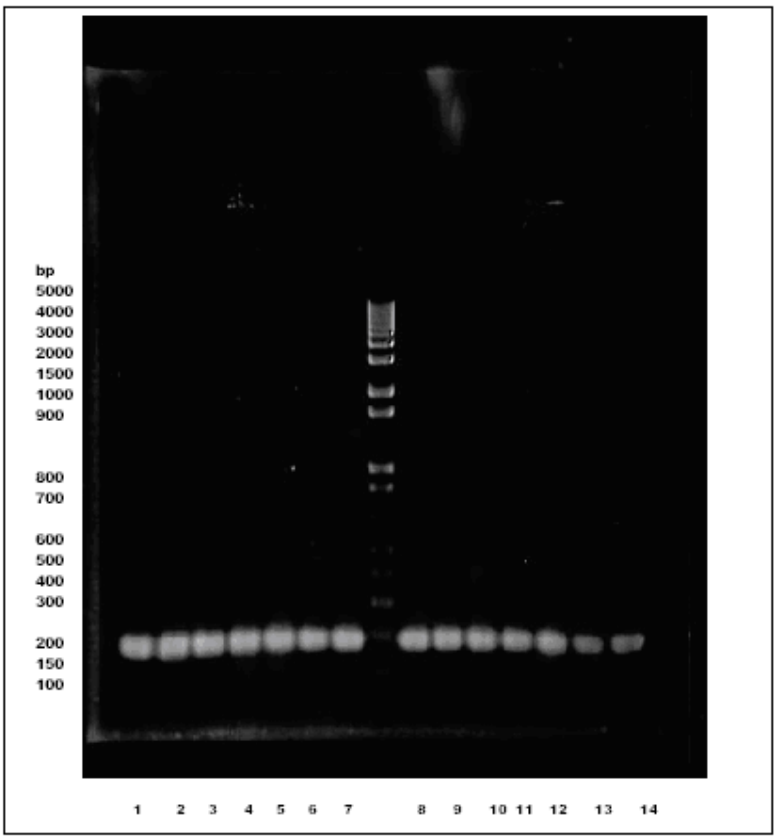


**Figure-2.3: Agarose gel showing cloning of the Par-4 SAC domain in H-MBP-3C vector. Lane 1 to 7 and 9 to 12 for 200 base pairs for colonies grown on the LB + amp agar plate; 13 and 14 control.**

**2.2b Cloning of the Par-4 SAC with pProEXHTb vector:**

The pProEXHTb Prokaryotic Expression system is 4779bp designed for the expression of foreign proteins in *E. coli*. The protein is expressed fused to a 6 histidine sequence $(His)_6$ for affinity purification. The gene of interest is cloned into the multiple cloning site (MCS) of pProExHTb vector (see Figure-2.4). Upon expression, the histidine sequence is at the amino terminus of the fusion protein. The histidine sequence has a strong affinity for Ni-NTA resin matrix, making it simple to purify the desired protein. The vector also encodes the sequence for the Tobacco Etch Virus (TEV) protease cleavage site. rTEV Protease can be used to remove the histidine from fusion protein. The pProExHTb vectors contain a 6x histidine affinity tag for ease of purification, a 7-amino acid spacer arm, the TEV protease recognition

site for cleavage of the 6x histidine affinity tag. The trc promoter and lacI gene enable inducible expression of a cloned gene with IPTG. The plasmid also contain the pBR322 origin of replication, the b-lactamase gene conferring ampicillin resistance ($Ap^r$), and the bacteriophage F1 origin of replication (F1 intergenic region). The Par-4(aa137-195)WT sequence is inserted between BamHI and HindIII of the MCS.

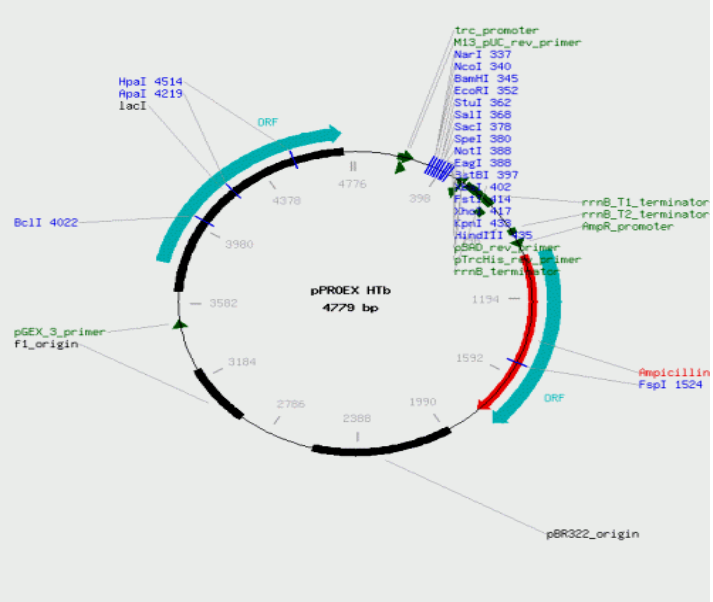


**Figure-2.4 pProEXHTb vector map**

The pProExHTb vector has been double digested with BamHI and HindIII restriction enzymes. The Figure-2.5 shows the corresponding double digestion and single digestion of the vector. The double digested Par-4 SAC domain has been ligated with the pProExHTb and transformed in to the *E. coli* Top10 Cells. The cells were grown on LB agar +amp (ampicillin) plate.

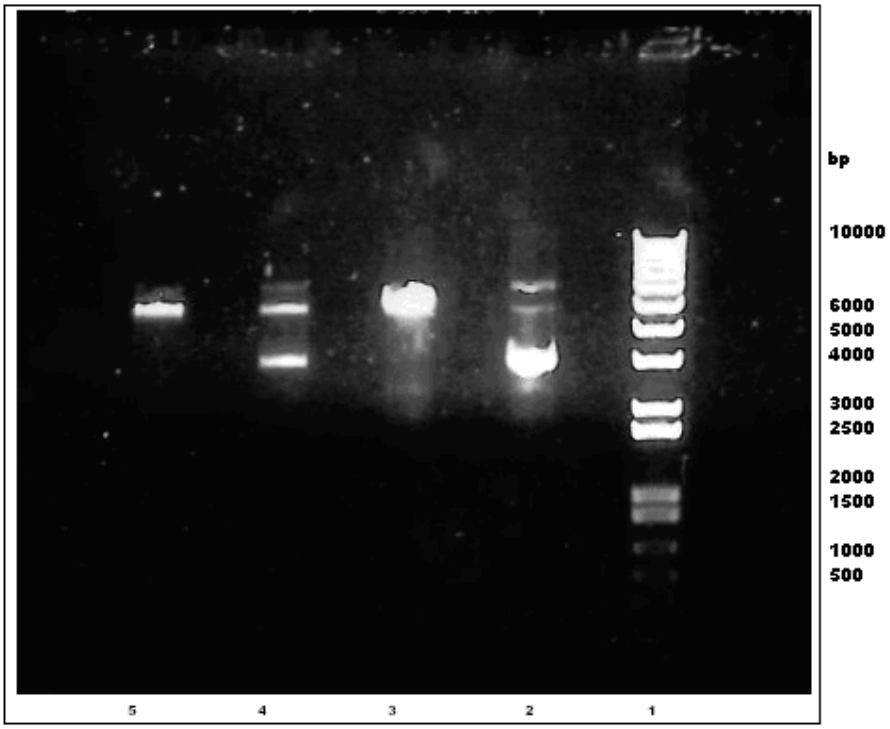


**Figure-2.5: Agarose gel analysis showing pPROEX-HTb showing lane 1 2 3 4 5 (Marker, pPROEXHTb, Double digestion, Single Digestion BamHI showing double bands caused by incomplete digestion, Single digestion HindIII respectively).**

**2.1c Cloning of the Par-4 SAC Domain with pET-TEV vector:**

The pET-TEV vector is a modified vector of pET32a. The pET32a series is designed for cloning and high level expression of peptide sequences fused with the 109aa Trx-Tag thioredoxin protein (see Figure-2.6)[36]. Cloning site is available for producing fusion proteins containing cleavable His.Tag and S.Tag sequences for detection and purification. The vector has got thioredoxin tag alongwith spacer region, His tag, spacer region and rTEV cleavage site. The pET32a vector was excised using MscI and BamHI [His-tag, thrombin, S-tag, enterkinase, NcoI-BamHi of the MCS]. Replaced with PCR-amplified section from pProExHTb [His-tag, spacer region, rTEV protease cleavage site]. Primers included sites for MscI (5') and BamHI (3'). The vector has got rTEV cleavage site, alongwith his tag and thioredoxin tag. The

DNA sequence for the Par-4 SAC domain was inserted between BamHI and HindIII site of multiple cloning sites.

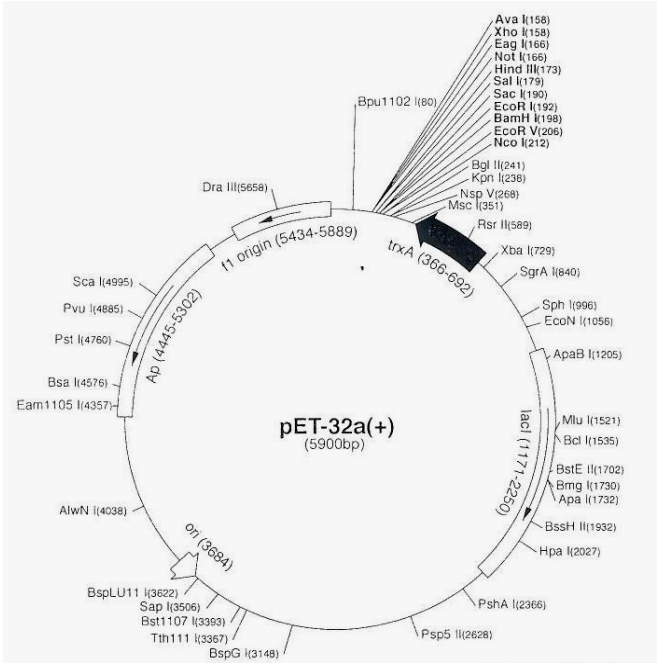


**Figure-2.6 Original pET-32a vector which was modified to make pET-TEV vector.**

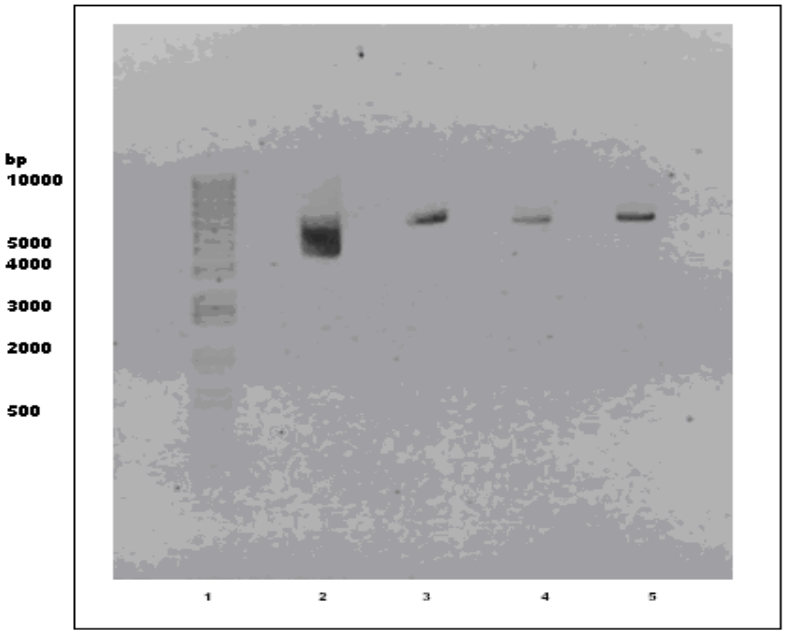


**Figure-2.7: Agarose gel analysis showing cloning of the Par-4 SAC with pET-TEV vector. Lane 1 2 3 4 5 corresponding to marker, pET-TEV plasmid, double digested plasmid, single digest BamHI and single digest HindIII**

The double digested plasmid has been ligated with the Par-4 SAC domain and transformed in *E. coli* BL21 CP cells for expression of protein. The pET-TEV vector used for expression of the Par-4 SAC protein was found to be giving good expression level and gives stable fusion protein.

# CHAPTER 3

## Expression And Purification of The Par-4 SAC Protein

### 3.1 Expression of the rrPar-4 SAC domain (aa137-195) with H-MBP-3C vector:

The Par4SAC in H-MBP-3C vector transformed in *E. coli* Rosetta cells. The protein expression was done in 100ml preculture Luria-Bertani (LB) media containing 100µl of ampicillin antibiotic (100µgml$^{-1}$). The cells were grown at 37$^0$C for overnight and the main culture was inoculated with the preculture. The main culture 500ml inoculated with 500µl ampicillin (100µgml$^{-1}$) was grown overnight with the cells induced with 500µl of 400mM IPTG at 0.6 $OD_{600}$. The cell culture was grown overnight and cells were harvested by centrifugation at 3800 rpm (Sorvall Evolution RC, Rotor GS3) for 20 min. The cell pellet was stored in -80$^0$C refrigerator. The cleavage reaction was performed at 4 ºC for at least 24 h. The cell pellet was resuspended in 15ml lysis buffer PBS of pH 7.5, the cells were lysed by using French Press three times. The cells after lysis were centrifuged at 15k rpm for 20min. Afterwards the supernatant was filtered through a 0.8 µm Filter.

Purification was done by using immobilized metal ion affinity chromatography (IMAC). The Ni-NTA Sepharose was prepared one day prior to purification. The supernatant solution was nutated with Ni-NTA column for overnight at 4$^0$C for His tag binding purpose to column. The column was washed two times column volume with 10ml 25mM imidazole, PBS of pH 7.5 wash buffer. The column was again washed with 10ml of 50mM imidazole, PBS of pH 7.5 wash buffer. The fusion protein was eluted with 250mM imidazole, PBS of pH 7.5 buffer. The eluted solution was buffer exchanged with 20mM $Na_2HPO_4$, 20mM $NaH_2PO_4$, 1mM EDTA, 1mM DTT buffer of pH 7.0 using 10 kDa cutoff Vivaspin concentrators (Vivascience). The solution was concentrated to ~1ml and the fusion protein was kept for cleavage reaction with HRV-14 3C protease at 4$^0$C overnight. The cleavage reaction was purified by nutating the 500µl of protein with Ni-NTA column to remove the His tagged MBP and 500µl with amylose resin. The cleaved protein was eluted with PBS

buffer of pH 7.5, the fusion protein and cleaved protein can be seen in the SDS-PAGE (Figure-3.1). The cleaved off Par-4SAC comprised the last 59 amino acids plus a four amino acids long cloning artifact at the N-terminus. As seen in figure 3.1 and 3.2 there is not much difference in the before cleavage and after cleavage Par-4 SAC fusion protein. It was observed that fusion protein was not stable, within a week the fusion protein started degrading. The MBP tag can be seen around 50 kDa. The protein before cleavage and after cleavage can be seen in SDS-PAGE and Tricine PAGE (see Figure-3.1 and Figure 3.2) respectively.

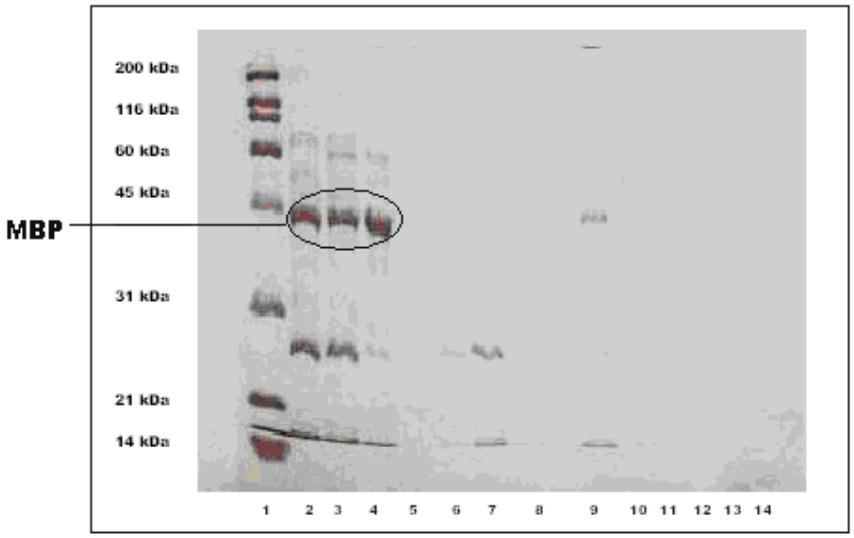


**Figure-3.1: SDS-PAGE with Coomassie BR250 staining. Lane: 2-before cleavage, 3-after cleavage, 4-amylose column beads, 5-8 eluent from amylose column, 9-Ni-NTA column beads, 10-14 eluent from Ni-NTA column.**

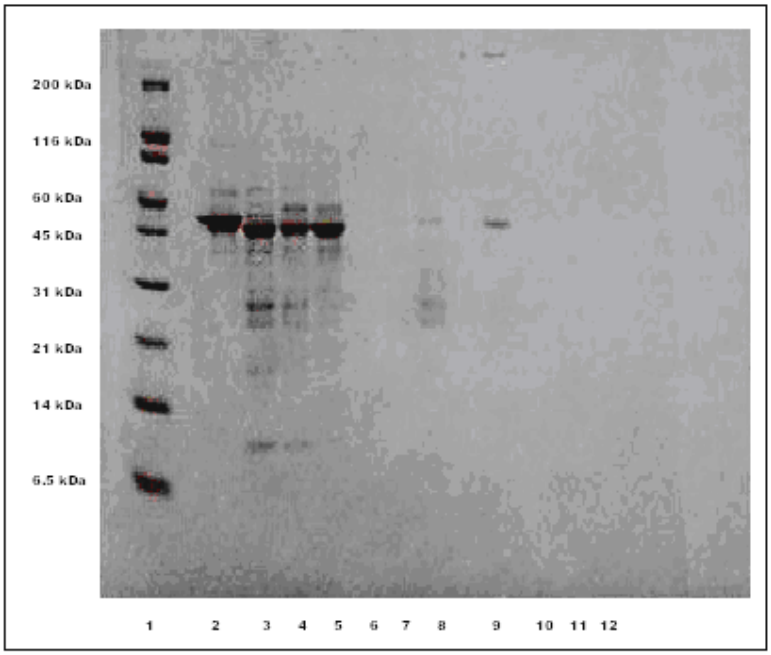


**Figure-3.2: 10% Tris-Tricine PAGE with silver staining. Lane: 2,3-before cleavage, 4-after cleavage, 5-amylose column beads, 6-8 amylose column elute, 9-Ni-NTA column beads, 10-12 Ni-NTA column elutes.**

Since the protein was not stable with the MBP vector we decided to try different expression vectors with fusion tag and without fusion tag for expression of stable rrPar-4 SAC domain protein.

### 3.2 Expression of the Par-4 SAC Domain with pProEXHTb vector:

The pProEXHTb vector was used for expression of the Par-4 SAC domain protein. The DNA fragment was inserted between BamHI and HindIII restriction site. The expression vector did not have any fusion tag for protein expression only his tag for affinity purification. The pProEXHTb vector was transformed in *E. coli* BL21 CP cells. The 100ml preculture LB media containing ampicillin and chloramphenicol was inoculated with one colony from the glycerol stock -80$^0$C. The cells were grown overnight at 37$^0$C in PC. The main culture was inoculated with PC and grown for

overnight. The expression level of the vector was checked by running SDS PAGE with before induction and after induction samples. The protein expression level after induction for six hours and overnight is too low as seen from the SDS PAGE.

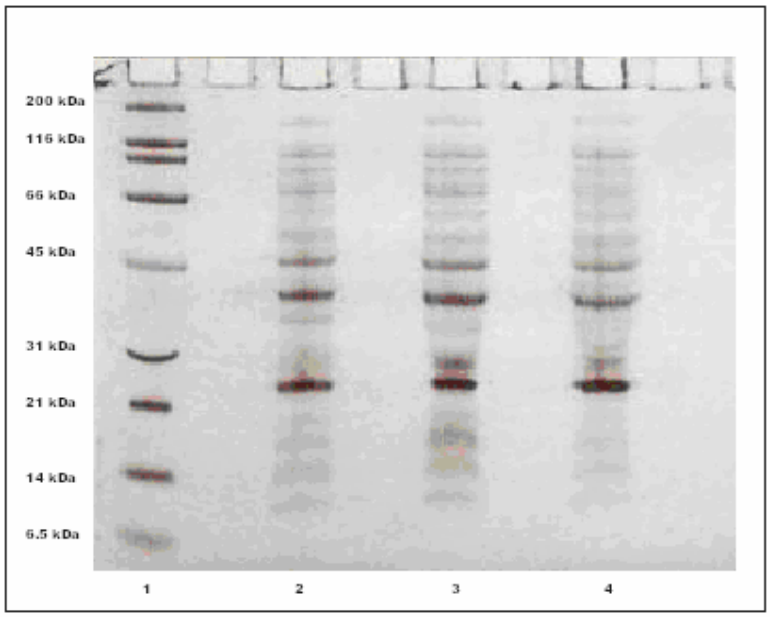


**Figure-3.3: Time control 10% Tris Tricine-PAGE gel showing before induction and after induction samples with Coomassie BR250 staining. Lane: 2-supernatant before induction, 3-six hours supernatant after induction, 4-eighteen hours supernatant after induction.**

Since the protein expression level was too low the further expression with this vector was not pursued.

### 3.3 Expression of the Par-4 SAC (137-195) Domain with pET-TEV vector:

The protein was expressed in *E.coli* BL21 CP cells. The 100ml LB preculture (PC) supplemented with 100µg/ml ampicillin and 34µg/ml chloramphenicol was inoculated with the glycerol stock of the plasmid containing the Par-4 SAC (aa137-195) domain. The PC was grown at $37^0$C on shaker for overnight. The main culture of 300ml LB supplemented with 100 µg/ml ampicillin and 34µg/ml chloramphenicol was then inoculated with 2ml of PC and grown continuously at $37^0$C. The cells were induced

with 300µl of 400mM IPTG at $OD_{600}$ of 0.6 and kept for overnight growth at $25^0$C. The time control samples to check the growth of cells in LB were collected at every two hours interval. The time control SDS PAGE analysis shown in figure-3.4. The cells were harvested after overnight growth and cell pellet was stored at $-80^0$C for further use.

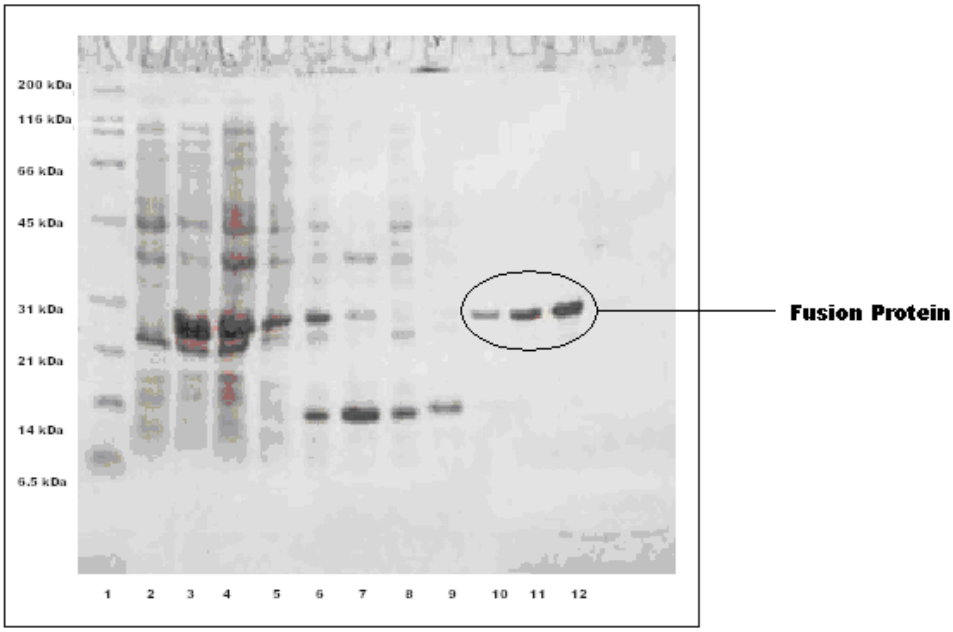


**Figure-3.4: 10% Resolving gel+ 4% Stacking gel SDS-PAGE with Coomassie BR250 staining showing expression of the Par-4 SAC with pET-Trx-rTEV vector. Lane: 2-before induction, 3-after induction six hours, 4-after induction eighteen hours, 5-lysate, 6-lysate supernatant, 7-pellet, 8-flow through, 9-wash1, 10-before eluting the column, 11-elution, 12-after eluting the column.**

As shown in Figure-3.4 the expression level of the Par-4 SAC protein with the pET-TEV vector after induction went well. The lane 3-4 shows good level of protein expression. The harvesting of cells was done as mentioned in chapter six. The lysis buffer used (50mM NaPhosphate, 25mM Imidazole, 10mM NaCl, 10% Glycerin pH = 8.0). The lane 10-12 shows the fusion protein on Ni-NTA column, in the elution buffer (50mM NaPhosphate, 500mM Imidazole, 10mM NaCl, 10% Glycerin pH = 8.0). The remaining fusion protein can be seen on Ni-NTA column was eluted by loading column with another 10ml elution buffer. The 10ml elution volume was

further concentrated with 10 kDa cutoff Vivaspin concentrators (Vivascience) and buffer exchanged with buffer (50mM Tris HCl, 0.5mM EDTA, 1mM DTT pH= 8.0) to a volume of 2-4ml. The one ball of rTEV protease was added and the cleavage reaction was kept at room temperature for four hours. The SDS-PAGE analysis with before cleavage and after cleavage sample (see Figure 3.5). The fusion protein can be seen between 21-31 kDa.

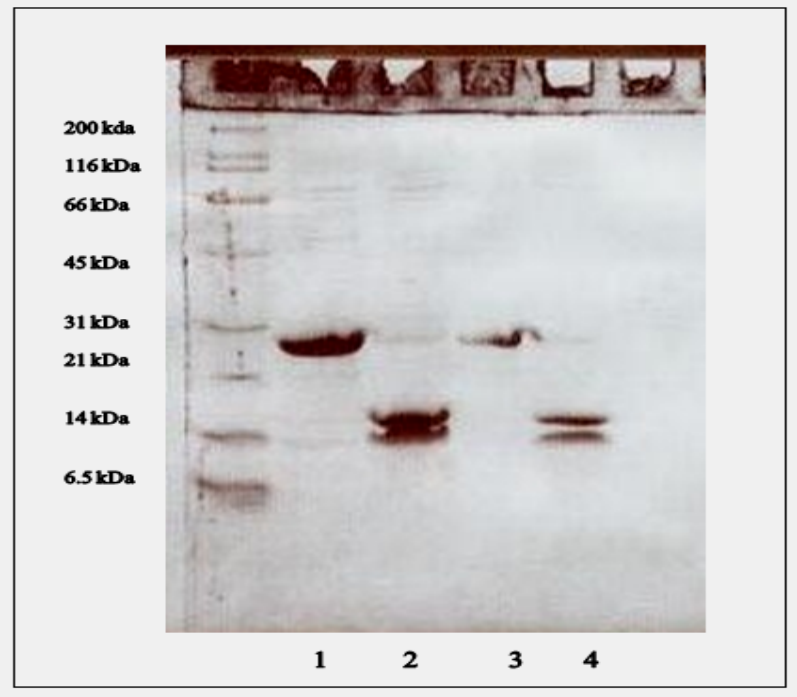


**Figure-3.5: SDS-PAGE 10% resolving gel + 4% stacking gel. Lane: 1- 5µl before cleavage sample, 2- 5µl after cleavage sample, 3- 3µl before cleavage sample, 4- 3µl after cleavage sample.**

The cleavage reaction was dialysed with 500ml of buffer (20mM Tris-HCl, 50mM NaCl, 0.3mM TCEP pH= 8.0) for overnight. The cleavage reaction of fusion protein was purified on 1ml Ni-NTA column. The thiredoxin (Trx) tag remains on the column and flow through (FT) gives purified protein the remaining protein on column was eluted with 5ml of buffer (20mM Tris-HCl, 50mM NaCl pH=8.0). The Tricine PAGE analysis shows (see Figure-3.6) fusion protein in the flow through (FT) of cleavage reaction therefore cleavage reaction along with all elutes was further purified with another 2ml Ni-NTA column. All solutions were mixed and nutated with Ni-NTA for overnight to get the good affinity.

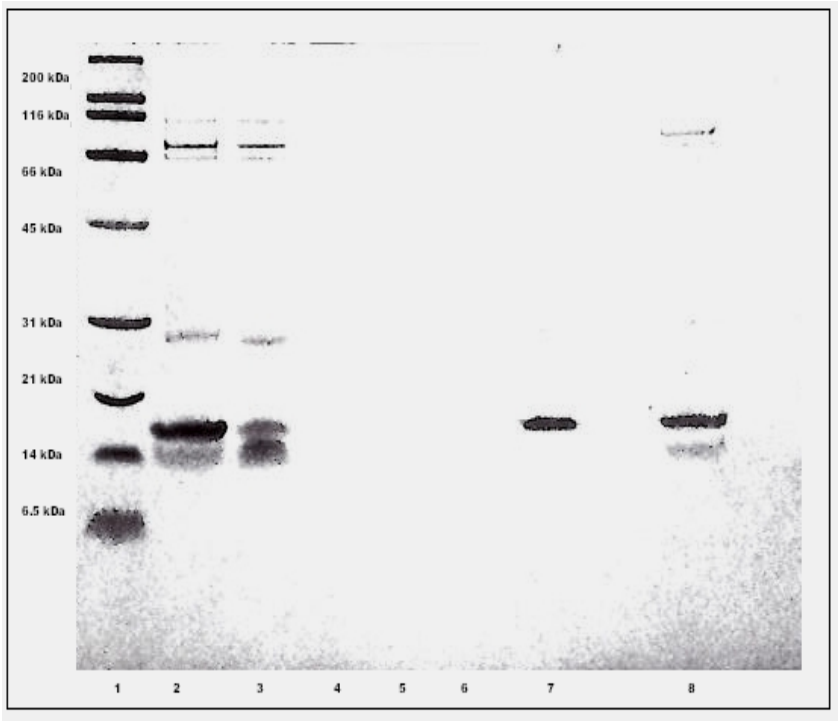


**Figure-3.6: Tricine-PAGE. Lane: 2- FT 5µl sample, 3-6 elution samples, 7- after elution column beads, 8- FT 3µl sample.**

The nutated mixture of Ni-NTA resin was reloaded on the column tube and the flow through (FT) was again collected. The remaining protein was purified with 5ml Tris-HCl buffer and collected in 1.5ml aliquots. The Tricine page analysis showed (Figure 3.7) the elutes lane 2 to 5 showing protein around 14 kDa the actual MW of the Par-4SAC protein is 7 kDa and thioredoxin tag remaining on the column lane 6.

The thioredoxin (Trx) tag runs on tricine page around 14 kDa because the fusion protein is expressed as Trx followed by a 10 residue linker region followed by 6 histidines which forms the His tag followed by 14 residue linker at whose end is the consensus sequence for the rtev to recognise. Therefore after cleavage around 29 residues is still left with the cleaved of thioredoxin tag that is why it runs higher.

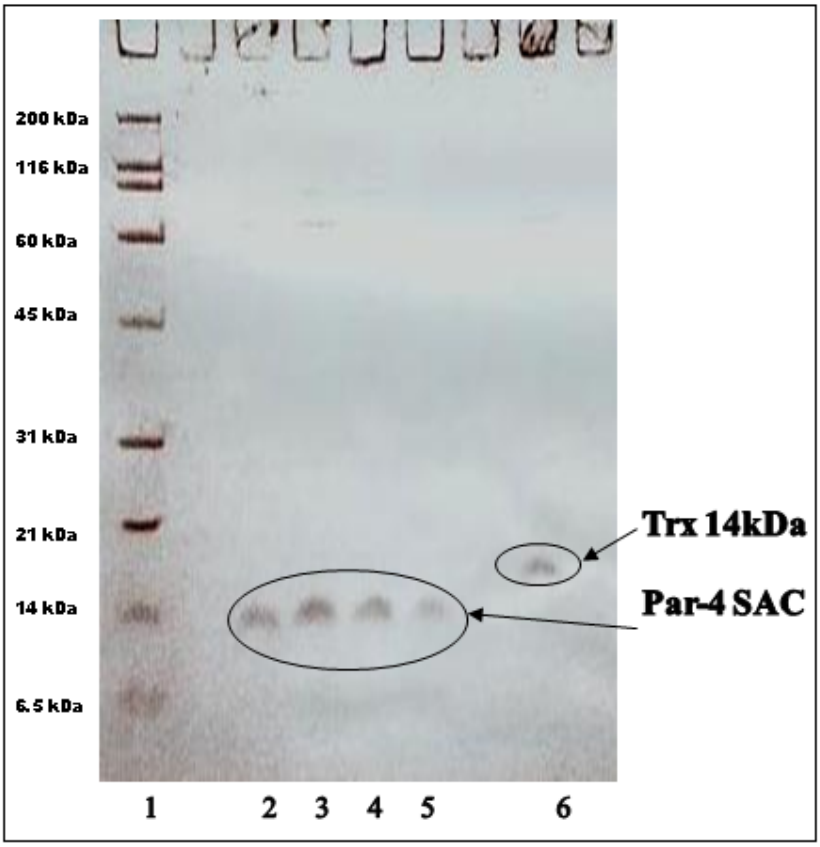


**Figure-3.7: Tricine-PAGE. Lane: 2- 5µl FT sample, 3-5 5µl elution samples, 6-5µl after eluting the column beads shows thioredoxin tag running above 14 kDa. Thioredoxin running above the apparent MW 12kDa is because the vector has got thioredoxin tag along with spacer region, His tag, spacer region and then rTEV cleavage site.**

The FT and elution sample were added together and concentrated with 3 kDa cutoff Vivaspin concentrators (Vivascience). The protein was stored at $4^0$C for further purification by gel filtration.

The protein was further purified with SEC (Size exclusion chromatography). The Par-4 SAC protein was filtered through 0.2 µm filter. Two separate run of the Par-4 SAC protein 100µl and 300 µl was loaded in the sample loop of a low pressure gel filtration unit with Superdex 75 10/300 GL column bed dimensions 10 x 300 mm with recommended sample 250 µl (Amersham) was used with AKTA Prime Plus Amersham Biosciences. All SEC was performed using 0.2 µm filtered buffer

20mMNaP, 100mM NaCl, 1mM DTT pH= 7.5 at a flow rate of 0.5 mL $min^{-1}$ for 60 minutes. The total number of peaks observed at retention time of peak-1 15min-25min, peak-2 26min-31.5min, peak-3 31.5-37.5 and peak-4 38min-42min. All elutes were collected in different tubes. The peak-1 to peak-4 was concentrated separately with 3 kDa cutoff Vivaspin concentrators (Vivascience) to the volume of 500µl.

The Tricine page analysis shows (Figure-3.8) presence of protein from SEC peak-3 of retention time $t_r$ = 31.5-37.5.

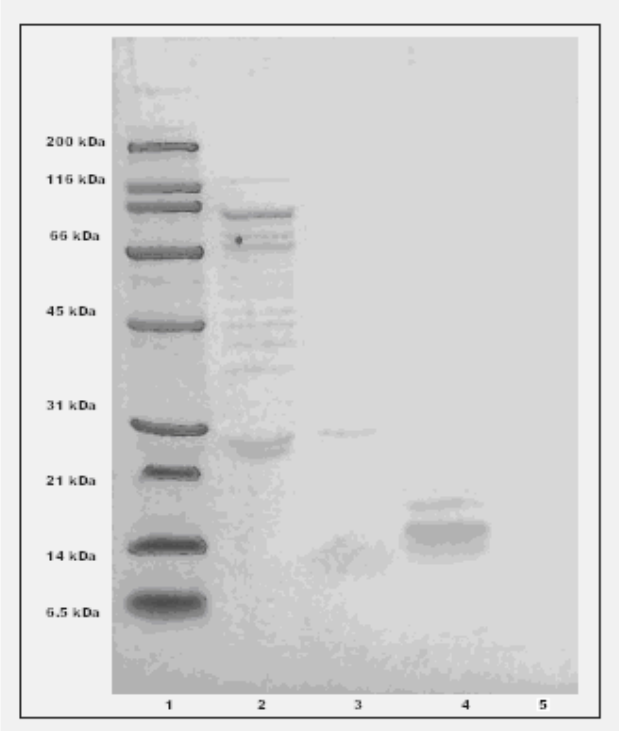


**Figure-3.8: Tricine-PAGE. Lane: 2- 5µl Peak-1 sample, 3- 5µl Peak-2, 4- 5µl Peak-3, 5- 5µl Peak-4.**

The protein purity has been checked by Tricine page analysis with silver staining (Figure 3.9) by loading different concentration of protein sample. The lane 2 to 5 shows different concentration gradient of protein along with impurities. The further characterization of protein by NMR spectroscopy and DLS (Dynamic light scatterning) needs highly pure sample. For further purification of protein sample reversed-phase high performance liquid chromatography (RP-HPLC) was used. RP-HPLC is a separation of molecules based on their hydrophobicity. A Dionex HPLC

system used with Phenomenex C18 column of 250 x 4.6mm and particle size of 5µm, column oven temperature was kept $25^0$C. The DAD UV detector was set to 205nm and 280nm. The mobile phase with gradient elution of solvent-A (0.1% TFA in water) and solvent-B (0.08% TFA in acetonitrile). The first gradient elution method used with 0-60% B for 60min at flow rate of 1ml/min. In order to elute protein of interest early the gradient for solvent B was changed to 0-25% for 25 min and then changed to 100% at 40min.

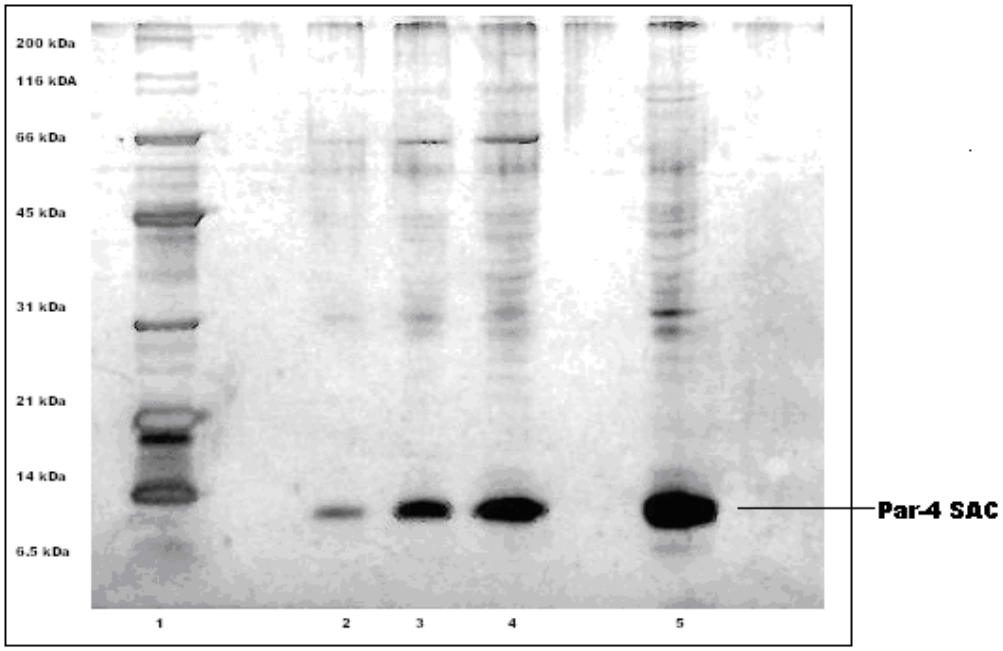


**Figure-3.9: Tricine-PAGE with silver staining. Lane: 2- 1µl of protein sample, 3- 2.5µl of protein sample, 4- 5µl of protein sample, 5- 7.5µl of protein sample.**

Protein was purified by RPHPLC with 250 µl injection volume and run time of 60 minute. Two peaks were observed of retention time (tr) peak-1 tr = 12.5 to 15 min and peak-2 of tr = 34.5 to 35 min. The total 1ml of protein volume was purified in this way. The Tricine page analysis shows peak-2 of protein sample. The other 1ml of impure protein was also prurified in similar way and the total aliquots of peak-2 were concentrated by speed vaccum to 500µl for eight hours. The acetonitrile from the mobile phase was removed by freeze drying for overnight. The protein sample was kept for freezing at $-80^0$C and submitted for freeze drying. The protein after freeze drying was resuspended in buffer (20mM NaP, 50mM NaCl, 1mM DTT pH = 7.5).

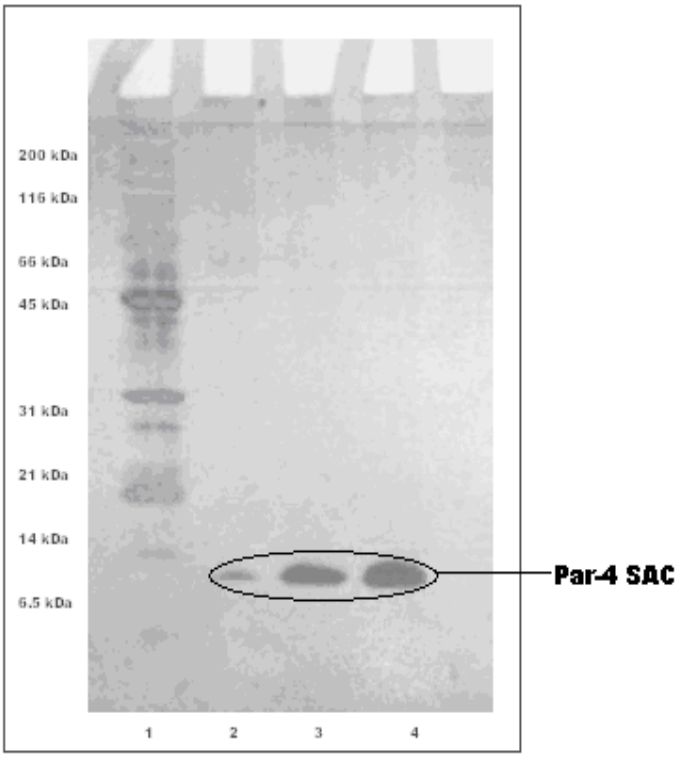


**Figure-3.10: 10% Tris PAGE of the Par-4 SAC protein samples after purifying with RP HPLC and freeze drying resuspended in 20mM NaP+50mM NaCl+1mM DTT pH = 7.5 buffer. Lane: 2 2µl protein sample, 3- 5 µl protein sample batch-1, 4- 5 µl protein sample batch-2.**

The Par-4 SAC protein after RPHPLC purification was characterized by NMR spectroscopy. The proton NMR spectra of the Par-4 SAC protein shows that protein is not folded see Figure- 4.1 in chapter four.

### 3.4 Expression of the Par-4 SAC (aa137-195) Domain with pGroESL plasmid:

To study the protein folding possibility with chaperonin we co-expressed pET-TEV with a plasmid that has previously been shown to synthesize groES and groEL (pGroESL)[37]. The pGroESL plasmid was given by Dr. Mark Patchett (Massey University, Palmerston North New Zealand). The chaperonin proteins could assist in

the assembly of [alpha] and beta subunits. The plasmid, pGroESL, utilizes two different promoters to simultaneously over-express groES and groEL. The lac promoter is induced by isopropyl-β-D-thiogalactopyrano-side (IPTG). The pGroESL plasmid chaperonin is useful for folding of the disordered protein. Molecular chaperones are a ubiquitous class of proteins that play an essential role in protein folding by helping other polypeptides reach a proper conformation or cellular location without becoming part of the final structure[38,39,40]. In *E. coli* two molecular chaperone "machines," DnaK-DnaJ-GrpE and GroEL-GroES, have been studied extensively. DnaK and its cofactors, DnJ and GrpE, have been proposed to interact with nascent polypeptides and to either directly facilitate the proper folding of the newly synthesized proteins[38,39,40] or to mediate their transfer to the "downstream" GroEL-GroES chaperonins. Overproduction of either GroEL-GroES chaperone machines has been shown to improve the cytoplasmic solubility and secretion of a number of aggregation prone heterologous proteins in *E. coli.* However, higher intracellular concentrations of molecular chaperones had little effect on the proper folding or localization of other recombinant polypeptides. The protein expression and purification was done as per the procedure discussed in 3.3. It was found that there is not much difference between the $^{1}H$ NMR spectra of protein expressed with and without pGroESL plasmid.

# CHAPTER 4

## Results and Discussion

### 4.1 Results:

The Par-4 SAC (aa137-195) protein after cleavage with rTEV protease and purification gives 62 aa 7.0 kDa protein. The relative mobility analysis of the electrophoretic profiles of the rrPar-4SAC using denaturing Tricine-PAGE analysis shown in Figure-3.9 Chapter-3 determined apparent molecular weight of 12.4 kDa. The size of the protein is significantly larger (77% larger for rrPar-4SAC) than that expected.

The pure protein characterized by $^{1}H$ NMR spectroscopy. The $^{1}H$ NMR spectra recorded in 95% buffer (20mM sodium phosphate, 50mM NaCl, 1mM DTT pH= 8.0) and 5% $D_2O$. Spectra were routinely recorded on a 500 MHz and 700 MHz Bruker Avance spectrometer using a room temperature QXI probe and Cryoprobe respectively. All NMR spectra were collected at $25^{0}C$. The spectral analysis was primarily focused on the positions of the individual NMR lines in the $^{1}H$ NMR spectrum, as given by the "chemical shift", δ, in parts per million (ppm) relative to a reference compound. The $^{1}H$ chemical shifts (CS) in Figure-4.1 show for aliphatic region of the NMR spectra corresponds to the random coil NMR CS. The CS shown in the aromatic and amide region of the proton NMR spectra also shows less CS dispersion corresponding to the random coil. The protein CS in whole spectra of proton NMR spectra corresponds to the presence of intrinsically disordered protein (IDP's) or random coil. The protein cannot be characterized further by using NMR spectroscopy due to unfolded nature of protein. The proton NMR spectra show absence of secondary structure for the protein.

**NMR pH titration of protein:**

Protein folding rates are sensitive to a wide variety of environmental conditions, including temperature, pH, buffer, ionic strength, and the concentration and nature of any residual denaturant.

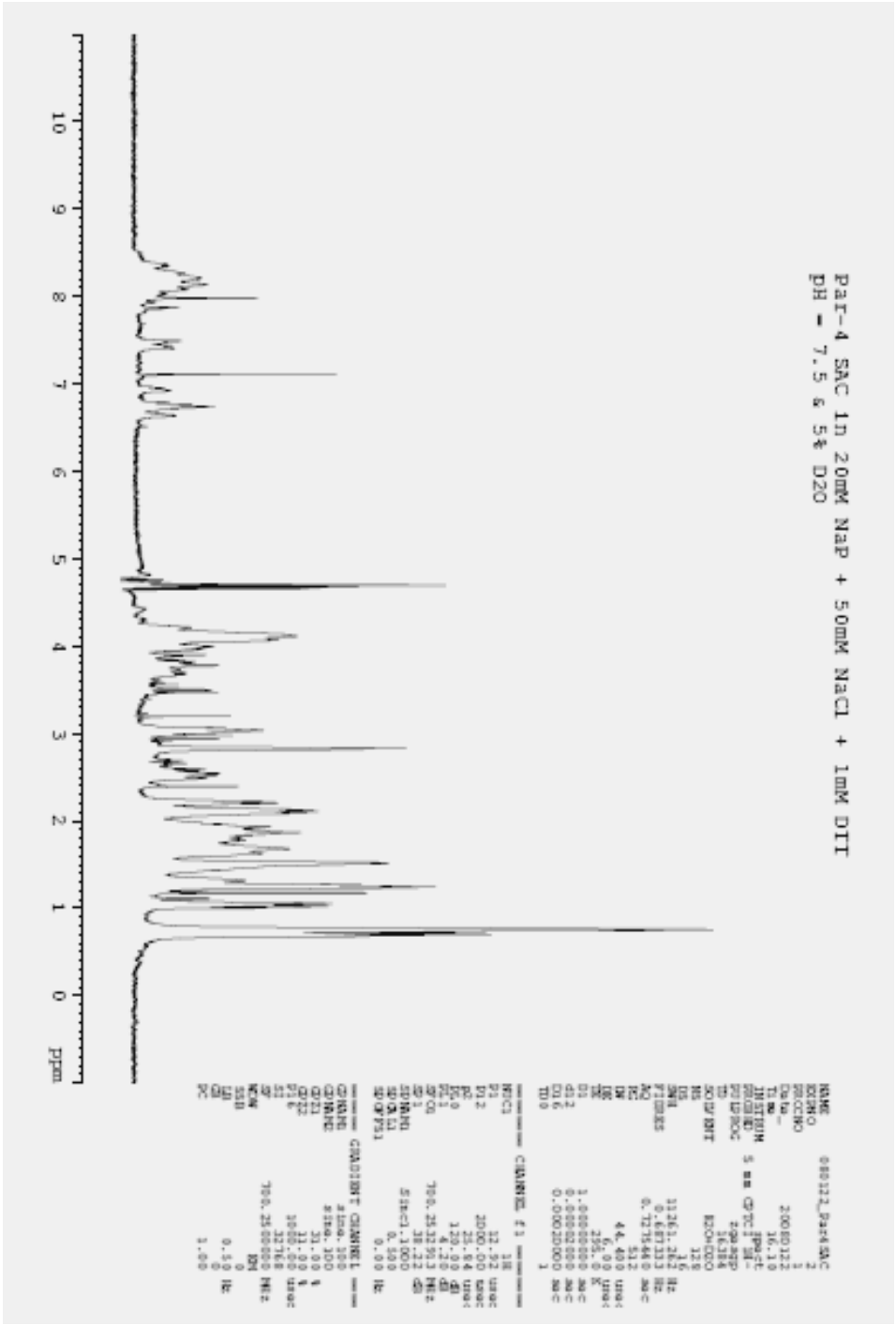


**Figure: 4.1 A one dimensional $^{1}H$ NMR spectra of the Par-4 SAC domain in 20mM phosphate buffer of pH 7.5 at $25^{0}C$. Proton chemical shifts not well resolved and expanded in amide region of the spectrum.**

The structure function of proteins is influenced by external parameters such as pH of the environment. The pH at which proteins are most stable is the optimum pH of stability and the pH at which the binding affinity of protein-protein complexes is maximal is called optimum pH of the complexes. The NMR sample used here was initially prepared by using 0.5ml of the Par-4 SAC domain protein in buffer (20mM Tris, 50mM NaCl pH = 8.0) and 5% $D_2O$. The pH of sample was adjusted by addition of small aliquots of HCl and NaOH solutions, and pH measurements were made by using Whatman pH indicator paper for acidic and basic range. The proton spectrum was recorded at each pH values on Bruker Avance 500 MHz NMR spectrometer with QXI probe at room temperature. The initial spectrum at pH 7.5 (Figure- 4.1) was identical to the final spectrum at the end of the titration down to 4.0 and back to the starting value. The spectrum showing comparison of NMR spectra at different pH values does not show presence of folded protein.

**HSQC ($^{1}$H-$^{15}$N) Heteronuclear NMR spectroscopy:**

$^{15}$N-labelled protein can be used to record the standard solution-NMR HSQC experiment. This gives a good spectrum with which to check whether your protein is folded and assess the quality of your spectrum to see whether it is worth recording other spectra and possibly proceeding on to other, more expensive labelling schemes. It is also possible to record a variety of dynamics experiments using $^{15}$N-labelled protein as well as N-H residual dipolar couplings (RDCs). If the protein is comparatively small (< ~150 amino acids) you can assign the $^{1}$H and $^{15}$N backbone resonances using $^{15}$N-NOESY and $^{15}$N-TOCSY epxeriments. $^{15}$N-labelled protein can also be useful for titrations with ligands or other proteins with which it forms a complex[41]. The $^{15}$N labelled protein sample was prepared using *E. coli* bacteria supplemented with labelled $^{15}$N ammonium sulphate more details experimental procedures. The sample after protein expression was purified with affinity, size exclusion chromatography. The sample was filtered through Millipore ultrafree 0.22μm filter for recording NMR spectrum on Bruker 700MHz spectrometer equipped with cryoprobe. NMR spectra was recorded with protein sample in a buffer of pH 7.0 containing of 20mM sodium phosphate, 100mM NaCl, 1mM DTT.

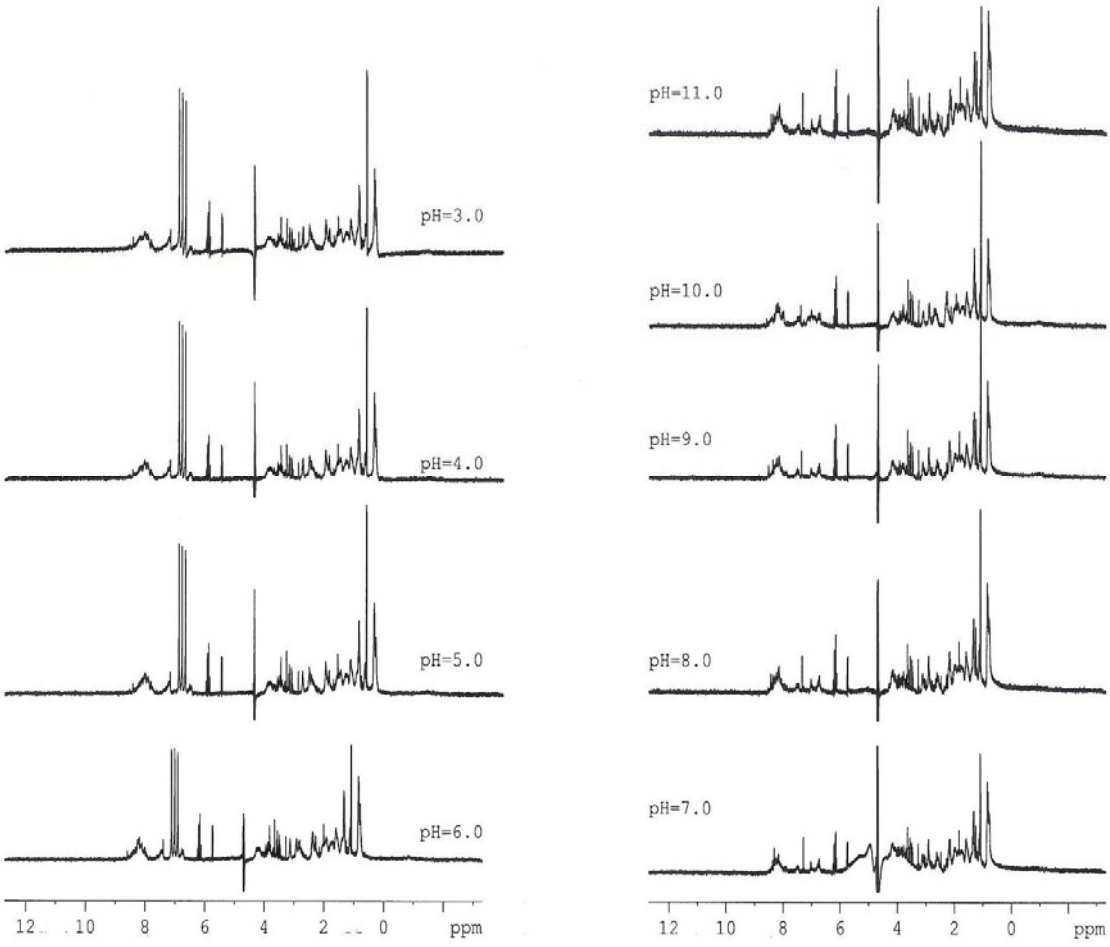


**Figure: 4.2 A one dimensional $^{1}H$ NMR spectra of the Par-4 SAC domain in 20mM phosphate buffer at different pH values at $25^{0}C$. Proton chemical shifts not well dispersed.**

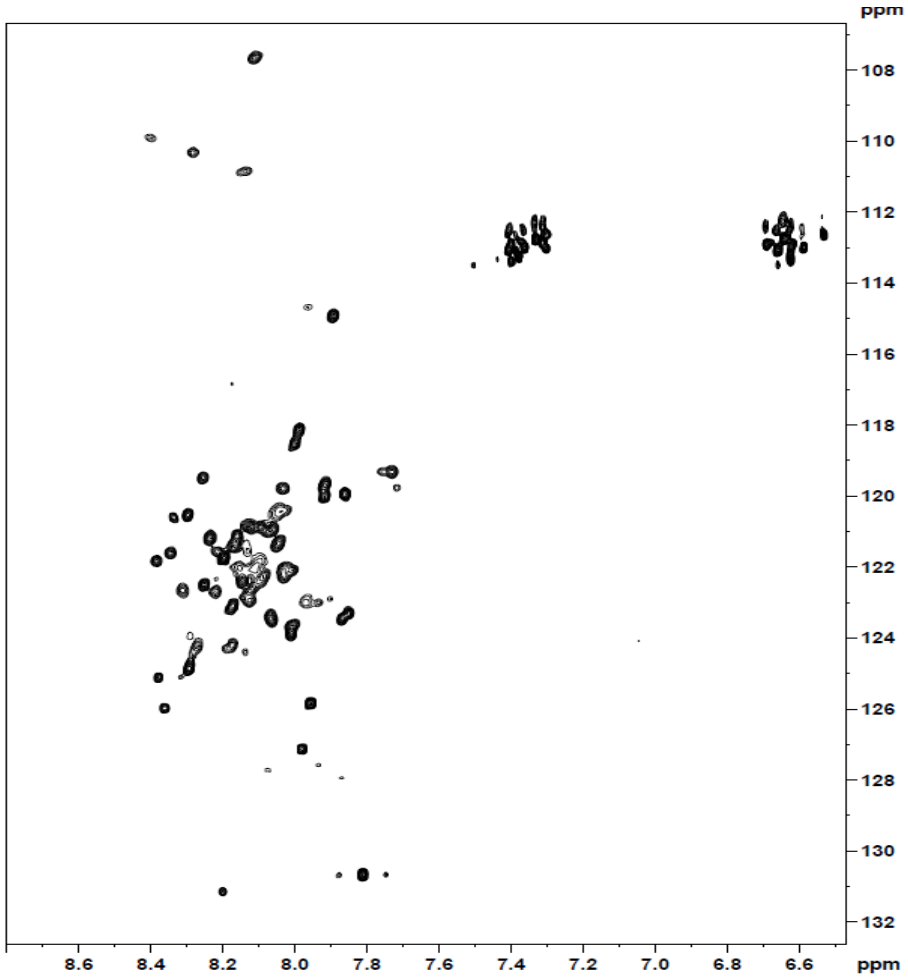


**Figure: 4.3 Contour plot of the HSQC spectrum of the purified Par-4 SAC domain acquired at 278K. The spectrum was recorded on a Bruker 700 MHz instrument with a cryoprobe. The protein sample was in a buffer of pH 7.0 containing of 20mM sodium phosphate + 100mM NaCl 1mM DTT.**

To make a more comparative study of the rrPar-4 SAC protein pairwise overlays of $^{1}H$-$^{15}N$ heteronuclear single quantum coherence (HSQC) spectra for rrPar-4FL, rrPar-4DLZ and rrPar-4SAC are shown in figure 4.4[42]. The spectra of all three proteins display the features that characterize disorder in proteins, namely sharp peaks and narrow $^{1}H$ chemical shift dispersion[43,44]. The rrPar-4FullLength and rrPar-4ΔLZ was also studied in our laboratory[45]. Chemical shift similarities indicate some structural similarity between rrPar-4FL and rrPar-4DLZ (Figure 4.4 A). Fewer peaks share similar chemical shifts when comparing rrPar-4FL or rrPar-4DLZ with rrPar-4SAC (Figure 4.4 A, B). Thus, the majority of residues in rrPar-4SAC experience a different local environment and possibly a different conformation than the SAC domain in the

context of either the rrPar-4FL or rrPar-4DLZ constructs. Only 160 of the 308 peaks expected (335 – N-terminal residue – 26 prolyl residues) for rrPar-4FL and 152 of the 266 expected peaks (293 – N-terminal residue – 26 prolyl residues) for rrPar-4DLZ are readily picked. Conversely, 58 peaks of the expected 60 (62 – N-terminal residue – one prolyl residue) were readily identifiable for rrPar-4SAC with only two glycyl residues being unobservable.

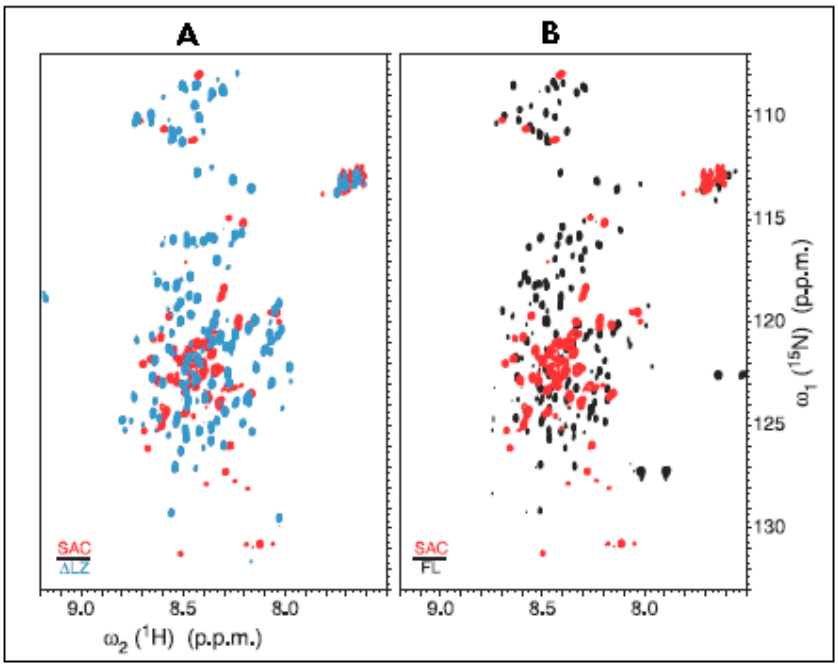


**Figure: 4.4 Pairwise overlays of $^1$H-$^{15}$N HSQC spectra of (A) rr Par-4ΔLZ (blue contours) and rrPar-4SAC (red contours and (B) rrPar-4FL (black contours) and rrPar-4SAC (red contours). The composition of the sample was $^{15}$N-rrPar-4SAC 0.34 mM in 10mM Tris (pH 7.0) 20mM NaCl, 5% D2O. All spectra were recorded at 5$^0$C and the processing parameters (See Experimental procedures) were identical for qualitative comparison.**

**CD Spectroscopy data**

To assess the degree of α-helicity in the rrPar-SAC a CD spectroscopy data was recorded. The full range (5 $^0$C steps from 5–75$^0$C) and a sub-set of four spectra (5, 25, 45 and 65$^0$C) of a temperature series recorded by CD spectroscopy for rrPar-4SAC (Figure 4.5A, B). Significant α-helical character is not seen rrPar-4SAC. The CD

spectra for rrPar-4SAC show a typical profile of IDPs with a deep transition at 200 nm[46].

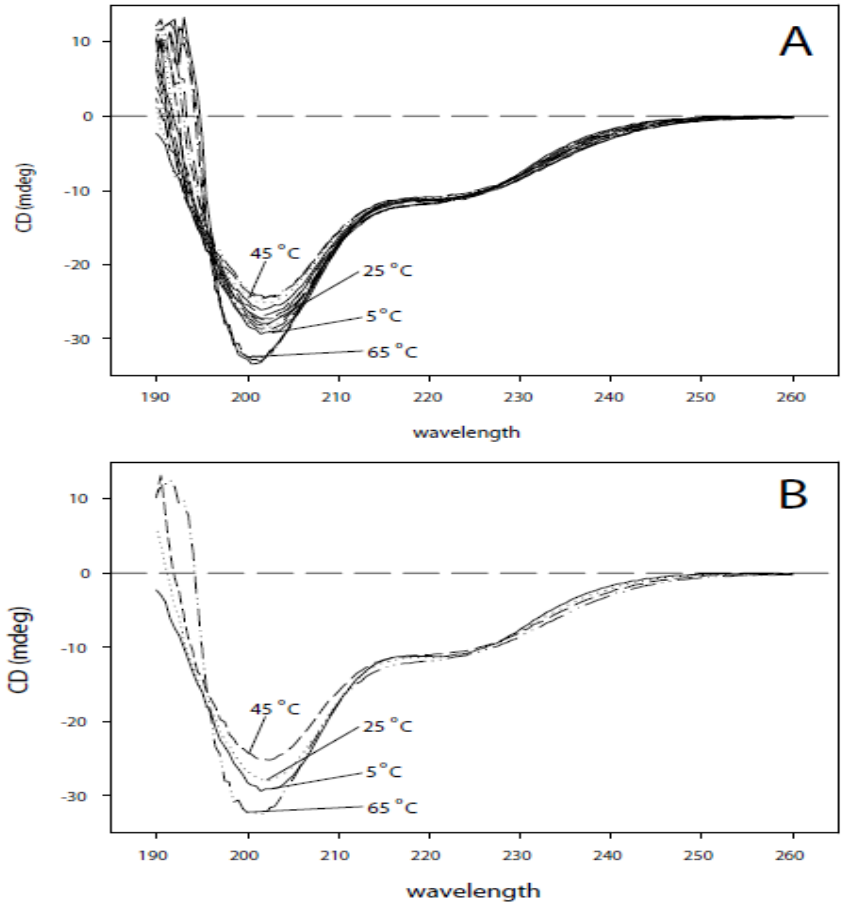


**Figure: 4.5 Temperature dependence of the CD spectrum of the Par-4 SAC recorded in 10 mM Tris (pH 7.0) 20mM NaCl over temperature range of 5-65 $^0$C (A). For clarity, four equally spaced temperatures from the sampled range are shown in (B).**

**Dynamic Light Scattering (DLS) measurement:**

As calculated (i.e. from sequence) and experimentally determined [i.e. from MS, Tricine-PAGE and dynamic light scattering (DLS)], the molecular weights for rrPar-4FL and rrPar-4SAC are given in Table 1. Because DLS measures the Stokes radius ($R_S$) of a particle, The equation $\log(R_S) = 0.357 \times \log(MW) - 0.204$ was used to convert

from $R_S$ values to MW for comparative purposes[47,48]. Although this approximate calculation does not take into account the shape of the particle (i.e. it assumes a sphere), the result is useful for illustrating the degree of extended structure in the protein. The results of DLS experiments are summarized in Table 1. The measured $R_S$ for rrPar-4FL was 189 Å, which is much larger than expected for a monomeric random coil, suggesting a polymeric state for rrPar-4FL under these conditions[42].

**Table 1 Hydrodynamic properties of rrPar-4 constructs using various biophysical techniques. MW (kDa) and hydrodynamic radius (Å) are shown in the format MW (RS) for two constructs using four techniques. RS and MW were calculated from the primary structure in reference to a folded conformation using $\log(R_S) = 0.357 \times \log(MW) - 0.204$.**

| Method of analysis | | | | |
|---|---|---|---|---|
| **Construct** | **Sequence** | **MS** | **PAGE** | **DLS** |
| **rrPar-4FL** | **36.1 (26.5)** | **36.2 (26.5)** | **49.5 (29.6)** | **8889(189)** |
| **rrPar-4SAC** | **7.0 (14.8** | **7.1 (14.8)** | **12.5 (18.1)** | **18.7(20.9)** |

The primary structure predicts MWs of 7.0 kDa for rrPar-4SAC. MALDI-TOF mass spectroscopy was used to assess the purity and determine the sizes of the constructs produced. The sizes determined for rrPar-4SAC (6.6 Da difference between expected and observed after accounting for $^{15}N$ labelling of the sample used for MS analysis) agree within error (approximately 0.1%) with the sizes predicted from sequence analysis (Table 1). Also the MS revealed that the rrPar-4FL construct is approximately 0.2 kDa larger than expected studied by a colleague in our laboratory.

**Sequence Complexity of The Par-4 SAC domain:**

A large proportion of proteins are found to be intrinsically disordered, which refers to segments of protein sequences that do not form well-defined three-dimensional structures in their native states[49]. Many proteins contain local regions of such disorder, and in some cases entire proteins appear to exist as ensembles of structures[50]. Since protein subsequence that lack the ability to form a well-defined three-dimensional structure may still be functionally related, identifying such disordered regions is getting more and more important for both understanding protein

functions and conducting structural analyses. Intrinsic disorder refers to segments or to whole proteins that fail to self-fold into fixed 3D structure, with such disorder sometimes existing in the native state.

It has been shown in many studies that protein disorders can be predicted from their primary sequences[51]. Intrinsically disordered proteins are identified by low hdrophobicity, high net charge, low compactness, absence of globularity, lack of secondary structure, and high flexibility[49,52]. Compared to sequences from ordered proteins, these variously characterized intrinsically disordered segments and proteins, typically have obviously higher levels of protein-specific subsets of the following amino acids: R, K, E, P, and S, and lower levels of subsets of the following: C, W, Y, I, and V. Thus one distinction of the amino acid sequences of "natively unfolded" proteins has been suggested in the literature, i.e. the presence of numerous uncompensated charged groups (often negative), producing a large net charge at neutral pH. A low content of hydrophobic amino acid residues has been also noted for several "natively unfolded" proteins. Consistent with these characteristics, Par-4 contains a low content of hydrophobic amino acids, which contributes to the low overall hydrophobicity. The hydrophobicity of each amino acid sequence was calculated by the Kyte and Doolittle approximation[53].

By using the Swiss Institute of Bioinformatics (SIB) server ExPASy the protein information was extracted. The mean hydrophobicity for the Par-4 SAC domain was calculated 0.30 by sum of the normalized hydrophobicities of all residues divided by the number of residues in the polypeptide. The mean net charge for the Par-4 SAC is defined as the net charge at pH 7.0, divided by the total number of residues (0.017). The plot of mean net charge versus mean net hydrophobicity for the set of folded and "natively unfolded" (Table 2) shows the Par-4 SAC domain also is a natively unfolded protein (Figure- 4.6).

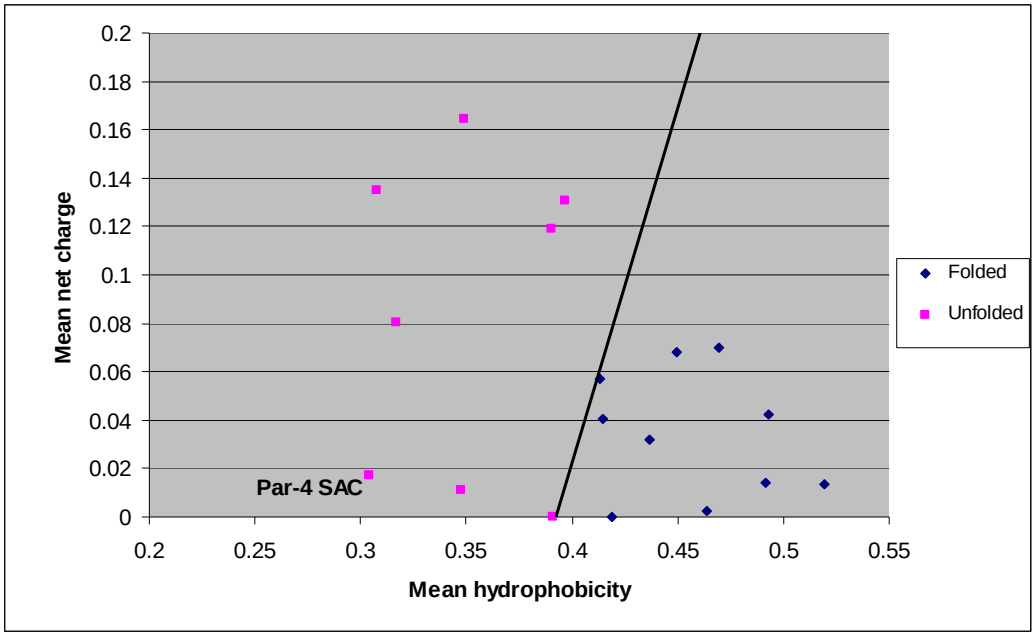


**Figure 4.6: Charge versus hydrophobicity plot of the Par-4 SAC. The mean overall hydrophobicity and mean net charge of the Par-4 SAC were calculated and plotted together with the values for a collection of natively unfolded and folded proteins. The unfolded proteins data was used from Uversky et.al. Proteins: Struct. Funct. And Gen. 41:415-427, 2000[52]. The solid black line represents the border between "natively unfolded" and natively folded proteins in the charge hydrophobicity space. Par-4 SAC localizes to the natively unfolded region of the charge hydrophobicity plot.**

Here rrPar-4SAC clearly fall into disordered space generally characterized by low mean hydrophobicity and high net charge. The construct representing the SAC domain (rrPar-4SAC), with 14 positively charged and 13 negatively charged residues but few hydrophobic residues, lies further in the disordered region. Based on DISEMBL analysis[54], the majority (> 70%) of the Par-4 is predicted to be disordered[42]. Par-4 SAC contains one tyrosine (Tyr), one cysteine (Cys), two valine (Val) and no tryptophan, shows depletion of amino acids contributing to protein folding. A recent study compared the mean hydrophobicity and mean net charge of natively folded and natively unfolded amino acid sequences, establishing that natively unfolded proteins occupy clearly demarcated region of the charge-hydrophobicity space[53].

**TABLE 2: Data used for charge vs. hydrophobicity plot**

| | Protein | Mean Hydrophobicity | Mean net Charge |
|---|---|---|---|
| LEF-1 C-terminal | **Unfolded** | 0.39 | 0.1188 |
| Antitermination | | 0.3968 | 0.1308 |
| Snc1, cytoplasmic | | 0.391 | 0 |
| Nonhistone | | 0.3075 | 0.1348 |
| Fibronectin-binding | | 0.3493 | 0.1642 |
| Secretogranin | | 0.3171 | 0.0806 |
| EMB-1 protein | | 0.3478 | 0.0108 |
| **Par-4 SAC** | | 0.3042 | 0.0169 |
| Actin | **Folded** | 0.4641 | 0.0024 |
| BSA | | 0.4368 | 0.0319 |
| ProteinG | | 0.4129 | 0.0569 |
| Lysozyme | | 0.4933 | 0.0422 |
| PKC | | 0.4498 | 0.0681 |
| Ubiquitin | | 0.4189 | 0 |
| HT-3C | | 0.5193 | 0.0134 |
| TrypsinBPTI | | 0.4692 | 0.07 |
| Calbindin | | 0.4148 | 0.0406 |
| Hemoglobin | | 0.4918 | 0.0141 |

Thus, the amino acid composition of the Par-4 SAC and its low mean hydrophobicity predict that its structure is intrinsically disordered protein (IDP).

A classically folded protein is less susceptible to proteolysis than an IDP upon equilateral exposure to a protease such as trypsin because some of its cleavage sites are protected by tertiary folding[55,56]. The results of a limited trypsin digest of rrPar-4FL, rrPar-4ΔLZ, rrPar-4SAC and BSA were studied[42]. After 15 min of exposure to trypsin rrPar-4DLZ was more than 95% digested, rrPar-4FL and rrPar-4SAC were over 80% digested, whereas BSA was only 10% digested. BSA was chosen for comparison because it has a similar percentage of predicted cut sites to that of the Par-4 constructs.

## 4.2 Discussion:

The structure-defines-function paradigm of molecular biology is currently under scrutiny because many proteins have been identified that are functional without the need for well-defined secondary or tertiary structure[51,57,58]**.** The occurrence of intrinsic disorder in proteomes is correlated to the complexity of the cell; thus, eukaryotic proteins have a higher proportion of disorder (35–51% of proteins with disordered regions of 40 residues or longer) than proteins from prokaryotes and archaea (6–33%)[59]. The prevalence of intrinsic disorder is higher in proteins that are involved in cell signalling, cytoskeletal organization and ribosomal or cancer-related processes[60]. Disorder in proteins that control these processes appears to be of functional importance because these events are often tightly controlled and highly dynamic and often become deregulated in cancerous cells[61]. Many signaling proteins function in pathways associated with cancer

### Sequence analysis of the Par-4

Bioinformatic analysis of rrPar-4FL reveals characteristic features of IDPs, including high net charge, low mean hydrophobicity and low sequence complexity[43, 23]. Relative to the amino acid usage observed in folded proteins, rrPar-4SAC is depleted in order-promoting amino acids and enriched in disorder promoting residues (Section 4.3). The lack of hydrophobic residues inhibits the formation of a hydrophobic core and thus the formation of stable tertiary structure[51]. More than 70% of rrPar-4FL is predicted to be disordered by DISEMBL, providing a strong argument against the formation of stable global tertiary structure. Hydrophobic cluster analysis is a method of displaying the primary structure such that the clustering of hydrophobic residues and thus regions of possible order become evident[42,63]. Intrinsic disorder in proteins is often erroneously considered to be a featureless random coil, although proteins do not achieve a completely random conformation even in strongly denaturing conditions[64]**.** A more accurate depiction is that IDPs exist as ensembles of rapidly interchanging conformers that sample varying regions of secondary structure space. IDPs can be broadly categorized into three non-exclusive groups (i.e. a single IDP may fall into more than one category): random coil, pre-molten globule or molten globule[65]**.**

Because of the high percentage of rrPar-4FL that is predicted as disordered, a random coil-like classification of the ensemble would appear to be the most appropriate.

rrPar-4 displays aberrant electrophoretic mobility and is susceptible to proteolysis. Aberrant electrophoretic mobility in a denaturing PAGE system is a hallmark of IDPs because their unique amino acid composition reduces the amount of sodium dodecyl sulphate that is able to bind[46,51,66]. Aberrant mobility on PAGE gels of rrPar-4 and rrPar-4 constructs (i.e. deletion mutants) has been demonstrated, although it is not known whether the effects of IDP amino acid composition were considered[67,68]. In the present study, slower than expected migration of the Par-4 constructs resulted in apparent MWs that were 1.3-(rrPar-4FL and rrPar-4DLZ) to 1.8- (rrPar-4SAC) fold larger than that predicted from sequence or measured using MS (Table 1).

Limited proteolysis can be used to distinguish ordered and disordered proteins based on their relative sensitivity to cleavage by proteases such as trypsin[43,56,69]. Although rrPar-4FL, rrPar-4DLZ and BSA contain an approximately equal percentage of trypsin cut sites, BSA is digested at a much slower rate[42]. This implies that a significant portion of the conformational ensemble of the rrPar-4 proteins are more exposed to the solvent than BSA and largely lack protection by folded and stable tertiary structure.

**Hydrodynamic radius of the Par-4 is larger than predicted by sequence analysis:**

An IDP will have an observable $R_S$ larger than a folded globular protein of the same MW[70,71]. Some examples of IDPs with large $R_S$ values relative to MW have been summarized previously[72]. In the present study, the $R_S$ measured for rrPar-4FL and rrPar-4SAC correspond to MWs of $8.9 \times 10^3$ and 18.7 kDa respectively, and are much larger (713 and 141%) than what would be expected for a folded globular protein of similar MW. The MW estimations shown in Table 1 are used as a point of reference to illustrate that the degree of 'unfoldedness' of the Par-4 is very high, which is similar to that expected for a coil like as opposed to a pre-molten globule ensemble[70].

**Secondary structure of the Par-4 assessed by CD and NMR**

The CD spectrum for rrPar-4SAC is exemplary of IDPs with a deep transition at 200 nm and a minor transition at 222 nm[46]. The CD spectra of IDPs are often complicated by minor contributions from secondary structure elements, such as alpha or poly-proline type II helices[73]. Deconvolution of the $25^0$C spectra estimates 18% of combined regular and distorted a-helix for rrPar-4SAC[74]. The construct of rrPar-4 SAC remain relatively stable throughout the heating cycle because the 5–$65^0$C traces exhibit similar features. Thus, in addition to the coiled-coil region of rrPar-4FL, other regions of these proteins may transiently populate α-helical or other secondary structures. Interestingly, although the overall temperature-induced changes are minor, an isodichroic point at 210 nm is observed for rrPar-4SAC, which may be interpreted as a two-state conformational change (Figure 4.5A). This could be the result of secondary and/or tertiary structure that is thermally disrupted.

In a protein structure the individual residues are packed into chemical environments, which are very different from the random coil situation. The nuclei in the folded form are subject to a many different types of microenvironments of chemical screens. For NMR spectra of proteins the NMR signals of the nuclei of the individual residues are in most cases seen in the vicinity of the random coil shift value. However, due to structural diversity in a protein structure, the signals are often shifted significantly away from the random coil value. The atomic resolution of NMR makes it uniquely suited to assess the 'orderedness' of an IDP[46]. Because most residues of an IDP are solvent exposed and inherently flexible, they share a similar chemical environment and, consequently, share similar NMR frequencies, resulting in significant overlap of resonances (particularly for $^1$H resonances) and population-weighted average chemical shifts[75]. Mobility also results in sharp peaks as a result of increased T2 values[76]. The spectra of rrPar-FL, rrPar-ΔDLZ and rrPar-4SAC shown in Figure 4.4 are characteristic of IDPs, with ensemble averages, narrow peaks and poor chemical shift dispersion.

If protein is folded the proton NMR spectra show CS of the amino acids in the upfield and downfield region of proton NMR spectra. The side chains of the folded protein in the centre are less exposed to the solvent (Figure-4.8) which shows

different chemical interaction and gives well expanded CS in proton NMR spectra. The chemical shifts are primarily determined by the covalent structure of the amino acid residue. They can also be significantly affected by the interactions with the solvent. Therefore, the exclusion of the solvent water from the interior of a globular protein causes the chemical shifts of the core residues to be different from those of the water-exposed amino acid residues, so that even NMR lines originating from multiple residues of the same amino acid type can be distinguished. This "conformation-dependent chemical shift dispersion" was found to be sufficiently large to enable $^{1}H$ NMR studies of protein denaturation.

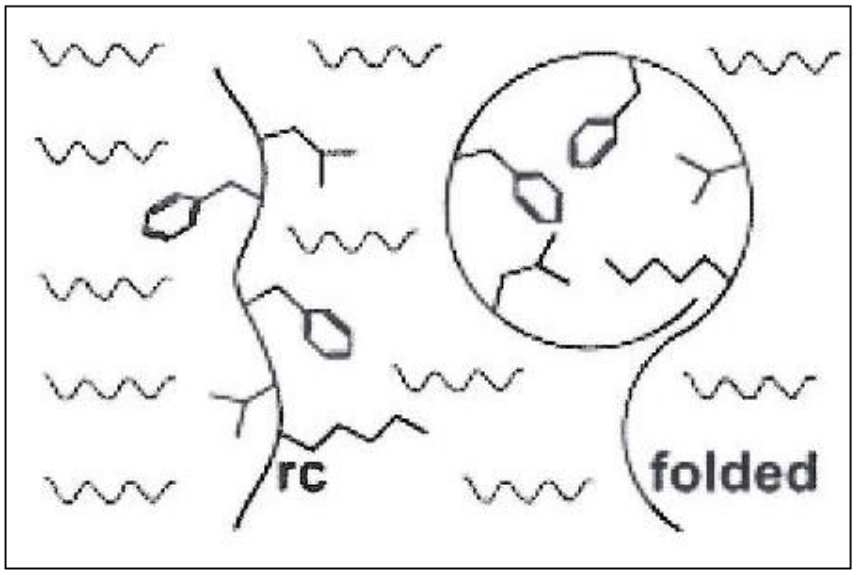


**Figure: 4.8 Schematic illustrating that the solvent ( ∿∿ ) has free access to all parts of a random coil polypeptide chain (rc), whereas it is excluded from the core of a folded globular protein. (Figure taken from Kurt Wüthrich, Nobel Lecture 2002.)**

The spectrum of rrPar-4SAC (Figure 4.4) is much more complete. A similar degree of disorder is suggested by the peak shape and chemical shift dispersion. The size of the rrPar-4SAC (7 kDa) relative to that of the other constructs (> 30 kDa) is likely to be a contributing factor in the observance of these resonances because fewer residues equate to less chance of spectral overlap and a lower likelihood of slow to intermediate exchange.

**4.3 Conclusion:**

The data obtained in the present study indicate that the rrPar-4 SAC [selective apoptosis of cancer cells (SAC) domain construct, (residues 137-195)] can be classified as predominantly intrinsically disordered proteins (IDP's) and cannot be further characterized by NMR spectroscopy. The intrinsically disordered proteins (IDP's) are proteins which do not show any protein folding and lacks stable tertiary or secondary structure under physiological conditions therefore structural assignments will be difficult.

The further studies conducted in our research laboratory on rrPar-4 FL and rrPar-4 Leucine zipper (residues 292-332 of rrPar-4) region shows that the N-terminal rrPar-4 FL (residues 1-332) as a predominantly intrinsically disordered protein whereas the C-terminus (residues 292-332 of racine Par-4) is predicted to form Leucine Zipper (LZ)[21,42,45]. Limited proteolysis and DLS experiments demonstrate that rrPar-4FL is primarily extended in solution, exhibiting high susceptibility to trypsin and a large hydrodynamic radius. Furthermore, CD and NMR experiments revealed characteristic spectral features of intrinsic disorder. The bioinformatics analysis shows that highly conserved Par-4 has low sequence complexity, is enriched in polar and charged amino acids and is classified as disordered when plotted in charge-hydrophobicity space. The SDS-PAGE results as mentioned in previous Chapter-3 show that the Par-4 SAC protein runs above its actual molecular weight. This is characteristic of the IDP's which bind less SDS than usual because of their unusual amino acid composition and their apparent Mw is often 1.2-1.8 times higher than the real one calculated from sequence data.

# CHAPTER 5

# SUMMARY AND FUTURE WORK

## 5.1 Summary of results:

The intrinsic disorder of the Par-4SAC protein has been predicted by bioinformatic analysis CD spectropolarimetry, DLS and NMR spectroscopy. The sequence analysis study done in this project shows that highly conserved Par-4SAC has low sequence complexity, is enriched in polar and charged amino acids and is classified as disordered when plotted in charge-hydrophobicity space. CD data shows rrPar-4SAC a typical profile of IDPs with a deep transition at 200 nm. The observable Stokes radius of a protein increases in proportion to its degree of 'unfoldedness'; thus, an IDP will have an observable $R_S$ larger than a folded globular protein of the same MW. The present study the $R_S$ measured for rrPar-4SAC corresponds to MW of 18.7 kDa. Also the limited proteolysis and DLS experiments demonstrate that rrPar-4SAC is primarily extended in solution, exhibiting high susceptibility to trypsin and a large hydrodynamic radius.

Furthermore NMR experiments revealed characteristic spectral features of intrinsic disorder. The proton NMR spectrum of the protein shows low chemical shift dispersion. The $^{1}H$-$^{15}N$ HSQC spectrum also show 58 peaks of the expected 60 (62 – N-terminal residue – one prolyl residue) were readily identifiable for rrPar-4SAC with only two glycyl residues being unobservable. Nonetheless, a similar degree of disorder is suggested by the peak shape and chemical shift dispersion. The size of the rrPar-4SAC (7 kDa) relative to that of the other constructs (> 30 kDa) is likely to be a contributing factor in the observance of these resonances because fewer residues equates to less chance of spectral overlap and a lower likelihood of slow to intermediate exchange. Thus the NMR spectrum of rrPar-4SAC is characteristic of IDPs, with ensemble averages, narrow peaks and poor chemical shift dispersion. The overall data demonstrate that rrPar-4SAC is intrinsically disordered protein (IDP).

**5.2 Future work:**

Often, intrinsically disordered proteins are involved in regulation, signaling and control pathways, where binding to multiple partners and high specificity interactions play a crucial role. It is suggested that functions of intrinsically disordered proteins may arise from the specific disorder form, from inter-conversion of disordered forms, or from transitions between disordered and ordered conformations. Intrinsic disorder imparts many advantages to a multifunctional protein such as the Par-4. Protein–protein and protein–ligand interactions can be highly specific yet readily reversible, whereas post-translational modifications allow for very tight control of the functions of the Par-4. The results obtained in the present study demonstrate that the combined intrinsically disordered and coiled coil nature of the Par-4 provides unique structural properties through which Par-4 can perform a multifunctional role in various tissues and cellular processes.

Par-4 forms a stable coiled-coil through a self-association event mediated by the C-terminus. The coiled-coil was probed using increasing concentrations of chaotropic agents and was found to be very stable[21]. Using NMR, the segment of the Par-4 not involved in the coiled-coil was shown to have spectral features that were similar to those of a C-terminal deletion mutant. This is important because it suggests that the Par-4 is able to bind more than one partner at a time and thus could function as a hub linking the functions of several proteins. The coiled-coil region of the Par-4 represents an important functional domain that is an example of a gain of structure upon binding transition, which is another important feature of IDPs[77].

Par-4 contains several conserved phosphorylation sites that are modified by kinases, such as PKA, PKC, casein kinase II and Akt1, adding a further level of regulation for the function of the Par-4[17]. Phosphorylation of an absolutely conserved threonine (rat T155, human T163 or mouse T156; Figure 1.2) by PKA is required for nuclear translocation[78]**.** The phosphorylation sites come within rrPar-4SAC domain. Phosphorylation of a C-terminal serine residue (rat S249, human or mouse S231; Figure 1.2) by Akt1 effectively inactivates the Par-4 by allowing the chaperone protein 14-3-3 to bind and sequester it in the cytoplasm, even if it is phosphorylated on T155[79]. These multiple interactions coupled with a high degree of sequence

conservation and post-translational modification suggest that the in vivo role(s) of the Par-4 are highly temporally and spatially regulated. The interaction study of the above mentioned partners can be done.

Due to time constraints we have not done crystallization trial for the Par-4 SAC domain protein. The future work for the Par-4 SAC protein study could be done by trial experiments for protein crystallization. The crystals can be further used for X-ray crystallography study. The most direct way of identifying the lack of structure is to look for residues with missing backbone coordinates in three-dimensional structures determined by X-ray crystallography. Unfolded conformations can also be detected and characterized by small-angle X-ray scattering and sedimentation analysis. The trial experiments for protein crystallization can be done by using the present buffer conditions.

# CHAPTER 6

# EXPERIMENTAL PROCEDURES

## 6.1 General methods

### *Media*

All *E. coli* cultures used during the research project were grown in Luria Broth (LB) made up (25 $gL^{-1}$) with Milli-Q water and sterilized by autoclaving at $121^0$C and 15 psi for 20 minutes. Agar media for agar plates was prepared by adding 1.5% agar (Invitrogen) and the liquid LB media autoclaved. Add 100 µg $mL^{-1}$ ampicillin, 34 µg $mL^{-1}$ chloramphenicol immediately after cooling the solution to ~$60^0$C and pour the solution on plates.

### *pH measurement*

The pH of buffers used in this project was measured using a Sartorius PP-15 Professional Meter equilibrated with 3M KCl.

### *Expression vector construction*

A versatile expression vector was designed to be used to either express the Par-4 constructs alone or in conjunction with putative binding partners in Escherichia coli cells. The pET23a vector from Novagen (Merck Biosciences, Darmstadt, Germany) was used as a template to create pCFETrxH-TEV, which enables the expression of targets as fusion proteins with thioredoxin. A hexa-histidine tag is included between thioredoxin and the N-terminus of the target. An rTEV-protease cleavage site was included for removal of the thioredoxin and hexa-histidine tags. The fusion tag was PCR amplified using pET32a (Merck Biosciences) as a template and subsequently inserted into the NdeI/BamHI restriction sites of pET23a. The primers were: forward 5′-CTGGCATATGAGCGATAAAATTATTCAC-3′ and reverse 5′-CCGGGGATC CCTGAAAATACAGGTTTTCGGTCGTTGGGATATCGTAATCGTGATGGTGAT GGTGATGCATATG-3′.

The rrPar-ΔDLZ (residues 1–290) construct lacking the leucine zipper was PCR amplified with the primers 5′-CAGGGATCCATGGCGACCGGCGGCTATCGGAG-3′ and 5′-CCGGAAGCTT TTATTCTTCTTTATCTTGCATCAG-3′ using the full-length construct as a template. The PCR product was then cloned into the BamHI / HindIII sites of pCFE-TrxH-TEV. The rrPar-4SAC (residues 137–195) construct representing the SAC domain was cloned in the same manner as rrPar-ΔDLZ using the primers 5′GAGGATCCAGGAAAGGCAAAGGGCAGATCG-3′ and 5′GCAA GCT TTTA TGCTTCATTCTGGATGGTG-3′.

***Growth of E. coli cells***

20 mL of LB containing 100 µg $mL^{-1}$ ampicillin or 34 µg$mL^{-1}$ chloramphenicol (depending upon cell type) was inoculated with a scraping glycerol stock of the appropriate strain of *E. coli* cells. The cells were grown overnight, shaking, at $37^0$C. 2.5 mL of the overnight culture was added to inoculate the 250 mL LB ( with 100 µg $mL^{-1}$ ampicillin and 34 µg $mL^{-1}$ chloramphenicol) and was grown until mid-logarithmic phase ($OD_{600}$ 0.4-0.6).

***Induction of protein expression***

Expression of genes inserted into the multiples cloning site of the plasmid is under control of the *lac* promoter and is therefore induced by the addition of lactose or isopropyl-β-D-thiogalactopyranoside (IPTG). IPTG was used because it is a non-physiological analogue of lactose which is not metabolized by the cell, unlike lactose. IPTG was added to a final concentration of 0.4µM from stock solution of 0.4M IPTG to induce protein expression.

***Harvesting and lysis of cells***

The cells were harvested after induction by centrifugation at 5000 rpm for 20 minutes at $4^0$C. Cell pellet were stored at $-80^0$C until lysis. Cell pellets were resuspended on ice in the appropriate lysis buffer. Cell pellets from 250 mL cultures were in resuspended in 10mL of lysis buffer, add 5µl of protease inhibitor (Calbiochem Cocktail set III) and 20mg of lysozyme (from chicken egg white, SIGMA) incubate for 1-3 hours at $37^0$C. The cells were lysed by French Press (8000psi) and cell debris was removed by centrifugation at 10000 rpm for 25 minutes at $4^0$C.

The lysis buffer used for protein expression with H-MBP-3C vector consisted of PBS containing 12mM $Na_2HPO_4$, 12 mM $NaH_2PO_4$, 140 mM NaCl, 3mM KCl pH= 7.4

Different buffers used for protein expression with pProExHTb and pET-TEV vector:
A: Lysis buffer 50mM Na-phosphate, 25mM Imidazole, 10% Glycerol, 100mM NaCl pH= 8.0
B: Wash buffer 50mM Na-phosphate, 50mM Imidazole, 10% Glycerol, 100mM NaCl pH= 8.0
C: Elution buffers used 50mM Na-phosphate, 500mM Imidazole, 10% Glycerol, 100mM NaCl pH= 8.0
D. Exchange buffer 50mM Tris HCl + 0.5mM EDTA + 1mM DTT pH= 8.0

***Sonication***

All sonication was performed using a VirTis VirsSonic digital 475 ultrasonic cell disrupter using a 1/8 inch probe at ~60 Watts.

***Polyacrylamide gel electrophoresis***

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using the method of Laemmli[80] with 4% (w/v) stacking gel and a 12% (w/v) resolving gel, using a Mini Protean III cell (Bio-Rad). All samples were added to SDS loading buffer and boiled for 5 minutes before being loaded to gels. A current of 200V was applied until the dye front ran off the bottom of the resolving gel. Bio-Rad low range molecular weight marker or Bio-Rad Precision Plus protein standards were used. To detect low molecular weight protein 16% resolving gel, 10% spacer gel and 4% stacking gel was used. The 16% resolving gel contained 0.83gm Glycerol, 1.675ml of 3M Tris-HCl + 0.3% SDS buffer of pH 8.45, 1.675ml of 49.5% Acrylamide/1.5% C, 50µl of 10% APS, 10µl of TEMED (N,N,N',N'-tetramethylethylenediamine) make upto 5ml with water. For 10% spacer gel use 1.0 ml of Tris buffer, 0.6 ml of 49.5% Acrylamide make up to 3ml with water. 4% stacking gel was made with 0.775 ml 3M Tris buffer and 0.25 ml of 49.5% Acrylamide and make up to 3.125 ml with water.

The Tris Tricine gels were run half an hour at current of 12mAmp and for 2 hours 30 minutes at 24mAmp current. Also Tris Tricine gels with SDS PAGE of 10% resolving gel and 4% stacking gel were also used. Protein in the gels were detected by staining the gels for at least 1-2 hours in a solution consisting of 1g $L^{-1}$ Coomassie Brilliant Blue R 250 (CBB, Park Scientific) in 40% (v/v) ethanol and 10% (v/v) acetic acid in water. CBB is the most popular protein stain. It is an anionic dye, which binds with proteins non-specifically. The structure of CBB is predominantly non-polar. The proteins are detected as blue bands on a clear background. As SDS is also anionic, it may interfere with staining process. Therefore, large volume of staining solution is used, approximately ten times the volume of the gel. Excess dye was removed by soaking gels for 20-30 minutes in an identical solution but without the Coomassie dye. Silver staining protocol was also used for detection of the protein with low concentration.

***Size-exclusion chromatography (SEC)***

SEC, or gel filtration chromatography, separates proteins according to their molecular mass. The matrix is porous, acting like a molecular sieve that excludes large proteins that are eluted from the column with the void volume and retains smaller proteins that are eluted between the void volume and the total column volume in order of decreasing molecular mass. SEC in this project was used as a final purification step before using sample for NMR spectroscopy. A Superdex 75 10/300 GL column bed dimensions 10 x 300 mm with recommended sample 250 µl (Amersham) was used with AKTA Prime Plus Amersham Biosciences low pressure gel filtration unit. All SEC was performed using 0.2 µm filtered buffer 20mMNaP, 100mM NaCl, 1mM DTT pH= 7.5 at a flow rate of 0.4 mL $min^{-1}$ for 60 minutes.

***Reversed-phase HPLC***

Reversed-phase High Performance Chromatography (RP-HPLC) is a separation of molecules based on their hydrophobicity. Hydrophilic molecules elute first from a reversed-phase column, strongly hydrophobic molecules last. The most common chromatographic mode used for reversed phase separation of peptides is gradient elution, using two different mobile phases. It is widely used method for purification of and analysis of peptides owing to its powerful resolving capability, reproducibility and recovery. In this operation mode the sample, dissolved in a hydrophilic, aqueous

eluent will be injected onto a column which is equilibrated with hydrophilic eluent. A hydrophobic molecule will bind to the column, because the affinity of the molecule to the hydrophobic stationary phase is large than to the hydrophilic mobile phase. A Dionex HPLC system used with Phenomenex C18 column of 250 x 4.6mm and particle size of 5µm, column oven temperature was kept $25^0$C. The DAD UV detector was set to 205nm and 280nm. The mobile phase with gradient elution of solvent-A (0.1% TFA in water) and solvent-B (0.08% TFA in acetonitrile). The first gradient elution method used with 0-60% B for 60min at flow rate of 1ml/min. In order to elute protein of interest early the gradient for solvent B was changed to 0-25% for 25 min and then changed to 100% at 40min.

***Determination of Protein concentration***

Protein concentration was determined by the method of BCA Assay or Bradford, 1976[81] using bovine serum albumin (BSA) as a standard. BSA protein standards were made to protein concentrations of 0 mg $mL^{-1}$, 0.05 mg $mL^{-1}$, 0.1 mg $mL^{-1}$, 0.2 mg $L^{-1}$ and 0.4 mg $L^{-1}$ by dilution of a 20 mg $mL^{-1}$ stock BSA solution (Bio-Rad). 1 mL of a 1/5 dilution of stock Bradford reagent (Bio-Rad) was added to 100 µl of each standard and a standard curve was produced by taking absorbance readings at 595 nm of the standard mixtures after 5 minutes at room temperature. Protein samples of unknown concentration were diluted 100 µl, 1 mL of Bradford reagent added and absorbance recorded in the same manner as above. Protein concentrations were determined in mg $mL^{-1}$ using Cary UV concentration software.

Protein concentration was determined by A280 and A205 measurements using the extinction coefficients 1490 $M^{-1}\cdot cm^{-1}$(rrPar-4SAC) and the relationship described by Scopes[82]. The purified samples were assessed by MALDI-TOF mass spectroscopy (Centre for Protein Research, University of Otago, Dunedin, New Zealand).

***NMR spectroscopy***

NMR experiments were performed on a Bruker Avance 700 MHz spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) equipped with a cryoprobe, four rf-channels and gradient pulse capabilities. All spectra were acquired at $5^0$C on 300µL samples containing 5% $D_2O$ in Shigemi NMR tubes. The rrPar-4FL sample

concentration was 0.48 mM in 10 mM Tris (pH 7.0), 20 mM NaCl. The rrPar-4DLZ construct was uniformly $^{15}N$ labelled with a protein concentration of 0.09 mM in 20 mM $NaPO_4$ (pH 7.5), 100 mM NaCl, 1 mM dithiothreitol. Similarly rrPar-4SAC was uniformly $^{15}N$ labelled at a concentration of 0.34 mM in 10 mM Tris (pH 7.0), 20 mM NaCl.

$^{1}H$-$^{15}N$ HSQC spectra were recorded with the settings: rrPar-4FL: 200 transients, 2048 x 128 points ($F_2$ x $F_1$) and spectral widths of 8389.2 and 2128.9 Hz for $F_2$ and $F_1$, respectively; rrPar-4DL*Z*: 20 transients, 2048 x 128 points ($F_2$ x $F_1$) and spectral widths of 8389.2 and 2128.9 Hz for $F_2$ and $F_1$, respectively; rrPar-4SAC: 24 transients, 2048 x 256 points ($F_2$ x $F_1$) and spectral widths of 8389.2 and 2128.9 Hz for $F_2$ and $F_1$, respectively. All data sets were linear predicted and zero-filled once in the indirect dimension before Fourier transformation and final processing. Spectra were apodised using a shifted ($\pi$/6) squared sinusoidal bell function using TopSpin 2.1 (Bruker BioSpin GmbH, Rheinstetten, Germany). The $^{1}H$ and $^{15}N$ chemical shifts were referenced to the water signal[83].

***Circular Dichroism (CD) Spectropolarimetry:***

CD spectropolarimetry was used to assess the proportion of stable secondary structure and the thermal stability of the Par-4 SAC. Spectra were recorded on a Chirascan CD spectropolarimeter (Applied Photophysics, Leatherhead, UK) equipped with a recirculating water bath. Samples were at a concentration of 0.3 mg-$mL^{-1}$ in native buffer (10 mM Tris, pH 7.0, 20 mM NaCl) or native buffer plus 1 or 6 m urea. Spectra were recorded in 0.5 nm steps from 260 to 190 nm with an integration time of 1 s at each wavelength. Three successive scans were recorded, the sample blank was subtracted and the scans were averaged and smoothed using a sliding window function. Thermal stability was determined by acquiring CD spectra as a function of temperature at $5^0$C intervals from 5-$75^0$C with 2 min of equilibration time at each temperature point. Deconvolution was performed using the contin ll algorithm[84] through the dichroweb server interface[85].

**Dynamic Light Scattering (DLS)**

The apparent Stokes radii of the rrPar-4 constructs were analysed using a Zetasizer Nano ZS (Malvern Instruments, Malvern, UK). Sample concentrations were 0.3

mg.mL$^{-1}$ in native buffer (10 mM Tris, pH 7.0, 20 mM NaCl) or native buffer plus 1 or 6 M urea. DLS data were obtained at 25$^0$C using a low-volume disposable 1 cm pathlength plastic cuvette (Sarstedt, Nürnbrecht, Germany) and five successive scans were collected and averaged for each protein sample. Samples were prepared 1 day in advance and maintained overnight at 4$^0$C to allow any bubbles to dissipate and were then allowed to equilibrate to 25$^0$C before measurements were made. The diffusion coefficients were extracted from the correlation curve and the hydrodynamic radius was calculated using the Stokes–Einstein equation[47,48]. The highest peak of the resulting histogram recorded for each sample was taken as the mean diameter for that particular sample.

## REFERENCES


1. Williams G. T., Programmed cell death: apoptosis and oncogenesis. *Cell*, **1991**, 65, 1097-1098.

2. Wyllie A. H., Apoptosis and the regulation of cell numbers in normal and neoplastic tissues: An overview. *Cancer Metast Rev.* **1992**, 11, 95-03.

3. Kerr JFR, Winterford C. M., Haroom V. Apoptosis: Its significance in cancer and cancer therapy. *Cancer*, **1994**, 73, 2013-2026.

4. Johnstone, R. W., A. A. Ruefli, and S. W. Lowe. Apoptosis: a link between cancer genetics and chemotherapy. *Cell*, **2002,** 108, 153-164

5. Guo, Q., Fu W., Xie, J., Sells, S. F., Geddes, J. W., Bondada, V., Rangnekar, V.M., and Mattson, M.P. Par-4 is a mediator of neuronal degeneration associated with the pathogenesis of Alzheimers disease. *Nat. Med.* **1998**, 4, 957-962

6. Mattson, M.P., Duan, W., Chan, S.L. and Camandola, S., Par-4 – An emerging pivotal player in neuronal apoptosis and neurodegenerative disorders. *J. Mol. Neurosci.* **1999**, 13, 17-30

7. Duan, W., Guo, Z., and Matson, M.P. Participation of par-4 in the degeneration of striatal neurons induced by metabolic compromise with 3-nitropropionic acid. *Exp. Neurol.* **2000**, 165, 1-11

8. Sells S. F., Wood D. P., Joshi-barve S. S., Muthukumar S., Jacob R. J., Crist S. A., et al. Commonality of the gene programs induced by effectors of apoptosis in androgen-dependent and –independent prostate cells. *Cell Growth Differ.* **1994**, 5, 457-466.

9. Sells S. F., Hans S. S., Muthukumar S., Maddiwar N., Johet. al. Nstone R.,

Boghaert E., et al. Expression and function of the leucine zipper protein Par-4 in apoptosis. *Mol. Cell. Biol.* **1997**, 17, 3823-3232.

10. Guo Q., Fu W., Xie J., Luo H., Sells S. F., Geddes J. W., et al. Par-4 is a mediator of neuronal degeneration associated with the pathogenesis of Alzheimer's disease. *Nature Med.* **1998**, 4, 957-962.

11. Chakraborty, M., S. G. Qui, K. M. Vasudevan, and V. M. Rangnekar, Par-4 drives trafficking and activation of Fas and Fasl to induce prostate cancer cell apoptosis and tumor regression. *Cancer Res.* **2001**, 61, 7255-7263.

12. Nalca, A., S. G. Qui, N. El-Guendy, S. Krishnan, and V. M. Rangnekar, Oncogenic Ras sensitizes cells to apoptosis by Par-4. *J. Biol. Chem.* **1999**, 274, 29976-29983.

13. R.W. Johnstone, R.H. See, S.F. Sells, J. Wang, S. Muthukumar, C. Englert, D.A. Haber, J.D. Licht, S.P. Sugrue, T. Roberts, V.M. Rangnekar, Y. Shi, A novel repressor, par-4 modulates transcription and growth suppression function of the Wilms' tumour suppressor WT1, *Mol. Cell. Biol.* **1996**, 16, 6945-6956.

14. M.T. Diaz-Meco, M.M. Municio, S. Frutos, P. Sanchez, J. Lozano, L. Sanz, J. Moscat, The product of par-4, a gene induced during apoptosis, interacts selectively with the atypical isoforms of protein kinase C, *Cell* 1996, 86,777-786.

15. M.P. Mattson, S.Camandola, NF-kappaB in neuronal plasticity and neurodeg-enerative disorders, *J. Clin. Invest.* **2001**, 107, 247-254.

16. Ranganathan, P. and Rangnekar, V.M., Regulation of cancer cell survival by Par-4. *Ann. N.Y. Acad. Sci.* **2005**, 1059, 76-85

17. El-Guendy, N., and V.M. Rangnekar, Apoptosis by Par-4 in cancer and neurodegenerative diseases. *Exp. Cell Res.* **2003**, 283, 51-66.

18. N. El-Guendy, V.M. Rangnekar, Nuclear translocation of Par-4 is essential for inhibition of RelA in a ζPKC dependent manner, in: Proceedings of the American Association for Cancer Research, San Francisco, CA, **2002**.

19. El-Guendy, N., Zhao, Y., Gurumuthy, S., Burikhanow, R. and Rangnekar, V.M., Identification of a unique core domain of Par-4 sufficient for selective apoptosis induction in cancer cells. *Mol. Cell. Biol.* **2003**, 23, 5516-5525

20. Sells S.F., *et.al.* Expression and function of the leucine zipper protein Par-4 in apoptosis. *Molecular and Cell Biology*, **1997**, 17, 3823-3832.

21. Dutta, K., Alexandrow, A., Huang, H. and Pascal, S.M., pH induced folding of an apoptotic coiled coil. *Protein Sci.* **2001**, 10, 2531-2540

22. Dutta, K., Engler, F.A., Cotton, L., Alexandrow, A., Bedi, G.S., Colquhoun and Pascal, S.M., Stabilization of a pH-sensitive apoptosis-linked coiled coil through single point mutations. *Prot. Sci.* **2002**, 12, 257-265

23. Diaz-Meco, M.T., Municio, M.M., Frutos, S., Sanchez, P., Lozano, J., Sanz, L. and Moscat, J., The product of par-4, a gene induced during apoptosis, interacts selectively with the atypical isoforms of protein kinase C. *Cell* **1996**, 86, 777-786

24. Sells, S.F., S.S. Han, S. Muthukumar, N. Maddiwar, R. Johnstone, E. Boghaert, D. Gillis, G. Liu, P. Nair, S. Monnig, P. Collini, M.P. Mattson, V.P. Sukhatme, S.G. Zimmer, D.P. Wood, Jr., J.W. McRoberts, Y. Shi, and V.M. Rangnekar, Expression and function of the leucine zipper protein Par-4 in apoptosis. *Mol. Cell. Biol.* **1997**, 17, 3823-3832.

25. Camandola S. and Mattson M.P., Pro-apoptotoc action of PAR-4 involves inhibition of NF-kappaB activity and suppression of BCL-2 expression. *J Neurosci Res.* **2000,** 61, 134-139.

26. J.R. Herman, S. Gurumurthy, M. Chakraborty, V.M. Rangnekar, Par-4 causes regression of orthotopic tumors in immunocompetent mouse prostate cancer model, in: Proceedings of the American Association for Cancer Research, Naples, F.L., **2001**.

27. Wagner, G. Prospects for NMR of large proteins. *J. Biomol. NMR* **1993,** 3, 375-385.

28. Kay, L.E. and Gardner, K.H. Solution NMR spectroscopy beyond 25kDa. *Curr. Opin. Struct. Biol.* **1997,** 7, 564-570.

29. Wüthrich, K. *NMR of Proteins and Nucleic Acids,* Wiley, New York **1986.**

30. Wagner, G., Anil-Kumar and Wüthrich, K. Systematic application of two dimensionnal 1H nuclear magnetic resonance techniques for studies of proteins 2: combined use of correlated spectroscopy and nuclear overhauser spectroscopy for sequential assignments of backbone resonances and elucidation of polypeptide secondary structures. *Eur. J. Biochem.* **1981**, 114, 375-384.

31. Lipstiz R. S., Tjandra N. Residual Dipolar couplings in NMR structure analysis. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, 33, 387-413.

32. Wüthrich, K., Billeter, M. and Braun, W. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J. Mol. Biol.* **1984**, 180, 715-740.

33. Alexandrov, A., Dutta, K, and Pascal, S.M. MBP-fusion with a viral protease cleavage site: One-step cleavage/purification of insoluble proteins. *Biotechniques*, **2001**, 30,1194-1198.

34. Cordingley, M.G., P.L. Callahan, V.V. Sardana, V.M. Garsky, and R.J. Colonno. Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J. Biol. Chem.* **1990**, 265, 9062-9065.

35. Walker, P.A., L.E. Leong, P.W. Ng, S.H. Tan, S. Waller, D. Murphy, and A.G. Porter. Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology*, **1994**, 12, 601-605

36. LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., and McCOy, J.M., *B*io/Techonology, **1993**, 11, 187-193

37. Goloubinoff, P., Gatenby, A.A and Lorimer, G.H. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **1989**, 337**,** 44-47.

38. Hartl, F. U., Hlodan, R., and Langer, T. *Trends Biochem. Sci.* **1994**, 19**,** 20-25.

39. Hendrick, J. P., and Hartl, F. U. *Annu. Rev. Biochem.* **1993**, 62**,** 349-384.

40. Baneyx, F. *Ann. N. Y. Acad. Sci.* **1994**, 745**,** 383-394.

41. L.P. McIntosh and F.W. Dahlquist *Quart. Rev. Biophys.* **1990**, 23, 1-38.

42. Libich D.S., Schwalbe M., Kate S., Venugopal H., Claridge J.K., Edwards, P.J., Dutta K., Pascal SM. Intrinsic disorder and coiled-coil formation in prostate apoptosis response factor 4. *FEBS J*, **2009**, 276, 3710–3728.

43. Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B & Longhi S Assessing protein disorder and induced folding. *Proteins*, **2006**, 62, 24–45.

44. Dyson HJ & Wright PE Elucidation of the protein folding landscape by NMR. *Methods Enzymol*, **2005**, 394, 299–321.

45. Martin Schwalbe, Kaushik Dutta, David S. Libich, Hariprasad Venugopal, Jolyon K. Claridge, David A. Gell, Joel P. Mackay, Patrick J. B. Edwards, and Steven M. Pascal, Two-state conformational equilibrium in the Par-4 leucine zipper domain, *Proteins,* **2010** Apr 15;78(11):2433-2449

46. Receveur-Brechot V., Bourhis J.M., Uversky V.N., Canard B. & Longhi S. Assessing protein disorder and induced folding. *Proteins*, **2006**, 62, 24–45.

47. Uversky V. N., What does it mean to be natively unfolded? *Eur. J. Biochem.*, **2002**, 269, 2–12.

48. Uversky V.N., Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule.*Biochemistry,* **1993**, 32, 13288–13298.

49. Dunker A.K., Lawson J.D., Brown C.J., Williams R.M., Romero P., Oh J.S., Oldfield C.J., Camp -en A.M., Ratliff CM, Hipps K.W. et. al. Intrinsically disordered protein. *J. Mol. Graph. Model*, **2001**, 19, 26–59.

50. Garnier J, Gibrat J. F. & Robson B., GOR method for predicting protein secondary structure from amino acid sequence. *Comput. Methods. Macromol. Seq. Ana.,* **1996,** 266, 540–553.

51. Tompa P., Intrinsically unstructured proteins. *Trends Biochem. Sci*., **2002**, 27, 527–533.

52. Uversky VN, Gillespie JR & Fink AL, Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins*, **2000**, 41, 415–427.

53. Jack Kyte and Russell F. Doolittle, A simple method for displaying the hydropath-ic character of a protein, *J. Mol. Biol.* **1982**, 157, 105-132

54. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ & Russell RB Protein disorder prediction: implications for structural proteomics. *Structure,* **2003**, 11, 1453–1459

55. Spolaore B, Bermejo R, Zambonin M & Fontana A., Protein interactions leading to conformational changes monitored by limited proteolysis: apo form and fragments of horse cytochrome c. *Biochemistry*, **2001**, 40, 9460–9468.

56. Hubbard S. J., The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta.*, **1998**, 1382, 191–206.

57. Dyson H.J. & Wright P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005** 6, 197–208.

58. Uversky V.N., Oldfield C.J. & Dunker A.K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.*, **2005,** 18, 343–384.

59. Dunker A.K., Obradovic Z., Romero P., Garner E.C. &Brown C.J., Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform.* **2000**, 11, 161 -171.

60. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z & Dunker AK Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol Biol.* **2002,** 323, 573–584.

61. Hanahan D. & Weinberg R.A. The hallmarks of cancer. *Cell,* **2000**,100, 57–70.

62. Romero P., Obradovic Z., Li X., Garner E.C., Brown C.J. & Dunker A.K., Sequence complexity of disordered protein. *Proteins,* **2001,** 42, 38–48.

63. Callebaut I., Labesse G., Durand P., Poupon A., Canard L., Chomilier J., Henrissat B. & Mornon J.P., Deciphering protein sequence information through Hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol. Life Sci*. **1997**, 53, 621–645.

64. Shortle D., The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J*. **1996,** 10, 27–34.

65. Uversky V.N. & Fink A.L., Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys. Acta*. **2004**, 1698, 131–153.

66. Hu C.C. & Ghabrial S.A., The conserved, hydrophilic and arginine-rich N-terminal domain of cucumovirus coat proteins contributes to their anomalous electrophoretic mobilities in sodium dodecylsulfatepolyacrylamide gels. *J. Virol. Methods* **1995**, 55, 367–379.

67. Goswami A., Qiu S., Dexheimer T.S., Ranganathan P., Burikhanov R., Pommier Y. & Rangnekar V.M. Par-4 binds to topoisomerase 1 and attenuates its DNA relaxation activity. *Cancer Res.* **2008**, 68, 6190–6198.

68. Vetterkind S., Illenberger S., Kubicek J., Boosen M., Appel S., Naim H.Y., Scheidtmann K.H. & Preuss U., Binding of Par-4 to the actin cytoskeleton is essential for Par-4/Dlk-mediated apoptosis. *Exp. Cell Res.* **2005**, 305, 392–408.

69. Lacy E.R., Wang Y., Post J., Nourse A., Webb W., Mapelli M., Musacchio A., Siuz dak G. & Kriwacki R.W. Molecular basis for the specificity of p27 toward cyclin-dependent kinases that regulate cell division. *J. Mol. Biol.* **2005**,349, 764–773.

70. Uversky V.N. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, **1993**, 32, 13288–13298.

71. Tcherkasskaya O. & Uversky V.N. Denatured collapsed states in protein folding: example of apomyoglobin. *Proteins*, **2001**, 44, 244–254.

72. Uversky V.N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **2002**, 11, 739–756.

73. Shi Z., Woody R.W. & Kallenbach N.R., Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem.* **2002**, 62, 163–240.

74. Whitmore L. & Wallace B.A. DICHROWEB, an online server for protein

secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **2004**, 32, W668–W673.

75. Dyson H.J. & Wright P.E. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* **2001** 339, 258-270.

76. Dyson H.J. & Wright P.E., Unfolded proteins and protein folding studied by NMR. *Chem Rev.* **2004**, 104, 3607–3622.

77. Oldfield C.J., Cheng Y., Cortese M.S., Romero P, Uversky V.N. & Dunker A.K., Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*, 2005,44, 12454–12470.

78. Gurumurthy S., Goswami A., Vasudevan K.M. & Rangnekar V.M., Phosphorylation of Par-4 by protein kinase A is critical for apoptosis. *Mol. Cell Biol.*, **2005**, 25, 1146–1161.

79. Goswami A., Burikhanov R., de Thonel A., Fujita N., Goswami M., Zhao Y., Eriksson J.E., Tsuruo T. & Rangnekar V.M., Binding and phosphorylation of par-4 by akt is essential for cancer cell survival. **2005**, *Mol.Cell,* 20, 33–44.

80. Laemmli, U. K., Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* **1970**, 227, 680-685.

81. Bradford, M. M., A rapid and sensitive method for the quantification ofmicrogram quantities of protein using the principles of protein-dye binding. *Anal. Biochem.* **1976**, 72, 248-254.

82. Scopes R.K., Measurement of protein by spectrophotometry at 205 nm. *Anal. Biohem.*, **1974**, 59, 277–282.

83. Wishart D.S., Bigam C.G., Yao J., Abildgaard F., Dyson H.J., Oldfield E., Markley J.L. & Sykes B.D. $^{1}H$, $^{13}C$ and $^{15}N$ chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **1995**, 6, 135–140.

84. van Stokkum I.H., Spoelder H.J., Bloemendal M., van Grondelle R. & Groen F.C. Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem.*, **1990**, 191, 110–118.

85. Whitmore L. & Wallace B.A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **2004**, 32, W668–W673.