Structural and functional characterisation of the protein inhibitor of activated STAT3 (PIAS3)

A thesis submitted in fulfilment of
the requirements for the degree

of

DOCTOR OF PHILOSOPHY
IN BIOCHEMISTRY

of

RHODES UNIVERSITY

in the

FACULTY OF SCIENCE

BIOMEDICAL BIOTECHNOLOGY RESEARCH UNIT (BioBRU)

DEPARTMENT OF BIOCHEMISTRY, MICROBIOLOGY & BIOTECHNOLOGY

by

NICODEMUS MAUTSA

JANUARY 2011

ABSTRACT

The signal transducer and activator of transcription (STAT) and protein inhibitor of STAT (PIAS) system represent an elegant regulatory mechanism of transcriptional control IN mammalian cytokine signalling. Abnormal activation of the system is associated with immune disorders and a large group of diverse tumours. PIAS3 is a multiple domain protein with distinct functions involved in regulation of cytokine-mediated gene activation pathways. Its over-expression significantly inhibits cell growth and renders cancer cells more sensitive to drugs. The objective of this study was to structurally and biochemically characterise the function of the PIAS3 protein using *in silico*, *in vivo* and *in vitro* analysis approaches. The conservation pattern of the PIAS protein family and critical conserved residues in the PINIT (Proline, Isoleucine, Asparagine, Isoleucine, Tyrosine) domain were identified. The PINIT domain model was generated based on the PINIT domain structure of yeast PIAS3 homologue Siz1 and structural determinants in the PIAS3-STAT3 interaction were evaluated. Guided by the *in silico* findings, *in vivo* analysis of the localisation of the PIAS3, mutant derivatives of PIAS3 (PIAS3-L97A, PIAS3-R99N, PIAS3-R99Q), PINIT and acidic domain was conducted. PIAS3 was completely localised in the nucleus while PIAS3 mutants appeared to exhibit diffuse cytoplasmic distribution. The PINIT domain was predominantly localised in the nucleus with some apparent perinuclear staining while the acidic domain exhibited a predominantly perinuclear staining pattern. Further analysis of the PINIT domain and the effect of the mutants on PIAS3-STAT3 interaction were assessed by *in vitro* analysis. Guided by *in silico* analysis, the PINIT domain and mutant derivatives of PINIT domain (PINIT-L97A, PINIT-R99N, and PINIT-R99Q) were heterologously expressed in *Escherichia coli* and subsequently purified using a combination of immobilized metal affinity and size exclusion based chromatography. The size and structural elements of the PINIT domain and its mutants were characterised. The 23 kDa PINIT domain was found to exist as a monomer in solution and its secondary structure was shown to consist of 66 % β-sheets by fourier transformed infrared spectroscopy consistent with the generated homology model. Using surface plasmon resonance spectroscopy (SPR) the PINIT domain was shown to bind to STAT3 in a specific concentration dependent manner. Recombinant PINIT-L97A, PINIT-R99N and PINIT-R99Q mutants, which exhibited similar structural integrity to the wildtype, were found to abrogate binding to STAT3. These findings suggest that these residues form part of a potential binding surface for stat3. In conclusion, this study has provided evidence

that the PINIT domain is an important determinant of PIAS3 interaction with STAT3 and that the interaction is mediated by defined conserved residues directly involved in the PINIT-STAT3 interaction.

DECLARATION

I, **NICODEMUS MAUTSA**, declare that this is my own unaided work. It is being submitted for the degree of Doctor of Philosophy in Rhodes University in the Faculty of Science. It has not been submitted for any degree for examination in any other university.

_____

NICODEMUS MAUTSA

DATED THIS _____ DAY OF _____ AT _____.

DEDICATION

This thesis is dedicated to my late father

Mr Andrew L. Mautsa

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Gregory L. Blatch for his endless support, guidance, encouragement, his passion and enthusiasm for my research project.

I wish to acknowledge my co-supervisors; Dr Earl Prinsloo and Dr Özlem Tastan Bishop for their technical assistance, encouragement and guidance.

I wish to extend further acknowledgements to; Professor Chris Whiteley, Dr Aileen Boschof for their guidance and helpful contributions as members of PhD steering committee.

I would like to recognise the following individual for their assistance in proof reading my thesis; Dr Eva Persce, Dr Earl Prinsloo, Wilbert Kadye and James Njunge.

I gratefully acknowledge Dr Hélène Boeuf (Université de Bordeaux, France) for the gift of the plasmid p513-flag-PIAS3 and Dr Christoph Müller (EMBL, Germany) for the kind donation of pET32b-STAT3β plasmid.

I wish to thank members of the Biomedical Biotechnology Research Unit (BioBRU) (Rhodes University) for their support.

I gratefully acknowledge the financial assistance of the National Research Foundation (NRF); the research was funded by a NRF and Research Niche Area (RNA) grant.

RESEARCH OUTPUTS

Mautsa N, Prinsloo E, Bishop OT and Blatch GL. 2010. The PINIT domain of PIAS3: structure-function analysis of its interaction with STAT3. *Journal of Molecular Recognition.* DOI: **10:1002**/jmr.1111

Mautsa N, Prinsloo E and Blatch GL. Functional characterisation and structural elucidation of the protein inhibitor of activated STAT3 (PIAS3) domain(s). 22nd SASBMB conference, 2010, Bloemfontein, South Africa

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A/T	Adenine/Thymine
CCD	Coiled–coil domain
CNTF	Ciliary neurotrophic factor
DBD	DNA–binding domain
DOPE	Discrete optimised protein energy
EGFP	Enhanced green fluorescent protein
FTIR	Fourier transform infrared resonance
GBM	Glioblastoma multiform
GST	Glutathione S–transferase
HeLa	Human cervix carcinoma cell line
His–tag	Histidine tag
IL-6	Interleukin–6
JAK	Janus Kinase
KChAP	K^+ channel associated protein
MAR	Matrix attached region
MITF	Microphthalmia transcriptional factor
Miz1	Msx–interacting zinc finger
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
PDB	Protein data base
PIAS3	Protein inhibitor of activated STAT3
PINIT	Proline, isoleucine, asparagine, isoleucine, threonine
RING	Real interesting gene

S/T	Serine/Theronine
SAP	SAF–A/B, Acinus and PIAS motif
SAR	Scaffolding attached region
SDS–PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SIM	SUMO interacting motif
SPR	Surface Plasmon resonance
STAT3	Signal transducers and activators of transcription 3
SUMO–1	Small ubiquitin–like modifier 1
ZIMIZ	Zinc finger MIZ domain–containing protein

CHAPTER 1

LITERATURE REVIEW, PROBLEM STATEMENT AND OBJECTIVES

The signal transducer and activator of transcription (STAT)/protein inhibitor of STAT (PIAS) system represents an elegant system of control of mammalian cytokine signalling. Abnormal activation of transcriptional factors such as STAT proteins is associated with immune disorders and is found in a large group of diverse tumours. PIAS3 regulates cytokine–mediated gene activation pathways and its overexpression significantly inhibits cell growth and also renders cancer cells more sensitive to drugs. This review captures the historic and current literature of the system, with a focus on Janus kinase (JAK)–STAT signalling, STAT3 and its natural regulator PIAS3. Critique of the available structure/function data provides a view of the problems and current knowledge gap.

1.1 INTRODUCTION

Signal transducers and activators of transcription (STAT) are a family of transcription factors that activate gene transcription in response to a number of different cytokines (O'Shea, 1997; Leonard and O'Shea, 1998; Hoey and Grusby, 1999). STAT proteins are expressed in diverse mammalian tissues and have been implicated in oncogenesis. The STAT protein family consists of several members all encoded by separate genes theorised to have diverged from a single gene through several consecutive duplications into three genetic *loci* (Copeland *et al*., 1995). STAT proteins share six conserved structural regions that mediate cooperative binding to multiple DNA sites (Xu *et al*., 1996; Vinkemeier *et al*., 1996). The protein inhibitor of activated STAT (PIAS) protein family is a group of nuclear proteins that modulate transcriptional activities of various transcriptional factors. PIAS proteins are expressed in multiple human tissues (Chung *et al*., 1997) that includes lungs and kidneys (Wible *et al*., 1998). PIAS are multidomain proteins with distinct functions (Chung *et al*., 1997; Levy *et al*., 2006; Yagil *et al*., 2009; Borghouts *et al*., 2010). PIAS proteins exist as splice variants of different PIAS isoforms and homologues were identified in animal, plant and yeast species (Johnson and Gupta, 2001; Takahashi *et al*., 2001; Zhao and Blobel, 2005; Cheng *et al*., 2006). Many different proteins, in particular transcription factors interact with members of the PIAS family (Shuai and Liu, 2005). Various regions of PIAS proteins are involved in different protein–protein interactions (Liao *et al*., 2000). The JAK (Janus kinase)–STAT pathway is activated by cytokine binding to its receptor and STAT is phosphorylated; these activation pathways are tightly controlled by positive and negative regulators such as the PIAS proteins. Uncontrolled cytokine signalling is associated with immune disorders and large group of diverse tumours (Brantley *et al*., 2008). The focus of this study is on the structural and molecular determinants of PIAS3-STAT3 interaction.

1.2 THE STAT PROTEIN FAMILY

The STAT protein family consists of seven members (STAT1, STAT2, STAT3, STAT4, STAT5α, and STAT5β AND STAT6) all encoded by separate genes (Darnell *et al*., 1994; Zhong *et al*., 1994). Due to the conserved nature of the family members, it is assumed that they have diverged from a single gene through several consecutive duplications into three genetic *loci* (Copeland *et al*., 1995). In humans, the STAT genes were mapped into distinct chromosomes (Copeland *et al*., 1995). STAT3, STAT5α and STAT5β were mapped to chromosome 17 (bands q11–1 to q22), whereas STAT1 and STAT4 were mapped to

chromosome 2 (bands q12 to q33) (Yamamoto *et al*., 2003). In mouse, STAT1 and STAT4 were located on chromosome 1 (band 1 C1.1) (Schindler *et al*., 1992; Yamamoto *et al*., 2003) and STAT2 and STAT6 were located on chromosome 10 (band 10 D3) (FU *et al*., 1992; Quel *et al*., 1995). STAT3, STAT5α and STAT5β were mapped to chromosome 11 (band 11 D) (Zhong *et al*., 1994a; Copeland *et al*., 1995; Shi *et al*.,1996; Levy *et al*.,1998). Although the expression patterns of STAT protein family members vary during cell development and in multiple tissue types, nevertheless, they are highly sequence conserved (Akira *et al*., 1999).

1.2.1 STRUCTURE AND FUNCTION OF STAT3 PROTEIN

STAT proteins share six structural regions **Figure 1.1**. The STAT3 N–terminal domain (ND) consist of 130 residues that mediate cooperative binding to multiple DNA sites (Xu *et al*., 1996; Vinkemeier *et al*., 1996). The 4–helix bundle coiled–coil domain (CCD) (residues 130 to 320) is immediately followed by the eight–stranded β–barrel forming the DNA binding domain (DBD) with residues 400–500 conferring DNA–binding specificity; however this domain is not sufficient for optimal DNA binding (Horvath *et al*., 1995). The DNA binding β–barrel domain is linked to the SH2 domain by a small helical domain, formed by two helix–loop–helix modules, called the 'connector' domain. This domain shows structural similarity to calcium–binding domains. However, the loops of the connector domain are longer and the connector domain shows no structural similarity to SH3 domains as was previously suggested through sequence identity (Becker *et al*., 1998). Residues 600–700 share homology with SH2 domains (SH2) and mediate dimerisation as a result of phosphotyrosine recognition (Shuai and Liu. 2005) and the transcriptional activation domain (**Figure 1.1B**). The phosphorylated tyrosine is located at the C–terminus and the position varies with family member (**Figure 1.1A**). In addition, the C–terminus is important for transcriptional activation, which can be regulated by serine phosphorylation (Wen *et al*., 1995; Akira, 1999).

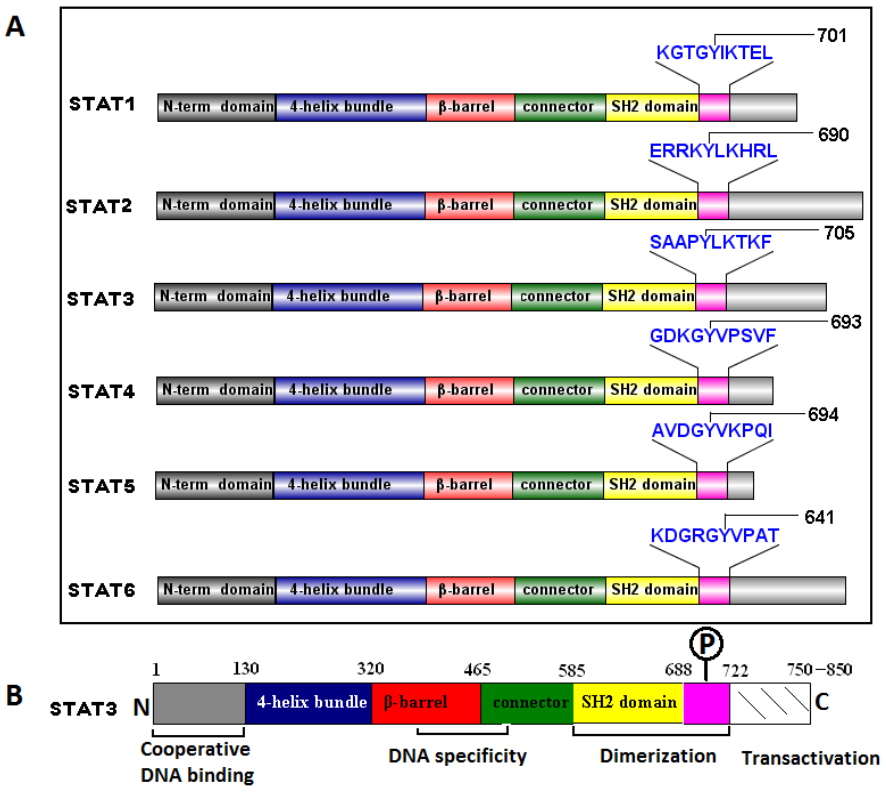

Figure 1.1 Schematic illustrations of STAT family members, their domains and structural features. A) N–terminal domain (ND); 4–helix bundle coiled–coil domain (CCD); β–barrel; connector; Src homology domain 2(SH2) and the tyrosine residue (**Y**) phosphorylation site indicated by a number in each particular schematic STAT protein (adapted from Akira, 1999). **B**) Schematic representation of STAT3 functional features; the N–terminal domain (residues 1–130) that mediates cooperativity in binding to multiple DNA sites; residues 400–500 confer DNA–binding specificity; Residues 600–700 share homology with Src–homology–2 (SH2) domains and mediate dimerization as a result of phosphotyrosine; the phosphorylated tyrosine is located around residue Tyr 705; the C–terminus is important for transcriptional activation (adapted from Becker *et al*., 1998).

These distinct functional domains within the STAT molecules were identified by sequence comparisons, biochemical assays and mutagenesis. Upon receptor activation, a single tyrosine residue (Y705 in STAT3) is phosphorylated (Akira, 1999; Shuai, 2006). Recombinant C–terminal fragment of STAT3β (residues 127 to 722) was phosphorylated at Tyr 705 causing it to dimerize and bind to specific DNA oligonucleotides, thus enabling the crystal structure of the STAT3β homodimer bound to DNA to be determined (**Figure 1.2A**) (Becker *et al*., 1998). Braunstein *et al*. (2003) reported that the C-terminal STAT3β fragment (residues 127–722) occurred as a monomeric species to a greater extent compared to full length STAT3. These findings showed that the full length dimer is mostly mediated by N–terminal domain interactions. This conclusion was substantiated by the crystal structures of unphosphorylated STAT1 and STAT5 which showed that the core fragment (residues ~130 to ~680) formed a reciprocal dimer involving CCD and DBD (Mao *et al*., 2005; Neculai *et al*.,

2005). Although different STAT proteins have similar domains, each has its unique structure and biochemical features and these differences directly correlate with their specific biological functions where STAT1 is a tumour suppressor whereas STAT3 is an oncogene (potential tumour promoter).

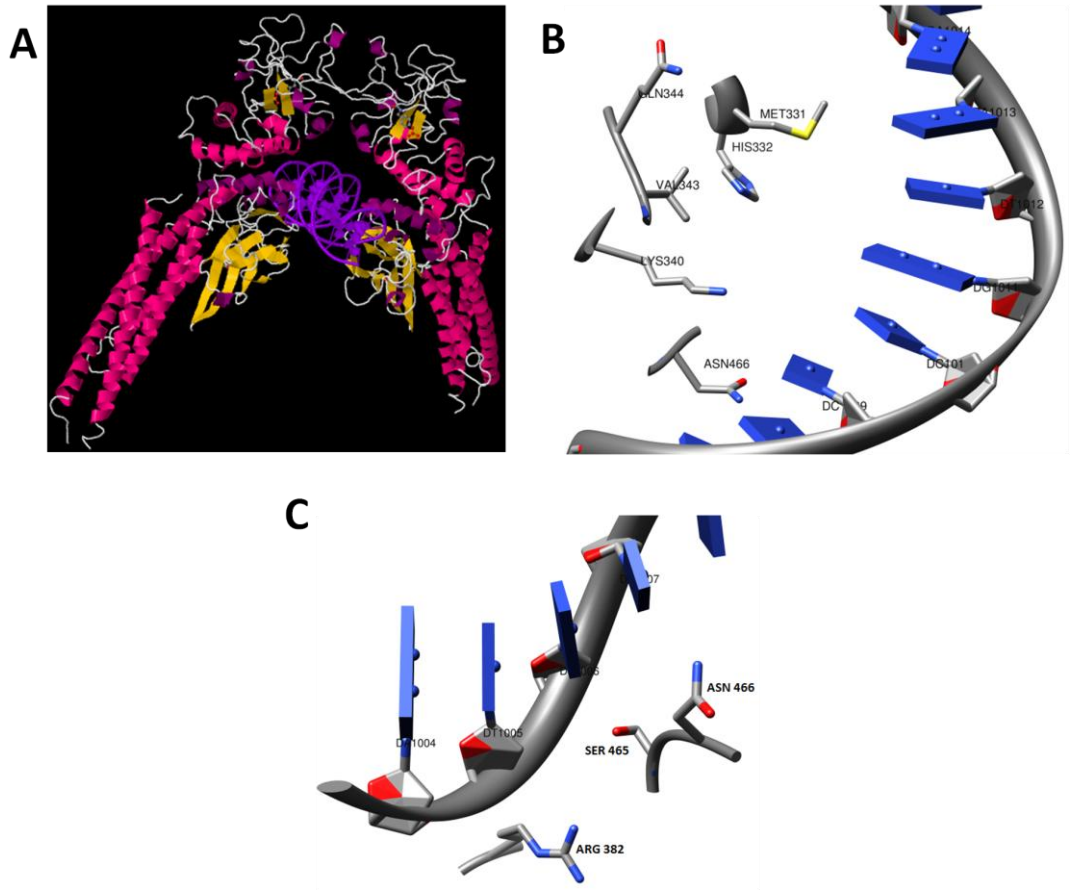

Figure 1.2 Three-dimensional representation of STAT3β homodimer–DNA complex and DB domain residues. A) STAT3β homodimer–DNA–complex showing the N–terminal 4–helix bundle (hot pink), the β–barrel domain (orange), the connector domain in and the SH2 domain and phosphotyrosine–containing region in yellow. Views are shown along the DNA axis running vertically Rendered in JmoL (www.jmol.org) (adapted from Becker *et al*., 1998). B) Three dimensional representation of the residues and DNA phosphate groups involved in STAT3β–DNA–complex formation in the first DNA strand, atoms are coloured by element where oxygen atoms (red), nitrogen atoms (blue), sulphur atoms (yellow) are colour coded and the rest of the atoms are in grey. C) Three dimensional representation of the residues and DNA phosphate groups involved in STAT3β–DNA–complex formation with the second DNA strand, atoms are coloured by element where oxygen atoms (red) nitrogen atoms (blue) are colour coded and the rest of the atoms are in grey The Figures were rendered in UCSF Chimera 1.5 (Goddard, *et al*., 2005).

STAT3 has been suggested to associate with PIAS3 protein through its DB domain and the CCD (Chung *et al*., 1997; Borghouts *et al*., 2010). Furthermore, the STAT3 DB domain has been implicated in the regulation of nuclear export in resting cells (Zhiyuan and Kone, 2004). The crystal structure determined by Becker *et al*. (1998) showed that the DB domain

residues, His 332, Lys 340 and Gln 344 formed polar contacts to the phosphate group of one DNA strand while residues Met 331 and Val 343 formed hydrophobic interactions with its backbone sugars. The base specificity was determined by residues Asn 466 (**Figure 1.2B**). The other stand of DNA was bound by residues Arg 382, Val 432, Ser 465 and Gln 469 (**Figure 1.2C**).

1.3 LOCALISATION, CO–LOCALISATION OF PIAS3 AND INTERACTING PROTEINS

The PIAS protein family is a group of nuclear proteins that modulate transcriptional activities of various transcriptional factors. In particular, PIAS3 regulates the transcriptional activity of STAT3 by inhibiting its DNA binding (Chung *et al*., 1997). Basal amounts of PIAS3 were found to be expressed in the nucleus of the majority of epithelial and endothelial cells (Wang and Banerjee, 2004). Also, 100/103 of samples examined by Wang and Banerjee (2004) showed that PIAS3 is expressed in a variety of human tumours including lung, breast, prostate, colorectal and brain cancer. The subcellular localisation of PIAS3 in the nucleus in both resting cells and stimulated cells have been shown in many studies (Kotaja *et al*., 2002; Duval *et al*., 2003; Sonnenblick *et al*., 2004; Man *et al*., 2006; Yamashina *et al*., 2006; Peng *et al*., 2010) (**Table 1.1**). However, the cytoplasmic localisation of PIAS3 has also been shown in both stimulated and unstimulated cells (**Table 1.1**), while the localisation in both nucleus and cytoplasm was shown in NIH3T3 mouse foetal fibroblast cells by the work of Rödel *et al*. (2000). Interestingly, live imaging of human pulmonary epithelial cells (A549 and H520) by Dabir *et al*. (2009) revealed PIAS3 trafficking from the cytoplasm to the nucleus and back into the cytoplasm after 30 minutes of stimulating the cells with epidermal growth factor (EGF). The findings were consistent with that of Man *et al*. (2006), where T47D breast cancer cells were stimulated with progesterone and PIAS3 was subsequently observed in both the cytoplasm and nucleus with predominant nuclear staining. However, HeLa cervical cancer cells under the same treatment were found to have complete nuclear localisation of PIAS3 (Man *et al*., 2006).

It has been suggested that localisation of the PIAS3 in the cytoplasm and nucleus is largely dependent on the associating proteins, and therefore, co–localise with these various proteins. Microphthalmia transcriptional factor (MITF) is a basic helix–loop–helix leucine zipper (bHLH–Zip) DNA–binding protein (Hodgkinson *et al*., 1993). MITF and PIAS3 co–localised

in the nucleus in resting cells (Sonnenblick *et al*., 2004). Glucocorticoid receptor–interacting protein 1 (GRIP1) is a transcriptional factor, a member of steroid receptor co-activator family and interacts with PIAS proteins. Substitution of the sumoylation sites in GRIP1 impaired its co–localisation resulting in diminished co–localisation with androgen receptor (AR) (Kotaja *et al*., 2002). Duplin is a negative regulator of β–catenin–dependent T–cell factor (Tcf) transcriptional activity in the Wnt signalling pathway and was identified as a PIAS3 binding protein(Yamashina *et al*., 2006). Co-expression of Myc–Duplin with Flag–PIAS3 in COS (CV-1 (simian) in origin and carrying the SV40 genetic material) cells resulted in nuclear co–localisation of the two proteins in punctate structures. Furthermore, biochemical immunoprecipitation assay indicated that HA–Duplin indeed formed a complex with Flag–PIAS3 in COS cells (Yamashina *et al*., 2006). Other PIAS family members were also evidently localised in nuclear punctate structures where they co–localised with their associating proteins. Of note is Dnmt3a, which is one of the three mammalian DNA methyltransferates that plays a crucial role in transcriptional silencing among other functions. GFP–PIAS1 and GFP–PIASxα were found co–localised with Dnmt3a in punctate structures exclusively in the nucleus of NIH3T3 cells (Ling *et al*., 2003). PIASxα, PIASxβ, PIAS1 and PIAS3, which interact with the small ubiquitin–related modifier SUMO–1 and its E2 conjugate, Ubc9 (Kotaja *et al*., 2002), were all found co–localised with SUMO–1 in COS–1 cell nuclei in punctate structures (Kotaja *et al*., 2002). Trim32, a RING domain ubiquitin–protein isopeptide ligase interacts and co–localises with PIASy and promotes PIASy ubiquitination and degradation (Albor *et al*., 2006). PIASy was shown to predominantly localise to the nucleus (Zoumpoulidou *et al*., 2004; Albor *et al*., 2006) and treatment with MG132 (Z-Leu-Leu-Leu-al is an inhibitor of proteasome) showed a reduction of nuclear localisation and accumulation in cytoplamic granules (Albor *et al*., 2006). PIASy co–localise with Trim32 concentrated around the nucleus (Reymond *et al*., 2001; Albor *et al*., 2006). Progesterone receptor (PR) is critical in cell proliferation and differentiation and its transcriptional activity is modulated by multiple protein co-factors. Endogenous PR was found to localise mainly in the nucleus and cytosol in the absence of progesterone treatment of T47D cells while PIAS3 resided mainly in the nucleus, as punctate structures (Man *et al*., 2006). Treatment of T47D cells with progesterone resulted in complete translocation of liganded–PR from cytoplasm to the nucleus where it co–localised with PIAS3 in the dot-like structures (Man *et al*., 2006).

Of significance to this study is the PIAS3–STAT3 association, which was first shown by protein immunoprecipitation with anti–PIAS3 from IL–6 treated M1 mouse myeloid leukemic cells (Chung *et al*., 1997). The PIAS3 was shown to freely interact with unphosphorylated and phosphorylated forms of STAT3 (Kotaja *et al*., 2002) and its movement from the cytoplasm to the nucleus has been shown to depend on the phosphorylated or unpohosphorylated state of STAT3 transcriptional factor (Dabir *et al*., 2009). However, Chung *et al*, (1997) and Borghouts *et al*, (2010) suggested that the PIAS3–STAT3 complex was only formed with the phosphorylated STAT3 as it translocated to the nucleus. It was shown that interleukin-6/interleukin-6 receptor (IL–6/IL–6R) treatment induced STAT3 to localise in the nucleus with MITF and PIAS3 whereas STAT3–Y705F was detected predominantly in the cytoplasm. However, STAT3–C (a constitutive form) showed co–localisation with PIAS3 and MITF in the nucleus (Sonnenblick *et al*., 2004). Cells transfected with mutant STAT3-Y705F and stimulated for 10 minutes showed very little nuclear PIAS3 compared to the cytoplasmic compartment (Dabir *et al*., 2009). The translocation of proteins, in particular STAT3, upon cytokine stimulation, is critical for the control of cell growth and proliferation.

Table 1.1 Localisation of PIAS3 in various cell lines.

Cell line	**Localisation**	**References**
NIH3T3 mouse foetal fibroblast	–predominantly nucleus –cytoplasm	Rödel *et al*., 2000
Hep3B human heptoma cells	–predominantly nucleus –cytoplasm	Dabir *et al*., 2009
COS–1	–complete nucleus localisation	Chung *et al*., 1997; Duval *et al*., 2003
HeLa cervical cancer cells	–predominantly nucleus –cytoplasm	Man *et al*., 2006
A549 and H520 human epithelial cells	–nucleus	Rödel *et al*., 2000
T47D Human ductal breast epithelial tumour cells	–nucleus	Man *et al*., 2006
M1 cells mouse myeloid leukemic cells	–predominantly nucleus –cytoplasm	Chung *et al*., 1997

1.4 JAK–STAT SIGNALLING PATHWAY REGULATION

JAK (Janus kinase)–STAT3 pathway is activated when a cytokine binds to its receptor leading to receptor phosphorylation. This site will then act as a docking site to recruit latent cytoplasmic STAT3. The STAT3 is phosphorylated by JAK and subsequently dimerises and leaves the receptor to translocate to the nucleus where it directs gene activation (Darnell *et al*., 1994; Levy *et al*., 2002). The JAK–STAT3 pathway can also be activated by growth factor receptors and certain kinases including IL–6, 10, FGFR (fibroblast growth factor), HGFR (hepatocyte growth factor also known as MET), VEGFR (VEGF receptor), EGF (epidermal growth factor), and PDGF (platelet–derived growth factor) (Darnell *et al*.,1994). Also, oncoproteins such as Src (sarcoma) and Abl (Abelson murine leukemia viral oncogene**)** are activators of STAT3 in the JAK–STAT3 signalling pathway. The cytokine IL–6 is a major systematic mediator of the acute phase response to infection and injury, which includes increased production of acute phase plasma proteins by the liver (e.g. fibrinogen, c–reactive proteins, α1–antichymotrysin) and enhanced immune function, increased B–cell differentiation and T–cell activation (Sehgal *et al*., 1989; Mickiewicz *et al*., 1995, 1998). IL–6 binds to the α–chain of the gp130 receptor to form a binary complex thereby inducing dimerisation of gp130. This dimerisation induces the activation of JAK family tyrosine kinases, (primarily JAK1, also JAK2 and Tyk) (Darnell *et al*., 1994, Heinrich *et al*., 1998) associated with the cytoplasmic tail of gp130, by tyrosine phosphorylation of the JAK kinases thus leading to their activation. Six discrete docking sites are created on the gp130 receptor by activated JAK kinases. These sites are proposed to recruit monomeric STAT3 in the cytoplasm via the SH2 domain of STAT3 (Heinrich *et al*., 1998). The event is followed by tyrosine phosphorylation of STAT3 (Tyr705) and subsequent dimerization and translocation into the nucleus allowing the dimeric STAT3 to bind target DNA motifs and modulate gene expression (Heinrich *et al*., 1998).

JAK–STAT3 signalling pathway is down-regulated at various steps (Shuai, 2000, 2006). SOCS (suppressor of cytokine signalling) can inhibit the JAK–STAT signalling pathway by inhibiting the tyrosine kinase activity of JAKs. Also, STAT signalling can be negatively regulated by PTPase (protein tyrosine phosphatase) by deactivating JAKs and STATs. In particular, protein tyrosine phophatase TC45 deactivates STAT1 by dephosphorylation of

STAT1 in the nucleus (Hoeve *et al*., 2002). PIAS proteins regulate the JAK–STAT pathway by inhibiting STAT transcriptional activation activity (Shuai and Liu. 2005; Shuai, 2000). PIAS proteins bind specifically to phosphorylated STAT dimers in the nucleus and prevent it from binding target DNA and thereby inhibiting STAT3–mediated gene activation. There is a specific PIAS inhibitor for each STAT signalling pathway (Liao *et al*., 2000). The regulation of the JAK–STAT3 signalling pathway is largely influenced by levels of either STAT3 or PIAS3 expression in any given cells. Furthermore, it is known that STAT protein arginine methylation (Mowen *et al*., 2001) affects the JAK–STAT pathway. Methylation of STAT1 prevented its association with PIAS1 resulting in increased amount of STAT1 available for DNA binding and gene induction.

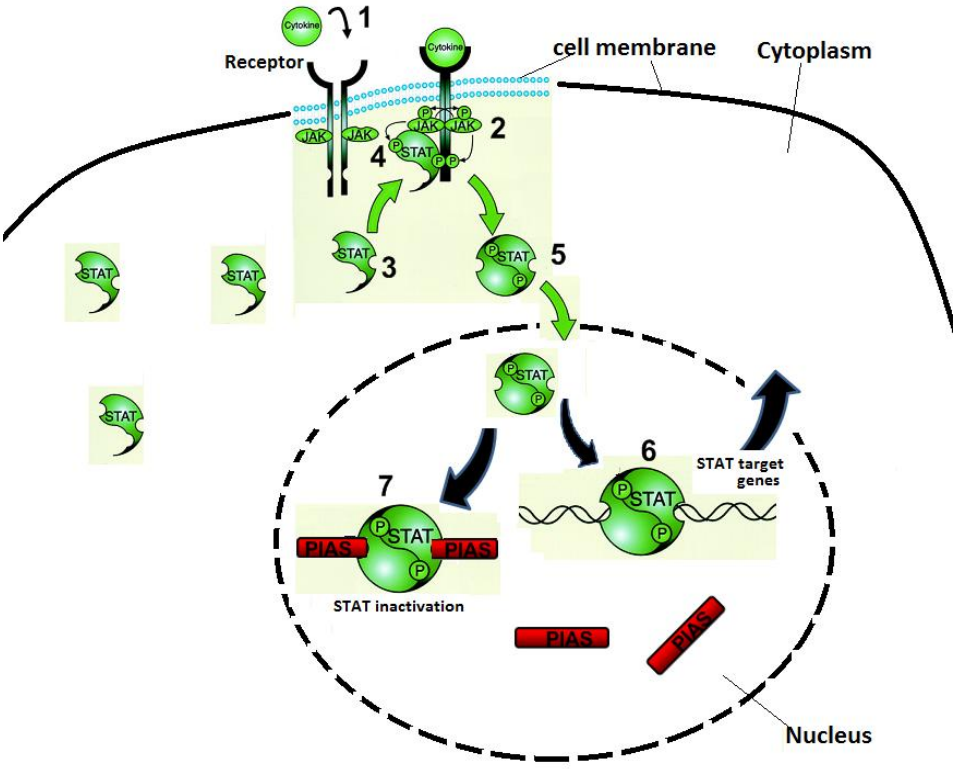

Figure 1.3 Schematic representation of canonical JAK–STAT pathway activation. Cytokine binds to the cytokine receptor at the cell surface (1) and activates JAK by phosphorylation (2). The activated JAK kinases recruit monomeric STAT to the receptor and (3) phosphorylate STAT at defined tyrosine residues (4). The phosphorylated STAT dimerises and translocates to the nucleus (5) where it induces the transcription of target genes. (6). Inhibition of STAT transcriptional activity by PIAS protein results in inactivation (7). Numbers indicate the flow of pathway on JAK–STAT activation and arrows indicate the direction of protein movement (adapted from Wormald and Hilton, 2004).

1.5 THE ROLE OF PIAS AND STAT PROTEINS IN CANCER AND OTHER HUMAN DISEASES

Cytokine–mediated gene activation pathways are tightly controlled by positive and negative regulators. Abnormal cytokine signalling is associated with cancer and immune disorders. STAT proteins activate transcription in response to numerous cytokines controlling proliferation, gene expression, and apoptosis. Aberrant activation of STAT3 and STAT5 and/or loss of STAT1 signalling is found in a large group of diverse tumours including Glioblastoma multiforme (GBM) (Brantley *et al*., 2008).

STAT proteins regulate many pathways important in oncogenesis including cell–cycle progression, apoptosis, tumour angiogenesis, and tumour–cell evasion of the immune system. The recent number of new cases of prostate cancer was estimated at 9,034,542 worldwide accounting for 7.1 % of all cancers (Globocan, 2008). Androgen plays an important role in the development and growth of prostate carcinoma (Kokontis and Liao, 1999). The transcriptional activity of AR is regulated by positive or negative transcriptional cofactors that include, PIAS1, PIAS3, and PIASy, which are expressed in the human prostate, and have distinct effects on AR–mediated gene activation in prostate cancer cells. While PIAS1 and PIAS3 enhance the transcriptional activity of AR, PIASy acts as a potent inhibitor of AR in prostate cancer cells (Mitchell *et al*., 1999). The N–terminal LXXLL signature motif of PIASy is essential for the trans–repression activity of PIASy (Mitchell *et al*., 1999).

Cystic fibrosis (CF), a disease caused by a defective gene encoding a protein called the cystic fibrosis trans–membrane conductance regulator (CFTR) (Welsh *et al*., 1995), is characterised by chronic lung infection resulting in inflammation and progressive lung damage. Increased levels of PIAS1 interrupt normal STAT1 cell signalling pathways, resulting in reduced IFN regulatory factor–1 (IRF–1) and nitric oxide synthase–2 (NOS2) expression in CF epithelial cells because NOS2 and IRF–1 expression are dependent on the activation of STAT1. This reduction in NOS2 expression and subsequent reduction in nitric oxide (NO) production has been postulated to play a role in the abnormal regulation of trans–epithelial sodium absorption observed in CF and its associated characteristic of susceptibility to bacterial infection (Meng *et al*., 1998; Kelly and Drumm, 1998). Phosphorylated STAT1 was subsequently found associated with PIAS1 in CF epithelial cells (Kelly and Elmer, 2000).

Lung cancer is the leading cause of death in adult men in Europe, the United States, and Japan. Lung cancer cells are relatively resistant to conventional chemotherapeutic drugs but undergo extensive apoptosis after treatment with pharmacological inhibitors of lipid kinase

(PI3–K/Akt) or Janus kinase (JAK/STAT) signalling pathways (Sordella *et al*., 2004). Treatment with LY294002, an inhibitor of phosphatidylinositol 3–kinase, retarded the growth of human lung cancer cells and rendered them more sensitive to chemotherapeutic agents (Ogata *et al*., 2006). Overexpression of PIAS3 not only significantly inhibited cell growth but also rendered cancer cells up to 12–fold more sensitive to the above drugs. However, the inhibition of JAK-STAT significantly suppressed cell growth but did not increase drug sensitivity (Ogata *et al*., 2006).

It has been reported that PIAS1 can regulate the transcriptional activity of the tumour suppressor p53, the p53–related protein p73, and the p53 regulator MDM2 (mouse double minute 2 homologue) (Melino *et al*., 2003; Urist and Prives, 2002). The tumour suppressor protein p53 and its two homologues p63 and p73 activates overlapping as well as specific sets of genes that have important roles in the regulation of the cell cycle and apoptosis (Urist and Prives, 2002; Melino *et al*., 2003). Under normal conditions, p53 is short lived and undergoes proteasome-mediated degradation. After exposure to various forms of stress and DNA damage, p53 is activated by post–transcriptional modifications which lead to p53 accumulation and downstream gene activation and ultimately cell–cycle arrest or apoptosis (Levine, 1997; Vousden and Prives, 2005). MDM2, which mediates negative–feedback control of p53, can repress the transcriptional activity of p53 and target it for degradation (Honda *et al*., 1997; Prives, 1998; Yang *et al*., 2004). It has been shown that PIAS1 and PIASx–β can promote the conjugation of SUMO to MDM2 in both *in vivo* and *in vitro* assays (Miyauchi *et al*., 2002).

1.6 THE PIAS PROTEIN FAMILY: STRUCTURE AND FUNCTION

The PIAS family members: PIAS1, PIAS2, PIAS3, PIAS4, PIASx (consisting of two splice variants: PIASxα and PIASxβ), and PIASy (Shuai *et al*., 1994) were initially identified in trying to understand the JAK–STAT signalling pathway (Chung *et al*., 1997; Liu *et al*., 1998). The existence of splice variants of PIAS isoforms adds to the complexity of PIAS protein family. They belong to the second class of SUMO E3 ligases and their homologues were found in non–vertebrate animal species, plants and yeast, including a protein encoded by the *Drosophila melanogaster* gene *Zimp* and the *Saccharomyces cerevisiae* proteins Siz1, Siz2/Nfi1, Mms21, and Zip3 (Johnson and Gupta, 2001; Takahashi *et al*.,2001; Zhao and Blobel, 2005; Cheng *et al*., 2006). A single *Drosophila* PIAS-encoding gene termed *Su (var) 2–10* was shown to be a gene required for normal chromosome function (Hari *et al.*, 2001).

This gene was described as *zimp*, has strong homology to the mammalian PIAS-encoding genes, and was named dPIAS/Zimp (Mohr and Boswell, 1999). Human Zimp10 and Zimp7 may be more distantly related mammalian homologues of PIAS proteins (Beliakoff and Sun, 2006). Nevertheless, evolutionary conservation suggests a common function which includes modulation of the activity of transcription factors either by protein–protein interactions or DNA binding or both. These two novel PIAS–like proteins contain the SP–RING/Miz domain (Sharma *et al*., 2003; Huang *et al*., 2005) resulting in the names Zimp7 and Zimp10 (Zinc finger containing, Miz–1, PIAS–like protein on chromosome 7 or 10). Other PIAS–like proteins such as the non–structural maintenance of chromosomes element 2 homolog (NSE2), zinc finger MIZ domain–containing protein 1 (ZMIZ1) or ZMIZ2, possess the PIAS RING (SP–RING) signature. KChAP (K^+ channel associated protein) was discovered by yeast two–hybrid screening using the rat brain cDNA library with full–length Kvβ1.2 subunit of the K+ channel as bait (Wible *et al*., 1998). The rat KChAP has high sequence identity with PIAS3 and was termed PIAS3β to distinguish it from mouse PIAS3 (Wible *et al*., 1998). KChAP is a potassium ion channel protein that acts as a chaperone to enhance expression of Kv2.1 protein and it belongs to the PIAS family (Chung *et al*., 1997). KChAP is homologous to PIAS3 and the two may be the products of alternative splicing of a single gene. KChAP contains an in–frame insertion of 35–amino acids at the N–terminal region (Kuryshev *et al*., 2000) which is lacking in human PIAS3. The Kv channel binding region, KChAP–M, is present in both KChAP and PIAS3, suggesting that KChAP and PIAS3 may interact with the potassium ion channels (Kuryshev *et al*., 2000).

1.6.1 PIAS Domains, Structure and Function

Members of the PIAS family share a high degree of sequence identity. Overall, five different domains have been identified in the PIAS protein family, namely DNA binding scaffold attachment factor–A/B/ACINUS/ PIAS (SAP) domain (Okubo *et al*., 2004), PINIT (proline, isoleucine, asparagine, isoleucine, threonine) domain (Duval *et al*., 2003), a putative really interesting new gene (RING)–type zinc–finger binding domain (Hanson *et al*., 1991; Hochstrasser, 2001), a SIM (SUMO–interacting motif), and the serine/theroine–rich (S/T) acidic domain (Minty *et al*., 2000) (**Figure 1.4**).

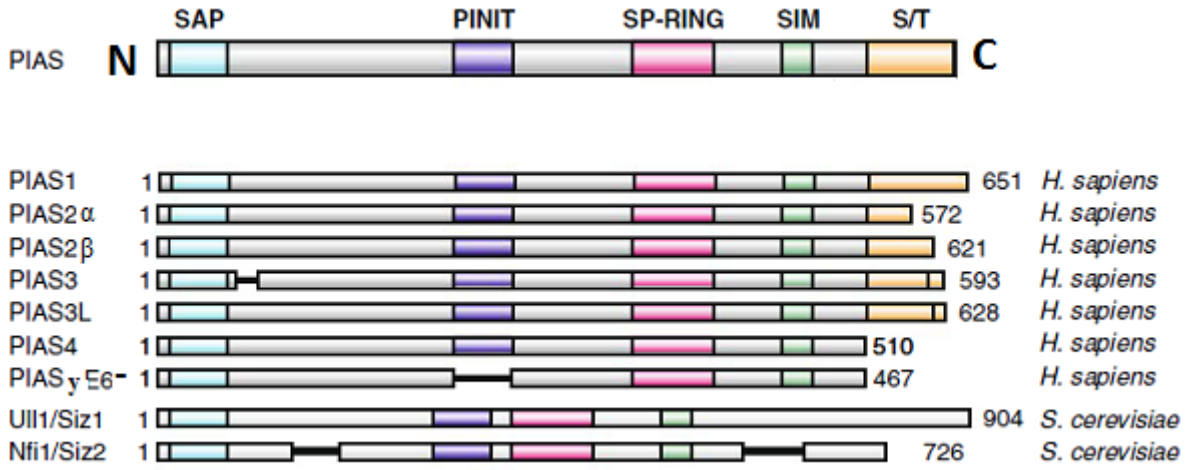

Figure 1. 4 Schematic representation of the domains of PIAS family members and their orthologs. Illustration of the domain structures of the different human PIAS proteins from *Homo sapiens* and their orthologous PIAS proteins from *S. cerevisiae* are also shown. PIAS3 is shorter that PIAS3L by 35 amino acids depicted after the SAP domain of PIAS3. Also, the PIASyE6⁻ lacks the PINIT motif. Numbers on the left and right depict the first and last amino acid in each PIAS protein. SIM and S/T. SAP domain is in turquoise colour, the PINIT domain is in purple colour and SP-RING domain is in pink colour, a SIM is in green colour and the C-terminal domain is in yellow colour. (adapted from Rytinki *et al*., 2009).

The S/T region is shorter in PIAS4 than in other PIAS proteins and the PINIT domain does not exist in the splice variant of PIASy (PIASyE6⁻). The splice variants PIASxα and PIASxβ have different lengths of the S/T region (Wu *et al*., 1997; Moilanen *et el*., 1999). Isoforms of PIAS3 and PIAS3L differ by a stretch of 35–amino acids between the SAP domain and the PINIT domain. The non–vertebrate PIAS orthologues essentially have the same motifs and domains as their mammalian PIAS counterparts.

1.6.2 The SAP domain

The sequence alignment of the four–helix SAP domain of PIAS1 and the SAP–domain of other PIAS family members (**Figure 1.5**) showed a high degree of sequence identity of 86, 77, and 57% for PIAS3, PIASxa, and PIASy respectively. Therefore, it is likely that all of the N–terminal domains of the PIAS family adopt a four–helix bundle conformation (Okubo *et al*., 2004).

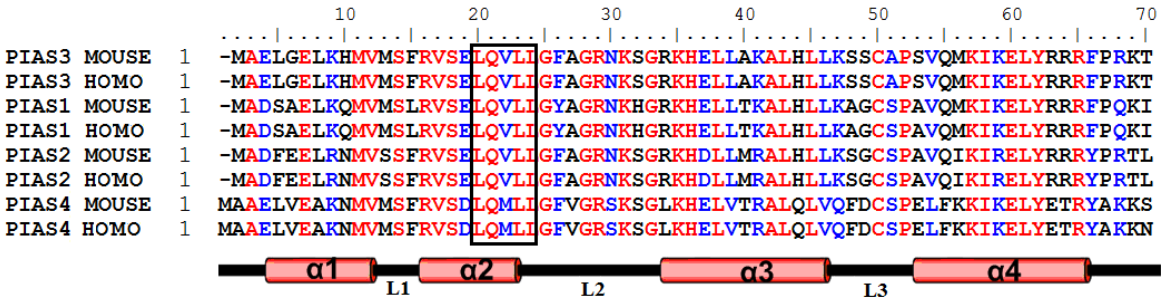

Figure 1.5 Sequence alignment of the SAP domains of the PIAS family members. Alignment of the C–terminal domain sequences of PIAS protein family using CLUSTAL W software (Larkin *et al.*, 2007). The numbers indicate the first and last residue in the particular sequence and the numbers at the top indicate the position of mouse PIAS3 residues. Shading represents the conservation of residues (scored according to Blosum62 score table). Identical residues are in red, and highly conserved residues are in blue, and of low residues conservation are in black. A black box indicates the SAP domain signature sequence LQVLL. Secondary structural elements for SAP domain are indicated below as helices and loops (L1, L2, and L3). The sequence accession numbers: PIAS3 Mouse (Mus musculus), AF034080; PIAS3 Homo (Homo sapiens), NP_006090.2; PIAS1 Mouse (Mus musculus), NP_062637.2; PIAS1 Homo (Homo sapiens), ABP49566.1; PIAS2 Homo (Homo sapiens), NP_004662.2; PIAS2 Mouse (Mus musculus), NP_032628.3; PIAS4 Mouse (Mus musculus), NP_067476.2; PIAS4 Homo (Homo sapiens), AAH10047.1.

The N–terminal SAP domain is found in many chromatin–associated proteins and is involved in sequence or structure–specific DNA binding (Aravind and Koonin, 2000). The N–terminal SAP domain of PIAS1 was found to be a four–helix bundle with crossover loops connecting the two pairs of helices (Okubo *et al.*, 2004) (**Figure 1.6**) and a putative DNA–binding motif involved in chromosomal organization (Romig *et al.*, 1992). The SAP domain binds to adenine/thymine rich (A/T) chromosomal regions known as scaffolding or matrix–attachment region (SAR/MAR) (Romig *et al.*, 1992) and performs a specific role in chromosomal organization that provide links between transcription repair, RNA processing and apoptotic chromatin degradation. The LXXLL signature sequence in the SAP domain has been shown to be important in the assembly of nuclear receptor co–activator complexes (Wu *et al.*, 1997). The SAP motif is predicted to have a helix bundle with two amphipathic helices that plays a crucial role in helix–helix interaction. The solution structure of the N–terminal domain (residues 1–65) of PIAS1 was determined to be a four–helix bundle (Okubo *et al.*, 2004) (**Figure 1.6**).

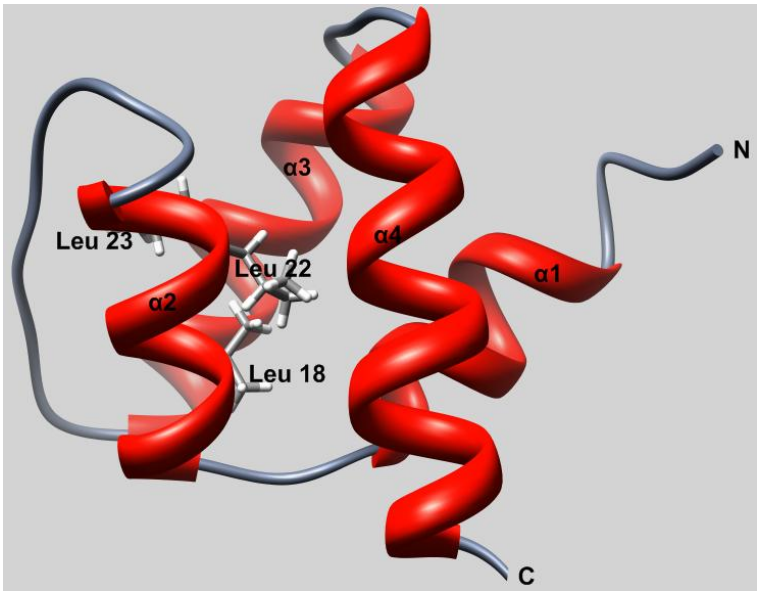

Figure 1.6 Three dimensional ribbon representation of the SAP domain of PIAS1. A four–helix bundle with a topology of an up–down–extended loop–down–up, a part of which the helix–extended loop–helix represents the SAP domain structure of PIAS. The indicated conserved leucine residues (Leu 18, Leu 22 and Leu 23) forms part of LXXLL motif which is the signature sequence of the SAP domain. The structure was determined by nuclear magnetic resonance spectroscopy (NMR) (PDB ID: 1v66) (adapted from Okubo *et al*., 2004).

The four–helix bundle adopts a topology of up–down–extended loop–down–up with two cross–over loops connecting the two pairs of helices as depicted in **Figure 1.6**. Although the domain exhibits strong DNA binding ability it does not resemble any of the known motifs of DNA–binding domains such as the helix–turn–helix or helix–loop–helix DNA binding motif and the leucine zipper or the zinc finger motif (Okubo *et al*., 2004).

1.6.3 The PINIT domain

The PINIT motif previously identified by Duval *et al*, 2003 is located in a highly conserved region of PIAS proteins **(Figure 1.7)**. The PINIT domain is present in all PIAS proteins except PIASy^{E6-}, which is a splice variant of PIASy that lacks exon 6 (Wong *et al*., 2004). The *holo*-PIAS3 protein structure has not been determined and the PINIT domain structure of the PIAS3 protein does not exist. However, the Siz1 X–ray crystallographic structure, determined by Yunus and Lima (2009), revealed that the PINIT domain of Siz 1 was formed by two antiparallel β–sheets connected by helix and loop (**Figure 1.8**). The PINIT motif was located at the hydrophobic core of the domain (Yunus and Lima, 2009) and mutation of the motif affected PIAS3 nuclear retention (Duval *et al*., 2003) and disruption of restricted nuclear localisation of PIAS3 (Wong *et al*., 2004).

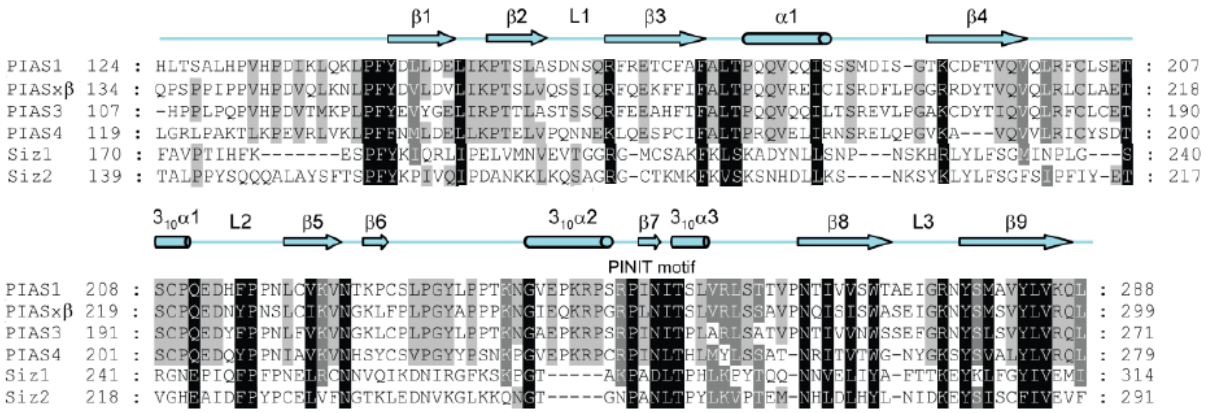

Figure 1.7 Sequence alignment of PINIT domains of the Siz and PIAS family of proteins. Amino acid alignment of sequences for PINIT domains from Siz and PIAS family members, *S. cerevisiae* (Siz1, Siz2), *Homo sapiens* (PIAS1, PIASxβ, PIAS3, and PIASy. Secondary structure elements of the PINIT domain are shown above the alignment; helices are indicated as bars and β–strands as arrows. The numbers indicate the first and last residue in the particular sequence. Identical residues in all sequences (black background) and highly conserved residues (grey) and low conserved residues (light grey). (adapted from Yunus and Lima, 2009).

The PINIT domain plays a role in the ligase function with some substrates that recognize Siz1–dependent substrates (Reindle *et al*., 2006). Experimental evidence revealed interactions between PCNA (proliferating cell nuclear antigen) and GST–PINIT (Yunus and Lima, 2009). The PINIT motif has been shown to be essential for the nuclear retention (Duval *et al*., 2003) and a short specific peptide sequence (V82-T104) within the PINIT domain binds to both STAT3 and MITF (Sonnenblick *et al*., 2004; Levy *et al*.,2006). This short peptide of 23 amino acids is capable of inducing apoptosis in both RBL–2H3 and mouse melanoma cells by inhibiting the transcriptional activity of both MITF and STAT3 (Yagil *et al*., 2009).

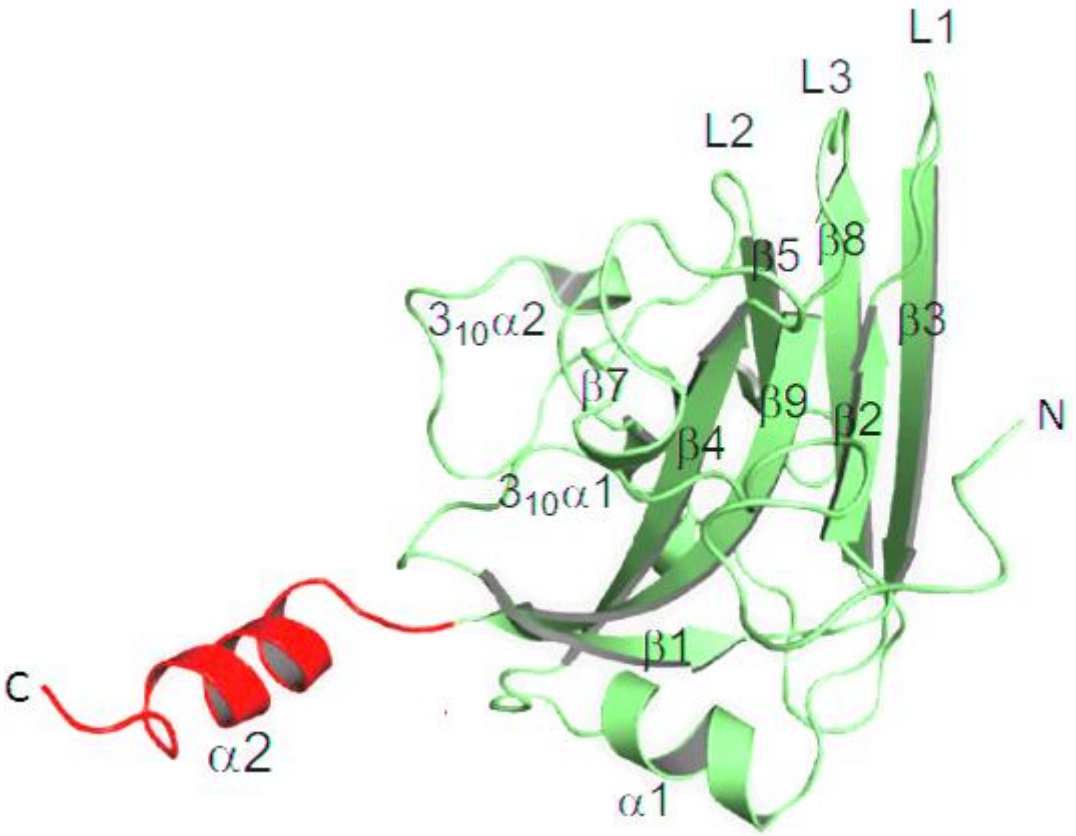

Figure 1.8 Ribbon representation of the three–dimensional structure of the PINIT domain of Siz 1 protein (PDB ID: 3i2D). The PINIT domain of Siz1 protein showing by two antiparallel β–sheets connected by loops joining strands at one end of the molecule and connected by helix and loop. The N–terminal PINIT domain (amino acids 172–315; in cyan colour) is formed by two antiparallel β–sheets; one includes β1, β2, β4, and β9, and the other includes β3, β5 and β8. The β–sheets are connected by protruding loops (L1, L2, and L3) that join strands β2–3, β4–5, and β8–β9 at one end of the molecule, while β3–4 and β5–8 are connected by a helix α1 and a loop, respectively, on the opposite surface. The C-terminus helix (α2) (red) connects the PINIT domain to the SP-RING domain. (adapted from Yunus and Lima, 2009).

The PINIT domain has been suggested to interact with the STAT3 specifically on the DNA binding domain of STAT3 (Chung *et al*., 1997). Interaction of STAT3 with various fragment lengths of PIAS3 has been shown. PIAS3 (1–123) fragment did not interact with STAT3 (Yamashina *et al*., 2006) despite carrying part of the PINIT domain. However, PIAS3 (224–584) formed a complex with STAT3 and it encompasses the C–terminus of the PINIT domain. The determination of the PINIT domain structure forms the critical basis of understanding the molecular mechanism of PIAS3 protein.

1.6.4 The SP–RING domain

The classical RING domain contains eight zinc–binding cysteine or histidine residues (Weissman, 2001). The residues can coordinate two zinc ions, creating a globular domain that can mediate protein–protein interactions. RING proteins function as ubiquitin E3 ligases bind substrates with their RING and directly interact with the E2 dependent–enzymes to govern the specificity of ubiquitylation. They facilitate the transfer of ubiquitin from the E2 to the substrate without formation of covalent intermediates (Weissman, 2001). The SP–RING motif lacks two zinc–coordinating cysteines found in the classical RING domain but has a similar fold as in the RING finger (Joazeiro *et al*., 2000). The RING–type zinc–finger domain is defined to be any small, functional, independently folded domain that requires coordination of one or more zinc ions to stabilise its structure (Laity *et al*., 2001). The zinc is complexed to four conserved cysteine residues and/or histidines stabilizing a finger–like structure that can coordinate two zinc atoms and mediate multi–protein complex formation and protein–protein interactions (Weissman, 2001). The RING–finger domain is mostly involved in DNA binding or in protein–protein interaction. The consensus sequence of a single finger is Cys–$X_{2–4}$–Cys–X_3–Phe–X_5–Leu–X_2–His–X3–His. The spacing between potential zinc co–ordinating residues and the amino acid composition of the mammalian PIAS RING–like structure differ substantially from the C3HC4 (RING–HC)–type RING finger found commonly in ubiquitin E3 ligase (Weissman, 2001).

An alignment of the SP–RING motifs from mammalian PIAS proteins and the RING finger region from PIAS yeast homologues and the c–cbl ubiquitin ligase showed conserved cysteine and histidine residues that formed the RING finger that bind to the zinc ion (**Figure 1.9**). When compared with the classical RING finger, the SP–RING motif lacks the second and sixth cysteine residues that are part of the first and third pair of cysteine/histidine residues in the RING motif. In a cross–brace arrangement the first and third pair of cysteine/histidine residues forms one zinc binding site, while the second and fourth pairs form the second binding site in the RING motif (Aravind and Koonin, 2000; Ohi *et al*., 2003).

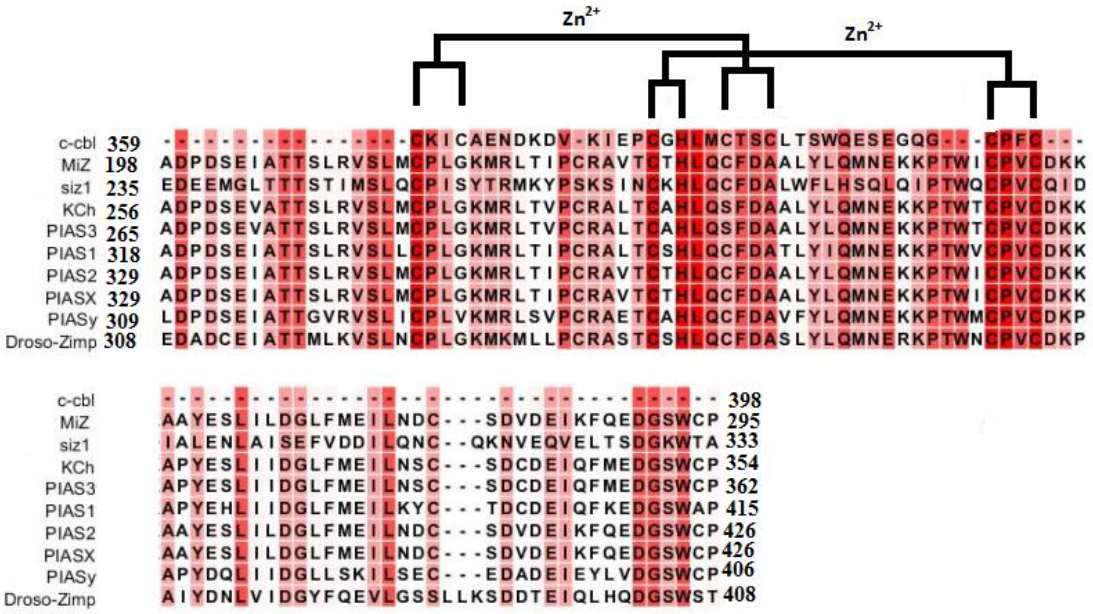

Figure 1.9 Sequence alignment of the SP–RING domains of the PIAS family and the orthologs. Sequence alignment of *S. cerevisiae* Siz proteins, human PIAS family, and *Drosophila* zimp. Identical residues have red background, highly conserved residues have pink background and residues of low conservation have white background. Cysteine/cysteine and cysteine/histidine zinc coordinating residues in 'cross–brass' arrangement are shown above the alignment. The numbers indicate the first and last residue in the particular sequence. Sequence accession numbers: *c–cbl* ubiquitin ligase (gi: 50315), Miz–zinc finger (MiZ1–2773148), Siz1 (gi: 258588585), KChAP (3127051), PIAS3 (2689028), PIAS1 (gi: 31543478), PIAS2 (gi: 56404605), PIASX (gi: 56699458), PIASy (gi: 45219874). (adapted from Schmidt and Müller. 2003).

The SP–RING motif of PIAS family is structurally conserved and is suggested to have a similar fold as in the RING finger, although it lacks two zinc co–ordinating cysteines found in the classical RING domain. The zinc ion is crucial for the stability of this domain type in the absence of the metal ion the domain unfolds as it is too small to have a hydrophobic core. Many zinc finger proteins are transcription factors that function by recognizing a specific DNA sequence (Laity *et al*., 2001). It is also required for sumoylation, which plays a role in targeting proteins to specific subcellular locations, stabilizing target proteins and modulating the transcriptional activity of substrate proteins. However, several structurally different types of proteins can exhibit E3–type ligase activity in sumoylation reactions (Kerscher *et al*., 2006). Also, proteins such as the non–structural maintenance of chromosomes element 2 homolog (NSE2) and the zinc finger MIZ domain–containing protein 1 (ZMIZ1) or ZMIZ2, that possess the PIAS RING (SP–RING) signature or its extended version, are considered only very distantly related to the PIAS proteins (Beliakoff and Sun, 2006).

1.6.5 The acidic domain and SUMO interacting motif (SIM)

The acidic domain at PIAS C–terminal region contains Serine/Theronine (S/T) residues. This region is the most diverse region in the PIAS family. The acidic domain was shown to be involved in the binding of PIAS3 to the nuclear receptor coactivator (TIF2) (Jimenez–Lara *et al*., 2002) and to bind to STAT3 (Chung *et al*., 1997, Borghouts *et al*., 2010). The C–terminal domain of PIAS proteins harbours a putative SUMO1 interacting motif (SIM) (Minty *et al*., 2000). The SIM is defined as a sequence of wxww or wwxw (where w is often V or I) and usually next to a C–terminal cluster of negatively charged residues. The SIM motif interacts non–covalently with SUMO proteins and has less consensus sequence (Minty *et al*., 2000). The SUMO1 interaction motif and S/T rich region are found in all PIAS proteins, except for PIAS4 (Shuai, 2006).

1.7 PIAS PROTEIN INTERACTION WITH TRANSCRIPTIONAL FACTORS AND OTHER PROTEINS

Different regions of PIAS proteins are involved in protein–protein interactions and regulate the transcription activation activities of transcription factors (**Figure 1.10**). More than 60 proteins, most of them transcription factors, have been suggested to interact with members of the PIAS family (Shuai and Liu, 2005). Regions of PIAS proteins that are involved in protein–protein interactions have been identified in many studies. Interestingly, various regions of PIAS proteins seem to be involved in different protein–protein interactions (**Figure 1.10**). For example, the N–terminal region of PIAS1 can interact with the p65 subunit of nuclear factor kappa B (NF–κB), whereas the C–terminal region of PIAS1 can bind STAT1 (Liao *et al*., 2000). These findings were results of targeted mutational analysis to dissect the functional role of PIAS proteins in various signalling events (Shuai and Liu, 2005).

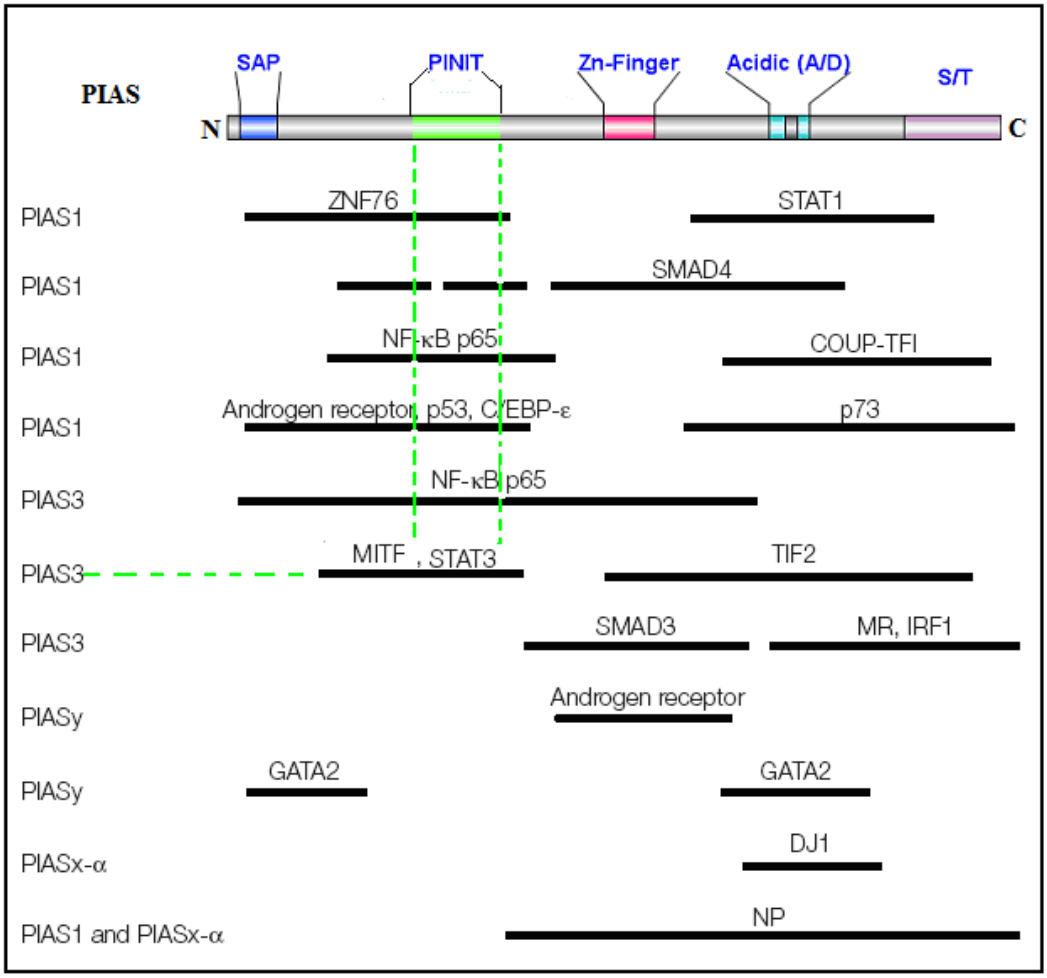

Figure 1.10 PIAS proteins interaction with transcriptional factors and other proteins. The top panel is the schematic illustration of PIAS protein domains and regions that are involved in interaction with transcriptional factors and other protein below as black horizontal lines. The PIAS family member that interacts with the transcriptional factor or other protein is shown at the left side. The broken line illustrates the PINIT domain of PIAS3 that associates with STAT3. C/EBP–ε, CCAAT/enhancer–binding protein–ε; COUP–TFI, chicken ovalbumin upstream promoter transcription factor 1; GATA2, GATA–binding protein 2; IRF1, interferon–regulatory factor 1; MITF, microphthalmia–associated transcription factor; MR, mineralocorticoid receptor; NF–κB, nuclear factor–κB; NP, nucleocapsid protein, from hantavirus; p53, tumour suppressor protein p53; p73, tumour protein p73; SMAD, SMA (small body size) and MAD (mothers against decapentaplegic)–related protein; DJ1, a Parkinson's disease protein; TIF2, transcriptional intermediary factor 2; ZNF76, zinc–finger protein 76. (adapted from Shuai and Liu, 2005)

1.8 PIAS3–STAT3 DOMAINS INVOLVED IN INTERACTION

PIAS3 is a multifunctional protein domain comprised of distinct functional domains that are capable of interacting with various proteins (**Figure 1.10**). The interaction of PIAS3 with STAT3 was first shown in IL–6 treated M1 cells by immunoprecipitation and no evidence of PIAS3–STAT3 association in untreated cells (Chung *et al.*, 1997). Furthermore, treatment of

the HepG2 cells with another STAT3 activation cytokine, ciliary neurotrophic factor (CNTF) and oncostatin M (OM), showed the association of PIAS3 and STAT3 but not in untreated cells (Chung *et al*., 1997). These data suggested that PIAS3 associates with STAT3 following STAT3 stimulation and activation. However, stimulation of STAT3 results in the formation of the dimer as described in section 1.4. Further analysis of the form of STAT3 that associates with PIAS3 was achieved by probing for phosphorylated STAT3 in the immunoprecipitated complex. Interestingly Chung *et al*. (1997) showed that the PIAS3 completely inhibited the DNA–binding activity of STAT3–STAT3 homodimer in HepG2 IL–6 treated cells. These findings confirmed the similar inhibitory effect of PIAS3 on the DNA–binding activity of STAT3 which was observed in nuclear extracts prepared from IL–6–treated M1 and MCF7 cells. These data suggest the possibility that PIAS3 interferes with the DNA binding domain of STAT3.

Yamamoto *et al*. (2003) showed by immunoprecipitation that $STAT3_{320–493}$ and $STAT3_{494–750}$ regions interact with PIAS3 and not the N–terminal domain ($STAT3_{1–137}$) or the CCD ($STAT3_{138–319}$). These results supported the findings by Chung *et al*. (1997) that PIAS3 interacts with DNA binding domain of STAT3 (**Figure 1.11**). Levy *et al*. (2006) showed that the $PIAS3_{82–132}$ region associates with STAT3 and this region is encompassed in the PINIT domain (**Figure 1.11**). However, it is important to mention that the different PIAS3 fragments used were *in vitro* translated [^{35}S] methionine labeled (Levy *et al*., 2006). The effect of folding and whether the different fragments were synthesised *in vitro* with similar efficiency was not validated. Furthermore, interpretation of the data from the mutational analysis performed to determine the region critical for PIAS3–STAT3 interaction was problematic. PIAS3–Y94P was found to be disruptive with respect to STAT3 binding (Levy *et al*., 2006). However, the loss of binding was probably as a result of structural changes and thus may not necessarily be due to the amino acid substitution. Protein-protein interaction analysis using PIAS3 domains which are characterised for their folding need to be conducted (e.g. using surface plasmon resonance spectroscopy [SPR]). In addition, mutations which maintain the structural integrity were necessary to determine critical residues in the PIAS3–STAT3 interaction. In an attempt to find the minimum fragment that can perform the same function as the full length PIAS3, Yagil *et al*. (2009) showed that the $PIAS_{82–104}$ fragment (**Figure 1.11**) was sufficient to induce apoptosis by arresting STAT3 transcriptional activity and this work was conducted *in vivo* with no further mutational analysis.

Dabir *et al*. (2009) focused on STAT3 residue Y705 with experimental results that suggested the PIAS3–STAT3 interaction was based on the phosphorylation of STAT3 Y705 and the disruption of PIAS3-STAT3 was based on dephosphorylation of STAT3 Y705 by PIAS3 in the nucleus.

Recently Borghouts *et al*. (2010) focused on the C–terminal region of the PIAS3 by performing yeast two–hybrid and coimmunoprecipitation experiments using recombinant peptides of both STAT3 and PIAS3. Yeast two–hybrid analysis with the N–terminal $PIAS3_{1–319}$ showed no interaction with STAT3. However, yeast two–hybrid showed that $PIAS3_{400–523}$ and $PIAS3_{400–543}$ fragments interacted with STAT3 (Borghouts *et al*., 2010). It should be noted that the C–terminal region of PIAS1 freely interacted with phosphorylated or unphosphorylated STAT1. However, the presence of the N–terminal region prevented PIAS1–STAT1 interaction (Liao *et al*., 2000). Nevertheless, Borghouts *et al*. (2010) showed that the interaction of these PIAS3 fragments were specifically with the CCD of STAT3.

Presently, the understanding of the PIAS3–STAT3 association is not clear. Various regions of PIAS3 have been suggested to be involved in the PIAS3–STAT3 association. Also, multiple regions of STAT3 have been suggested to associate with PIAS3. The main problem with the understanding of the PIAS3–STAT3 interaction is the limited structural data available for the PIAS family members. Hence, in this study using structural bioinformatics, mapping of the regions and residues potentially important for the PIAS3–STAT3 interaction was based on the conservation of residues and their orientation in the tertiary structure. Furthermore, using biophysical and biochemical approaches not previously employed, the PIAS3–STAT3 association was further investigated in this study.

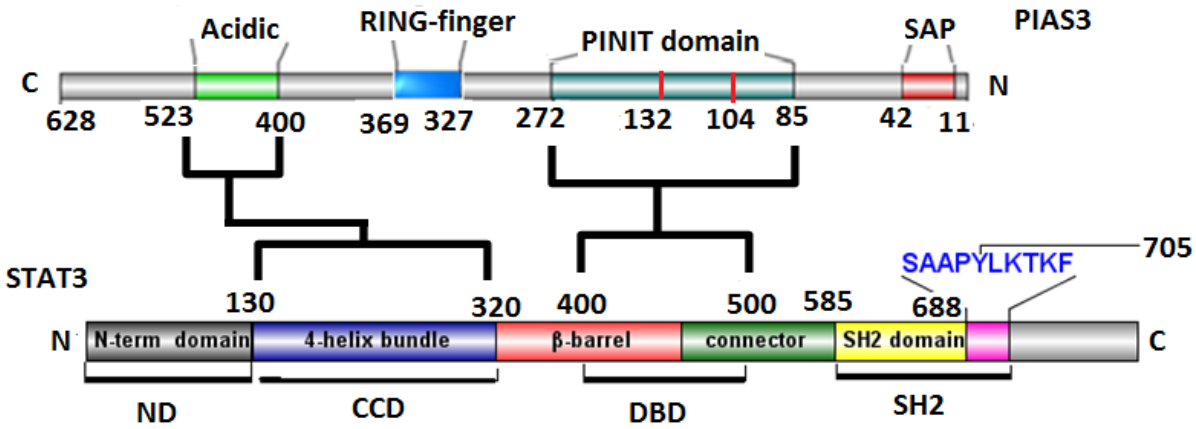

Figure 1.11 Schematic representation of STAT3 and PIAS3 binding regions. Various PINIT domain regions that have been suggested to associate with STAT3 DNA binding domain. The PINIT domain (residues 82–272) of PIAS3 associates with STAT3 DNA binding domain (residues 400–500). The short region of PINIT domain (residues 82–132) associates with STAT3 (Levy *et al*., 2006). The shortest region of the PINIT domain (residues 82–104) has been suggested to associate with STAT3 (Yagil *et al*., 2009). PIAS3 (residue 400–523) has been suggested to associate with the CCD of STAT3 (Borghouts *et al*., 2010).

1.9 PROBLEM STATEMENT

A number of regions of PIAS3 have been suggested to interact with STAT3 using *in vivo* and *in vitro* experiments. Various peptides of PIAS3 have been characterised as the determinants of PIAS3–STAT3–interaction. The multidomain structure of PIAS3 makes it a major challenge to identify a specific domain when studied in isolation that confers the same activity as full length PIAS3. These problems are exacerbated by difficulties around production of protein quantity and quality suitable for structural analysis, and furthermore, by the lack of *holo*-PIAS3 structure or the structure of any PIAS family member that can be used as template.

1.10 HYPOTHESIS

PIAS3 individual domains alone are necessary but not sufficient for its functional interaction with STAT3.

1.11 BROAD OBJECTIVES

A. Bioinformatics analysis of PIAS3 and its domains.

B. Analysis of the cellular localisation of PIAS3 and its domains.

C. Development of a suitable expression system for the heterologous expression and purification of PIAS3 or its domains.

D. The *in vitro* analysis of the PIAS33–STAT3 interaction.

1.12 SPECIFIC OBJECTIVES

A1. Bioinformatics analysis of PIAS3, the PINIT domain and the acidic domain. Homology modeling of the PINIT domain and its mutants and validation of the models.

B1. Investigation of the localisation of PIAS3, the PINIT domain, the acidic domain, mutant PIAS3, and their co–localisation with STAT3.

C1. Plasmid construction for the heterologous expression and purification of the PINIT domain, characterisation of the size and folded state of the PINIT domain and its mutant derivatives.

D1. PINIT–STAT3 binding analysis using SPR, including the generation of PINIT domain mutants to assess effects of the mutations on PINIT–STAT3 association.

CHAPTER 2

BIOINFORMATICS ANALYSIS OF THE PIAS3, PINIT DOMAIN AND THE ACIDIC DOMAIN OF PIAS3 PROTEIN.

In human there are four different PIAS isoforms encoded by separate genes that modulate the activities of transcriptional factors. The high degree of sequence identity among PIAS proteins predicts similar functions. The conserved multiple domains have distinct functions and have been suggested to modulate the transcriptional activity of the STAT protein family. The broad objective of this study was to analyse the conservation pattern of the PIAS protein family and thereby identify residues conserved in individual domains capable of addressing PIAS3 biological function. Secondly, the prediction of secondary structure elements of the PINIT domain allowed for the generation of the PINIT domain homology model to evaluate the structural determinants in PIAS3–STAT3 interaction. Furthermore, the study aimed to predict and mutate critical residues that potentially determine PIAS3 function.

2.1 INTRODUCTION

The molecular mechanism of the PIAS3 interaction with STAT3 can be understood better by having structural insight of the protein and its domains. Structures can be determined using X–ray crystallography and nuclear magnetic resonance spectroscopy (NMR), but these techniques are dependent on factors such as production of sufficient protein and the size of the protein. NMR spectroscopy resolution has a size limitation of 30 kDa. To date the PIAS3 structure does not exist in available structural databases. Of the PIAS family members the SAP domain of PIAS1 has been solved by NMR (Okubo *et al*., 2004), and the PINIT domain of Siz1, a yeast ortholog of PIAS family, was recently elucidated using X–ray crystallography (Yunus and Lima, 2009). Advances in the computational systems have allowed for the development of bioinformatics approaches to calculate *in silico* homology models from available templates using known protein sequences of the protein of interest. This is referred to as the template based approach to determine structure based on sequence and fold similarity.

Comparative or homology modeling approaches are based on the conservation of protein structures and sequences. Proteins with high sequence identity are likely to share similar structure (Sancheze and Sali, 1997). Homology modeling depends heavily on the accuracy of alignment, which allows detection of conserved domains. The confidence of the model increases with increase in similarity between the target and template sequences.

Three–dimensional structure generally provides more information about a protein function than the sequence because interactions of a protein with other molecules are determined by amino acids that are close in space but are frequently distant in sequence. Homology modeling is a multistep process that is described by four iterative steps: template recognition, target–template sequence alignment, model generation and model validation (**Figure 2.1**).

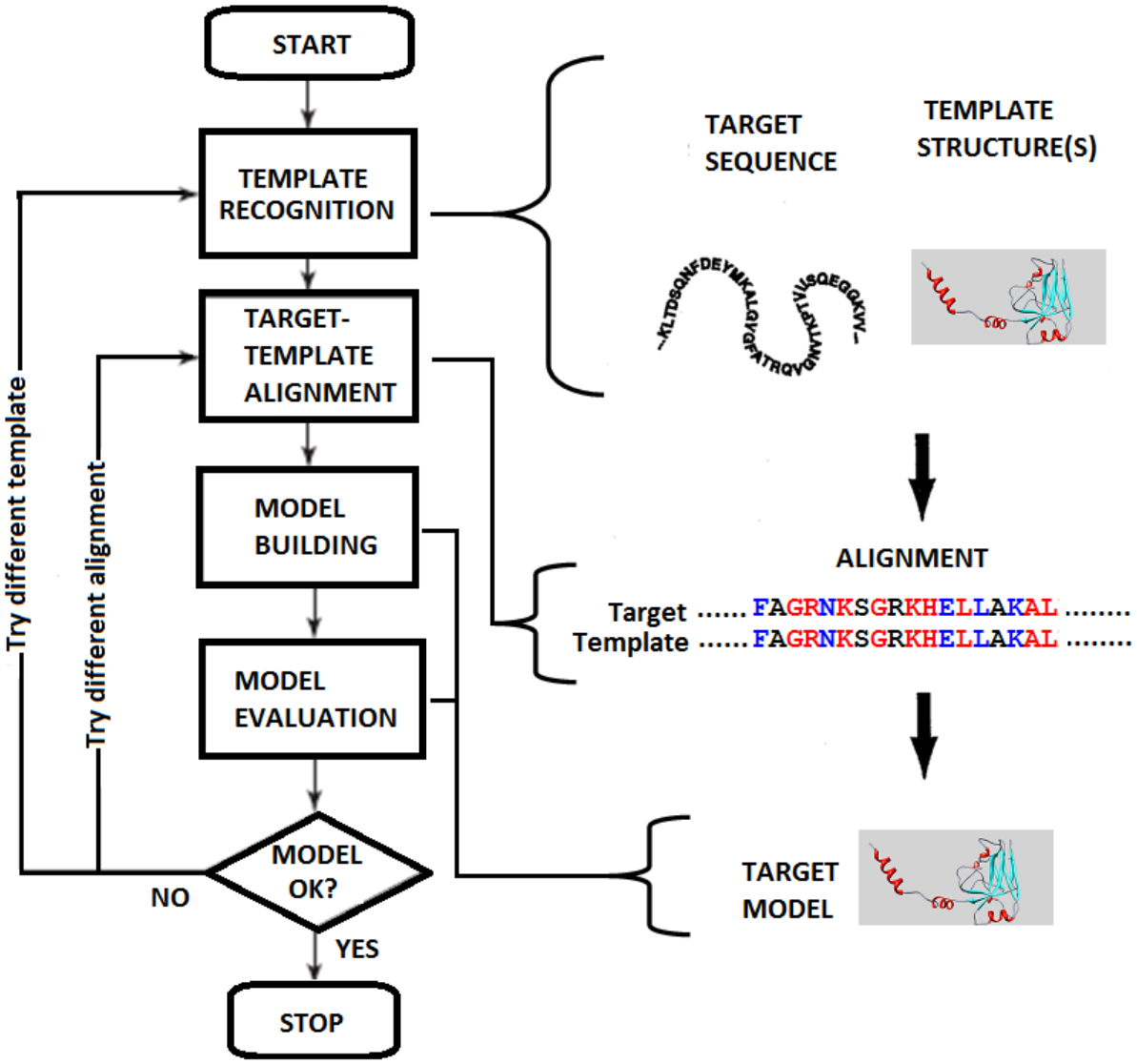

Figure 2.1 Flowchart illustrating the steps in comparative protein structure modeling. Steps in construction of a homology model as implemented in MODELLER 9v3 (adapted from Sanchez and Sali, 1997).

Step 1: Template recognition

Template searches are performed using known protein sequences and structure databases; this is done by comparing the query sequence with the sequence of each of the known structures in the available database. Algorithms employed include Basic Local Alignment Search Tools (BLAST) or sequence profile methods using position specific iterative (PSI)–BLAST (Altschul and Koonin, 1998) which are based on pairwise comparative methods. Sensitive search with profile methods such as Hidden Markov Models (HMM) (Söding *et al.*, 2005) and HHpred (Söding *et al.*, 2005) available as web based servers

(http://toolkit.tuebingen.mpg.de) may be required and also it is necessary to directly evaluate the compatibility between the target sequence and each structure in the data base using "Threading" (Jones *et al.,* 1992). From the list of templates, the most suitable template with high percentage sequence identity to the query is selected for modeling.

Step 2: Target–template sequence alignment correction

Specialised methods are used to align the template sequence with the target sequence such as CLUSTALW (Higgins and Sharp, 1998; Thompson *et al.*, 1994). The use of the multiple sequence alignment to derive position specific scoring matrices (*profiles*) (Taylor, 1986; Dodge *et al.*, 1998) can be used during manual intervention to correct gaps in the alignment and also the alignment can be improved by including structural information from the template, for example gaps should be avoided in secondary structure elements and in buried hydrophobic regions (Sancheze and Sali, 2000). Manual alignment by visual inspection of the template is important, as mis–alignment of a single amino acid may result in spatial errors of approximately 4 Å in the calculated model.

Step 3: Model Building

MODELLER is an automated computer program that models a protein by satisfaction of spatial restraints; it uses either distance geometry or optimization techniques to satisfy spatial restraints obtained from alignment of the target sequence with the template structure (Sali and Blundell, 1993). MODELLER extracts atom–atom distance and dihedral angle restraints on the target from the template structure(s) and combines them with general rules of protein structure such as bond length and angle preference (Sancheze and Sali, 2000).

Step 4: Model evaluation

Model errors mainly depend on the percentage sequence identity between template and the target. If the sequence identity is below 30 %, the alignment will be the key to the accuracy of the homology modeling (Chothia and Lesk, 1986; Sippl, 1993). Also, errors from the template affect the accuracy of the homology. Model evaluation can be performed using programs such as PROCHECK (Laskowski *et al.,* 1993) and WHATCHECK (Hooft *et al.,* 1996) to assess the model stereochemistry, bonds, bond angles, dihedral angles, and nonbonded atom–atom distances. To predict whether or not a template is correct, it is necessary to compare the calculated modeller Z–score (Sippl, 1993) for the model and the

template structure(s). The Z–score of a protein is defined as the energy separation between the native fold and the average of an ensemble of misfolds in the units of the standard deviation of the ensemble (Zhang and Skolnick, 1998) and should be comparable to that of the template. External evaluation is the prediction of unreliable regions in the model using programs like MetaMQAPII (Pawlowski *et al.*, 2008) and Verify3D (Eisenberg *et al.*, 1997) (available as web based servers **Appendix F1**).

The model can also be improved by iterations consisting of template selection, alignment, and model building, guided by model assessment. The iterations may be repeated until no improvement in the model is detected (Guenther *et al.*, 1997; Sanchez and Sali, 1997).

Loops vary among homologs while the core regions are conserved and accurately aligned. Loop refinement is achieved by optimizing a scoring function (Spassov *et al*., 2008) using methods exploiting different protein representations, objective functions, and optimization algorithms thereby improving the quality of the predicted model.

Based on the systems described above, the objective of the current study was to use bioinformatics tools to analyse available PIAS protein sequences and identify conserved regions and motifs to map specific residues critical for PIAS3 function. The homology model of the PINIT domain of PIAS3 was generated using MODELLER–9v3. Conserved residues deemed potentially important to the structural and functional integrity of PIAS3 were identified. Interesting domains capable of addressing biological questions may be investigated based on knowledge of protein structure, sequence conservation pattern and prediction of the secondary structures elements. The outcome of the study presented here guided the *in vitro* and *in vivo* work in the subsequent chapters (Chapter 3–5)

2.2 PROCEDURES

2.2.1 Sequence retrieval and alignment

Searches for non–redundant sequences in databases (nr) were performed using default parameters at the NCBI using PSI–BLAST (Altschul and Koonin, 1998) (www.ncbi.nlm.nih.gov/blast) and HHpred (Söding *et al.*, 2005) (http://toolkit.tuebingen.mpg.de/hhpred). The available sequence of *mouse* PIAS3 (accession number: AF034080) was used as query. Profile based alignments of sequences from the PIAS family were performed using ClustalW (Higgins, *et al.* 1994), set to Blosum62 matrix; all other parameters were set to default (Thompson *et al.*, 1994). The final multiple sequence alignment was manually refined to ensure that there were no unwarranted gaps introduced

within potential structural elements like α–helices and β–strands. Accession numbers for the protein sequences used in the multiple sequence alignments are: PIAS3 Mouse (*Mus musculus*), AF034080; PIAS3 Homo (*Homo sapiens*), NP_006090.2; PIAS1 Mouse (*Mus musculus*), NP_062637.2; PIAS1 Homo (*Homo sapiens*), ABP49566.1; PIAS1 Gal (*Gallus gallus*), NP_001025797.1; PIAS2 Homo (*Homo sapiens*), NP_004662.2; PIAS2 Mouse (*Mus musculus*), NP_032628.3; PIASx Gal (*Gallus gallus*), NP001025797.1; PIASx Homo (*Homo sapiens*), AAC36705.1; PIASy Homo (*Homo sapiens*), AAC36703.1; PIAS4 Rat (*Rattus norvegicus*), NP_001094227.1; PIAS4 Homo (*Homo sapiens*), AAH10047.1; Siz1 (*Saccharomyces cerevisiae*), (PDB ID: 3I2D).

2.2.2 Secondary structure prediction and assessment

Secondary structure prediction of the PINIT domain of PIAS3 protein was achieved by application of the following software packages: PSIPRED (Jones, 1999); JNET (Cuff and Burton, 1999); Prof (Rost) (Rost, 2001); and Prof (Ouali) (Ouali and King, 2000). Matches between the query sequence and known protein structures identified by alignment of the predicted secondary structures was carried out using 3DPSSM (Kelly *et al*., 2000) and HHpred (Söding *et al*., 2005). Predicted fold recognitions produced were compared and evaluated; template sequences remotely related to target sequences were eliminated.

2.2.3 PINIT domain model building

The PINIT domain region, consisting of residues 85 to 272 of the PIAS3 protein (accession number: AF034080), was submitted to the HHpred web server to identify templates (Söding *et al.,* 2005). The server identified only one template, Siz/PIAS SUMO E3 ligase Siz1 (PDB ID: 3I2D) (Yunus and Lima, 2009) with 17 % sequence identity. A target–template sequence alignment was automatically built with the automated HHpred alignment function. Since sequence identity was low, the target–template alignment was compared with multiple sequence alignments produced with ClustalW and manually corrected. The secondary structure was matched between the predicted secondary structure of the target protein and the calculated secondary structure of the template. The alignment file was submitted to MODELLER–9v3 and 100 models of the PINIT domain ($PIAS_{85–272}$) were built using python scripts run in MODELLER-9V3 (Appendix F2). DOPE Z (normalized Discrete Optimised Protein Energy) scores were calculated for all models (Shen and Sali, 2006).

2.2.4 Model evaluation

The DOPE Z–scores of the 100 generated models were calculated using an automated python script run in MODELLER–9v3 (**Appendix F2**). The model with the best DOPE Z–score was subjected to loop refinement using automated python scripts run in MODELLER–9v3. During refinement, 100 models were generated and their DOPE Z–score calculated and best model was used for further refinement of other loops. After loop refinement and picking the model with the best DOPE Z–score, the model quality assessment was further performed by MetaMQAPII (Pawlowski *et al.,* 2008) which assesses the model accuracy according to temperature scheme. Furthermore, the PINIT model was evaluated by quality assessment using Verify3D (Eisenberg *et al.,* 1997). Visual evaluations were performed by inspecting the models with UCSF Chimera (Goddard, *et al.,* 2005) and PyMOL (Delano, 2004) molecular visualisation programs.

2.2.5 Homology modeling of the mutant PINIT domain models

Mutant models PINIT–L97A, PINIT–R99N and PINIT–R99Q were calculated and validated as described above from the final validated PINIT domain model as template structure.

2.3 RESULTS

2.3.1 PIAS protein family sequence alignment and analysis

Analysis of the sequence alignment of the PIAS protein family revealed conserved regions and domains of PIAS3. Analysis of the sequence alignment (**Figure 2.2**) identified four conserved regions corresponding to the four known conserved domains of PIAS family: N–terminal SAP domain (position 1 to 70) with the "**LQVLL"** signature sequence; the PINIT domain with its signature **"PINIT"** motif (position 149 to 336, **Figure 2.2**); the RING–type zinc–finger binding domain showing residues that forms a zinc–finger binding (position 337 to 420, **Figure 2.2**); and the C–terminal acidic domain (position 421 to 556, **Figure 2.2**) rich with acidic residues and the signature SIM binding motif or SUMO–1 binding motif. The PIAS protein family sequence alignment validated the previous findings of the domains described in the PIAS protein family (Hanson *et al.,* 1991; Romig *et al.,* 1992; Minty *et al* 2000; Hochstrasser, 2001; Duval *et al.,* 2003).

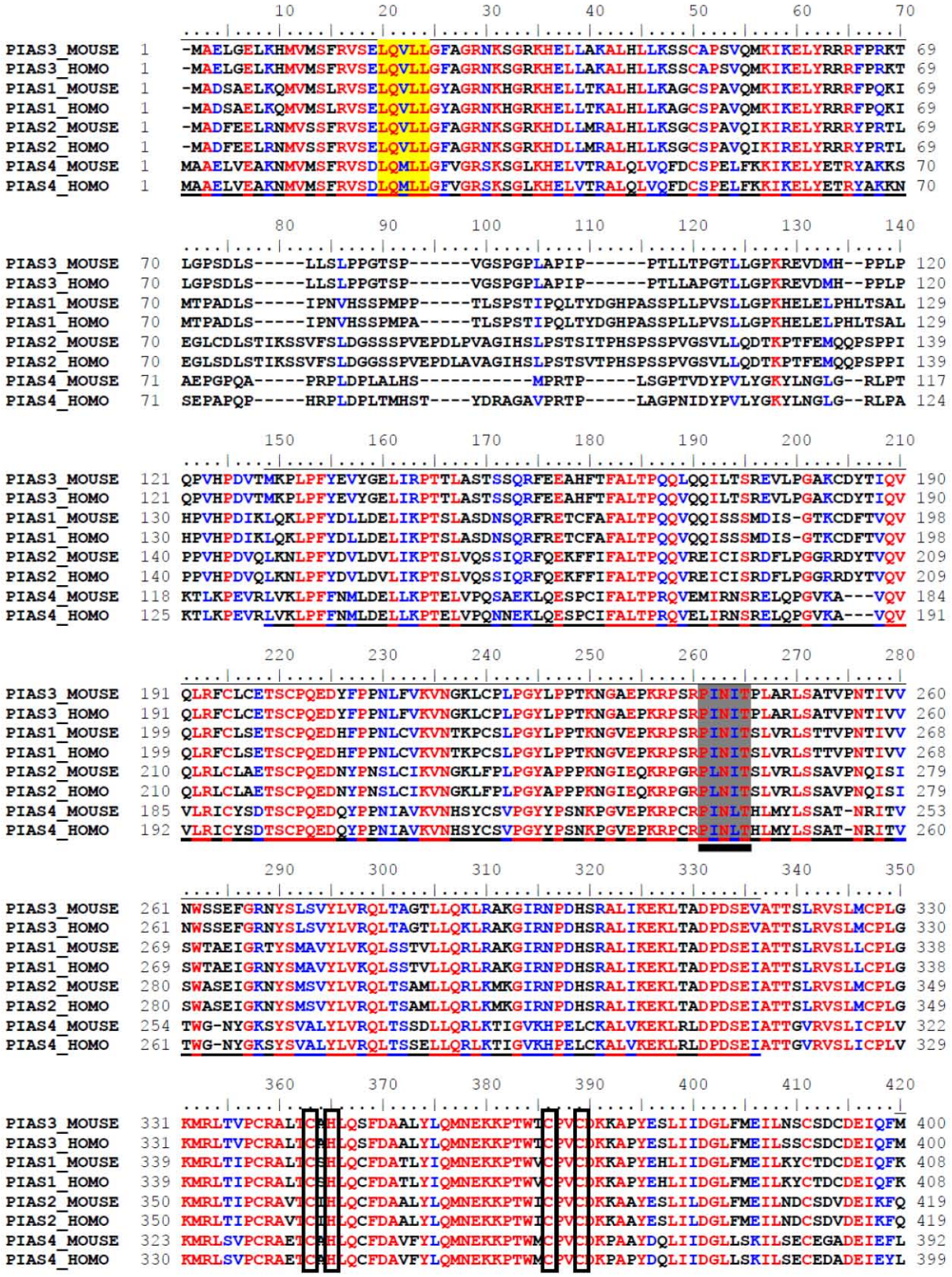

Figure 2.2 Sequence alignment and analysis of PIAS family members from different organism (continued on next page).

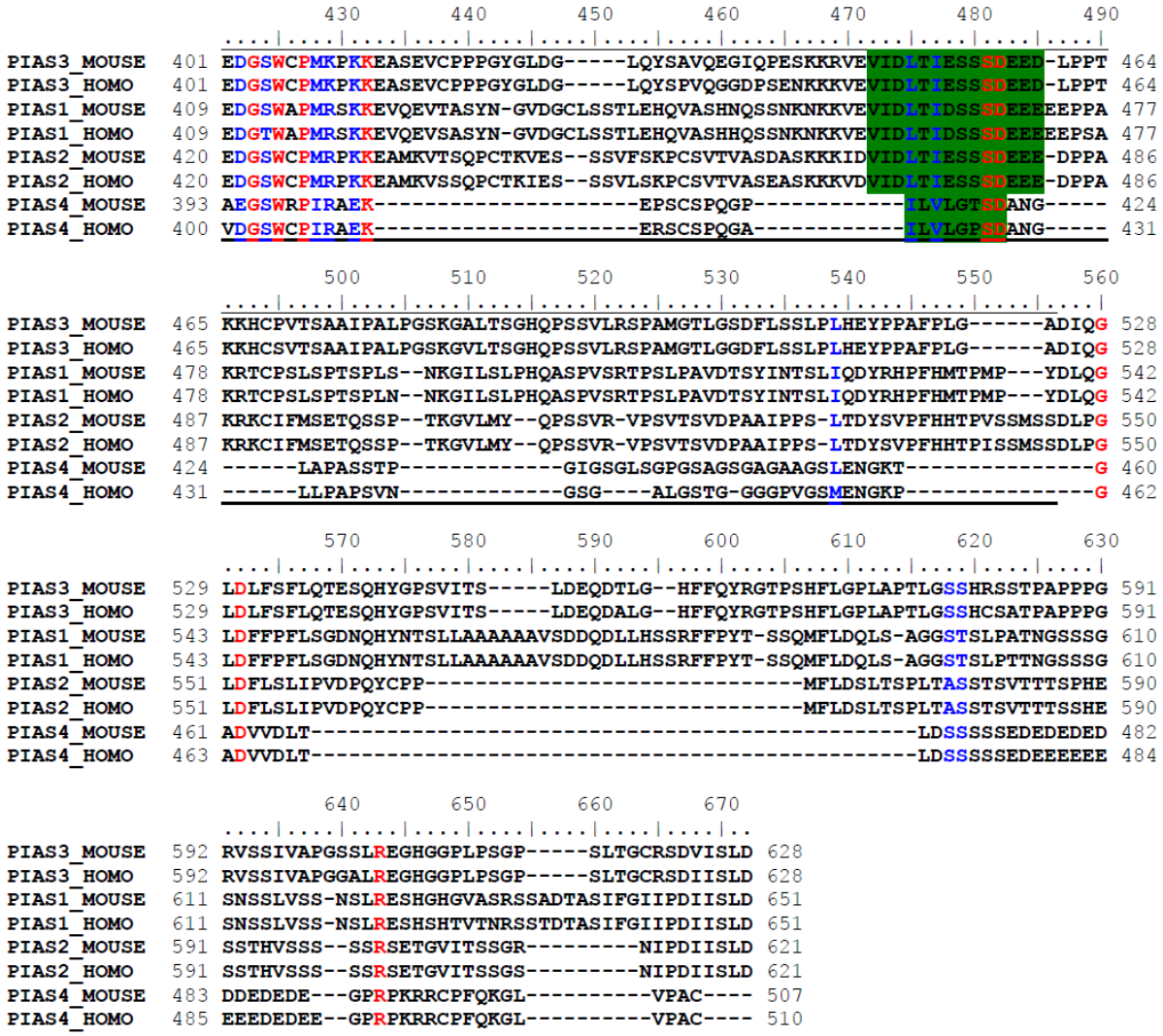

Figure 2.2 Sequence alignment and analysis of PIAS family members from different organisms (continued from previous page). Alignment of the PIAS protein family aligned with 'CLUSTALW' software (Larkin *et al.*, 2007). The numbers at the beginning and at the end of each sequence indicate the position of the first and last of the aligned residues in the respective protein sequence. Shading represents the conservation of residues (scored according to Blosum62 score table). Identical residues are in red, highly conserved residues are in blue and with low conservation are residues in black. The accession numbers for the protein sequences used in the multiple sequence alignments are shown in procedures section **2.2.1**. The broken lines at the bottom and top of the alignment indicate the beginning and end of the domains. The first domain is the SAP domain marked by its signature sequence "LQVLL" in yellow background. The second domain is the PINIT domain and is marked with its highly conserved signature sequence "PINIT" in grey backgrounds and underlined with a bold black line. The third domain is the SP–RING domain and is marked with its zinc ion binding residues indicated in boxes. The last domain is the diverse C–terminal acidic domain, which is marked with its signature SUMO–1 binding motif and acidic residues "VIDLTIESSSDEED" indicated with a green background.

2.3.2 PINIT domain sequence alignment and analysis

Comparative sequence analysis of the PINIT domain of the PIAS3 family with the PINIT domain of the Siz1 yeast revealed differential distribution of the Pro, Leu, Thr residues in the 'PINIT' motif (**Figure 2.3**). Although the alignment of the PINIT domain of the PIAS3 with the PINIT domain of Siz1 showed low sequence similarity, the two proteins have common functions (Yunus and Lima, 2009). However, sequences may be distant with low similarity but the tertiary fold of the protein is often highly conserved, hence similar biological activity.

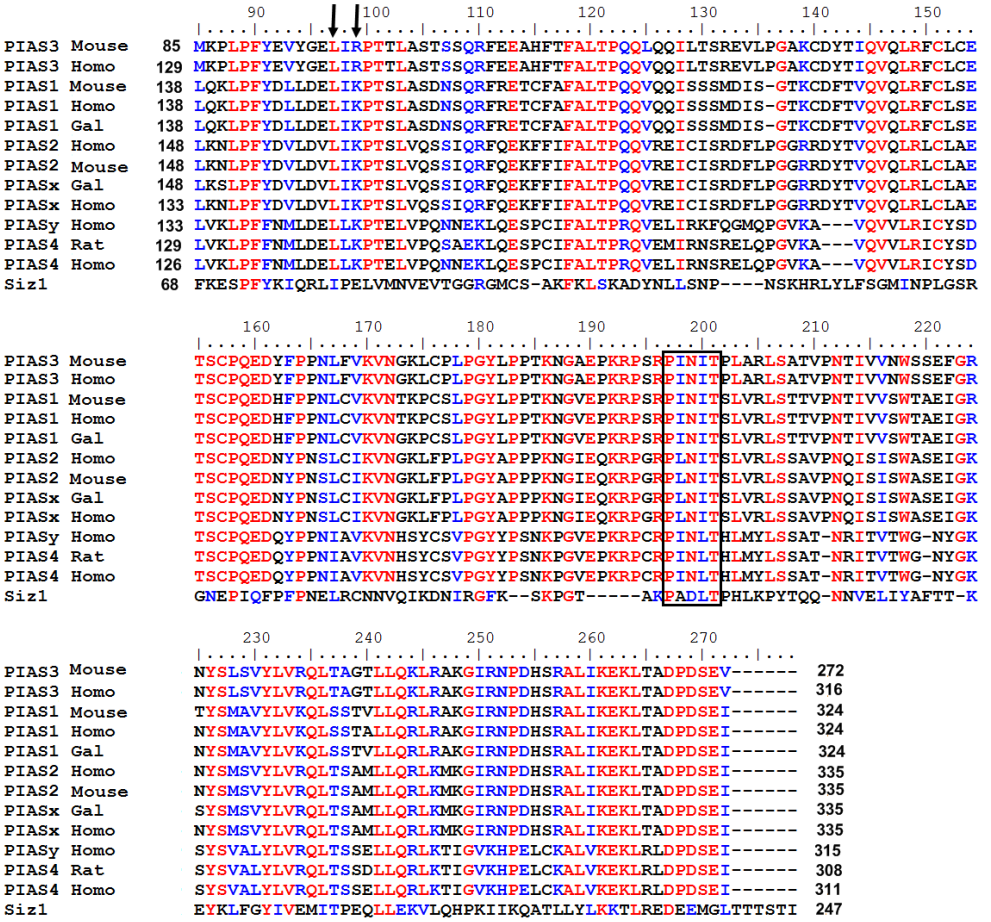

Figure 2.3 Sequence alignment and analysis of the PINIT domains of the PIAS family and yeast Siz1 protein. The alignment of PINIT domain sequences of the PIAS family proteins and the PINIT domain sequence of Siz1 using CLUSTALW software (Larkin *et al*., 2007). The numbers indicate the first and last residue in the particular sequence and the numbers at the top indicate the position of mouse PIAS3 residues. Shading represents the conservation of residues (scored according to Blosum62 score table). Identical residues are in red, highly conserved residues are in blue and residues of low conservation are in black. The conserved "PINIT" residues are indicated in the box and the two arrows indicate PINIT residues subjected to mutational analysis in this study (L97 and R99). The accession numbers for the protein sequences and the organism used in the multiple sequence alignments are shown in procedures section **2.2.1**. (adapted from Mautsa *et al*., 2010)

2.3.3 Acidic domain sequence alignment and analysis

The alignment of the C–terminal domain ($PIAS3_{400–523}$) of PIAS proteins shows that it is more diverse, however, the distribution of several serines (Ser) at the N–terminal border of the domain is highly conserved (**Figure 2.4**). Previous studies have identified similar consensus sequences which include clusters of Val, Ile, Leu and the acidic residues, Asp and Glu (Hannich *et al.*, 2005; Minty *et al.*, 2000). Although the alignment showed a diverse C–terminal domain of PIAS3, of significance is the highly conserved motif "VIDLTIESSSDEED" (indicated on **Figure 2.4**). Previous studies predicted the motif as a SUMO–1 binding motif (Song *et al.*, 2005) that recognises the SUMO moiety of modified proteins in sumoylation–dependent cellular functions.

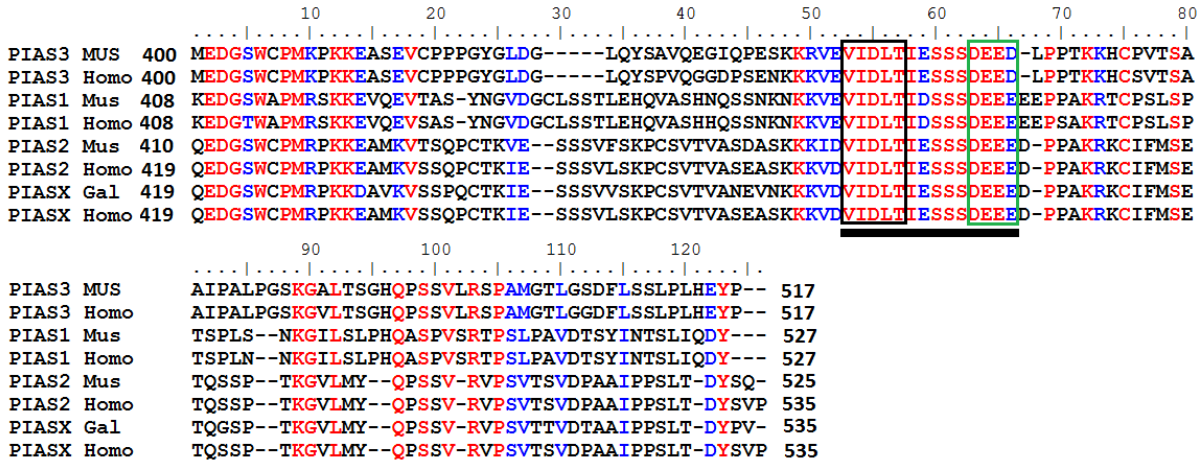

Figure 2.4 Sequence alignment and analysis of the acidic domains PIAS family. Alignment of the C–terminal domain sequences of PIAS protein family using CLUSTALW software (Larkin *et al.*, 2007). The numbers indicate the first and last residue in the particular sequence and the numbers at the top indicate the position of mouse PIAS3 residues. Shading represents the conservation of residues (scored according to Blosum62 score table). Identical residues are in red and highly conserved residues are in blue and residues of low conservation are in black. The predicted SUMO binding motif (black box) and the acidic amino acids rich region (green box) are indicated in the boxes. The highly conserved motif "VIDLTIESSSDEED)" of the C–terminal domain of PIAS is underlined with a bold black line. The accession numbers for the protein sequences and the organism used in the multiple sequence alignments are shown in procedures section **2.2.1**.

Other non–covalent SUMO binding sites of several proteins also contain similar sequences such as the Epstein–Barr virus nuclear antigen 3C (EBNA3C) with the sequence 'DDVIEVIDVETTE" (Rosendorff *et al.*, 2004). These sequences are similar to that of the SUMO–1 binding site of PIASx 'VDVIDLT' (**Figure 2.4 boxed**) determined by Song *et al.* (2004, 2005). Substitution of the "VILT" individually with alanine reduced the binding

affinity of the mutated peptide to SUMO–1, particularly the threonine which resulted in reduction of affinity by ~10 fold (Song *et al.*, 2005). The conservation of these residues at the C–terminal domain of PIAS3 protein may suggest a common function. Interestingly as mentioned previously, $PIAS3_{400–523}$ has been suggested to interact with the coiled coil domain of the STAT3 protein (Borghouts *et al.*, 2010).

2.3.4 PINIT domain secondary structure prediction analysis

The alignment in **Figure 2.2** and **2.3** showed high conservation of the PINIT domain in the PIAS protein family. Although highly conserved residues exist elsewhere within the domain, focus on the N–terminal region of the PINIT domain is ideal since previous mutations elsewhere within the domain did not abrogate its biochemical functions. The "PINIT" motif is highly conserved (**Figure 2.2 and 2.3**) and secondary structure predictions of the PINIT domain of PIAS3 by four independent prediction software packages (**Figure 2.5**) indicate a high consensus. The N–terminal region of the PINIT domain (23 amino acids) was identified as the minimal epitope that can trigger apoptosis (Yagil *et al.*, 2009).

```
PINIT       MKPLPFYEVYGELIRPTTLASTSSQRFEEAHFTFALTPQQLQQILTSREVLPGAKCDYTIQVQLRFCLCETSCPQEDYFPPNLFVKVNGKLCPLP
SS PSIPRED         EEEEEE   EEE          EEEEEE  HHHHHHHHH             EEEEEEEEE                 EEEEE  EE
CONF        988987066554349889524567731122489984999999886384445555578517999999842688774445897539999999870267
SS JNET            EEEEEE   EEE          EEEEEE  HHHHHHHHH             EEEEEEEEE                 EEEEE  EE
CONF        988987066554349889524567731122489984999999886384445555578517999999842688774445897539999999870267
SS ProfO         EEEEE      EEEE         EEEEEEE  HHHHHHHH             EEEEEEEEEE                EEEEEE
CONF        888875345433357755664577777654578875665677776467777777777764788888875577887777787647887557656777
SS ProfR           EEEEEE   EEE          EEEEEE  HHHHHHHHH             EEEEEEEEE                 EEEEE  EE
CONF        988987066554349889524567731122489984999999886384445555578517999999842688774445897539999999870267

PINIT       GYLPPTKNGAEPKRPSRPINITPLARLSATVPNTIVVNWSSEFGRNYSLSVYLVRQLTAGTLLQKLRAKGIRNPDHSRALIKEKLTADPDSEV
SS PSIPRED                        HHH         EEEEEE       EEEEEEEEEEE  HHHHHHHHHH      HHHHHHHHHHHH
CONF        767888888667889898744422203789973699996478895799999999827999999998659879899999999886139867779
SS JNET                           HHH         EEEEEE       EEEEEEEEEEE  HHHHHHHHHH      HHHHHHHHHHHH
CONF        767888888667889898744422203789973699996478895799999999827999999998659879899999999886139867779
SS ProfO                        EEE         EEEEEEE      EEEEEEEEEE   HHHHHHHHH       HHHHHHHHHHH
CONF        788888888888888888776553344467776478888764777478888888756767888888874788875788888887647777788
SS ProfR                          HHH         EEEEEE       EEEEEEEEEEE  HHHHHHHHHH      HHHHHHHHHHHH
CONF        767888888667889898744422203789973699996478895799999999827999999998659879899999999886139867779
```

Figure 2.5 Prediction of the secondary structure of the PINIT domain. Secondary structure prediction of the PINIT domain using three independent web based algorithms (**Appendix F1**): PSIPRED (Jones, 1999); JNET (Cuff and Burton, 1999); ProfR (Rost, 2001); and ProfO (Ouali and King, 2000). **H (**red**)** indicates α–helix; **E** (blue) indicates β–strand; CONF is the prediction confidence level indicated by numbers from 0 to 9 as the lowest and highest confidence level respectively; SS, secondary structure; and gaps indicates loops.

Previous mutational studies of PIAS3 Y94P mutant protein have been found to abrogate the association of PIAS3 with either MITF or STAT3 (Sonnenblick *et al.*, 2004; Levy *et al.*, 2006). The *de novo* predicted model by Levy *et al*, 2006 showed that the position of Y94 on

the helix. Furthermore, Y94P mutation resulted in disruption of the helix due to the nature of the proline residue. Proline does not fit into the regular part of either helix because it does not have a backbone-NH available to take part in an H-bonding (Williams and Deber, 1991). In the helix center, the ring pushes away the proceding (N-terminal) turn of the helix producing a bend and breaking the next H-bond. The resulting loss of protein function observed by Levy *et al*, (2006) might possibly be due to disruption of the helix. Secondary structure prediction analysis of PINIT mutant proteins PINIT–L97A, PINIT–R99N and PINIT–R99Q showed that the secondary structure predictions compare well to the unmodified PINIT domain for PINIT–R99N and PINIT–R99Q, with a small change in PINIT–L97A. These data suggest that the mutations are likely not to affect the local secondary structure features of the protein (**Figure 2.6**).

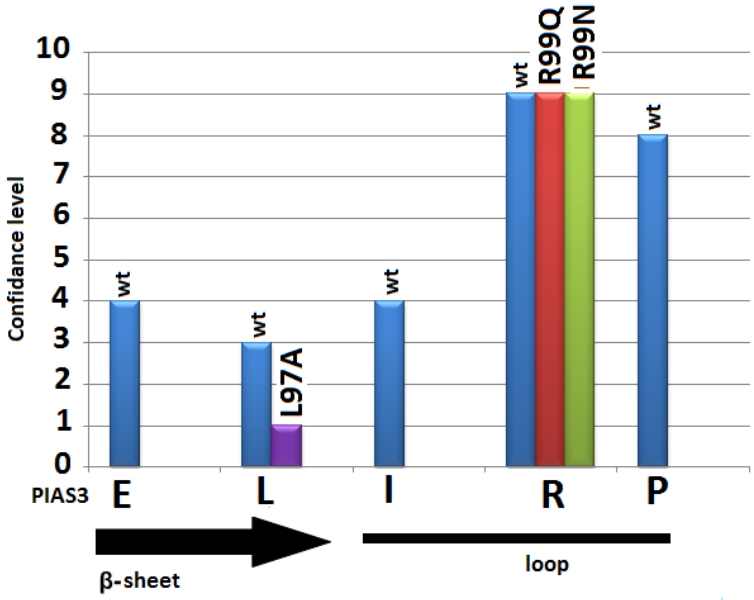

Figure 2.6 Assessment of the effect of the mutation on the local structure of the PINIT domain. *In silico* secondary structure prediction of each residue in PINIT domain before and after mutation with confidence values ranging from 0 to 9. wt, represent the unmodified PINIT domain; L97A, represent PINIT–L97A mutant protein; R99N represent PINIT–R99N mutant; R99Q represent PINIT–R99Q mutant. The confidence levels from 0 to 9 is shown in the Y–axis and the secondary structure features of the unmodified PINIT domain and the residues in the region are shown in the X–axis. L97A mutation resulted in the lower confidence level of Alanine being in the β–sheet and the R99N and R99Q mutations predicted with the highest confidence level that the residues remain in the loop suggesting no local structural change.

2.3.5 The PINIT domain modeling with MODELLER

The PINIT domains of Siz1 and PIAS3 have 17 % sequence identity. Quality and confidence of any generated model is dependent on sequence identity. However, the overall assessment

of the secondary structure prediction of the PINIT domain of Siz1 compared to the secondary structure prediction of the PINIT domain of PIAS3 (**Figure 2.7**) showed a high consensus and a high degree of fold similarity (**Figure 2.7**).

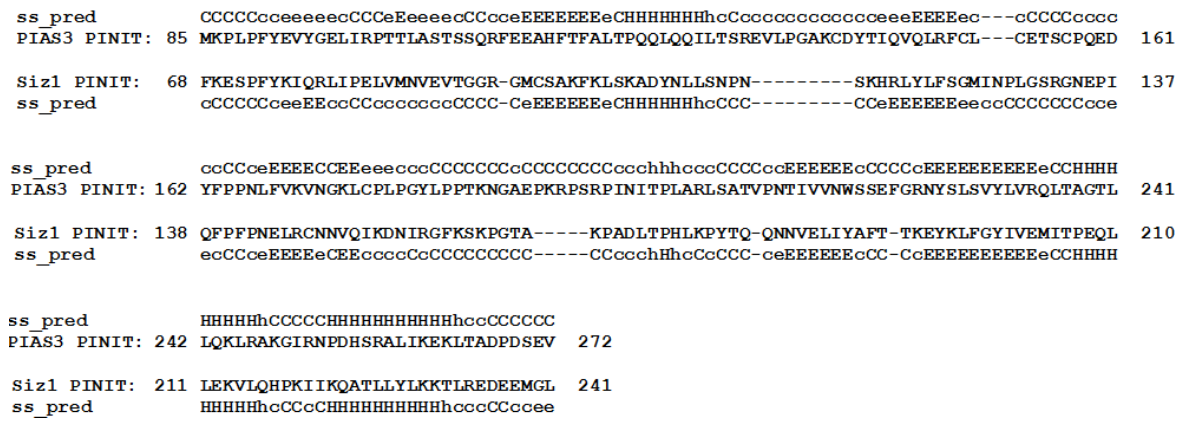

Figure 2.7 Sequence–structure alignment and secondary structure prediction of the PINIT domain of PIAS3 and Siz1 protein. Alignment was generated using CLUSTALW and the secondary structure predicted by PSI–PRED (Jones *et al*., 1999) was performed in the web based programme HHpred (Söding *et al*., 2005). H, helices prediction; E, β–strands prediction; C, coil prediction; ss–pred, secondary structure prediction; (H,E,C capital letter indicates prediction with high confidence and h,e,c indicates prediction with lower confidence) Siz1, PDB ID: 3I2D.

The predicted PINIT domain model (**Figure 2.8A**) revealed two antiparallel β–sheets; one includes β1, β2, β4, and β7, and the other includes β3, β5, and β6. The β–sheets are connected by protruding loops (L1, L2, and L3) that join strands β2–3, β4–5, and β6–7 at one end of the molecule, while β3–4 and β5–7 are connected by a helix, α1, and a loop, respectively, on the opposite surface (**Figure 2.8A**). Superimposition of the predicted model with the X–ray structure shows a close agreement of all the secondary structure features with a minor mis–alignment on the flexible parts of the model (**Figure 2.8B**) (loops and C–terminal) even though the predicted model has undergone rigorous loop refinement.

.

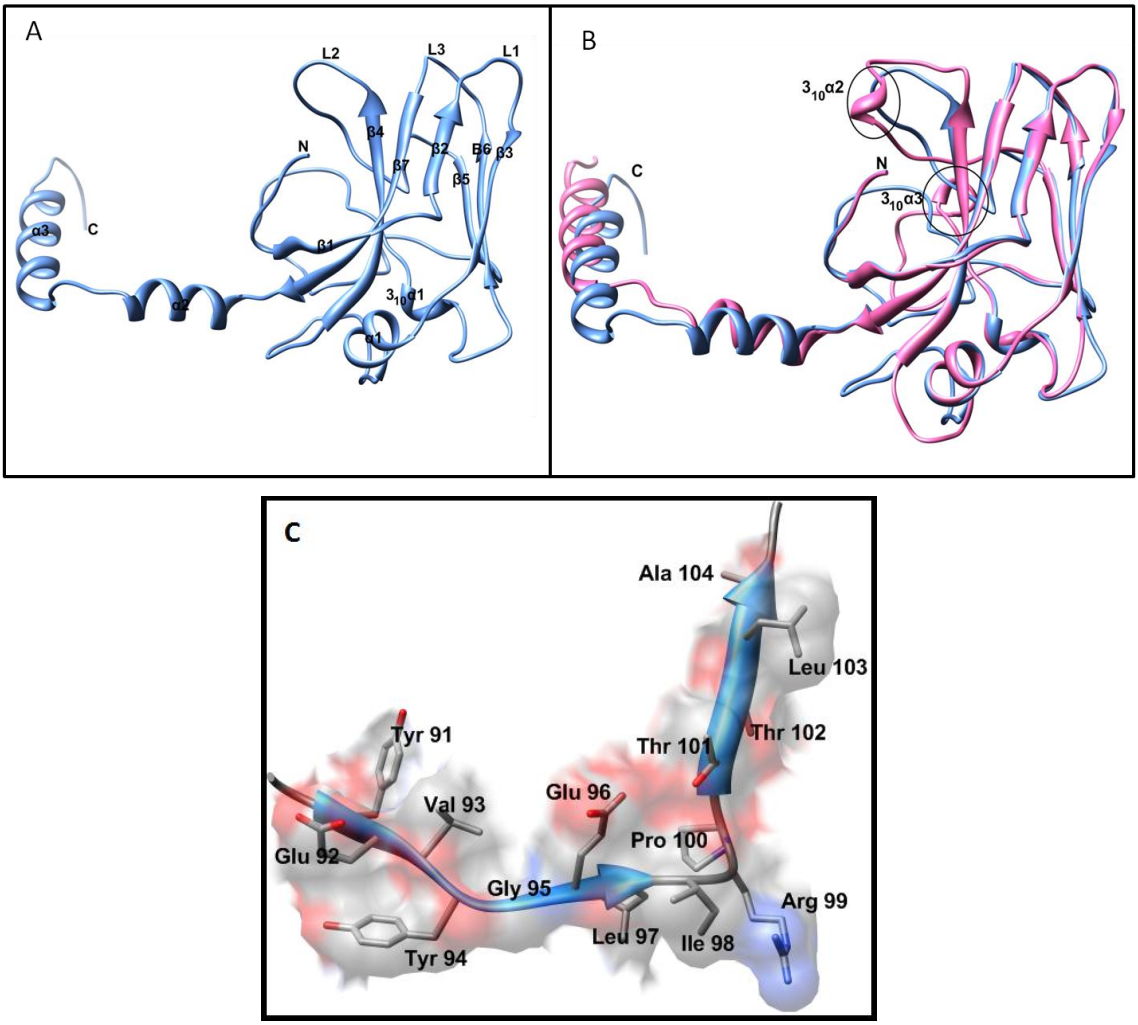

Figure 2.8 Ribbon representation of the PINIT domain model of PIAS3 and its superimposition on the PINIT domain of Siz1. A) The PINIT domain model created using MODELLER–9v3 and evaluated as described below. The model revealed three antiparallel β–sheets connected by protruding loops and the helix–loop–helix at the C–terminus. The model was generated using the PINIT domain X–ray structure of the Siz1 (PDB ID: 3I2D) (Yunus and Lima, 2009). **B**) Siz1–PIAS3 PINIT domain structural alignment. The superimposed generated model of PINIT domain of PIAS3 (cornflower blue) with the PINIT domain structure of Siz1 (hot pink). The superimposition was performed in UCSF Chimera 1.5 (Goddard, *et al*., 2005) and the structure shows alignment in the helix structural features. **C**) PINIT domain N–terminal β–strand–loop–β–strand enveloped by a transparent molecular surface, residues mutated are L97 and R99, also residue Y94 mutated by Levy *et al*. (2006).

2.3.6 Model quality assessment and evaluation

The 100 generated PINIT models were visually inspected in PyMol and their quality assessed. Furthermore, all models were evaluated with a normalised DOPE Z–score (Discrete Optimized Protein Energy) from modeller package that used standard Modeller energy function and the best model with DOPE Z–score of –0.363 was chosen, however a good quality model has a DOPE Z–score of -0.5 (Shen and Sali, 2006). To further improve the quality of the model the loops were refined using python script for loop refinement

(**Appendix F2**) as described above (section 2.2.5). Loops Glu–134 to Thr–144, Glu–155 to Phe–164 and Leu–179 to Ile–199 were refined and the model was further improved to give a final DOPE Z–score of -0.369. The assessment of the PINIT model quality by visual identification of potential errors using colouring was performed using MetaMQAPII (Pawlowski *et al.*, 2008) which predicts the local deviation of residues in the model for their counterpart in the native structures. The temperature (B–factor) fields in PINIT domain model file were replaced with the MetaMQAP score corresponding to linear scaling of values of 0.00 (predicted no deviation) and 99.99 (predicted deviation of ≥10Å) ((Pawlowski *et al.*, 2008). The results were visualised with PyMol version 0.98 (Delano, 2004) that allowed colouring of the structure according to the B–factor values. The per–residue accuracy was visualised as a colour in a spectrum between blue (predicted high accuracy) and red (predicted low accuracy) (**Figure 2.9**). The assessment shows that the PINIT model has high accuracy prediction in the core structure β–sheet compared to the template structure (Siz1) indicated by blue regions on β–sheet structures of the PINIT model. However, the N–terminal β–sheet and the C–terminal helix–loop–helix of the PINIT model deviated from the native structure indicated by the shift of colour spectrum from blue towards yellow and red (**Figure 2.9B**).

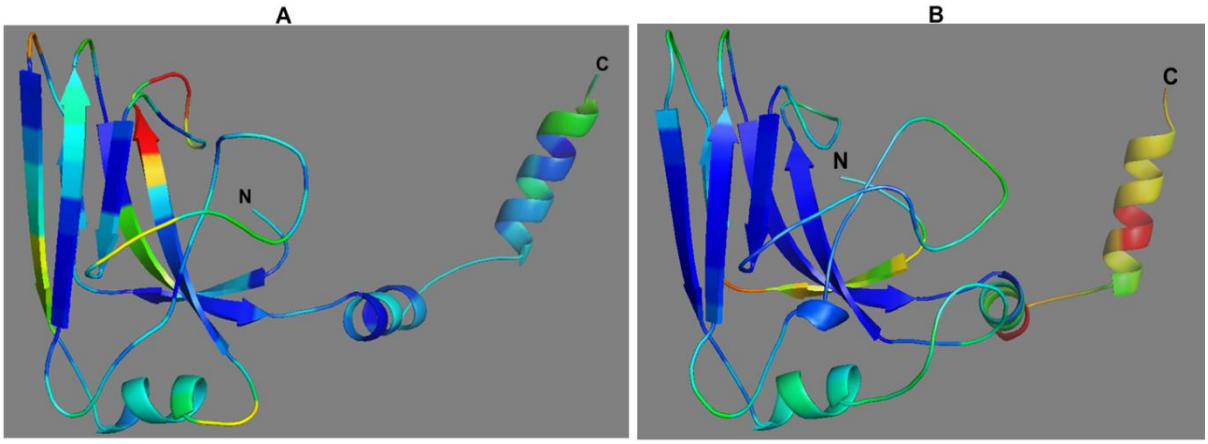

Figure 2.9 Visual identification of potential errors in the PINIT domain structure of Siz1 and the generated PINIT domain model using 'colouring' by MetaMQAPII. **A**) Quality assessment of the PINIT domain structure of Siz1 protein using web based programme MetaMQAPII (Pawlowski *et al.*, 2008). The spectrum of colours from blue to red indicates the spectrum of residues predicted to be correct to incorrect. **B**) Quality assessment of the PINIT domain model of PIAS3 protein generated using the PINIT domain structure of Siz1 (PDB ID: 3I2D). The predicted deviation from the native structure is indicated by a shift of the colour spectrum from blue towards yellow and red. The images were rendered in PyMol version 0.98 (Delano, 2004).

Evaluation of the model by Verify3D (Eisenberg *et al.*, 1997) revealed results that were consistent with the evaluated results from MetaMQAPII (**Figure 2.10**). The N–terminal β–sheet and the C–terminal helix–loop–helix region were shown to deviate from the native structures (**Figure 2.9B, Figure 2.10**). Furthermore, Verify3D (Eisenberg *et al.*, 1997) revealed residues in the N–terminal β–sheet and C–terminal helix–loop–helix local secondary structure were below a threshold score (0.1 indicated by a red line) and indicating a slight deviation from the native structure. In particular Ser 132 (indicated by a blue circle **Figure 2.10**) was below 0 score and indicated a large deviation from the expected environment of the native structures (**Figure 2.10**). These regions correspond to the N–terminal β–sheet (β1) and the C–terminal α–helix (α2 and α3) of the predicted PINIT domain model (**Figure 2.9B**)

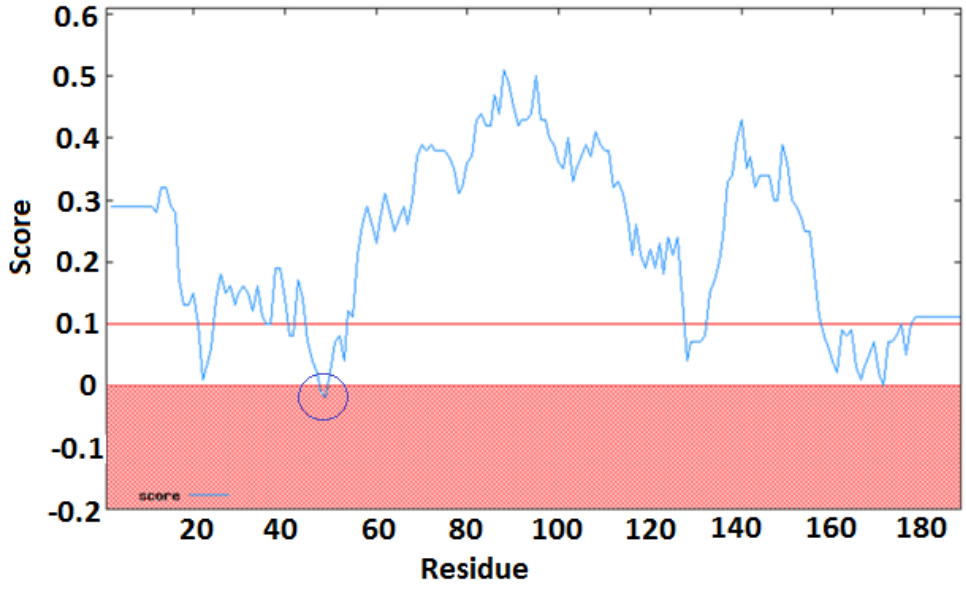

Figure 2.10 Profile score for each residue in the PINIT model by Veryfy3D. Evaluation of the environment of each residue in the predicted PINIT domain model with respect to the expected environment as found in high resolution X–ray structures (Luthy *et al.*, 1992). The N–terminal and the C–terminal residues show a deviation from the expected environment of a native protein and the residue Ser132 indicated by a blue circle has a negative score.

2.3.7 Analysis of the predicted structure of the PINIT domain and mutated PINIT domain derivatives

To identify potential structural determinants of the PINIT domain that enable it to bind to STAT3, the PINIT domain model was generated from the crystallographic structure of Siz1 (PDB ID: 3I2D) as template (Yunus and Lima, 2009). The predicted structure of the PINIT domain revealed a conserved "PINIT" motif buried in the hydrophobic core, the R99 residue

is surface exposed and the L97 residue is also buried in the hydrophobic environment (**Figure 2.11**). While the PINIT domain had low sequence identity to Siz1, a comparison of the secondary structure predictions for the two proteins showed a good match and the DOPE Z–score of -0.369, after loop refinement indicated that the model was relatively accurate. Homology models of the mutated PINIT domains were generated using the unmodified PINIT model as a template and evaluated in the same manner. The first helix of the PINIT domain is the focus of mutational studies. Previous mutations focused on the first helix gave little attention to the structural bases on the loss of ability to bind STAT3 (Levy *et al.,* 2006; Yagil *et al.,* 2009). The PINIT–L97A model (**Figure 2.12A**) showed that the replacement of leucine with alanine potentially resulted in a loss of hydrophobic or van der Waals contacts within a 3.5 Å sphere. The loss of non–polar contacts possibly results in local perturbations of the PINIT conformation. The PINIT–R99Q and PINIT–R99N models illustrated how the replacement of Arg with Gln (**Figure 2.12B**) or Asn (**Figure 2.12C**) resulted in both a reduction in side–chain length and a loss of charge.

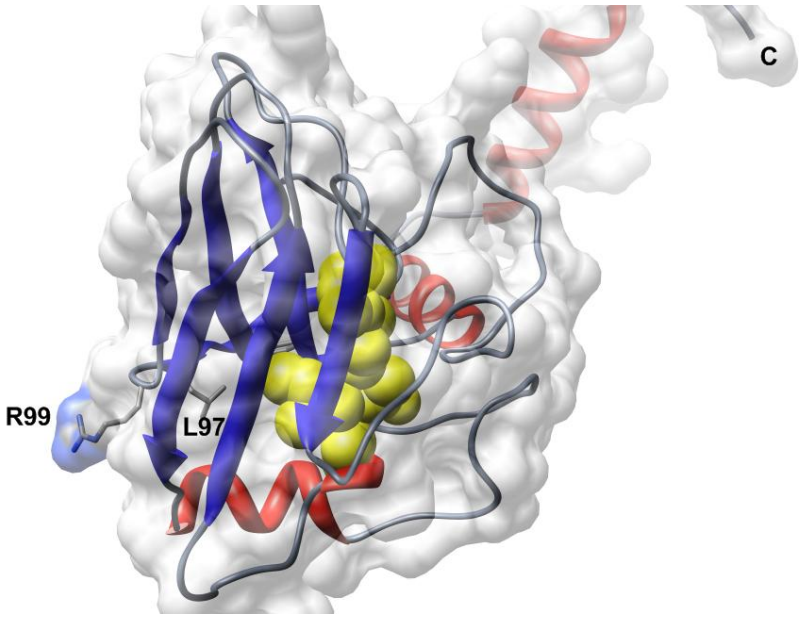

Figure 2.11 Ribbon representation of the PINIT domain of PIAS3 enveloped by a transparent molecular surface. The PINIT model shows the position of the mutated residues (L97 and R99); the R99 residue protruding outside, and the L97 buried in the hydrophobic core. The PINIT motif is shown as space fill (yellow), and is buried in a deep cleft in the hydrophobic core. The antiparallel β–sheets are connected by protruding loops and the helix–loop–helix at the C–terminus. The model was generated from the crystal structure of the PINIT domain of Siz1 as template (PDB ID: 3I2D) (Yunus and Lima, 2009) using MODELLER 9v3 and rendered in UCSF Chimera 1.5 (Goddard, *et al.*, 2005). (adapted from Mautsa *et al*., 2010).

The L97A PINIT mutant model (**Figure 2.13**) revealed an increase in contact distance with neighbouring atoms when leucine is replaced with alanine residue. In solution this may result in possible loss of van der Waals and hydrophobic interactions and local collapse of the secondary structure features due to tight packing, thereby affecting the function of neighbouring atoms.

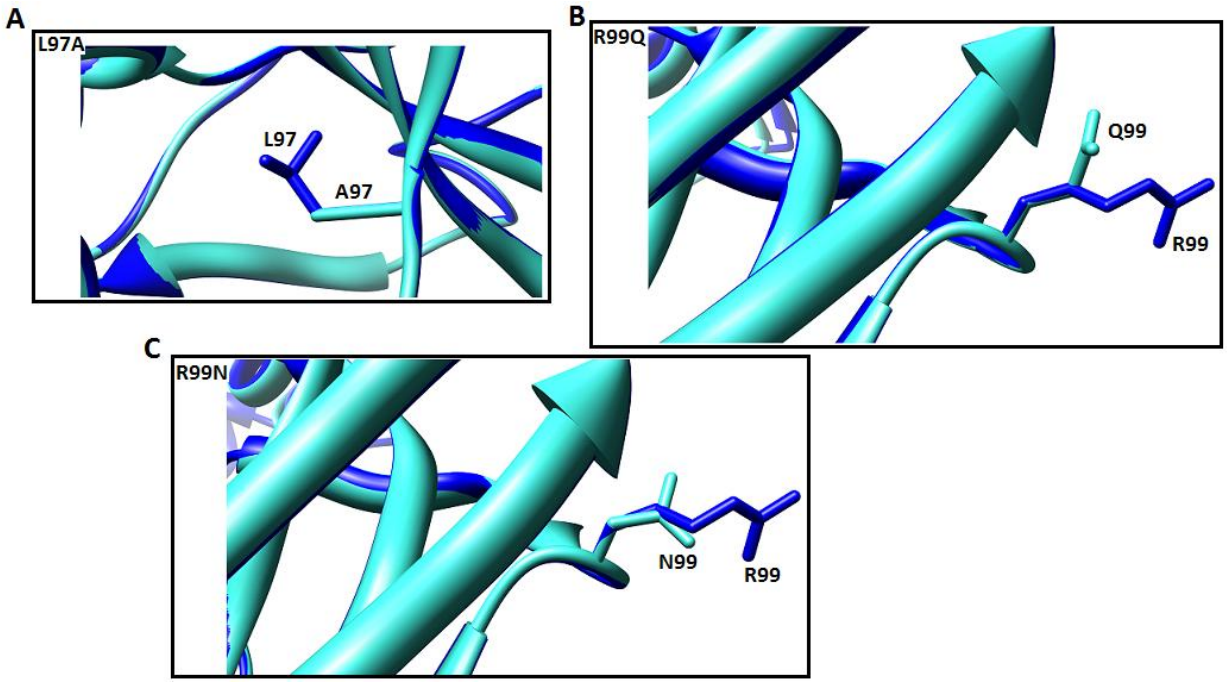

Figure 2.12 Superimposition of the unmodified and mutant PINIT domain models. (A) PINIT–L97A mutant model (turquoise) superimposed with unmodified PINIT domain (blue); **(B)** PINIT–R99Q mutant model (turquoise) superimposed with the unmodified PINIT domain (blue); **(C)** PINIT–R99N mutant model (turquoise) superimposed with the unmodified PINIT domain (blue). (adapted from Mautsa *et al*., 2010)

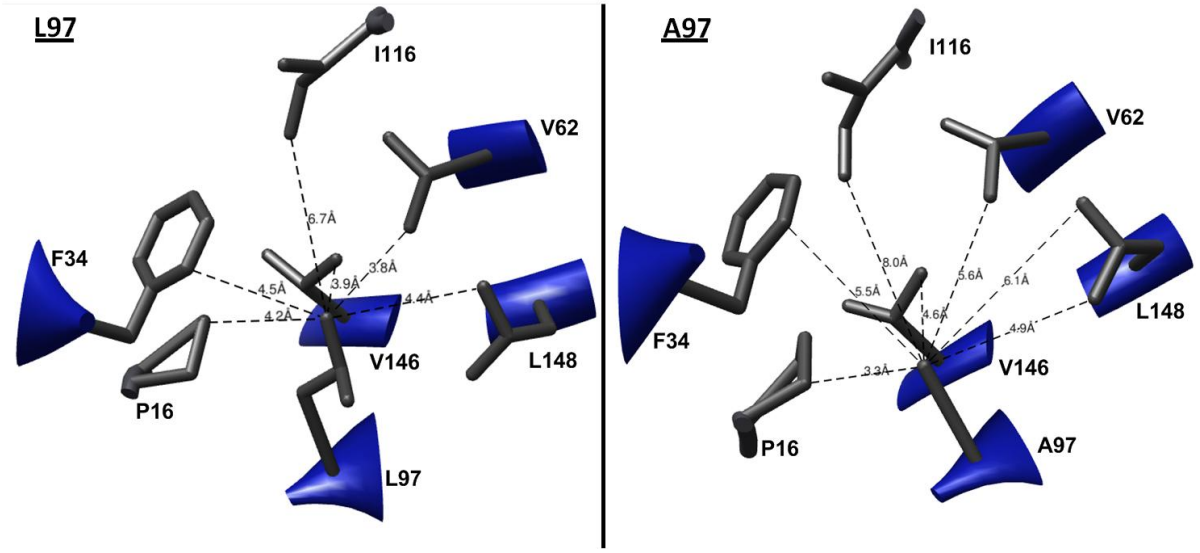

Figure 2.13 Assessment of the contact distance of the L97 of the unmodified PINIT domain model and A97 of the mutant PINIT domain model with the neighbouring atoms. The distance in Angstrom between the L97 and the nearest neighbouring atoms increased when this residue was replaced with Ala.

2.4 DISCUSSION AND CONCLUSION

Computational analysis of the PIAS protein family revealed four conserved structural domains and motifs previously identified. The PINIT and acidic domains or the PIAS3 protein were previously shown to interact with the STAT3. Sequence analysis of the separate domains showed that the PINIT domains are highly conserved. The Acidic domain showed sequence diversity within the PIAS family. Although the SUMO–1 binding motif is highly conserved within the acidic domain, no previous studies have shown involvement in PIAS3–STAT3 interaction. The study presented here focused on the PINIT domain structural determinants in PIAS3–STAT3 interaction.

While there is a very low global sequence similarity between Siz1 and other PIAS proteins (**Figure 2.3**) (17 %), sequence similarity within the PIAS family is on average 80 %. However, there is structural conservation across the homologs as shown by the secondary structure prediction of the PIAS3 PINIT domain and the Siz1 structure from HHpred (**Figure 2.7**). This shows that although the sequence similarity is so diverse within the PIAS orthologs, structure is conserved. Comparison of predicted secondary structure features of the PINIT domain of PIAS3 and the predicted secondary structure features of the PINIT domain of Siz1 show a high degree of similarity and consensus on prediction (**Figure 2.7**). The PINIT domain of Siz1 structure was then used to generate the model of the PINIT domain of

PIAS3. The generated PINIT model (**Figure 2.11**) showed the "PINIT" motif on a loop buried in the hydrophobic core. However, its location on the loop indicates the possibility of a certain degree of flexibility to achieve its functional purpose. Furthermore, the model shows interesting surface exposed residues that are within the region that was previously the study focus of PIAS3–STAT3 interactions (Levy *et al.,* 2006; Yagil *et al.,* 2009). Previous studies on the disruption of this motif resulted in delocalisation of the PIAS3 proteins. Furthermore, Levy *et al*. (2006) determined that 50 amino acids at the N–terminus of the PINIT domain are sufficient to bind to STAT3 while Yagil *et al*. (2009) showed that 23 amino acids at the N–terminus of the PINIT domain achieve the same function. The predicted PINIT homology model showed that the N–terminal 23 amino acid epitope is located on a loop–sheet–loop structure (**Figure 2.8C**). The PINIT domain showed that the N–terminal antiparallel β–sheet formed by β1, β2 and β3 (**Figure 2.8**) formed a stable structure. The final predicted PINIT domain model assessed by MetaMQAPII (Pawlowski *et al.,* 2008 (**Figure 2.8 and Figure2.9 and Figure 2.10**) was shown to be a good quality model that can be adopted with high confidence for use as a template to generate the mutant models.

A previous mutation Y94P resulted in loss of PIAS3 function, probably due to disruption of secondary structure (Levy *et al.,* 2006). The generated PINIT model showed R99 amino acid surface exposed (**Figure 2.11C** β–sheet–loop–β–sheet) and potentially able to take part in hydrogen bonding and ionic interactions whereas amino acid L97 on the same loop is buried inside. Therefore, these conserved residues were targeted for mutational analysis in this study. Furthermore, the sequence analysis results (**Figure 2.2 and Figure 2.3**), together with the previous *in vivo* and *in vitro* mutational analysis of PIAS3 (Levy *et al.*, 2006; Yagil *et al.,* 2009) highlighted the need to investigate the involvement of these residues in the PIAS3–STAT3 interaction. Replacement of the surface exposed R99 with Asn and Gln residues (which are less bulky and lack charge) and the L97 residue, with Ala (a relatively small residue) (**Figure 2.13**) could possibly affect PIAS3–STAT3 interaction. The assessment of the structural effect of these mutated residues (**Figure 2.6 and Figure 2.12**) indicated no potential gross disruption of the local structure of the PINIT domain. However, the PINIT domain model revealed that the replacement of the L97 residue buried in the hydrophobic core with Ala increased the contact distance between the residues within the hydrophobic core (**Figure 2.13**) and this may result in tight packing under physiological conditions affecting the orientation of the neighbouring residues and possibly the biochemical function of the PIAS3 protein. Previous PIAS3–STAT3 interactions studies focused mainly on the N–

terminus of the PINIT domain. However, PIAS3 is a multidomain protein and other domains may be involved in PIAS3–STAT3 interaction. In particular, findings by Borghouts *et al.* (2010) suggested that the C–terminal domain of PIAS3 is involved in PIAS3–STAT3 association. These findings were based on yeast two–hybrid experiments also, mutational analysis in the context of PIAS3–STAT3 direct interaction were not performed. However, performed mutational analysis were on Val, Ile, Leu, Thr residues in the context of PIAS3–SUMO–1 association and resulted in reduced binding affinity of the SUMO–1 (Song *et al.,* 2005). These findings highlight the need to investigate the acidic domain of PIAS3 relative to PINIT domain and PIAS3 full length.

This chapter employed bioinformatics tools to accurately predict the structural elements of PIAS3. A high quality PINIT domain model was generated and adopted with confidence to predict key residues governing the PINIT domain function and to assess the structural effect of mutation of identified residues. Furthermore, the information revealed by the sequence alignment and secondary structure predictions will guide the determination of the PINIT domain boundaries for *in vivo* and *in vitro* expression of a folded and functional protein. Insights into *in vitro* and *in vivo* activity of *in silico* predicted mutants presented here will be discussed further in Chapter 3 and 5.

CHAPTER 3

ANALYSIS OF THE CELLULAR LOCALISATION OF PIAS3 AND ITS DOMAINS

Regulation of STAT3 localisation by PIAS3 requires complete elucidation. It is critical to understand the molecular determinants governing protein–protein (PIAS3–STAT3) interaction and the binding interface between PIAS3 and STAT3 in order to understand PIAS3 potential in the regulatory control mechanism. Here, in vivo analysis of PIAS3–STAT3 interaction was investigated by comparative localisation analysis of full length PIAS3, the PINIT and acidic domains with endogenous STAT3 in HeLa human cervical cancer cells in the presence and absence of IL–6 stimulation.

3.1 INTRODUCTION

PIAS3 subcellular localisation is contentious. PIAS3 was originally found to be localised mainly in the nucleus, seemingly governed by the presence of the PINIT motif (Duval *et al*., 2003). More recently Dabir *et al*. (2009) provided evidence for a mainly cytoplasm distribution of PIAS3. In their description nuclear translocation was reliant on stimulation of the gp130/JAK/STAT3 pathway and STAT3 phosphorylation at tyrosine 705. Furthermore, Dabir *et al*. (2009) showed that phosphorylated STAT3 levels were significantly reduced in the nucleus and this reduction was found to be PIAS3 dose–dependent. The nuclear localisation signal has not been fully described. The localisation and co–localisation of PIAS3 and activated STAT3 suggested PIAS3 role in regulation or mediation of phosphorylated STAT3 transcriptional activity. Knowledge of the interaction interface of PIAS3–STAT3 would further our understanding of PIAS3 activity in this tightly regulated system. The PINIT and the acidic domains of PIAS3 were suggested to interact with STAT3, the former interacting with the DNA binding domain of STAT3 and the latter interacting with the coiled coil domain of STAT3 (Chung *et al*.,1997; Duval *et al*., 2003; Yamamoto *et al*, 2003; Borghouts *et al*., 2010).

Using the HeLa cervical cancer cell model, the study presented here attempted to show the comparative localisation of full length PIAS3, PINIT and acidic domain of PIAS3 with endogenous STAT3.

3.2 METHODS

3.2.1 Construction of expression plasmid encoding for Flag–PIAS3, Flag–PINIT, Flag–acidic and mutants Flag–PIAS3

The p513–flag–PIAS3 mammalian expression construct with expression controlled by simian virus 40 (SV40) promoter (Duval *et al.*, 2003) was a kind donation from Dr Hélène Boeuf (Université de Bordeaux, France). The presence of the PIAS3 coding region in p513–flag–PIAS3 plasmid was confirmed by diagnostic restriction endonuclease analysis and automated DNA sequencing (Rhodes University DNA sequencing unit). The plasmid was used as the template for amplification of the PINIT and acidic domain coding regions (as described in Chapter 2) using the following primer sets: PINIT domain forward primer (5'–CAT ATG AAG CCC CTG CCC TTC –3') with *Nde*I site (underlined) and the PINIT domain reverse primer (5'- AAG CTTATTA CAC TTC ACT GTC GGG GTC - 3') *Hind* III site (underlined) and acidic domain forward primer (5'- CAT ATG GAA GAT GGA TCC TGG TGT C –3')

with *Nde*I site (underlined) and acidic domain reverse primer (5'–AAGCTTTAAGCCCCCAGTGG –3') with *Hind*III site (underlined). The polymerase chain reaction (PCR) amplification of the PINIT and acidic domains coding regions were performed as follows: one cycle of denaturation (95°C for 30 seconds), 30 cycles of denaturation, annealing and extension (95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds), one cycle of final extension (72°C for 7 minutes) followed by a 4°C hold. The PCR–amplified fragments were purified from 0.8 % TBE agarose gel electrophoresis and ligated into the pGEM–T–Easy vector to generate pGEM–T–PINIT and pGEM–T–Acidic and transformed into *E. coli* JM109. The PINIT and acidic domain–encoding *Nde*I–*Hind*III fragment was restricted from pGEM–T and purified from 0.8 % TBE agarose gel electrophoresis and ligated downstream of Flag–tag coding sequence of *Nde*I–*Hind*III restricted p513–flag expression vector to generate p513–flag–PINIT and p513–flag–acidic. The restricted p513–flag vector was also purified from 0.8 % TBE agarose gel electrophoresis. The constructs were confirmed by restriction analysis and automated DNA sequencing (Rhodes University DNA sequencing unit).

3.2.2 Mutagenesis of PIAS3

P513–flag–PIAS3 plasmid was used as a template for site–directed mutagenesis using the double stranded whole plasmid linear non–PCR amplification procedure (according to the QuikChange mutagenesis kit; Stratagene). Complementary mutagenesis primers were designed for the introduction of single point mutations at L97A, R99N and R99Q (as previously described in Chapter 2) into the PINIT domain coding sequence (**Appendix G2**). Each mutagenesis reaction was comprised of 100 ng of p513–flag–PIAS3 parental plasmid template, 2.5 µl of dimethyl sulfoxide (DMSO), 10 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP mix , 125 ng of the forward primer, 125 ng of the reverse primer, 5 µl of 10x *Pfu* DNA polymerase buffer (200mM Tris–HCl [pH 8.8], 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$), 1 U of *Pfu* DNA Polymerase and sterile double distilled water to a final volume of 50 µl. Thermal cycling was allowed to proceed as follows: one cycle of denaturation (95°C for 30 seconds), 18 cycles of denaturation, annealing and extension (95°C for 30 seconds, 52°C for 60 seconds, 68°C for 5 minutes), one cycle of final extension (68°C for 7 minutes) and a 4°C hold at the end. Digestion of the parental p513–flg–PIAS3 plasmid in the amplification product was achieved by the addition of 5U of *Dpn*I restriction endonuclease to the reaction mixture and incubation at 37°C overnight. Pre– and post–*Dpn*I samples were

analysed by 0.8% TBE agarose gel electrophoresis (**Appendix D4**). An aliquot of 10 µl of post–*Dpn*I samples was transformed into *E. coli* JM109 supercompetent cells (Promega) for screening purposes. Plasmid DNA was isolated from the resulting colonies and screened for the desired mutation by automated DNA sequencing (**Appendix D7**) using the designed forward and reverse sequencing primer set (**Appendix H3**). The p513–flag–PIAS3 plasmid was used to generate the p513–flag empty vector by engineering an *Nde*I site and a STOP codon upstream of *Hind*III site by the following primers: forward primers (5'–CCATTTCCTTGGAC**CATATG***TAA*GCTTCCTAGGTC –3') with *Nde*I site (bold and underlined) and the STOP codon (italics) and the reverse primer (5'–GACCTAGGAAGCTTA**CATATG**GTCCAAGGAAATGG –3') with *Nde*I site (bold and underlined). The PIAS3 coding region was removed by restriction with *Nde*I enzyme and the p513–flag vector was purified from 0.8 % TBE agarose gel electrophoresis and re–ligated.

3.2.3 Cell culture and maintenance

Cell lines used in this study were MCF7 human breast cancer cell line, NIH3T3 immortalised mouse fibroblast cell line and HeLa cervical cancer cell line. MCF7 cell line is a breast epithelial adenocarcinoma cell line derived from a metastatic pleural effusion in a 69 year old Caucasian adult female (Soule *et al.*, 1973). NIH3T3 cell line was originally established from the primary mouse embryonic fibroblast cells (Todaro and Green, 1963). HeLa cells are human epithelial cells from a fatal cervical carcinoma. The cell line was derived from cervical cancer cells taken from Henrietta Lacks 1951. All cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 5 % heat inactivated foetal calf serum (FCS) 100 $U.ml^{-1}$, penicillin–streptomycin and 2 mM L–Glutamine, and were incubated in a humidified atmosphere at 37°C with 10.0 % (v/v) CO_2 in air. Cell were passaged at a ratio 1:3 every 3 to 4 days through trypsinisation by aspirating the medium prior to addition of 300 µl 1x Trypsin/EDTA (ethylenediaminetetraacetic acid) followed by wash with 1X PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM $K_2H_2PO_4$, pH 7.4) and incubating the cells at 37^oC at 10% CO_2 until cells lifted. Cells were seeded in fresh culture dishes containing fresh culture medium. The DMEM with supplements described above medium was changed every second day. Cell viability was assessed by trypan blue (Sigma) dye exclusion using a hemocytometer and observation by inverted light microscopy.

3.2.4 Transient transfection

Endotoxin–free plasmids encoding for PIAS3, PINIT & acidic domains and mutant PIAS3 proteins were isolated using GenElute^{TM} endotoxin–free plasmid midiprep kit (Sigma–Aldrich) (**Appendix D2**) before transfection of HeLa cells. HeLa cells were plated onto glass–bottomed culture well plate. The next day, the media was aspirated and cells washed once with Opti–MEM media (invitrogen). Transfection mixes contained 250 µl/well of serum–free drug–free Opti–MEM and 4 µg/well of each plasmid, and 10 µl/well lipofectamine–2000 were mixed and allowed to form DNA–lipofectamine complexes for 20 minutes at room temperature. Transfection mixtures were added to washed cells in dishes and incubated for 5 h at 37 °C, 10 % CO_2. After incubation, 1 ml of DMEM supplemented with 2 % FCS antibiotics–free was added to each well and after 12 hours the medium was replaced with a complete media with antibiotics.

3.2.5 Preparation of cell lysates

Protein lysates were prepared 48 h post–transfection by washing cells twice in ice–cold phosphate–buffered saline. Cells were trypsinized by 250 µl 1x Trypsin/EDTA in calcium and magnesium–free phosphate buffered saline (PBS), and incubated for 3–5 min at room temperature. The trypsinized cells were resuspended in equal volume of DMEM containing 10% (v/v) FCS to stop the trypsin reaction. The cells were washed in PBS buffer prior to resuspension in RIPA lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1% Triton X–100, 0.1% SDS, 0.8% deoxycholic acid, 10% glycerol, 1 mM EDTA, 1 mM PMSF, 5 µg/ml leupeptin–pepstatin–aprotinin, 0.15 mM $NaVO_3$, and 1 mM DTT) and incubated on ice for 15 minutes. Lysates were clarified at 10,000 × *g* for 5 minutes at 4 °C and the soluble protein lysate concentration quantified by NanoDrop 2000 (Thermo scientific).

3.2.6 Detection of the expressed protein by immunofluorecence staining

Transfected cells grown on a glass cover slip in a 24 well plate were starved for 12 hours in serum free media prior to 10 minutes and 30 minutes stimulation with 20 ng/ml IL–6 (Sigma Aldrich) before washing with PBS. Cells were fixed by incubation with PBS containing 4% paraformaldehyde and 4 % sucrose for 15 minutes at room temperature. The cells were then washed twice with PBS for 5 minutes and permeabilized by incubation with 0.25 % Triton^{TM} X–100 in PBS for 5 minutes followed by two washes with PBS for 5 minutes. The cells were blocked using 5% BSA in PBS for 30 minutes at 37^{o}C followed by incubation with mouse anti–FLAG M2 monoclonal antibody (Sigma Aldrich) and rabbit polyclonal anti-STAT3 IgG

antibodies (Santa Cruz biotechnology) in 3 % BSA in PBS at a dilution of 1:500 for 2 hours at 37^{o}C. The cells were washed twice with PBS for 5 minutes followed by incubation in FITC–labeled secondary antibodies, Alexa Fluor^{R} 488 donkey anti–mouse IgG (Invitrogen, USA) and Alexa Fluor^{R} 546 chicken anti–rabbit IgG (Invitrogen, USA) at a 1: 1000 dilution in 3 % BSA in PBS for 45 minutes at 37^{o}C. The cells were washed twice with PBS followed by nuclear counterstaining with Hoescht at 1:1000 dilution. The moisture was allowed to evaporate in a dark cupboard before the glass coverslip was mounted with cells side down on glass slides using a small drop of DEKO mounting medium (Invitrogen). Immunofluorescence images were captured using a confocal laser–scanning Zeiss LSM 510 microscope with 40X oil objective. Images were recorded and processed digitally with Zeiss LSM Image Browser software (Carl Zeiss Gmb H Jena). The excitation wavelength for Alexa Fluor^{R} 488 was 488 nm, and the emission was captured at 500–550 nm. The excitation wavelength for Alexa Fluor^{R} 546 was 546 nm and the emission was captured at 560–600 nm.

3.2.7 Detection of proteins by western blot analysis

Protein lysates were prepared as described above (section 3.3.5) and protein concentrations were quantified by NanoDrop 2000 (Thermo Scientific). Approximately 15 µg of protein from each sample was separated on 11 % SDS–PAGE, electro blotted onto a Hybond^{TM}–C extra nitrocellulose membrane (Amersham USA). The membrane was blocked for 1 h at room temperature using 5.0 % (w/v) non–fat powder milk in Tris–buffered saline–Tween (TBST); (50 mM Tris, 150 mM NaCl, 0.1% (v/v) Tween 20, pH 7.6); the membrane was incubated overnight at 4^{o}C on a rocking platform in 5.0 % (w/v) non fat powder milk in TBST, pH 7.6 containing specific primary antibodies; washed once with TBST followed by incubation for 1 h at room temperature on a rocking platform in 5.0 % (w/v) non fat powder milk in TBST, pH 7.6 containing specific secondary antibodies. The membrane was washed for 15 minutes three times with TBST before chemiluminescence–based protein detection. Chemiluminescence–based protein detection was achieved using the ECL^{TM} western blotting kit (GE Healthcare) as per the manufacturer's instructions, and captured with a Chemidoc chemiluminescence imaging system (BioRad, UK). Specific antibodies and the dilutions were as follows: 1:1000 for mouse anti–FLAG M2 monoclonal antibodies, 1:500 for mouse monoclonal anti-STAT3 IgG antibodies, 1: 1000 for mouse monoclonal anti-PIAS3 IgG antibodies (Santa cruz biotechnology) and Horseradish peroxidise conjugated (HRP) goat anti–mouse IgG (GE Healthcare) at 1:5000 dilution.

3.2.8 Quantitative co–localisation analysis

Immunofluoresence images captured by immunofluorescence microscopy were used to quantitatively analyse potential co–localisation of PIAS3 and STAT3. Images were processed using ImageJ 1.43 (McMaster Biophotonics Facility (MBF), McMaster University, Canada). The images were separated into component red and green channels. Background signal correction was performed by selecting areas of interest outside the cell. The red and green images were selected and used for intensity correlation analysis, and the colour scatter plots and the Pearson's correlation coefficient (MBF manual http://www.macbiophotonics.ca/imagej/) were automatically calculated.

3.3 RESULTS

3.3.1 Construction and verification of plasmids encoding for PIAS3, PINIT and the acidic domains of PIAS3.

The following expression plasmids: p513–Flag–PINIT, p513–Flag–acidic, p513–Flag–PIAS3 and p513–Flag were successfully constructed and confirmed by restriction endonuclease analysis (**Appendix B1**) and automated DNA sequencing. PIAS3 point mutation corresponding to L97A, R99N and R99Q were made on p513–Flag–PIAS3–L997, p513–Flag–R99N, and p513–Flag–PIAS3–R99Q plasmids respectively and verified by automated DNA sequencing.

3.3.2 Protein expression in Cell lines

HeLa cells express endogenous STAT3 and low but detectable levels of PIAS3, therefore endogenous STAT3 was stained for co–localisation studies. Western blot analysis was performed to confirm the expression of endogenous STAT3 and PIAS3 in HeLa, MCF7 and NIH3T3 cell lines. All cell lines expressed endogenous PIAS3 at low levels compared to STAT3 (**Figure 3.1**).

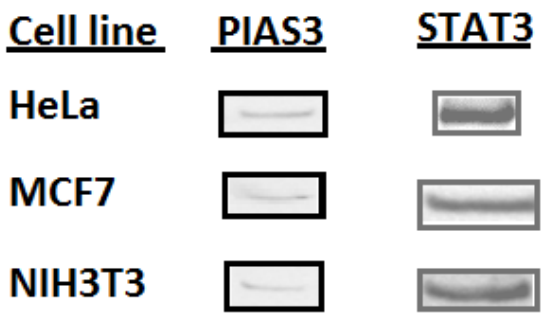

Figure 3.1 Expression of endogenous STAT3 and PIAS3 in various mammalian cell lines. Detection of STAT3 and PIAS3 in HeLa, MCF7 and NIH3T3 cells by western blot analysis. Equal amounts of 15 µg of soluble proteins were loaded in each well of 12 % SDS–PAGE and electroblotted to the membrane. Endogenous PIAS3 and STAT3 protein at 68 kDa and 85 kDa respectively were detected. The presence of STAT3 and PIAS3 were determined by western blot analysis using 3 mouse monoclonal anti-STAT3 IgG antibodies (Santa cruz biotechnology), mouse monoclonal anti-PIAS3 IgG and HRP–conjugated goat anti–mouse IgG (GE Healthcare) at 1:5000 dilution and a chemiluminescence based detection system. The signal was detected using a Chemidoc XR imaging system (BioRad, UK).

3.3.2.1 *In vivo* expression of Flag–tagged proteins

Endotoxin–free plasmids encoding for the PIAS3, PINIT and acidic proteins were successfully isolated and characterised. Human HeLa cervical cancer cells were successfully cultured and transfected. All analyses were done by transiently transfection of plasmids encoding for PIAS3, PINIT and acidic domains and the transfection efficiency of approximately 5 % was observed in all transfections. The plasmids that were transfected into the cells encoded the flag–tag, facilitating detection. The expression of Flag–PIAS3, Flag–PINIT and Flag–acidic proteins in the transfected HeLa cells were confirmed by western blot analysis of the soluble protein lysates of transfected HeLa cell lysates prepared 48 hours post transfection. Molecular mass species of 68 kDa, 23 kDa and 19 kDa corresponding to Flag–PIAS3, Flag–PINIT and Flag–acidic, respectively, were detected (data not shown).

3.3.3 Localisation of the PIAS3, PINIT domain and acidic domain in IL–6 stimulated and unstimulated HeLa cells

HeLa cells were successfully transfected with flag–tagged plasmid encoding for PIAS3, PINIT and acidic domains. The transfection efficiencies were estimated by immunofluorescence microscopy and approximately averaged at 5%. PIAS3 localisation was examined in the presence and absence of IL–6 stimulation. Flag–PIAS3 was found completely localised in the nucleus of all transfected HeLa cells (**Figure 3.2**). PIAS3 exhibits a similar pattern after stimulation with IL–6 for 10 and 30 minutes. PIAS3 displayed a speckled staining pattern of nuclear distribution but no definitive conclusions could be drawn

from this (**Figure 3.2**). Localisation of the PIAS3 protein in the nucleus was consistent with previous findings from various cell lines **Table 1.1** (Kotaja *et al*., 2002; Sonnenblick *et al*., 2004; Duval *et al*., 2003). However, in this study no response in PIAS3 was observed on treatment of transfected HeLa cells with IL–6, unlike IL-6 treatment on human pulmonary epithelial cell lines (A549 and H520) with EGF (Dabir *et al*., 2009).

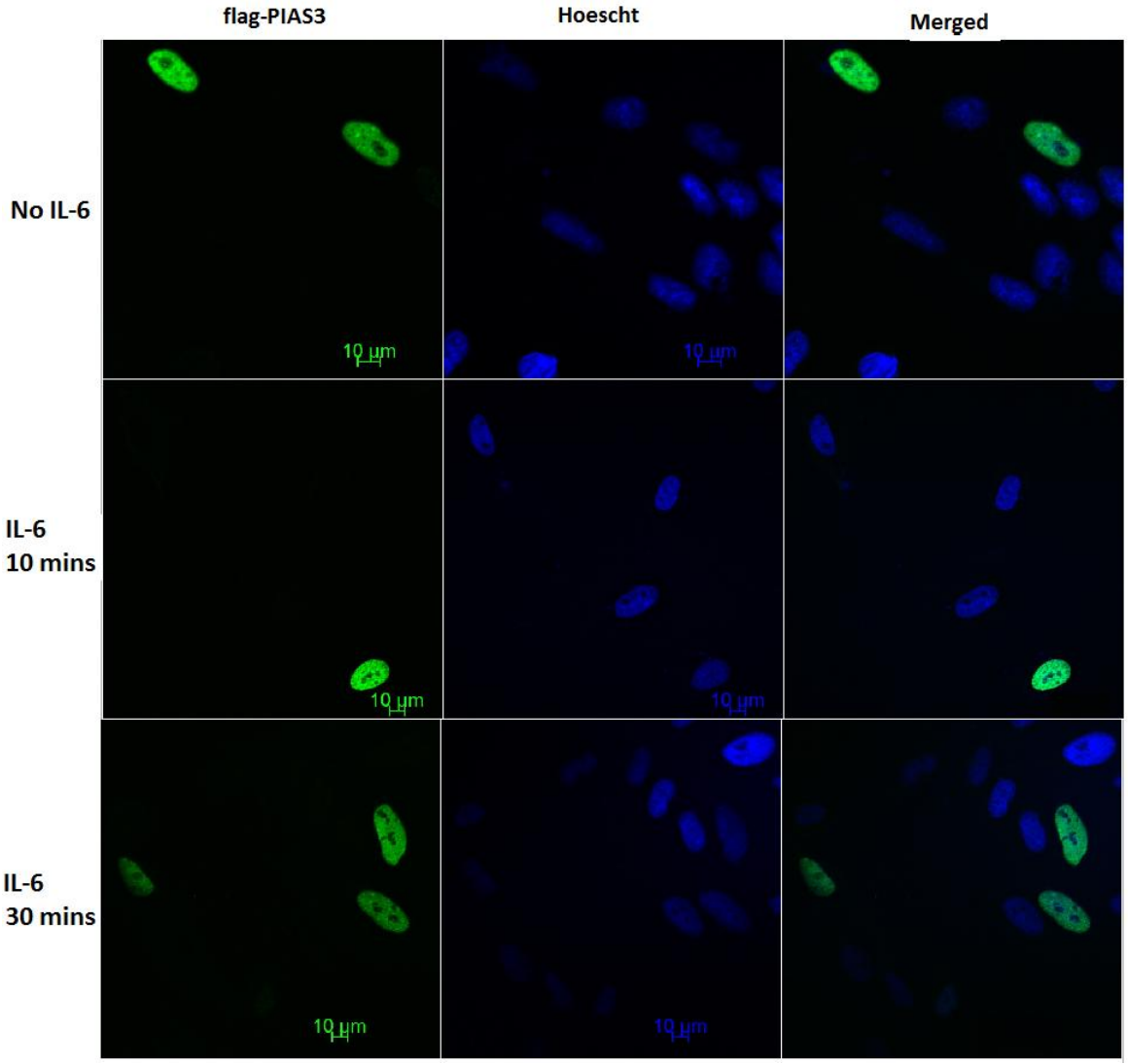

Figure 3.2 The subcellular localisation of PIAS3 in HeLa cells. HeLa cells were seeded in a glass bottomed culture dish for 24 hours and transfected with a p513–flag–PIAS3 plasmid encoding flag–PIAS3 protein. After 48 hours the cells were starved for 12 hours in serum free medium before being left unstimulated or stimulated with 20 ng/ml IL–6 for 10 minutes and 30 minutes. The cells were fixed and permiabilised as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the flag–PIAS3 image with nuclei hoescht staining image using Zeiss LSM Image Browser software (Carl Zeiss Gmb H Jena).

To examine if the nuclear retention "PINIT" motif may also act as a nuclear localisation signal, HeLa cells were transfected with a plasmid encoding the PINIT domain alone and treated under the same conditions as that of full length PIAS3. The PINIT domain was found to be predominantly localised at the periphery of the nucleus in most cells however, the PINIT protein was also observed in the cytoplasm in lower amount in all cells (**Figure 3.3**).

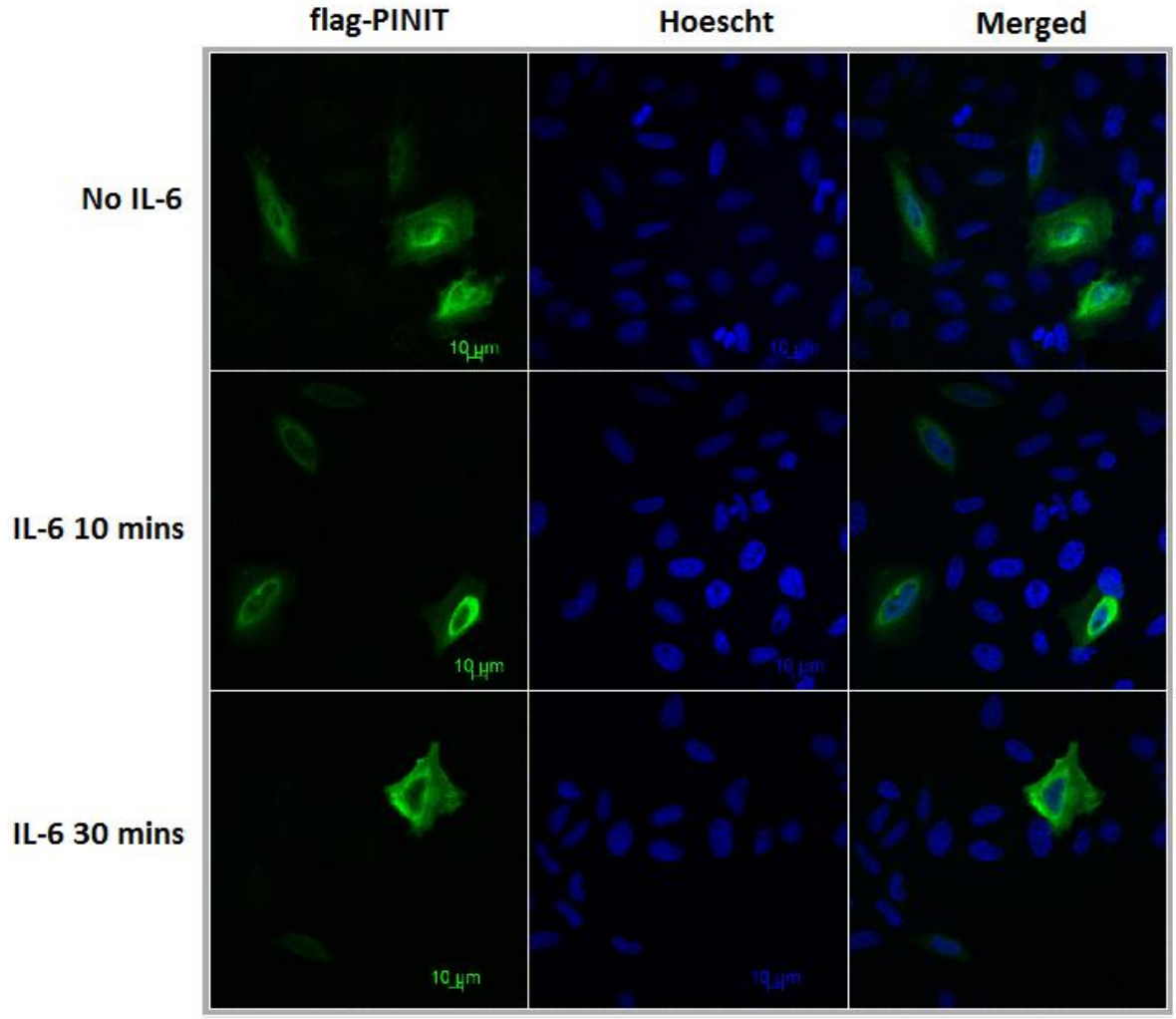

Figure 3.3 The subcellular localisation of PINIT domain of PIAS3 in HeLa cells. HeLa cells were seeded in a glass bottomed culture dish for 24 hours and transfected with a p513–flag–PINIT plasmid encoding flag–PINIT protein. After 48 hours the cells were starved for 12 hours in serum free medium before being left unstimulated or stimulated with 20 ng/ml IL–6 for 10 minutes and 30 minutes. The cells were fixed and permiabilised as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the flag–PINIT image with nuclei Hoescht staining image using Zeiss LSM Image Browser software (Carl Zeiss Gmb H Jena).

Upon stimulation with the IL–6, the PINIT domain showed a similar pattern. Unlike the full length PIAS3 protein, the PINIT domain, despite carrying the nuclear retention motif, does not appear to localise exclusively to the nucleus in all transfected cells as observed for full length PIAS3.Upon stimulation with IL–6 STAT3 was phosphorylated, and translocated into the nucleus potentially either associating with nuclear PIAS3 or forming complexes with cytoplasmic PIAS3 prior to nuclear translocation (Dabir *et al*., 2009). The distribution of the PINIT domain alone in HeLa cells appeared to exhibit nuclear and cytoplasmic distribution.
To examine its localisation in comparison to the PINIT domain, the flag–acidic domain was expressed in HeLa cells.

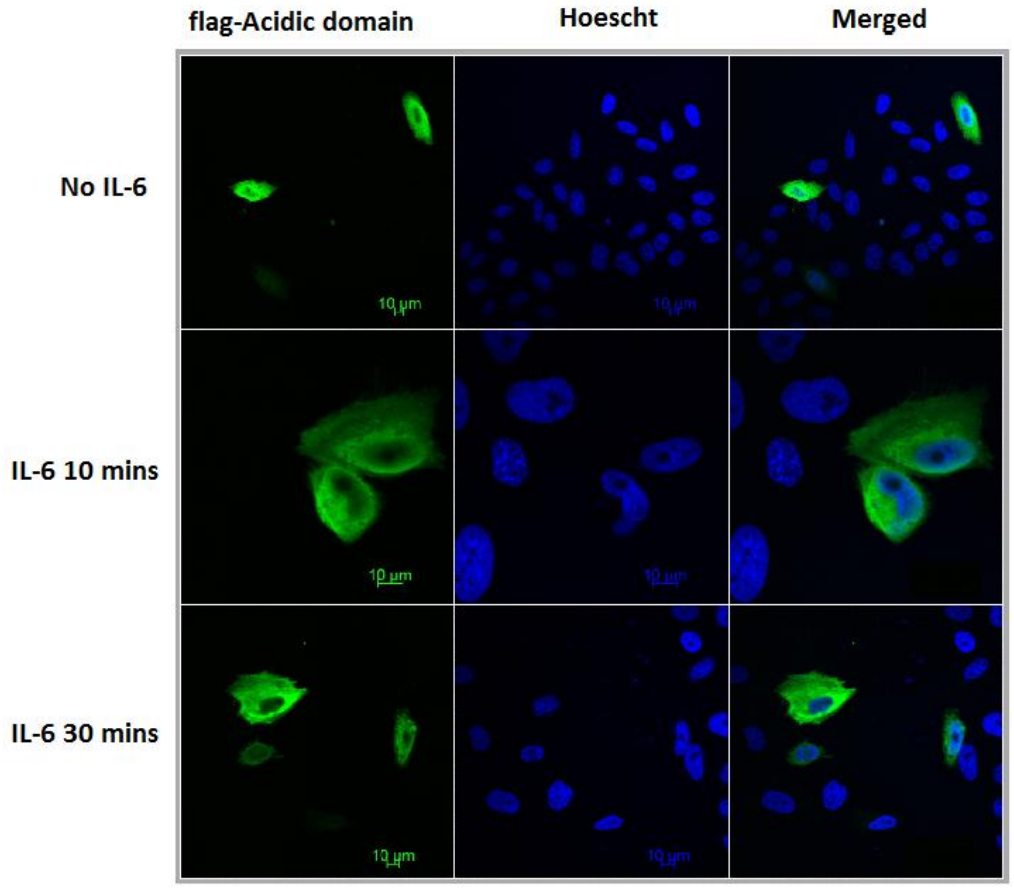

Figure 3.4 The subcellular localisation of the acidic domain of PIAS3 in HeLa cells. HeLa cells were seeded in a glass bottomed culture dish for 24 hours and transfected with a p513–flag–acidic plasmid encoding flag–acidic protein. After 48 hours the cells were starved for 12 hours in serum free medium before being left unstimulated or stimulated with 20 ng/ml IL–6 for 10 minutes and 30 minutes. The cells were fixed and permiabilised as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the flag–acidic image with nuclei Hoescht staining image using Zeiss LSM Image Browser software (Carl Zeiss Gmb H Jena).

The cellular localisation of the acidic domain showed diffuse cytoplasm distribution with apparent accumulation in the nuclear periphery in all the cells (**Figure 3.4**). The acidic and the PINIT domains exhibited similar localisation in terms of the perinuclear accumulation. However, the acidic domain showed more cytoplasmic localisation than PINIT domain while the PINIT domains showed more nuclear localisation than acidic domain. The acidic and PINIT domains had localisations that were different from the PIAS3 which showed a complete nuclear localisation.

3.3.4 Localisation of the mutant PIAS3 in IL–6 stimulated and unstimulated HeLa Cells

Conserved residues L97 and R99, potentially important for PINIT domain function in PIAS3–STAT3 association and predicted structural effects upon mutation (L97A, R99N and R99Q), were shown in chapter 2. To determine whether the mutations affected the localisation by structural disruption of the PINIT domain and *holo* PIAS3, the flag–PIAS3 mutants were expressed in HeLa cells. The expressed flag–PIAS3–L97A was found to predominately localise diffusely in the nucleus with punctate pattern in most of the transfected cells. However, unlike wildtype PIAS3, PIAS3–L97A exhibited traces in the cytoplasm in most cells (**Figure 3.5**). PIAS3–L97A showed a similar pattern after stimulation with IL–6 for 10 and 30 minutes. The L97A mutation reduced complete nuclear localisation as observed in wildtype PIAS3 (**Figure 3.2**).

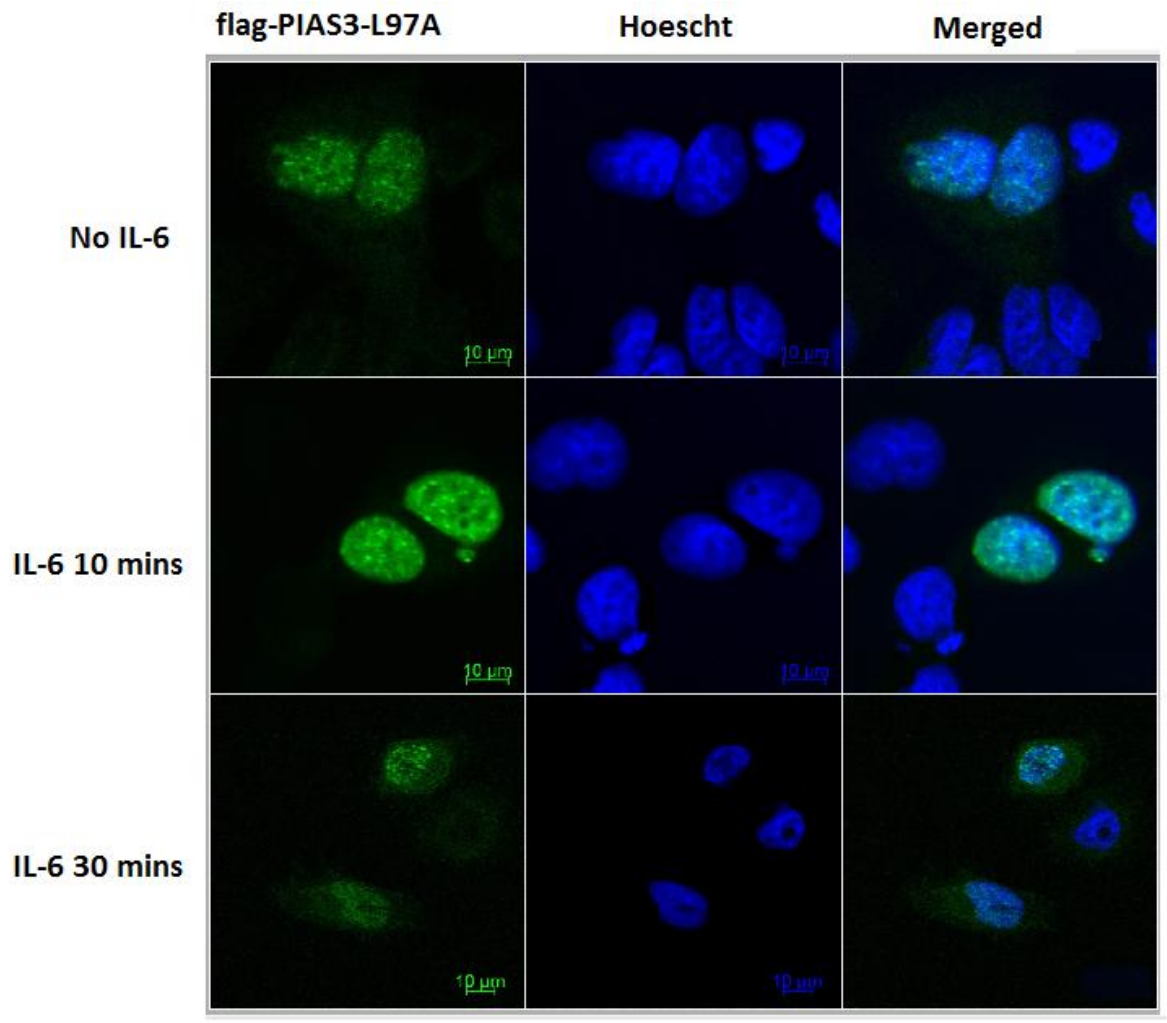

Figure 3.5 The subcellular localisation of PIAS3–L97A in HeLa cells. HeLa cells were seeded in a glass bottomed culture dish for 24 hours and transfected with a p513–flag–PIAS3–L97A plasmid encoding flag–PIAS3–L97A protein. After 48 hours the cells were starved for 12 hours in serum free medium before being left unstimulated or stimulated with 20 ng/ml IL–6 for 10 minutes and 30 minutes. The cells were fixed and permiabilised as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the flag–PIAS3–L97A image with nuclei hoescht staining image using Zeiss LSM Image Browser software (Carl Zeiss GmbH Jena).

To examine the effect of mutation of residue R99 of PIAS3, flag–PIAS3–R99N protein was expressed in HeLa cells under the same conditions as previously described. The flag–PIAS3–R99N showed a complete nuclear localisation like the wildtype PIAS3 (**Figure 3.6**). Unlike wildtype PIAS3 and PIAS3–L97A, PIAS3–R99N showed a diffuse nuclear localisation and no punctate pattern was observed in most cells. This observation suggests a possibility of destabilisation of PIAS3 function due to the mutation. The residue R99 was further analysed

by replacement with Q and flag–PIAS3–R99Q analysed in the same manner.

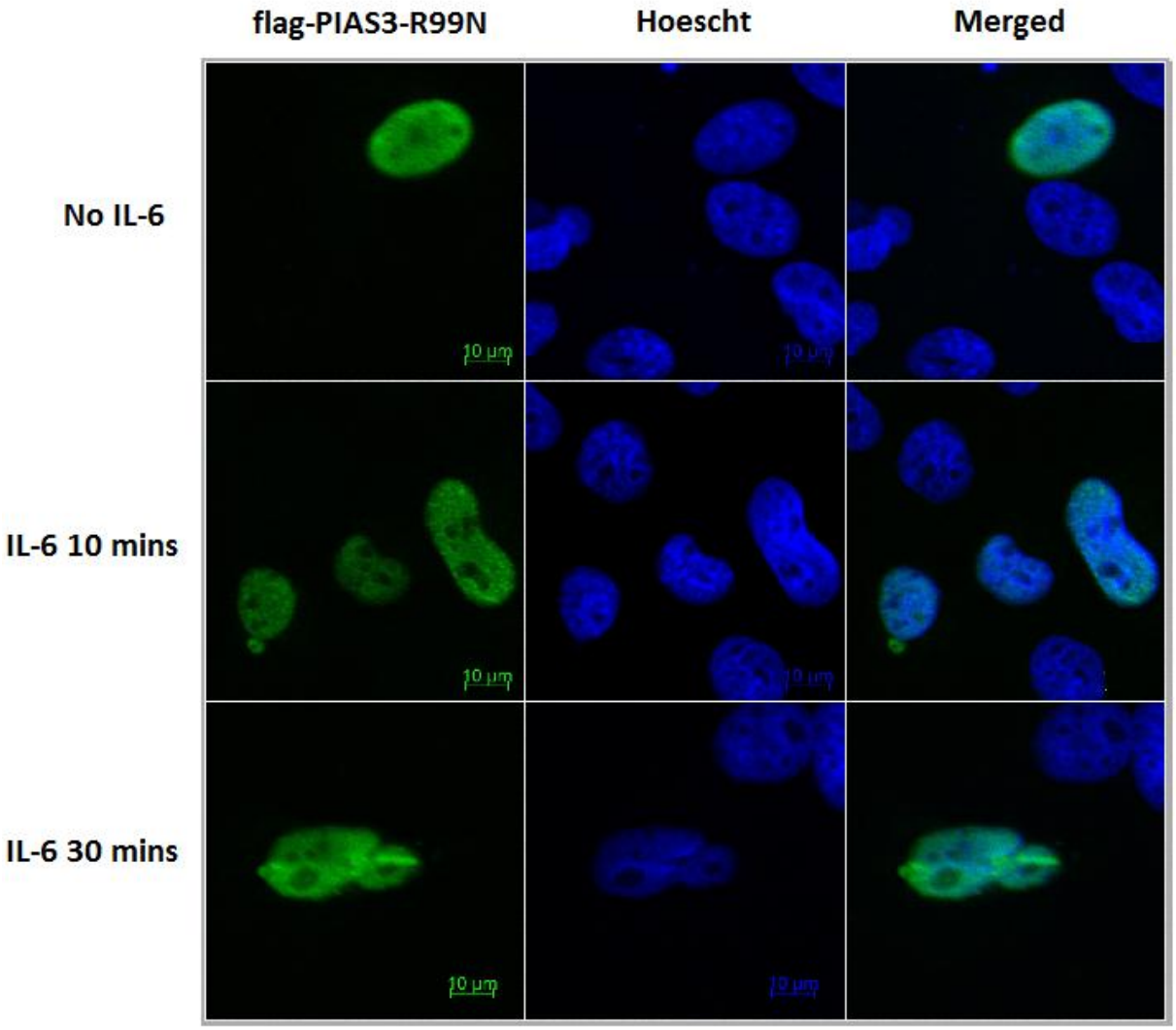

Figure 3.6 The subcellular localisation of PIAS3–R99N in HeLa cells. HeLa cells were seeded in a glass bottomed culture dish for 24 hours and transfected with a p513–flag–PIAS3–R99N encoding flag–PIAS3–R99N protein. After 48 hours the cells were starved for 12 hours in serum free medium before being left unstimulated or stimulated with 20 ng/ml IL–6 for 10 minutes and 30 minutes. The cells were fixed and permiabilised as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the flag–PIAS3–R99N image with nuclei Hoescht staining image using Zeiss LSM Image Browser software (Carl Zeiss GmbH Jena).

Flag–PIAS3–R99Q expressed in HeLa cells showed a diffuse nuclear distribution pattern which persisted under IL–6 stimulation (**Figure 3.7**). Traces of PIAS3–R99Q in the cytoplasm were observed on IL–6 stimulation in all of the transfected cells. Complete localisation of PIAS3–R99Q observed in resting cells was similar to the wildtype PIAS3. The

existence of cytoplasmic PIAS3–R99Q and the diffuse localisation after IL–6 stimulation in all cells may suggest destabilisation of PIAS3 function due to the mutation.

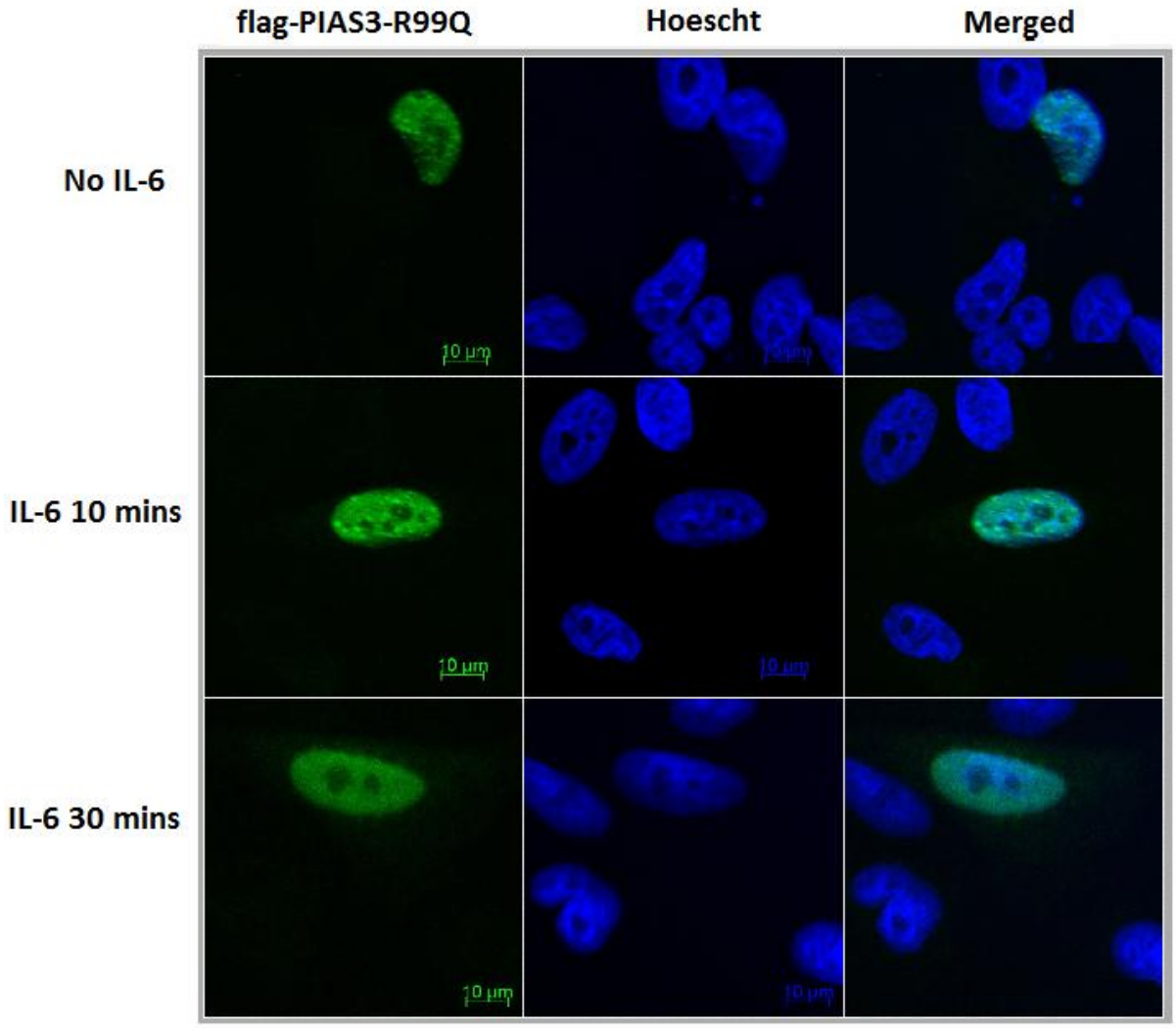

Figure 3.7 The subcellular localisation of PIAS3–R99Q in HeLa cells. HeLa cells were seeded in a glass bottomed culture dish for 24 hours and transfected with a p513–flag–PIAS3–R99Q encoding flag–PIAS3–R99Q protein. After 48 hours the cells were starved for 12 hours in serum free medium before being left unstimulated or stimulated with 20 ng/ml IL–6 for 10 minutes and 30 minutes. The cells were fixed and permiabilised as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the flag–PIAS3–R99Q image with nuclei Hoescht staining image using Zeiss LSM Image Browser software (Carl Zeiss GmbH Jena).

The differences observed in localisation of mutants PIAS3 compared to the wildtype PIAS3 suggests that a mutation of L97 and R99 causes a subtle shift in PIAS3 localisation. PIAS3–L97A and PIAS399Q showed a slightly more cytoplasmic localisation compared to wildtype PIAS3.

3.3.5 Co–localisation of PIAS3, PINIT domain and acidic domain with STAT3

STAT3 was predominantly localised in the nucleus in all cells and was less distributed in the cytoplasm in all cells. Also punctate structures of STAT3 localisation in the cytoplasm were observed (**Figure 3.8B).** PIAS3 displayed speckled pattern and diffuse distribution in the nucleus but completely localised in the nucleus unlike STAT3. The distribution of PINIT domain was observed predominantly in the perinucleus and nucleus in most cells, but also less distributed in cytoplasm in all cells. The acidic domain was distributed predominantly in the perinucleus, with some traces in the cytoplasm in all cells. There was more cytoplasmic localisation of acidic domain compared to PINIT domain in most cells.

Co–localisation of PIAS3, PINIT and acidic with STAT3 was investigated in transfected HeLa cells 30 minutes post IL–6 stimulation using immunofluroscence staining as described in section 3.2.3. STAT3 and PIAS3 showed a different distribution; however, the PIAS3–STAT3 merged picture (**Figure 3.8A**) showed potential co–localisation of PIAS3 and STAT3 in the nucleus. Co–localisation of PINIT and STAT3, merged red (STAT3) and green (PINIT) image observed on yellow regions of the image (**Figure 3.8B**) showed a high degree of co–localisation in the nucleus. The acidic domain was observed co–localised with STAT3 in cytoplasm and perinucleus region indicated by yellow region of the merged picture (**Figure 3.8C**).

Also observed was the change of cell morphology with time after transfection (data not shown). Comparing cells in **Figure 3.8 (A–C)** , it was observed that 48 hours post transfection cells expressing PIAS3, PINIT and acidic domain displayed different morphology which probably suggested that expressed proteins selectively affects the growth and survival of cells. In particular the PINIT domain and the acidic domain transfected cells displayed similar morphology but as a whole different from the PIAS3 transfected cells.

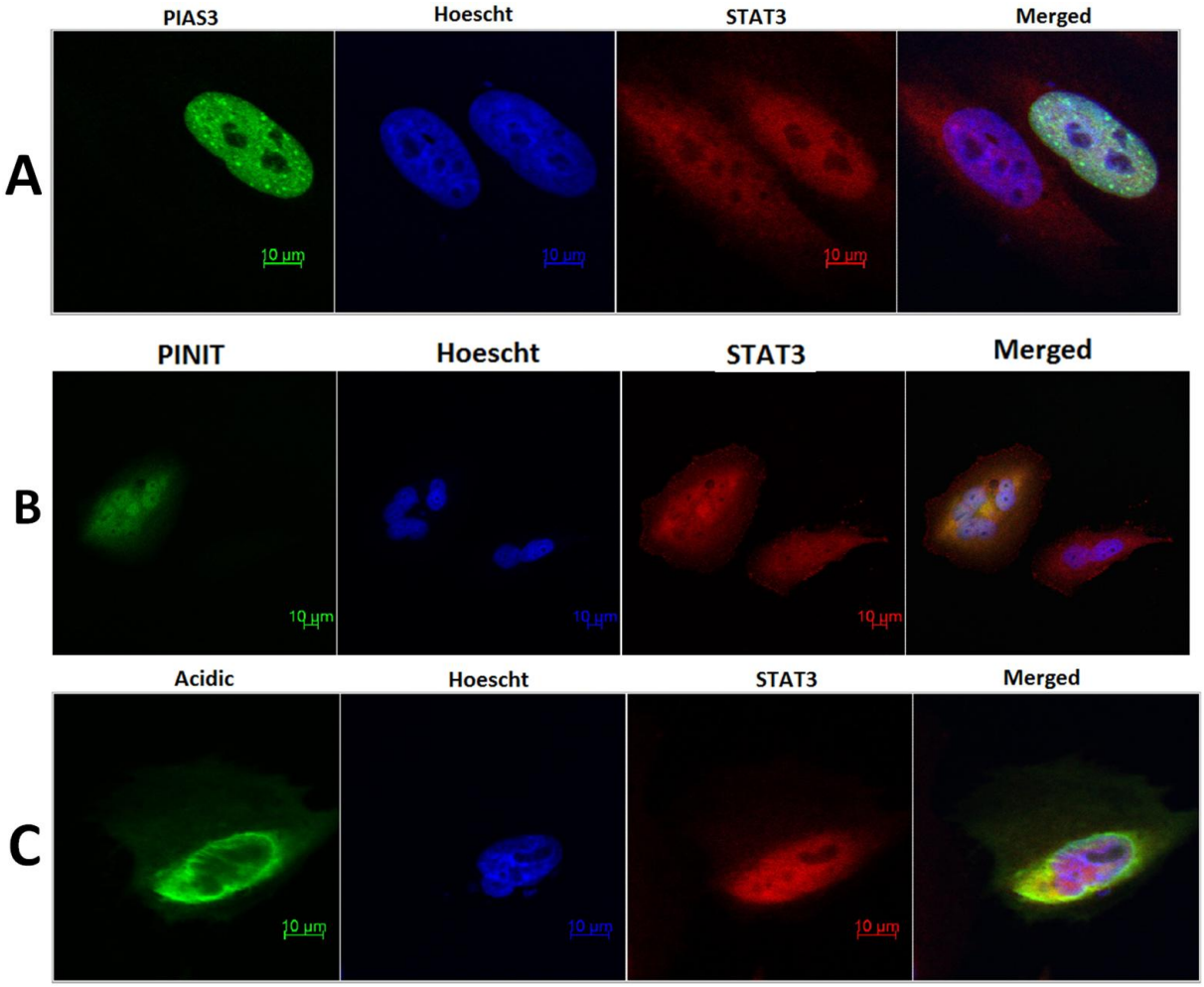

Figure 3.8 Co–localisation analysis of PIAS3, PINIT and acidic domain with STAT3 in HeLa cells. A) HeLa cells grown on glass cover slips were transfected for 48 hours with a p513–flag–PIAS3 plasmid. **B**) Hela cells grown on glass cover slips were transiently transfceted for 48 hours with p513–flag–PINIT plasmid. **C**) HeLa cells grown on cover slips were transiently transfected for 48 hours with p513–flag–acidic. Cells were starved for 12 hours and stimulated for 30 minutes with IL–6 before fixing as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The endogenous STAT3 was stained with rabbit polyclonal anti-STAT3 IgG antibodies and followed by one hour incubation Alexa Fluor 546 chicken anti–rabbit IgG. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the green image with nuclei Hoescht staining image and STAT3 (red) image using Zeiss LSM Image Browser software (Carl Zeiss GmbH Jena).

The degree of co–localisation of PIAS3 and its domains with STAT3 in HeLa cells was quantitatively assessed using Pearson's correlation coefficient (PCC). This is the relationship between the red and green pixels in an image and was displayed as an intensity–scatter plot where the red image component is represented along the x–axis and green image component along the y–axis (**Appendix B3**). The co–localization results were shown in a pixel

distribution along a linear correlation line and the deviation from the linear distribution is quantified by the Pearson's correlation coefficient (PCC). The PCC defines the quality of the linear relationship between two signals as the red and green channel intensity distributions are linked. For confocal images PCC value close to 1 indicate reliable co–localisation. Quantitative analysis of co–localisation showed a lower degree of co–localisation of acidic domain (PCC of 0.574) with STAT3 in comparison with PINIT (PCC of 0.732) and PIAS3 (PCC of 0.638).

3.4 DISCUSSION AND CONCLUSION

The localisation studies showed that the PIAS3 protein completely localised in the nucleus in all cells under both unstimulated and IL–6 stimulated conditions. PIAS3 exhibited nuclear punctate pattern in all cells. The PINIT domains alone showed a predominant localisation in the perinucleus and nucleus in most cells. However, cytoplasmic PINIT protein was also observed in all cells. The acidic domain alone was observed predominantly in the perinucleus in all cells; nevertheless more cytoplasmic acidic domain was observed compared to the PINIT domain. The PINIT and acidic domains exhibit cytoplamic distribution compared to wildtype PIAS3 which had complete nuclear localisation. The findings on PIAS3 localisation in this study were consistent with previous researchers (Kotaja *et al*., 2002; Sonnenblick *et al*., 2004; Duval *et al*., 2003; Man *et al*., 2006; Yamashina *et al*., 2006; Peng *et al*., 2010). However, previous studies by Dabir *et al*. (2009) showed cytoplasm localisation of endogenous PIAS3 which translocation into the nucleus upon EGF stimulation. Also, cytoplasmic localisation of endogenous PIAS3 was observed and its nucleus translocation upon proestrogene stimulation (Man *et al*., 2006).

Differences in findings by previous researchers on PIAS3 localisation may be due to influences by different cell line. HeLa cells constitutively express IL–6 (Eustace *et al*., 1993) and IL–6 acts in an autocrine manner (Eustace *et al*., 1993). Also, it was observed that exogenous IL–6 did not activate STAT3 in cervical carcinoma cells (Hess *et al*., 2000). The non effect on IL–6 stimulation on HeLa cells observed on PIAS3, PINIT and acidic domains might be either due to stimulated pathway by autocrine system or a specific defect in the signalling pathway. It was also observed that glycoprotein 80 (gp80) was not detected on the surface of carcinoma cell lines that include HeLa cell line (Hess *et al*., 2000), but shedded sgp80 was detected, hence should be able to bind IL–6 and complete IL–6R–signaling

complex. This explains the weak response of HeLa cells to exogenous IL–6 (Hess *et al*., 2000).

The nuclear retention of the PIAS3 was suggested to be a result of the "PINIT" motif (Duval *et al*., 2003). The *in vivo* expression of the PINIT domain showed that the PINIT is diffusely localised in the perinucleus and cytoplasm unlike the wildtype PIAS3 despite both carrying the "PINIT" motif. These data suggest that the "PINIT" motif may not be critical for nuclear retention as suggested by previous studies conducted by deleting the "PINIT" motif (Duval *et al*., 2003). However, the delocalisation observed by Duval *et al*. (2003) may have been a result of disruption of the PIAS3 structure due to the deletion mutation of the "PINIT" motif. The acidic domain showed a diffuse localisation in the cytoplasm unlike the wildtype PIAS3 which was completely localised in the nucleus.

All the PIAS3 mutants exhibited a slight difference on the localisation pattern compared to the wildtype PIAS3. Unlike the wildtype PIAS3, the existence of the cytoplasmic PIAS3–L97A and PIAS3–R99Q, and the diffuse nuclear localisation of PIAS3–R99N suggested a mutational effect on PIAS3 localisation. Residue R99 was predicted *in silico* to be surface exposed and has been predicted to be strongly involved in the PIAS–STAT3 interaction. Therefore abrogation of the punctate structures on PIAS3–R99N and PIAS3–R99Q suggest a functional purpose of the residue. Furthermore, diffusely localised PIAS3–L97A in the cytoplasm suggests an effect of the mutation. In chapter 2, this residue was shown buried in the hydrophobic core, although it was not predicted to be directly involved in the PIAS3–STAT3 interaction. Replacement of this residue with a less bulk residues may have affected the neighbouring residues as described in chapter 2. Similar punctate granule pattern in PIAS3 localisation were observed in PIAS3–L97A, unlike wildtype PIAS3, PIAS3–L97A was diffusely localised in the cytoplasm.

Despite the minor discrepancies described above, no major delocalisation effects were observed for all the PIAS3 mutants. However, observed traces of diffuse cytoplasmic staining of the PIAS3–L97A and PIAS3–R99Q could possible be due to abrogation of mutant PIAS3–STAT3 interaction. Therefore, it is necessary for detailed *in vitro* analysis of the PIAS3–STAT3 interaction with major focus on the region carrying the mutated residues in particular the PINIT domain. Also, further analysis (e.g. immunoprecipitation) of the mutant ability to interact with STAT3 *in vivo* is required.

Co–localisation of the PIAS3, PINIT and acidic domains with STAT3 was qualitatively and quantitatively analysed. The qualitative assessment was performed by overlaying the flag–

tagged image (green) and the STAT3 (red) image and the yellow regions shows a complete co–localisation of the red and green pixels. The quantitative assessment was performed by plotting the flag–tagged (green, y–axis) and the STAT3 (red, x–axis) image pixels and calculation of the Pearson's correlation coefficient to determine the extent of co–localisation (**Appendix B3**). The localisation of endogenous STAT3 was observed by staining using STAT3 rabbit polyclonal IgG antibodies that target both the phosphorylated and unphosphorylated forms of STAT3. The observed predominantly nuclear localised STAT3 was most probable the phosphorylated STAT3. Therefore, staining with anti–phospho–STAT3 rabbit monoclonal antibodies could validate these findings.

Qualitative and quantitative analysis of co–localistion revealed a high degree of PINIT domains co–localisation with STAT3 and these data suggest a possibility of interaction with STAT3. This domain was of interest in this study because it was previously suggested to interact with STAT3 (Chung *et al*., 1997; Borghouts *et al*., 2010). Furthermore, the predominance of STAT3 nuclear localisation and the perinuclear and nuclear localisation of the PINIT domain in all transfected cells suggested the necessity to further investigate the nature of PINIT–STAT3 interaction. The difference in morphological changes in cells expressing PIAS3, PINIT and acidic domain suggested a selective effect of the protein in cells. Similar morphologic changes were observed in late apoptotic murine high–grade glioblastma cells (Tu–9648, Tu–2449) transfected by recombinant purified PIAS3 (400–523) peptide (Borghouts *et al*., 2010). The apoptotic effect of PIAS3 (82–132) and PIAS3 (82–104) peptides was observed in RBL–2H3 cell lines (Levy *et al*., 2006; Yagil *et al*., 2009), and these peptide regions are encompassed in the PINIT domain. All these findings suggest further investigation of PINIT–STAT3 interaction using *in vitro* assays.

The study concludes that the PIAS3 completely localised in the nucleus and mutation of residues R99 and L97 did not affect the nuclear retention signal. The PINIT domain, despite carrying the "PINIT" motif, was not completely localised in the nucleus, but had more co–localisation with the STAT3 compared to full length PIAS3 and acidic domain. *In vitro* characterisation of the PINIT domain and the mutants will reveal further information. The investigation of the PINIT domain and its mutants using recombinant proteins by employing biophysical techniques will reveal more information about the PINIT domain and its binding affinity. Also, mutant PINIT–STAT3 direct interaction studies will reveal the extent of the effect of residues R99 and L97 predicted *in silico*.

CHAPTER 4

CLONING, EXPRESSION AND PURIFICATION OF $(HIS)_7$–PINIT PROTEIN

PIAS proteins located in distinct chromosomes in the human genome are negative regulators of cytokine–mediated gene activation pathways. The PINIT domain of the PIAS3 has been suggested to be the determinant domain for PIAS3–STAT3 interaction. Here, molecular techniques were employed to design expression constructs encoding the PINIT domain of PIAS3 for heterologous expression and purification of recombinant PINIT protein. Using size exclusion chromatography, Fourier Transformed Infrared Spectroscopy and simple immunological assays, recombinant proteins were characterised both in terms of structural integrity and functional biological activity.

4.1 INTRODUCTION

Although PIAS3 protein is a multidomain protein, only the PINIT and acidic domains were suggested to interact with STAT3 (Chung *et al*., 1997; Levy *et al*., 2006; Yagil *et al*., 2009; Borghouts *et al*., 2010). *In silico* and *in vivo* analysis of the PIAS3, PINIT and acidic domains showed that findings about the PINIT domain highlighted the need for its further investigation using an *in vitro* approach.

The objective of recombinant PINIT protein expression is to produce a protein with detectable biological activity. In most cases the desired activity is supported by a discrete domain and it is often not necessary to express the full–length protein to address particular biological questions. For successful expression of a fragment of the *holo*–protein, the choice of domain boundary is critical. Small structural differences often have great influence in the expression and solubility. Therefore, prior knowledge of protein structure, sequence conservation pattern and prediction of the secondary structures or unfolded/ disordered regions must be considered (Ginalski *et al.*, 2003; Ward *et al.*, 2004). The use of secondary structure prediction software can avoid disruption of the predicted secondary structural elements and borders should be engineered to encompass these secondary structure features (Yang *et al.*, 2005). In addition, it has been advised to avoid inclusion of low complexity regions or hydrophobic residues at the termini (Yang *et al.*, 2005).

Practical key points to consider are: expression strain; expression vector plasmid; size of the fragment to express; the affinity tag; and the purification strategy to use. Among the many systems available for heterologous protein expression, the gram-negative bacterium *Escherichia coli* remains one of the most attractive because of its ability to grow rapidly to high densities on inexpensive substrates. Methodical solubility studies with variable induction and growth strategies need to be employed to yield soluble and correctly folded protein. The use of affinity tags often improves the expression and purification of high quality proteins. Some tags potentially interfere with protein folding and function. Certain affinity tags have a beneficial effect on protein solubility especially in bacterial protein expression (Kapust *et al.*, 1999; Chong, 2001). Small tags, such as His–tag (Hochuli *et al.*, 1988), bear a smaller risk of steric interference than larger tags such as glutathione–S–transferase (GST) (Chong, 2001) or maltose binding protein (MBP) (Terpe, 2003). In comparison to His–tag, GST have combinations of disadvantages that includes its homo–dimeric nature which affect the protein solubility and folding (Kaplan *et al.*1997). Furthermore, the solvent exposed cysteine residues of GST can lead to oxidative aggregation

and expenditure of more metabolic energy during expression compared to His–tag (Kaplan *et al.*, 1997). However, successful expression of a protein is by performing empirical trials through altering expression conditions such as temperature, inducer concentration and observing the solubility and stabilities of the recombinant proteins that are obtained (Riggs *et al.*, 1994).

The Fourier transform infrared spectroscopy (FTIR) technique is applied to structurally characterise the protein in aqueous and non-aqueous environments. FTIR provides information about the secondary structure content of protein and indicates the conformation of the protein (Lamba *et al*., 1993; Haris and Severcan, 1999). Also, circular dichroism spectroscopy is used to gain information about the secondary structure of proteins and polypeptides in solution (Alder *et al*., 1973). The advantages of both techniques are that they use very little sample (200 µl of 0.5 mg/ml solution in standard cells) and are non–destructive. The disadvantages of both techniques are the interference with solvent absorption in the UV region. However, this can be digitally subtracted in FTIR technique. Furthermore, the disadvantage of CD is that it is operated below 200 nm wavelength where only very dilute and non–absorbing buffers allow measurements.

Here, the successful cloning, heterologous expression and purification of functional, folded recombinant PINIT domain is described. In addition, the PINIT domain protein was characterised by size exclusion chromatography and FTIR and its preliminary functional analysis were performed by dot blot association assay.

4.2 METHODS

4.2.1 Materials

Specialised materials and reagents utilised are listed in the **Appendix E.** Reagents were obtained from Sigma Chemicals (USA), Roche Molecular Biochemical (USA) and Merck Chemicals (Germany). Optimized oligonucleotides were synthesized by IDT Oligo (USA) and distributed by Whitehead Scientific (South Africa). Nickel–chelating Sepharose Fast Flow matrix was obtained from Pharmacia Biotech (Sweden). The p513–Flag–PIAS3–WT plasmid was a kind donation by Dr Hélène Boeuf (Université de Bordeaux, France) (Duval *et al*., 2003), and the pET32b–STAT3b–tc plasmid was a kind donation by Dr Christoph Müller (EMBL, Germany) (Becker *et al*., 1998). The pQE2 and pQE60 plasmids were purchased from Qiagen (USA) and the pGEM–T–EASY vector and pGEX–4T–1 were purchased from

Promega (USA). Bacterial expression strains *E.coli* XL1 Blue, *E. coli* BL 21(DE3) and JM109 were purchased from Novagen (USA). Rosseta and *E. coli* M15[pREP4] were purchased from Qiagen (USA). The Hybond^{TM}–C Extra nitrocellulose membranes, and the size exclusion column, Superdex 200 HR 10/30 were purchased from Amersham–Phamarcia Biotech, (USA). The mouse monoclonal anti–His primary antibody, horseradish peroxidase (HRP)–conjugated goat anti–mouse IgG antibody, mouse Anti–Glutathione–S–transferase (GST) monoclonal antibody and the Enhanced Chemiluminescence (ECL) kit were purchased from GE Healthcare (UK). Amicon^{R} Ultra Ultracel^{R} (10K) Centrifugal Filters were purchased from Millipore (Ireland). HisTrap columns were purchased from GE Healthcare (UK).

4.2.2 Construction of double tagged pGEX4T–PINIT plasmid encoding GST–PINIT–$(His)_6$ protein

The PINIT domain coding region was amplified from purified p513–flag–PIAS3 construct using the following primers: forward primer (5'CATATGAAGCCCCTGCCCTTCTATGAAGTCTATGGG –3') (annealing temperature 64.4^{o}C) with *Nde*I (underlined) and reverse primer (5'GTCGACTTA***GTGATGGTGATGGTGATGCAC***CTTCACTG –3') (annealing temperature 64.6^{o}C) with a *Sal*I site (underlined) and His_6–tag (bold and italics). Polymerase chain reaction (PCR) was used to amplify the PINIT domain coding region (initial denaturation, 94^{o}C for 2 minutes; cycles of denaturation, annealing and elongation, 30 cycles of 94^{o}C for 30 seconds, 55^{o}C for 30 seconds and 72^{o}C for 30 seconds respectively; and final elongation, 1 cycle at 72^{o}C for 7 minutes). The PCR–amplified fragment was ligated into the pGEM–T–Easy vector to generate pGEM–T–PINIT. The PINIT–encoding *Nde*I–*Sal*I fragment was restricted from pGEM–T–PINIT and ligated downstream of the GST–tag coding region of *Nde*I–*Sal*I restricted pGEX4T–1 expression vector to generate pGEX4T–1–PINIT with two tags (GST and His–tag). Restriction endonuclease analysis was performed using *Nde*I and Sal*I* restriction enzymes and the fidelity of the constructs were confirmed by automated direct sequencing (Rhodes University DNA sequencing unit, South Africa).

4.2.3 Construction of single tagged pGEX4T–PINIT plasmid encoding GST–PINIT protein

A GST–tagged PINIT single tagged construct was generated from the double tagged pGEX4T–1–PINIT (GST–PINIT–$(His)_6$) by insertion of a stop codon upstream of the His–tag coding region. The insertion mutagenesis was performed by PCR using the double

stranded whole plasmid linear non–PCR amplification procedure (QuikChange mutagenesis kit; Stratagene). With the following primers: (5'–CCCCGACAGTGAAGTGTAACATCACCATCACCATC–3') and (5'–GATGGTGATGGTGATGTTACACTTCACTGTCGGGG–3') (stop codon underlined). Each mutagenesis reaction was comprised of 100 ng of pGEX4T–PINIT parental plasmid template, 2.5 µl of dimethyl sulfoxide (DMSO) 10 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP mix , 125 ng of the forward primer, 125 ng of the reverse primer, 5 µl of 10x *Pfu* DNA polymerase buffer (100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, and 1 U of *Pfu* DNA polymerase and sterile distilled water to a final volume of 50 µl. Thermal cycling was allowed to proceed as follows: one cycle of denaturation (95°C for 30 seconds), 18 cycles of denaturation, annealing and extension (95°C for 30 seconds, 52°C for 60 seconds, 68°C for 5 minutes), one cycle of final extension (68°C for 7 minutes) and a 4°C hold at the end. Digestion of the parental pGEX4T–PINIT plasmid in the amplification product was achieved by the addition of 5U of *Dpn*I restriction endonuclease to the reaction mixture and incubation at 37°C overnight. Pre– and post–*Dpn*I samples were analysed by 0.8% TBE agarose gel electrophoresis. An aliquot of 10 µl of post–*Dpn*I samples was transformed into *E. coli* JM109 supercompetent cells (Promega) for screening purposes. Plasmid DNA was isolated from the resulting colonies and screened for the desired mutation by DNA sequencing using the designed forward sequencing primer (**Appendix H3**). The fidelity of the construct was verified by automated direct sequencing (Rhodes University DNA sequencing unit, South Africa).

4.2.4 Construction of PQE60–PINIT plasmid encoding PINIT–$(His)_6$ protein

The PINIT domain coding region was amplified form p513–flag–PIAS3 construct using the following primers: forward primer (5'– CCA TGG AGC CCC TGC CCT TCT ATG –3') (annealing temperature 63.5^{o}C) with *Nco*I (underlined) and the reverse primer (5'–AGATCTTCACTGTCGGGGTCAGCGG –3') (annealing temperature 64.1^{o}C) with a *Bgl* II site (underlined. The PINIT domain coding region was amplified using polymerase chain reaction (PCR) (stage1, 94^{o}C for 2 minutes; Stage 2, 30 cycles of 94^{o}C for 30 seconds, 55^{o}C for 30 seconds and 72^{o}C for 30 seconds; and Stage 3, 1 cycle at 72^{o}C for 7 minutes). The PCR–amplified fragment was ligated into the pGEM–T–Easy vector to generate pGEM–T–PINIT. The PINIT–encoding *Nco*I–*Bgl*II fragment was restricted from pGEM–T–PINIT and ligated upstream of the His–tag coding region of *Nco*I–*Bgl*II restricted pQE60 expression

vector to generate pQE60–PINIT. Restriction digestion using *Nco*I and *Bgl*II as well as automated direct sequencing was used to confirm the fidelity of the construct (Rhodes University DNA sequencing unit, South Africa).

4.2.5 Construction of pQE2–PINIT plasmid encoding $(His)_7$–PINIT protein

The p513–Flag–PIAS3–WT plasmid was used as the template for amplification of the PINIT domain coding region. The polymerase chain reaction (PCR) amplification of the PINIT domain coding region was performed with the following primers: the forward primer, 5'– CAT ATG AAG CCC CTG CCC TTC –3' (annealing temperature 58^{o}C) with a *Nde*I site (underlined); and the reverse primer, 5'– AAG CTT ATTA CAC TTC ACT GTC GGG GTC 3' (annealing temperature 60^{o}C) with *Hind* III site (underlined). The PCR consisted of three stages: stage 1, 94°C for 5 min; stage 2, 25 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds; and stage 3, 1 cycle at 72°C for 5 min. The PCR–amplified fragment was purified after agarose gel electrophoresis and ligated into the pGEM–T–Easy vector to generate pGEM–T–PINIT that was transformed into *E. coli* JM109. The PINIT domain–encoding *Nde*I – *Hind*III fragment was restricted from pGEM–T–PINIT and purified after agarose gel electrophoresis before ligation downstream of His–tag coding sequence of *Nde*I–*Hind*III restricted pQE2 expression vector to generate pQE2–PINIT. The fidelity of the constructs and the mutants were verified by automated direct sequencing (Rhodes university DNA sequencing unit, South Africa).

4.2.6 Expression of GST–PINIT–$(His)_6$, GST–PINIT and PINIT–$(His)_6$ proteins

The plasmids constructs: pGEX4T–PINIT (GST–PINIT–$(His)_6$); pGEX4T–PINIT (GST–PINIT); pQE60–PINIT (PINIT–$(His)_6$), where each transformed separately into various *E coli* expression strains (**Table 4.1**); *E .coli* BL21 (DE3), *E .coli* XLI Blue, *E .coli* M15 [pREP4] and Rosetta. The cells were spread onto Luria Bertani (LB) agar plates containing 100 µg/ml ampicillin. *E .coli* M15 [pREP4] transformed with the plasmid were spread onto LB plates containing 100 µg/ml ampicillin and 50 µg/ml kanamycin followed by incubation at 37^{o}C for approximately 16 hours to select successful transformants. Starter cultures were prepared by inoculation of a single colony of each transformants into 25 ml Luria Bertani (LB) broth containing appropriate antibiotic (*E .coli* BL21 (DE3), *E .coli* XLI Blue and *E .coli* Rosetta transformants, ampicillin; *E .coli* M15 [pREP4] transformants, ampicillin and kanamycin) and incubated overnight at 37^{o}C. The starter culture was inoculated into 225 ml fresh LB with 100 µg/ml appropriate antibiotic and grown for 2–3 hours at 37^{o}C until $A_{600\ nm}$

of 0.4 to 0.45. Temperatures were adjusted as in **Table 4.1** before inducing with isopropyl β–D–1–thiogalactopyranoside (IPTG) at a final concentration of 0.8 mM. Six hourly samples and an overnight sample were collected and analysed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) (Laemmli, 1970). The presence of PINIT–$(His)_6$ and GST–PINIT–$(His)_6$ proteins were determined by western blot analysis using mouse monoclonal anti–His primary antibody (1: 5000 dilution) and HRP–conjugated anti–mouse IgG antibodies (1:5000 dilution). The presence of GST–PINIT was determined using mouse Anti–Glutathione–S–transferase (GST) monoclonal antibody (1:5000 dilution) and HRP–conjugated anti–mouse IgG antibodies (1:5000 dilution). Chemiluminescence–based protein detection was achieved using the ECL™ western blotting kit (GE Healthcare) as per the manufacturer's instructions and captured with a Chemidoc chemiluminescence imaging system (BioRad, UK).

4.2.7 Expression and batch purification of $(His)_7$–PINIT protein by batch nickel affinity chromatography

Starter cultures were prepared by inoculating of *E. coli* XL1 Blue [pQE2–PINIT] into 25 ml Luria Bertani (LB) broth containing 100 µg/ml ampicillin, and incubating overnight at 30°C. The starter culture was inoculated into 1 liter fresh LB with 100 µg/ml ampicillin, and grown for two hours at 37°C until $A_{600\ nm}$ of 0.4 to 0.45 and switched the temperature to 20°C and inducing with isopropyl β–D–1–thiogalactopyranoside (IPTG) at final concentration of 0.8 mM. After 4 hour post inductions cells were harvested by centrifugation at 4000 g for 20 minutes. The protein was purified under native conditions using cells from a 1 liter (4 x 250 ml) culture of *E.coli* XL1 Blue [pQE2–PINIT]. Cells were lysed under native conditions in lysis buffer (10 ml; 40 mM Tris–HCl pH 8.0, 100 mM NaCl, 10 mM imidazole), containing lysozyme (100 µg/ml), 1 mM phenylmethanesulphonyl fluoride (PMSF), aprotinin 800 nM, and pepstatin 0.8 µg/ml. The cells were stored at –80°C overnight and thawed the following morning. Sonication was carried out (3 X 15 sec) and the cell debris was removed by centrifugation at 4000xg for 25 minutes at 4°C. The supernatant was added to 50% (w/v) slurry of nickel–chelating sepharose beads (1ml) in lysis buffer and allowed to bind for 2 hrs at 4°C with gentle agitation. The beads were washed three times in native wash buffer (40 mM Tris–HCl pH 8.0, 100 mM NaCl, and 50 mM imidazole) and the bound $(His)_7$–PINIT protein eluted with native elution buffer (40mM Tris–HCl pH 7.5; 100 mM NaCl; 500 mM imidazole). The eluted protein was purified further and buffer exchanged by size exclusion chromatography (section 4.2.10) as the final step of purification. Protein concentrations were

determined using the Bradford method (Bradford, 1976). Protein purity was visualized by coomassie stained 12% SDS–PAGE gels (Laemmli, 1970). The presence of $(His)_7$–PINIT was determined by western blot analysis using mouse monoclonal anti–His primary antibody (1:5000 dilution), HRP–conjugated anti–mouse IgG antibodies (1:5000 dilution). Chemiluminescence–based protein detection was achieved using the ECL™ western blotting kit (GE Healthcare) as per the manufacturer's instructions, and captured with a Chemidoc chemiluminescence imaging system (BioRad, UK).

4.2.8 HisTrap nickel affinity column protein purification of $(His)_7$–PINIT protein

A 1 ml HisTrap nickel affinity column (GE healthcare) (dynamic binding capacity of 40 mg of His–tagged protein per ml of the medium) was washed with 10 column volume (CV) of distilled water to remove storage ethanol. The column was equilibrated with 5 CV of binding buffer (20 mM sodium phosphate, pH 7.4 containing 0.5 M NaCl). Protein samples were filtered through a 0.45 µm filter and 2 ml was applied to the column, followed by washing with 10 CV wash buffer (20 mM sodium phosphate buffer, pH 7.4 and 50 mM imidazole) at 0.4 ml/min flow rate prior to elution. Bound protein was eluted with 10 CV of elution buffer (20 mM sodium phosphate, pH 7.4, containing 0.5 M NaCl and 0.5 M imidazole), using linear gradient from 0 to 100 % and followed by a further 5 CV 100 % elution buffer at 0.4 ml/min flow rate. The HisTrap column purification procedure described above was performed using a ÄKTA (BASIC) FPLC system (Amersham Pharmacia Biotech, UK). Protein purity was visualized by coomassie stained 12% SDS–PAGE gels (Laemmli, 1970). The presence of $(His)_7$–PINIT was determined by western blot analysis using mouse monoclonal anti–His primary antibody (1:5000 dilution), HRP–conjugated anti–mouse IgG antibodies (1:5000 dilution). Chemiluminescence–based protein detection was achieved using the ECL™ western blotting kit (GE Healthcare) as per the manufacturer's instructions, and captured with a Chemidoc chemiluminescence imaging system (BioRad, UK).

4.2.9 Expression and purification of STAT3 protein

Mouse STAT3β was expressed and purified as described by Becker *et al.* (1998). Briefly, pET32–STAT3β–tc vector was transformed into *E. coli* BL21 (DE3). Overnight starter cultures (25ml) in LB broth containing 200µg/ml amplicillin were inoculated into 975ml of LB broth (200µg/ml ampicillin) for expression. Cultures were grown to $A_{600\ nm}$ of 0.3 at 37°C prior to a temperature change to 20°C until $A_{600\ nm}$ of 0.5–0.6. Expression was induced with 1mM IPTG for 5–6 hours at 20°C. Cells were pelleted by centrifugation at 4 000 g and

resuspended in ice–cold lysis buffer (20mM HEPES, pH 7.5, 150mM NaCl, 10% (v/v) Glycerol). Cells were lysed by lysozyme treatment at 37°C and sonication (3 x 1min pulses at 50% power). Cell debris was pelleted by centrifugation at 27000*g* for 40min at 4°C. Supernatants were treated with 0.1% (v/v) polyethylimine (incubated on ice for 15minutes with gentle shaking) and further centrifugation at 27000*g* to remove nucleic acids followed by addition of 35% (w/v) powdered ammonium sulphate to precipitate soluble protein by centrifugation at 27000*g* (4°C) after 15 minute on ice with gentle agitation. Protein pellets were resuspended in 10ml dialysis buffer (20mM HEPES, pH 7.0, 200mM NaCl) and dialysed against 2 changes (2 hours each) of dialysis buffer at a ratio of 1:100 followed by a final exchange at 4°C overnight. STAT3 protein was further purified to homogeneity by gel filtration chromatography using Superdex 200pg 16/60HR on a ÄKTA (BASIC) FPLC system (Amersham Pharmacia Biotech, UK) system at 1ml/min. Fractions were collected, analysed by 12% SDS–PAGE and concentrated by Amicon^{R} Ultra Ultracel^{R} (10K) Centrifugal Filters (Millipore, Ireland) into 10 mM HEPES, pH 7.4, 150 mM NaCl. Protein was quantified by Bradford method (Bradford, 1976).

4.2.10 Molecular mass characterisation by size exclusion chromatography

Size exclusion chromatography of $(His)_7$–PINIT protein and $(His)_7$–PINIT mutants was performed on ÄKTA (BASIC) FPLC system (Amersham Pharmacia Biotech, UK) with a Superdex 200 HR 10/30 column at 25^{o}C. The mobile phase/elution buffer (20 mM HEPES buffer, pH 7.5 containing 150 mM NaCl) was used at a flow rate of 0.5 ml/min. Elution volumes of 1 ml were collected by peak detection at 280 nm absorbance. The molecular mass of the eluted proteins were calculated from a calibration curve using bovine serum albumin (BSA), 67 kDa, albumnin, 43 kDa, ferritin, 440 kDa, and lysozyme, 14.6 kDa as standards. Blue dextran was used to determine the void volume. All fractions were concentrated by Amicon^{R} Ultra Ultracel^{R} (10K) Centrifugal Filters (Millipore, Ireland) and analyzed by 12 % SDS–PAGE (laemmli, 1970). The presence of $(His)_7$–PINIT was determined by western blot analysis using mouse monoclonal anti–His primary antibody (1:5000 dilution), HRP–conjugated anti–mouse IgG antibodies (1:5000 dilution). Chemiluminescence–based protein detection was achieved using the ECL^{TM} western blotting kit (GE Healthcare) as per the manufacturer's instructions, and captured with a Chemidoc chemiluminescence imaging system (BioRad, UK).

4.2.11 Structural and folding analysis of $(His)_7$–PINIT protein by FTIR spectroscopy

The FTIR studies were conducted with a PerkinElmer Spectrum 100 FTIR spectrometer (PerkinElmer Inc.) and data acquisition by Spectrum version 6.3.5 software. The FTIR instrument scans automatically for 50 scans for one spectrum at 4 cm^{-1} spectral resolution. Subtraction of solvent contribution was performed by subtraction of blanking buffer HBS buffer (10 mM Hepes buffered saline, 150 mM NaCl, pH 7.4) spectra recorded under the same scanning conditions as the sample. The protein sample at 0.6 mg/ml in HBS buffer (10 mM Hepes buffered saline, 150 mM NaCl, pH 7.4) and 10µl of protein sample was used for each spectrum scan. Hydrogen exchange was performed by incubation with 20% D_2O overnight at 4°C before recording the spectra.

4.2.12 FTIR data processing and analysis

The amide I band of proteins consists of overlapping bands representing structural element such as α–helices, β–sheets, turns and non–ordered structures. Individual components were resolved by Fourier deconvolution procedure. This involved narrowing the widths of infrared bands, allowing increased separation of the overlapping components present within the broad band envelope (Kauppinen *et al.*, 1981; Haris and Severcan, 1999). Bands revealed by this technique were used to identify the different structures present in a protein and also to detect conformational changes by monitoring alterations in the frequency and intensity of these bands (Haris *et al.*, 1986; Surewicz Mantsch, 1988; Arrondo *et al.*, 1993). The absorption associated with Amide I band leads to stretching vibrations of the C=O bond of the amide. Absorption associated with Amide II bands leads primarily to bending vibration of the N–H bending with a contribution from C–N stretching. Amide I bands in the spectral ranges from 1620–1640 cm^{-1}, with proteins is attributed from β–sheet structure. Experimental studies on proteins of known structure showed that α–helices conformation gave raises to infrared absorption in the range 1650–1658 cm^{-1}. (Haris *et al.*, 1986; Surewicz Mantsch, 1988; Arrondo *et al.*, 1993; Surewicz and Mantsch, 1988; Susi and Byler, 1986; Tamm and Tatulian, 1997). In some cases, bands around 1655 cm^{-1} were attributed to large loop structures rather than to α–helices (Prestelski *et al.*, 1991) (**Appendix C1**). Analysis of the peptide group vibration in the polypeptide system allowed assignment of the characteristic bands. Secondary structure quantification by measuring the relative areas of amide I band components was performed by Gaussian curve fitting in the amide I region using PeakFit ID (SySTAT Software Inc, USA). Bands were assigned to various elements of secondary

structure (α–helix, β–sheet and unordered motifs as (Fu *et al.*, 1994; Griebenow *et al.*, 1995; Singh *et al.*, 1993). The frequencies of band centres were those obtained from the resolution enhanced spectra and Gaussian fitted curve.

4.2.13 Biochemical function of $(His)_7$–PINIT protein by dot blot association assay

$(His)_7$–PINIT–STAT3 binding studies were performed by dot blot association assay using chemiluminescence–based immunodetection for visualization. Recombinant mouse STAT3 was expressed and purified as previously described using *E.coli* BL21 (DE3) as a host strain (Becker *et al.*, 1998). STAT3 protein (50 µg, 100 µg, 200 µg, 400 µg and 500 µg,) was spotted on Hybond^{TM}–C Extra Nitrocellulose membrane in a Bio–Dot ST (BioRad, UK) apparatus connected to a vacuum pump. BSA (600 ng) was used as a negative control and $(His)_7$–PINIT protein (40 µg) as a positive control. The membrane was blocked with 5% non–fat powdered milk in TBS (50 mM Tris, 150 mM NaCl, pH 7.5) for one hour at room temperature and overlayed with 100 ng/ml PINIT protein for two hours at 4^{o}C. The membrane was washed twice for 10 minutes in TBST buffer (50 mM Tris–HCl, 150 mM NaCl, 1% Tween–20 (v/v) pH 7.5), before incubation for one hour in mouse monoclonal anti–His primary antibody (1:5000 dilution). The membrane was washed with TBST before incubation for one hour in HRP–conjugated goat anti–mouse antibody (1:5000 dilution). The membrane was washed with TBST before incubation with the chemiluminescence reagents (ECL, Amersham, UK), and detection of the signal using chemiluminescence–based protein detection. Chemiluminescence–based protein detection was achieved using the ECL^{TM} western blotting kit (GE Healthcare, UK) as per the manufacturer's instructions, and captured with a Chemidoc chemiluminescence imaging system (BioRad, UK).

4.3 RESULTS

4.3.1 Design and construction strategy of PINIT protein expression plasmids

The region coding for the PINIT domain was amplified from p513–flag–PIAS3 construct (a kind donation from Dr Hélène Boeuf, Université de Bordeaux, France) (Duval *et al.*, 2003). The construct was confirmed by diagnostic restriction endonuclease analysis (**Appendix B8**) and further verified by automated direct sequencing (Rhodes University DNA sequencing unit, South Africa). The size of the PIAS3 construct was confirmed to be 2040 bp and direct

sequencing also confirmed the presence of full length PIAS3 coding region between *Nde*I and *Hind*III sites.

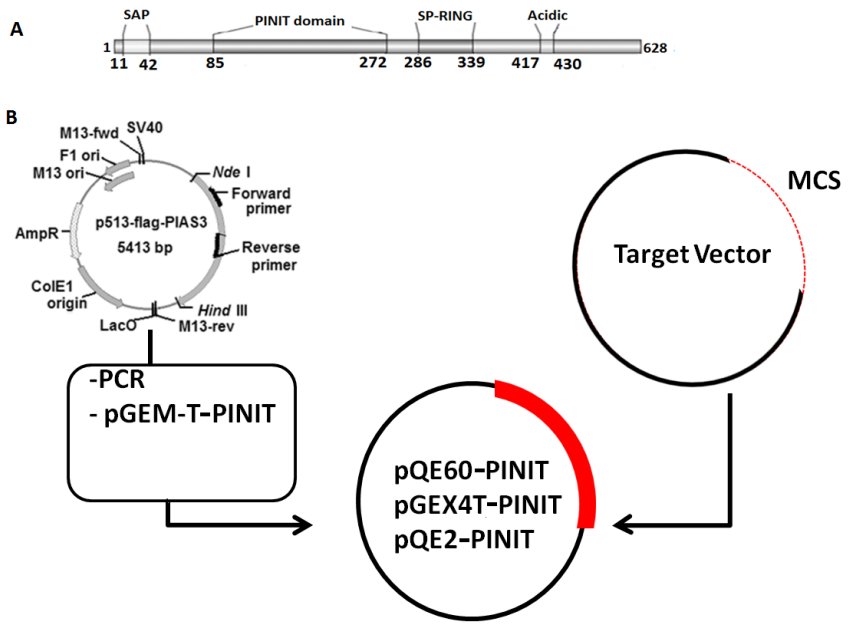

Figure 4.1 Design strategy for construction of the expression plasmids. (**A**) Conserved domains in PIAS3 showing the PINIT domain from amino acid position 85 to 272 (Accession No. AF034080). (**B**) Construction of the expression plasmids. The region encoding the PINIT domain (amino acids 85–272) was PCR amplified from p513–Flag–PIAS3 plasmid and ligated into pGEM–T–Easy to give pGEM–T–PINIT. The PINIT–encoding fragment was restricted from pGEM–T–PINIT and ligated on the multiple cloning sites (MCS) of the target expression vector to generate PINIT domain expression plasmids (pQE60–PINIT, pGEX4T–PINIT and pQE2–PINIT).

Bioinformatics tools were employed in chapter 2 to analyse the PIAS proteins and further analyse the conserved PINIT domain. The results were critical to define the choice of PINIT domain boundary. The knowledge of PINIT domain structure, sequence conservation pattern and prediction of the secondary structures or unfolded/ disordered regions were carefully considered to influence the expression and solubility of the PINIT domain.

Furthermore, analysis of the nucleotide sequences of the PINIT domain revealed the existence of arginine (AGG/AGA) and proline (CCC) rare codons. This could result in problems of early translation termination and suitable expression strains that co–expresses the tRNA for these rare codons were used. Using PCR, the region encoding the PINIT domain was amplified and restriction sites were designed to facilitate excision and insertion into a selected target vector via pGEM–T (**Figure 4.1**) to yield a construct encoding a PINIT protein.

4.3.2 PINIT–$(His)_6$ expression analysis

C–terminal tags offer certain advantages that may contribute to their effectiveness. Because the His–tag is at the end of the protein, only full–length PINIT domain proteins will be tagged and any PINIT protein truncated by premature termination of translation will not be purified by the Ni–IMAC column. Such premature termination can occur because PIAS3 protein is a mammalian protein with rare codons and is being expressed in a prokaryote system. The region encoding the PINIT domain of $PIAS3_{85–272}$ protein was amplified by PCR with *Nco*I and *Bgl*II restriction sites to facilitate insertion into pQE60 (via a pGEM–T based strategy; **Figure 4.1**) to give pQE60–PINIT. The pQE60–PINIT construct (**Figure 4.2A**) encoding a PINIT–$(His)_6$ protein was confirmed by diagnostic restriction endonuclease analysis (**Figure 4.2B**) and further verified by automated direct sequencing confirmed that PINIT domain was cloned in frame between *Nco*I and *Bgl*II sites.

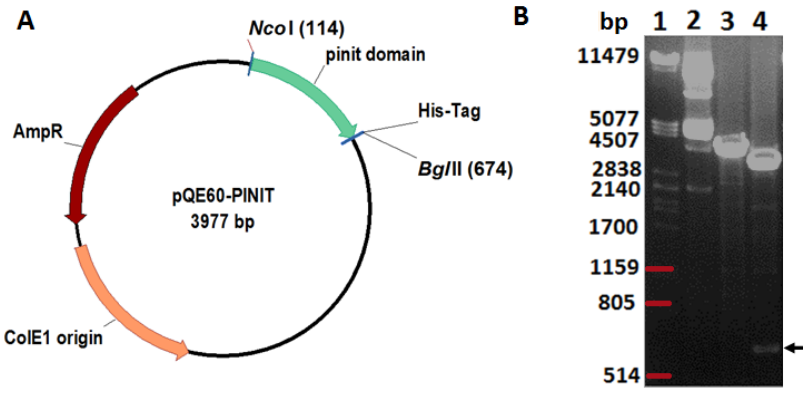

Figure 4.2 Diagnostic restriction analysis of pQE60–PINIT. A) plasmid map of pQE60–PINIT rendered using Vector NTI Advance™ software package (version 10.3; Invitrogen). The β–lactamase–encoding gene for ampicillin resistance (Amp^R) is indicated. The position of the PINIT domain between *Nco*I and *Bgl*II restriction sites are indicated; the region encoding the C–terminal His tag segments is indicated downstream of the PINIT domain coding region. The origin of replication (ColE1 origin) is indicated. **B)** Restriction analysis of the pQE60–PINIT plasmid with *Nco*I and *Bgl*II restriction enzymes. DNA was loaded on 0.8% TBE agarose gel containing ethidium bromide in the following order: lane 1, lambda DNA molecular markers digested with Pst I; lane 2, undigested pQE60–PINIT; lane 3, pQE60–PINIT plasmid digested with *Bgl*II restriction enzyme; lane 4, pQE60–PINIT plasmid digested with *Nco*I and *Bgl*II restriction enzymes. Arrow indicates the position of the PINIT domain DNA fragment. The expected size of the PINIT domain DNA fragment is 561 bp.

The PINIT–$(His)_6$ production in *E. coli* XL1 Blue [pQE60–PINIT] was observed one hour post induction on SDS–PAGE analysis and confirmed by western blot analysis ((**Figure 4.3**). Low expression levels were persistently observed despite optimisation of temperatures and use of different expression strains (**Table 4.1**). The low expression levels of PINIT–$(His)_6$ could be results of early translation termination or the protein being toxic to the expression

strain. Expression of the PINIT domain in a different expression vector and strain that could potentially improve the expression, folding and solubility was necessary.

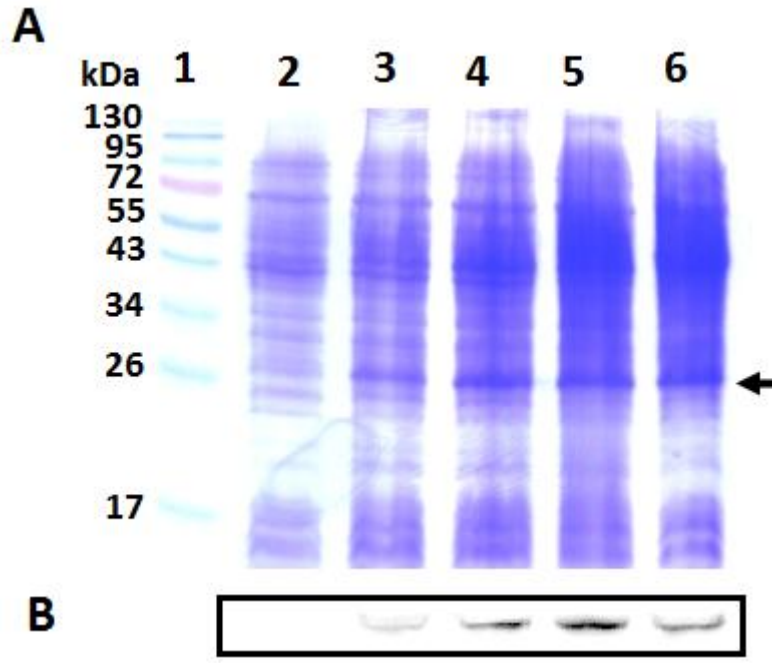

Figure 4.3 Heterologous expression levels of PINIT–$(His)_6$ were low. A) 12% SDS–PAGE gel of the PINIT–$(His)_6$ containing total protein extracts from *E.coli* XL1 Blue [pQE60–PINIT]: lane 1, molecular mass marker; lane 2, non–induced sample; lanes 3 – 5, IPTG induced samples 1, 3, 5 hour post induction; and lane 6, overnight post induction sample. . **B**) Western blot analysis for the detection of PINIT–$(His)_6$ protein in *E. coli* XL1 Blue [pQE60–PINIT] total protein extracts prepared for the expression analysis in (A) using mouse monoclonal anti–His primary antibody and HRP–conjugated goat anti–mouse IgG antibodies. The arrow indicates the position of PINIT–$(His)_6$.

4.3.3 GST–PINIT–$(His)_6$ expression analysis

Double tagging the PINIT domain using N–terminal GST–tag and C–terminal His–tag was done to maximise the benefit of the tags. pGEX4T–1 expression vector was constructed as described in section 4.2.2. The primary advantage of GST–tag is it will potentially increase the solubility of insoluble or semi–soluble proteins expressed in *E. coli* and both tags facilitate purification. Therefore, the region encoding the PINIT domain of $PIAS3_{85–272}$ protein was amplified by PCR with *Nde* I and *Sal*I restriction sites to facilitate insertion into pGEX4T (via a pGEM–T based strategy; **Figure 4.1**) to give pGEX4T–PINIT. The reverse prime was designed with codons encoding for the His–tag upstream of the *Sal*I restriction site to introduce a C–terminal His–tag. The pGEX4T–PINIT construct (**Figure 4.4A**) encoding a GST–PINIT–$(His)_6$ protein was confirmed by diagnostic endonuclease restriction analysis (**Figure 4.4B**). The size of the PINIT domain encoding region was confirmed to be 561 bp

and automated direct sequencing (Rhodes University DNA sequencing unit, South Africa) confirmed that the PINIT domain coding region was inserted between *Nde* I and *Sal*I sites and was in frame.

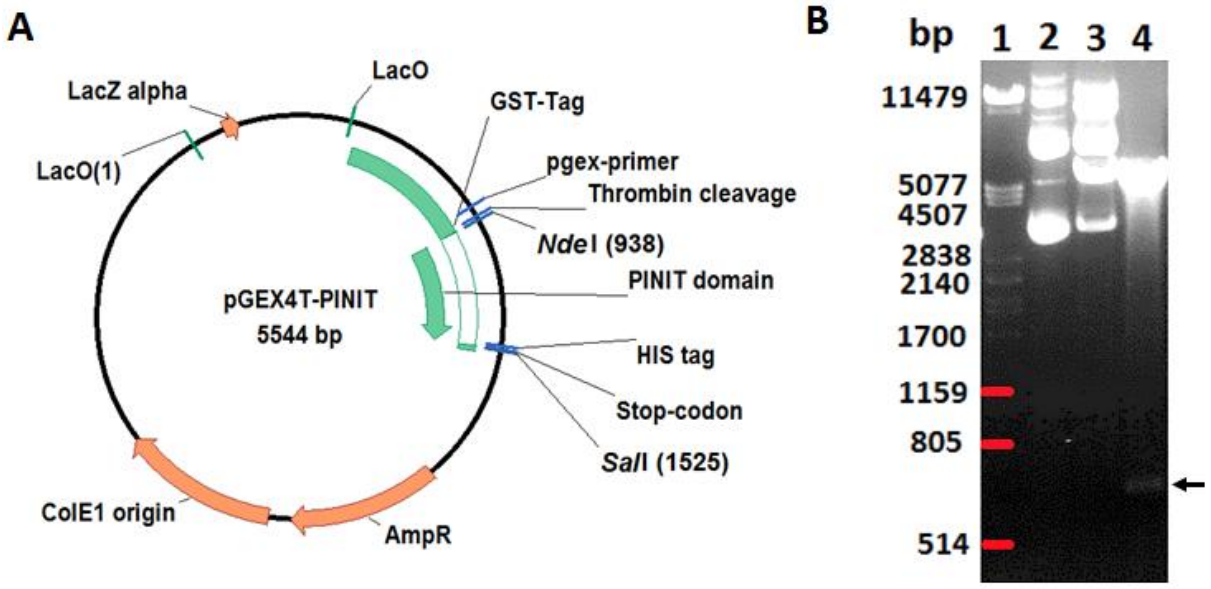

Figure 4.4 Diagnostic restriction analysis of pGEX4T–PINIT. A) Plasmid map of pGEX4T–PINIT rendered using Vector NTI Advance™ software package (version 10.3; Invitrogen). The plasmids confer ampicillin resistance to transformed *E. coli* cells as indicated (Amp^R; β–lactamase gene). The position of the PINIT domain between *Nde*I and *Sal*I restriction sites is indicated; The region encoding the GST tag segments is indicated upstream of the PINIT domain coding region and the region encoding for the His Tag is indicated downstream of the PINIT domain coding region. The origin of replication (ColE1 origin) is indicated and regions coding for *Lac*Z alpha, *Lac*O genes and the thrombin cleavage site are indicated. **B**) Restriction analysis of the pGEX4T–PINIT plasmid with *Nde*I and *Sal*I restriction enzymes. DNA was loaded on 0.8% TBE agarose gel containing ethidium bromide in the following order: Lane1, lambda DNA molecular markers digested with *Pst*I; lane 2, undigested pGEX4T–PINIT; lane 3, pGEX4T–PINIT plasmid digested with *Nde*I restriction enzyme; lane 4, pGEX4T–PINIT plasmid digested with *Nde*I and *Sal*I restriction enzymes. Arrow indicates the position of the PINIT domain DNA fragment. The expected size of the PINIT domain DNA fragment is 561 bp.

GST–PINIT–$(His)_6$ protein was observed one hour after induction and the presence of the protein was confirmed by western blot analysis at 45 kDa molecular mass. Low level of protein expression was observed on SDS–PAGE (**Figure 4.5**). Western blot analysis was performed using mouse monoclonal anti–His primary antibody to detect the PINIT domain protein C–terminal His tag. **Figure 4.5B** showed that the full length GST–PINIT–$(His)_6$ was expressed but at insufficient levels to be observed clearly on SDS–PAGE. Probably overexpression of the protein was affected by the disadvantages of the tags as discussed in section 4.1 and also the existence of the rare codons resulted in early translation termination. The existence of the arginine (AGG/AGA) and proline (CCC) rare codons in the PINIT

domain coding region cause early translation termination. Therefore *E. coli* BL21 (DE3) CodonPlus–RP expression strain that co–express the tRNA for these rare codons was used (**Table 4.1**). However, expression trials were performed at various expression temperatures and using various *E. coli* expression strain (**Table 4.1**) with no improved expression levels.

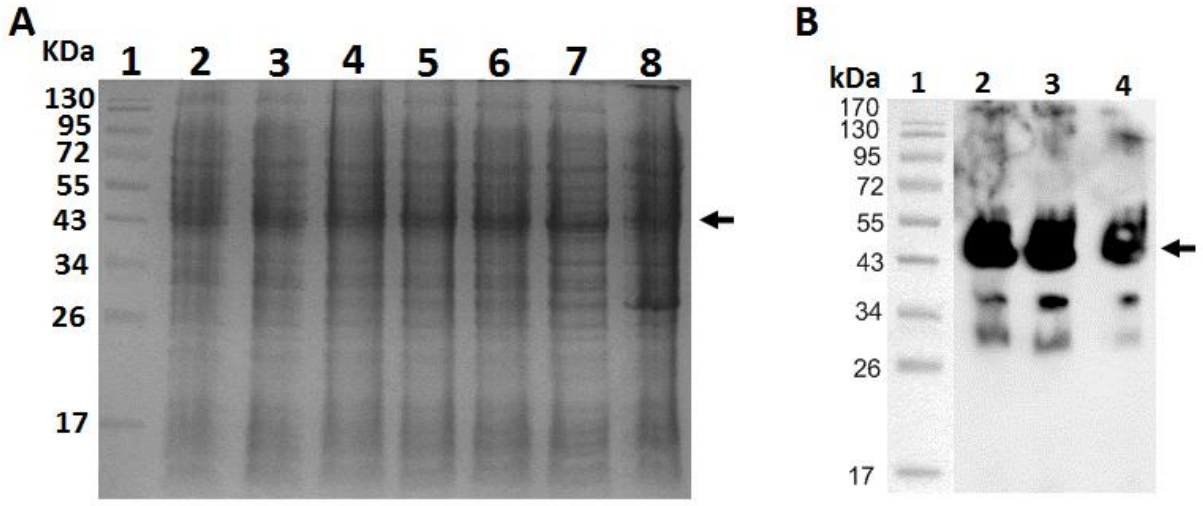

Figure 4.5 GST–PINIT–$(His)_6$ was expressed at low levels. A) 12% SDS–PAGE gel of the GST–PINIT–$(His)_6$ containing total protein extracts from *E.coli* BL21 (DE3) [pGEX4T–PINIT]: lane 1, molecular mass marker; lane 2, non–induced sample; lanes 3 – 7, induced samples using IPTG taken hourly; and Lane 8, overnight sample. **B**) Western blot analysis of GST–PINIT–$(His)_6$ protein in *E.coli* BL21 (DE3) [pGEX4T–PINIT] using mouse monoclonal anti–His primary antibody, HRP–conjugated goat anti–mouse IgG antibodies: lane 1, molecular mass marker; lane 2, induced samples using IPTG taken after one hour; lane 3, four hours induction; lane 4, overnight (16 hours) induction. The arrow indicated the position of the GST–PINIT–$(His)_6$ protein. The arrow indicates the position of GST–PINIT–$(His)_6$ protein.

Although the GST–PINIT–$(His)_6$ protein was expressed, the low level of expression requires further optimisation of the expression condition. Possible reasons for low expression were straining of the PINIT protein folding resulting in possibly in a toxic protein due to the two tags. The existence of a C–terminal His–tag possibly resulted in a toxic protein fold. Removing the C–terminal His–tag would possibly improve the expression levels.

4.3.4 GST–PINIT expression analysis

A stop codon (TAA) was introduced by site directed mutagenesis upstream of the C–terminal His–tag coding region to express GST–PINIT protein. The verified construct was expressed in *E. coli* BL21 (DE3) [pGEX4T–PINIT–TAA–$(His)_6$] and expression analysed by SDS–PAGE and western blot. Low expression levels of GST–PINIT protein were observed one hour post induction (**Figure 4.6A, lane 1**). Molecular mass species at approximately 47 kDa

and 26 kDa were confirmed by western blot analysis to be a GST–PINIT and possible GST–tag (**Figure 4.6A and 4.6B**). The expression analysis repeatedly showed consistent expression of the 47 kDa and 26 kDa species.

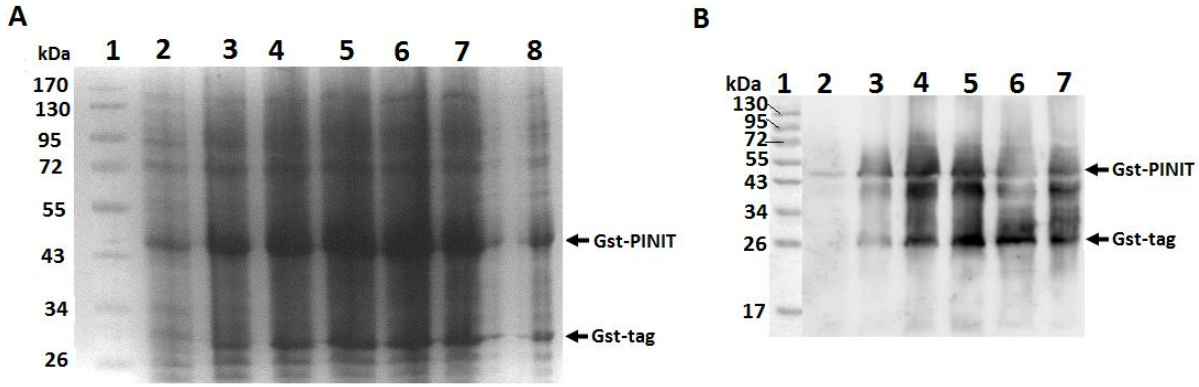

Figure 4.6 GST–PINIT was expressed at low levels. A) 12% SDS–PAGE gel of the GST–PINIT containing total protein extracts from *E.coli* BL21 (DE3) [pGEX4T–PINIT–TAA–$(His)_6$]: lane 1, molecular mass marker; lane 2, non–induced sample; lanes 3 – 7, induced samples using IPTG taken hourly; and Lane 8, overnight sample. **B**) Western blot analysis of GST–PINIT protein in *E.coli* BL21 (DE3) [pGEX4T–PINIT–TAA–$(His)_6$] using mouse anti– GST monoclonal antibody and HRP–conjugated goat anti–mouse IgG antibody: lane 1, molecular mass marker; lane 2, non–induced sample; lane 3–6 induced samples using IPTG taken after 1, 2, 4 and 6 hours induction; lane 7, overnight (16 hours) induction. The arrow indicated the position of the GST–PINIT protein and the GST–tag protein.

A possible reason for the low expression level was attributed to the folding of the GST–PINIT protein. The molecular weight of the GST is 26 kDa and the PINIT domain is approximately of the same size, 23 kDa. Possibly the GST–tag alters the folding of the PINIT domain resulting in cleavage and degradation of the PINIT domain protein, hence the observed GST–tag protein on SDS–PAGE and by western blot analysis (**Figure 4.6A and 4.6B**). An attempt to purify the GST–PINIT using a GST–Trap column (GE healthcare) resulted in degradation of the PINIT protein and eluted GST–tag protein only (data not shown). Further expression optimisation at different temperatures and use of various expression strains (**Table 4.1**) did not improve expression levels.

4.3.5 $(His)_7$–PINIT protein expression analysis

The region encoding the PINIT domain of $PIAS3_{85–272}$ protein was amplified by PCR with *Nde* I and *Hind* III restriction sites to facilitate insertion into pQE2 (via a pGEM–T based strategy; **Figure 4.7**) to give pQE2–PINIT. The pQE2–PINIT construct encoding a $(His)_7$–PINIT protein was confirmed by diagnostic restriction endonuclease analysis and the PINIT

domain coding region of size 561 bp was confirmed (**Figure 4.7**) and further verified by automated direct sequencing to show that the coding region was inserted in–frame with *Nde* I and *Hind* III sites.

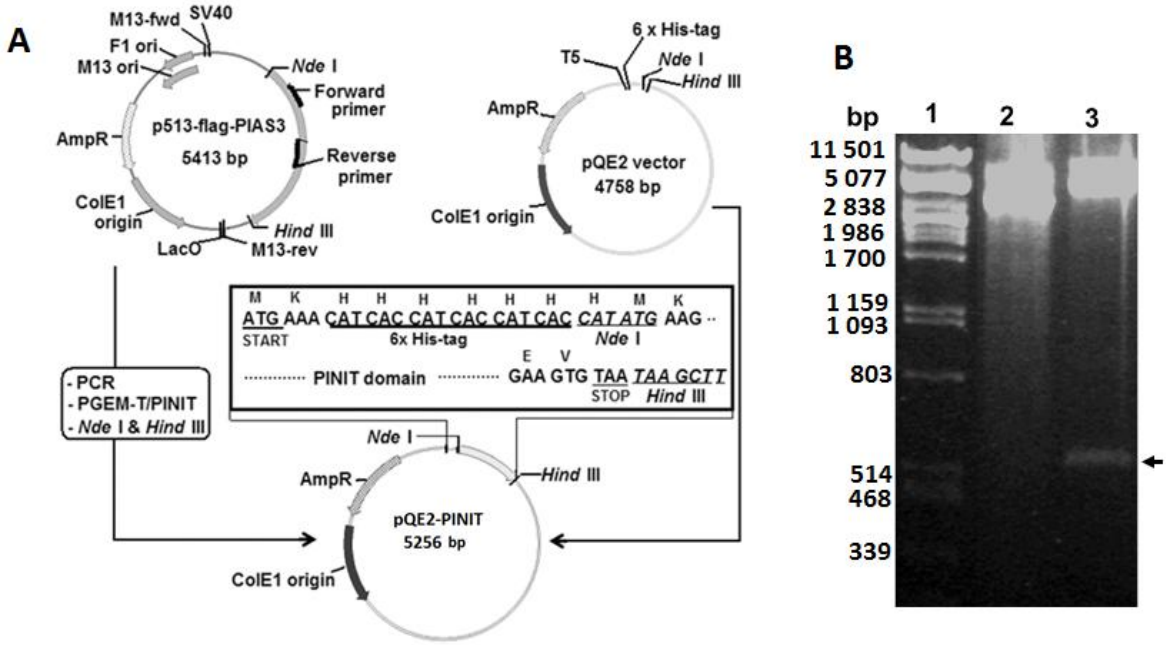

Figure 4.7 Construction and diagnostic analysis of pQE2–PINIT plasmid. A). The region encoding the PINIT domain (amino acids 85–272) was PCR amplified from p513–Flag–PIAS3 plasmid and ligated into pGEM–T–Easy to give pGEM–T–PINIT. The PINIT–encoding fragment was restricted from pGEM–T–PINIT and ligated downstream of the His–tag coding sequence of *Nde*I–*Hind*III restricted pQE2 expression vector to generate pQE2–PINIT. **B**) Restriction analysis of the pQE2–PINIT plasmid with *Nde*I and *Hind*III restriction enzymes. DNA was loaded on 0.8% TBE agarose gel containing ethidium bromide in the following order: Lane1, lambda DNA molecular markers digested with *Pst*I; lane 2, undigested pQE2–PINIT plasmid; lane 3, pQE2–PINIT plasmid digested with *Nde*I and *Hind*III restriction enzymes. Arrow indicates the position of the PINIT domain DNA fragment. The expected size of the PINIT domain DNA fragment is 561 bp. (adapted from Mautsa *et al*., 2010).

4.3.5 Expression optimisation of $(His)_7$–PINIT protein in *E. coli* XL 1 Blue [pQE2–PINIT]

$(His)_7$–PINIT production in *E. coli* XL1 Blue [pQE2–PINIT] was evident as an over–expressed protein one hour after induction, with the highest expression levels overnight post induction (**Figure 4.8A, lane 8**). Western blot analyses confirmed the expression of $(His)_7$–PINIT protein (**Figure 4.8B**). The predicted subunit molecular mass of the $(His)_7$–PINIT protein is 23 kDa, and this correlated with its observed subunit molecular mass on a SDS–PAGE gel (slightly below the 26 kDa marker; **Figure 4.8A)**. Although the highest expression was evident on overnight post induction sample, higher order species of $(His)_7$–PINIT were

observed by SDS-PAGE and western blot analysis of the whole cell lysates (data not shown) suggesting formation of protein aggregates. Therefore, all subsequent studies were conducted with a four hour post induction expressed $(His)_7$–PINIT protein.

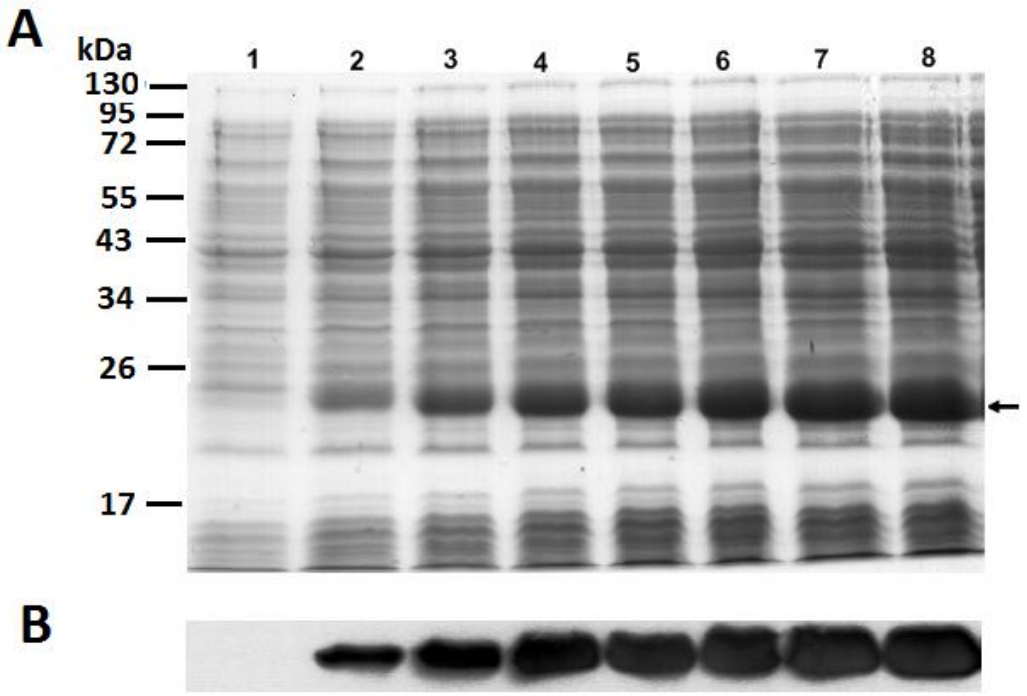

Figure 4.8 Heterologous over-expression of $(His)_7$-PINIT was successful. A) 12% SDS–PAGE analysis of the $(His)_7$–PINIT–containing total protein extracts prepared from *E. coli* XL1 Blue [pQE2–PINIT] during expression analysis: Molecular mass markers are indicated on left hand side; lane 1,non–induced sample; lanes 2–7, induced samples taken hourly; and lane 8, overnight induced sample. Arrow indicate the position of the $(His)_7$–PINIT protein. **B**) Western blot analysis for the detection of $(His)_7$–PINIT protein in *E. coli* XL1 Blue [pQE2–PINIT] total protein extracts prepared for the expression analysis in (A) using mouse monoclonal anti–His primary antibody and HRP–conjugated goat anti–mouse IgG antibodies. The arrow indicates the position of $(His)_7$–PINIT protein. (adapted from Mautsa *et al*., 2010).

Table 4. 1 Cloning vectors, expression strains and expression condition trials for optimum production of the PINIT domain fusion proteins.

Expression vector	**Expressed protein**	*E. coli* **Expression strains**	**Temperature ^{o}C**	**IPTG mM**	**Optimum expression conditions**
pQE60	PINIT–$(His)_6$	XL 1 Blue; BL 21(DE3); M15 [pREP4]	37; 30; 20	0.8 ; 1.0	Low expression observed
pGEX4–1	GST–PINIT–$(His)_6$	XL 1 Blue; BL 21(DE3); Rosseta; M15[pREP4]	37; 30; 20	0.8; 1.0	Expression at 20^{o}C in BL21 (DE3) induced by 1 mM IPTG
pGEX4T–1	GST–PINIT	XL 1 Blue; BL 21(DE3); Rosseta; M15[pREP4]	37 ; 30 ; 20	0.8 ; 1.0	Expression at 20^{o}C in BL21 (DE3) induced with 1mM IPTG
pQE2	$(His)_7$–PINIT	XL 1 Blue; BL 21(DE3)	37; 30; 20	0.8; 1.0	Expression at 20^{o}C in XL 1 Blue with 0.8 1mM IPTG

4.3.6 Solubility and batch purification studies of $(His)_7$–PINIT protein by nickel affinity chromatography

Both denaturing and native conditions were employed to investigate solubility and purification of $(His)_7$–PINIT. Denaturing conditions were considered necessary in order to investigate the nature and condition of the higher order species detected by western blot analysis in whole cell lysates from SDS–PAGE gel. Cell lysis and purification was carried out in the presence of the 8 M urea based buffers. Analysis of the total, soluble and insoluble protein by SDS–PAGE showed the presence of an overexpressed $(His)_7$–PINIT protein that was confirmed by western blot analysis (**Figure 4.9A, 4.9B, 4.9C**). The SDS–PAGE analysis

showed that the protein was soluble as most of the protein was found in the soluble fraction (**Figure 4.9A, lane 6**). However, the quantitative recovery of the protein in **Figure** 4.9B showed that the protein was not stable under denaturing purification conditions as shown by $(His)_7$–PINIT degradation products below the 17 kDa marker (**Figure 4.9C**). Western blot analysis detected the presence of the monomeric species of $(His)_7$–PINIT protein and not higher order species. However analysis of nickel affinity batch purified protein on a SDS–PAGE gel revealed different molecular mass proteins at 45 kDa and 100 kDa (**Figure 4.9B**), possibly representing co–purifying *E.coli* chaperone proteins (Flynn *et al.*, 1991; Blond–Elgundi *et al.*, 1993). Mass spectroscopy revealed that the species at 26 kDa, 45 kDa and 95 kDa were $(His)_7$–PINIT protein, elongation factor Tu (EFTU) and ClpB respectively.

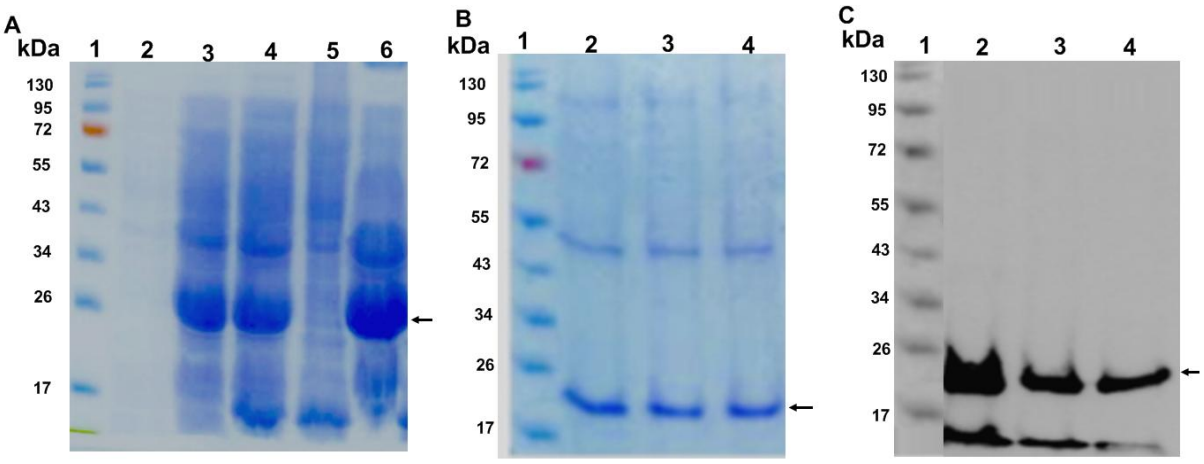

Figure 4.9 Solubility and batch nickel affinity purification studies under denaturing conditions of $(His)_7$–PINIT. A) SDS–PAGE (12 %) gel of expression of $(His)_7$–PINIT protein from *E.coli* XL1 Blue[pQE2–PINIT]: Lane 1, molecular mass marker; lane 2, uninduced whole cell lysates; lane 3, whole cell lysates after 4 hours induction; lane 4, total protein lysates after sonication; lane 5, Insoluble protein; lane 6, soluble protein, supernatant after centrifugation of the cell lysates. **B) SDS–PAGE (12 %) gel of purified $(His)_7$–PINIT:** lane 1, molecular mass marker; lane 2, first elution batch with 500mM imidazole; lane 3, second elution batch with 500mM Imidazole; lane 4, third elution batch with 500mM imidazole. **C) Western blot analysis of the purified $(His)_7$–PINIT from *E.coli* XL1 Blue [pQE2–PINIT]:** Lane 1, molecular mass marker, lane 2, first elution batch with 500mM imidazole; lane 3, second elution batch with 500mM Imidazole; lane 4, third elution batch with 500mM imidazole. Arrows indicate the position of $(His)_7$–PINIT protein.

To biochemically characterize the protein, a soluble folded protein is necessary and therefore purification of the protein in its native state was performed using nickel affinity batch purification chromatography. Analysis of the purified protein by SDS–PAGE showed predominantly monomeric species of $(His)_7$–PINIT indicated by the arrow (**Figure 4.10A**). The SDS–PAGE showed that $(His)_7$–PINIT is soluble under native purification conditions as most of the protein was shown to be in the soluble fraction (**Figure 4.10 A, lane 3**) and not in

the insoluble fraction (**Figure 4.10A, lane 4**). The quantitative recovery of the protein shown on the SDS–PAGE gel showed that the protein was stable under native purification conditions compared to the denaturing purification condition. The western blot analysis indicated the presence of the purified $(His)_7$–PINIT protein (**Figure 4.10 B**). However, co–purifying *E. coli* species were observed at 55–72 kDa of the elution fractions and these corresponded to DnaK and GroEL molecular masses. Western blot analysis of the purified products with Anti-DnaK antibodies indicated that the contaminating species were different molecular mass species of *E. coli* DnaK (**Appendix B6**).

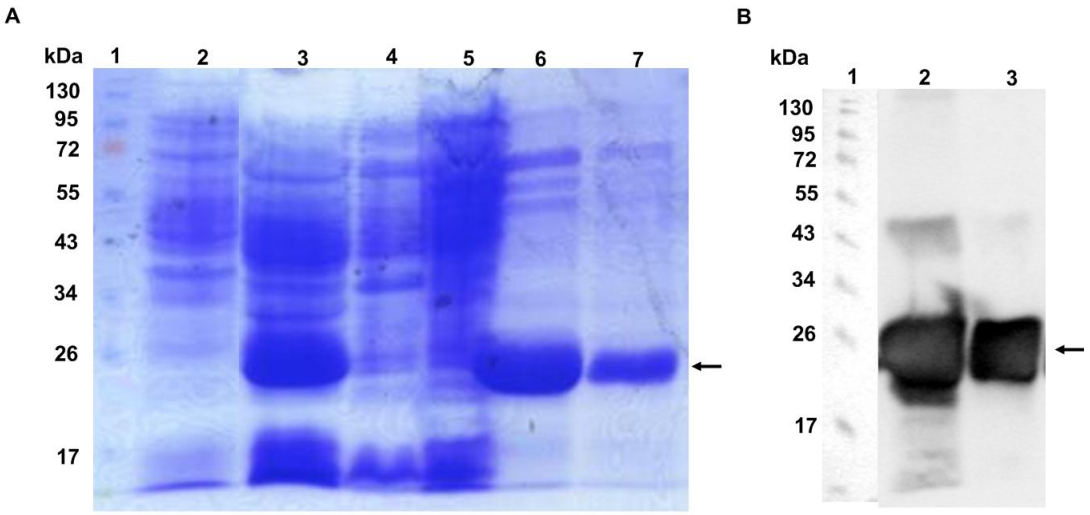

Figure 4.10 Solubility and batch nickel affinity purification studies under native conditions of $(His)_7$–PINIT. A) SDS–PAGE (12 %) gel of expression and purification of $(His)_7$–PINIT protein from *E.coli* XL1 Blue [pQE2–PINIT]: Lane 1, molecular mass marker; lane 2, uninduced whole cell lysates; lane 3, soluble protein, supernatant after centrifugation of the cell lysates; lane 4, insoluble protein; lane 5, flow through; lane 6, first elution batch with 500 mM imidazole; lane 7, second elution batch with 500mM Imidazole. **B) Western blot analysis of the purified $(His)_7$–PINIT from *E.coli* XL1 Blue [pQE2–PINIT]:** Lane 1, molecular mass marker; lane 2, first elution batch with 500 mM imidazole; lane 3, second elution batch with 500 mM Imidazole. Arrows indicate the position of $(His)_7$–PINIT protein.

4.3.7 HisTrap nickel affinity column purification of $(His)_7$–PINIT protein

A pre–packed HisTrap nickel affinity column was employed for the purification of $(His)_7$–PINIT protein. The column was connected to the FPLC and fast flow washes at 4ml/min with 60 mM imidazole wash buffer was used to reduce non specific binding *E. coli* protein. The chromatogram (**Figure 4.11A**) showed the 5 ml peak of the non-specific *E. coli* proteins washed out with 60 mM imidazole wash buffer. The gradient elution was switched on after 10 CV washes and $(His)_7$–PINIT protein was eluted at 125 mM imidazole concentration at 14 ml peak (**Figure 4.11A)**. SDS–PAGE gel analysis of the 14 ml peak fraction showed $(His)_7$–PINIT protein as the major species (**Figure 4.11B**) at molecular mass below 26 kDa

consistent with the batch purification processes. The amount of *E. coli* protein contaminants co–purifying with the $(His)_7$–PINIT were significantly less compared to batch nickel affinity purification.

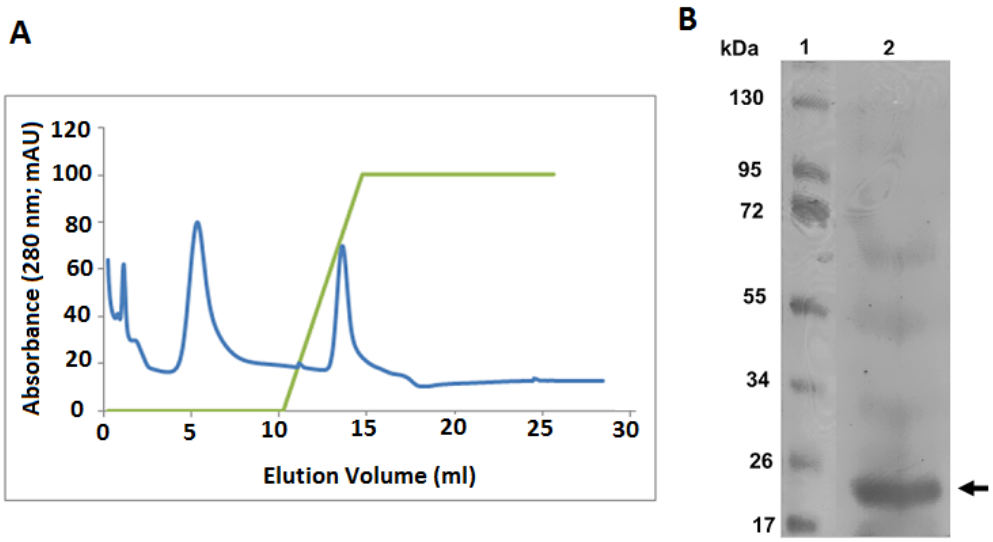

Figure 4.11 Column based nickel affinity purification of $(His)_7$–PINIT. A) Chromatogram of $(His)_7$–PINIT purification from 1ml HisTrap Nickel affinity column using a ÄKTA (BASIC) FPLC system (Amersham Pharmacia Biotech, UK). The peak at 6 ml indicate protein washed out with 60 mM imidazole wash buffer and peak at 14 ml indicate $(His)_7$–PINIT protein elution with 500 mM imidazole elution buffer. The blue line indicate the chromatogram trace at absorbance 280 nm and the green line indicate the gradient percentage increase of the imidazole concentration in the elution buffer. The complete purification process was performed at 4 ml/min flow rate **B) SDS–PAGE (12 %) gel analysis of the peak fraction elution 14 ml**. Lane 1, molecular mass marker; lane 2, elution volume fraction 14 ml. Arrow indicate the position $(His)_7$–PINIT protein.

4.3.8 Purification with Adenosine Tri–phosphate (ATP) based buffers

Although $(His)_7$–PINIT was successfully expressed at 20^{o}C, the level of *E–coli* proteins co–purifying required further purification or optimisation of the purification conditions. The quantitative recovery of these higher molecular mass species suggested that they were binding partners. Furthermore, mass spectroscopy and western blot analysis of the purified $(His)_7$–PINIT protein samples revealed that these higher order contaminating species were EFTU, ClpB and DnaK respectively. Other possible proteins that could co–purify with the PINIT domain were GroEL and GroES. Owing to the distinct ability of DnaK chaperone to bind and release substrate in the presence of ADP/ATP, respectively (Rohman and Harrison-Lavoie, 2000), pre–incubation of the protein solution with ATP prior to purification dissociates protein–DnaK complex (Amersham Pharmacia Biotech, The recombinant protein handbook, 18–1142–75, 2001). Also this is known to reduce but not completely remove the contaminating GroEL (Rohman and Harrison-Lavoie, 2000).

The total soluble protein was pre–incubated in binding buffer containing 5 mM ATP for 20 minutes at room temperature before applying the sample to the HisTrap column equilibrated with binding buffer that contained 5 mM ATP. The elution fractions were analysed by SDS–PAGE gel and showed decrease of the higher molecular species (**Figure 4.12**). Lower molecular weight species present on the SDS–PAGE could possibly have been degradation products from sample treatment or early translation termination products.

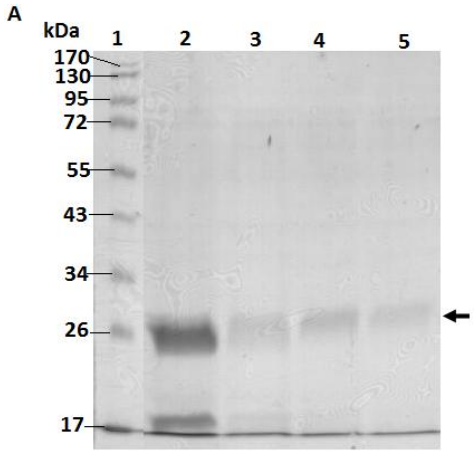

Figure 4.12 Purification of $(His)_7$–PINIT using ATP during column based nickel affinity chromatography. A) SDS–PAGE (12 %) analysis of $(His)_7$–PINIT purified from ***E.coli* XL1 Blue [pQE2–PINIT]**. Lane 1, molecular mass marker; lane 2 – 5, $(His)_7$–PINIT elution fraction 13 to 16 ml respectively from the chromatogram of $(His)_7$–PINIT (chromatogram not shown). Arrow indicate the position of the $(His)_7$–PINIT.

4.3.9 $(His)_7$–PINIT protein characterisation by size exclusion chromatography

Characterisation of the size and determination of oligomeric state of the $(His)_7$–PINIT domain protein in solution was performed by size exclusion chromatography. This method was also used to further purify protein after nickel affinity chromatography. The size exclusion column was calibrated using commercial standards. Blue dextran was used to determine the void volume. The molecular weight of the PINIT domain protein was calculated from the standard curve (**Appendix B8 and B9**).

A two stage size exclusion chromatography was performed. First stage, partially purified $(His)_7$–PINIT from column-based native nickel affinity chromatography in the presence of ATP was applied to the size exclusion column. The chromatogram (**Figure 4.13A**) showed multiple peaks of higher order species of *E. coli* contaminants shown on the SDS–PAGE (**Figure 4.13B**). However, the size exclusion column was equilibrated with ATP based buffer and the contaminants were separated from the $(His)_7$–PINIT by size. SDS–PAGE and western blot analysis of the eluted fractions showed that the $(His)_7$–PINIT protein eluted at

fractions 16 to 19 ml. Western blot analysis of the fractions indicated that bulk of the $(His)_7$–PINIT protein was eluted at fraction 16, 17 and 18 ml (**Figure 4.13C**). In the second stage, fractions 16 to 18 ml from the first size exclusion analysis were concentrated using Amicon^R Ultra Ultracel^R (10K) Centrifugal Filters (Millipore, Ireland) and loaded into the size exclusion column. A single distinct peak was observed and found to contain $(His)_7$–PINIT by western blot analysis (**Figure 4.13D**).

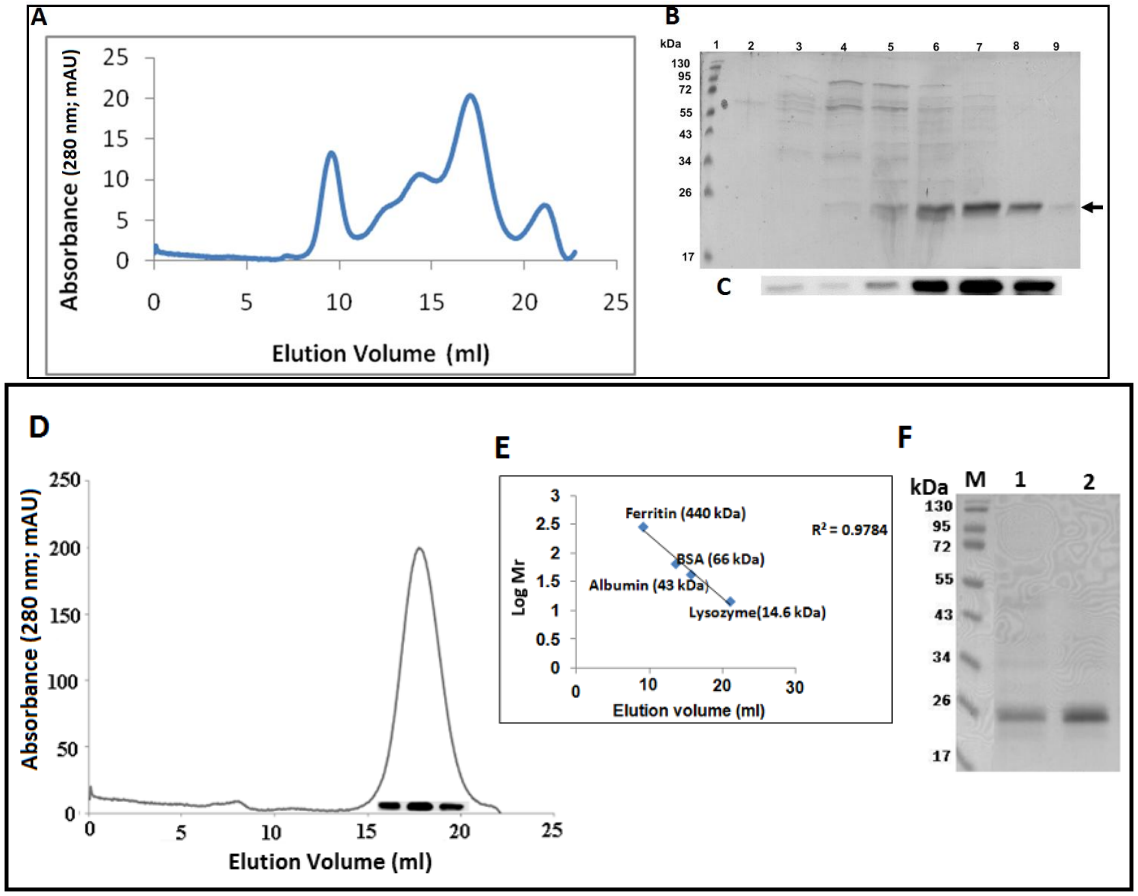

Figure 4.13 Purification of $(His)_7$–PINIT protein by two stage size exclusion chromatography. A) Size exclusion chromatography of $(His)_7$–PINIT protein fraction eluted from HisTrap column based native purification with ATP. **B)** SDS–PAGE (12 %) gel analysis of size exclusion fractions 12 ml to19 ml; lane 1, molecular mass marker; lane 2–9, fraction 12 to 19 ml respectively from the size exclusion column. Arrow indicate the position of the $(His)_7$–PINIT.**C**) Western blot analysis of fraction 13 to 18 ml resolved by SDS–PAGE analysis. **D**) Stage two of size exclusion chromatography of concentrated $(His)_7$–PINIT fractions (16 to 19 ml) from first stage of size exclusion chromatography and western blot analysis of the peak fractions. **E**) Standard curve plot for $(His)_7$–PINIT molecular mass determination. The molecular mass of the $(His)_7$–PINIT peak fraction was found to be approximately 23 kDa. **F**) SDS–PAGE analysis of the size exclusion fractions 18 and 19 ml. (adapted from Mautsa *et al*., 2010).

The mobility of this peak indicated a species of molecular mass of approximately 23 kDa implying that the PINIT domain existed as a monomeric species in solution (**Figure 4.13E**).

The $(His)_7$–PINIT protein was purified to homogeneity as shown by a single major protein band on SDS–PAGE (**Figure 4.13F**). The fractions under the peak were concentrated and the protein concentration quantified by Bradford assay to be approximately 0.6 mg per litre of original culture.

4.3.10 Secondary structure analysis of $(His)_7$–PINIT protein by FTIR spectroscopy

The PINIT domain protein was analysed in an H_2O environment and in 20 % D_2O based HBS buffer (**Figure 4.14)**. Amide I band arises mainly from the overlapping of stretching vibrations of the C=O bonds of the protein backbone (Susi and Byler, 1986) and Amide II bands arises from vibration in NH side chains. Comparison of the PINIT domain in H_2O environment and in the D_2O environment (**Figure 4.14**) reflected spectral shift as a consequence of the isotropic replacement of the exchangeable hydrogen of the PINIT domain amino groups.

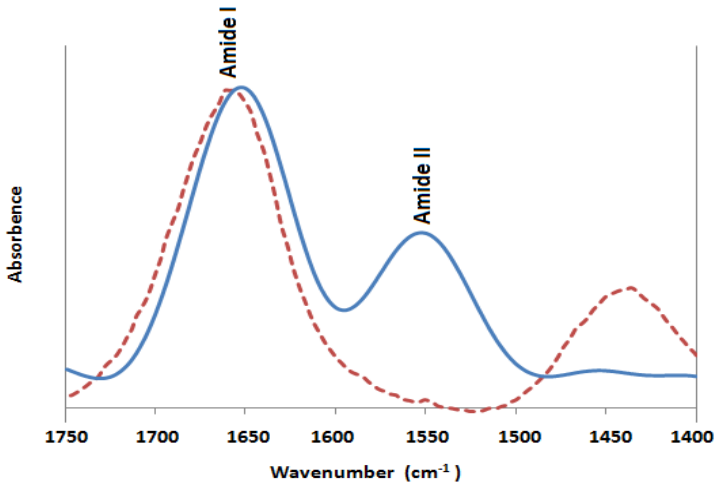

Figure 4.14 FTIR analysis of purified $(His)_7$–PINIT. Infrared spectrum of the PINIT domain protein in H_2O solution (blue continuous trace) and in 20 % D_2O solution (red dashed trace) Difference spectra after digital subtraction of the buffer spectrum. Amide I, 1600−1690 cm^{-1} arises from C=O stretching; Amide II, 1480−1575cm^{-1} arises from CN stretching and NH bending (Miyazawa *et al.*, 1956; Krimm and Bandekar, 1986). Amide II frequencies shift for PINIT domain in D_2O is due to deuterated exchange.

In a well–defined folded three–dimensional structure protein, many of the amide hydrogen were buried in the interior of the protein and this resulted in marked decrease in the rate of amide hydrogen to water molecules. However, with time they eventually exchanged since proteins are flexible. Previous researchers have used hydrogen isotope exchange to show protein fold (Kunihiro *et al.*, 1984). The addition of heavy water affected the vibration frequencies of Amide I and Amide II bands of the infrared spectrum of the protein.

Superimposition of the spectrums over different times of deuterium exchange showed the change of the Amide II band (**Figure 4.14**).

Peaks centred at 1625, 1634.7,1644 cm^{-1} for PINIT protein in the H_2O environment (**Figure 4.15**) were inside the region which was a characteristic of the Amide group implicated in β–sheets (Byler and Susi, 1986; Susi and Byler, 1986; Surewicz and Mantsch, 1988). The bands around 1625 cm^{-1} and 1691 cm^{-1} reflected a particular β–structure which involved stronger hydrogen bonding. Bands at 1653 cm^{-1} in H_2O environment can in principle be assigned to α–helical and 1662.9 cm^{-1} can be assigned to 3_{10} α–helix (Krimm and Bandekar, 1986; Susi and Byler, 1986; Dong *et al.*, 1990) even though this structure is rarely observed in proteins.

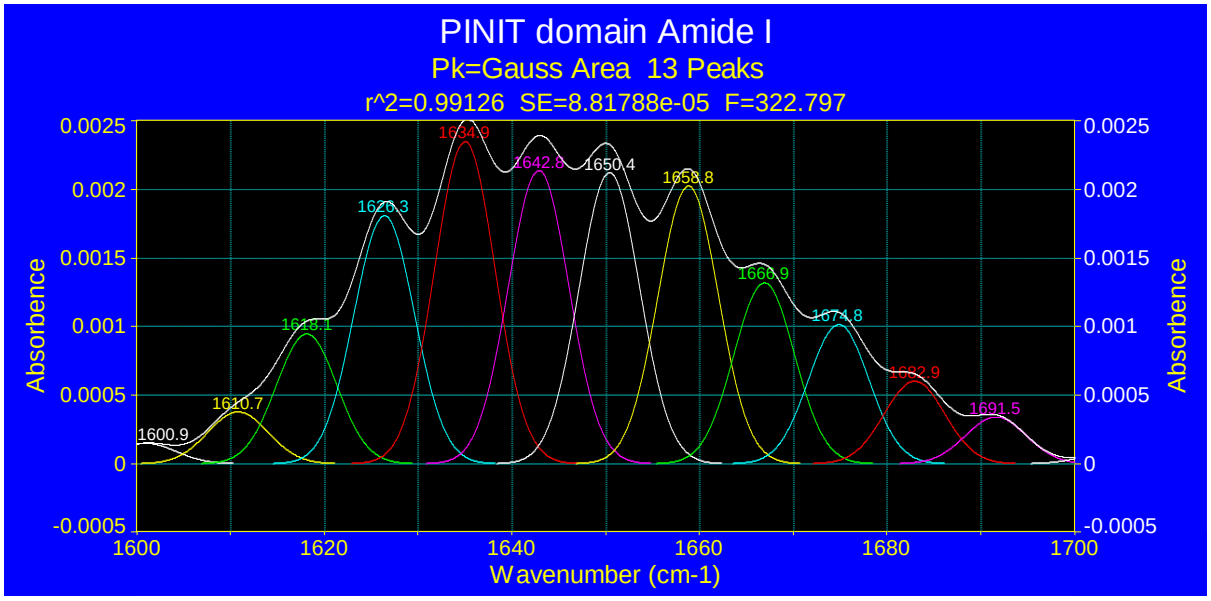

Figure 4.15 Secondary structure analysis of $(His)_7$–PINIT protein in H_2O environment. The Amide 1 region of the infrared spectra of PINIT domain protein in H_2O environment was deconvoluted and the peaks fitted with the Gaussian curve. Gaussian bands are shown as symmetrical peaks underneath the deconvolved infrared (IR) spectra. The peak wavelength numbers are shown at the top of each peak.

The PINIT protein in the D_2O environment had overlapping bands arising from deuterated and the unexchanged population of the different extended structures, resulting in exposure of hidden bands at 1637 cm^{-1} (**Figure 4.16**). The observed Amide I bands within 1672 cm^{-1} to 1681 cm^{-1} in H_2O could be assigned to β–turns (Byler and Susi, 1986; Krimm and Bandekar, 1986). Large vibration frequency shifts were predicted for a given type of β–turn as dihedral angles can vary (Krimm and Bandekar, 1986). Bands at 1655 cm–1 and 1646 cm–1 could be assigned to the deuterated α–helix and random segments or loops, respectively.

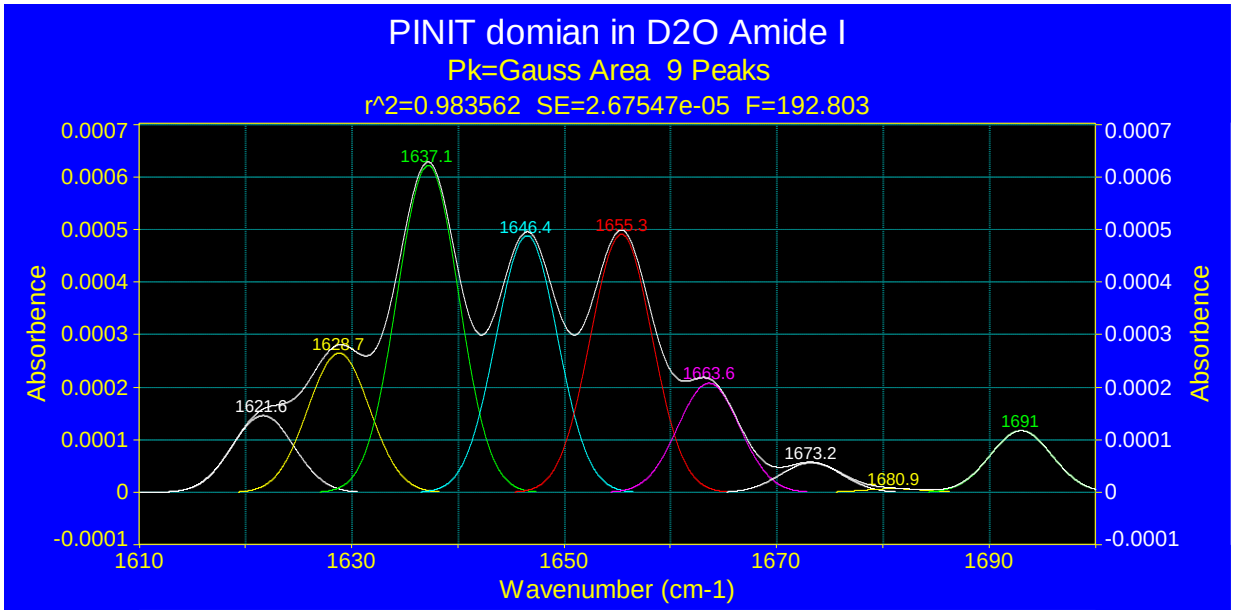

Figure 4.16 Secondary structure analysis of $(His)_7$–PINIT protein in D_2O environment. Amide I region infrared spectra of $(His)_7$–PINIT protein in D_2O environment was deconvoluted and the peaks fitted with the Gaussian curve Gaussian bands are shown as symmetrical peaks underneath the deconvolved IR spectra. The peak wavelength numbers are shown on each peak.

Table 4.2 Fractional band areas (% Area), Frequencies (wave number) and band assignments of FTIR Amide I component bands of native PINIT domain protein in H_2O and 20 % D_2O environment.

% Area	Wave number (cm^{-1}) H_2O	Assignment	% Area	Wave number (cm^{-1}) D_2O	Assignments
12.1	1625.1	β-sheet	6.1	1621.7	β-sheet
16.9	1634.7	β-sheet	11.0	1628.7	β-sheet
17.9	1644.0	β-sheet	25.9	1637.1	β-sheet
15.9	1653.3	α-Helix	20.3	1646.4	β-sheet
13.2	1662.9	3_{10}-Helix	20.4	1655.3	α-Helix
8.6	1672.8	β-Turn	8.7	1663.6	β-Turn
4.6	1681.2	β-Turn	2.4	1673.2	β-sheet
2.1	1691.5	β-sheet	0.3	1680.9	β-sheet
			4.9	1689.3	β-Turn

The fractional areas of the Amide I component bands are directly proportional to the relative content of the secondary structure types yielding them (Byler and Susi, 1986). The positions and the percentage of band areas for the different fitted bands from the Gaussian curve fit results in H_2O and D_2O (**Table 4.2**). The PINIT domain in H_2O solution; 49 % of the amide C=O groups are involved in β–sheets, 29 % α–helical, 14 % in β–turns and 8 % should be

random or unordered and irregular structures. Quantification in D_2O yields more realistic results because of deuteration shifts allowing identification of other structures. The β–sheets are estimated to 66 %, and α–helical are quantified to be 20.4 % and 13.6 % β–turns.

The generated PINIT model (**Figure 4.17A**) and its systematic representation of the secondary structures (**Figure 4.17B**) showed that the bulk part of the secondary structure features are β–sheets and the D_2O results (**Table 4.2**) indicated that 66 % of the amide C=O groups are involved in β–sheets.

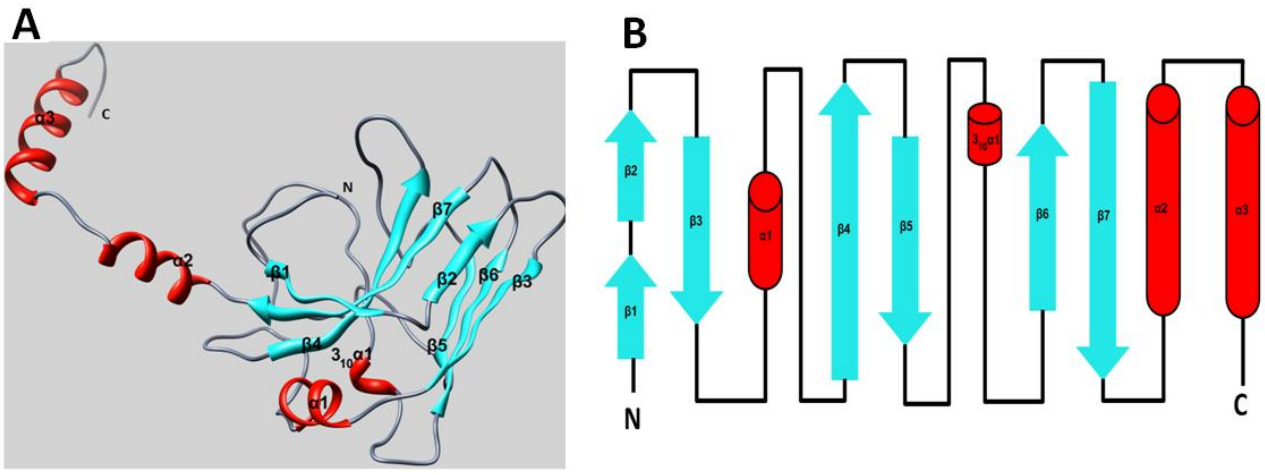

Figure 4.17 Cartoon representation of the PINIT domain model and the systematic secondary structure representation. A) The model of the PINIT domain generated from the X–ray structure of the PINIT domain of Siz1 (PDB ID: 3i2D) showing the two antiparallel β sheets; one includes β1, β2, β4, and β7, and the other includes β3, β5, and β6. The β sheets are connected by protruding loops (L1, L2, and L3) that join strands β2–3, β4–5, and β6–7 at one end of the molecule, while β3–4 and β5–7 are connected by a helix α1 and a loop, respectively, on the opposite surface. **B)** The systematic representation of the flow of the secondary structure features of the generated PINIT domain model.

4.3.11 Preliminary characterisation of the PINIT–STAT3 interaction

Preliminary characterisation of PINIT–STAT3 interaction was performed by dot blot association assay before employing sensitive methods like SPR. The $(His)_7$–PINIT protein was found to interact with varying amounts of recombinant STAT3 (50 µg – 500 µg) blotted onto nitrocellulose membrane. The interaction was concentration dependent as indicated by the increasing intensity of the signal (**Figure 4.18**). When the signal intensity was plotted against the amount of STAT3, a typical saturation curve was generated, suggesting that interaction was potentially specific. The saturation point was reached at 200 µg of STAT3 (S_3), with no further PINIT protein binding evident with increasing STAT3. The absence of any detectable interaction of PINIT with BSA protein also suggested that PINIT bound to STAT3 in a specific manner.

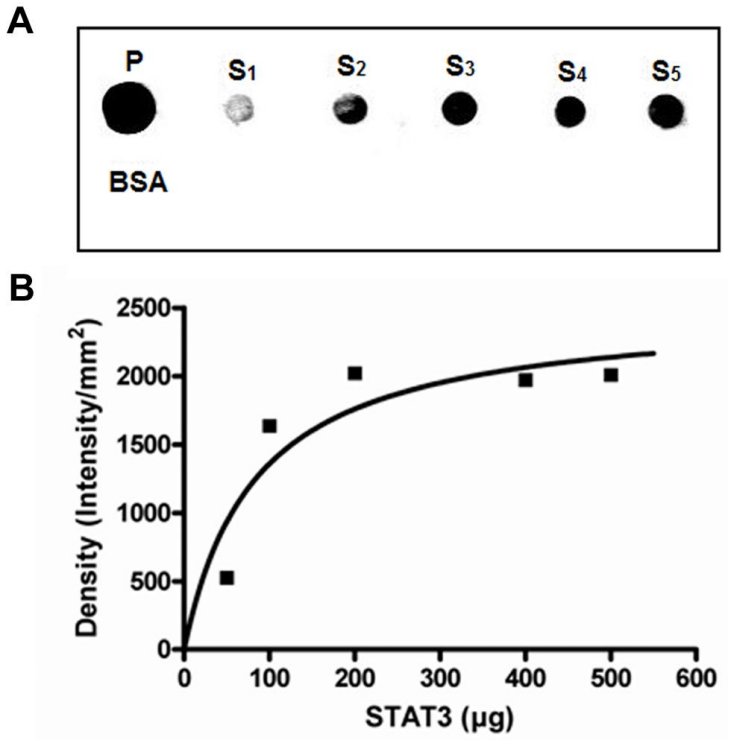

Figure 4.18 (His)$_7$–PINIT interaction with STAT3 protein at increasing concentration of STAT3. A) Dots blot assay; STAT3 was vacum blotted onto a nitrocellulose membrane, incubated with (His)$_7$–PINIT (100 µg/ml), and the bound (His)$_7$–PINIT detected by western blot analysis using mouse monoclonal anti–His primary antibody and HRP–conjugated goat anti–mouse IgG antibodies; P, (His)$_7$–PINIT (40 µg); STAT3: S_1 = 50 µg; S_2 = 100 µg; S_3 = 200 µg; S_4 = 400 µg; S_5 = 500 µg; and BSA = 600 µg. **B**) **Graphical representation of the dot blot assay; t**he intensity of (His)$_7$–PINIT bound to STAT3 detected by western blot analysis was plotted against increasing amounts of STAT3 on the nitrocellulose membrane. This experiment was repeated three times giving similar results; a typical example is shown here.

4.4 DISCUSSION AND CONCLUSION

The successful expression and purification of (His)$_7$–PINIT protein from *E.coli* XL1 Blue [pQE2–PINIT] was achieved. Also, (His)$_7$–PINIT protein size and secondary structure content were characterised and preliminary functional investigation showed that the (His)$_7$–PINIT is folded and functional as it was capable of associating with STAT3.

Expression of the PINIT domain of a mammalian PIAS3 protein in a prokaryote system required a carefully designed strategy. This was because of the existence of rare codons that were likely to cause early translation termination and or misfolding of the protein. The problem was overcome through trying various expression vectors and combination of fusion tags and expression strains that co–express the tRNA for those rare codons (**Table 4.2**). Although GST–tag was suggested to improve the protein folding (Kaplan *et al*., 1997) the folding of the PINIT domain was affected by the GST–tag. Two species of the GST–tagged

protein were observed suggesting PINIT protein degradation. The double tagging of the protein was suggested to improve the protein yield and enhances its solubility as well as facilitating its purification (Pryor and Leiting, 1997). However, the GST–tag and His–tag did not successfully over–expressed GST–PINIT–$(His)_6$ protein. Probably, the size of the PINIT domain greatly affected the folding because GST–tag and PINIT domain had molecular mass of 26 kDa and 23 kDa respectively. His–tag is small and was suggested to enhance protein expression. Although the expression was successful, the position of the His–tag was critical for the expression of the PINIT protein. pQE expression vector systems with T5 promoter containing two *lac* operator sequences for tight regulation were used. The difference was that pQE2 had the N–terminal His–tag and pQE60 had a C–terminal His–tag. Successful overexpression was achieved by use of the pQE2 vector with an N–terminal His–tag.

ATP based nickel affinity column and a two stage size exclusion chromatography was determined as the standard $(His)_7$–PINIT purification procedure to achieve high protein purity. *E. coli* contaminating proteins which were persistently co–purifying with $(His)_7$–PINIT were identified by mass spectroscopy and western blot analysis as *E. coli* chaperone proteins which have higher affinity for ATP. The use of ATP based buffers in nickel affinity column purification eliminated most of the contaminating proteins. Also, size exclusion chromatography with ATP based buffers further facilitated the separation of the remaining contaminating protein. A second stage of the size exclusion was necessary to characterise the size and further purify the $(His)_7$–PINIT protein. Due to the rigorous purification stages, $(His)_7$–PINIT protein degraded to a low yield (0.6 mg per litre of original culture). For structural elucidation by either NMR or X–ray crystallography, uniformly purified protein of at least 10 mg was required.

However, FTIR was used to characterise the secondary structure features of the PINIT domain and determined its folding state by hydrogen–deuterium exchange. Approximately 66 % of the secondary structure features of $(His)_7$–PINIT domain were β–sheets consistent with the homology model generated from the PINIT domain of Siz1 (**Figure 4.17A** and **4.17B**). Hydrogen–deuterium exchange of the amide bonds showed that the protein was folded because a large reduction in the intensity of the Amide II band in D_2O due to the hydrogen–deuterium exchange of the amide bonds resulted in the shift of the amide II band towards lower frequencies (approximately 1455 cm^{-1}). This was because buried hydrophobic pockets of β–sheets amino acids in Amide I band were not easily accessible by D_2O for deuterium exchange compared to amide II bands primarily from N–H stretching vibrations (**Figure**

4.14). Furthermore, the folding of the $(His)_7$–PINIT was confirmed by its ability to associate with STAT3. Dot blot association assay revealed that the purified protein can bind specifically to STAT3 protein and saturate the STAT3 binding site or in a concentration dependent manner (**Figure 4.18**).

To conclude, the PINIT domain of the PIAS3 was successfully heterologously over-expressed at 20^{o}C in *E. coli* XL 1 Blue expression strain. The expression and purification was achieved by use of the N–terminal $(His)_7$–tag and the purity of the $(His)_7$–PINIT protein was improved by use of ATP containing washbuffer to remove contaminating proteins of *E.coli* origin. Furthermore, biophysical and biochemical techniques were employed to characterise the size, secondary structure content and its association with STAT3 protein. The results showed successful production of a functional and folded protein and this set a platform for further investigation of PINIT–STAT3 interaction.

CHAPTER 5

THE *IN VITRO* ANALYSIS OF THE PIAS3–STAT3 INTERACTION: ROLE OF THE PINIT DOMAIN

Protein–protein interactions are essential for the functioning of living cells. Biomolecular interactions can be visualised in real time, using the principle of surface plasmon resonance spectroscopy on a BIAcore instrument, and kinetic rate and affinity constants can be determined. In this study SPR spectroscopy has been applied to the interaction of the PINIT domain with recombinant STAT3 protein to determine kinetic and affinity constants. Based on previous studies and structural bioinformatic analysis (Chapter 2), mutants were generated of the PINIT domain which were further characterised based on molecular mass and structural integrity. Furthermore, the significance of the replaced residues in PINIT mutants was evaluated by PINIT mutant–STAT3 interaction analysis.

5.1 INTRODUCTION

The basic concept of SPR is the detection of biospecific adsorption of an analyte due to changes in reflective index close to the chip surface, in continuous flow, to an immobilized ligand at the same spot and at the same time as it occurs, i.e. real–time interaction kinetics (Liedberg *et al*., 1983). SPR technology is an efficient and sensitive method to analyse binding kinetics of an interaction (Rich and Myzka, 2000; 2004). It allows the measurement of analyte association rate (k_{on}) and dissociation rate (k_{off}), which are indicators for the degree of recognition and binding stability of the ligand–analyte interaction. The ratio of the two is a measure of the affinity or binding strength of the analyte for the ligand ($K_D = k_{off} / k_{on}$) meaning that a high affinity can be caused by a high association rate or slow dissociation rate (Jönsson *et al*., 1991; Myszka, 1997). The basic multistep protocol for examining molecular interactions may be described by the following four steps: (1) immobilisation or capture of a ligand; (2) injection of the test analyte and real–time recording of an interaction curve; (3) step 2 is repeated with increasing concentrations of test analyte; and (4) the data fit is performed on the collected sensorgrams for determination of rate constants (Beseničar *et al*., 2006). The association and dissociation of the analyte to the immobilised ligand are followed in real–time and presented in collected sensorgrams of response versus time. The response signal from the SPR detector is proportional to the mass of protein per surface area (Stenberg *et al*., 1991). The advantages of SPR over other techniques are: direct and rapid determination of association and dissociation rates of the binding process; no need for labelling of protein; and small amount of sample used in the assay (often nanomolar concentrations of protein) (Beseničar *et al*., 2006). Assay types performed using the BIAcore are: binding specificity, i.e. qualitative studies to confirm the specificity of interactions; and quantitative measurements for determination of affinity and kinetic rate constant evaluation. While kinetic association (k_a) and dissociation (k_d) rate constants from real–time measurements of binding interactions provide information regarding complex formation and complex stability, the rate constants provide a link between protein function and structure through the evaluation of the impact of amino acid substitution.

Characterisation of the PIAS3–STAT3 interaction by direct biophysical techniques such as SPR has not been done. Studies of protein–protein interaction using biophysical techniques are dependent on the production of folded, biologically active protein. In the previous chapter

(Chapter 4), the production of folded and functional wildtype PINIT domain was described. Here, the direct interaction with recombinant STAT3 will be shown in real–time by quantitative SPR spectroscopy for determination of kinetic association and dissociation rate constants based on the concentration dependency of interaction. Furthermore, mutant PINIT proteins were characterised by FTIR and used in a qualitative SPR assay to evaluate the significance of the substituted amino acids in the PINIT–STAT3 interaction.

5.2 METHODS

5.2.1 Preconcentration of STAT3 to determine optimal immobilization conditions

STAT3β used in this study was expressed and purified as previously described (section 4.2.9**).** Preconcentration analysis of STAT3β was performed at a flow rate of 10µl/min. Recombinant STAT3β (20µg/ml) was diluted in 10 mM sodium acetate buffer solutions at four pH values (pH 4.0, pH 4.5, pH 5.0, pH5.5) to determine a suitable pH for immobilisation. The optimal pH for immobilization should be 1 unit below the estimated pI of STAT3 in order to ensure that the immobilised ligand protein has a net positive charge. The pI and molecular mass of the protein were estimated using the pI/MW tool found at http://au.expasy.org/tools/pi_tool.html (**Appendix F1**). The pH that resulted in the maximum surface retention was used for immobilization.

5.2.2 Immobilisation of STAT3 on CM5 sensor chip

SPR was performed using a BIAcore X instrument (GE Healthcare, Sweden). STAT3β was immobilized on flow cell one (Fc1) of a primed CM5 sensor chip at approximately 8500 RU using amine coupling. The CM5 dextran matrix on the sensor chip surface was first activated with a 1:1 mixture of 0.1 M N–hydroxysuccinimide (NHS) and 0.4 M 1–ethyl–3–[dimethylaminopropyl] carbodiimide (EDC) to create reactive succinimide esters. Hepes–buffered saline (HBS; 10 mM HEPES, 150 mM NaCl, pH 7.4) was used as the running buffer at 25°C. Flow cells were activated for 7 minutes by injecting 35 µl of 1:1 NHS/EDC (0.1 M NHS and 0.4 M EDC). An aliquot of 10 µl of 20 µg/ml STAT3 protein in 10 mM sodium acetate, pH 4.5 was injected at a constant flow rate of 5 µl/min, followed by a 70 µl injection of ethanolamine (1.0 M; pH 8.5) to block any remaining activated groups on the surface. This method resulted in approximately 8 500 RU STAT3 immobilised. Flow cell two (FC2) was blocked with ethanolamine after NHS/EDC activation and used as an inline reference cell.

5.2.3 Quantitative analysis of PINIT–STAT3 protein binding by SPR spectroscopy

Injections of 10 µl of $(His)_7$–PINIT (5–25 µM) and the $(His)_7$–PINIT mutants were performed at a flow rate of 10 µl/min using HBS buffer as the running buffer for 60 s association and 60 s dissociation. All sensorgrams were collected at 25°C. Triplicate injections of $(His)_7$–PINIT were performed for each concentration to account for statistical variability. Kinetic evaluation of the data was performed based on the 1:1 Langmuir association for determination of the observed rate constant, k_{obs} and R_{eq}, the steady state binding level. Rate constants were calculated following linear regression fitting of k_{obs} versus concentration of analyte ([$(His)_7$–PINIT]) plots according to the equation, $k_{obs} = k_a.Conc_{analyte} + k_d$. The affinity constant (K_D) was calculated from the ratio of the dissociation (k_d) and the association rate constants (k_a), (i.e. $K_D = k_d/k_a$). Data and statistical analysis were performed using BIAevaluation 3.2 (BIAcore, Sweden) and Prism 5.03 software (Graphpad Software, USA).

5.2.4 Mutagenesis, expression and purification of the $(His)_7$–PINIT mutants

Site–directed mutagenesis was performed using the double stranded whole plasmid linear non–PCR amplification procedure (QuikChange mutagenesis kit; Stratagene). Complementary mutagenesis primers were designed for the introduction of single point mutations at L97A, R99N and R99Q mutation into the PINIT domain coding sequence (**Appendix G2**). The primers were synthesised by Integrated DNA Technologies (USA). Each mutagenesis reaction was comprised of 100 ng of pQE2–PINIT parental plasmid template, 2.5 µl of dimethyl sulfoxide (DMSO), 10 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP mix , 125 ng of the forward primer, 125 ng of the reverse primer, 5 µl of 10x *Pfu* DNA polymerase buffer (200 mM Tris–HCl [pH 8.8], 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$), and 1 U of *Pfu* DNA Polymerase and sterile distilled water to a final volume of 50 µl. Thermal cycling was allowed to proceed as follows: one cycle of denaturation (95°C for 30 seconds), 18 cycles of denaturation, annealing and extension (95°C for 30 seconds, 52°C for 60 seconds, 68°C for 5 minutes), one cycle of final extension (68°C for 7 minutes) and a 4°C hold at the end. Digestion of the parental pQE2–PINIT plasmid in the amplification product was achieved by the addition of 5U of *Dpn*I restriction endonuclease to the reaction mixture and incubation at 37°C overnight. Pre– and post–*Dpn*I samples were analysed by 0.8% TBE agarose gel electrophoresis (**Appendix D4**). An aliquot of 10 µl of post–*Dpn*I samples was transformed into *E. coli* JM109 supercompetent cells (Promega) for screening

purposes. Plasmid DNA was isolated from the resulting colonies and screened for the desired mutation by DNA sequencing (**Appendix D7**) using the designed forward sequencing primer (**Appendix H3**). The expression and purification of the $(His)_7$–PINIT domain mutants were performed as previously described (**Chapter 4** section **4.2.7** and **4.2.8**).

5.2.5 Molecular mass characterisation of the $(His)_7$–PINIT mutants by size exclusion chromatography

Size exclusion chromatography of the $(His)_7$–PINIT mutants were performed as previously described in section 4.2.10.

5.2.6 Structural and folding analysis of $(His)_7$–PINIT mutants by FTIR spectroscopy

Structural and folding analysis of the $(His)_7$–PINIT mutants and the FTIR spectra data processing and analysis were performed as previously described (**Chapter 4,** section 4.2.11 and 4.2.12).

5.2.7 Assessment of the importance of *in silico* predicted R97 and R99 residues by SPR

SPR was performed using BIAcore X (GE Healthcare, Sweden) as previously described (section 5.2.3) using 20 µM of each $(His)_7$–PINIT mutant. $(His)_7$–PINIT mutants were expressed, purified and quantified as previously described (Chapter 4).

5.3 RESULTS

5.3.1 Immobilisation of the STAT3 on CM5 sensor chip

Immobilization of STAT3 via covalent amine coupling was performed in sodium acetate pH 4.5. Although pH 4 resulted in a sharper gradient for preconcentration, pH 4.5 was selected to avoid any deleterious effects to STAT3 tertiary structure at low pH (**Figure 5.1**). Typically, the optimal immobilization pH is one unit below the pI; the predicted theoretical STAT3 pI was determined to be 6.85 (http://expasy.org/tools/pitool.html) (Gasteiger *et al*., 2005). Under these conditions the ligand displays a positive charge and is effectively preconcentrated into the negatively charged carboxymethyl dextran matrix. The unreacted esters are "blocked" with ethanolamine. Inline reference surfaces are prepared in the same manner, except that all carboxyls are blocked and no ligand was added. The final immobilisation level of STAT3 was calculated by subtracting the reference cell (Fc2) from Fc1; STAT3 was immobilised at 8511 RU (27045.2–18534.1 = 8511) (**Figure 5.2A** and **5.2B**).

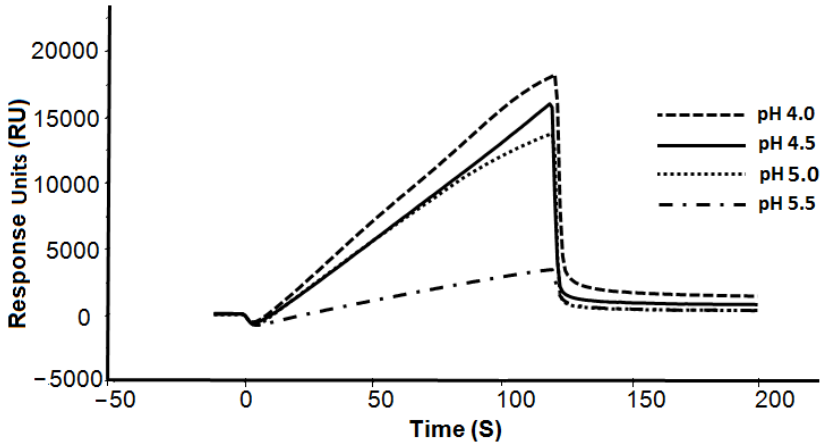

Figure 5.1 STAT3 preconcentration test to determine the optimum pH and concentration needed to reach a targeted level of response. STAT3 (20 µg/ml) in 10 mM sodium acetate buffer solution at pH 4.0 pH4.5 pH5.0 pH5.5 was passed over CM5 sensor chip at flow rate of 5 µl/minutes. The surface was regenerated by passing over 10 µl of glycine pH 1.5 followed by 20 µl of 0.05 % SDS at 5 µl/minute flow rate. **(RU, response units).**

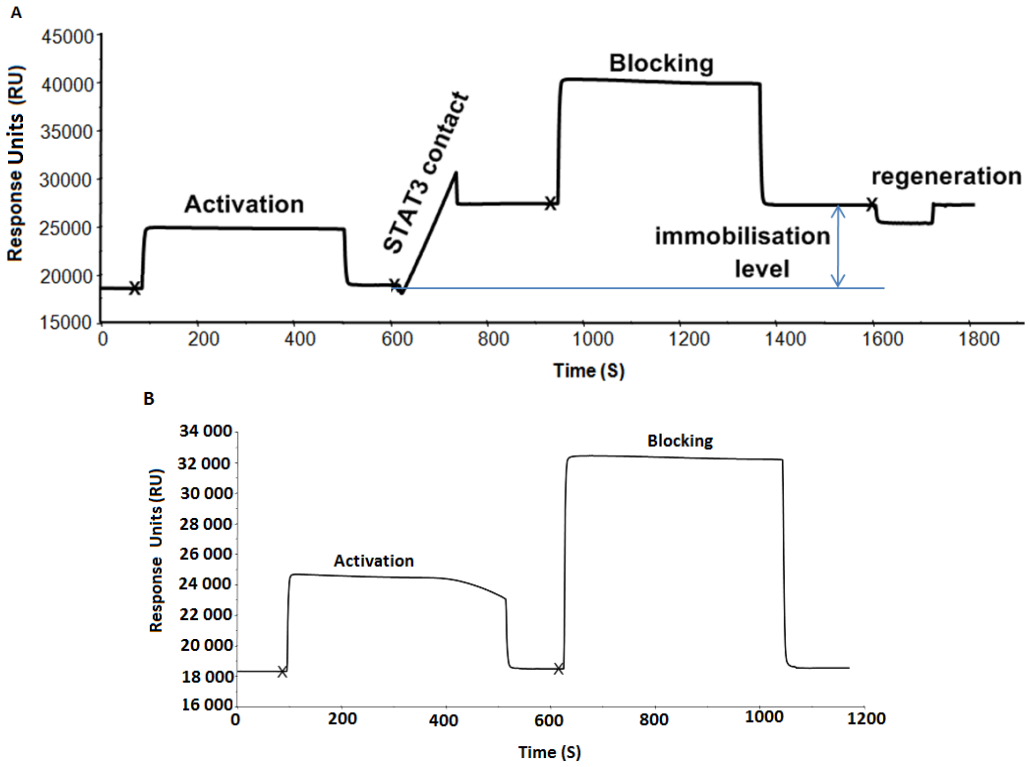

Figure 5.2 The immobilisation of STAT3 on the surface of the sensor chip. A) The immobilization of STAT3 was performed by activation of carboxymethyl groups on a dextran–coated chip by reaction with N–hydroxysuccinimide. This was followed by covalent bonding of STAT3 to the chip surface via amide linkages and excess activated carboxyls were blockaged with ethanolamine. The surface was regenerated with 10 mM glycine pH 1.5. **B**) The reference surface was prepared by activating the carboxymethyl groups by reacting with N–hydroxysuccinimide or 1–ethyl–3–(3–dimethylamino–propyl) carbodimide hydrochloride but no STAT3 was added and all carboxyls were blocked with ethanolamine. The x on the graph represents the injection points.

5.3.2 Quantitative analysis of PINIT–STAT3 interaction

Real time interaction analysis of $(His)_7$–PINIT binding to STAT3 was determined by injection of varying concentrations of $(His)_7$–PINIT (5 µM to 25 µM) over immobilised STAT3 (**Figure 5.3A**). The chip surface was regenerated by removal of $(His)_7$–PINIT with regeneration buffer (10 mM glycine, pH 1.5). Curve fits were performed under the assumption of an $A + B \leftrightarrow AB$ binding model to calculate association kinetics (A, represents the analyte; B, represents the immobilised ligand and AB, represents ligand–analyte complexes). The binding responses revealed a concentration dependency for the interaction of $(His)_7$–PINIT with STAT3 (**Figure 5.3A**).

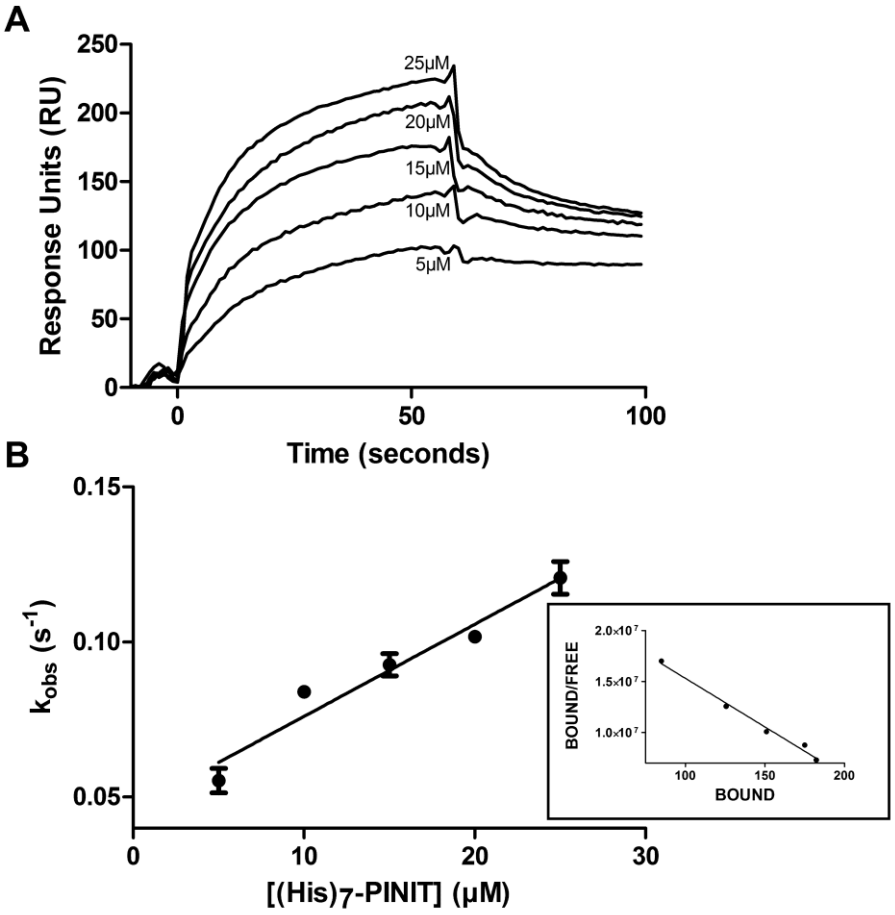

Figure 5.3 SPR analysis of $(His)_7$–PINIT–STAT3 interaction. A) SPR analysis of $(His)_7$–PINIT interaction with immobilised STAT3. Representative sensorgrams were obtained by injecting 10 µl of $(His)_7$–PINIT (5, 10, 15, 20 and 25 µM) at a flow rate of 10 µl/min. **B)** Linear plot of k_{obs} vs [$(His)_7$–PINIT] allowed for determination (see text for details) of kinetic rate and affinity constants, k_a ($2.97 \times 10^3 \pm 268.9$ $M^{-1}.s^{-1}$) and k_d ($0.046 \pm 0.0045.s^{-1}$) and K_D (15.7 ± 0.2 µM). The inset shows that the binding follows 1:1 Langmuir binding kinetics after data transformation via Scatchard plot. (adapted from Mautsa *et al*., 2010).

The apparent affinity constant (K_D) for $(His)_7$–PINIT–STAT3 interaction was found to be 15.7 ± 0.2 µM (n=3) as calculated from k_a and k_d values obtained from plotting k_{obs} vs [$(His)_7$–PINIT] (see **Figure 5.3B).** Linear transformation of data and analysis via Scatchard plot (see **Figure 5.3B** INSET, Bound/Free vs Bound, where Bound is R_{eq} and Free is [$(His)_7$–PINIT]) confirmed $(His)_7$–PINIT–STAT3 interaction follows 1:1 Langmuir binding kinetics. Also, the linear association kinetics and the concentration range of the analyte $(His)_7$–PINIT strongly suggest that the simple binding model is applicable to describe the interaction between PINIT and STAT3.

5.3.3 Expression and purification of the $(His)_7$–PINIT mutants

The PINIT domain mutants were produced to elucidate the function of highly conserved residues (L97 and R99). Based on previous knowledge of conserved residues, homology modeling, and previous mutational studies (Duval *et al*., 2003; Levy *et al*., 2006; Yagil *et al*., 2009), mutants of $(His)_7$–PINIT were generated by substitutions as described in the methodology (section 5.2.3). Expression studies were performed at the same conditions as the $(His)_7$–PINIT protein. $(His)_7$–PINIT–L97A, $(His)_7$–PINIT–R99N and $(His)_7$–PINIT–R99Q mutant proteins production in *E. coli* XL1 Blue [pQE2–PINIT] was evident one hour post induction with the maximum expression levels at overnight post induction (**Figure 5.4, A–C**). Although the highest expression levels were observed overnight, SDS–resistant higher order molecular mass species of $(His)_7$–PINIT were evident on western blot analysis (data not shown). Therefore a four hour post induction expression was considered for purification and characterisation because single species of $(His)_7$–PINIT protein was observed and the expression level was considered sufficient. The expressions of $(His)_7$–PINIT domain mutants were consistent with the wild type $(His)_7$–PINIT domain. PINIT domain mutants were purified in the same manner as the wildtype PINIT domain by ATP based nickel affinity column and a two stage size exclusion chromatography. The size exclusion chromatogram of the mutants showed protein elution at 17 to 19 ml elution volume (**Appendix B9**). Elution fractions of individual mutant $(His)_7$–PINIT after the second size exclusion were concentrated by Amicon^R Ultra Ultracel^R (10K) Centrifugal Filters (Millipore, Ireland) and analysed by 12% SDS–PAGE (**Figure 5.4D**). The protein yield of the mutants, cultured in one litre of LB broth, respectively, was determined: $(His)_7$–PINIT–L97A, 0.71 mg/litre of original culture; $(His)_7$–PINIT–R99N, 0.66 mg/litre of original culture; and $(His)_7$–PINIT–R99Q, 0.76 mg/litre of original culture.

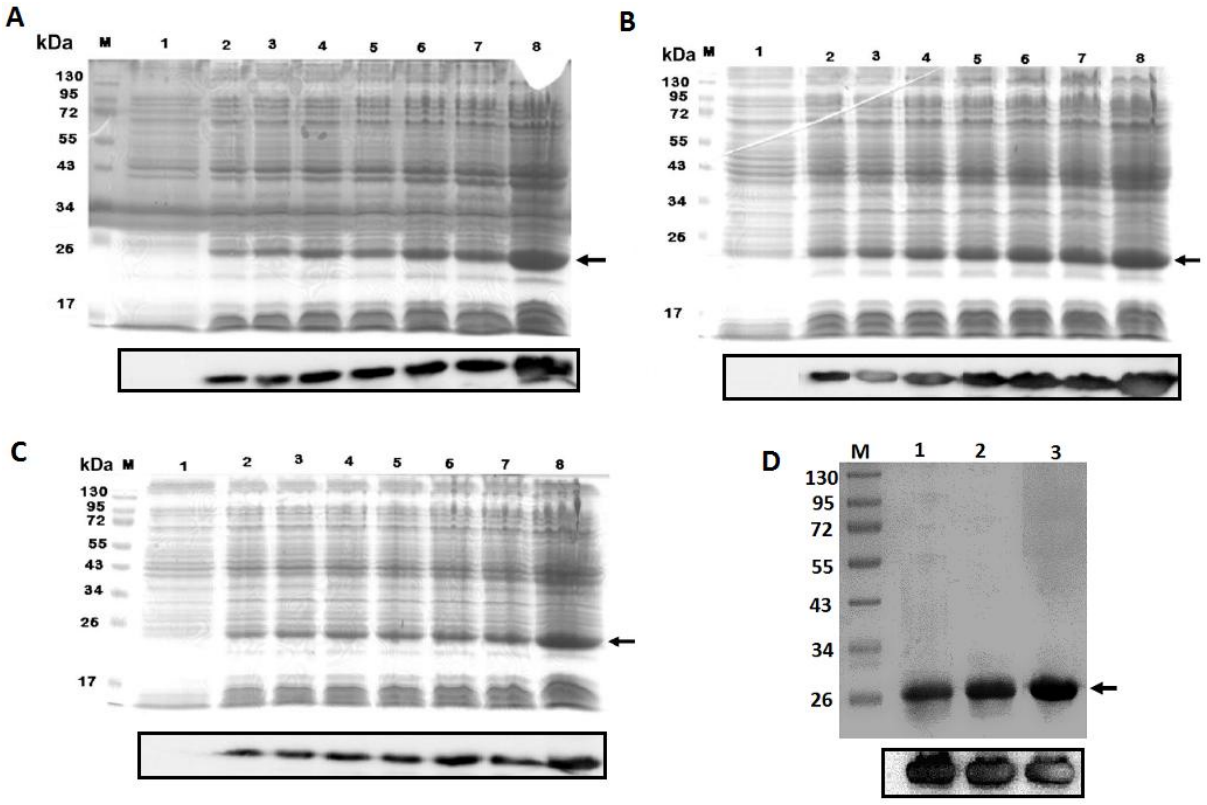

Figure 5.4 Production of recombinant PINIT domain mutants. A) upper panel, **12 % SDS–PAGE gel analysis of $(His)_7$–PINIT–L97A containing total protein extracts prepared from *E. coli* XL1 Blue[pQE2–PINIT–L97A]:** Molecular mass markers are indicated on left hand side M; lane 1, non–induced sample; lanes 2 – 7, induced samples using IPTG taken hourly; and Lane 8, overnight 16 hours induction sample. Lower panel, Western blot analysis of $(His)_7$–PINIT–L97A corresponding to the lanes of the upper panel $(His)_7$–PINIT–L97A was detected by western blot analysis using mouse monoclonal anti–His primary antibodies (GE Healthcare) and HRP–conjugated anti–mouse secondary antibodies. **B**) upper panel, **12 % SDS–PAGE gel analysis of $(His)_7$–PINIT–R99N containing total protein extracts prepared from *E. coli* XL1 Blue[pQE2–PINIT–R99N]:** Molecular mass markers are indicated on left hand side M; lane 1, non–induced sample; lanes 2 – 7, induced samples using IPTG taken hourly; and Lane 8, overnight 16 hours induction sample. Lower panel, Western blot analysis of $(His)_7$–PINIT–R99N corresponding to the lanes of the upper panel $(His)_7$–PINIT–R99N was detected by western blot analysis using mouse monoclonal anti–His primary antibodies (GE Healthcare) and HRP–conjugated anti–mouse secondary antibodies. **C**) upper panel, **12 % SDS–PAGE gel analysis of $(His)_7$–PINIT–R99Q containing total protein extracts prepared from *E. coli* XL1 Blue[pQE2–PINIT–R99Q]:** Molecular mass markers are indicated on left hand side M; lane 1, non–induced sample; lanes 2 – 7, induced samples using IPTG taken hourly; and Lane 8, overnight 16 hours induction sample. Lower panel, Western blot analysis of $(His)_7$–PINIT–R99Q corresponding to the lanes of the upper panel $(His)_7$–PINIT–R99Q was detected by western blot analysis using mouse monoclonal anti–His primary antibodies (GE Healthcare) and HRP–conjugated anti–mouse secondary antibodies. **D**) Upper panel–12 % SDS–PAGE gel analysis of purified mutants of $(His)_7$–PINIT: Molecular mass markers are indicated on left hand side M; lane1 $(His)_7$–PINIT–L97A; lane 2, $(His)_7$–PINIT–R99N; lane 3, $(His)_7$–PINIT–R99Q. Lower panel, Western blot analysis of mutants of $(His)_7$–PINIT corresponding to the lanes of the upper panel were detected by western blot analysis using mouse monoclonal anti–His primary antibodies (GE Healthcare) and HRP–conjugated anti–mouse secondary antibodies. Arrows indicate the position of the mutants of $(His)_7$–PINIT.

5.3.4 Structural and folding analysis of mutant $(His)_7$–PINIT proteins by FTIR spectroscopy

FTIR analysis of the mutants $(His)_7$–PINIT in comparison with the wildtype (His)–PINIT revealed similar secondary structure features. The percentage areas occupied by specific secondary structure features were the same as in all mutants of $(His)_7$–PINIT (**Appendix C3**), furthermore, infrared spectra of the wild type $(His)_7$–PINIT and mutants, ($(His)_7$–PINIT–L97A, $(His)_7$–PINIT–R99N and $(His)_7$–PINIT–R99Q) (**Figure 5.5**) revealed similar folds and peaks. Assignment of the frequencies showed that the three proteins had similar wave numbers and assignments and the percentage area occupied by β–sheets was approximately the same (**Appendix C2**). The data on percentage of areas occupied by the secondary structure features did not suggest conformational changes caused by the substitution mutations. The FTIR data revealed no shift in areas occupied by β–sheets and α–helices in both wildtype PINIT domain and the mutant PINIT domains.

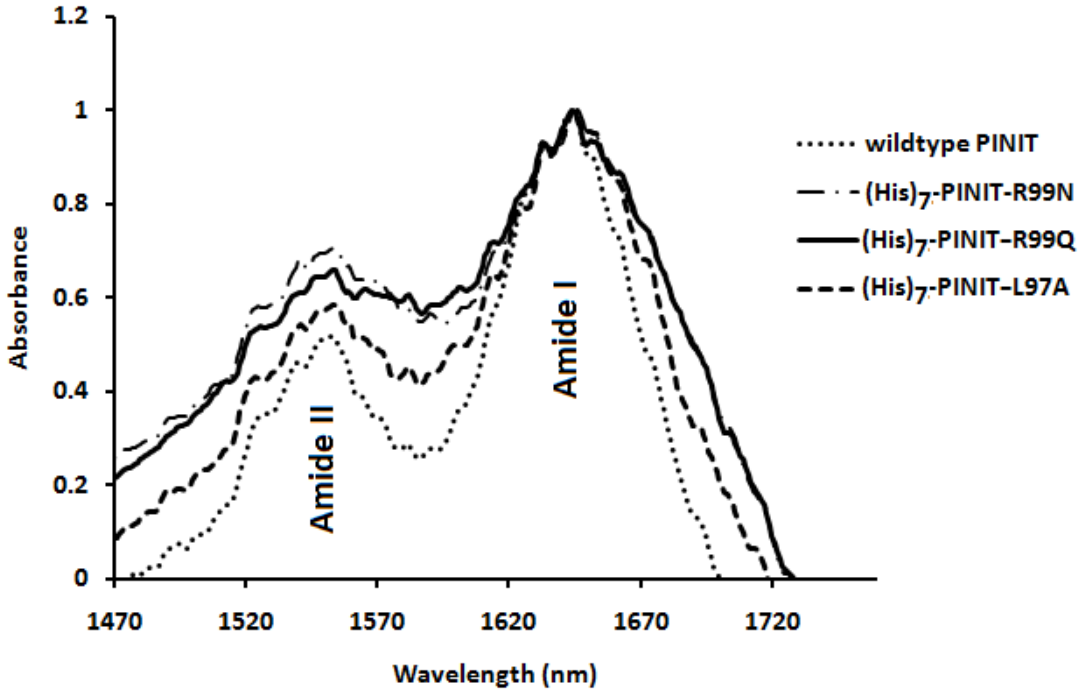

Figure 5.5 Spectra analysis of the mutants $(His)_7$–PINIT. Absorbance spectra comparison of wildtype $(His)_7$–PINIT and its mutant derivates. The spectra similarity showed similar fold and secondary structure contents between the wildtype and mutants of $(His)_7$–PINIT. All spectra were recorded under the same conditions and settings.

5.3.5 The importance of L97 and R99 residues of the PINIT domain on PINIT–STAT3 interactions

STAT3 was immobilized on the sensor chip and wild type $(His)_7$–PINIT and its mutants were passed over the chip at the same concentration (20 µM). Comparison of the response curves for the mutant domains versus the wild type domain indicated that there was limited binding by PINIT–R99N and PINIT–L97A mutants and no binding, but rather an apparent bulk shift response for the PINIT–R99Q mutant (**Figure 5.6**). Fourier transform infra–red spectroscopy indicated that the three mutants and the wild type $(His)_7$–PINIT had superimposable spectra, and therefore were structurally similar (**Figure 5.5**). The *in silico* analysis of these residues showed that the R99 residue was surface exposed (Chapter 2 **Figure 2.8C**) whereas the L97 residue was buried in the hydrophobic pocket (Chapter 2 **Figure 2.8C**). The sequence conservation and structural analysis revealed the potential involvement of these residues in protein–protein interaction. *In vitro* analysis of these mutant proteins by SPR revealed that these residues were potentially critical for interaction of the PIAS3 PINIT domain with STAT3.

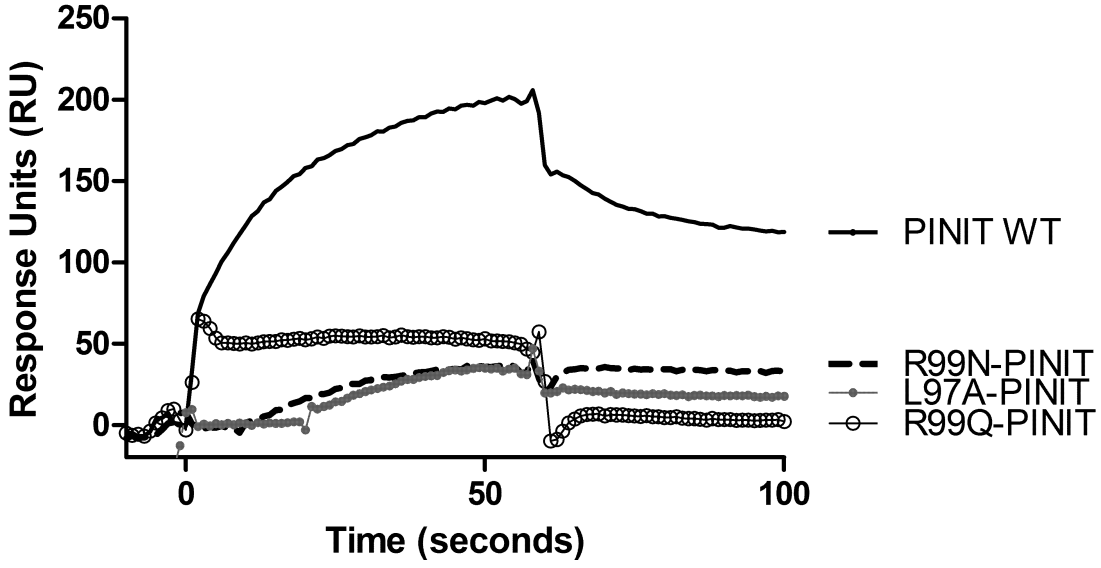

Figure 5.6 SPR analysis of the interaction of wildtype $(His)_7$–PINIT and $(His)_7$–PINIT mutants with STAT3. SPR analysis of the interaction of wild type $(His)_7$–PINIT and its mutant derivatives (PINIT–R99N, PINIT–R99Q and PINIT–L97A) with immobilized STAT3. The curves for a single concentration of the $(His)_7$–PINIT proteins (20 µM) are shown. The residues L97 and R99 are critical for the ability of the PINIT domain to interact with STAT3.

5.4 DISCUSSION AND CONCLUSION

A quantitative assessment of the kinetics of PINIT–STAT3 interaction provided evidence for the function of the PINIT domain in the interaction mechanism of PIAS3 with STAT3 protein. The study showed that biomolecular interaction analysis using SPR was a suitable technique for the analytical description of PINIT–STAT3 interaction. The binding kinetics observed strongly suggest that the interaction at the sensor surface can be sufficiently described by a 1:1 molar interaction model as shown by the Scatchard linear plot (**Figure 5.3B**). The SPR analysis confirmed the preliminary evidence of the biochemical function of the purified $(His)_7$–PINIT performed by dot blot association assay (**Chapter 4**). Both experiments revealed the concentration dependency of the PINIT–STAT3 interaction.

Using rational site directed mutagenesis, mutant PINIT proteins were produced, quantified and biochemically characterised in the same manner as wild type $(His)_7$–PINIT (**Chapter 4**). The expression of the mutant $(His)_7$–PINIT protein (**Figure 5.4**) and the yield per litre of culture was in the range of 0.6 to 0.7 mg, which was consistent with the wild type $(His)_7$–PINIT (0.6 mg/litre of culture).

An investigation of conformational changes in the mutant $(His)_7$–PINIT proteins in aqueous solution revealed predominantly β–sheet proteins and peak assignments consistent with that of wild type $(His)_7$–PINIT. If there was significant conformational change due to mutation, evidence of drastic reduction of the β–sheet content would be detected by change in percentage area occupied by the β–sheet. FTIR of the PINIT domain and its mutants in H_2O environment revealed that the protein had approximately 45 % β–sheet content (**Appendix C3**). Mutations performed were on the loop between the β1 and β2 sheet (**Chapter 2. Figure 2.8C**) and spectra analysis showed no effect on the sheet content. Structural analysis (chapter 2) and FTIR analysis found that the L97A, R99N and R99Q mutations did not disrupt the overall tertiary structure of the PINIT domain, suggesting that any functional effects of the mutations were not an indirect effect of misfolding.

SPR analysis of the predicted residues on PINIT–STAT3 interaction revealed that the L97A, R99N and R99Q mutations resulted in abrogation of PINIT domain binding to STAT3 suggesting that L97 and R99 were directly involved in binding. The abrogation of PINIT–R99N and PINIT–R99Q interaction with STAT3 was possibly due to the reduced bulk and lack of charge on the Asn and Gln residues compared to Arg. Loss of functionality for PINIT–L97A mutant was possibly due to substitution with Ala, a small residue compared to

Leu which resulted in a more compact packing conformation causing a change in orientation of the neighbouring R99 residue.

In silico analysis of the PINIT domain was described in Chapter 2 and revealed that L97 and R99 residues were highly conserved and indicated potential importance in PIAS3 structure and function contributed to PINIT–STAT3 interaction. Furthermore, the homology model of the PINIT domain showed that the positively charged R99 residue of the PINIT domain was predicted to be surface exposed (**Chapter 2 Figure 2.8C**), and thus potentially able to form electrostatic contacts with, as of yet unidentified, negatively charge residues on STAT3.

Previous mutational studies within the same region, in particular, PIAS3–Y94P mutant protein, has been found to not associate with STAT3 (Levy *et al*., 2006). However, its functional effects were probably the result of disruption in secondary structure and hence misfolding of PIAS3. The Tyr residue is shown on the β–sheet (**Figure 2.8C**) and therefore, its substitution with Pro may results in secondary structure disruption possible because Pro does not fit into the regular part of either helix or sheet structures due to its lack of backbone–NH.

PINIT–STAT3 interaction using SPR has been shown for the first time and these data validated the previous *in vivo* findings that the PINIT domain alone can interact with STAT3 (Levy *et al*., 2006; Yagil *et al*., 2009). Furthermore, structure function analysis using point mutations suggest that the PINIT domain ($PIAS3_{85–272}$) is potentially a major determinant in PIAS3–STAT3 interaction. In particular the N–terminal region β1–loop–β2 region of PINIT domain (**Chapter 2, Figure 2.8**) which was also a major focus by previous researchers.

CHAPTER 6

CONCLUSIONS AND FUTURE PROSPECTS

The study attempted to address the knowledge gap of PIAS3–STAT3 interaction. The work presented here focused on a previously identified critical domain, PINIT, that was capable of performing the same function as the full length PIAS3 and attempted to address an essential question of the requirements of the PIAS3 interface for PIAS3–STAT3 interaction. This chapter critically discusses the in silico; in vivo and in vitro findings of PIAS3–STAT3 interaction described in previous chapters and suggests future prospects for study.

6.1 CONCLUSIONS

The research findings detailed in this study have for the first time shown the following: a PINIT domain homology model based on the PINIT domain structure; the heterologous overexpression and purification of a folded functional recombinant PINIT domain; biophysical characterisation of the PINIT-STAT3 interactions using SPR; and the mutation of some of the residues critical for PINIT-STAT3 interaction and potentially PIAS3-STAT3 interaction (Mautsa *et al*., 2010). Furthermore, this study has shown for the first time the localisation effect of the critical residues that determine the PINIT-STAT3 interaction on the full length PIAS3 and also the localisation effect of the individual domains of PIAS3.

The work presented here focused on the structural and functional characterisation of PIAS3 *in silico, in vivo* and *in vitro. In silico* PIAS3 analysis and the key findings were used to guide the experimental design strategies. In particular, the determination of PINIT domain (M85–V272) boundaries based on the conserved regions and predicted secondary structure features. The predicted PINIT domain homology model showed the orientation of residues in three–dimensional space. Of particular interest was the N–terminus β–sheet–loop–β–sheet peptide (**Figure 2.8C**) of the generated PINIT domain model. The region was previously suggested as the minimum fragment that could perform the same function as the PIAS3 (Levy *et al*., 2006; Yagil *et al*., 2009). Based on the *in silico* analysis discussed in Chapter 2, L97 and R99 residues were predicted to be critical for PIAS3 structure and function. It should be noted that this prediction was based on the model that was generated using low target–template sequence identity (PINIT domain of PIAS3 and Siz1 respectively). Nevertheless, the secondary structure prediction showed that the fold was conserved across species, in this case mammalian (target) and yeast (template). Also, L97A, R99N and R99Q mutations were predicted to have no structural effect on the PINIT domain. Therefore, the model was validated and adopted for further *in vivo* and *in vitro* structure–function analysis. The *in silico* predictions were first evaluated *in vivo* using HeLa human cervical cancer cells. The localisation effect of individual domains in comparison to the full length was evaluated under IL–6 stimulation. PIAS3 was completely localised in the nucleus in all cells while the PINIT domain was predominantly perinuclear and nuclear localised in most cells with some in the cytoplasm in all cells. The acidic domain was predominantly perinuclear localised in all cells and also found in the cytoplasm. Comparison of the localisation differences of full length

PIAS3 with PINIT and acidic domains suggested a degree of functional disruption; hinting at an undefined or incomplete (i.e. a multidomain NLS) nuclear localisation signal.

It has been suggested previously that the PINIT motif is necessary for nuclear localisation of the PIAS3 protein (Duval *et al*., 2003). However, the localisation differences of the PIAS3 and the PINIT domain in HeLa cells (**Figure 3.3** and **Figure 3.4**) suggests that the conserved PINIT motif may possibly be critical for structural stability of the PIAS3 protein. Co–localisation analysis of PIAS3, PINIT and acidic domains with STAT3 suggested good co–localisation of PIAS3 and PINIT domain with STAT3 compared to acidic domain co–localisation with STAT3. These findings were not conclusive because only a single cell line was used; also, HeLa cells were shown to constitutively express IL–6(Maleczyk *et al*., 1991, Eustace *et al*., 1993) suggesting autocrine activation of the JAK–STAT3 pathway. Use of an IL–6 inducible cell line would allow for a more dynamic picture of PIAS3–STAT3 association and trafficking as opposed to the largely static results obtained from the HeLa cells. Furthermore, it may be necessary to perform the study in a normal cell line in comparison to the cancer cell line because of differences in STAT3 regulation and PIAS3 expression levels.

Observed differences of localisation between the mutants (PIAS3–L97 and PIAS3–R99) and the wildtype PIAS3 suggested possible functional disruption due to the mutations. These residues were predicted to be directly involved in PIAS3–STAT3 interaction. The mutations potentially resulted in a loss of the non covalent contacts involved in the association of PIAS3 with STAT3. However, taken together, the *in silico* and *in vivo* findings, suggested the necessity to further investigate the PINIT domain and the residues L97 and R99 within the PINIT domain.

Guided by these findings, the main study focused on the heterologous over–production of the PINIT domain in a bacterial system. Previous approaches involved production of PIAS3 peptides in mammalian and yeast cells (Chung *et al* 1997; Levy *et al*., 2003; Yagil *et al*., 2009; Borghouts *et al*., 2010). Heterologous over–production of the PINIT domain in sufficient quantities allowed application of biophysical techniques to characterise the protein in terms of structure and function. Size exclusion chromatography showed that the protein existed as a monomer in solution. Although the PINIT domain protein was successful expressed, the yield typical for a one litre culture was low due to the number of purification stages performed to get a pure protein. Nevertheless, the protein yield was sufficient for the biophysical and biochemical functional characterisation. The over–expression was dependent

on the $(His)_7$–tag because other tags (GST-tag and His-tag) were attempted with no successful over–expression. PIAS3 structural determination by X–ray crystallography or NMR spectroscopy was limited by production of the protein, and in this study the production of the protein for structural study was limited by the loss of the protein due to rigorous purification stages. However, secondary structural content characterisation with FTIR was in agreement with the generated PINIT model in that, 66 % of the secondary structures were β–sheets.

Preliminary investigation of the purified PINIT domain protein showed that the protein was able to bind STAT3 in a specific concentration dependent manner. These findings formed the basis for further investigation of the PINIT–STAT3 interaction using SPR and also the evaluation of the importance of L97 and R99 residues on PINIT–STAT3 interaction. The quantitative assessment of the kinetics from SPR data suggested a strong PINIT–STAT3 interaction which followed 1:1 Langmuir binding kinetics. This suggested that the PINIT domain followed a typical mechanism of a classic biochemical inhibitor, i.e. binding to a single site on STAT3. PINIT–L97A, PINIT–R99N and PINIT–R99Q mutants were expressed and purified in the same manner as the wildtype PINIT domain. The mutations performed were predicted *in silico* and validated by FTIR to have no apparent structural effect on the PINIT domain. Therefore, it was expected that any loss of function was as a result of direct involvement of the L97 and R99 residues. The L97 and R99 residues were determined to be directly involved in the PIAS3–STAT3 interaction due to abrogation of PINIT–STAT3 interaction when mutated.

The research findings in this study, in the context of finding of others (Chung *et al*., 1997; Borghouts *et al*., 2010), suggested that two separate domains of PIAS3 (PINIT and acidic domains) possibly contribute to STAT3 binding. Nevertheless, it has been suggested that protein domain contributions to binding are often additive (Prinsloo *et al*., 2009). But this can only be addressed by conducting studies on *holo*–PIAS3–STAT3. Previous studies showed that mutation of the PINIT domain within full–length PIAS3 abrogated PIAS3–STAT3 interaction (Levy *et al*., 2006). This evidence was consistent with data from this study and also suggests that the PINIT domain is an important determinant of PIAS3–STAT3 interaction. Furthermore, this study was performed with unphosphorylated monomeric STAT3 and provided experimental evidence that the association of the PINIT domain of PIAS3 was not dependent on phosphorylation of STAT3. This calls for the evaluation of the notion that PIAS3 protein binds to phosphorylated STAT3 and translocates to the nucleus as

a complex where it dephosphorylates STAT3 and returns to the cytoplasm (Dabir *et al*., 2009).

The interaction of PIAS3 with STAT3 is critical in the regulation of aberrant activation of STAT3. To date, PIAS3 protein or PIAS family member structures have not been determined. *In silico* predictions which depend on the existence of appropriate template structures allow for *in silico* investigation of PIAS3 structure. The existence of the structure was critical to guide the design *of in vivo* and *in vitro* experiments. These studies were performed using generally similar molecular approaches of indirect interaction of PIAS3–STAT3. Also there was limited structural analysis and application of *in silico* tools to analyse PIAS3 structural functions. Although mutations performed resulted in loss of PIAS3 function, these mutations were based on random conserved residues in the full length PIAS3 rather than based on specified individual domains. Hence, various random binding site of PIAS3 were suggested by individual researchers (Levy *et al*., 2006; Yagil *et al*., 2009; Borghouts *et al*., 2010). The findings in this study could form the basis of chemotherapeutic drug design to inhibit STAT3 aberrant activation.

6.2 FUTURE PROSPECTS

If multiple domains of PIAS3 are involved in PIAS3–STAT3 interaction, detailed *in vitro* and *in vivo* studies on individual domains are required to elucidate the multiple molecular determinants of PIAS3 regulation of STAT3. It is important to assess the efficacy of the PINIT domain on suppression of STAT3 transcriptional activity in comparison to full length PIAS3. Also this would allow for the validation of L97 and R99 residues and the role they may play in PIAS3–STAT3 interaction. Furthermore, *in vivo* analysis of the functional effects of mutating L97 and R99 in the negative regulation of STAT3 could be assessed by performing luciferase promoter reporter assays. The luciferase reporter plasmid containing STAT3 promoter region could be co-transfected with the expression plasmid containing either PIAS3, mutant PIAS3 (L97, R99), PINIT or acidic domains and then measure the luciferase activity. Furthermore the effect of expression of the mutants PIAS3 and the PINIT and acidic domain on the cell cycle could be determined by analysis of DNA content using propidium iodide staining (Gazitt and Hu, 1998) and analysis by flow cytometry. These may answer questions of the sufficiency of the PINIT domain alone in regulation of STAT3.

Although localisation studies performed in the current study have shown a degree of perturbation on PIAS3 function by mutation of L97 and R99, immunoprecipitation studies could be performed to investigate the extent of the effect of the mutations on the perturbation of PIAS3-STAT3 interaction. Localisation studies would be more effective to assess the effect mutation or individual domains and PIAS3 full length if stable transfectants were generated.

Also, of interest, would be the comparison of the binding affinity of the phoshorylated and unphosphorylated STAT3 for the PINIT domain and full–length PIAS3. For example, performing SPR studies using a STAT3 Y705E mutant (i.e. a phospho–mimic) will verify the importance of this residue in the PIAS3–STAT3 interaction. When these questions are fully addressed and taken together with the knowledge of the PINIT domain binding interface shown in this study, it is necessary to investigate the STAT3 binding interface. *In silico* PINIT–STAT3 docking and mutational analysis of STAT3 identified residues using SPR can be employed. While the mutational analysis of the PINIT domain have been shown to abrogate the PINIT-STAT3 interaction. There are more conserved residues that could be mutated to determine the number of PINIT interacting surfaces and assess the importance of the L97 and R99 residues in the PINIT-STAT3 interaction.

Although the expression and purification of the PINIT domain was successful, other expression strategies and expression systems (e.g. pET vector system) could be employed in order to produce sufficient quantities for structural determination by X-ray crystallography and NMR. Co–crystallisation of the PINIT–STAT3 complex to elucidate the complex using X–ray crystallography will contribute significantly to the mechanism of interaction, potentially revealing distinct contact points. Also, the NMR solution structure of the PINIT domain structure would allow for PINIT–STAT3 interaction studies by NMR titration experiments. This method will further identify other residues that can potentially interact with STAT3 in solution and these residues can be mapped into the PINIT domain model to gain insight of their orientation in three-dimensional space.

REFERENCES

Akira S. 1999. Functional roles of STAT family proteins lessons from knockout mice. *Stem cell*. **17**:138–146.

Albor A, El–Hizawi S, Horn EJ, Laederich M, Frosk P, Wrogemann K, Kulesz–Martin M. 2006. The interaction of PIASy with Trim32, an E3–ubiquitin ligase mutated in limb–girdle muscular Dystrophy type 2h, promotes PIASy degradation and regulates UVB–induced keratinocyte apoptosis through NFκB. *J. Biol. Chem*. **35**: 25850–25866.

Alder AJ, Greenfield NJ, Fasman GD. 1973. Circular dichroism and optical rotary dispersion of proteins and polypeptides. *Meth. Enzymology*. **27**: 675.

Altschul S.F. and Koonin E.V. 1998. Iterated profile searches with PSI–BLAST, a tool for discovery in protein databases. *Trends Biochem. Sci.* **23**: 444–447.

Aravind L, Koonin EV. 2000. SAP a putative DNAbinding motif involved in chromosomal organization. *Trends Biochem. Sci*. **25:** 112–114.

Arrondo J L, Muga A, Castresana J, Goni FM. 1993. Quantitative studies of the structure of proteins in solution by Fourier–transform infrared spectroscopy. *Prog. Biophys. Mol. Biol.* **59**: 23 –56.

Becker S, Corthals GL, Aebersold R, Groner B, Müller CW. 1998. Expression of tyrosine phosphorylated DNA binding STAT3b dimer in bacteria. *Fed. Europ. Biochem. Soc. Let.* **441**: 141–147.

Becker S, Groner B, Muller CW. 1998. Three–dimensional structure of the STAT3β homodimer bound to DNA. *Nature*. **394.**

Beliakoff J and Sun Z. 2006. Zimp7 and Zimp10, two novel PIAS-like proteins, function as androgen receptor coregulators. *Nucl. Recep. Sign.* **4**: e017

Beseničar M, Maček P, Lakey JH, Anderluh G, 2006. Surface plasmon resonance in protein–membrane interactions. *Chem. Phys. Lips*.**141**:169–178.

Blond–Elguindi S, Cwirla SE, Dower WJ, Lipshatz RJ, sprang SR, Sambrook JE, Gething MJH. 1993. Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell*. **75**: 717–728.

Borghouts C, Tittmann H, Delis N, Kirchenbauer M, Brill B, Groner B. 2010. The intracellular delivery of a recombinant peptide derived from the acidic domain of PIAS3 inhibits STAT3 transactivation and induces tumour cell death. *Mol. Cancer Res*. **8**: 539–553.

Bradford MM. 1976. A rapid and sensitive method for quantitation of micro quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem*. **72**: 248–254.

Brantley EC, Nabors LB, Gillespie CY, Choi YH, Palmer CA, Harrison K, Roarty K, Benveniste EN. 2008. Loss of protein inhibitors of activated STAT–3 expression in glioblastoma multiforme tumours: implications for STAT–3 activation and gene expression .*Clin. Cancer Res.* **14***:* 4694.

Byler DM, Susi H. 1986. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymer*. **25**: 469−487.

Cheng CH, Lo YH, Liang SS, Ti SC, Lin FM, Yeh CH, Huang HY, Wang TF. 2006. SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of Saccharomyces cerevisiae. *Genes Dev*. **20**: 2067–2081.

Chothia C, Lesk AM. 1986. The relationship between the divergence of sequence and structure in proteins. *Europ. Mol. Biol.org. J.* **5**: 823–826.

Chong S. 2001. Proteind: affinity tags. *Encyclo. Life Sci*.**10**:1038

Chung CD, Liao J, Liu B, Rao X, Jay P, Berta P, Shuai K. 1997. Specific inhibition of STAT3 signal transduction by PIAS3. *Science*. **278**: 1803–1805.

Copeland NG, Gilbert DJ, Schindler C, Zhong Z, Wen Z, Darnell JE Jr, Mui AL, Miyajima A, Quelle FW, Ihle JN, Jenkins NA. 1995. Distribution of the mammalian STAT gene family in mouse chromosomes. *Genomics.* **29:** 225–228.

Cuff JA, Barton GJ. 1999. Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*. **40**: 502–511.

Dabir S, Kluge A, Dowlati A. 2009. The association and nuclear translocation of the PIAS3–STAT3 complex is ligand and time dependent. *Mol. Cancer Res*. **11**: 1854–60.

Darnell JE Jr, Kerr IM, Stark GR. 1994. Jak–STAT pathways and transcriptional activation in response to IFNs and other extracellular signalling proteins. *Science.* **264**: 1415–1421.

DeLano WL. 2004. The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org

Dodge C, Schneider R, Sander C. 1998. The HSSP database of protein structure–sequence alignments and family profiles. *Nucl. Acid Res.* **26:** 313–315.

Dong A, Caughey B, Caughey WS, Bhat KS, Coe JE. 1992. Secondary structure of the pentraxin female protein in water determined by infrared spectroscopy: effects of calcium and phosphorylcholine. *Biochemistry*. **31**: 9364−9370.

Duval D, Duval G, Kedinger C, Poch O, Boeuf H. 2003. The "PINIT" motif of a newly identified conserved domain of the PIAS protein family is essential for nuclear retention of PIAS3L. *Fed. Europ. Biochem. Soc. Lett.* **554**: 111–118.

Eisenberg D, Luthy R, Bowie JU. 1997. VERIFY3D: assessment of protein models with three–dimensional profiles. *Methods Enzymol*. **277**:396–404

Eswar N, Webb B, Marti–Renom MA, Madhusudhan MS, Eramian D, Sali A. 2007. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci* . Chapter 2, Unit 2.9.

Eustace D, Han X, Gooding R, Rowbottom A, Riches P, and Heyderman E. 1993. Interleukin–6 (IL–6) functions as an autocrine growth factor in cervical carcinomas in vitro. *Gynecol. Oncol.* **50:**1939–1948

Flynn GC, Pohl J, Flocco MT, Rothman JE. 1991. Peptide–binding specificity of the molecular chaperone BiP. *Nature*. **353**: 726–730.

Fu FN, DeOliveira DB, Trumble WR, Sarkar HK, Singh BR. 1994. Secondary structure estimation of proteins using the amide III region of Fourier transform infrared spectroscopy: Application to analyze calcium–binding–induced structural changes in calsequestrin. *Appl. Spectrosc*. **48**: 1432–1441.

Fu XY, Schindler C, Improta T, Aebersold R, Darnell JE Jr. 1992. The proteins of ISGF–3, the interferon a–induced transcriptional activator, define a gene family involved in signal transduction. *Proc. Natl. Acad. Sci*. USA. **89:** 7840–7843.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A, Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005).

Gazitt Y, Hu WX. 1998. Fas (APO-1/CD95)-mediated apoptosis is independent of bcl-2: a study with cell lines overexpressing bcl-2 and with bcl-2 transfected cell lines. *Int. J. Oncol*. **12**: 211-220.

Ginalski K, Elofsson A, Fischer D, Rychlewski L, 3D–Jury L. 2003. A simple approach to improve protein structure predictions. *Bioinformatics*. **19**: 1015–1018.

Globocan .2008. Globocan 2008 Cancer Fact Sheet. http://globocan.iarc.fr.

Goddard TD, Huang CC, Ferrin TE. 2005. Software extensions to UCSF Chimera for interactive visualization of large molecular assemblies. *Structure*. **13**: 473–482.

Griebenow K, Klibanov AM. 1995. Lyophilization–induced changes in the secondary structure of proteins. *Proc. Natl. Acad. Sci.* USA. **92**: 10969–10972.

Grönberg MD H. 2003. Prostate cancer epidemiology. *Series*. **361**: 859–864.

Guenther B, Onrust R, Sali A, O'Donnell M, Kuriyan J. 1997. Crystal structure of the delta subunit of the clamp-loader complex of *E. coli* DNA polymerase III. *Cell*. **91**: 335–345.

Hannich, J T, Lewis A, Kroetz M B, Li SJ, Heide H, Emili A, Hochstrasser M. 2005. Defining the SUMO–modified proteome by multiple approaches in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**: 4102–4110.

Hanson IM, Poustka A, Trowsdale J. 1991. New genes in the class II region of the human major histocompatibility complex. *Genomics*. **10**: 417–424.

Hari KL. *et al*. 2001. The *Drosophila Su(var)2–10* locus regulates chromosome structure and function and encodes a member of the PIAS protein family. *Gene. Dev*. **15**: 1334–1348.

Haris PI, Lee DC, Chapman D. 1986. A fourier transform infrared investigation of the structural differences between ribonuclease A and ribonuclease S. *Biochim. Biophys. Acta*. **874**: 255.

Haris PI, Severcan F. 1999. FTIR spectroscopic characterization of protein structure in aqueous and non–aqueous media. *J. Mol. Catalysis*. **7**: 207–221.

Heinrich PC, Behrmann I, Müller–Newen G, Schaper F, Graeve L. 1998. Interleukin–6–type cytokine signalling through the gp130/Jak/STAT pathway. *Biochem. J.* **334:** 297–314.

Hess S, Smola H, Sandaradura de Silva U, Hadaschik D, Kube D, Baldus SE, Flucke U, Pfister H. 2000. Loss of IL–6 receptor expression in carcinoma cells inhibits autocrine IL–6 Stimulation: abrogation of constitutive monocyte chemoattractant protein–1 production. *Immunology*. **165**:1939–1948.

Higgins DG, Sharp PM. 1998. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*. **73**: 237–244.

Hochstrasser M. 2001. New structural clues to substrate specificity in ubiquitin system. *Mol. Cel. Biol.* **9**: 453–454.

Hodgkinson CA, Moore KJ, Nakayama A, Steingrímsson E, Copeland NG, Jenkins NA, Arnheiter H. 1993. Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic–helix–loop–helix–zipper protein. *Cell*. **74:** 395–404.

Hoeve JT, Ibarra-Sanchez MJ, Fu Y, Zhu W, Tremblay M, David M, Shuai K. 2002. Identification of a nuclear STAT1 protein tyrosine phosphatase. *Mol. Cell. Biol.* **16**: 5662-5668.

Hoey T. Grusby MJ. 1999. STATs as mediators of cytokine–induced responses. *Immunology*, **71**:145–162.

Honda R, Tanaka H, Yasuda H. 1997. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumour suppressor p53. *Fed. Europ. Biochem. Soc. Let.* **420**: 25–27.

Hooft R, Vriend G, Sander C, Abola E. 1996. Errors in protein structures. *Nature*. **81**:272.

Horvath CM, Wen Z, Darnell JE Jr. 1995. A STAT protein domain that determines DNA sequence recognition suggests a novel DNA–binding domain. *Gene. Dev.* **9**: 984–994.

Huang CY, Beliakoff J, Li X, Lee J, Sharma M, Lim B, Sun Z. 2005. hZimp7, A novel PIAS-like protein, enhances androgen receptor-mediated transcription and interacts with SWI/SNF-like BAF complexes. *Mol. Endocrinol*. **12**:2915-29.

Jones DT, Taylor WR, Thornton, JM. 1992 A new approach to protein fold recognition. *Nature*. **358**:86-89.

Jimenez–Lara AM, *et al*. 2002 PIAS3 (protein inhibitor of activated STAT–3) modulates the transcriptional activation mediated by the nuclear receptor coactivator TIF2. *Fed. Europ. Biochem. Soc. Lett*. **526**: 142–146

Joazeiro CA, Weissman AM. 2000. RING finger proteins: mediators of ubiquitin ligase activity. *Cell*. **102**: 549–552.

Johnson ES, Gupta AA. 2001. An E3–like factor that promotes SUMOconjugation to the yeast septins. *Cell*. **106:** 735–744.

Jones DT. 1999. Protein secondary structure prediction based on position–specific scoring matrices. *J. Mol Biol*. **292**: 195–202.

Jönsson U, Fägerstam L, Ivarsson B, Johnsson B, Karlsson R, Lundh K, Löfas S, Persson B, Roos H, Rönnberg I, Sjölander S, Stenberg E, Stahlberg R, Urbaniczky S, Östlin H, Malmqvist M. 1991. Real–time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *Bio. Techniques.* **11**:620–627.

Kabish W, Sander C. 1983. Dictionary of protein secondary structure : pattern recognition of hydrogen–bonded and geometrical features. *Biopolymers*. **22**:2577–2637.

Kaplan W, Hüsler P, Klump H, Erhardt J, Sluis–Cremer N, Dirr H. 1997. Conformational stability of pGEX–expressed *schistosoma japonicum* glutathione S–transferase: a detoxification enzyme and fusion–protein affinity tag. *Protein Sci*. **6**: 399–406.

Kauppinen JK, Moffatt DJ, Mantsch HH, Cameron, DG. 1981. Fourier self–deconvolution: A method for resolving intrinsically overlapped bands. *Appl. Spectrosc*. **35**: 271.

Kelley TJ, Drumm ML. 1998. Inducible nitric oxide synthase expression is reduced in cystic fibrosis murine and human airway epithelial cells. *J. Clin. invest*. **102**: 1200–1207.

Kelley TJ, Elmer HL. 2000. *In vivo* alterations of IFN regulatory factor–1 and PIAS1 protein levels in cystic fibrosis epithelium. *J. Clin Invest*. **106**: 403–410.

Kelly LA, MacMallum RM, Sternberg MJ. 2000. Ehanced genome annotation using structural profiles in the program 3D–PSSM. *J Mol. Biol*. **299**: 499–520.

Kerscher O, Felberbaum R, Hochstrasser M. 2006. Modification of proteins by ubiquitin and ubiquitin–like proteins. *Annu. Rev. Cell. Dev. Biol.* **22:** 159–180.

Klingmuller U, Lorenz U, Cantley LC, Neel BG, Lodish HF. 1995. Specific recruitment of SH–PTP1 to the erythropoietin receptor causes inactivation of JAK2 and termination of proliferative signals. *Cell*. **80**: 729–738.

Kokontis JM, Liao S. 1999**.** Molecular action of androgen in the normal and neoplastic prostate. *Vit. and Horm.* **55**: 219–307.

Kotaja N, Karvonen U, Janne OA, Palvimo JJ. 2002. The nuclear receptor interaction domain of GRIP1 is modulated by covalent attachment of SUMO–1. *J. Biol. Chem*.**277**: 30283–30288.

Krimm S, Bandekar J. 1986. Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem*. **38**: 181−364

Kunihiro K, Kim P, Baldwin RL. 1984. Strategy for trapping intermediates in the folding of ribonuclease and for using 1H–NMR to determine their structures. *Biopolymers.* **22:** 59–67.

Kuryshev YA, Gudz TI, Brown AM, Wible BA. 2000. KChAP as a chaperone for specific K^+ channels. *Am J. Cell Physiol.*. **278**: c931–c941.

Laity JH, Lee BM, Wright PE. 2001, Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. in struct. Biol*. **11**:39–46.

Lamba OP, Borchman D, Sinha SK, Renugopalakrishnan V, Tappert MC. 1993. Estimation of the secondary structure and conformation of bovine lens crystallins by infrared spectroscopy: quantitative analysis and resolution by fourier self–deconvolution ad curve fit. *Biochem. Biophys. Acta.* **116**:113-123.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.2007. Clustal W and Clustal X version 2.0. *Bioinformaticss*. **21**: 2947–8.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of the protein structures. *J. Appl. Cryst*. **26**: 283–291.

Leonard JW, O'Shea JJ. 1998. JAKS AND STATS: Biological Implications. *Immunology*, **16**:293–322.

Levine AJ. 1997. p53 the cellular gatekeeper for growth and division. *Cell*. **88**:323–331.

Levy C, Lee NY, Nechushtan H, Schueler–Furman O, Sonnenblick A, Hacohen S, Razin E. 2006. Identifying a common molecular mechanism for inhibition of MITF andSTAT3 by PIAS3. *Blood*. **107:** 2839–2845.

Liao J, Fu J, Shuai K. 2000. Distinct roles of the NH_2– and COOH–terminal domains of the protein inhibitor of activated signal transducer and activator of transcription STAT 1 PIAS1 in cytokine–induced PIAS1–STAT1 interaction. *Proc. Nat. Acad. of Sci.* USA. **97:** 5267–5272.

Liedberg B, Nylander C, Lundstorm I. 1983. Surface plasmon resonance for gas detection and biosensing. *Sens. Actuators*. **B.4**:229-304.

Ling Y, Sankpal UT, Robertson AK, McNally JG, Karpova T, Robertson KD. 2003. Modification of de novo DNA methyltransferase 3a (Dnmt3a) by SUMO–1 modulates its interaction with histone deacetylases (HDACs) and its capacity to repress transcription. *Nucl. Acids Res*. **32**: 598–610.

.Liu B, Liao J, Rao X, Kushner SA, Chung CD, Chang DD, Shuai K. 1998. Inhibition of STAT1–mediated gene activation by PIAS1. *Proc. Natl. Acad. Sci*. USA. **95**: 10626–10631.

Mackiewiez A, Koj A, Sehgal PB. 1995. Interleukin–6–type cytokines. *Ann. N. Y. Acad. Sci.***762:** 1–520.

Man JH, Li HY, Zhang PJ, Zhou T, He K, Pan X, Liang B, Li AL, Zhao J, Gong WL, Jin BF, Xia Q, Yu M, Shen BF, Zhang XM. 2006. PIAS3 induction of PRB sumoylation represses PRB transactivation by destabilizing its retention in the nucleus. *Nucl. Acids Res*. **34**: 5552–5566.

Mao JJ. 2005. Stem–cell–driven regeneration of synovial joints. *Biol. Cell*. **97**:289–301.

Mautsa N, Prinsloo E, Bishop OT and Blatch GL. 2010. The PINIT domain of PIAS3: structure-function analysis of its interaction with STAT3. *J. Mol. Recog.* DOI: **10:1002**/jmr.1111.

Melino G, Lu X, Gasco M, Crook T, Knight RA. 2003. Functional regulation of p73 and p63: development and cancer. *Trends Biochem Sci*. **28**: 663–670.

Meng QH, Springall DR, Bishop AE, Morgan K, Evans TJ, Habib S, Gruenert DC, Gyi KM, Hodson ME, Yacoub MH, Polak JM. 1998. Lack of inducible nitric oxide synthase in bronchial epithelium: a possible mechanism of susceptibility to infection in cystic fibrosis. *J. Pathol*. **184:** 323–331.

Minty A, Dumont X, Kaghad M, Caput D. 2000. Covalent modification of p73α by SUMO–1. Two–hybrid screening with p73 identifies novel SUMO–1–interacting proteins and a SUMO–1 interaction motif. *J. Biol. Chem*. **275**: 36316–36323.

Mitchell SH, Zhu W, Young CY. 1999. Resveratrol inhibits the expression and function of the androgen receptor in LNCaP prostate cancer cells. *Cancer Res*. **59**: 5892–5895, 1999.

Miyauchi Y, Yogosawa S, Honda R, Nishida T, Yasuda H. 2002. Sumoylation of Mdm2 by Protein Inhibitor of Activated STAT (PIAS) and RanBP2 Enzymes. *J. Biol. Chem.* **277**: 50131–50136.

Mohr SE, Boswell RE. 1999. Zimp encodes a homologue of mouse Miz1 and PIAS3 and is an essential gene in *Drosophila* melanogaster. *Gene*. **229**: 109–116.

Moilanen AM, Karvonen U, Poukka H, Yan W, Toppari J, Jänne OA, Palvimo JJ. 1999. A testis–specific androgen receptor coregulator that belongs to a novel family of nuclear proteins. *J. Biol. Chem.* **274**: 3700–3704.

Mowen KA, Tang J, Zhu W, Schurter BT, Shuai K, Herschman HR. 2001. Arginine methylation of STAT1 modulates IFNalpha/beta–induced transcription. *Cell*. **104:** 731–741.

Myszka DG. 1997. Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Curr. Opin. Biotechnol***. 8:**50–57.

Neculai D, Neculai AM, Verrier S, Straub K, Klumpp K, Pfitzner E, Becker S. 2005. Structure of the unphosphorylated STAT5α dimer. *J. Boil. Chem.* **280**: 40782–40787.

O'Shea JJ. 1997. Jaks-STATs cytokine signal review transduction, and immunoregulation: are we there yet?. *Immunology*. **7:**1–11.

Ogata Y, Osaki T, Naka T, Iwahori T, Mitsugi F, Nagatomo I, Kijima T, Kumagai T, Mitsuhiro Y, Tachibana I, Kawase I. 2006. Overexpression of PIAS3 suppresses cell growth and restores the drug sensitivity of human lung cancer cells in association with PI3–K/Ajt inactivation. *Noeplasia.***10:** 817–825.

Ohi MD, Vander Kooi CW, Rosenberg JA, Chazin WJ, Gould KL. 2003. Structural insights into the U–box, a domain associated with multi–ubiquitination. *Nat. Struct. Biol*. **3:** 3.

Okubo S, Hara F, Tsuchida Y, Shimotakahara S, Suzuki S, Hatanaka H, Yokoyama S, Tanaka H, Yasuda H, Shindo H. 2004. NMR Structure of the N–terminal domain of SUMO ligase PIAS1 and its interaction with tumour suppressor p53 and A/T–rich DNA oligomers, *J. Biol. Chem.* **279**: 31455–31461.

Ouali M, King RD. 2000. Cascaded multiple classifiers for secondary structure prediction. *Prot. Sci*. **9**: 1162–1176.

Pawlowski M, Gajda M, Matlak R, Bujnicki JM. 2008. MetaMQAP: A meta–server for the quality assessment of protein models. *Bioinformatics*. **9**: 403.

Peng Y, Lee J, Zhu C, Sun Z. 2010. A novel role for protein inhibitor of activated STAT (PIAS) proteins in modulating the activity of Zimp7, a novel PIAS–like protein, in androgen receptor–mediated transcription. *J. Biol. Chem*. **285**: 11465–11475.

Prinsloo E, Oosthuizen V, Muramoto K, Naude RJ. 2006. In vitro refolding of recombinant human free secretory component using equilibrium gradient dialysis. *Prot. Expr. Puri*. **47**: 179–185.

Prives C. 1998. Signaling to p53: breaking the MDM2–p53 circuit. *Cell*. **95**: 5–9.

Pryor KD, Leiting B. 1997. High–level expression of soluble protein in *Escherichia coli* using a $(His)_6$–tag and maltose–binding protein double–affinity fusion system. *Protein Expr. Purif.* **10**: 309–319.

Reindle A, Belichenko I, Bylebyl GR, ChenXL, Gandhi N, Johnson ES. 2006. Multiple domains in Siz SUMO ligases contribute to substrate selectivity. *J. Cell Sci*. **119**: 4749–4757.

Reymond A, Meroni G, Fantozzi A, Merla G, Cairo S, Luzi L, Riganelli D, Zanaria E, Messali S, Cainarca S, Guffanti A, Minucci S, Pelicci PG, Ballabio A. 2001. The tripartite motif family identifies cell compartments. *Europ. Mol. Biol.org. J.* **20**: 2140–2151.

Rich RL, Myszka DG. 2000. Advances in surface plasmon resonance biosensor analysis. *Curr. Opin.in Biotechnol.* 11: 54-61

Rich RL, Myszka DG. 2004. Why you should be using more SPR biosensor technology. *Drug Discov. Today.* **1**:301–308.

Rödel B, Tavassoli K, Karsunky H, Schmidt T, Bachmann M, Schaper F, Heinrich P, Shuai K, Elsässer HP, Möröy T. 2000. The zinc finger protein Gfi–1 can enhance STAT3

signaling by interacting with the STAT3 inhibitor PIAS3. *Europ. Mol. Biol.org. J.* **19**:5845–5855.

Rohman M, Harrison-Lavoie KJ. 2000. Separation of copurifying GroEL from glutathione-S-transferase fusion proteins. *Protein Expr. Purif.* **20**: 45–47.

Romig H, Fackelmayer FO, Renz A, Ramsperger U, Richter A. 1992. Characterization of SAF–A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear/scaffold attachment DNA elements. *Europ. Mol. Biol.org. J.* **1**:3431–3440.

Rosendorff A, Illanes D, David G, Lin J, Kieff E, Johannsen E. 2004. EBNA3C coactivation with EBNA2 requires a SUMO homology domain. *J. Virol*. **78**: 367–377.

Rost B. 2001. Protein secondary structure prediction continues to rise. *J. Struct. Biol.* **134**: 204–218.

Rytinki MM, Kaikkonen S, Pehkonen P, Jääskeläinen T, Palvimo JJ. 2009. PIAS proteins: pleiotropic interactors associated with SUMO. *Cell. Mol. Life Sci*. **66**:3029–3041.

Sali A, Blundell TL.1993. Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol*. **234:** 779-815.

Sachdev S, Bruhn L, Sieber H, Pichler A, Melchior F, Grosschedl R. 2001. PIASy, a nuclear matrix–associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Gene. Dev.* **15**: 3088–3103.

Sancheze R, Sali, A. 1997. Advances in comparative protein–structure modeling. *Cur. Opin Struct. Biol.* **7**: 206–214.

Schindler C, Fu XY, Improta T, Aebersold R, Darnell JE Jr. 1992. Proteins of transcription factor ISGF–3: one gene encodes the 91–and 84–kDa ISGF–3 proteins that are activated by interferon alpha. *Proc. Natl. Acad. Sci*. **89:** 7836–7839.

Schmidt D, Muller S. 2003. PIAS/SUMO: new partners in transcriptional regulation. *Cell Mol Life Sci*. **60**: 2561–74.

Sehgal PB, Grieninger G, Tosato G 1989. Regulation of the acute phase and immune responses: interleukin 6. *Ann. N. Y. Acad. Sci.* **557:** 1–582.

Sharma M, Li X, Wang Y, Zarnegar M, Huang CY, Palvimo JJ, Lim B, Sun Z**.** 2003. hZimp10 is an androgen receptor co–activator and forms a complex with SUMO–1 at replication foci. *Europ. Mol. Biol.org. J.* **22:** 6101–6114.

Shen MY, Sali A. 2006. Statistical potential for assessment and prediction of protein structure. *Prot. Sci*. **15**: 2507–2524.

Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) FUGUE: sequence–structure homology recognition using environment–specific substitution tables and structure dependent gap penalties. *J. Mol. Biol.* **310:** 243–257

Shuai K, Liu B. 2005. Regulation of gene–activation pathways by PIAS proteins in the immune system. *Nat. Rev. Immunology*. **5**: 593–605.

Shuai K. 1999. The STAT family of proteins in cytokine signaling. *Prog. Biophys. Mol. Biol.* **71**: 405–422.

Shuai K. 2000. Modulation of STAT signaling by STAT–interacting proteins. *Oncogene* **19**: 2638–2644.

Shuai K. 2006. Regulation of cytokine signaling pathways by PIAS proteins. *Cell Res*. **16**: 196–202

Singh BR, DeOliveria DB, Fu FN, Fuller MP. 1993. Fourier transform infrared analysis of amide III bands of proteins for secondary structure estimation. *Proc. Biomol. Spectrosc III. Los Angeles*, CA.

Sippl MJ. 1993. Recognition of errors in three dimensional structure of proteins. *Proteins*. **17**: 355–362.

Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucl. Acids Res*. **33**: W244–W248.

Song J, Durrin LK, Wilkinson TA, Krontiris TG, Chen Y. 2004. Identification of a SUMO–binding motif that recognizes SUMO–modified proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 14373–14378.

Song J, Zhang Z, Hu W, Chen Y. 2005. Small Ubiquitin–like Modifier (SUMO) recognition of a SUMO binding motif. *J. Biol. Chem*. **280**: 48.

Sonnenblick A, Levy C, Razin E. 2004. Interplay between MITF, PIAS3, and STAT3 in mast cells and melanocytes. *Mol. Cell. Biol.* **24:** 10584–10592.

Sordella R, Bell DW, Haber DA, Settleman J. 2004. Gefitinib–sensitizing EGFR mutations in lung cancer activate anti–apoptotic pathways. *Science.* **305**:1163–1167.

Soule HD, Vazguez J, Long A, Albert S, Brennan M. 1973. A human cell line from a pleural effusion derived from a breast carcinoma. *J. Nat. Cancer*. **51**: 1409–1416

Spassov VZ, Flook PK, Yan L. 2008. LOOPER: a molecular mechanics-based algorithm for protein loop prediction. *Prot. Eng. Design and Sel.* **2**: 91-100.

Surewicz WK, Mantsch HH.1988. New insight into protein secondary structure from resolution–enhanced infrared spectra. *Biochim. Biophys. Acta*. **952**: 115−130.

Susi H, Byler DM. 1986. Resolution–enhanced fourier transform infrared spectroscopy of enzymes. *Methods Enzymol.* **130**: 290−311.

Stenberg E, Persson B, Roos H, Urbaniczky C. 1991. Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins *J. Col. Interface Sci.***143**: 513-526.

Takahashi K, Taira T, Niki T, Seino C, Iguchi–Ariga SMM, Ariga H. 2001. DJ–1 Positively regulates the androgen receptor by impairing the binding of PIASxα to the receptor. *J. Bio. Chem.* **276**: 37556–37563.

Tamm LK, Tatulian SA, Rev Q. 1997. Infrared spectroscopy of proteins and peptides in lipid bilayers. *Biophys*. **30**: 365.

Taylor WR. 1986. Identification of proteins sequences homology by consensus template alignment. *J Mol. Biol*. **188**: 233–58.

Terpe K. 2003. Overview of tag protein fusions:from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol*. **60**:523–533

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position–specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**: 4673–4680.

Todaro GJ, Green H. 1963. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol*. **17**: 299–313.

Urist M, Prives C. 2002. p53 Leans on its siblings. *Cancer Cell.* **1**: 311−313.

Vinkemeier U, Cohen SL, Moarefi I, Chait BT, Kuriyan J, Darnell JE Jr. 1996. DNA binding of *in vitro* activated STAT1 alpha, STAT1 beta and truncated STAT1: interaction between NH2–terminal domains stabilizes binding of two dimers to tandem DNA sites. *Europ. Mol. Biol.org. J.* **15**: 5616–5626.

Vousden KH, Prives C. 2005. P53 and prognosis: new insights and further complexity. *Cell.* **120**: 7–10.

Wang L, Banerjee S. 2004. Differential PIAS3 expression in human malignancy. *Oncol. Rep*. **11**: 1319–1324.

Ward JJ, McGuffin LJ, Buxton BF, Jones DT. 2004. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. **20**: 2138–2139.

Weissman AM. 2001. Themes and variations on ubiquitylation. *Nat. Rev. Mol. Cell Biol.* **2**: 169–178.

Welsh MJ, Tsui LC, Boat TF, and Beaud 1995. Cystic fibrosis. In the metabolic and molecular basis of inherited diseases. *Valle, eds. (New York: McGraw–Hill, Inc.)*, 3799–3876.

Wen Z, Zhong Z, Darnell JJ Jr. 1995. Maximal activation of transcriptionby STAT1 and STAT3 requires both tyrosine and serine phosphorylation. *Cell*. **82**:241–250.

Wible BA, Yang Q, Kuryshew YA, Accili EA, Brown AM. 1998. Cloning and expression of a novel K^+ channel regulatory protein, KChAP. *J. Biol. Chem*. **273**: 11745–11751.

Williams KA, Deber CM. 1991. Proline residues in transmembrane helices: structure or dynamics role?. *Biochemistry*. **30**: 8919-8923.

Wong AK, Kim R, Christofk H, Gao J, Lawson G, Hong G. 2004. Protein inhibitor of activated STAT (PIASy) and a splice variant lacking exon 6 enhance sumoylation but are not essential for embryogenesis and adult life. *Mol. Cell. Biol*. 24: 5577–5586.

Wormald S, Hilton JD. 2004. Inhibitors of cytokine signal transduction. *J. Bio. Chem*. **279**: 821–824.

Wu L, WU H, Sangiorgi F, WU N, Bell JR, Lyons GE, Maxson R. 1997. Miz 1, a novel zinc finger transcription factor that interacts with Msx2 and enhances its affinity for DNA. *Mech. Dev*. **65**: 3–17.

Xu X, Sun YL, Hoey T. 1996. Cooperative DNA binding and sequence–selective recognition conferred by the STAT amino–terminal domain. *Science*. **273**: 794–797.

Yagil Z, Kay G, Nechushtan H, Razin E. 2009. A specific epitope of protein inhibitor of activated STAT3 is responsible for the introduction of apoptosis in rate transformed mast cells. *Immunology*. **182**: 2168–2175.

Yamamoto T, Sato N, Sekine Y, Yumioka T, Imoto S, Junicho A, Fuse H, Matsuda T. 2003. Molecular interactions between STAT3 and protein inhibitor of activated STAT3 and androgen receptor. *Biochem. Biophys. Res. Commun*. **306:** 610–615.

Yamashina K, Yamamoto H, Chayama K, Nakajima K, Kikuchi A. 2006. Suppression of STAT3 activity by duplin, which is a negative regulator of the Wnt signal. *J. Biochem*. **139**: 305–314.

Yang Y, Li CH, Weissman AM. 2004. Regulating the p53 system through ubiquitination. *Oncogene*. **23:** 2096–2106.

Yang ZR, Thomson R, McNeil P, Esnouf RM. 2005. RONN: the basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*. **21**: 266–273.

Yunus AA, Lima CD. 2009. Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO Modification of PCNA. *Mol. Cell*. **35**: 669–682.

Zhang L, Skolnick J. 1998. What should the Z–score of native protein structures be?. *Protein Sci*.**7**: 1201–1207.

Zhao X, Blobel G. 2005. A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 4777–82.

Zhiyuan Y, Kone BC. 2004. The STAT3 DNA–binding domain mediates interaction with NF–κB p65 and inducible nitric oxide synthase transrepression in mesangial cell *J. Am. Soc. Nephrol.* **15**: 585–591.

Zhong Z, Wen Z, Darnell JE Jr. 1994. STAT3 and Stat4: members of the family of signal transducers and activators of transcription. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 4806–4810.

Zoumpoulidou G, Jones MC, De Mattos SF, Francis JM, Fusi L, Lee YS, Christian M, Varshochi R, Lam EW, Brosens JJ. 2004. Convergence of interferon–gamma and progesterone signaling pathways in human endometrium: role of PIASy (protein inhibitor of activated signal transducer and activator of transcription–y). *Mol. Endocrinol.* **20**:1988–1999.

APPENDICES

APPENDIX A: AMINO ACID AND NUCLEOTIDE NOMENCLATURE

One and three-letter codes were used to represent amino acids, and single letter codes were used to represent nucleotides as set forward by the Joint Commission of Biochemical Nomenclature (JBNC) of IUPAC (International Union of Pure and Applied Chemistry) and the IUBMB (International Union of Biochemistry and Molecular Biology):

NUCLEOTIDE	SINGLE-LETTER CODE
Adenine	A
Cytosine	C
Guanine	G
Thymine	T
Uracil	U
Any Nucleotide	(A, C, G, T or U) N

AMINO ACID	1-LETTER CODE	3-LETTER CODE	DNA CODONS
Alanine	A	Ala	GCT, GCC, GCA, GCG
Arginine	R	Arg	CGT, CGC, CGA, CGG, AGA, AGG
Asparagine	N	Asn	AAT, AAC
Aspartic acid	D	Asp	GAT, GAC
Cysteine	C	Cys	TGT, TGC
Glutamine	Q	Gln	CAA, CAG
Glutamic acid	E	Glu	GAA, GAG
Glycine	G	Gly	GGT, GGC, GGA, GGG
Histidine	H	His	CAT, CAC
Isoleucine	I	Ile	ATT, ATC, ATA
Leucine	L	Leu	CTT, CTC, CTA, CTG, TTA, TTG
Lysine	K	Lys	AAA, AAG
Methionine	M	Met	ATG
Phenylalanine	F	Phe	TTT, TTC
Proline	P	Pro	CCT, CCC, CCA, CCG
Serine	S	Ser	TCT, TCC, TCA, TCG, AGT, AGC
Threonine	T	Thr	ACT, ACC, ACA, ACG
Tryptophan	W	Trp	TGG
Tyrosine	Y	Tyr	TAT, TAC
Valine	V	Val	GTT, GTC, GTA, GTG
Stop	-	-	TAA, TAG, TGA
Any Amino Acid	X	-	-

APPENDIX B: SUPPLEMENTARY FIGURES

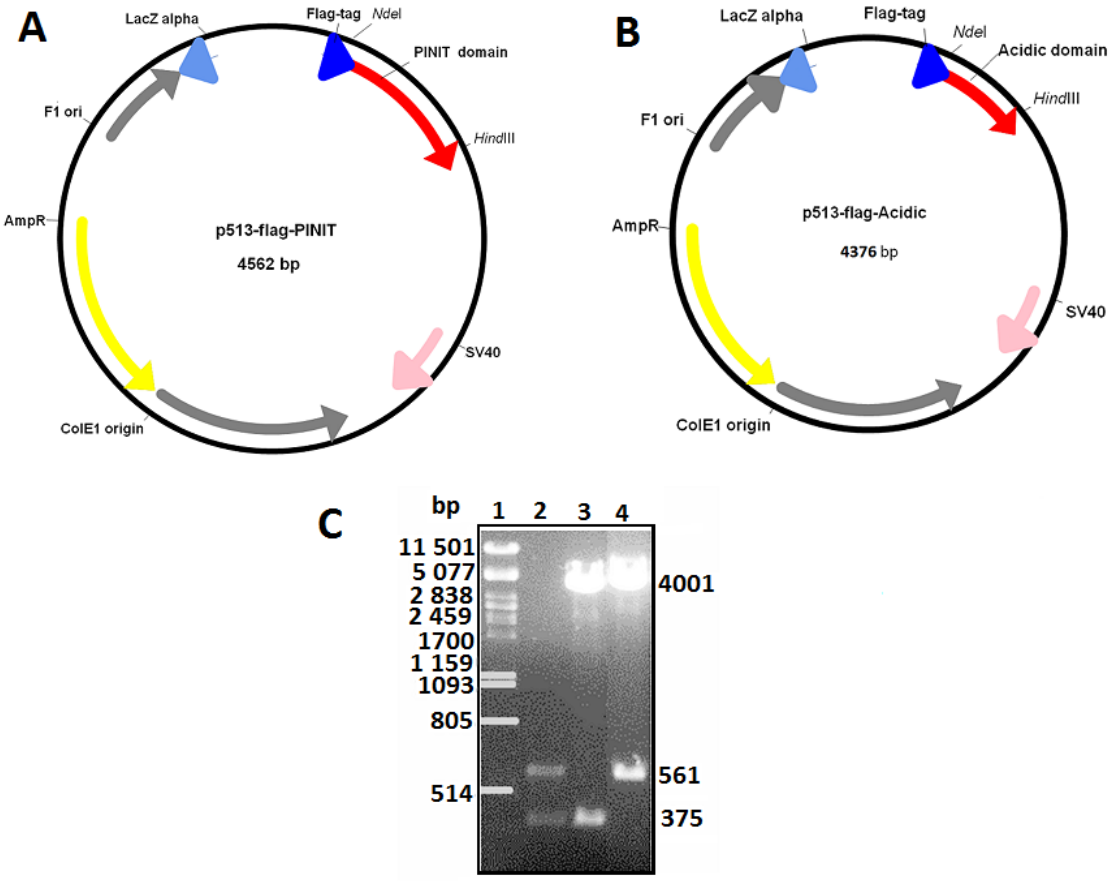

Figure B1 Restriction endonuclease analysis of PINIT and acidic domain constructs. A) Restriction map of p513–flag–PINIT. The regions on the plasmids encoding for PINIT domain, the N–terminal flag–tag encoding segments are indicated by red. **B) Restriction map of p513–flag–Acidic plasmid.** The regions on the plasmids encoding for Acidic domain and the N–terminal flag–tag encoding segments are indicated. **C) Ethidium bromide stain of 0.8% TBE agarose gels electrophoresis.** Restriction analysis of p513–flag–PINIT and p513–flag–Acidic plasmids with *Nde*I and *Hind*III restriction enzymes. DNA was loaded on 0.8% TBE agarose gel containing ethidium bromide in the following order: Lane1, lambda DNA molecular markers digested with *Pst*I; lane 2, a cocktail of PCR amplified fragments encoding for the PINIT and Acidic domains; lane 3, p513–flag–Acidic plasmid digested with *Nde*I and *Hind*III restriction enzyme; lane 4, p513–flag–PINIT plasmid digested with *Nde*I and *Hind*III restriction enzyme. The expected sizes of the PINIT and acidic domain DNA fragments are 561 and 375 bp respectively.

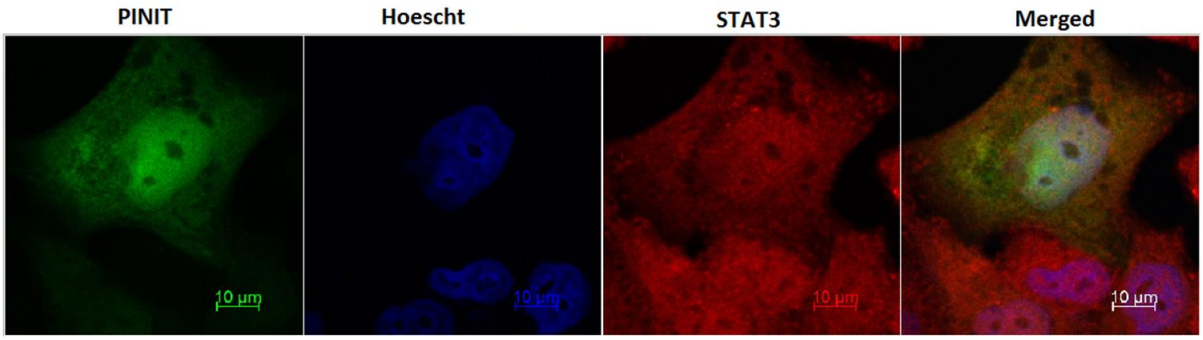

Figure B2 Co–localisation analysis of PINIT domain with STAT3 in HeLa cells. A) HeLa cells grown on glass cover slips were transiently transfected for 48 hours with a p513–flag–PIAS3 plasmid. Cells were starved for 12 hours and stimulated for 30 minutes with IL–6 before fixing as described in section 3.2.3. Immunofluorescent labeling was performed with mouse anti–FLAG M2 monoclonal antibodies followed by one hour incubation with Alexa Fluor^{R} 488 donkey anti–mouse IgG (green). Cell nuclei (blue) were directly labelled with Hoescht 33258. The endogenous STAT3 was stained with STAT3 rabbit polyclonal IgG antibodies and followed by one hour incubation Alexa Fluor 546 chicken anti–rabbit IgG. The immunofluoresence images were captured using a confocal fluorescence microscopy on a laser–scanning Zeiss LSM 510 confocal microscope. Merged images were automatically created by merging the green image with nuclei hoescht staining image and STAT3 (red) image using Zeiss LSM Image Browser software (Carl Zeiss GmbH Jena).

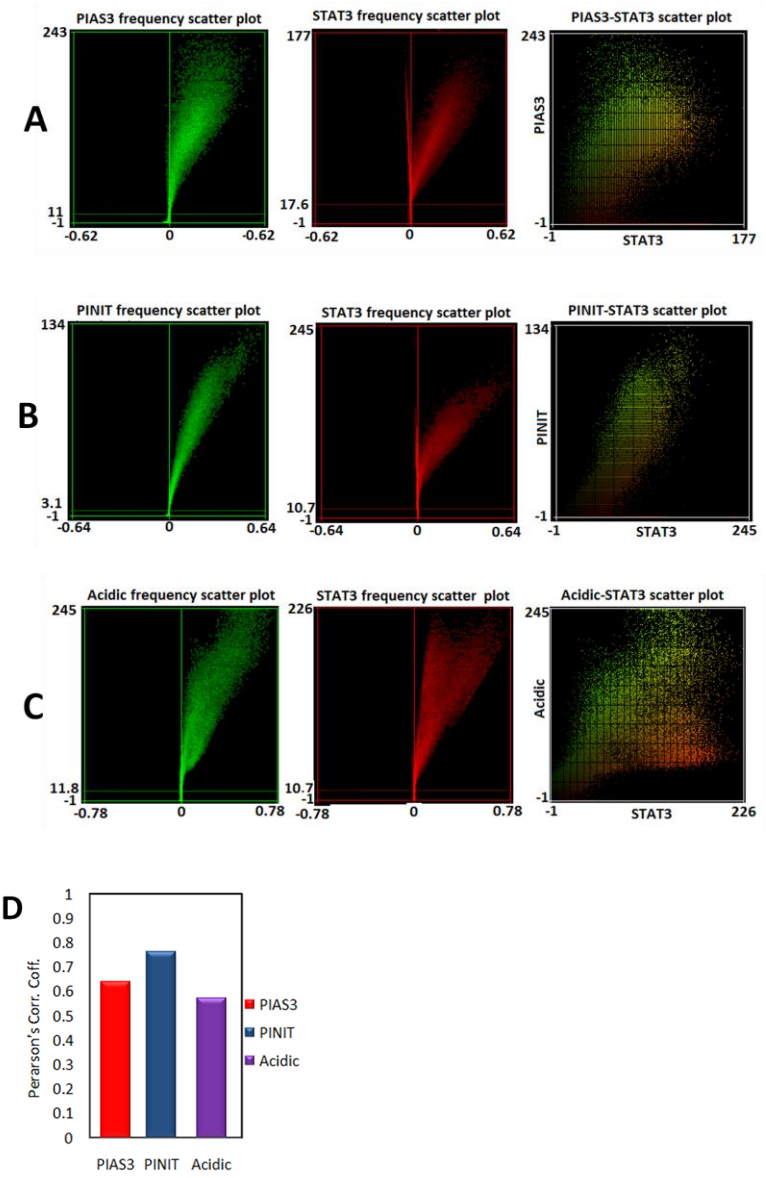

Figure B3 Quantitative co–localisation analysis of the PIAS3, PINIT, and acidic domains and STAT3. A) **Frequency scatter plots of PIAS3 and STAT3 and the PIAS3–STAT3 scatter plot**. PIAS3–STAT3 scatter plot, the intensity of a given pixel in PIAS3 (green) image was used as the x–coordinate of the scatter plot and the intensity of the corresponding pixel in STAT3 (red) image as the y–coordinate. **B**) **Frequency scatter plots of PINIT and STAT3 and the PINIT–STAT3 scatter plot**. PINIT–STAT3 scatter plot, The intensity of a given pixel in PINIT domain (green) image was used as the x–coordinate of the scatter plot and the intensity of the corresponding pixel in STAT3 (red) image as the y–coordinate. **C) Frequency scatter plots of Acidic domain and STAT3 and the scatter plot of Acidic–STAT3.** Acidic–STAT3 scatter plot, The intensity of a given pixel in Acidic domain (green) image was used as the x–coordinate of the scatter plot and the intensity of the corresponding pixel in STAT3 (red) image as the y–coordinate. **D**) Comparison of Pearson's correlation coefficient (Pearson's corr. coeff.) calculated for PIAS3, PINIT and Acidic domains. Co–localization results shown in a pixel distribution along a straight line whose slope will depend on the fluorescence ratio between the two channels and whose spread is quantified by the Pearson's correlation coefficient (PCC) which is close to 1 as red and green channel intensity distributions are linked. Co–localisation analysis was performed using MacBiophotonics ImageJ software (MBF–ImageJ; http://www.macbiophotonics.ca). PCC: PIAS3–STAT3 was 0.638; PINIT–STAT3 was 0.732; Acidic–STAT3 was 0.574.

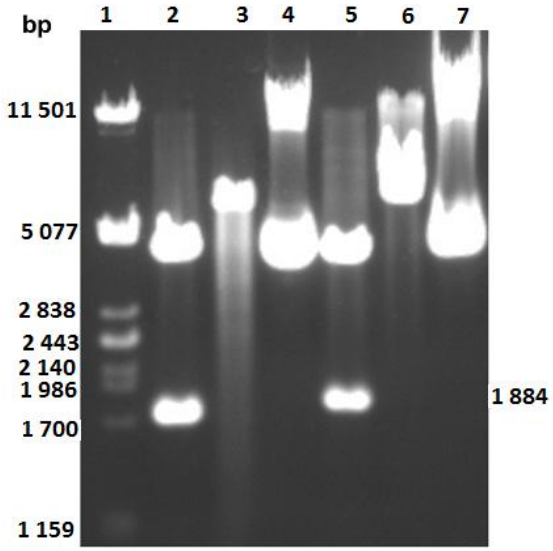

Figure B4 Restriction endonuclease analysis of p513–flag–PIAS3, p513–flag–PIAS3L, pQE2–PIAS3, pQE2–PIAS3L. DNA was loaded on 0.8% TBE agarose gel containing ethidium bromide in the following order. Lane 1, lambda DNA molecular markers digested with *Pst*I; lane 2, p513–flag–PIAS3 construct digested with *Nde*I and *Hind*III; lane 3, undigested p513–flag–PIAS3; lane 4, p513–flag–PIAS3 digested with Hind III; lane 5, p513–flag–PIAS3L digested with *Nde*I and *Hind*III; lane 6, p513–flag–PIAS3L digested with *Hind*III; lane 7, undigested p513–flag–PIAS3L.

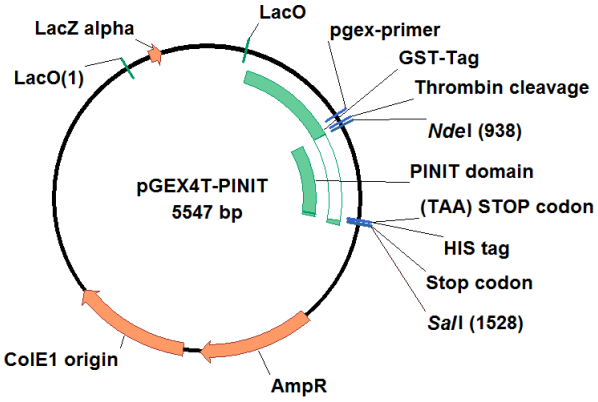

Figure B4 The plasmid map of the pGEX4T–PINIT–TAA–$(His)_6$. B5 Plasmid rendered using Vector NTI Advance™ software package (version 10.3; Invitrogen). The plasmids confer ampicillin resistance to transformed *E. coli* cells as indicated (Amp^R; β–lactamase gene). The position of the PINIT domain between NdeI and SalI restriction sites is indicated; the plasmid has a stop codon (TAA) between the PINIT coding region and the His–tag coding region. The region encoding the GST tag segments is indicated upstream of the PINIT domain coding region and the region encoding for the His-tag is indicated downstream of the PINIT domain coding region after the TAA stop codon. The origin of replication (ColE1 origin) is indicated and regions coding for *Lac*Z alpha, *Lac*O genes and the thrombin cleavage site are indicated.

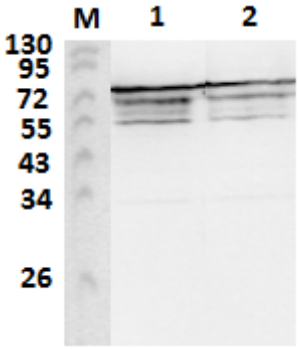

Figure B6 Western blot analysis of DnaK contaminants in purified $(His)_7$–PINIT protein. Western blot analysis of DnaK using mouse anti–DnaK monoclonal antibody and HRP–conjugated goat anti–mouse IgG antibodies: lane 1, molecular mass marker; lane 2 first elution batch of $(His)_7$–PINIT; lane 2, second elution batch of $(His)_7$–PINIT. The western blot analysis showed multiple species of contaminating DnaK protein in the purified $(His)_7$–PINIT.

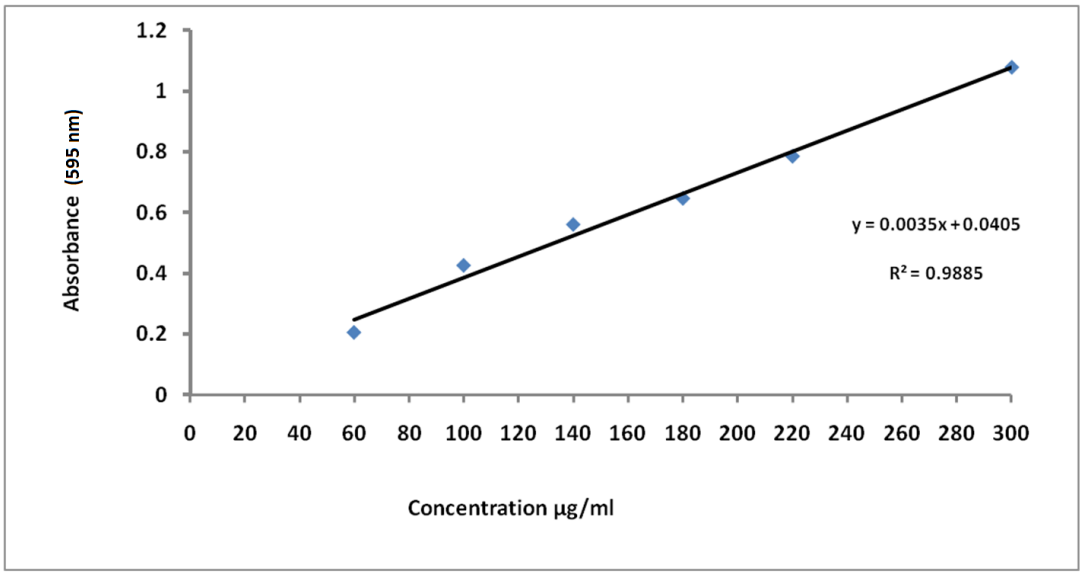

Figure B7 Bradford standard curve for protein concentration determination. Bradford standard curve for protein concentation determination prepared using varying BSA concentartion and Bradford reagents. Absorbance of the samples was read at 595 nm in a PowerWave^{TM} Microplate spectrophotometer (Biotek). The curve is represented by the linear equation: $y = 0.0035x - 0.0405$; $R^2 = 0.9885$.

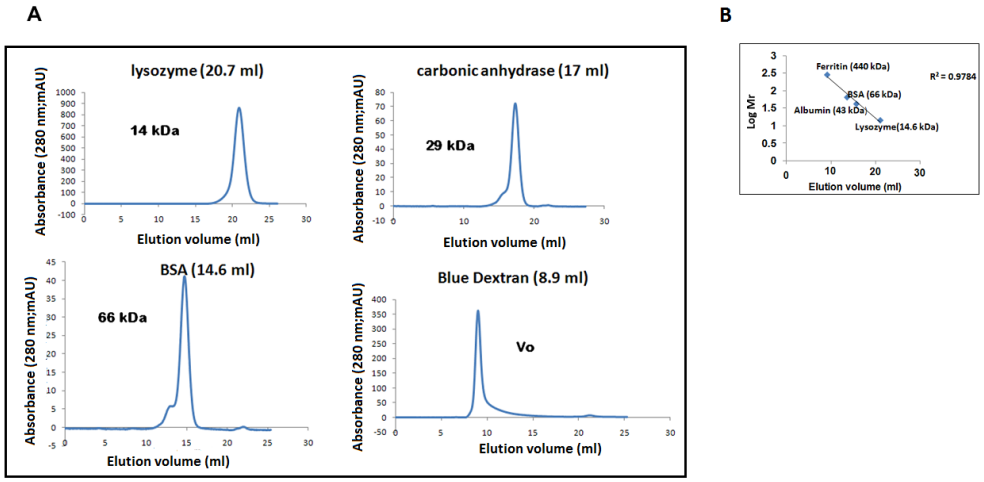

Figure B8 Size exclusion chromatography of the protein standards. A) gel filtration chromatography traces of lysozyme, carbonic anhydrase, BSA, and Blue dextran for standard curve plot to determine the molecular mass of the PINIT domain protein. **(B)** Protein standard curve plot of log molecular weight of the proteins in **Figure** (3A) against their retention volume for $(His)_7$–PINIT molecular mass determination. The molecular mass of the $(His)_7$–PINIT peak fraction was found to be approximately 23 kDa.

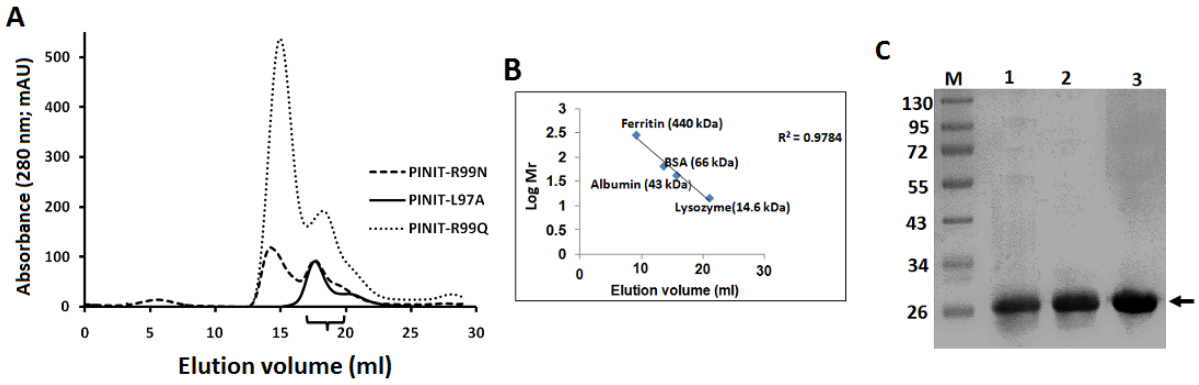

Figure B9 Size exclusion chromatography of PINIT domain mutants. A) Size exclusion chromatography of $(His)_7$–PINIT–L97A, $(His)_7$–PINIT–R99N and $(His)_7$–PINIT–R99Q proteins proteins. All mutants were eluted at the same elution volumes under the second peak indicated by a bracket. **B**) Standard curve plot used for molecular mass determination of the eluted mutant proteins. The molecular mass of the mutant $(His)_7$–PINIT proteins were determied as approximately 23 kDa. **C**) 12 % SDS–PAGE analysis of the concentrated fractions 18 and 19 ml of the mutant proteins. M, molecular mass marker; lane 1, $(His)_7$–PINIT–R99N; lane 2, $(His)_7$–PINIT–L97A; lane 3, $(His)_7$–PINIT–R99Q.

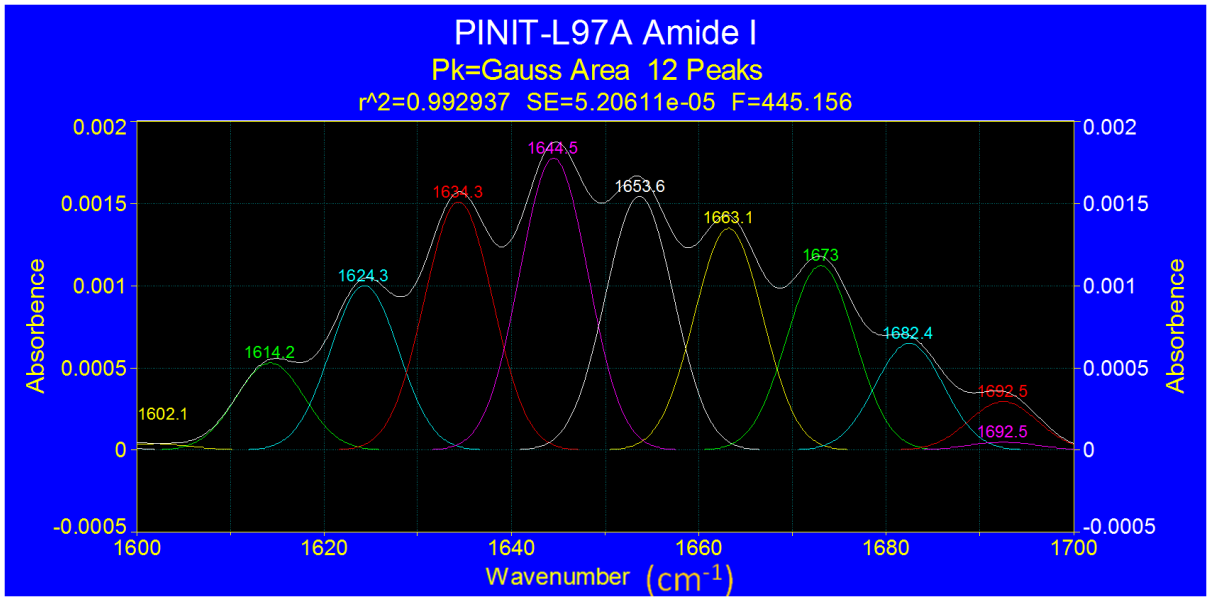

Figure B10 Secondary structure analysis of $(His)_7$–PINIT–L97A protein in H_2O environment. Amide I region infrared spectra of $(His)_7$–PINIT–L97A protein in H_2O environment was deconvoluted and the peaks fitted with the Gaussian curve Gaussian bands are shown as symmetrical peaks underneath the deconvolved IR spectra. The peak wavelength numbers are shown at the on each peak.

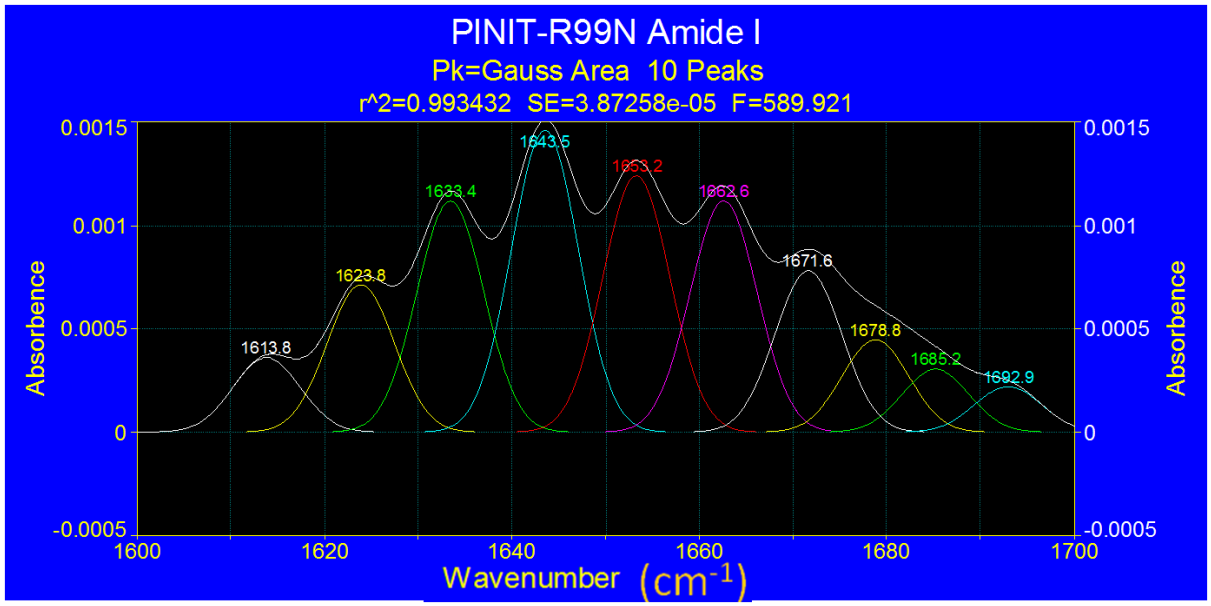

Figure B11 Secondary structure analysis of $(His)_7$–PINIT–R99N protein in H_2O environment. Amide I region infrared spectra of $(His)_7$–PINIT–R99N protein in H_2O environment was deconvoluted and the peaks fitted with the Gaussian curve Gaussian bands are shown as symmetrical peaks underneath the deconvolved IR spectra. The peak wavelength numbers are shown at the on each peak.

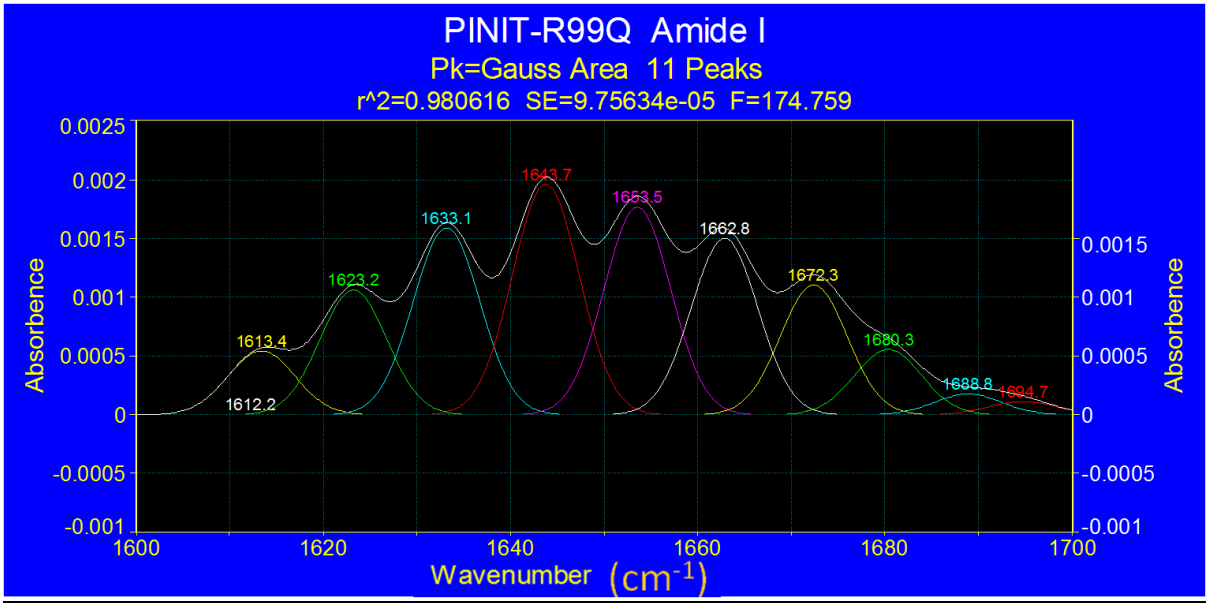

Figure B12 Secondary structure analysis of $(His)_7$–PINIT–R99Q protein in H_2O environment. Amide I region infrared spectra of $(His)_7$–PINIT–R99Q protein in H_2O environment was deconvoluted and the peaks fitted with the Gaussian curve Gaussian bands are shown as symmetrical peaks underneath the deconvolved IR spectra. The peak wavelength numbers are shown at the on each peak.

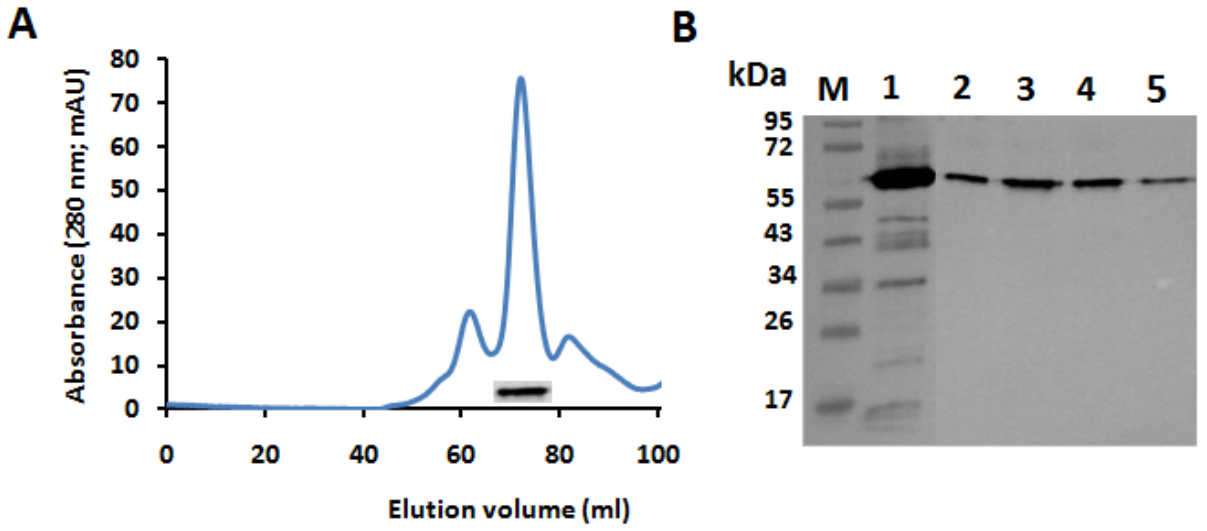

Figure B14 Size exclusion chromatography of ammonium sulphate purified STAT3β protein. A) Size exclusion chromatography of STAT3β protein fraction purified by ammonium sulphate precipitation. Western blot analysies to detect STAT3β was performed usisng mouse monoclonal anti-STAT3 IgG antibodies and HRP conjugated goat anti-mouse IgG antibodies. **B)** SDS-PAGE analysis of purified STAT3β and size exclusion fractions; molecular mass markers are indicated on the left side; lane 1, STAT3β purified by ammonium sulphate precipitation as decdribed in section 4.2.9; line 2–5, fraction 68, 70, 72 and 74 ml respectively respectively from the size exclusion column.

APPENDIX C: SUPPLEMENTARY TABLES

Table C1 Deconvoluted amide I band frequencies and assignments to secondary structure for protein in D_2O and H_2O media. Data adapted form Dong *et al*. (1992); Susi *et al*. (1986).

$H_2O^{\dagger}$		$D_2O^{\ddagger}$	
Mean frequencies	Assignment	Mean frequencies	Assignment
1624±1.0	β-sheet	1624±4.0	β-sheet
1627±2.0	β-sheet		
1633±2.0	β-sheet	1631±3.0	β-sheet
1638±2.0	β-sheet	1637±3.0	β-sheet
1642±1.0	β-sheet	1641±2.0	3_{10} Helix
1648±2.0	Random	1645±4.0	Random
1656±2.0	α Helix	1653±4.0	α-Helix
1663±3.0	3_{10} Helix	1663±4.0	β-Turn
1667±1.0	β-Turn	1671±3.0	β-Turn
1675±1.0	β-Turn	1675±5.0	β-sheet
1680±2.0	β-Turn	1683±2.0	β-Turn
1685±2.0	β-Turn	1689±2.0	β-Turn
1691±2.0	β-sheet	1694±2.0	β-Turn
1696±2.0	β-sheet		

Table C2 Fractional band areas (% Area), frequencies (wave number) and band assignments of FTIR Amide I component bands of PINIT mutants; PINIT–L97A, PINIT–R99N, PINIT–R99Q and the native PINIT domain protein in H_2O.

PINIT-L97A			PINIT-R99N			PINIT-R99Q			PINIT		
% Area	Wave number (cm^{-1}) H_2O	Assignment	% Area	Wave number (cm^{-1}) H_2O	Assignment	% Area	Wave number (cm^{-1}) H_2O	Assignment	% Area	Wave number (cm-1) H_2O	Assignment
9.9	1624.3	β-sheet	9.2	1623.8	β-sheet	10.2	1623.2	β-sheet	12.1	1625.1	β-sheet
15.0	1634.3	β-sheet	14.4	1633.4	β-sheet	15.3	1633.1	β-sheet	16.9	1634.7	β-sheet
17.6	1644.5	β-sheet	18.8	1643.5	β-sheet	18.9	1643.7	β-sheet	17.9	1644.0	β-sheet
15.3	1653.6	α-Helix	16.0	1653.2	α-Helix	17.0	1653.5	α-Helix	15.9	1653.3	α-Helix
13.4	1663.1	3_{10}-Helix	14.4	1662.6	3_{10}-Helix	14.5	1662.8	3_{10}-Helix	13.2	1662.9	3_{10}-Helix
11.1	1673.0	β-Turn	10.0	1671.6	β-Turn	10.6	1672.3	β-Turn	8.6	1672.8	β-Turn
6.5	1682.4	β-Turn	5.8	1678.8	β-Turn	5.4	1680.3	β-Turn	2.9	1691.5	β-sheet
2.9	1692.5	β-sheet	3.9	1685.2	β-Turn	1.7	1688.8	β-Turn			
			2.9	1692.6	β-sheet	1.1	1693.9	β-sheet			

Table C3 Relative content of secondary structures quantified from fractional band areas (% Area) of FTIR amide I secondary structure features of the PINIT domain, PINIT–L97A, PINIT–R99N and PINIT–R99Q in H2O solution.

	β–sheet (% Area)	**α–helix (% Area)**
PINIT–L97A	45.4	15.3
PINIT–R99N	45.3	16
PINIT–R99Q	45.5	17
PINIT	45.2	16

APPENDIX D: GENERAL EXPERIMENTAL PROCEDURES OF STANDARD MOLECULAR BIOLOGY TECHNIQUES

D1: ISOLATION OF PLASMID DNA

The protocol for isolation of plasmid DNA was adapted from that described QIAprep Miniprep handbook (QIAGEN) plasmid isolation kit. In brief, *E. coli* cells transformed with the plasmid of interest were grown overnight at suitable temperature and in 5 ml cultures of LB media (1% tryptone, 0.5% yeast extract, 1% NaCl) supplemented with the appropriate antibiotic for plasmid selection at concentration of 100 µg/ml. The cells were harvested in a microcentrifuge (~17900 x *g* for 1 minute) and the cell pellet was resuspended in 250 µl of resuspension buffer (P1). Lysis buffer 250 µl (LyseBlue reagent P2) was added and mixed by inverting the tube 4–6 times, and subsequently 350 µl of neutralisation buffer (N3) was added. This was followed by centrifugation ((~17900 x *g* for 10 minutes) and the supernatant was applied to the QIAprep spin column by pipetting and centrifuged for 30–60 seconds and discard the flow through. The column was washed with by adding 500 µl of wash buffer (PB) and centrifuging for 30–60 seconds and discards the flow–through. The second was performed by adding 750 µl buffer (PE) and centrifuging for 30–60 seconds. The flow–through was discarded and centrifuged for an additional 1 minute to remove residual wash buffer. The DNA was eluted in a clean 1.5 ml microcentrifuge tube by adding 50 µl elution buffer (EB) or water to the centre of the QIAprep spin column and let it stand for 1 minute and centrifuge for 1 minute at ~17900 x *g*.

D2: ISOLATION OF ENDOTOXIN FREE PLASMID DNA FROM *E. COLI* FOR TRANSFECTION IN MAMMALIAN CELLS

The protocol for isolation of endotoxin free plasmid DNA was adapted from that described in the GenElute^{TM} Endotoxin–free plasmid Midiprep kit (Sigma Aldrich). In brief, pellet 40 µl of overnight recombinant *E. coli* culture by centrifugation at 5 000 x *g*. The bacterial pellet was resuspended to homogenous with 1.2 ml of resuspension solution. The resuspended cells are lysed by adding 1.2 ml of lysis solution and immediately mix the contents by gentle inversion 6–8 times. The debris were precipitated by adding 0.8 ml of neutralization solution and contents were immediately mixed thoroughly by gentle inversion followed by centrifuging at ≥15 000 x *g* for 15 minutes at 2–8^{o}C. The supernatant was transferred to clean

tube. The removal of endotoxin was performed by adding 300 µl of endotoxin removal solution to the supernatant and thoroughly mixed by inversion for 1 minute followed by incubation on ice for ≥5 minutes; with 1–2 times mixing during the ice incubation. The tube was then warmed at 37°C in water bath for 5 minutes followed by centrifugation at 3 000–5 000 x *g* for 5 minutes at room temperature. The clear upper phase which contains plasmid DNA was transferred to the clean tube and the process repeated. The clear upper phase was transferred to a clean tube and 0.8 ml of DNA binding solution was added and mixed thoroughly by inversion or by votexing and transferred to a GenElute Midprep binding column in a collection tube and centrifuged at 3 000–5 000 x *g* for 1–2 minutes. The flow through was discarded and washed by 2 ml of optional was solution by centrifuging at 3 000–5 000 x *g* for 2 minutes followed by a 3 ml was with washing solution concentrate at 3 000–5 000 x *g* and discard the flow through and spin for an extra 1 minutes to remove residual wash solution. The DNA was eluted by transferring the column to a clean collection tube and adding 0.8 ml of pre–warmed (at 65^{o}C) endotoxin–free water directly to the binding filter allowing the water to soak for 10 minutes before centrifuging at 3 000–5 000 x *g* for 3 – 5 minutes to elute the DNA.

D3: DNA DIGESTION WITH RESTRICTION ENZYMES

Plasmid DNA was digested with the appropriate restriction endonuclease(s) for three hours at the 37°C in a digestion reaction comprising: 200 – 500 ng of plasmid DNA, 1X restriction buffer and 1 – 2 U of restriction endonuclease enzyme and distilled water to a final volume of 20 µl. The digested DNA was resolved by agarose gel electrophoresis. Restriction buffers for single and double restriction enzyme digestions were selected as per the supplier's recommendations. PstI–digested λDNA marker was prepared by the digestion of 20 µl of 526 µg.ml–1 λDNA (Promega) for three hours at 37°C in a reaction containing 5 U of *Pst*I restriction enzyme (Fermentas), 20 µl of the appropriate 10x restriction enzyme buffer (Fermentas) and distilled water to a final volume of 200 µl. The digested λDNA was treated with 6x DNA gel loading buffer (0.25% (w/v) bromophenol blue, 30% (v/v) glycerol) for use in subsequent agarose gel electrophoresis.

D4: AGAROSE GEL ELECTROPHORESIS

Agarose gels were prepared by melting molecular grade agarose (0.8 % or 1 % (w/v)) in TBE Buffer (45 mM Borate, 1 mM EDTA, 45 mM Tris–Cl, pH 8.3) and supplementing ethidium bromide to a final concentration of 0.5 µg/ml on cooling prior to casting. DNA samples for electrophoresis were treated with 6x DNA gel loading buffer (0.25% (w/v) bromophenol blue, 30% (v/v) glycerol) and loaded onto the gel with an appropriate marker of *Pst*I–digested λDNA. The samples were resolved at 100 V and visualised under ultra–violet light with a Chemidoc Imaging System (Bio–Rad).

D5: EXTRACTION AND PURIFICATION OF DNA FROM AN AGAROSE GEL

Resolved DNA fragments were isolated subsequent to agarose gel electrophoresis using the Zymoclean Gel DNA recovery kit (ZYMO RESARCH) as per the manufacturer's instructions. In brief, the DNA fragment of interest was identified by brief exposure to long–wave UV light, excised from the gel using a sterile razor blade and transferred to a 1.5 ml microcentrifuge tube. To the microcentrifuge tube, 3 volumes of ADB buffer to each volume of agarose excised from the gel was added and incubated at 55°C for 10 minutes until the gel slice was completely dissolved. The melted agarose solution was transferred to a Zymo–spin column in a collection tube. This was centrifuged at ≥10 000 x *g* for 30 – 60 seconds and the flow–through was discarded. The column was washed twice with was buffer and centrifuged at ≥ 10 000 x *g*. The flow–through was discarded and centrifuged for an additional 1 minute to remove residual wash buffer. Water or elution buffer ≥ 6 µl was added directly to the column matrix and placed into a 1.5 ml tube and centrifuged at ≥10 000 x *g* for 30 – 60 seconds to elute DNA.

D6: LIGATION OF DNA FRAGMENTS

DNA fragments intended for ligation (typically 500 ng of insert fragment to 100 ng of target plasmid) were incubated overnight at 4°C in a ligation reaction comprising 1 µl of 10x ligation buffer (Roche Applied Sciences), 1 U of T4 DNA Ligase (Roche Applied Sciences) and distilled water to a final volume of 10 µl. The ligation reaction was transformed into competent *E. coli* cells.

D7: DNA SEQUENCING

Plasmid DNA was isolated for DNA sequencing using the QiaPrep^{R} Miniprep Kit (Qiagen) as per the manufacturer's instructions (**APPENDIX**). Sequencing reactions comprised the plasmid DNA (350 ng), 3.2 pmol of primer (forward or reverse primer), 2 µl of 5 x Big DyeR Terminator Sequencing Buffer (Big Dye Terminator Cycle Sequencing Kit version 3.1, Applied Biosystems), 4 µl of Big Dye Terminator (Applied Biosystems) and distilled water to a final volume of 10 µl. Thermal cycling was perfomed in a GeneAmp PCR System 9700 (version 3.05; Applied Biosystems) as follows: one cycle of denaturation (96 °C, 2 minutes), 30 cycles of denaturation, annealing and extension (96 °C for 30 seconds, appropriate annealing temperature for 30 seconds, and 72°C for 30 seconds) and a final elongation at 72^{o}C for 7 minutes. Purification of the amplification product from unincorporated big dye terminators was achieved with Zymo–Spin I TM columns (Zymo Research) as per the manufacturer's instructions. The DNA was eluted in 15 ul of water and vacumm dried. The purified DNA resuspended in Hi–Di buffer for sequencing in ABIPRISM 3100 Genetic Analyser (Applied Biosystems, USA) and analysed by capillary electrophoresis. DNA sequencing results were analysed using BioEdit Sequence Alignment Editor (version 7.0.4.1).

D8: DNA CLEAN AND CONCENTRATOR

DNA cleaning up and concentration was performed using a DNA clean and Concentrator–5^{TM} kit (ZYMO RESEARCH) as per the manufacturer's instructions. In brief, in 1.5 ml microcentrifuge tube 2 volumes of DNA binding buffer to each volume of DNA was added and mixed briefly by votexing. The mixture is then transferred to a Zymo–spin^{TM} column in a collection tube and centrifuged at ≥ 10 000 rpm for 30 seconds and the discard the flow–through. The column was washed twice with 200 µl Wash buffer, centrifuging at ≥ 10 000 rpm for 30 seconds. The DNA was eluted by adding 6–10 µl of water directly to the column matrix and transfer the column to the 1.5 ml microcentrifuge and centrifuged at ≥ 10 000 rpm for 30 seconds to elute the DNA.

D9: NICKEL AFFINITY COLUMN STRIPPING AND RECHARGING

The column was stripped by washing with at least 5–10 column volumes of stripping buffer (20 mM sodium phosphate, 0.5 M NaCl, 50 mM EDTA, pH 7.4) followed by 5 –10 column volumes of distilled water before recharging the column. The column is recharged by loading 0.5 ml or 2.5 ml of 0.1 M NiSO4 in distilled water on HisTrap HP 1 ml and 5 ml column, respectively. The column is then washed with 5 column volumes of distilled water and 5 columns of binding buffer (20 mM sodium phosphate, 0.5 NaCl, 20–40 mM imidazole, pH 7.4) (to adjust the pH) before storage in 20 % ethanol.

D10: PREPARATION OF COMPETENT *E. COLI* CELLS

The strain of interest was grown overnight at 37°C in 5 ml of LB media (1% tryptone, 0.5% yeast extract, 1% NaCl). The resulting overnight culture was diluted into 50 ml of LB media to an A_{600} of 0.1 and allowed to grow until early log phase (A_{600} of 0.3 – 0.6). The cells were harvested by centrifugation at 5000 x g for 5 minutes at 4°C and resuspended in 50 ml of ice–cold 0.1 M $MgCl_2$ and incubated on ice for 2 minutes. The cells were then pelleted by centrifugation as before and resuspended in 25 ml of ice–cold 0.1 M $CaCl_2$ and incubated on ice for one hour. The cells were centrifugation as before and resuspended in 5 ml of 0.1 M $CaCl_2$ and 5 ml of 30% (v/v) glycerol. The competent cells were aliquoted in 300 µl volumes in microcentrifuge tubes and stored at –80°C prior to use.

D11: TRANSFORMATION OF COMPETENT *E. COLI* CELLS

Competent *E. coli* cells 300 µl were incubated with 100 ng of the plasmid DNA of interest 1–2 µl of ligation product at 4°C for 30 minutes, followed by heat shock at 42°C for 45 seconds and incubation on ice for 5 minutes. The cells were diluted with 700 µl of LB media (1% tryptone, 0.5% yeast extract, 1% NaCl) and incubated at 37°C for one hour with shaking. Transformed cells are pelleted by centrifugation and dicard the supernatant and resuspend the pellet in 100 µl fresh LB media. The bacterial suspension (100 µl) was plated onto 1.5% agar in LB media supplemented with the appropriate antibiotics at desired concentration. The plates were incubated overnight at 37°C. Transformation controls included a sterile control with sterile distilled water replacing the plasmid DNA in the incubation mixture, and a competence control with plasmid DNA of known concentration (10 ng of plasmid pUC18; Promega) transformed into the *E. coli* cells.

D12: PROTEIN CONCENTRATION DETERMINATION BY BRADFORD'S ASSAY

Protein concentration determination was performed by Bradford's assay (Bradford, 1976). A volume of 200 µl of Bradford's reagent (Bio–Rad; diluted 1:4 with distilled water) was added to 10 µl of undiluted, 1:10 and 1:100 diluted protein samples of unknown concentration. Following incubation at room temperature for 10 minutes, the Absorbance of the samples was read at 595 nm in a PowerWave^{TM} Microplate spectrophotometer (Biotek). Protein concentration was determined with the corresponding 0 – 250 µg.ml–1 Bovine Serum

D13: BRADFORD ASSAY FOR PROTEIN CONCENTRATION DETERMINATION

Protein concentrations were determined by Bradford assay method (Bradford 1976). Bovine serum albumen (BSA) was used as the standard contain a range of 20 to 300 µg/ml concentration to volume of 100 µl and 100 µl of Bradford reagent was added the protein samples were prepared in the same manner. Standards were prepared in triplicate and incubated at room temperature for 5 minutes and absorbance was read at 595 nm using a PowerWave (PowerWave^{TM} Microplate spectrophotometer (Biotek)).

D14: SDS–PAGE

Protein samples were treated with 5x SDS–PAGE sample buffer (10% glycerol, 2% SDS, 5% β–mercaptoethanol, 0.05% bromophenol blue, 0.0625 M Tris, pH 6.8) in a ratio of 4:1 respectively and loaded onto a polyacrylamide gel constituted by a resolving gel (10–12% (w/v) acrylamide, 0.1% (w/v) SDS, 0.05% (w/v) ammonium persulphate (APS), 0.005% (v/v) N,N,N',N'–tetramethylethylenediamine (TEMED), 0.375 M Tris, pH 8.8) and a stacking gel (4% (w/v) acrylamide, 0.1% (w/v) SDS, 0.05% (w/v) APS, 0.005% (v/v) TEMED, 0.125 M Tris, pH 6.8). The gel was resolved in a Mini Protean^{R} II system (Bio–Rad) at 160 V for one hour and stained or used for Western blot analysis. Staining of the SDS–PAGE gel was performed using Coomassie Blue stain (40% (v/v) methanol, 7% (v/v) acetic acid, 0.25% (w/v) Coomassie Blue R250 in distilled water) for one hour and destained for two hours using destaining solution (40% (v/v) methanol, 7% (v/v) acetic acid in distilled water).

D15: PROTEIN DETECTION BY WESTERN BLOT ANALYSIS

The protocol for the detection of proteins by Western blot analysis was adapted from Amersham ECL Advanced Western blotting detection Kit (GE Healthcare). Proteins were resolved by SDS–PAGE and transferred onto nitrocellulose membrane (Hybond C–extra; GE Healthcare) in transfer buffer (20% (v/v) methanol, 192 mM glycine, 25 mM Tris) at 100 V for two hours in a PowerPack western transfer blotting system (BioRad). Protein transfer was verified with Ponceau stain; (0.5 % (w/v) Ponceau, 1% (v/v) glacial acetic acid). The membrane was subsequently destained with distilled water and blocked overnight at 4°C in blocking solution (5% (w/v) fat–free milk powder in Tris Buffered Saline (TBS; 50 mM Tris, 150 mM NaCl, pH 7.5). The membrane was incubated with appropriate primary antibody (at desired dilution ration ranging from 1: 500–5 000) for one hour at room temperature or overnight at 4^{o}C. The membrane was washed three times with Tris Buffered Saline–Tween buffer (TBS–T; TBS containing 0.1% (v/v) Tween 20) and incubated with the appropriate horse–radish peroxidise (HRP)–conjugated secondary antibody (1:5000 in blocking solution) for one hour at room temperature and washed three times for 15 minutes each wash with TBS–T. The protein was detected by chemiluminescence–based protein detection ECL Western blotting kit (GE Healthcare) following manufacturer's instructions, and images were captured with a Chemidoc chemiluminescence imaging system (BioRad, UK).

D16: MEMBRANE STRIPPING AND REPROBING

The protocol for membrane stripping and reprobing for proteins by Western blot analysis was adapted from Amersham ECL Advanced Western blotting detection Kit (GE Healthcare). In brief, the membrane is submerged in stripping buffer (100 mM 2–mercaptoethanol, 2 % (w/v) SDS, 62.5 mM Tris–HCl pH 6.7) and incubates at 50^{o}C for 30 minutes with occasional agitation. The membrane was washed twice for 10 minutes in PBS–T (1 M Tris HCl, pH 7.6; 100 mM NaCl, 0.1 % (v/v) Tween 20) at room temperature. The membrane was blocked in blocking solution (5 % (w/v) no–fat powder milk and PBS–T) for one hour at room temperature. The immunodetection was carried on as described in **Appendix C14.**

APPENDIX E: LIST OF MATERIALS AND SPECIALISED REAGENTS

ANTIBODIES	SUPPLIER
Alexa Fluor^{R} 488 donkey anti–mouse	Invitrogen, USA
Alexa Fluor^{R} 488 chicken anti–rabbit	Invitrogen, USA
Alexa Fluor^{R} 546 donkey anti–mouse	Invitrogen, USA
HRP–goat anti–mouse IgG	GE Healthcare, UK
HRP–conjugated sheep anti–mouse	GE Healthcare, UK
HRP–conjugated donkey anti–rabbit	GE Healthcare, UK
Mouse anti–GST monoclonal antibody	Santa Cruz biotechnology, USA
Mouse anti–FLAG M2 monoclonal antibodies	Sigma–Aldrich, USA
Mouse monoclonal anti–His primary antibody	GE Healthcare, UK
Mouse Anti–DnaK Monoclonal Antibody	Sigma–Aldrich, USA
mouse monoclonal ant-PIAS3 IgG	Santa Cruz biotechnology, USA
mouse monoclonal anti-STAT3 IgG antibody	Santa Cruz biotechnology, USA
rabbit polyclonal anti-STAT3 IgG	Santa Cruz biotechnology, USA

REAGENT	SUPPLIER
β–mercaptoethanol	Merck, Germany
λDNA	Promega, USA
Acetic Acid	Saarchem, South Africa
Adenosine triphosphate (disodium salt)	Sigma–Aldrich, USA
Agar (Bacteriological)	Biolab Diagnostics, South Africa
Agarose	Hispanagar, Spain
Ammonium per sulphate	Saarchem, South Africa
Ampicillin	Fisher Scientific, UK
30% Bis–Acrylamide	Bio–Rad, US
Bovine Serum Albumin	Sigma–Aldrich, USA

Boric Acid	Saarchem, South Africa
Bradford's Reagent	Bio–Rad, USA
Bromophenol Blue	Sigma–Aldrich, Germany
Calcium chloride	Saarchem, South Africa
Amicon^{R} Ultra Ultracel^{R} Centrifugal Filters	Millipore, Ireland
Coomassie Brilliant Blue R250	Sigma–Aldrich, Germany
Hoescht	Invitrogen, USA
dNTP mix	Roche Applied Sciences, Germany
EDTA, sodium salt	Saarchem, South Africa
Ethanol	Saarchem, South Africa
Ethidium bromide	Sigma–Aldrich, Germany
Glacial acetic acid	Saarchem, South Africa
Glycerol	EMD Chemicals, USA

REAGENT	**SUPPLIER**
Glycine	Sigma–Aldrich, Germany
HEPES	Fisher Scientific, UK
Hybond C–extra	GE Healthcare, UK
Hydrochloric Acid	Saarchem, South Africa
Imidazole	Sigma–Aldrich, Germany
Isopropyl–1–thio–β–D–galactopyranoside	Roche Applied Sciences, Germany
Kanamycin sulphate	Roche Applied Sciences, Germany
Lysozyme	Sigma–Aldrich, USA
Methanol	Saarchem, South Africa
Pepstatin A	Sigma–Aldrich, USA
Phenylmethylsulphonyl fluoride (PMSF)	Sigma–Aldrich, USA
Polyacrylamide	Bio–Rad, USA
Ponceau S	Sigma–Aldrich, Germany
Potassium chloride (KCl)	Saarchem, South Africa

Potassium hydroxide (KOH)	Saarchem, South Africa
Potassium phosphate (K_2HPO_4)	Merck, Germany
Potassium dihydrogen phosphate (KH_2PO_4)	Merck, Germany
Protein A/G PLUS Agarose	Santa Cruz Biotechnology, USA

REAGENT	**SUPPLIER**
Q–Sepharose Fast Flow™	Sigma–Aldrich, USA
Sepharose Fast Flow™	GE Healthcare, UK
Sodium chloride (NaCl)	Saarchem, South Africa
Sodium dodecyl sulphate (SDS)	Sigma–Aldrich, USA
Sodium phosphate (NaH_2PO_4 / Na_2HPO_4)	Saarchem, South Africa
Sodium hydroxide (NaOH)	Saarchem, South Africa
Snakeskin™ dialysis tubing	Thermo Scientific, USA
TEMED (N,N,N',N'–tetramethylethylenediamine)	Sigma–Aldrich, Germany
Tris (Tris–2–amino–2–hydroxymethyl–1,3–propanol)	Sigma–Aldrich, Germany
Tryptone	Oxoid, UK
Tween 20	Saarchem, South Africa
Urea	Sigma–Aldrich, Germany
Yeast extract	Oxoid, UK

RESTRICTION ENZYMES	**SUPPLIER**
*BamH*I	Fermentas, Lithuania
*Bgl*II	Promega, USA
*Dpn*I	Promega, USA
*Hind*III	Fermentas, Lithuania
Nde I	Promega, USA
*Nhe*I	GE Healthcare, UK
*Pst*I	Fermentas, Lithuania
*Sal*I	Fermentas, Lithuania

***E. COLI* STRAINS**	**SUPPLIER**
E. coli BL21(DE3)	Promega, USA
E. coli JM109	Promega, UK
E. coli Rosetta	Novagen, USA
E. coli XL1–Blue	Stratagene, USA
E. coli M15[pREP4]	Qiagen (USA)

PLASMIDS	**SUPPLIER**
pGEM–T Easy^{R}	Promega, USA
pQE2	Qiagen, USA
pQE60	Qiagen, USA
pGEX4T–N1	Promega, USA

ENZYMES	**SUPPLIER**
Expand High Fidelity Taq Polymerase	Roche Applied Sciences, Germany
10x Buffer	Roche Applied Sciences, Germany
T4 DNA Ligase	Roche Applied Sciences, Germany
Ligation buffer	Roche Applied Sciences, Germany
Pfu Dna Polymerase and 10x Buffer	Promega, USA

COMMERCIAL KITS	**SUPPLIER**
Big Dye^{TM} Terminator Cycle Sequencing Kit	Applied Biosciences, UK
ECL Western Blotting Kit	GE Healthcare, UK
Zymoclean^{TM} Gel DNA Recovery Kit	Zymo Research,USA
DNA Clean & Concentrator–5^{TM}	Zymo Research, USA
GenElute^{TM} Endotoxin–free Plsmid MidiPrep Kit	Sigma–Aldrich, USA
QIAPrep^{R} Miniprep Kit	Qiagen, USA
HisTrap HP	GE Healthcare, UK

PROTEIN MARKERS	SUPPLIER
Pageruler^{TM} Protein Ladder	Fermentas, USA
Protein marker II peqGold	Fermentas, USA
Protein marker IV pre–stained	Invitrogen, USA

PRIMERS

All primers were synthesised by :Integrated DNA Technologies (IDT, USA) through WhiteSci, South Africa.

:Inqaba Biotechnology, South Africa

APPENDIX F: WEB BASED BIOINFORMATICS ANALYSIS TOOLS AND PYTHON SCRIPTS.

F1: WEB BASED BIOINFORMATIC ANALYSIS TOOLS

http://expasy.org/tools/pi_tool.html– Compute the theoretical isoelectric point (pI) and molecular weight (Mw) from a UniProt Knowledgebase entry or for a user sequence

http://blast.ncbi.nlm.nih.gov/Blast.cgi– Search **protein** database using a **protein** query *Algorithms:* blastp, psi–blast, phi–blast

http://toolkit.tuebingen.mpg.de/hhpred– Homology detection & structure prediction by HMM–HMM comparison

http://toolkit.tuebingen.mpg.de/clustalw– ClustalW is a general purpose multiple alignment program for DNA or proteins.

http://bioinf.cs.ucl.ac.uk/psipred/psiform.html– Secondary structure prediction using neural networks. PSIPRED is one of the most popular and accurate methods around.

http://toolkit.tuebingen.mpg.de/psi_blast– Search with an amino acid sequence against protein databases for locally similar sequences

https://genesilico.pl/toolkit/unimod?method=MetaMQAPII– MetaMQAPII is a metaserver for quality assessment of protein structures optimized for theoretical models

https://genesilico.pl/toolkit/unimod?method=Verify3D– David Eisenberg's Verify3D method for the assessment of protein models with three–dimensional profiles."

https://genesilico.pl/toolkit/unimod?method=Prosa– Knowladge based mean fields based method.

http://www.pymol.org/– a highly extensible program for interactive visualization and analysis of molecular structures

http://www.cgl.ucsf.edu/chimera/– UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures

F2: PYTHON SCRIPT USED IN MODELLER 9V3

Python script to generate 100 models from the template structure

```
# Homology modeling by the automodel class

from modeller.automodel import *    # Load the automodel class

log.verbose()                       # request verbose output
env = environ()# create a new MODELLER environment to build this model in

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'


a = automodel(env,
 alnfile = 'alignment_hhpred.pir',                    # alignment filename
 knowns = ('3i2D'), # codes of the templates
       sequence = 'PIAS3_PINIT_domain', assess_methods=(assess.DOPE,
assess.GA341)) # code of the target
a.starting_model= 1 # index of the first model
a.ending_model = 100 # index of the last model
 # (determines how many models to calculate)
a.final_malign3d = True             # generate superimposed templatesand
model (*_fit.pdb files)
a.make() # do the actual homology modeling
ok_models = filter(lambda x: x['failure'] is None, a.outputs)
      # Get a list of all successfully built models from a.outputs
key = 'DOPE score'                  # Rank the models by DOPE score
ok_models.sort(lambda a,b: cmp(a[key], b[key]))
m = ok_models[0]                    # Get top model
print "1st top model: %s (DOPE score %.3f)" % (m['name'], m[key])
ms = ok_models[1]        # Get 2nd top model
print "2nd top model: %s (DOPE score %.3f)" % (ms['name'], ms[key])
mss = ok_models[2]                  # Get 3rd top model
print "3rd top model: %s (DOPE score %.3f)" % (mss['name'], mss[key])
```

Python script to calculate Zdope scores

```
import subprocess

ofile = open("zdope_scores.txt","w")
ofile.write("z-DOPE-score filename\n")
ofile.close()
models = []
for model in open("modellist").readlines():
 models.append(model.strip())
for model in models:
 subprocess.call("mod9v7 zdope_single.py "+model,shell=True)
 subprocess.call("mv zdope_single.log zdope."+model[:-4],shell=True)
exit
#print models
```

Python script to sort the calculated Zdope scores

```
ifile=open("zdope_scores.txt")
lines = ifile.readlines()
ifile.close()
scores = []
for line in lines:
 scores.append(line.rsplit())

scores.sort()
ofile=open("sorted_zdope_scores.csv","w")
for line in scores:
 ofile.write(str(line[0])+","+line[1]+",\n")
```

Python script for loop refinement of the existing model

```
# Loop refinement of an existing model
from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = 'Model_HHpred'

# Create a new class based on 'loopmodel' so that we can redefine
# select_loop_atoms (necessary)
class MyLoop(loopmodel):
 # This routine picks the residues to be refined by loop modeling
 def select_loop_atoms(self):
 # 10 residue insertion
 return selection(self.residue_range('70', '79'))

m = MyLoop(env,
 inimodel='PIAS3_PINIT_domain.B99990046.pdb', # initial model of the
target
 sequence='PINIT_46') # code of the target

m.loop.starting_model= 1 # index of the first loop model
m.loop.ending_model = 50 # index of the last loop model
m.loop.md_level = refine.very_slow # loop refinement method; this
yields
 # models quickly but of low quality;
 # use refine.slow for better models

m.make()
```

APPENDIX G: PROTEIN AND NUCLEOTIDE SEQUENCES

G1: NUCLEOTIDE SEQUENCES

>PIAS3 coding sequence

ATGGCGGAACTGGGCGAACTGAAACATATGGTGATGAGCTTTCGCGTGAGCGAACTGCAGGTGCTGCT
GGGCTTTGCGGGCCGCAACAAAAGCGGCCGCAAACATGAACTGCTGGCGAAAGCGCTGCATCTGCTGA
AAAGCAGCTGCGCGCCGAGCGTGCAGATGAAAATTAAAGAACTGTATCGCCGCCGCTTTCCGCGCAAA
ACCCTGGGCCCGAGCGATCTGAGCCTGCTGAGCCTGCCGCCGGGCACCAGCCCGGTGGGCAGCCCGGG
CCCGCTGGCGCCGATTCCGCCGACCCTGCTGACCCCGGGCACCCTGCTGGGCCCGAAACGCGAAGTGG
ATATGCATCCGCCGCTGCCGCAGCCGGTGCATCCGGATGTGACCATGAAACCGCTGCCGTTTTATGAA
GTGTATGGCGAACTGATTCGCCCGACCACCCTGGCGAGCACCAGCAGCCAGCGCTTTGAAGAAGCGCA
TTTTACCTTTGCGCTGACCCCGCAGCAGCTGCAGCAGATTCTGACCAGCCGCGAAGTGCTGCCGGGCG
CGAAATGCGATTATACCATTCAGGTGCAGCTGCGCTTTTGCCTGTGCGAAACCAGCTGCCCGCAGGAA
GATTATTTTCCGCCGAACCTGTTTGTGAAAGTGAACGGCAAACTGTGCCCGCTGCCGGGCTATCTGCC
GCCGACCAAAAACGGCGCGGAACCGAAACGCCCGAGCCGCCCGATTAACATTACCCCGCTGGCGCGCC
TGAGCGCGACCGTGCCGAACACCATTGTGGTGAACTGGAGCAGCGAATTTGGCCGCAACTATAGCCTG
AGCGTGTATCTGGTGCGCCAGCTGACCGCGGGCACCCTGCTGCAGAAACTGCGCGCGAAAGGCATTCG
CAACCCGGATCATAGCCGCGCGCTGATTAAAGAAAAACTGACCGCGGATCCGGATAGCGAAGTGGCGA
CCACCAGCCTGCGCGTGAGCCTGATGTGCCCGCTGGGCAAAATGCGCCTGACCGTGCCGTGCCGCGCG
CTGACCTGCGCGCATCTGCAGAGCTTTGATGCGGCGCTGTATCTGCAGATGAACGAAAAAAAACCGAC
CTGGACCTGCCCGGTGTGCGATAAAAAAGCGCCGTATGAAAGCCTGATTATTGATGGCCTGTTTATGG
AAATTCTGAACAGCTGCAGCGATTGCGATGAAATTCAGTTTATGGAAGATGGCAGCTGGTGCCCGATG
AAACCGAAAAAAGAAGCGAGCGAAGTGTGCCCGCCGCCGGGCTATGGCCTGGATGGCCTGCAGTATAG
CGCGGTGCAGGAAGGCATTCAGCCGGAAAGCAAAAAACGCGTGGAAGTGATTGATCTGACCATTGAAA
GCAGCAGCGATGAAGAAGATCTGCCGCCGACCAAAAAACATTGCCCGGTGACCAGCGCGGCGATTCCG
GCGCTGCCGGGCAGCAAAGGCGCGCTGACCAGCGGCCATCAGCCGAGCAGCGTGCTGCGCAGCCCGGC
GATGGGCACCCTGGGCAGCGATTTTCTGAGCAGCCTGCCGCTGCATGAATATCCGCCGGCGTTTCCGC
TGGGCGCGGATATTCAGGGCCTGGATCTGTTTAGCTTTCTGCAGACCGAAAGCCAGCATTATGGCCCG
AGCGTGATTACCAGCCTGGATGAACAGGATACCCTGGGCCATTTTTTTCAGTATCGCGGCACCCCGAG
CCATTTTCTGGGCCCGCTGGCGCCGACCCTGGGCAGCAGCCATCGCAGCAGCACCCCGGCGCCGCCGC
CGGGCCGCGTGAGCAGCATTGTGGCGCCGGGCAGCAGCCTGCGCGAAGGCCATGGCGGCCCGCTGCCG
AGCGGCCCGAGCCTGACCGGCTGCCGCAGCGATGTGATTAGCCTGGAT

>PINIT domain coding sequence (PIAS3)

ATGAAACCGCTGCCGTTTTATGAAGTGTATGGCGAACTGATTCGCCCGACCACCCTGGCGAGCACCAG
CAGCCAGCGCTTTGAAGAAGCGCATTTTACCTTTGCGCTGACCCCGCAGCAGCTGCAGCAGATTCTGA
CCAGCCGCGAAGTGCTGCCGGGCGCGAAATGCGATTATACCATTCAGGTGCAGCTGCGCTTTTGCCTG
TGCGAAACCAGCTGCCCGCAGGAAGATTATTTTCCGCCGAACCTGTTTGTGAAAGTGAACGGCAAACT
GTGCCCGCTGCCGGGCTATCTGCCGCCGACCAAAAACGGCGCGGAACCGAAACGCCCGAGCCGCCCGA
TTAACATTACCCCGCTGGCGCGCCTGAGCGCGACCGTGCCGAACACCATTGTGGTGAACTGGAGCAGC
GAATTTGGCCGCAACTATAGCCTGAGCGTGTATCTGGTGCGCCAGCTGACCGCGGGCACCCTGCTGCA
GAAACTGCGCGCGAAAGGCATTCGCAACCCGGATCATAGCCGCGCGCTGATTAAAGAAAAACTGACCG
CGGATCCGGATAGCGAAGTG

>**Acidic domain coding sequence (PIAS3)**

ATGGAAGATGGCAGCTGGTGCCCGATGAAACCGAAAAAAGAAGCGAGCGAAGTGTGCCCGCCGCCGGG
CTATGGCCTGGATGGCCTGCAGTATAGCGCGGTGCAGGAAGGCATTCAGCCGGAAAGCAAAAAACGCG
TGGAAGTGATTGATCTGACCATTGAAAGCAGCAGCGATGAAGAAGATCTGCCGCCGACCAAAAAACAT
TGCCCGGTGACCAGCGCGGCGATTCCGGCGCTGCCGGGCAGCAAAGGCGCGCTGACCAGCGGCCATCA
GCCGAGCAGCGTGCTGCGCAGCCCGGCGATGGGCACCCTGGGCAGCGATTTTCTGAGCAGCCTGCCGC
TGCATGAATATCCG

G2: PROTEIN SEQUENCES

>PIAS3 mouse

MAELGELKHMVMSFRVSELQVLLGFAGRNKSGRKHELLAKALHLLKSSCAPSVQMKIKEL
YRRRFPRKTLGPSDLSLLSLPPGTSPVGSPGPLAPIPPTLLTPGTLLGPKREVDMHPPLP
QPVHPDVTMKPLPFYEVYGELIRPTTLASTSSQRFEEAHFTFALTPQQLQQILTSREVLP
GAKCDYTIQVQLRFCLCETSCPQEDYFPPNLFVKVNGKLCPLPGYLPPTKNGAEPKRPSR
PINITPLARLSATVPNTIVVNWSSEFGRNYSLSVYLVRQLTAGTLLQKLRAKGIRNPDHS
RALIKEKLTADPDSEVATTSLRVSLMCPLGKMRLTVPCRALTCAHLQSFDAALYLQMNEK
KPTWTCPVCDKKAPYESLIIDGLFMEILNSCSDCDEIQFMEDGSWCPMKPKKEASEVCPP
PGYGLDGLQYSAVQEGIQPESKKRVEVIDLTIESSSDEEDLPPTKKHCPVTSAAIPALPG
SKGALTSGHQPSSVLRSPAMGTLGSDFLSSLPLHEYPPAFPLGADIQGLDLFSFLQTESQ
HYGPSVITSLDEQDTLGHFFQYRGTPSHFLGPLAPTLGSSHRSSTPAPPPGRVSSIVAPG
SSLREGHGGPLPSGPSLTGCRSDVISLD

> **PINIT domain (PIAS3)**

MKPLPFYEVYGELIRPTTLASTSSQRFEEAHFTFALTPQQLQQILTSREVLPGAKCDYTIQVQLRFCL
CETSCPQEDYFPPNLFVKVNGKLCPLPGYLPPTKNGAEPKRPSRPINITPLARLSATVPNTIVVNWSS
EFGRNYSLSVYLVRQLTAGTLLQKLRAKGIRNPDHSRALIKEKLTADPDSEV

>**Acidic domain (PIAS3)**

MEDGSWCPMKPKKEASEVCPPPGYGLDGLQYSAVQEGIQPESKKRVEVIDLTIESSSDEEDLPPTKKH
CPVTSAAIPALPGSKGALTSGHQPSSVLRSPAMGTLGSDFLSSLPLHEYP

> **PINIT domain (Siz1)**

FAVPTIHFKESPFYKIQRLIPELVMNVEVTGGRGMCSAKFKLSKADYNLLSNPNSKHRLYLFSGMINP
LGSRGNEPIQFPFPNELRCNNVQIKDNIRGFKSKPGTAKPADLTPHLKPYTQQNNVELIYAFTTKEYK
LFGYIVEMI

APPENDIX H: SEQUENCES OF PRIMERS USED IN PCR, MUTAGENESIS, AND DNA SEQUENCING

H1: PCR PRIMERS USED TO AMPLIFY THE PIAS3 DOMAINS CODING SEQUENCE

VECTOR	DOMAIN	PRIMER NAME	SEQUENCE (5' TO 3')
pQE2	PINIT	PIAS3_132_F	CATATGAAGCCCCTGCCCTTC
pQE2	PINIT	PIAS3_132–R	AAGCTTATTACACTTCACTGTCGGGGTC
pGEX4T	PINIT	pGEX4T–PINIT–F	CATATGAAGCCCCT CCCTTCTA GAAGTCTATGGG
pGEX4T	PINIT	pGEX4T–PINIT–R	GTCGACTTAGTGATGGTGATGGTGATGCACTTCACTG
pQE60	PINIT	pQE60–PINIT–F	CCATGGAGCCCCTGCCCTTCTATG
pQE60	PINIT	pQE60–PINIT–R	AGATCTTCACTGTCGGGGTCAGCGG
pEGFP–PIAS3		EGF–PIAS3–F	GCTAGCATGGCGGAGCTG
pEGFP–PIAS3		EGFP–PIAS3–R	GACGTCATTTCCTTGGACAAGCTT
pEGFP–PINIT		PINIT–F	GCTAGCATGAAGCCACTGCC
pEGFP–PINIT		PINIT–R	CCCCGACAGTGAAGTGAAGCT T
p513–flag –acidic		PIAS3–C–F	CATATGGAAGATGGATCCTGGTGTC
p513–flag-acidic		PIAS3–C–R	AAGCTTTAAGCCCCCAGTGG

H2: PRIMERS USED IN MUTAGENESIS

MUTATION	PRIMER	PRIMER NAME	PRIMER SEQUENCE (5' to 3')
PIAS3 R99Q	Sense	R99Q-sense	GGGGAGCTCATCCAACCCACCACCCTT
	Antisense	R99Q-antisense	AAGGGTGGTGGGTTGGATGAGCTCCCC
PIAS3 R99N	Sense	R99N-sense	TCTATGGGGAGCTCATCAATCCCACCACCCTTGCGTC
	Antisense	R99N-antisense	GACGCAAGGGTGGTGGGATTGATGAGCTCCCCATAGA
PIAS3 L97A	Sense	L97A-sense	GAAGTCTATGGGGAGGCCATCCGACCCACCAC
	Antisense	L97A-antisense	GTGGTGGGTCGGATGGCCTCCCCATAGACTTC
PINIT-TAA-$(His)_6$	Sense	PINIT-TAA-F	CCCCGACAGTGAAGTGTAACATCACCATCACCATC
	Antisense	PINIT-TAA-R	GATGGTGATGGTGATGTTACACTTCACTGTCGGGG
(P513-flag)*Nde*I	Sense	p513-*Nde*I-sense	CCATTTCCTTGGACCATATGTAAGCTTCCTAGGTC
	Antisense	p513-*Nde*I-anti	GACCTAGGAAGCTTACATATGGTCCAAGGAAATGG

H3: PRIMERS USED IN DNA SEQUENSES

TARGET	PRIMER NAME	PRIMER SEQUENCE (5’ to 3’)
pQE2	pQE2–F	CCCGAAAAGTGCCACC
pQE2	pQE2–R	TTAGCTCCTGAAAATCTCG
pQE60	pQE60–F	GACCGCTGACCCCGACAGTAGATCTCATCA
PQE60	pQE60–R	TGATGAGATCTACTGTCGGGGTCAGCGGTC
p513	p513–PIAS3–F	GCATTGTGGCTCCTGGGAGC
p513–flag	p513–flag–F	CCT CTG CTA ACC ATG TTC ATG CC