SIN1 AND SIN1 ISOFORMS:

An investigation into the biological significance of a novel human protein family.

Nicole Cloonan, B.Sc. (Hons)

ESKITIS Institute for Cell and Molecular Therapies

and

the School of Biomolecular and Biomedical Science

Faculty of Science
Griffith University

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

February 2006

Abstract

Stress activated protein kinase (SAPK) interacting protein 1 (Sin1) is a member of a recently characterized gene family, conserved from yeast to humans. The gene copy number is strictly conserved (one Sin1 gene per genome), and the protein may be expressed ubiquitously in mammalian tissues. The Sin1 family has been implicated in several different signal transduction pathways. Originally identified as a partial cDNA and candidate Ras inhibitor, recent functional studies have revealed interactions with an interferon (IFN) receptor subunit (IFNAR2), and the SAPK JNK. Interactions have also been described between the yeast orthologues and the phosphatidylinositol kinase TOR2. Collectively, these data suggest that Sin1 has an important cellular role, and this study has investigated possible functions for this protein.

As human Sin1 proteins have no paralogues within the genome, secondary structure homology was used to identify major domains within the protein. Four major domains within human Sin1 were deduced: an N-terminal domain containing a functional nuclear localization signal, a functional nuclear export signal, and a coiled-coil region; the conserved region in the middle that is likely to be a ubiquitin-like β-grasp protein binding domain; a Ras binding domain; and a pleckstrin homology-like domain that targets Sin1 to the plasma membrane and lipid rafts *in vivo*. Full and partial length EGFP constructs were used to examine the localization of human Sin1, and several isoforms derived from alternative splicing. All isoforms localized to the nucleus and nucleolus. Beyond this, Sin1α and Sin1γ had cytoplasmic staining, while Sin1 and Sin1β were also found at the plasma membrane and lipid rafts. Both the N-terminal

domain and the conserved region in the middle were found to contribute to nuclear localization.

Comparative genomic analysis between human, mouse, rat, dog, and chicken *Sin1* genes revealed a number of conserved intronic regions, and the putative functions of these were predicted. Additionally, a putative promoter module within a CpG island and encompassing the transcription start site was predicted in all species. The human CpG island was found to have promoter activity in HEK293 cells. Using bioinformatics, genes that may be co-regulated with Sin1 were identified. These genes contained the Sin1 promoter module, and were found to co-express in large scale gene expression studies. Most of these genes were directly involved in the cellular response to pathogen infection, suggesting a conserved role for Sin1 in this pathway.

Key biochemical functions of the Sin1 proteins were also identified, including the ability of Sin1 proteins to form dimers, and the ability of over-expressed Sin1 to induce apoptosis (mediated through the conserved region in the middle). Additionally, endogenous Sin1 protein levels were found to change following serum deprivation and hypoosmotic stress. Together, these studies have provided significant insight into the cellular role of Sin1, suggesting a role in inducing apoptosis as part of the IFN response to viral infection. The biological significance of the Sin1 proteins is discussed in the context of their predicted functions and the evolution of the protein family.

Statement of Originality

This work has never been previously submitted for a degree or diploma in any University. To the best of my knowledge and belief, the report contains no material previously published or written except where due reference is made in the report itself.

Nicole Cloonan

Table of Contents

List of Figures

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

List of Tables

CHAPTER THREE

List of Abbreviations

aa	amino acid
Akt	Murine thymoma viral (v-akt) oncogene homolog-1 (PKB)
ARE	A/T-rich regulatory element
ATF(1-2)	activating transcription factor (1-2)
AVO(1-3)	adheres voraciously to TOR (1-3)
bp	base pair
cAMP	cyclic AMP
CDC25	cell division cycle gene 25
CDK1	cyclin dependent kinase 1
cDNA	complementary DNA
CEG1	capping enzyme guanylyltransferase 1
CPBP	Core promoter binding protein
CpG	cytosine - phosphodiester bond - guanosine
cREL	v-rel reticuloendotheliosis viral oncogene homolog
CRIM	conserved region in the middle
CPBP	Core promoter binding protein
DABCO	1,4-diazabicyclo-[2,2,2]-octane
DMSO	dimethyl sulfoxide
dNTPs	dideoxynucleotides (dATP, dCTP, dGTP, dTTP)
DNA	deoxyribonucleic acid
DOE	department of energy

DTT	dithiothreitol
EBI	European Bioinformatics Institute
EBV	Epstein Barr virus
EDTA	ethylenediamine tetraacetic acid
EGF	epidermal growth factor
EGFP	enhanced green fluorescent protein
EGRF	Early Growth Response Family
ERK	Extracellular signal-regulated kinase
EST	expressed sequence tag
FCS	foetal calf serum
GAPVD1	GTPase activating protein and VPS9 domains 1
gDNA	genomic DNA
GDP	guanosine diphosphate
GNB1	guanine nucleotide binding protein (G protein), beta polypeptide 1
GTP	guanosine triphosphate
GST	glutathione S-transferase
h	hour(s)
HEAT	huntingtin-elongation-A subunit-TOR motif
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
HES1	Drosophila hairy and enhancer of split homologue 1
HUGO	The Human Genome Organization
IFA	immunofluorescence assay
IFN	interferon
IFNAR2	interferon (alpha, beta and omega) receptor 2

IRES	internal ribosome entry site
JNK	c-Jun N-terminal Kinase
kb	kilo base (pairs)
kcal	kilo-calorie
kDa	kilodaltons
LB	Luria-Bertani medium
LST8	Lethal with Sec Thirteen 8
M	molar
MAPK	Mitogen Activated Protein Kinase
MAPKAP1	MAPK associated protein 1
MAZ	Myc associated zinc finger protein
mb	mega base (pairs)
MEKK2	mitogen-activated protein kinase kinase kinase 2
mg	milligram
min	minute(s)
mL	millilitre
mM	millimolar
µg	microgram
µl	microlitre
µM	micromolar
mol	mole
MSRA	methionine sulfoxide reductase A
mRNA	messenger RNA
mTOR	mammalian target of rapamycin

NCBI	national center for biotechnology information
NES	nuclear export signal
NLS	nuclear localization signal
NR	non redundant
nt	nucleotide
NTD	N-terminal domain
OLF	Olfactory neuron-specific factor
ORF	open reading frame
PBS	phosphate buffered saline
PBX3	pre-B-cell leukemia transcription factor 3
PCR	polymerase chain reaction
PCSK6	Proprotein convertase subtilisin/kexin type 6 precursor
PH	pleckstrin homology domain
PHL	PH-like domain
PI3K	phosphatidylinositol 3 kinase
PtdInsP	phosphatidylinositol phosphate
pJNK	phospho-c-Jun N-terminal Kinase
PRDM15	PR domain containing 15
QIMR	Queensland Institute of Medical Research
RBD	Ras binding domain
RIP3	Ras interacting protein 3
RNA	ribonucleic acid
RNAse	ribonuclease
RPMI	Roswell Park Memorial Institute medium

rRNA	ribosomal RNA
RRP9	ribosomal RNA processing 9
RT-PCR	reverse-transcriptase PCR
Sam68	Src-associated in mitosis 68 kDa protein
SAPK	Stress activated protein kinase
SCD25	Suppressor of cdc25
SCD(I-V)	Sin1 conserved domain (I-V)
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
Sin1	SAPK interacting protein 1
S/MAR	scaffold/matrix attachment region
snoRNA	small nucleolar RNA
SP1	Stimulating Protein 1
TOR	target of rapamycin
Tris	Tris(hydroxymethyl)-enediamine
TRITC	tetramethylrhodamine isothiocyanate
TSS	transcription start site
U	units
UV	ultraviolet
UTR	untranslated region
WWP1	WW domain containing E3 ubiquitin protein ligase 1
YBX1	Y-box transcription factor 1

List of Publications

Publications arising from work described in this thesis:

N Cloonan, K Tay, Z Barnard, T Sculley, and G Bushell. "Comparative genomic analysis links Sin1 to the cellular response to viral invasion". Manuscript in preparation.

N Cloonan, M Crampton, T Sculley, and G Bushell. *"Multiple domains control the sub-cellular localization of the SAPK-interacting protein 1 (SIN1) isoforms"*. Manuscript submitted, 2006.

WA Schroder, M Buck, **N Cloonan**, JF Hancock, A Suhrbier, G Bushell, and T Sculley. *"Human Sin1 contains Ras-binding and pleckstrin homology domains and suppresses MAPK and Akt activation by Ras"*. Manuscript submitted, 2005.

W Schroder, **N Cloonan**, G Bushell, and T Sculley. *"Alternative polyadenylation and splicing of mRNAs transcribed from the human Sin1 gene."* Gene 2004; 339:17-23

Conference presentations (oral) arising from work described in this thesis:

N. Cloonan, M Crampton, T Sculley, and G Bushell. *"Multiple domains control the cellular localization and protein-protein interactions of human stress activated protein kinase interacting protein 1 (Sin1)."* ComBio 2005. 25th-29th September, 2005. Adelaide, Australia.

Conference presentations (poster) arising from work described in this thesis:

N Cloonan, M Crampton, and G Bushell. *"Is strict regulation of mRNA levels critical for translational regulation?"* 13th Annual International conference on Intelligent Systems for Molecular Biology (ISMB 2005). 25th-29th June, 2005. Detroit, USA.

W Schroder, **N Cloonan**, T Pirayesh, T Sculley, and G Bushell. *"The human Stress activated protein kinase-interacting 1 (Sin1) gene encodes JNK-binding proteins."* 30th FEBS Congress and 9th IUBMB conference. 2nd-7th July, 2005 Budapest, Hungary

Acknowledgements

I would first like to say thank you to my supervisors, Professor Gillian Bushell and Dr Tom Sculley, who really should have known better than to accept me as a student. What can I say? There's just no telling some people. Regardless, their support and input into this project was highly valued, and I deeply appreciate their efforts. I would also like to thank my friends and colleagues in the lab, past and present: Zeke, Matt, Shelley, Vicky, Nicole, Tim, and Kelvin. It has certainly been entertaining working (and bludging) with you all.

I acknowledge the academic and financial support of Griffith University, both directly through International Travel Grants, and indirectly through the School of BBS and ESKITIS. I'd also like to thank the International Society for Computational Biology, and the Australian and New Zealand Society for Cell and Developmental Biology for travel grants, and the Department of Education, Science, and Training for providing me with an Australian Postgraduate Award.

Without the love and support of my family, I would have been unable to complete (or even start) my PhD studies. I thank my parents, Jan and Ric, my siblings Greg, Michelle, and Kristin, and my grandparents Scottie and Terry. Thanks also to my good friend Mathew Taylor, for all the little things that add up to an awful lot. I am enormously grateful for the support of my husband, Darrin Taylor, throughout this project. I thank him for his emotional, financial, intellectual, and (on at least one drunken occasion) physical support. I am also very grateful for the patience of my children, who make everything we do so much more fun.

CHAPTER ONE

Introduction

1.1 General Introduction

In order for a complex multicellular organism to survive, its cells need to be able to grow, reproduce, change, and die in an appropriate manner. These crucial processes are highly regulated, and rely on proper interpretation of both physical and chemical cues from their environment to elicit a suitable response. Defects in transducing these signals cause a broad spectrum of diseases, including cancers (Rodenhuis, 1992), autoimmune disorders (Ohsako and Elkon, 1999), and various neuropathologies (Loddick *et al.*, 1997; Sibson *et al.*, 2002). It is unsurprising then that mechanisms of signal transduction have been the subject of intense study. Of particular interest have been members of the Ras family of small guanosine triphosphatases (small GTPases), as they act on a diverse range of signalling pathways, including cell differentiation, proliferation, and apoptosis (Spaargaren *et al.*, 1995). Their significance is highlighted by the finding that approximately 30% of all human tumours contain mutated Ras genes (Bos, 1989; Rodenhuis, 1992).

In 1991 a partial human cDNA (termed JC310) was shown to rescue the phenotype caused by constitutively activated $RAS2^{G19V}$ in yeast, suggesting that JC310 might be a novel component of the mammalian RAS signalling pathways (Colicelli *et al.*, 1991). Recently, the gene corresponding to JC310 was characterized, and full length mRNA and several splice variants were cloned in our laboratory (Schroder *et al.*, 2004). This gene was found to be a member of a family widely conserved throughout evolution, termed SAPK-interacting protein 1 (Sin1), after the *S. pombe* orthologue of

the same name (Wilkinson *et al.*, 1999). While the Sin1 proteins are thought to be involved in signal transduction, and several interesting molecular interactions have been described, the biological role of Sin1 remains unclear. This project focuses on investigating the biological significance of human Sin1 proteins.

1.2 The *Sin1* gene (MAPKAP1)

The *Sin1* gene spans approximately 280 kb on chromosome 9 at position q34.1 (Schroder *et al.*, 2004). The HUGO approved symbol for this gene is MAPKAP1, not to be confused with the X-linked Coffin-Lowry Syndrome protein of the same name. The acronym Sin1 is widely accepted (Wilkinson *et al.*, 1999; Loewith *et al.*, 2002; Schroder *et al.*, 2004; Wang and Roberts, 2004; Schroder *et al.*, 2005; Wang and Roberts, 2005), and is used throughout this thesis. The longest putative *Sin1* mRNA (derived from EST sequences; 3393 nt) is comprised of a non-coding exon at the 5'-end followed by 11 coding exons. An alternative exon (7a) is found between exons 6 and 7, is not transcribed in the full length sequence, and contains its own stop codon and polyadenylation signal (Figure 1.2.1). The 3' UTR of mRNA contains a potential A/T Rich Element (ARE), and two alternative polyadenylation sites. AREs decrease the half-life of mRNAs and are usually found in so called “rapid response” genes (Chen and Shyu, 1994). As *Sin1* transcripts can utilize either polyadenylation site *in vivo* (Schroder *et al.*, 2004), *Sin1* translation may be subject to tight regulation.

Three *Sin1* splice variants have been confirmed by RT-PCR in human cell lines

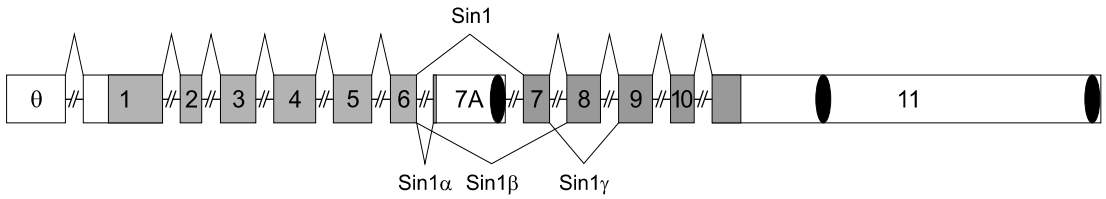


Figure 1.2.1. Schematic of *Sin1* gene structure. The constitutive splicing is indicated above the cartoon, and the alternate splicing events are indicated below. The coding regions within exons are grey, and polyadenylation sites are indicated by black ovals. Figure adapted from Schroder *et al.* (2004).

(Figure 1.2.1). Full length *Sin1* encodes a 522 aa ORF. *Sin1*α is formed from the splicing of exon 6 to the alternative exon 7A, and results in a truncated ORF of 323 aa. *Sin1*β mRNA lacks exon 7 and *Sin1*γ mRNA lacks exon 8, but both are otherwise uninterrupted and code for 486 aa and 475 aa ORFs, respectively (Schroder *et al.*, 2004). Full length *Sin1*, *Sin1*α, and *Sin1*β have been independently identified in a different study in human cells (Cheng *et al.*, 2005), and both *Sin1* and *Sin1*β have also been detected in sheep (Wang and Roberts, 2004).

Sin1 mRNA is present in a wide variety of tissues, and analysis of the EST database suggests that the gene may be ubiquitously expressed. The highest levels of *Sin1* mRNA are expressed in the heart and skeletal muscle (Loewith *et al.*, 2002; Schroder *et al.*, 2004; Cheng *et al.*, 2005). The distribution of *Sin1* splice variants has not been experimentally verified, although EST analysis suggested *Sin1*α is highly expressed in kidneys, whereas *Sin1*β is more evenly distributed through a range of tissues. *Sin1*γ is not present in the EST database, indicating that this may be a minor splice variant (Schroder *et al.*, 2004).

1.3 The conserved Sin1 protein family

Sin1 is the human member of a family of proteins widely distributed in the metazoan, fungal, and amoeboid kingdoms (Schroder *et al.*, 2004). Other family members include chicken Sin1 (Christiansen *et al.*, 2001), *S. pombe* Sin1 (Wilkinson *et al.*, 1999), *S. cerevisiae* AVO1 (Loewith *et al.*, 2002), and *Dictyostelium* RIP3 (Lee *et al.*, 1999). Sin1 has not been identified in prokaryotes or plants, and is probably not present in these organisms. It appears to be a unique gene, with only one copy per genome, and no obvious homologues (Schroder *et al.*, 2004; Wang and Roberts, 2005).

The signature sequence of the Sin1 family is the so-called conserved region in the middle (CRIM). As the only region present in every family member (Schroder *et al.*, 2004; Wang and Roberts, 2005), the CRIM is likely to be central to the function(s) of Sin1 proteins. No significant sequence similarity was detected between the CRIM and known protein domains, suggesting that the CRIM is either a novel domain, or a highly divergent form of a known domain. Four other regions of Sin1 are also widely conserved across species: SCDI (Sin1 conserved domain I), SCDII, SCDIV, and SCDV (Wang and Roberts, 2005). The SCDIII identified in Wang and Roberts' study is equivalent to the previously described CRIM (Figure 1.3.1). Again, no similarity to known protein domains was observed for these regions. A Ras Binding Domain (RBD) has been identified in the Sin1 orthologues RIP3 and AVO1 (Loewith *et al.*, 2002; Lee *et al.*, 2005), and Colicelli *et al.* (1991) reported that JC310 aligned non-significantly to the RBD of yeast CDC25 and SCD25. Collectively, these findings suggest that

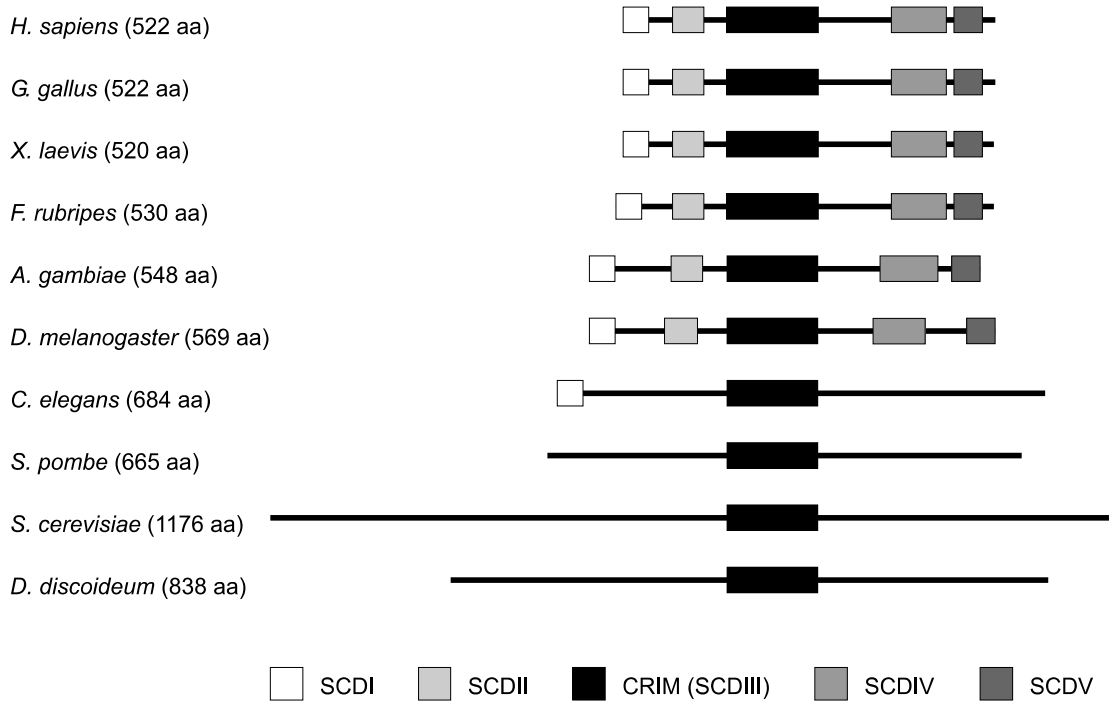


Figure 1.3.1. Five highly conserved regions of Sin1 proteins. Regions of sequence similarity identified by Wang and Roberts' (SCD = "Sin1 Conserved Domain") I to V are shown. SCDIII is identical to the previously described CRIM. Regions of sequence similarity do not necessarily correlate with full protein domains.

conserved domains are likely to be present, although their primary sequences may be highly divergent. Therefore, approaches that look beyond primary sequence homology are required to identify the protein folds within Sin1.

1.4 Sin1 molecular interactions

Often, the biological significance of a novel protein can be inferred by comparison with homologous proteins with known functions. Although no such proteins

exist for the Sin1 family, several interesting molecular interactions have been described for Sin1 and Sin1 orthologues. The wide evolutionary conservation of both the gene and the gene copy number suggests that the observed functions of the Sin1 orthologues are likely to be relevant to the function(s) of human Sin1 (Schroder *et al.*, 2004; Wang and Roberts, 2005). These interactions are outlined in the following sections.

1.4.1 Sin1 proteins interact with RAS

Human Sin1 was originally identified as a partial cDNA that inhibited signalling by constitutively activated $RAS2^{G19V}$ in yeast (Colicelli *et al.*, 1991). *Dictyostelium* RIP3 is a Ras interacting protein that is required for both chemotaxis, and the synthesis and relay of cyclic AMP (cAMP). RIP3 preferentially interacts with the activated form of RasG, and cells lacking RIP3 are unable to polarize appropriately in a cAMP gradient (Lee *et al.*, 1999). Human Sin1 also preferentially interacts with activated Ras. As there is significant co-localization of Sin1 and either H-Ras or K-Ras at the plasma membrane, and both Ras isoforms could be immunoprecipitated when over-expressed with FLAG-Sin1 (Schroder *et al.*, manuscript submitted), this interaction is likely to occur *in vivo*.

1.4.2 Sin1 proteins interact with SAPKs

S. pombe Sin1 is a Sty1 (Spc1/Spk1)-interacting protein. Sty1 is a member of the stress activated protein kinase (SAPK) family, and Sin1 derives its name from this

interaction (SAPK interacting protein 1). Yeast cells depleted of Sin1 are phenotypically identical to cells depleted of Sty1 (Wilkinson *et al.*, 1999): they are sensitive to certain environmental stresses, experience cell cycle delays, and lack the ability to initiate sexual conjugation. Wilkinson *et al*. (1999) showed that yeast Sin1 is constitutively phosphorylated, and hyper-phosphorylated upon stress, but that it is not a target of Sty1 *in vitro*. As the nuclear accumulation of stress-activated Sty1 was incomplete in cells depleted of Sin1, Sin1 may be important for translocation of Sty1 into the nucleus. Sin1 was also found to be important for the for the function of Atf1, an orthologue of human ATF2.

Sty1 is sensitive to UV radiation, osmotic stress, and other environmental stimuli, and is closely related to the mammalian SAPKs p38 and JNK (Millar, 1999). Accordingly, exogenous human Sin1 was found to co-precipitate and co-localize with endogenous JNK, but not p38 or ERK1/2 (Schroder *et al.*, 2005). Interestingly, exogenous Sin1α was also found to co-precipitate with Sin1, and both isoforms were able to suppress JNK phosphorylation in response to UV light. As the Sin1α isoform lacks the last 202 aa of full length Sin1, this data suggests that the C-terminal region is dispensable for binding JNK *in vivo*. These results contradict the previous finding (Wilkinson *et al.*, 1999) that the C-terminal region of *S. pombe* Sin1 is both sufficient and necessary for both Sty1 interaction, and Sty1 function. This region is also functionally conserved, as the C-terminal region of chicken Sin1 (94% identical to human Sin1) restored function to a *S. pombe* yeast strain where Sin1 was C-terminally

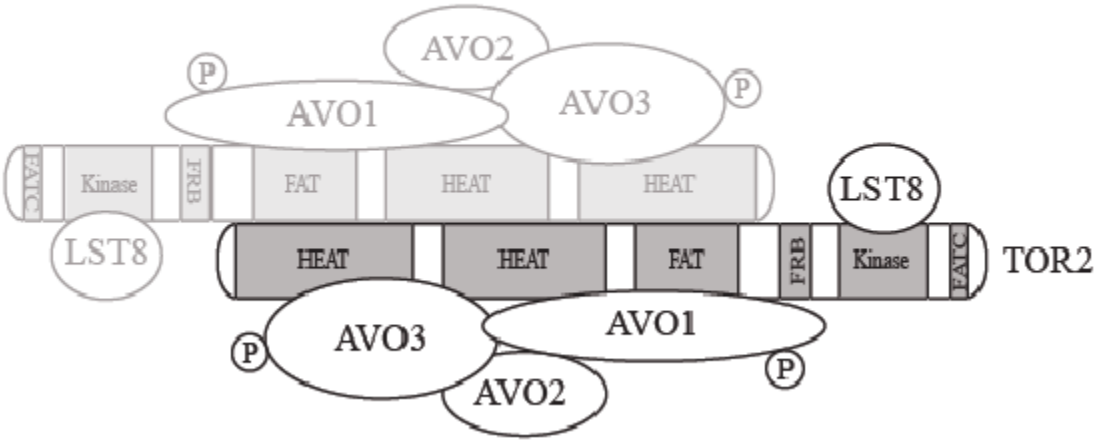


Figure 1.4.1. Model of TORC2 architecture. Figure adapted from Wullschleger *et al.* (2005). AVO1 and AVO3 bind cooperatively to the HEAT repeat region of TOR2. TORC2 autophosphorylates sites in AVO1 and AVO3. TORC2 is oligomeric, and likely to be a homodimer.

deleted (Wilkinson *et al.*, 1999). It is possible that there are multiple JNK docking sites within full length Sin1, but further experimental evidence is needed to resolve this issue.

1.4.3 Sin1 proteins interact with TOR

The *S. cerevisiae* Sin1 orthologue, AVO1, was found to bind specifically and avidly to the phosphatidylinositol kinase TOR2 (target of rapamycin 2). AVO1, TOR2, and three other proteins (AVO2, AVO3, and LST8) form the complex TORC2, which controls actin cytoskeleton organization through rapamycin-insensitive signalling (Loewith *et al.*, 2002; Wedaman *et al.*, 2003). This complex may be oligomeric (Wullschleger *et al.*, 2005), and a model of the *S. cerevisiae* TORC2 architecture is

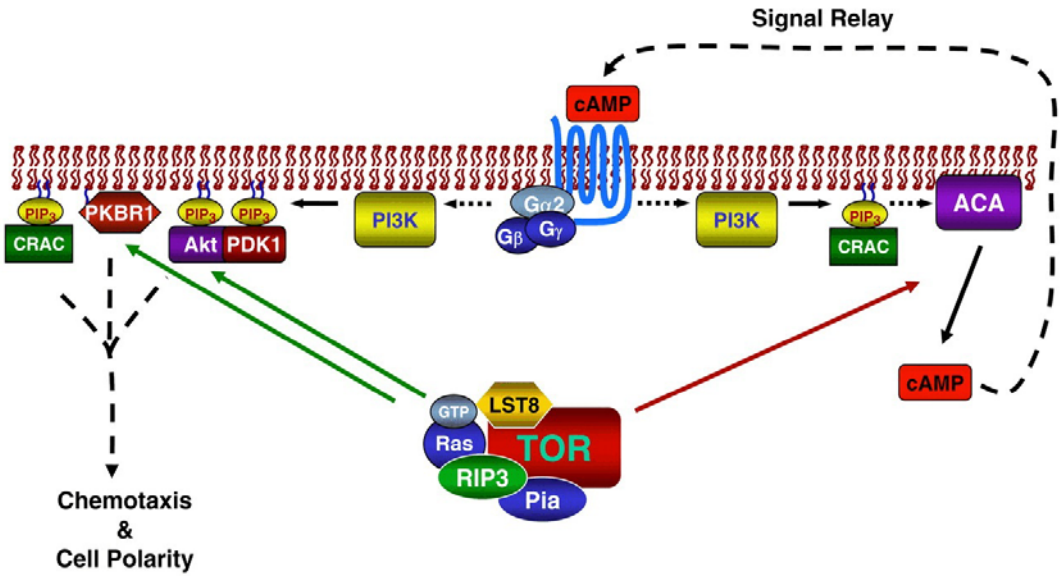


Figure 1.4.2. Model of TORC2 function during chemotaxis. Figure from Lee *et al.* (2005). Binding of TORC2 to activated Ras is necessary for downstream signalling.

presented in Figure 1.4.1. A RIP3-containing TORC2 complex was also identified in *Dictyostelium*, where it regulates both chemotaxis and signal relay (Lee *et al.*, 2005). This complex was also found to be rapamycin-insensitive, and regulated in part by Ras (Figure 1.4.2).

Even though the mammalian TOR (mTOR) is structurally and functionally orthologous to the yeast TORs, the interaction between mTOR and Sin1 could not be demonstrated in mammalian cells (Loewith *et al.*, 2002). While it is possible that Sin1 and mTOR do not interact, these authors used the Sin1β isoform in their assays, which has a significant sequence deletion as compared to the full length Sin1 (Schroder *et al.*, 2004). Considering that the TORC2 complex may be evolutionarily conserved (Jacinto

et al., 2004; Sarbassov *et al.*, 2004; Lee *et al.*, 2005; Wullschleger *et al.*, 2005), it still remains possible that Sin1 also binds mTOR in mammalian systems.

1.4.4 Sin1 proteins interact with IFNAR2

Ovine Sin1 was identified in a yeast two-hybrid screen as an interacting partner of IFNAR2, a subunit of a type I interferon receptor (Wang and Roberts, 2004). This interaction was confirmed by immunoprecipitation and co-localization. Like human Sin1, ovine Sin1 was found to be alternatively spliced, and ubiquitously expressed in the sheep tissues studied. Although the function of this interaction is unclear, the authors hypothesized that Sin1 could link type I IFN signalling to the SAPK pathways. Such a link is plausible, as type I IFNs have been shown to regulate SAPK signalling in some mammalian cells (Goh *et al.*, 1999; Uddin *et al.*, 1999; Doualla-Bell and Koromilas, 2001).

1.4.5 Sin1 proteins interact with MEKK2

The most recent addition to the list of Sin1-interacting proteins is MEKK2 (Cheng *et al.*, 2005). In this study, both transiently expressed and endogenous Sin1β were shown to interact with MEKK2, and negatively regulate MEKK2 signalling. Over-expression of Sin1β inhibited the MEKK2 activation of ERK5, and also inhibited MEKK2- (but not MEKK1)-dependent JNK1 activation. Further experiments demonstrated that this was not due to a direct blockade of JNK activity, but was caused

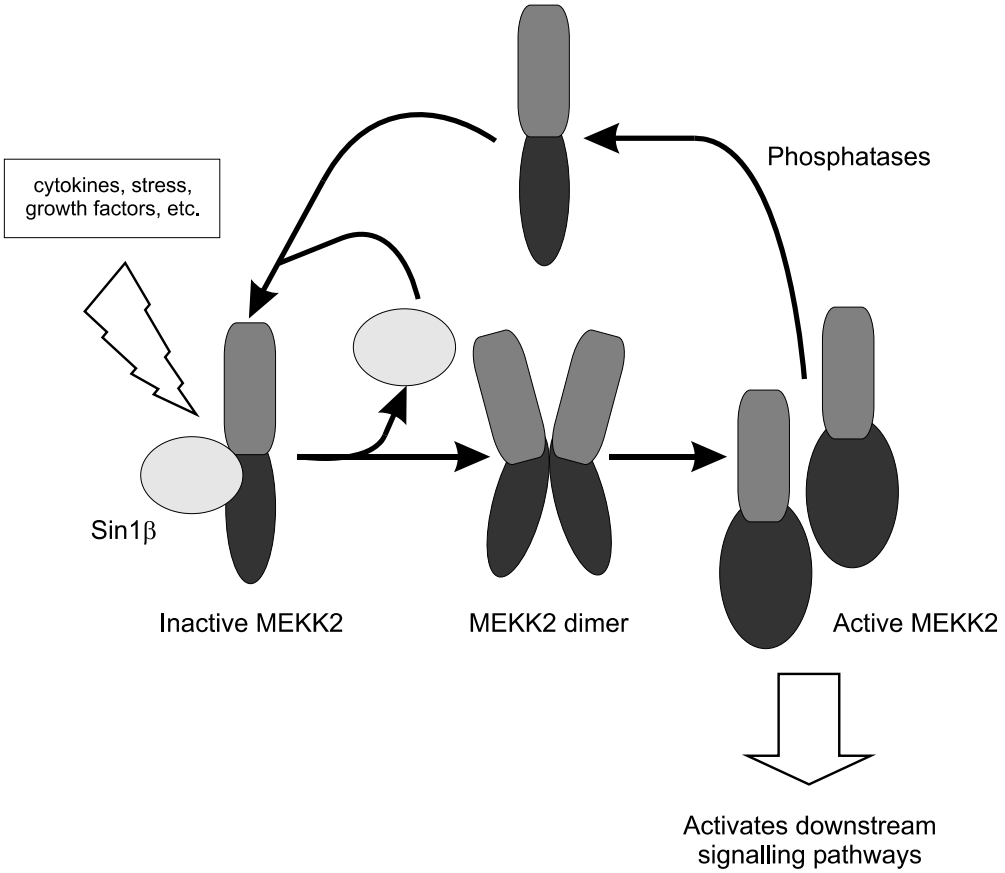


Figure 1.4.3. Model of MEKK2 regulation and activation. Figure adapted from Cheng *et al.* (2005).

by direct binding of Sin1β to inactive MEKK2, preventing MEKK2 dimerization and auto-phosphorylation. The Sin1β-MEKK2 complex was found to disassociate upon stimulation with epidermal growth factor (EGF). Additionally, knockdown of Sin1 expression activated JNK signalling, confirming that Sin1β is a negative regulator of MEKK2 *in vivo*. A model for Sin1-MEKK2 interaction is presented in Figure 1.4.3.

The interacting site for MEKK2 lies within the first 184 aa of Sin1, an area

conserved in all four confirmed Sin1 isoforms (Schroder *et al.*, 2004), suggesting that any of the isoforms might be able to bind MEKK2 and induce the same effect. Although inhibition of JNK signalling was used as evidence for the direct binding of Sin1 and Sin1α to JNK (Schroder *et al.*, 2005), it is possible that this was instead due to the interaction of these forms with MEKK2. Again, this question requires further experimental evidence for its resolution.

1.4.6 Sin1 proteins interact with Sam68

Sam68 acts as a scaffold in signalling pathways (Najib *et al.*, 2005), and was identified as a potential binding partner of Sin1 in a yeast two hybrid assay. Endogenous Sin1 proteins were found to co-localize with endogenous Sam68 (Schroder *et al*, unpublished data). This interaction has not been confirmed or characterized further, but may represent an interesting new aspect of Sin1 function.

1.4.7 Sin1 as an evolutionarily conserved scaffold molecule?

The single most engaging question about Sin1 concerns its function. There are no obvious sequences suggesting enzymatic activity; there are many different binding partners; and over-expression of this protein results in the inhibition of signalling pathways. Together, these results suggest that Sin1 may act as a scaffold or adaptor protein in a signalling pathway (Wang and Roberts, 2004; Schroder *et al.*, 2005). However, this would only explain the abilities of Sin1, not its purpose.

While interesting functions have been described for Sin1 orthologues, it is as important to consider the differences between these proteins as the similarities. For example, *S. pombe* Sin1 contains an additional 71 aa in the proposed binding regions to Sty1 that are not present in the human sequence (Wilkinson *et al.*, 1999), and both *S. cerevisiae* AVO1 and *D. discoideum* RIP3 contain significantly longer N-terminal regions than mammalian Sin1 proteins (Figure 1.3.1). It is possible that the biological roles of Sin1 proteins have evolved to fit the needs of very different organisms, and molecular functions of the protein may have been gained and/or lost over time.

1.5 Project Aims

Together, the interactions with important signalling molecules, the near ubiquitous expression in mammalian tissues, and the wide evolutionary conservation of both the gene and the gene copy number suggest a conserved and important role for Sin1 in a signal transduction-related process. However the specifics of this role remain elusive, and the physiological relevance of this protein remains unknown. Thus, while the overall aim of this thesis is to determine the biological significance of human Sin1 proteins, the specific aims focus on examining the localization of human Sin1 proteins (Chapter Two), the genomic regulation of human Sin1 gene expression (Chapter Three), and the biochemical functions of human Sin1 (Chapter Four). Together, these results have provided significant insights into the biological role of Sin1, and these are discussed in the final chapter.

CHAPTER TWO

An investigation of the localization of Sin1

2.1 INTRODUCTION

Like many human genes, mRNA transcripts derived from the human *Sin1* gene undergo alternative splicing. One variant, Sin1α, has been reported to bind JNK and affect its activity *in vivo* (Schroder *et al.*, 2005). Two other confirmed splice variants, Sin1β and Sin1γ (Schroder *et al.*, 2004), have not been studied in detail. As little is known about the protein domains comprising Sin1, functional predictions regarding these isoforms are not yet possible. Identifying domains within Sin1 is therefore an important step in understanding the localization, and thus the function of this protein family.

Although several regions of sequence conservation within Sin1 have been reported (Schroder *et al.*, 2004; Wang and Roberts, 2005), no significant homology to any other described protein or domain has been previously detected within the vertebrate members of the Sin1 family. However, a Ras Binding Domain (RBD) has been identified in the Sin1 orthologues RIP3 and AVO1 (Loewith *et al.*, 2002; Lee *et al.*, 2005), and Colicelli *et al* (1991) reported that a partial human Sin1 cDNA, JC310, aligned non-significantly to the RBD of yeast CDC25 and SCD25. These data suggest that conserved domains are likely to be present, although highly divergent in sequence. In this chapter the domains of Sin1 were defined by identifying the inter-domain regions and secondary structure homology. Four major domains within the full length Sin1 protein were identified, and their contribution to the cellular localization of full length Sin1 was determined. Together, these results provide an insight into the biological role of the Sin1 protein family.

2.2 MATERIALS AND METHODS

2.2.1 Databases, sequences, and sequence analysis

Blast (with low complexity filters), TblastN, and the NR and EST databases were accessed through NCBI (http://www.ncbi.nih.gov). Alignments, non-conservative non-synonymous amino acid substitutions, and amino acid composition were assessed using GeneDoc (http://www.psc.edu/biomed/genedoc/). Protein disorder was predicted using DomCut (http://www.bork.embl-heidelberg.de/~suyama/domcut) and DisEMBL 1.4 (http://dis.embl.de). Protein domain sequences were submitted to 3D-Jury (http://bioinfo.pl/Meta/) for fold prediction. Coiled-coil regions were predicted using COILS (http://www.ch.embnet.org/software/COILS_form.html). The NetNES server (http://www.cbs.dtu.dk/services/NetNES/) was used to predict nuclear export signals. The vertebrate Sin1 protein sequences used in this study are as follows: human, orangutan, mouse, rat, sheep, chicken, and frog sequences were obtained from NCBI (Genbank accession numbers: AAS90839, CAH90052, AAH96618, AAH79073, AAS55637, Q9W6S3, and NP_796319, respectively). Pig and cow sequences were assembled from the EST database. The chimpanzee and dog genomes were searched at NCBI, and the *Fugu rubripes* genome was located at the DOE Joint Genome Institute (http://genome.jgi-psf.org).

2.2.2 Plasmids and plasmid construction

Sin1, *Sin1*α, *Sin1*β, and *Sin1*γ splice variants were obtained from previously

described pGEM-T constructs (Schroder *et al.*, 2004). To remove the stop codon and add restriction enzyme sites, *Sin1*, *Sin1*β, and *Sin1*γ were amplified using Sin1-F9 (Schroder et al., 2005) and Sin1-R11 (5'-gcggtacccgctgctgcccggatttctt-3'), whereas *Sin1*α was amplified using Sin1-F9 and Sin1-R31 (5'-gcggtacccggtcacaagcacctgaaac-3'). Restriction sites are underlined. PCR products were purified, and cloned into the KpnI and EcoRI sites of pEGFP-N1 (Clontech) to generate C-terminally tagged Sin1-EGFP fusions. N-terminally tagged Sin1 was generated by subcloning the insert from pGEM-T-Sin1 into the EcoRI site of pEGFP-C3 (Clontech). EGFP-Sin1α was created by digestion of EGFP-Sin1 by BamHI, and self-ligation. N-terminally tagged partial Sin1 constructs were created by using PCR to amplify segments from the pGEMT-Sin1 plasmid, cloning into pGEMT, and then sub-cloning into either pEGFP-C3 or pGEX-2T (Amersham) to create either EGFP or GST tagged constructs. The primer pairs used were as follows: EGFP-NTD (Sin1-F37:5'-gaattcggatggccttcttggaca-3' and Sin1-R44:5'-gaattcgaatttctttctttccactg-3'); EGFP-SCDI (Sin1-F32:5'-ggaagatctgatggccttcttggacaatcc-3' and Sin1-R33:5'-tcaagatcttctaggtcaacatcatgatc-3'); EGFP-SCDII (Sin1-F38:5'-gaattcatgcccagtcagtcgata-3' and Sin1-R44); EGFP-CRIM (Sin1-F39:5'-gaattcctccaatttctgggaagc-3' and Sin1-R43:5'-gaattccttttcaaccagggccaa-3'); EGFP-CRIM-RBD (Sin1-F39 and Sin1-R42:5'-gaattcctccaaagtgctgtccag-3'); and EGFP-PHL (Sin1-F12:5'-gaattcttaaggacatagccacagtacagg-3' and Sin1-R8:5'-gaattcggccagtgtcactgctg-3'). The GST-Sin1 fusion expression plasmid was described previously (Schroder *et al.*, 2005). All constructs were verified by sequencing.

2.2.3 Tissue culture and transfections

COS7, HEK293, DG75, JAM, SHSY5Y, and HeLa cells were maintained in RPMI1640 medium (GibcoBRL) containing Antibiotic-Antimycotic (Invitrogen), and 10% (v/v) foetal calf serum (GibcoBRL), in a 5% CO_2 atmosphere at 37^0C. Cells were transfected using the Nucleofector system (Amaxa), using 1 μg of DNA and 1-2 X 10^6 cells per reaction.

2.2.4 Immunofluorescence assays, lipid raft labelling, and microscopy

Cells were harvested at 24 h. DG75 cells were washed three times in PBS, dispensed onto slides, and air-dried. Adherent cells grown on cover slips were washed three times with PBS. All cells were fixed in freshly prepared 4% paraformaldehyde for 30 min at 4^0C. Cells were washed again three times in PBS, permeabilized in 0.25% Triton X100 for 5 min, and the PBS washes repeated. Cells were stained with 5 μg/mL propidium iodide for 30 min at room temperature, and washed three times more. Finally, cells were mounted using 50% (v/v) glycerol and 2% (w/v) 1,4-diazabicyclo-[2,2,2]-octane (DABCO; Sigma) in PBS. Lipid raft labelling was performed on live DG75 cells using a Vybrant Lipid Raft Labelling Kit (Molecular Probes) according to the manufacturer's instructions. Slides were analysed on a Leica SP2 confocal microscope.

2.2.5 Lipid binding assays

PIP-strips (Echelon Biosciences) were incubated with affinity-purified GST-PHL (0.5 μg/mL) in the presence or absence of 1 mM neomycin (neo). The membranes were immunoblotted according to manufacturer's instructions, using an anti-GST antibody (Sigma) at 1:2000. Cos7 cells were transiently transfected with EGFP-PHL in the presence or absence of 10 mM neomycin, and harvested after 24 h. Cells were fixed, stained with propidium idodide, and analysed on a Leica SP2 confocal microscope.

2.2.6 SDS-PAGE and Western blots

Cells were resuspended and lysed in ice cold RIPA buffer at 1 X 10^7 cells/mL. Cell lysate (10 μL) was added to 5 μL 3 X Sample Buffer (180 mM Tris-Cl, pH 6.8; 30% (v/v) glycerol; 6% (w/v) SDS; 15% β-mercaptoethanol; 0.015% (w/v) bromophenol blue). Cells were sonicated, boiled for 5 min, and then separated by SDS-PAGE (Laemmli, 1970). Samples were run through a 4% polyacrylamide stacking layer, then a 10% polyacrylamide separating layer using a Mini-PROTEAN II system (BioRad) at 100 V for 2 h. Proteins were transferred to nitrocellulose membranes (BioRad) at 100 V for 1 h at 4°C (BioRad), and stained with 0.1% Ponceau S (w/v) in 5% acetic acid to confirm even protein loading. Immunoblots were carried out according to established methods (Sambrook, 1989). The anti-GFP (Sigma) was used at 1:4000, and anti-Sin1 (Schroder *et al.*, 2005) was used at 1:100.

2.3 RESULTS

2.3.1 Prediction of interdomain regions within Sin1 proteins

The vertebrate orthologues of Sin1 were examined for regions of non-synonymous and non-conservative amino acid substitution, low complexity, protein disorder, and amino acid content (Pro, Gly, Asp, and Asn). Regions high in these qualities are good candidates for inter-domain linkers (Wootton and Drummond, 1989; Dyson and Wright, 2002; Tanaka *et al.*, 2003). Based on the consensus of these properties, three linker regions were predicted within the Sin1 protein: N112 to Q139, Y268 to T274, and S356 to Q368. This results in four major domains predicted within the full length Sin1 protein (Figure 2.3.1). These potential domains were submitted to 3D-Jury to identify protein folding.

There were no significant similarities identified by 3D-Jury for the first N-terminal domain (M1-R111). The consensus prediction was primarily alpha-helical, suggesting that this region might fold as a coiled-coil. The full length protein sequence was submitted to COILS, which identified a single region in the N-terminal domain predicted to fold as a coiled-coil (I80 to Q107). The next domain (S140-K267) corresponds to the Conserved Region In the Middle (CRIM) identified by Schroder et al (2004). The CRIM is most similar to the Ubiquitin-like beta-Grasp fold, a domain important for protein-protein interactions (Orengo *et al.*, 1994). The region S275-N355 scored as significantly similar (>50) to the RBDs of c-Raf and a-Raf. This section is analogous to the RBDs previously predicted for the RIP3 and AVO1 proteins. The final

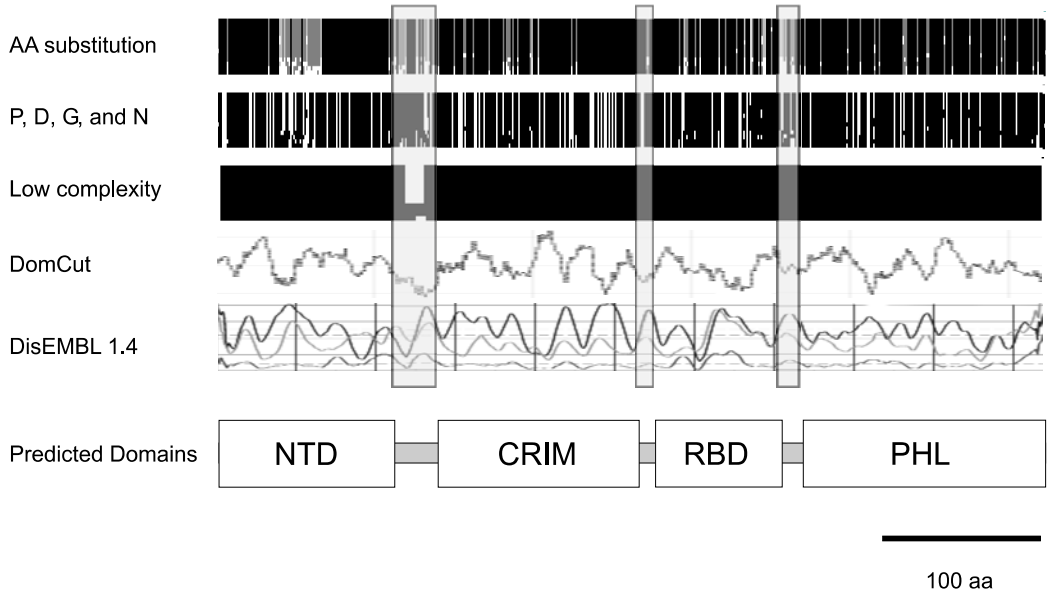


Figure 2.3.1. Prediction of interdomain regions in the full length Sin1 proteins. Consensus-based prediction of domain linkers, based on non-conservative and non-synonymous amino acid substitution, Pro, Asp, Asn, and Gly content, low complexity regions, and the output of DomCut and DisEMBL 1.4. Predicted interdomain regions are highlighted with the grey transparent box. The putative domains were submitted to the 3D-Jury meta server for protein fold prediction. NTD = N-terminal domain, CRIM conserved region in the middle, RBD = Ras binding domain, PHL = Pleckstrin homology like domain.

domain (I369-Q522) appeared similar to two superfamilies: the cupredoxin-like fold, and the PH domain-like folds, particularly the family of phosphotyrosine-binding domains. Although the cupredoxin folds scored higher over a small section (T398-K481), the overall structural alignment was better for the PH-like domain.

Thus, these results predict four domains in Sin1: an N-terminal domain that possibly contains a coiled-coil, the Conserved Region in the Middle, a Ras-binding domain, and a PH-like domain (Figure 2.3.1). Each of the splice variants used in this study has a different combination of domains. Sin1α has a disrupted RBD and lacks a

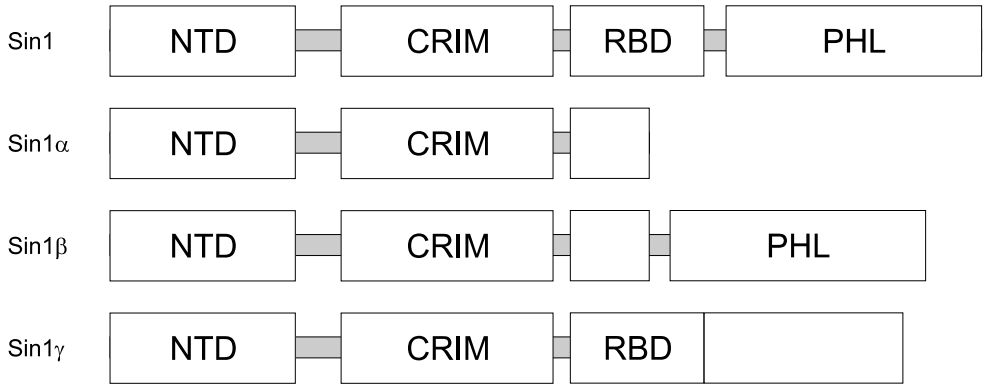


Figure 2.3.2. Different Sin1 isoforms comprise different combinations of domains. NTD = N-terminal domain, CRIM conserved region in the middle, RBD = Ras binding domain, PHL = Pleckstrin homology like domain.

PH-like domain altogether. Sin1β has a disrupted RBD but an intact PH-like domain, whereas Sin1γ has an intact RBD but a disrupted PH-like domain. These different domain combinations suggest different, and possibly opposing biological roles for these isoforms (Figure 2.3.2).

An examination of the evolutionary conservation of domains between the Sin1 orthologues also suggests diversity of function, as the CRIM is the only domain conserved in every member of the Sin1 family (Figure 2.3.3). The RBD is present in the worm (*C. elegans*) and fish species (*F. rubripes*), but absent in both insect species (*A. gambiae* and *D. melanogaster*), suggesting that interaction with Ras is not part of Sin1 function in these species. Similarly, the PH-like domain which is present in most

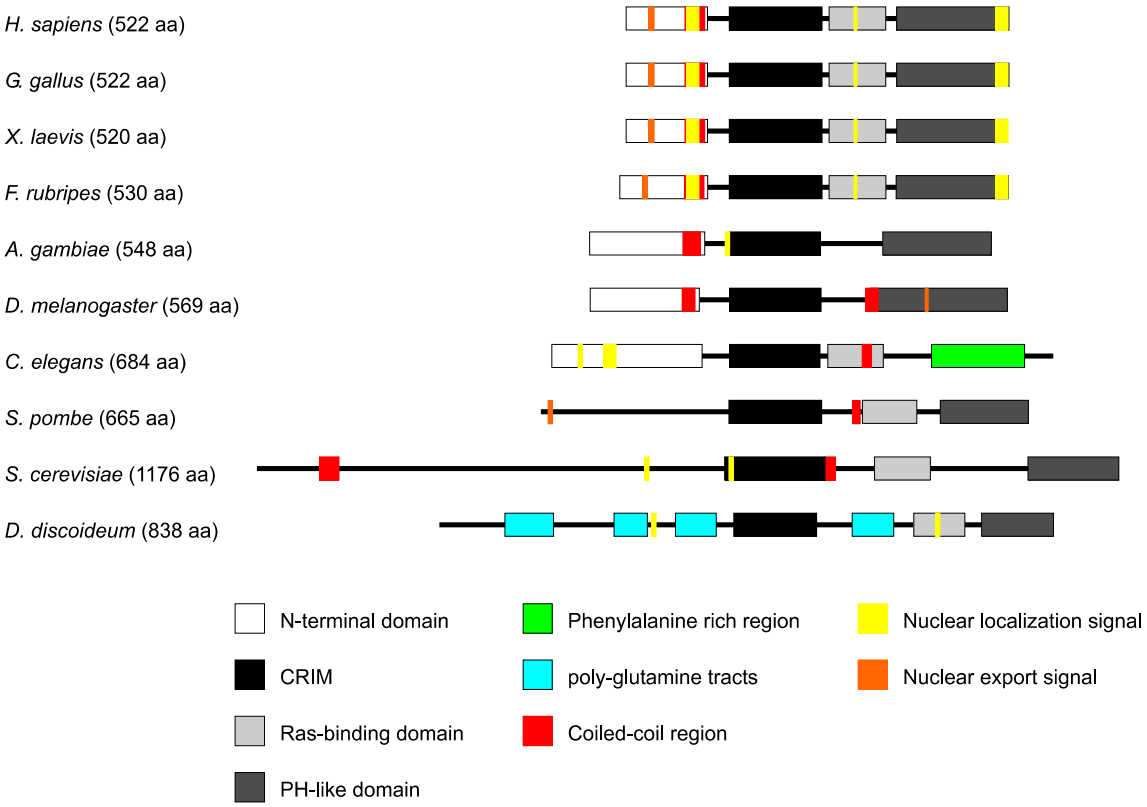


Figure 2.3.3. The evolutionary conservation of domains in Sin1 orthologues. A schematic representation of the domains and signals identified in this chapter, and their position in the Sin1 proteins. The species name is given to the left, and the length of the protein is shown in brackets. The full length protein is represented by the black line, and the identified domains or signals are represented by the coloured rectangles.

species, appears to be absent in *C. elegans*. Instead, this region in *C. elegans* is enriched with phenylalanine residues (38%). The largest difference between the Sin1 orthologues resides in the N-terminal region, where remnants of the domain cannot be detected in either of the yeast species or *D. discoideum*. Together, these data suggest that Sin1 may function differently in different species, although molecular interactions mediated by the

CRIM are likely to be conserved. Multiple sequence alignments of the conserved protein domains of Sin1 orthologues can be found in Appendix I. A complete amino acid alignment of insect and vertebrate Sin1 orthologues was published by Wang and Roberts (2005).

2.3.2 Sub-cellular localization of Sin1 and Sin1 splice variants

The localization pattern of endogenous Sin1 proteins was determined by IFA in HEK293 cells, where a complex pattern of nuclear (including nucleoli), cytoplasmic, and plasma membrane staining was observed (Figure 2.3.4A). To determine the contribution of each of the splice variants to the localization pattern observed for endogenous Sin1, EGFP tagged Sin1 fusion proteins were constructed. As the attachment of EGFP to Sin1 may affect its ability to fold appropriately, it may therefore affect its cellular localization. To address this, two Sin1 isoforms (Sin1, Sin1α) were constructed with EGFP fused to the N-terminus or the C-terminus, and their localization in Cos7 cells was compared with previously published studies using the small tags FLAG (Schroder *et al.*, 2005) and myc (Wang and Roberts, 2004). The N-terminally tagged proteins showed significantly less nuclear localization than the C-terminally tagged counterparts (Figure 2.3.5). As previous localization results using small N-terminal tags were most similar to those obtained by the C-terminally tagged EGFP constructs (Figure 2.3.4C), it is likely that fusing EGFP to the N-terminal of Sin1 isoforms interferes with appropriate nuclear localization of these proteins. Therefore, C-

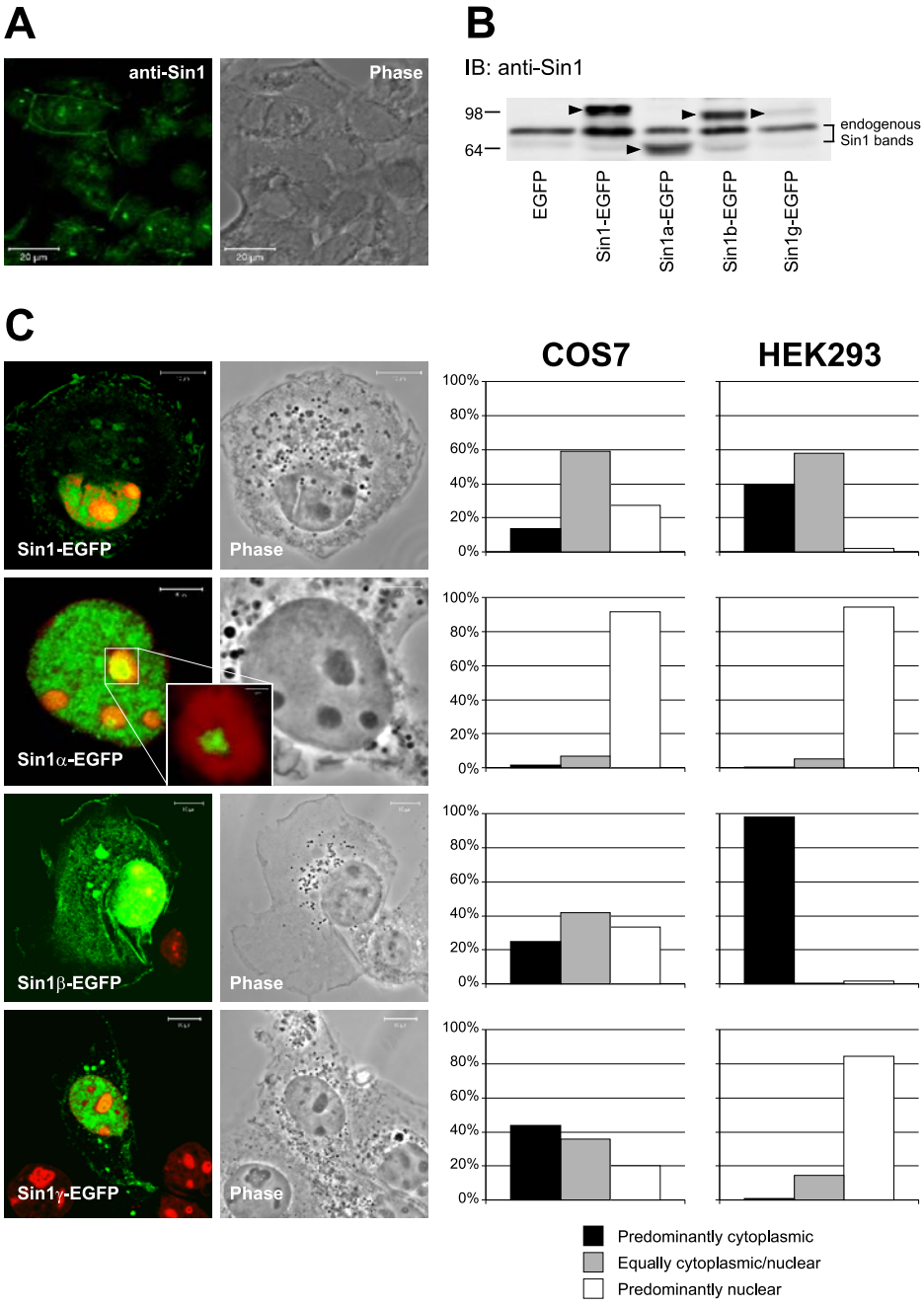


Figure 2.3.4. Localization of endogenous and over-expressed Sin1 isoforms. (A) HEK293 cells stained with an anti-Sin1 rabbit polyclonal antibody. (B) Expression of appropriately sized Sin1 constructs was confirmed by Western blot (black triangles). (C) C-terminally tagged EGFP constructs were transfected into COS7 cells, and harvested after 24 h. The slides were fixed, co-stained with propidium iodide, and analysed by confocal microscopy. The inset for Sin1α-EGFP shows a higher magnification of the nucleolar Sin1 staining, representative of that found in all Sin1 isoforms. The comparison of localization patterns between COS7 and HEK293 cells is shown on the right for each construct. At least 200 cells were counted from random fields of two or more independent transfections.

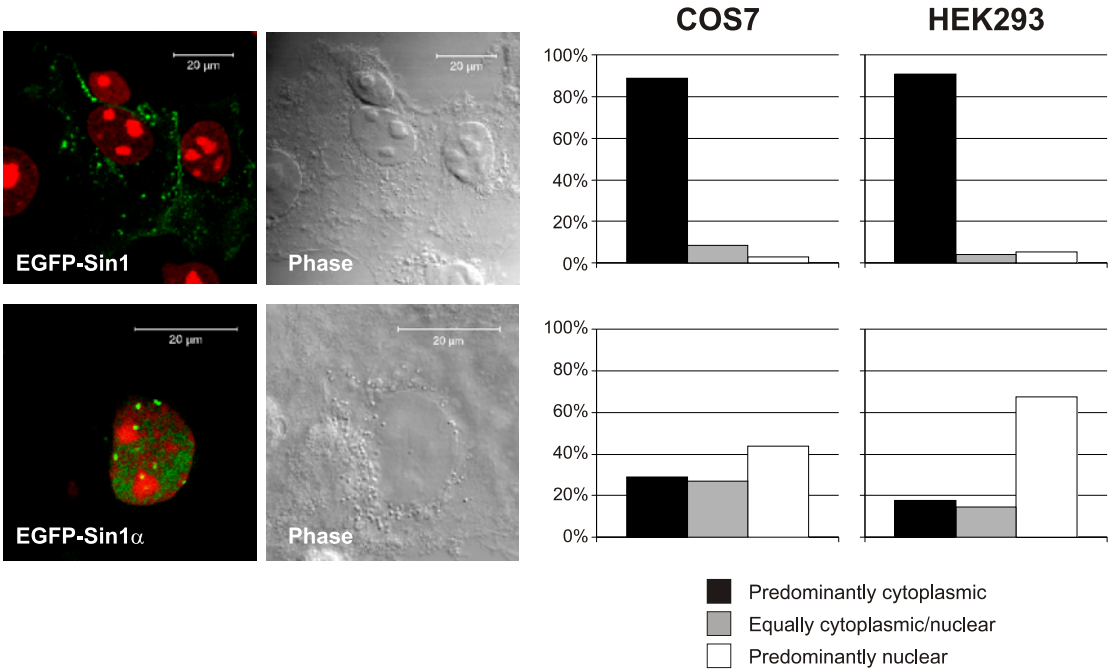


Figure 2.3.5. Comparison of N-terminally tagged EGFP Sin1 constructs. N-terminally tagged EGFP constructs were transfected into COS7 cells, and harvested after 24 h. The slides were fixed, co-stained with propidium iodide, and analysed by confocal microscopy. The comparison of localization patterns between COS7 and HEK293 cells is shown on the right for each construct, calculated as for Figure 2.2.3C.

terminally tagged constructs were used to examine the localization of Sin1 splice variants.

Expression of C-terminally tagged Sin1, Sin1α, Sin1β, and Sin1γ in transiently transfected Cos7 cells was confirmed by Western blot (Figure 2.3.4B). Full length Sin1 was largely membrane associated, although many cells also exhibited nuclear fluorescence. The same pattern of nuclear staining was observed for all isoforms studied (Figure 2.3.4C). This nuclear fluorescence was granular, and there was very little co-

localization with DNA (the propidium iodide stained regions), although co-localization was more pronounced in the nucleoli. Upon higher magnification, it became apparent that even in the nucleolus, intense fluorescence was largely restricted to areas that lack DNA (Figure 2.3.4C, inset). Unlike full length Sin1, no instances of plasma membrane localization were observed for the predominantly cytoplasmic Sin1γ isoform. Largely, this cytoplasmic localization was punctate, sometimes found in or near phase-dark organelles. Sin1α was predominantly nuclear, although a small percentage of cytoplasmic staining was also observed. Plasma membrane localization occurred only with Sin1 and Sin1β, both of which contain intact PH-like domains. The dual cytoplasmic/nuclear localization of the Sin1 proteins suggested that there may be motifs controlling the import and export of this protein between the nucleus and the cytoplasm, and that both are likely to be present in the N-terminal half of the protein (as this region is conserved in all of the isoforms).

To ensure that the results obtained in Cos7 cells were representative, the localization of these constructs in HeLa, DG75, and HEK293 cells was also examined. Similar cellular localization was found in all cell lines for each of the Sin1 isoforms. The largest differences were seen for Sin1β and Sin1γ in HEK293 cells, where the proteins were predominantly cytoplasmic and predominantly nuclear, respectively. The quantitative comparison between Cos7 and HEK293 is shown in Figure 2.3.4C. In DG75 and HeLa cells (data not shown), the localization distributions resembled those found in Cos7 cells.

2.3.3 Multiple domains control the nuclear localization of the Sin1 proteins

Several nuclear localization signals have been previously predicted for the Sin1 protein (Schroder *et al.*, 2004), two of which reside in the N-terminal half of the protein (R81-R97: bipartite NLS; and K310-K313: pattern 4 NLS). NetNES predicted a potential nuclear export signal in the N-terminal domain (L31-L36), suggesting that this domain may be involved in shuttling Sin1 between the nucleus and the cytoplasm. To examine the ability of this domain to target EGFP in Cos7 cells, we constructed three N-terminally tagged fusion proteins that covered the full length of the N-terminal domain (EGFP-NTD: M1-S113), and two previously identified conserved regions (Wang and Roberts, 2005) that fall within this domain (EGFP-SCDI: M1-E39; EGFP-SCDII: A67-S113). Sin1 Conserved Domain I (SCDI) contained the putative NES, and SCDII contained the bipartite NLS (Figure 2.3.6A). The expression of the constructs was confirmed by Western blot (Figure 2.3.6B), and the qualitative and quantitative results of their localization are shown in Figure 2.3.6C. The full length N-terminal domain was excluded from the nucleus, whereas the EGFP-SCDII construct was nuclear, particularly staining the nucleoli. The EGFP-SCDI construct had a distribution very similar to EGFP alone. In HEK293 cells however, a larger proportion of EGFP-SCDI fluorescence was excluded from the nucleus, suggesting that this motif may be functional in different cellular contexts.

The CRIM has no predicted nuclear localization signals. However, a C-terminally tagged fusion construct of the CRIM (EGFP-CRIM: P133-K267) was often targeted to the nucleus in Cos7 and HEK293 cells (Figure 2.3.6C), suggesting either the

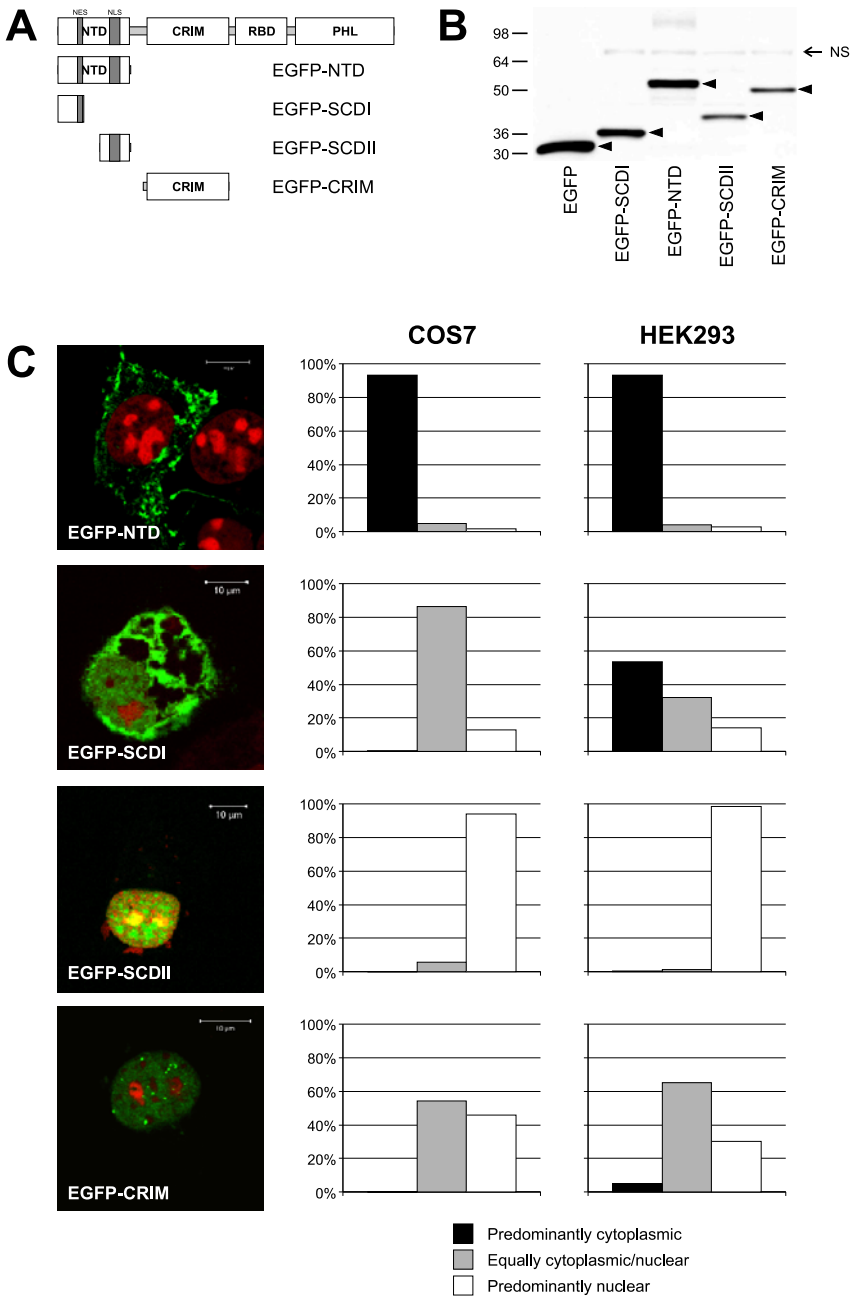


Figure 2.3.6. Multiple domains contribute to the nuclear localization of Sin1. (A) A schematic diagram of the N-terminally tagged partial Sin1 constructs used in this study. (B) Expression of appropriately sized EGFP constructs was confirmed by Western blot, identified by the black triangles. In this blot, one tenth of the EGFP cell lysate was loaded (compared with the other lanes) to avoid swamping the signal. A non-specific band often seen with this antibody is identified (NS). (C) N-terminally tagged partial Sin1 constructs were transfected into COS7 cells, and harvested after 24 h. The slides were fixed, co-stained with propidium iodide, and analysed by confocal microscopy. The comparison of localization patterns between COS7 and HEK293 cells is shown on the right for each construct, calculated as for Figure 2.3.4C.

presence of a novel NLS, or that it was transported to the nucleus by binding partners. In contrast to the nuclear localization for the N-terminal domain, the CRIM was largely excluded from the nucleolus, and was found in intense foci throughout the rest of the nucleus. Together, the localization of the NTD and the CRIM accounted for the nuclear localization observed in all over-expressed protein isoforms, and also the endogenous Sin1 proteins.

2.3.4 The PH-like domain of Sin1 targets fusion proteins to the plasma membrane and lipid rafts

The localization of the Sin1 splice variants indicated that an intact PH-like domain appeared to be necessary for targeting Sin1 proteins to the plasma membrane. To determine whether this domain was sufficient for plasma membrane localization, a fusion protein of the Sin1 PH-like domain (EGFP-PHL: D370-Q522) tagged at the N-terminal with EGFP was constructed and transfected into Cos7 cells. This protein localized to the plasma membrane (Figure 2.3.7A), indicating that the PH-like domain was both necessary and sufficient for membrane localization of Sin1 proteins.

The Sin1 PH-like domain was previously shown to bind to phosphatidylinositol derivatives *in vitro* (Schroder *et al*, manuscript submitted). To determine whether lipids target Sin1 to the plasma membrane *in vivo*, neomycin was used to mask the polar heads of phosphoinositides (Wang *et al.*, 1984; Welch and Feramisco, 1984; Bompard *et al.*, 2003). Protein-lipid overlay assays were used to confirm that neomycin treatment significantly reduced the *in vitro* association of GST-PHL with phosphoinositides

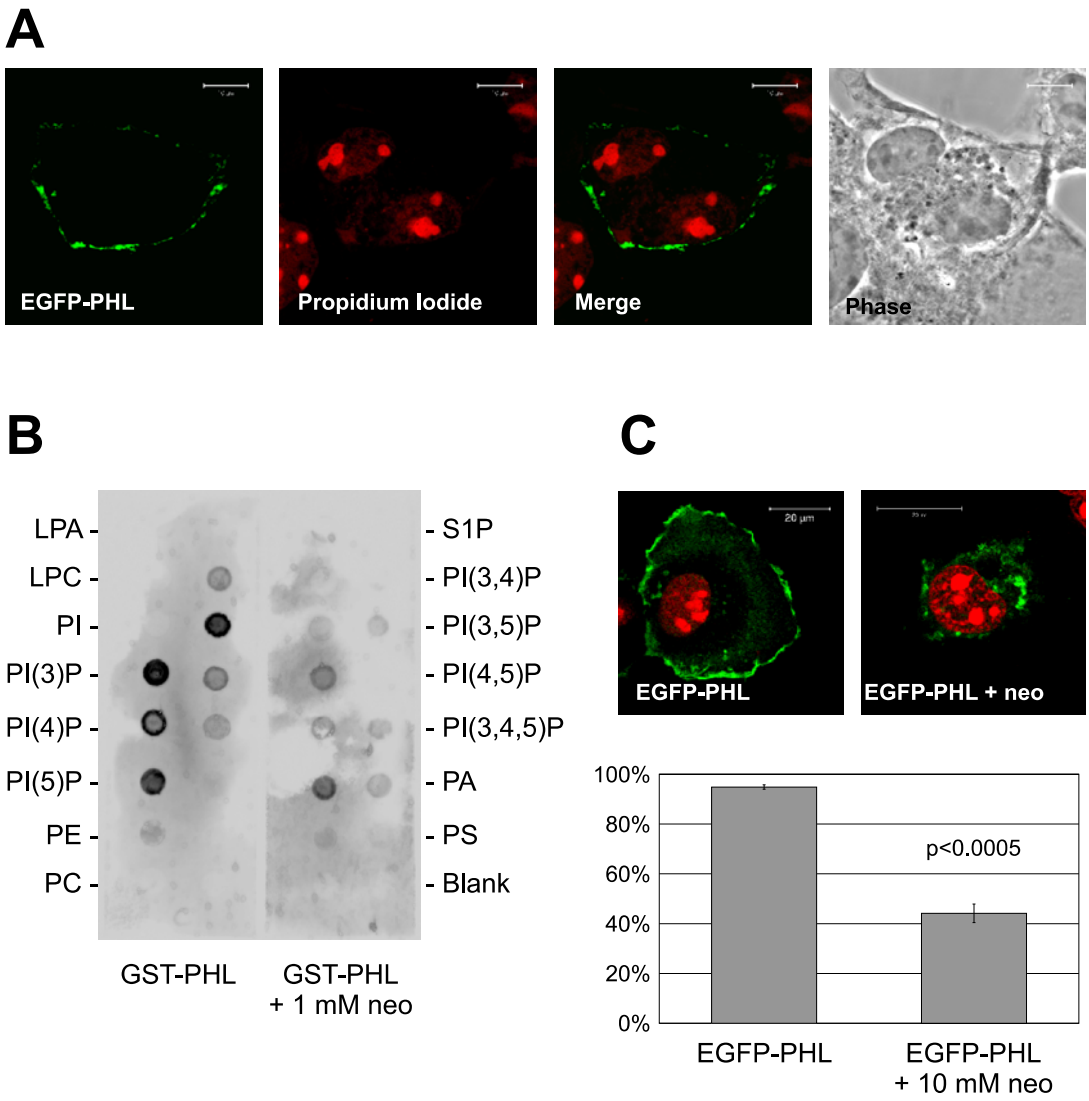


Figure 2.3.7. Phosphatidylinositol-phosphate (PIP) derivatives target Sin1 to the plasma membrane *in vivo*. (A) EGFP-PHL was transfected into COS7 cells, and harvested after 24 h. The slides were fixed, co-stained with propidium iodide, and analysed by confocal microscopy, revealing plasma membrane staining of this construct. (B) Various lipids arrayed on a membrane (PIP-strips; Echelon Biosciences) were incubated with affinity purified Sin1 PH-like domain (GST-PHL) in the presence or absence of 1 mM neomycin (neo). The presence of protein was detected using an anti-GST antibody. (C) Cos7 cells were transiently transfected with EGFP-PHL in the presence or absence of 10 mM neomycin. Confocal microscopy was used to image at least 200 cells from random fields of three independent experiments. The graph shows the average percentage of cells with significant membrane localization (and standard error). Representative cells are shown above the graph.

(Figure 2.3.7B). Cos7 cells transiently transfected with EGFP-PHL were then used to assess the effect of neomycin on fusion protein localization. Treatment with neomycin reduced the proportion of cells showing significant plasma membrane by more than 50% (Figure 2.3.7C), suggesting that it is likely to be the lipid binding property of the Sin1 PH-like domain that targets Sin1 proteins to the plasma membrane. As Sin1 is implicated in signalling pathways, and lipid rafts are important scaffolds for the initiation of signal transduction, the possible targeting of Sin1 proteins to lipid rafts was examined. DG75 cells transiently transfected with Sin1-EGFP, Sin1β-EGFP, or EGFP-PHL were co-stained for lipid rafts using a Vybrant Lipid Raft Labelling Kit (Molecular Probes). All constructs were targeted to lipid rafts (Figure 2.3.8), indicating that the Sin1 PH-like domain was sufficient for this localization.

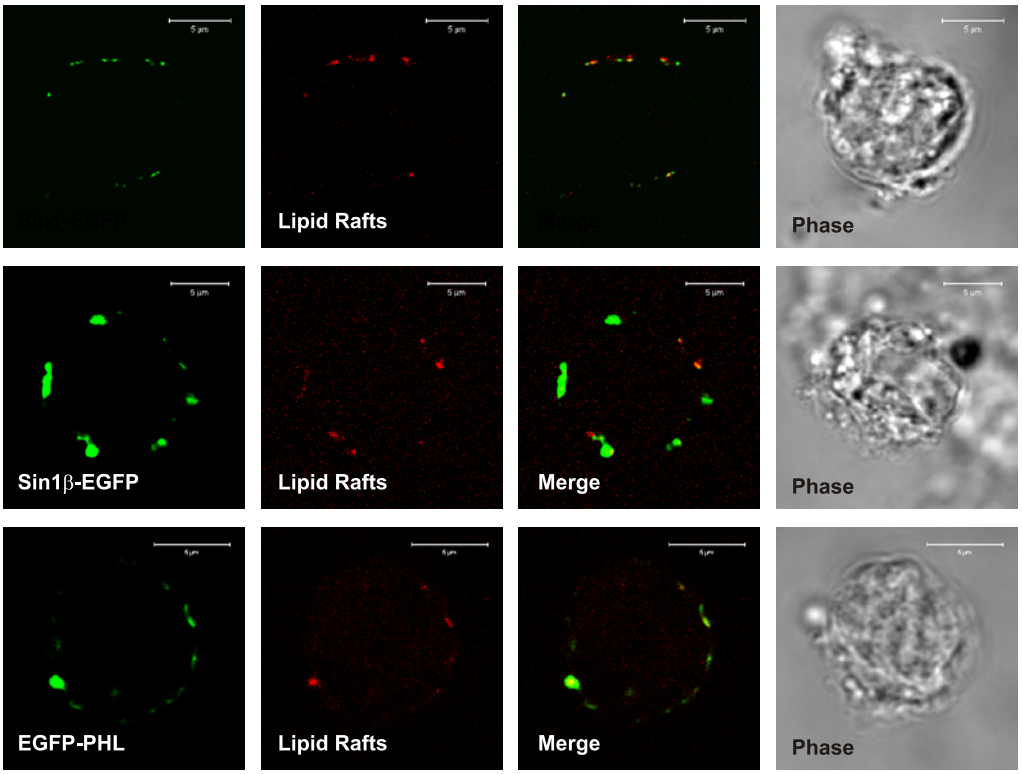


Figure 2.3.8. The PH-like domain targets Sin1 to lipid rafts. DG75 cells were transfected with either EGFP-Sin1, EGFP-Sin1β, or EGFP-PHL, and co-stained using a Vybrant lipid raft labelling kit (Molecular Probes). Confocal microscopy shows co-localization of the fluorescence for each construct used.

2.4 DISCUSSION

We have identified the major domains present in the Sin1 protein by defining inter-domain regions, and using structural homology-based searches to classify the folds. Two of the domains identified here, the RBD and the PH-like domain, were also identified in a detailed sequence homology-based study (Schroder *et al*, manuscript submitted). However, that study could not make predictions for the N-terminal half of the protein, emphasizing the need to look beyond primary sequence homology when identifying divergent domains. None of the domains identified appear to be catalytic, which is consistent with the idea that Sin1 proteins function as signalling adaptors.

Although brief descriptions of Sin1 localization have been reported previously (Wang and Roberts, 2004; Schroder *et al.*, 2005), this study is the first in depth examination of Sin1 localization and the motifs that control it. We have shown that Sin1 proteins are targeted to the plasma membrane by the PH-like domain, as the result of interactions with phospholipids *in vivo*. A sequence pattern (KVSMIHRLRF), very similar to a consensus phosphatidylinisitol-phosphate (PIP) binding motif reported by Isakoff *et al* (1998), was identified in the region deleted in the Sin1γ isoform (S357-G403). As Sin1γ does not localize to the plasma membrane, it is very likely that this PIP binding motif is responsible for the plasma membrane localization of Sin1 proteins. The PH-like domain also targets Sin1 proteins to lipid rafts in DG75 cells, although whether this results from the binding of phospholipids or from protein-protein interactions is not yet known. Both possibilities are conceivable: IFNAR2 may localize

to lipid rafts (Takaoka *et al.*, 2000) by interacting with Sin1 in the PH-like domain (Wang and Roberts, 2004), or Sin1 might bind the phosphatidylinositol derivatives enriched in lipid rafts (Hope and Pike, 1996). In either case, the localization of Sin1 and Sin1β in lipid rafts further strengthens the hypothesis that Sin1 is involved in a signal transduction-related process.

We also present the first localization data for two Sin1 isoforms: Sin1β and Sin1γ. All four Sin1 isoforms have both cytoplasmic and nuclear locations, suggesting that these proteins may shuttle between the cytoplasm and the nucleus. The N-terminal domain may regulate this process, as it contains a potential NES and a functional NLS, and the localization pattern of Sin1 can be disturbed by the N-terminal fusion of EGFP. The N-terminal domain sequence is also likely to be sufficient to explain the difference in the localization patterns of Sin1 splice variants in HEK293 and Cos7 cells (Fig. 2.3.6C), although the mechanisms for this are not yet known. It is also interesting that mammalian TOR (mTOR) also has a different localization in HEK293 cells as compared to the other cell lines (Zhang *et al.*, 2002). Binding of Sin1 orthologues to TOR2 has been shown in yeast and *Dictyostelium*, but has not yet been demonstrated for human Sin1 and mTOR. Co-expression of Sin1 and TOR in mammalian systems has been reported (Loewith *et al.*, 2002), and both are expressed in analogous regions of the developing vertebrate hindbrain (Hentges *et al.*, 1999; Christiansen *et al.*, 2001). The possible interaction between these two proteins clearly deserves further investigation.

The use of EGFP tags allowed detection of more structure in the nucleoplasmic

pattern than previously reported, and for the first time, the nucleolar localization of Sin1 proteins has been revealed. While the nucleolus is primarily known for its role in ribosome biogenesis, it also appears to function in RNA transport, RNA modification, and regulation of the cell cycle (reviewed in Olson *et al.*, 2002; Pederson, 1998; Visintin and Amon, 2000). In this context, it is worthwhile noting that the *S. Cerevisiae* Sin1 orthologue (AVO1) has been affinity precipitated with both CEG1 (Gavin *et al.*, 2002) and RRP9 (Ho *et al.*, 2002). CEG1 is a subunit of a nuclear mRNA capping enzyme (Shibagaki *et al.*, 1992), and RRP9 is a nucleolar protein that binds snoRNA and forms part of a complex involved with rRNA modification and processing (Venema *et al.*, 2000). This association with the nucleoli may be transitive, and dependent upon the cellular context, as not all cells with nuclear Sin1 fluorescence show staining of the nucleoli. There is some literature supportive of this idea, with certain stress response proteins having been shown to accumulate in the nucleoli of stressed cells (Welch and Feramisco, 1984). The role of Sin1 in nucleoli is unclear, but the idea that Sin1 may be involved with the transduction of signals to these compartments seems reasonable.

This chapter has delineated the multiple domains and motifs that contribute to the localization of Sin1 proteins, providing some insight into the possible functions of this protein. The next chapter employs bioinformatics to predict the regulation and pathways in which Sin1 lies.

CHAPTER THREE

An investigation of the regulatory regions of Sin1

3.1 INTRODUCTION

The human *Sin1* gene undergoes alternative polyadenylation and alternative splicing, and four human *Sin1* mRNA variants have been confirmed *in vivo* (Schroder *et al.*, 2004). The transcription and splicing of these isoforms are likely to be regulated, but identification of a functional *Sin1* promoter or other regulatory elements has not previously been attempted. The conservation of protein domains across Sin1 orthologues (Chapter 2), and recent evidence from yeast and *Dictyostelium* (Lee *et al.*, 2005; Wullschleger *et al.*, 2005) suggests that some Sin1 functions may be conserved throughout evolution. It is therefore reasonable to speculate that the regulation of *Sin1* expression, and therefore genomic *Sin1* regulatory elements, may also be conserved. Such regions of conservation can often be identified through comparative genomic analysis.

The power of comparative genomic analysis to identify regulatory elements in the non-coding regions of genes is sensitive to the conservation of gene expression patterns (Williams *et al.*, 2003). Thus, it is important to establish whether orthologous *Sin1* genes are structurally similar, and whether the *Sin1* splice variants observed in humans are likely to be conserved in other species. During the course of this project, a study was published examining the conservation of gene structure between *Sin1* orthologues (Wang and Roberts, 2005), finding that human, mouse and rat genes were similarly structured in terms of splicing and relative intron length, consistent with our own data. However, the authors did not examine the non-coding exon θ, which can be

difficult to identify in the absence of EST data. In the two fish examined (*F. rubripes* and *D. rerio*) the overall splicing pattern was conserved, although the proportional intron lengths were quite different between the two species, as well as between fish and mammals. This data suggests that while comparison of fish and human genomes can be of great benefit for some genes, the evolutionary distance may be too great for a useful comparison of *Sin1* regulatory elements.

In this chapter, the genomic structures and sequences of human, mouse, rat, chicken, and dog *Sin1* genes are compared, and a number of potential regulatory regions and their genomic context are described. The conservation of splicing between human and mouse is also examined, and a functional promoter located within a conserved CpG island is identified. The regulatory elements identified with approach, along with an analysis of gene expression data, were used to identify similarly regulated genes. Collectively, these results provide important insights into the biological role of Sin1.

3.2 MATERIALS AND METHODS

3.2.1 Sequences and sequence analysis

The contigs containing the *Sin1* genomic sequences used in this study were as follows: Human (Genbank Accession Number NT_008470), mouse (NT_039206), rat (NW_047652), and chicken (NW_060627). Dog *Sin1* genomic sequence was identified in a TblastN search of the NCBI's Reference Dog Genome (boxer), and encompasses two non-overlapping contigs (GenBank Accession Numbers AAEX01025409 and AAEX01025408). The Tetradon genome was browsed at http://www.genoscope.cns.fr/externe/tetranew/.

BLAST2Sequences and EMBOSS Dotmatcher (EBI; http://www.ebi.ac.uk) were used to identify conserved intronic regions larger than 100nt and greater than 75% identity, ignoring elements not found in the same relative position in all species. MAR-Wiz (http://www.futuresoft.org/MAR-Wiz/) and SMARtest (http://www.genomatix.de) were used to search for potential scaffold/matrix attachment regions. For MAR-Wiz, the cut-off threshold was set to 0.4, clip peaks was enabled, and the defaults for all other settings were used. Repetitive sequence was detected using Repeatmasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). CPGplot and CPGreport from EBI (http://www.ebi.ac.uk) were used to identify CpG islands. Prediction of secondary structures and free energy values of the 5'-UTRs were performed at the MFOLD server (http://www.bioinfo.rpi.edu/applications/mfold/). The Genomatix tools MatInspector, and FrameWorker (http://www.genomatix.de) were used to identify

conserved transcription factor binding sites, and ModelInspector was used to identify human genes containing the conserved *Sin1* promoter module. UTRScan (http://bighost.area.ba.cnr.it/BIG/UTRScan/) was used to identify functional sequence patterns located in the *Sin1* 5' and 3' UTRs. Tmm (http://microarray.cpmc.columbia.edu/tmm/index.html) was used to explore the list of genes co-expressing with *Sin1* based on multiple microarray data sets.

3.2.2 RNA isolation and RT-PCR

OLF422 cells (a mouse olfactory neuronal cell line (Zehntner *et al.*, 1998)) were maintained in DMEM (GibcoBRL) with 10% foetal calf serum (GibcoBRL). Total RNA was isolated from these cells using Trizol reagent (GibcoBRL) according to the manufacturer's instructions. RT-PCR was performed using the SuperScript One-Step RT-PCR kit (Invitrogen). *Sin1*α was detected using the Sin1-F11 and Sin1-R4 primer pair, and *Sin1*β was detected using Sin1-F11 and Sin1-R5 primer pair (Schroder *et al.*, 2004). The identity of detected bands was confirmed by direct sequencing. The author acknowledges Zeke Barnard, who assisted with the RNA isolation and RT-PCR for this part of the study.

3.2.3 Plasmid construction and luciferase assays.

PCR products generated with either CpG-F2 (5' - gctagcagtctccgacccaggcct - 3'), CpG-F3 (5' - gctagctccatattacagatattataataagagaa - 3'), CpG-F4 (5' - gctagcggttagacccac

cgtg - 3'), or CpG-F6 (5' - gctagcgacctcagaaaatcaaatatccaa - 3') and the reverse primer CpG-R2 (5' - aagcttgtgagagggccgccaggt - 3') were cloned into pGEM-T (Promega), and then subcloned into the NheI and HindIII sites of pGL3-Basic (Promega). HEK293 cells were cultured and transfected as per section 2.2.3. Test plasmid (1 μg) was co-transfected with 1 μg of pEGFP-C3 as a transfection efficiency control. Cells transfected with pGL3-Basic were used as a background control. Transfected cells were cultured for 48 hr prior to analysis of luciferase expression using a Dual-Luciferase Reporter Assay System (Promega). The author acknowledges Kang Wee Tay, who assisted with the plasmid construction and luciferase assays for this part of the study.

3.3 RESULTS

3.3.1 Genomic location, synteny, and structure of vertebrate Sin1 genes

The previously described 3393 nt sequence encoding full length human *Sin1* (GenBank Accession Number AY524430) maps to a gene located on chromosome 9 at position q34.1, which spans approximately 270kbp. It is flanked by the PBX3 gene on the telomere side, and the GAPVD1 gene on the centromere side. The PBX3 and GAPVD1 genes are transcribed from the same strand, whereas the *Sin1* gene is transcribed from the opposite strand. The structure of the human *Sin1* gene, including introns and splicing, is shown in Figure 3.3.3A. The size and relative conservation of the *Sin1* introns is detailed in Table 3.3.1.

The mouse and chicken *Sin1* genes share a similar arrangement in their genome to human *Sin1*. The mouse *Sin1* gene (GenBank Accession Number NM_177345) spans 210 kb on chromosome 2, flanked by Pbx3 and Gapvd1. The chicken *Sin1* gene (GenBank Accession Number NM_204826) spans 85 kb on chromosome 17, also flanked by PBX3 and a gene similar to mouse Gapvd1 (LOC417095). In both genomes, the telomere to centromere synteny is maintained. The region corresponding to *Sin1* in the rat has been automatically predicted and annotated as LOC296648, which theoretically encodes a protein 601 amino acids long. Using NCBI's Model Maker to examine the EST evidence for this gene suggests a structure with greater homology to human and mouse *Sin1*: a gene spanning 203 kb with 12 constitutively spliced exons coding for a 522 aa polypeptide with 96% identity to human Sin1 and 99% identity to

mouse Sin1. Dog *Sin1* genomic sequence spans two non-overlapping contigs on chromosome 9, so the size of the gene cannot be accurately determined, although it is no smaller than 245 kb. The predicted polypeptide is 522 aa with 99% identity to human Sin1. The exon/intron boundaries are very similar in all species, with the only exception being the exon θ and exon 1 splice. In rodent mRNA, extra non-coding sequence is apparent at the 5' end of exon 1 (data not shown), but as overall splicing is unaffected, the significance of this is unknown.

Due to a lack of canine EST sequences for this region, it was not possible to properly map the non-coding exon θ in the dog genome. However, based on blast searches, a putative exon θ located approximately 29 kb upstream from exon 1 could be identified. The size of this first putative intron is in accordance with the relative sizes of the other canine *Sin1* introns (between 85-95% of the respective human intron size; Table 3.3.1). Further, a 10 nt motif (TTCCGGGTCG), 100% conserved in human, mouse, rat, and chicken *Sin1* genes, is present in this exon. Typically found at +3 relative to the earliest transcription start site (TSS), this sequence allows the prediction of a canine TSS at position 19776 of the AAEX01025409 sequence. The predicted exon is 258 nt long and is 81% identical to the human sequence.

3.3.2 Structure and conservation of the 5'-UTR

The 5'-UTR of human *Sin1* is comprised of the non-coding exon θ and part of exon 1, and is at least 330 nucleotides long. It contains two short upstream ORFs of 9

and 5 amino acids in length, and a 241 nt stretch with a very high GC content (77 %) at the 5'-terminus. An IRES (internal ribosome entry site) pattern located at nucleotides 175-261 of the human *Sin1* sequence was identified using UTRscan (Pesole *et al.*, 1999). Potential secondary structures of the *Sin1* 5'-UTR were predicted using MFOLD (Mathews *et al.*, 1999; Zuker, 2003), and these have calculated free energy values ranging from -145.4 to -152.9 kcal/mol, indicating the presence of stable secondary structure.

The conservation of sequence in the 5'-UTR across all species studied is significantly lower than any of the coding regions (data not shown), and consequently the differences in 5'-UTR lengths between chicken (193 nt), human (330 nt), mouse (462 nt), and rat (472 nt) are unsurprising. The free energy values of predicted secondary structures (Mathews *et al.*, 1999; Zuker, 2003) for *Sin1* orthologues reflect these differences in length. The much shorter chicken 5' UTR has values ranging from -87.7 to -91.8 kcal/mol, significantly higher than those predicted for mouse (-219.4 to -230.5 kcal/mol), or rat (-221.7 to -233.2 kcal/mol).

3.3.3 Structure and conservation of the 3'-UTR

Alignment of the final exon (11) from human, mouse, rat and dog *Sin1* genes revealed four conserved regions in the 3'-UTR, three larger than 100 nt and one 39 nt long. The two largest regions surround functional polyadenylation signals previously identified in human *Sin1* (Schroder *et al.*, 2004). The remaining 3'-UTR sequence

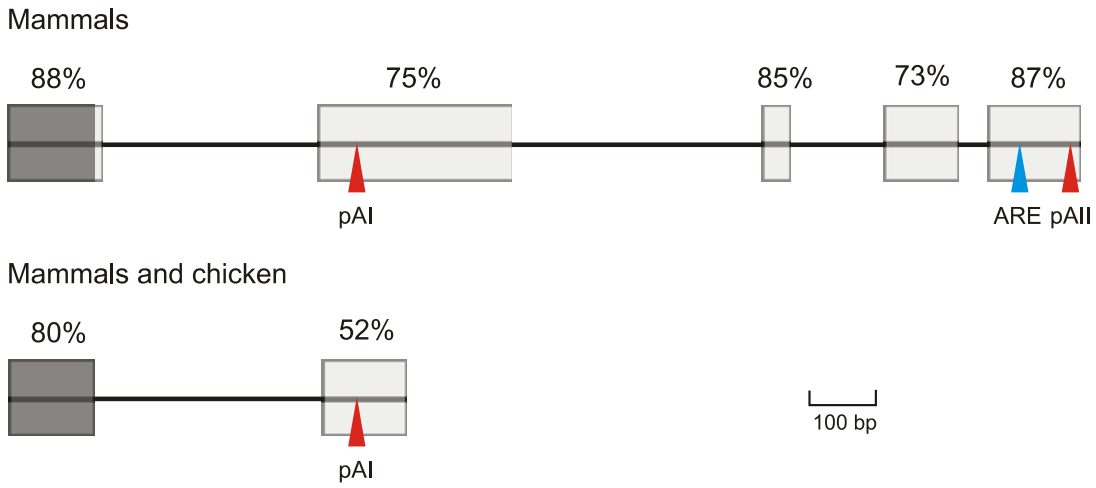


Figure 3.3.1 The conservation of *Sin1* exon 11 between orthologous *Sin1* genes was studied using Human, mouse, rat, and dog genes, with or without chicken *Sin1*. The boxes represent conserved sequences, dark grey for ORFs or light grey for untranslated regions. The overall identities between the species compared are listed above each region of conservation. The first and second polyadenylation sites (pAI and pAII, respectively) and the A/T rich Regulatory Element (ARE) are indicated.

shows little conservation. The region surrounding the second polyadenylation signal (pAII) is very A/T-rich (79%), and contains a ATTTA pentanucleotide motif. This is similar to the A/T-rich regulatory element (ARE), a cis-element that can decrease the half-life of mRNAs and is usually found in genes subject to tight regulation (Wilusz *et al.*, 2001). The structure of the 3'-UTR of chicken *Sin1* is different to mammalian *Sin1* (Figure 3.3.1), lacking a second polyadenylation site and other conserved sequence blocks, including the potential ARE. Only one region of similarity was detected, surrounding the first polyadenylation signal (pAI). This suggests that the alternative polyadenylation detected in humans occurs in other mammalian species, but not in the

chicken, and any regulatory function which may be associated with this region is specific to mammals.

3.3.4 Conservation of splice variants in the mouse transcriptome

Aside from full length *Sin1*, three other splice variants (confirmed by RT-PCR) are transcribed from the human *Sin1* gene: *Sin1*α is formed from the splicing of exon 6 to the alternative exon 7a, which contains both a stop codon and a polyadenylation signal; *Sin1*β mRNA which lacks exon 7; and *Sin1*γ lacks exon 8. *Sin1*β has been identified *in vivo* from sheep (Wang and Roberts, 2004). EST evidence suggests additional (non-confirmed) splice variants may also be conserved between the human, mouse, and cow transcriptomes (Schroder *et al.*, 2004). Although there is no EST support for *Sin1*α in any other organism, there may be insufficient sampling in the current EST databases to detect minor transcripts.

Exon 7a was identified by BLAST in all four non-human species with an overall identity of 78%. To investigate whether exon 7a is functional in the mouse, RT-PCR was performed on RNA isolated from OLF422 cells. Using primers specific for exon 5 and exon 7a, a band corresponding to the expected size of a *Sin1*α amplicon (409 bp) was detected, confirming the transcription of this variant in the mouse (Figure 3.3.2). The conservation of *Sin1*β was also investigated, using primers specific to exon 5 and exon 9, and this RT-PCR detected bands matching both the full length transcript (442 bp) and the β variant (334 bp). The identity of the PCR bands was confirmed by direct

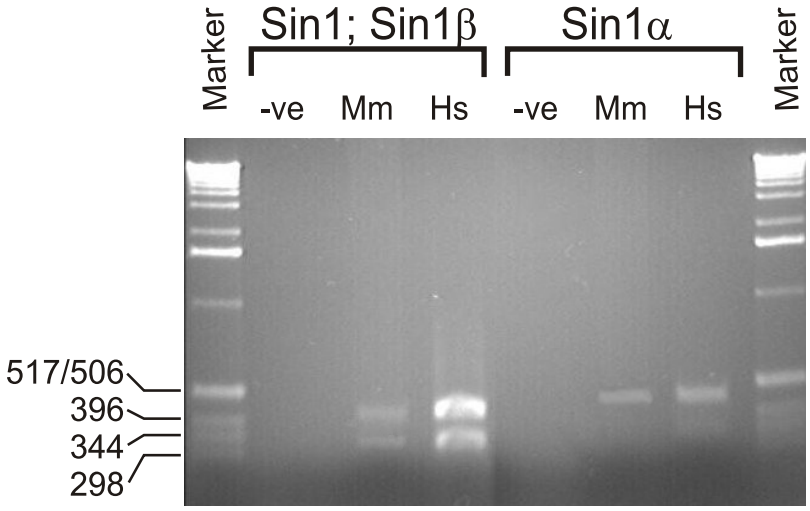


Figure 3.3.2 RT-PCR from mouse and human total RNA to detect splice variants. Primers binding exon 5 and exon 9 were used to detect *Sin1* and *Sin1*β transcripts, and primers binding exon 5 and exon 7a were used to detect *Sin1*α transcripts. Mm = Mus musculus; Hs = Homo sapiens; -ve = no DNA control.

sequencing. This data demonstrates that the *Sin1*α and *Sin1*β splice variants are conserved between the human and mouse transcriptomes, and this suggests that other splicing events may also be conserved.

3.3.5 Conserved elements within Sin1 introns

Regions of conservation between human and chicken *Sin1* introns were identified using dotmatcher (EBI) and blast2sequences (NCBI), and the relative regions of all five species were then aligned. Twenty five regions of high conservation were identified, with

identified, with overall identities ranging between 76% and 97% (Table 3.3.1). None of these conserved elements contained repetitive sequence (such as ALUs) that could be detected by RepeatMasker. Most introns contain at least one conserved element, the exception being intron 3. Intron four is the largest intron and also contains the highest number of conserved elements. The longest conserved element is located within intron 2 (598 nt; 96% identity). This intron also has the highest overall identity between all species. Conserved intronic elements represent 2.89% of the total human *Sin1* genomic sequence, and 4.89% of the non-coding, non-repetitive genomic sequence. The relative positions of these elements are presented graphically in Figure 3.3.3B. The conservation of these elements from chicken to human suggests that these are ancient regulatory elements, likely to be important for the conserved function of *Sin1*. Some of these elements may act as enhancers or silencers of the *Sin1* promoter, others as structural motifs, while some may regulate splicing or even code for regulatory RNA. Although the functions of these elements cannot be resolved without experimentation, the genomic context of these elements provides clues to their putative roles.

3.3.6 S/MAR analysis of human Sin1 genomic sequence

Scaffold/matrix attachment regions (S/MARs) are important DNA regulatory elements, where genomic DNA attaches to the nuclear matrix and brings it in close proximity to transcription factors, mediating a critical step in transcription (Bode *et al.*, 1995; Boulikas, 1995; Bode *et al.*, 2000). Potential scaffold/matrix attachment regions

Table 3.3.1. Conserved elements within *Sin1* introns. Intron sizes, percent identity , and regions of conservation within the *Sin1* genes of human (Hs), mouse (Mm), rat (Rn), dog (Cf), and chicken (Gg) are listed with their EST matches.

#[a]	Hs	Mm[b]	Rn[b]	Cf[b]	Gg[b]	Regions of High Identity[c]	Location[d]	EST/cDNA matches (species)
1	34327	24748 (23%)	25351 (20%)	~29kb (51%)	7311 (6%)	1.1: 299 nt (88%)	35779005-35778707	BY717877 (Mm)
						1.2: 212nt (90%)	35763568-35763357	DT887683 (Bt), DT889079 (Bt)
2	2408	2802 (55%)	2550 (59%)	2400 (81%)	3456 (37%)	2.1: 598nt (96%)	35753988-35753391	Not detected
3	12018	9035 (33%)	9003 (32%)	13722 (46%)	4148 (12%)	Not detected	-	-
4	71923	53294 (31%)	49333 (29%)	Part 1: 9064 (57%)	19893 (9%)	4.1: 251nt (95%)	35735984-35735734	Not detected
				Part 2: 60091 (50%)		4.2: 299nt (89%)	35730623-35730325	Not detected
						4.3: 355nt (97%)	35728372-35728018	Not detected
						4.4: 557nt (93%)	35717824-35717268	Not detected
						4.5: 220nt (85%)	35689056-35688837	BQ550613 (Mm)
						4.6: 233nt (77%)	35687890-35687658	BQ550612 (Mm), AA759295 (Hs), BM198047 (Mm)
						4.7: 128nt (87%)	35685165-35685038	BU424434 (Gg)
5	25745	20799 (39%)	16104 (16%)	23729 (47%)	3879 (7%)	5.1: 229nt (87%)	35644763-35644535	Not detected
6	16464	14938 (52%)	14925 (26%)	15278 (62%)	7401 (18%)	6.1: 516nt (92%)	35633592-35633077	Not detected
						6.2: 382nt (89%)	35629634-35629253	AJ814883 (Bt)
						6.3: 203nt (79%)	35628771-35628569	Not detected

#[a]	Hs	Mm[a]	Rn[a]	Cf[a]	Gg[a]	Regions of High Identity[b]	Location[c]	EST/cDNA matches (species)
7	21293	18907 (47%)	18332 (46%)	20228 (64%)	8711 (16%)	7.1: 587nt (90%)	35625748-35625162	CA426059 (Hs), AW271794 (Hs), AI760204 (Hs), AI309016 (Hs), AA604325 (Hs)
8	15056	9973 (37%)	10585 (40%)	12815 (59%)	6010 (15%)	8.1: 450nt (92%)	35599223-35598774	Not detected
9	21726	18029 (46%)	16984 (22%)	19618 (57%)	6441 (11%)	9.1: 184nt (90%)	35579122-35578903	Not detected
						9.2: 260nt (81%)	35571476-35571191	Not detected
10	16333	16090 (39%)	13788 (17%)	15455 (62%)	4779 (12%)	10.1: 139nt (76%)	35557018-35556880	BX108539 (Hs), AW511271 (Hs), BU140466 (Gg)
						10.2: 439nt (91%)	35554061-35553623	Not detected
11	23373	22284 (45%)	20442 (25%)	22680 (57%)	8698 (16%)	11.1: 220nt (80%)	35548835-35548616	Not detected
						11.2: 184nt (77%)	35547372-35547189	Not detected
						11.3: 444nt (85%)	35546687-35546244	Not detected
						11.4: 199nt (90%)	35545061-35544863	DA031869 (Hs)
12	5488	3480 (26%)	3427 (19%)	5367 (60%)	1483 (15%)	12.1: 222nt (87%)	35524668-35524447	BX482879 (Hs)

Table Notes:

a - intron number

b - percent identity (in brackets) compared to human sequence

c - Conserved regions are numbered sequentially from the 5' end of the intron. The nucleotide lengths of the conserved regions refer to the human sequence, and the overall identity (between all five species) is listed in brackets.

d - locations are mapped to the human contig GenBank Accession Number **NT_008470**

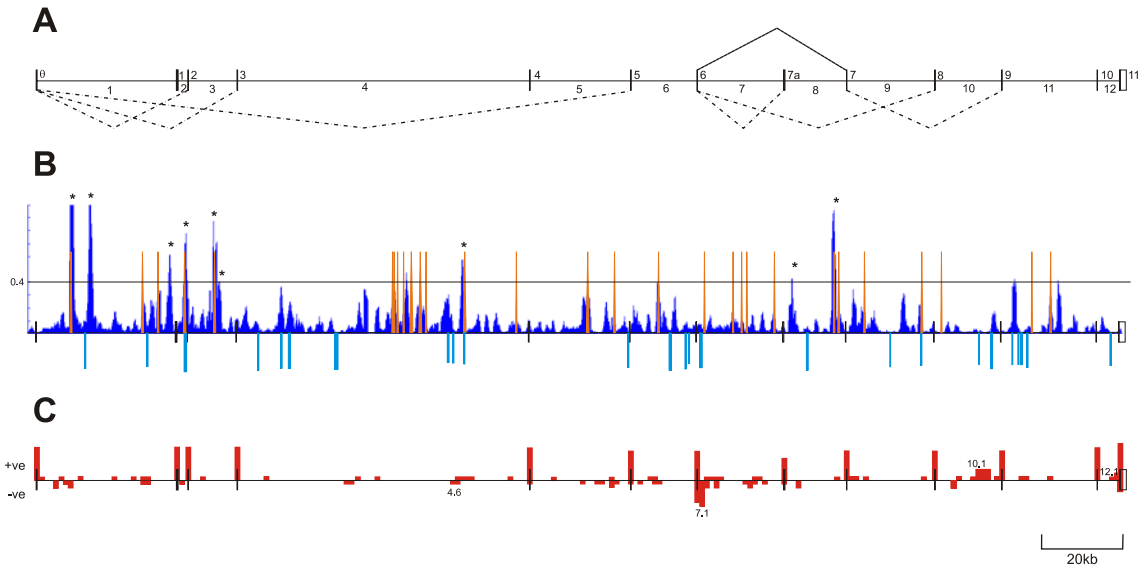


Figure 3.3.3. The human *Sin1* genomic structure, mapped with: (A) constitutive splicing (above the gene) and alternative splicing (below the gene); (B) conserved elements (below) and the output of MAR-Wiz (above - graph) and SMARtest (above - spikes). MAR-Wiz peaks called as S/MARs based on a threshold of 0.4 are marked with an asterisk. Conserved elements that overlap with predicted S/MARs are indicated; (C) human EST evidence. Conserved elements overlapping with EST evidence from the sense strand (above) and anti-sense strand (below) are indicated. The height of the bars indicates the relative quantity of ESTs.

(S/MARs) were predicted using MAR-Wiz (Singh *et al.*, 1997) and SMARtest (Frisch *et al.*, 2002). SMARtest predicted 29 potential sites, MAR-Wiz identified 9 potential sites, and there were 5 common predictions between the two programs. By far the largest peak identified by Mar-Wiz sits closest to the 5' end of the *Sin1* gene, where it is presumably required for transcription initiation. Three of the conserved intronic elements identified above (2.1, 4.7, and 9.2) are located at predicted S/MARs (Figure 3.3.3B), and it is reasonable to speculate that these elements play a structural role in *Sin1* gene regulation.

3.3.7 Potential Regulatory RNAs within the human Sin1 gene

As it is possible that the conserved elements identified in the *Sin1* introns represent undiscovered exons, the NR and EST databases (NCBI) were searched to determine whether any of the conserved intronic elements were transcribed. Ten elements (1.1, 1.2, 4.5, 4.6, 4.7, 6.2, 7.1, 10.1, 11.4, and 12.1) were identified that had EST or cDNA evidence from at least one of the species studied to support their transcription (Table 3.3.1). Each EST containing a match to a conserved element was examined to determine whether it might be spliced to a known exon, indicating the presence of a previously undetected exon. No such splicing events were detected. BlastX (NCBI) was then used to determine whether there was any similarity between the 25 conserved elements and known proteins. Only element 7.1 returned a hit with any significance (1e-06), and this was to a hypothetical protein from *Tetraodon nigroviridis* (spotted green puffer fish). This protein is translated from a predicted gene model, and is not supported by Tetraodon mRNA. Further, the genomic region containing this model also contains the likely Tetraodon *Sin1* orthologue, suggesting that element 7.1 is also conserved in this genome. The relative level of EST expression for both strands of the entire *Sin1* region (derived from NCBI's ModelMaker) can be seen in Figure 3.3.3C. It is interesting to note that kidney is the tissue of origin for 3 of 5 of the ESTs overlapping with element 7.1. This element lies between exon 6 and the alternately spliced exon 7a which encodes the *Sin1*α isoform, where EST evidence is largely from kidneys. It is possible then, that this element may play a role in regulating kidney specific patterns of

Sin1 splicing. Together, these data suggest that while some of the conserved intronic elements are transcribed, it is unlikely that these are also translated and they may therefore represent regulatory RNA sequences.

3.3.8 Sin1 genes contain a conserved CpG island with promoter activity

CpG islands are associated with the 5' end of most"housekeeping" genes, as well as many regulated genes (Antequera and Bird, 1993). Cpgplot (Gardiner-Garden and Frommer, 1987; Rice *et al.*, 2000) was used to analyse the 275 kb of genomic sequence encompassing the human *Sin1* gene, which identified the presence of seven CpG islands. Only one of these islands is conserved in the mouse, rat and chicken *Sin1* genomic sequences. While the length of these islands differs between species, in all cases it encompasses the transcription start site (TSS) as determined by EST evidence (Table 3.3.2). There is only one CpG island present in the canine AAEX01025409 sequence, and it encompasses the region that is predicted to contain exon θ and the TSS.

MatInspector (Quandt *et al.*, 1995) was used to search for conserved transcription factor binding sites within the CpG islands, which revealed a module of binding sites conserved between all four species. A similar module was found within the canine CpG island identified above. Figure 3.3.4 depicts the structure of this promoter model in all species. In this model, the *Sin1* promoters are TATA-less, but contain a binding site for the ubiquitous transcription factor Stimulating Protein 1 (SP1). This is typical of many genes, allowing basal transcription across a wide range of tissues

Table 3.3.2. The conserved CpG islands from orthologous *Sin1* genes. For each island, the length, C+G content, and ratio of observed to expected CpG dinucleotides are detailed, as well as their location in the sequences used.

	Length (bp)	C+G (%)	Obs/Exp ratio	Location (nt)	Genbank Accession
Human	454	72.47	0.77	35790886-35790433	**NT_008470**
Mouse	512	71.88	0.82	11835310-11835822	**NT_039206**
Rat	388	69.59	0.8	3747668-3748055	**NW_047652**
Chicken	694	74.06	1.08	725672-726365	**NW_060627**
Dog	647	70.48	0.95	19493-20139	**AAEX01025409**

(Samson and Wong, 2002). Several studies have previously found that *Sin1* is widely, if not ubiquitously, transcribed at low levels (Loewith *et al.*, 2002; Schroder *et al.*, 2004; Cheng *et al.*, 2005), and SP1 may be important in regulating this transcription. Three other conserved binding sites, HES1 (Drosophila hairy and enhancer of split homologue 1), Elk-1, and members of the Early Growth Response Family (EGRF) were also identified. HES1 is a transcriptional repressor, and probably acts as a negative regulator of myogenesis (Sasai *et al.*, 1992). Previous experiments have demonstrated that *Sin1* is highly expressed in skeletal muscle and the heart (Loewith *et al.*, 2002; Schroder *et al.*, 2004; Cheng *et al.*, 2005), and HES1 may therefore be important in muscle specific *Sin1* expression. The EGR family of transcription factors are induced by a wide range of extracellular stimuli, and depending on the cellular context, can be either stimulators or

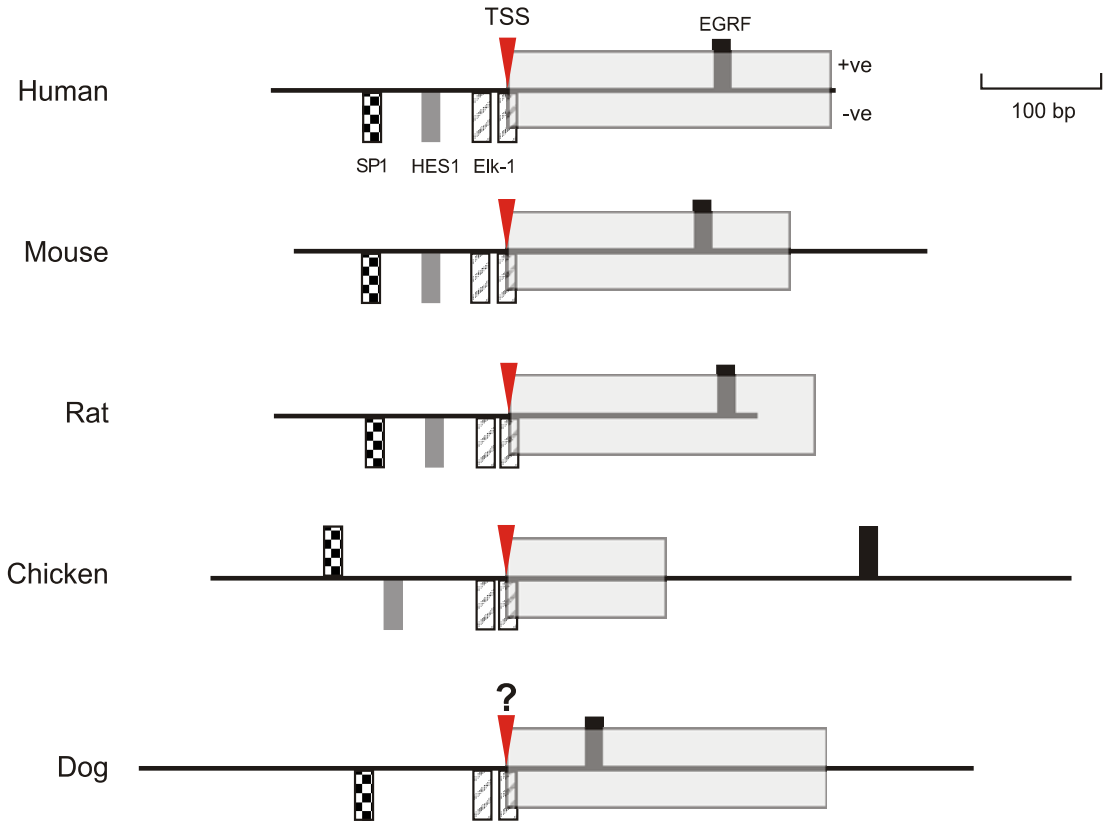


Figure 3.3.4 Putative *Sin1* promoter module. The conserved CpG islands in each species (black line) are labelled with the conserved transcription factors (rectangles) on the sense strand (above) or anti-sense strand (below). The transcription start site (TSS), or putative transcription start site (?) is shown. The grey boxes represent exon θ.

inhibitors of expression. They are generally immediate early response genes that may play a role in the regulation of cellular proliferation (Thiel and Cibelli, 2002). Like the EGR family, Elk-1 is phosphorylated by Mitogen Activated Protein Kinases (MAPKs) in response to extracellular stimuli (Shaw and Saxton, 2003), and Elk-1 binding domains are common in the promoters of immediate early response genes (Li *et al.*, 2000).

There were five additional transcription factor binding motifs identified as common to all five species, where the relative positions were not conserved. These were MAZ (Myc-associated zinc finger protein), c-Rel, OLF (Olfactory neuron-specific factor), and two members of the vertebrate Zinc finger binding protein family, ZBP89 and CPBP (Core promoter binding protein). The entire promoter module is suggestive of a gene responsive to external stimuli.

The prediction of promoter activity within the *Sin1* CpG island was confirmed by luciferase assays of genomic *Sin1* constructs. The *Sin1* CpG island was found to have strong promoter activity in HEK293 cells (Figure 3.3.5), but this is likely to be repressed in a genomic context, as the luciferase signal of longer constructs was found to be substantially lower. Our laboratory is currently investigating the contribution of each of the conserved transcription factor binding sites to the observed pattern of promoter activity.

3.3.9 Genes potentially co-regulated with Sin1

Two approaches were used to identify genes that might be co-regulated with *Sin1*. First, genes containing the *Sin1* promoter module near to their TSS were identified using the ModelInspector function of Genomatix. Additionally, genes whose expression was found to be positively correlated with *Sin1* in multiple large scale gene expression studies were identified through Tmm (Lee *et al.*, 2004). Figure 3.3.6 diagrammatically shows the combined results of these two searches. In total, 6 genes were found to be

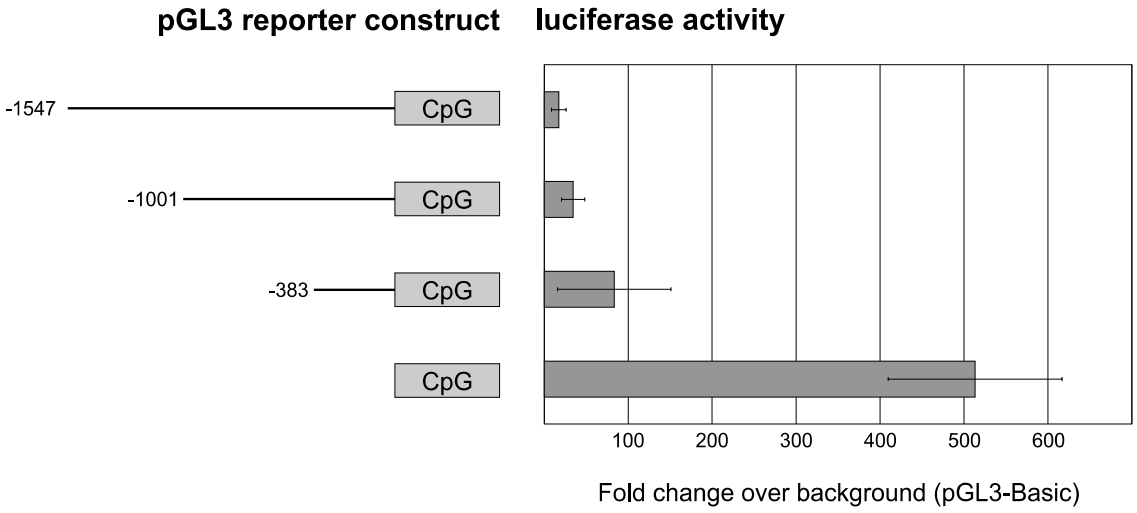


Figure 3.3.5. The *Sin1* CpG island displays strong promoter activity. pGL3 derived reporter constructs used in this assay are shown in the left panel. In this case, the +1 position is considered to be the start of the human CpG island. In the right panel, the luciferase activity of each of the constructs is shown, indicating that the CpG island alone had substantially higher promoter activity than the remaining constructs. The averages and standard errors from three independent experiments are shown.

common between the two sets (detailed in Table 3.3.3). Four of these genes (GNB1, YBX1, WWP1, and MSRA) are directly linked to the mammalian host response to pathogen invasion. GNB1 mRNA is upregulated in response various pathogenic and non-pathogenic Mycobacteria (Greenwell-Wild *et al.*, 2002; McGarvey *et al.*, 2004), and also appears to be induced in response to IFN-α (Yan *et al.*, 2004). YBX1 (otherwise known as NSEP1 and YB1) also appears to be induced in the host response to non-pathogenic Mycobacteria (McGarvey *et al.*, 2004), and both YBX1 and WWP1

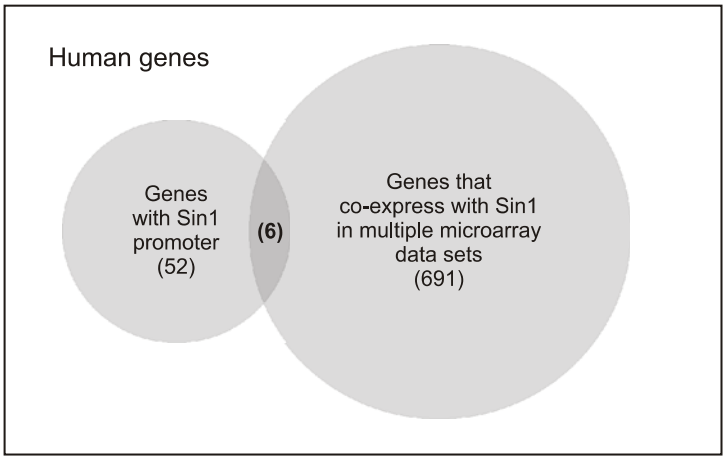


Figure 3.3.6. Venn diagram of genes that may be co-regulated with *Sin1*.

Table 3.3.3. Genes that may be co-regulated with *Sin1*. Genes found to be both co-expressed with *Sin1* and to possess the same promoter module found in *Sin1* are listed with their functional annotation derived from Gene Ontology (GO) terms.

Gene Symbol	Description	Functional annotation (from GO terms)
GNB1	guanine nucleotide binding protein (G protein), beta polypeptide 1	signal transduction, G-protein coupled receptor protein signaling pathway
MSRA	methionine sulfoxide reductase A (aka Peptide methionine sulfoxide reductase (EC 1.8.4.6))	methionine metabolism, response to oxidative stress, protein-methionine-S-oxide reductase activity
YBX1	Y-box transcription factor 1	DNA binding, RNA binding, transcription factor activity
PCSK6	Proprotein convertase subtilisin/kexin type 6 precursor (EC 3.4.21.-)	subtilase activity, calcium ion binding, proteolysis and peptidolysis, cell-cell signaling
PRDM15	PR domain containing 15	transcription factor, regulation of transcription
WWP1	WW domain containing E3 ubiquitin protein ligase 1	ubiquitin-protein ligase activity, signal transduction, central nervous system development

are induced in reponse to viral pathogens (Holm *et al.*, 2002; Challacombe *et al.*, 2004; Yim *et al.*, 2004; Martin-Serrano *et al.*, 2005). MSRA is involved in the reactive oxygen intermediates (ROI) response to microbial pathogens, which forms the basis of cellular resistance to microbes (Nathan and Shiloh, 2000). Of the remaining genes, PRDM15 has not been studied in detail, and no direct link could be found between PCSK6 (otherwise known as PACE4 or SPC4) and host response to invasion. Interestingly, there was an indirect link detected for PCSK6, in that other cellular subtilisins are involved in the proteolytic processing of viral proteins necessary for the entry of viruses into the mammalian host cell (Nakayama, 1997). Collectively, these results imply a role for Sin1 in the cellular response to pathogens.

3.4 DISCUSSION

The power of comparative genomic analysis to identify regulatory elements in the non-coding regions of genes is sensitive to the conservation of gene expression patterns. It was therefore important to establish whether orthologous *Sin1* genes were structurally similar, and whether the *Sin1* splice variants observed in humans were likely to be conserved in other species. Overall, the genomic structure and local synteny (with the neighbouring genes PBX3 and GAPVD1) of *Sin1* orthologues is highly conserved. Significantly, neither PBX3 nor GAPVD1 (which are transcribed from the opposite strand) were identified as co-regulated with *Sin1*, confirming that this gene is regulated independently of the broader genomic locus. There are several splice variants conserved between the human, mouse, and sheep transcriptomes, which are probably present in all mammals. This study identified a 10nt motif that is completely conserved between human, mouse, rat, dog, and chicken *Sin1* genes. With a CpG island, and an intron length within the range documented in this study, this sequence should provide a way to locate the non-coding exon θ and potential transcription start sites of vertebrate *Sin1* genes lacking EST data.

Most eukaryotic mRNAs have short 5'-UTRs (20–100 nt) with no upstream ORFs, allowing efficient translation. Only five to ten percent of vertebrate transcripts have long and highly structured 5' UTRs, and these usually encode regulatory proteins such as transcription factors and proteins involved in signal transduction (Kozak, 1992). The *Sin1* 5'-UTR is at least 330nt long with two upstream ORFs, and is highly

structured. Secondary structures with a stability of around -50 kcal/mol can inhibit translation by 85-95% (Kozak, 1986), so the *Sin1* 5'-UTRs from any of the species studied here could be potent inhibitors of translation, providing stringent post-transcriptional regulation. The presence of an IRES pattern in the 5'-UTR is significant in this context, as it may allow the ribosome to bypass the upstream ORFs and secondary structures under certain conditions. Many cellular stresses lead to a general inhibition of protein synthesis, however the presence of IRES patterns may allow certain mRNAs involved in stress responses to be actively translated (Holcik *et al.*, 2000; Nevins *et al.*, 2003). Considering the interactions with important signalling molecules such as Ras, MEKK2, and JNK, theorizing that *Sin1* may be regulated in response to cellular stresses is reasonable.

Intronic elements conserved between human, mouse, rat, chicken, and dog represent 4.89% of the non-coding, non-repetitive human genomic *Sin1* sequence. This number is higher than the 3.5% of the human genome predicted to be under negative selection pressure (where mutations in the sequence reduce the organism's fitness) (Smith *et al.*, 2004), and suggests that *Sin1* transcription may be more tightly regulated than the average human gene. This is consistent with a role in a signal transduction pathway, where all components must have appropriate expression levels to ensure both sufficient and specific amplification of signals.

This study has identified and isolated a conserved CpG island that exhibits strong promoter activity in HEK293 cells, and has also identified a number of conserved

intronic regions. Both may contain regulatory *cis*-elements that direct the transcription, splicing, and post-transcriptional regulation of *Sin1* mRNA. The putative transcription factor binding sites identified here - SP1, HES1, Elk1, and members of the EGR family - can now be tested for their role in *Sin1* regulation. Similarly, the role of the conserved intronic regions in *Sin1* regulation can also be tested, with the genomic context described here providing a more directed approach to these studies. Conserved elements overlapping with predicted scaffold/matrix attachment regions (S/MARs) are likely to have a structural role, whereas those that appear to be transcribed may function as regulatory RNAs. Other elements may positively or negatively regulate the activity of the *Sin1* promoter, and our laboratory is currently working toward elucidating the role of these elements.

By combining two different bioinformatics approaches, six genes that are potentially co-regulated with *Sin1* have been identified. All have been found to co-express with *Sin1* in large scale gene expression studies, and all contain the conserved promoter module identified by comparative genomic analysis of *Sin1* genes. The common link between these genes may be in the host response to pathogens. It is worth remembering that Sin1 has been shown to interact with IFNAR2 (a subunit of the type I IFN signalling receptor; Wang and Roberts, 2004), as type I IFNs play an important protective role against pathogen infection. The co-regulation of *Sin1* with genes involved in this pathway suggests that Sin1 may be involved in modulating the response to the IFN ligands.

CHAPTER FOUR

An investigation of the biochemical functions of Sin1

4.1 INTRODUCTION

The previous two chapters have examined the structure and localization of human Sin1 proteins, and the genomic regulation of the human *Sin1* gene. Another useful approach for the determination of the biological significance of a novel protein is to examine its biochemical functions, including protein-protein interactions, changes in protein expression levels, and physiological effects associated with the over- and under-expression of proteins in cells.

Many interacting partners of Sin1 and Sin1 orthologues have already been described (reviewed in Chapter One), and physiological changes associated with an absence of expressed Sin1 have been reported in lower eukaryotes. These changes include delays in entering mitosis (Wilkinson *et al.*, 1999), defects in the actin cytoskeleton (Loewith *et al.*, 2002), and defects in chemotaxis (Lee *et al.*, 1999). Although several studies have over-expressed Sin1 in mammalian cells (Wang and Roberts, 2004; Cheng *et al.*, 2005; Schroder *et al.*, 2005), none have reported phenotypic differences in these cells as compared to wild type cells. No changes in protein expression have been reported, although Sin1 proteins in *S. pombe* were found to be hyper-phosphorylated in response to multiple cellular stresses. Additionally *S. pombe* cells lacking *Sin1* were unable to survive under conditions of osmotic stress or heat stress, but could proliferate when under oxidative stress (Wilkinson *et al.*, 1999), suggesting that Sin1 is required for proper functioning of some, but not all stress pathways.

This chapter reports a major physiological effect of Sin1 over-expression, the response of endogenous Sin1 proteins to various cellular stresses, and an initial investigation into the binding partner(s) of the Conserved Region In the Middle (CRIM). Additionally, the presence of a coiled-coil domain in the N-terminal region of Sin1 (Chapter Two) prompted an investigation into the ability of Sin1 proteins to dimerize *in vivo*. The consequences of these results are discussed.

4.2 MATERIALS AND METHODS

4.2.1 Tissue culture, transfections and microscopy

Tissue culture, transfections, and microscopy were carried out as per sections 2.2.3 and 2.2.4. To select stable cell lines, cells were passaged 48-72 h after transfection, and split 1 in 10 into growth media containing an appropriate concentration of G418 (Sigma). To stain filamentous actin, TRITC-Phalloidin (Sigma) was used at 50 μg/mL. The anti-Sam68 (7-1) antibody (Santa Cruz) was used at 4 μg/mL.

4.2.2 Differential dye uptake assay

The morphological changes indicative of apoptosis were evaluated using a differential dye uptake assay (Belloc *et al.*, 1994). Transiently transfected cells were incubated for 15 min with 5 μg/mL each of Hoechst 33342 and propidium iodide, and analysed immediately under an Olympus BX50 fluorescence microscope. Cells that showed EGFP fluorescence were scored as either viable (blue intact nuclei), necrotic (pink intact nuclei), early apoptotic (blue condensed or fragmented nuclei), or late apoptotic (pink condensed or fragmented nuclei). At least 200 cells from three independent experiments were counted for each construct.

4.2.3 Cellular stress assays

Each assay used 1 X 10^6 healthy, mid-log phase cells. Where DG75-Sin1 cells were used (Schroder *et al*. 2005), expression of FLAG-Sin1 was induced by incubating

the cells with 5mM of cadmium chloride overnight prior to treatment. To examine the effect of UV irradiation, cells were washed twice with PBS and then exposed to 80 J/m^2 of UV-C radiation. Growth medium was replaced, and the cells incubated under normal conditions for various times as described in the text. To examine the response to H_2O_2 stress, cells were washed twice with PBS and resuspended in growth media containing 0 μM, 5 μM, 10 μM, 100 μM, 500 μM, 1 mM, 2.5 mM, or 5 mM H_2O_2, and incubated under normal conditions. To examine the effect of hyperosmotic stress, cells were washed twice with PBS and resuspended in growth media containing 200 mM of either sorbitol or sucrose, and incubated under normal conditions. Similarly, to examine hypoosmotic stress, cells were washed twice with PBS and resuspended in 10% FCS and either 30% or 70% RPMI in H_2O, and incubated under normal conditions. To examine the effect of temperature stress, cells were washed twice in PBS and resuspended in growth media containing 25 mM HEPES. Cells were then incubated at 44°C or 20°C for the times indicated in the text. Finally, to examine the effects of serum on Sin1 protein levels, cells were washed twice in PBS and resuspended in either 20% FCS and RPMI, 0.1% FCS and RPMI, or 0.1% FCS and PBS, and incubated under normal conditions. To determine the requirement of *de novo* RNA or protein synthesis for cellular stress, stock solutions of actinomycin D (50 mg/mL) and cycloheximide (5 mg/mL) were prepared in DMSO. Cells were pre-treated for 5 h with a final concentration of 50 μg/mL actinomycin D, 5 μg/mL cycloheximide, or 1 μL/mL DMSO alone. At harvesting, cells were washed twice with PBS and stored at -80°C until

required for analysis by SDS-PAGE and immunoblot (as per section 2.2.6). Protein bands were quantified using Scion Image (Scion Corporation).

4.2.4 GST pull-down assays

E. Coli Sure cells (Stratagene) expressing GST fusion proteins were resuspended in ice cold lysis buffer (20 mM Tris-Cl pH 8.0, 200 mM NaCl, 1.0% (v/v) Triton X-100, 0.1% (w/v) SDS) and lysed by French press. Cell debris was cleared by centrifugation at 14,000 g for 10 min at 4^0C. Fusion proteins were purified using glutathione agarose beads (Sigma), according to the manufacturer's instructions. Purified protein (10 μg) was used to pull down EGFP fusion proteins from 2 X 10^6 transiently transfected HEK293 cells, harvested after 24 h. Cells were washed with PBS before lysis in 200 μL of ice cold RIPA buffer (0.1% (w/v) SDS; 1% (v/v) Triton X-100; 0.5% sodium deoxycholate; 150 mM NaCl; 50 mM Tris-Cl, pH 8.0; 1 mM Na_3VO_4; 1 mM DTT; and a protease inhibitor cocktail (Roche)) on ice. Samples were cleared by centrifugation at 10,000 g for 10 min at 4^0C.

4.2.5 Immunoprecipitations

Each immunoprecipitation was performed at 4^0C using 5 X 10^6 transiently transfected HEK293 cells lysed in 500 μL of RIPA buffer, and cleared by centrifugation. Anti-GFP (2 μg of N-terminal rabbit polyclonal antibody; Sigma) was added to the cell lysates and incubated for 2 h with gentle agitation. Pre-washed EZview red protein G

affinity gel (20 μL bed volume; Sigma) was added to the sample, and the incubation was continued overnight. The beads were washed extensively in RIPA buffer, and analysed by SDS-PAGE and immunoblot as described in section 2.2.6.

4.3 RESULTS

4.3.1 Over-expression of Sin1 induces apoptosis

Despite repeated attempts with Cos7, HeLa, and DG75 cells, stable cell lines constitutively over-expressing any of the EGFP tagged Sin1 isoforms could not be generated under conditions that successfully selected the EGFP control, even when the initial efficiency of transfection was over 70%. In ME180 cells, selection for constitutive EGFP-Sin1 over-expression resulted in cells expressing very low amounts of fluorescent protein (data not shown). While these cells were viable, they did not proliferate, were much larger than the EGFP control cells, were often multinuclear, and did not generally survive more than six to eight weeks from the time of transfection (Fig. 4.3.1A). This phenotype was also observed in JAM cells (Figure 4.3.1C), and is consistent with cells experiencing mitotic catastrophe (Miranda *et al.*, 1996). Interestingly, the pattern of Sin1 localization in these cells is different to that observed for phenotypically normal cells. Normally, EGFP-Sin1 is strongly associated with the plasma membrane (Chapter Two: Figure 2.3.4); the fluorescence in these cells, however, is predominantly peri-nuclear (Figure 4.3.1B).

In transiently transfected cells, proliferation of cells over-expressing EGFP-Sin1 could not be detected, with the fluorescent cells often rounded and detached from the surface of the flask. This suggests that over-expression of Sin1 might induce cellular death. To investigate the possibility that transient-Sin1 over-expression results in apoptosis, HEK293 cells were transfected with EGFP-labelled Sin1 isoforms (described

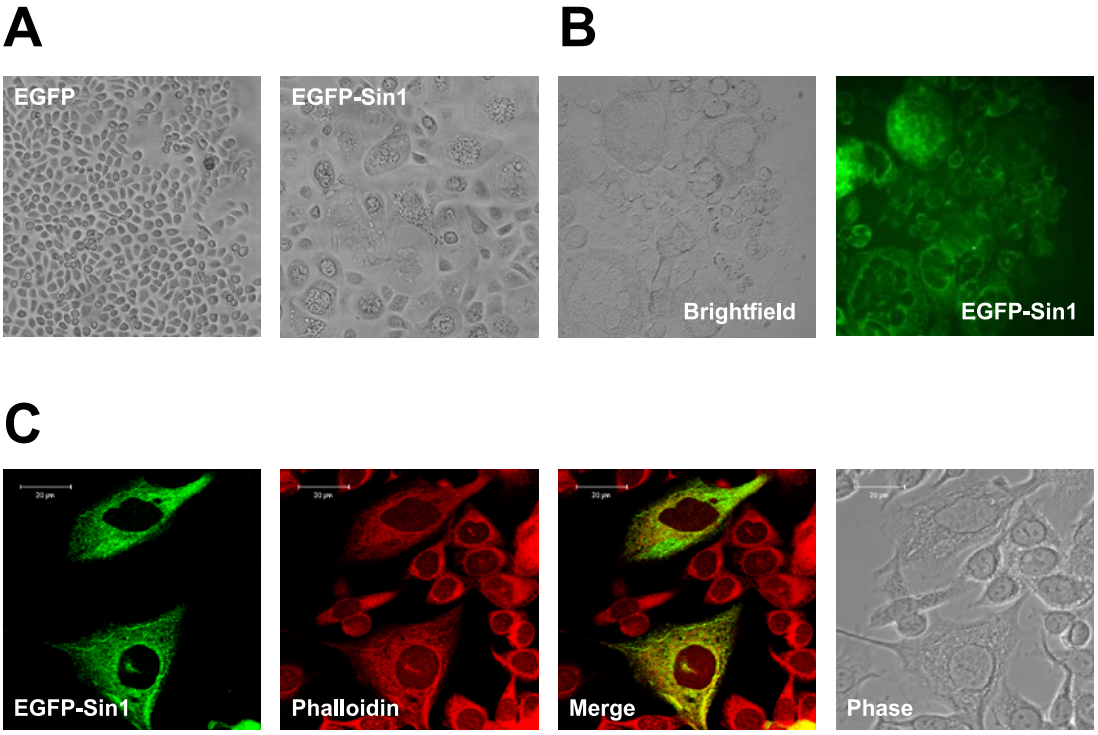


Figure 4.3.1. Constitutive over-expression of EGFP-Sin1 causes an enlarged, and often multinuclear phenotype in ME180 and JAM cell lines. (A) ME180 cells were transfected with either EGFP or EGFP-Sin1, and stable cell lines were selected using geneticin (400 µg/mL). Selection of cells constitutively expressing EGFP-Sin1 results in cells with visual characteristics of mitotic death. (B) Fluorescence microscopy of cells constitutively expressing EGFP-Sin1 reveals a non typical localization pattern for this plasmid. (C) Confocal microscopy of JAM cells transiently transfected with EGFP-Sin1 and co-stained with Phalloidin-TRITC. These cells also appear enlarged and multinuclear after 48 h.

in section 2.2.2). After 24 h, levels of apoptosis were determined by a differential dye uptake assay using both Hoechst 33342 (membrane permeable) and propidium iodide (membrane impermeable). Transient expression of most Sin1 C-terminally tagged isoforms could significantly induce apoptosis in HEK293 cells (Figure 4.3.2A).

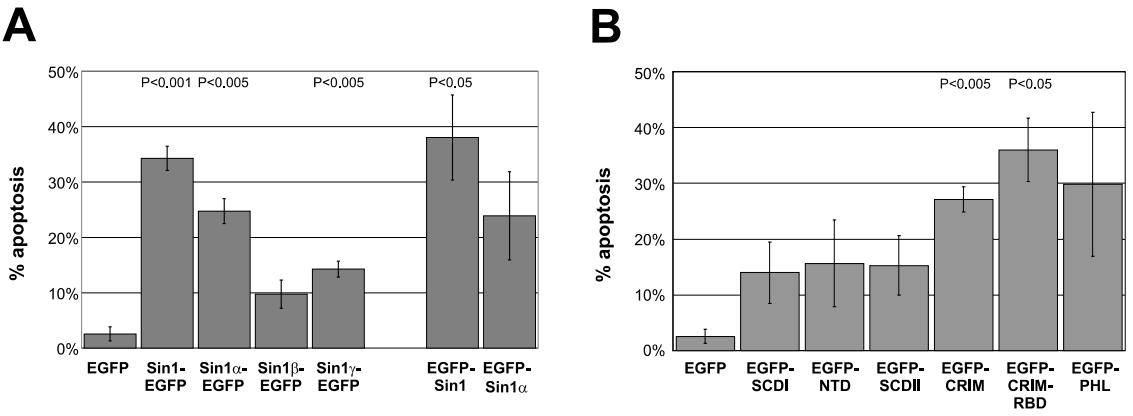


Figure 4.3.2. Over-expression of Sin1 induces apoptosis. (A) The percentage of HEK293 cells transiently expressing EGFP tagged Sin1 isoforms that are undergoing apoptosis. The average of three independent experiments is shown with standard errors. The statistical significance of the difference between fusion proteins and EGFP alone is listed above the bars (P value; student's t-test). C-terminally tagged isoforms are grouped to the left of the graph, whereas the two N-terminally tagged isoforms are grouped to the right. (B) The percentage of HEK293 cells, transiently expressing EGFP tagged Sin1 partial constructs, that are undergoing apoptosis.

Although Sin1β appeared to also increase basal apoptosis levels, this did not reach a statistically significant level. Interestingly, while the N-terminally tagged Sin1 isoforms could also induce apoptosis (Figure 4.3.2A), the standard errors of these experiment were much larger than the errors found for C-terminally tagged isoforms. As N-terminally tagged Sin1 isoforms were inhibited in their ability to enter the nucleus, nuclear localization might be required to induce apoptosis. The tolerance of various cell lines for Sin1 over-expression does not appear to be identical. For example, transiently transfected Cos7 cells expressed Sin1-EGFP well, and fluorescence was still detectable (although low) after 72 h, whereas only cells in late apoptosis had Sin1-EGFP

fluorescence 24 h post transfection in the SHSY5Y cell line. To determine which domain of Sin1 was primarily responsible for inducing apoptosis, differential dye uptake assays were performed on HEK293 cells transiently over-expressing EGFP tagged deletion mutants of Sin1 (described in section 2.2.2). Constructs containing the CRIM were able to induce apoptosis (Figure 4.3.2B), suggesting that Sin1-induced apoptosis may be primarily mediated through this domain. Although the PH-like domain also had an affect on apoptosis, this was not statistically significant, as it resulted from widely distributed data (as opposed to a single outlier). This suggests that other factors are likely to contribute to the apoptosis induced from the over-expression of the Sin1 PH-like domain. Further experiments are required to elucidate the contribution of this domain to apoptosis induced by Sin1 over-expression.

4.3.2 Sin1 protein levels change in response to serum and hypoosmotic stress

It has been suggested that Sin1 proteins may function as scaffold or adaptor proteins in signal transduction pathways (Schroder *et al.*, 2005). As the efficiency of signal transduction pathways depends on the appropriate concentration of scaffold molecules (Burack and Shaw, 2000; Ferrell, 2000), regulating the expression of scaffold proteins provides a way to switch on or off signalling pathways. Thus, possible changes in the protein levels of endogenous Sin1 proteins following cellular stresses were examined.

HEK293 cells were subjected to a range of cellular stresses, including

hyperosmotic stress (sucrose and sorbitol), hypoosmotic stress (30% and 70% RPMI in H_2O), temperature stress (44°C and 20°C), oxidative stress (H_2O_2), UV irradiation, and changes in serum concentration. At defined time points following these stresses, total cell lysates were prepared and Sin1 proteins were detected by immunoblot, with the expression levels of the two major Sin1 isoforms (p75 and p70) quantified. No significant changes in Sin1 expression levels were detected following temperature stress, oxidative stress, or hyperosmotic stress (Figure 4.3.3 and data not shown). In contrast, hypoosmotic stress induced a rapid and significant increase of the Sin1 p70 isoform, whereas full length Sin1 (p75) was largely unchanged. Serum addition (20% FCS), but not serum deprivation (0.1% FCS), induced significant decreases in the levels of full length Sin1. The combination of serum and media deprivation (0.1% FCS in PBS) produced a rapid, significant, and sustained decrease in the levels of full length Sin1, but not the p70 isoform, suggesting that different Sin1 isoforms have different cellular roles.

To determine the effect of serum and media deprivation on Sin1 protein expression in other cell lines, Cos7 cells were examined for changes in endogenous Sin1 expression following treatment with 0.1% FCS in PBS. A stable DG75 cell line containing a $CdCl_2$ inducible FLAG-tagged Sin1 expression vector (DG75-Sin1; Schroder *et al.*, 2005) was induced to over-express Sin1 and treated with 0.1% FCS in PBS. Following these stresses, total cell lysates were prepared, Sin1 proteins were detected by immunoblot, and the levels of full length Sin1 isoforms were quantified.

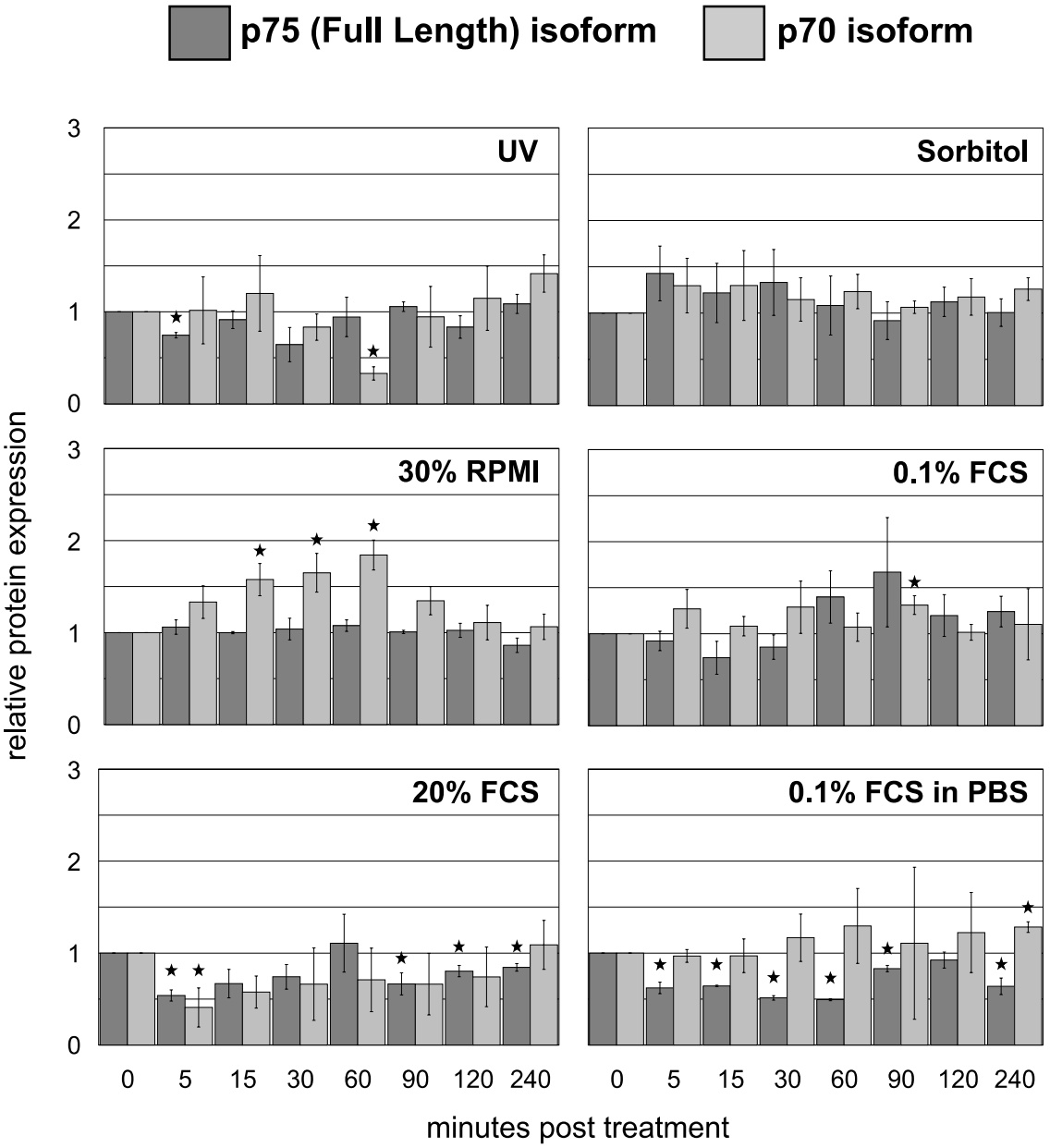


Figure 4.3.3. Endogenous Sin1 expression levels following stress. HEK293 cells were treated with various cellular stresses, and harvested at the times indicated. Total cell lysates were separated by SDS-PAGE, and Sin1 expression levels were measured by immunoblot and quantified. Expression levels were calculated as relative to the 0 min time point. Standard errors are shown for each time point in each treatment, and represent at least three independent experiments. Stars represent statistically significant differences between a given time point and time 0.

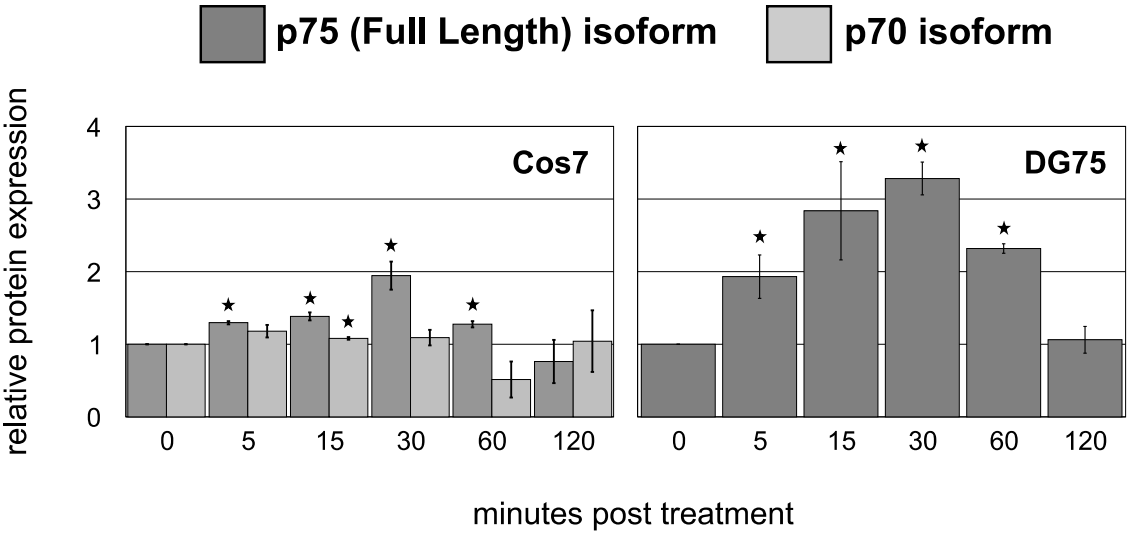


Figure 4.3.4. Response of Sin1 proteins to serum and media deprivation in Cos7 (endogenous Sin1) and DG75 (FLAG-Sin1) cell lines. Total cell lysates were separated by SDS-PAGE, and Sin1 expression levels were probed via immunoblot and quantified. Expression levels were calculated as relative to the 0 min time point. The average and standard errors of expression are shown for each time point in each treatment, and represent at least three independent experiments. Stars represent statistically significant differences between a given time point and time 0.

Interestingly, in both these cell lines full length Sin1 expression was found to rapidly and significantly increase in response to serum and media deprivation (Figure 4.3.4), contrasting with the decrease in expression of Sin1 following the same treatment in HEK293 cells (Figure 4.3.3). For the DG75-Sin1 cell line, immunoblots of cell lysates with an anti-Sin1 polyclonal antibody revealed a change in expression levels (Figure 4.3.4), whereas immunoblots of the same lysates with an anti-FLAG antibody did not. This data indicates that the increase in Sin1 protein levels comes from the increased expression of endogenous proteins, most likely to be regulated transcriptionally by

signals present in either the *Sin1* genomic DNA, or in the unspliced pre-mRNA. Alternatively, Sin1 protein levels may be determined at translation or post-translationally, by a mechanism that is inhibited by the fusion of the FLAG tag to the N-terminal of Sin1.

To determine whether the observed changes of full length Sin1 expression are controlled primarily at the level of transcription or translation, serum and media deprivation assays were carried out on HEK293 cells in the presence of actinomycin D (to inhibit transcription), or cycloheximide (to inhibit translation). HEK293 cells were pre-treated with either actinomycin D, cycloheximide, or DMSO alone as a control. Following this, cells were incubated in 0.1% FCS in PBS, and the levels of Sin1 proteins from total cell lysates were immunoblotted and quantified. The rapid decrease in Sin1 protein expression was again seen in the presence of DMSO alone, but was blocked by the use of either actinomycin D or cycloheximide (Figure 4.3.5), suggesting that *de novo* RNA and protein synthesis is required for the decrease in Sin1 protein levels in response to this stress. Together, these results suggest that Sin1 protein levels are regulated by cellular stresses, and that these responses vary in distinct cellular contexts. This study has also identified the first example of different cellular roles for the different Sin1 isoforms.

4.3.3 Sin1 proteins form dimers, mediated by the N-terminal domain

As Sin1 was predicted to contain a coiled-coil region within the N-terminal

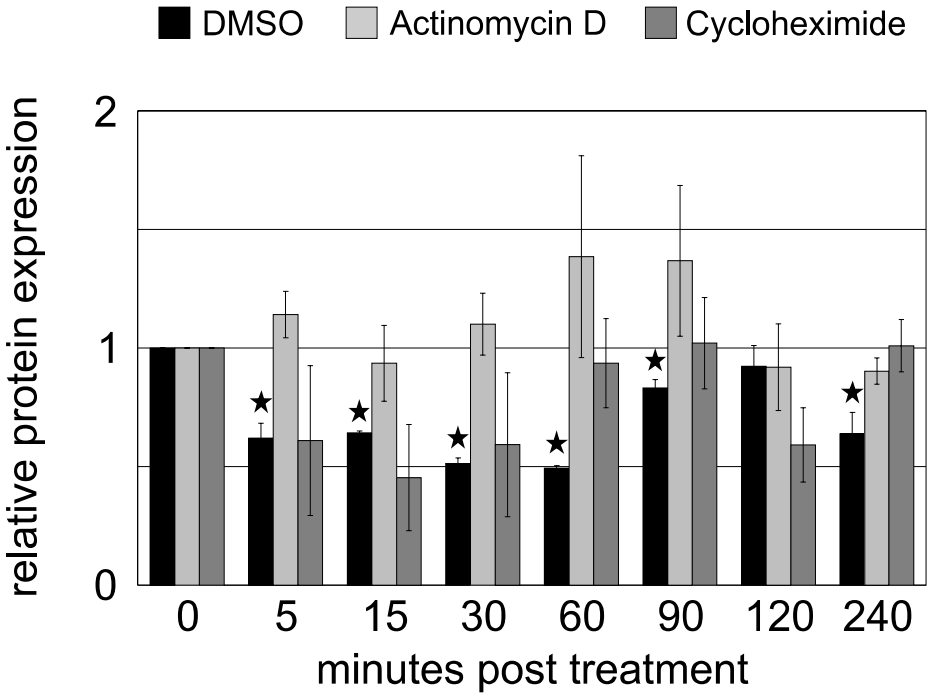


Figure 4.3.5. The Sin1 response to serum and media deprivation requires de novo RNA and protein synthesis. HEK293 cells were serum and media deprived for the times indicated in the presence of DMSO alone, actinomycin D (to inhibit transcription), or cycloheximide (to inhibit translation). The average and standard errors of expression are shown for each time point in each treatment, and represent at least three independent experiments. Stars represent statistically significant differences between a given time point and time 0.

domain (Chapter Two), and coiled-coil regions are often involved in protein-protein interactions (Burkhard *et al.*, 2001), the ability of Sin1 to bind to itself and form dimers was investigated. This was accomplished by "pull-downs" of Sin1 fusion proteins using bacterially expressed Sin1 fusion proteins. Affinity purified GST, GST-Sin1, or GST-NTD (N-terminal domain) attached to agarose beads were incubated with total cell lysates of HEK293 cells over-expressing EGFP-Sin1, EGFP-PHL (PH-like domain)

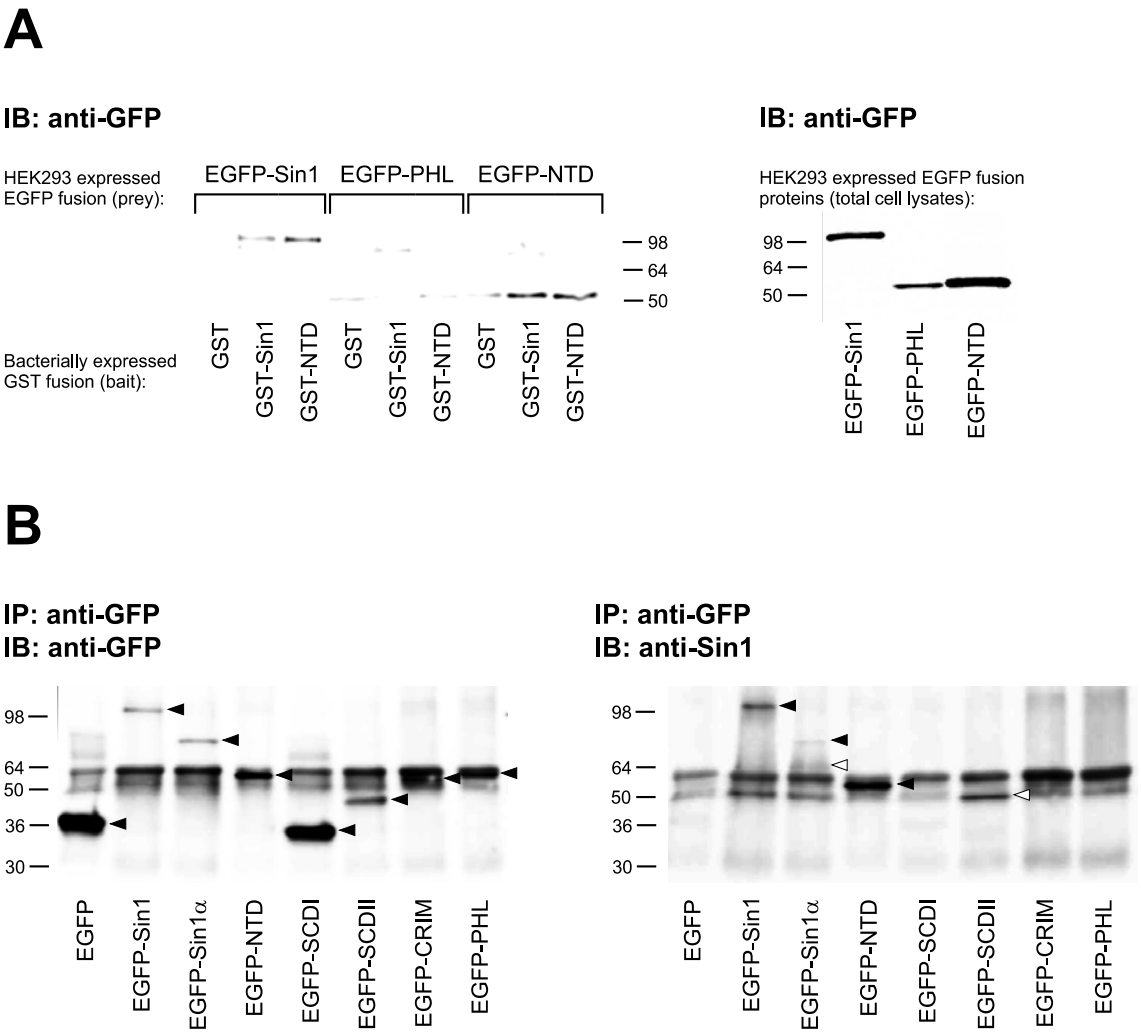


Figure 4.3.6. The N-terminal domain can interact with Sin1 proteins. (A) Bacterially expressed GST-Sin1, and GST-NTD are able to pull down over-expressed EGFP-Sin1 and EGFP-NTD from the cell lysate of transiently transfected HEK293 cells. (B) Immunoprecipitations of EGFP tagged Sin1 constructs can precipitate endogenous Sin1 proteins. The left panel shows the expected precipitation of EGFP tagged constructs with the anti-EGFP antibody, indicated by the black triangles. The right panel shows the same immunoprecipitations, this time immunoblotted with anti-Sin1 antibody. The EGFP constructs identified by the anti-Sin1 antibody are labelled with a black triangle. The specific endogenous Sin1 bands are labelled with a white triangle.

or EGFP-NTD. Bacterially expressed GST-Sin1 and the GST-NTD proteins, but not GST alone, were able to interact with EGFP-Sin1 and EGFP-NTD (Figure 4.3.6A). Neither were able to pull down EGFP-PHL fusion proteins, consistent with a self-association site in the N-terminal domain. To further narrow the region responsible for this interaction, immunoprecipitations were used to co-purify endogenous Sin1 proteins with EGFP fusion proteins. Total cell lysates of HEK293 cells over-expressing EGFP tagged deletion mutants of Sin1 were immunoprecipitated with an anti-GFP antibody. Both EGFP-Sin1α and EGFP-SCDII (containing the coiled-coil region) could precipitate endogenous Sin1 bands (Fig. 4.3.6B), confirming that the coiled-coil region could mediate Sin1 interaction *in vivo*.

4.3.4 Preliminary investigations into the binding partners of the CRIM

As the conserved region in the middle (CRIM) is the only protein domain of Sin1 that is present in every member of the Sin1 protein family, identifying the binding partners of the CRIM is likely to provide important clues to the function of Sin1. In order to identify potential binding partners of the CRIM, Cos7 cells were transiently transfected with plasmids encoding EGFP, EGFP-Sin1, or EGFP-CRIM. Cells were harvested after 24 hours, and complexes containing EGFP-CRIM or EGFP-Sin1 were immunoprecipitated from the total cell lysates using an anti-GFP antibody. The precipitated proteins were separated by SDS-PAGE, stained with Coomassie Blue, and compared to proteins precipitated by EGFP alone. Protein bands that were present in

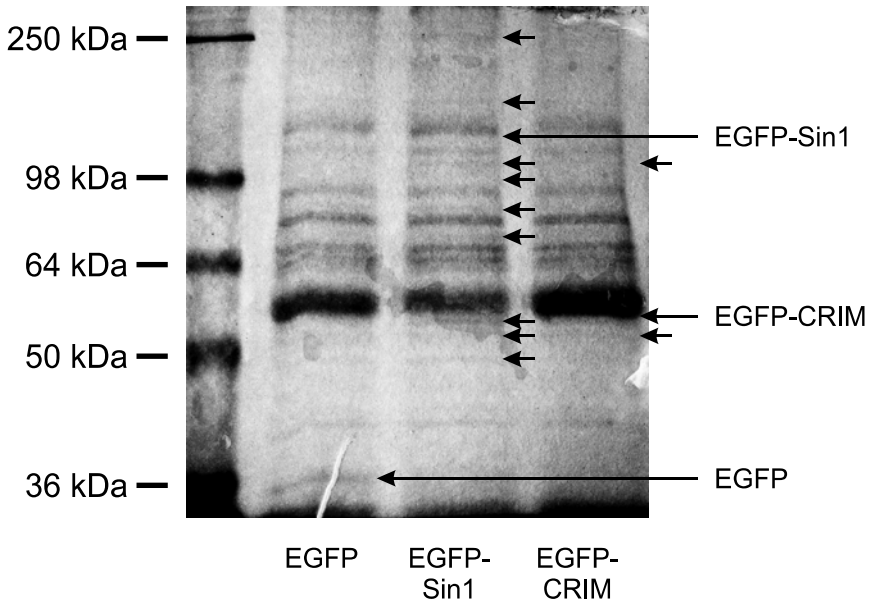


Figure 4.3.7. Immunoprecipitation of complexes containing EGFP-Sin1, or EGFP-CRIM. Cos7 cells were transiently transfected with EGFP, EGFP-Sin1, or EGFP-CRIM and harvested after 24 h. Immunoprecipitation was carried out using an anti-GFP antibody, and the precipitants were separated by SDS-PAGE and stained by coomassie. Protein bands appearing in the EGFP-Sin1 or EGFP-CRIM lanes, but not in the EGFP alone are identified by a black arrow. The target proteins of each precipitation are also indicated.

EGFP-Sin1 or EGFP-CRIM were excised from the gel, and were given to Ms. Shelley McRae (Griffith University) for trypsin digest and mass spectrophotometry fingerprint analysis. At the time of writing this thesis, the results of this analysis were not yet available. This aspect of the project, including the repetition of this experiment using the more sensitive silver stain, will now be followed up by KangWee Tay, a new PhD student in our laboratory.

A representative gel is shown in Figure 4.3.7. Two proteins bands appear

consistently in the CRIM lane that are not present in the EGFP lane. These bands, approximately 53 kDa and 102 kDa, were also present in the EGFP-Sin1 lane, suggesting that these proteins could be binding partners of the Sin1 CRIM. As discussed in Chapter One, orthologues of Sin1 in lower eukaryotes have been shown to bind to TOR (Loewith *et al.*, 2002; Wedaman *et al.*, 2003; Lee *et al.*, 2005), but the parallel interaction in mammalian cells has not yet been demonstrated (Loewith *et al.*, 2002). Interestingly, EGFP-Sin1 precipitated a band of approximately 270 kDa, which is close to the predicted size of mTOR (289 kDa).

As the CRIM when expressed alone localizes to the nucleus (Chapter Two), and there are no obvious nuclear localization signals in this domain, it is reasonable to speculate that one of the binding partners of the CRIM may be a nuclear protein. Of the known or potential protein partners of Sin1 (discussed in Chapter One), only Sam68 is a predominantly nuclear protein. Sam68 was identified as a candidate binding partner of Sin1 in a yeast two-hybrid screen (Schroder *et al*., unpublished data), and the potential interaction between these two proteins is currently being investigated by immunoprecipitation (Marion Buck, QIMR). To determine whether EGFP-CRIM and Sam68 co-localize, Cos7 cells were transiently transfected with EGFP-CRIM and analysed by IFA using an anti-Sam68 antibody. Despite both proteins localizing to the nucleus, Sam68 and EGFP-CRIM did not co-localize in healthy cells (Figure 4.3.8). Some co-localization was apparent in cells with signs of early apoptosis, but this was not extensive.

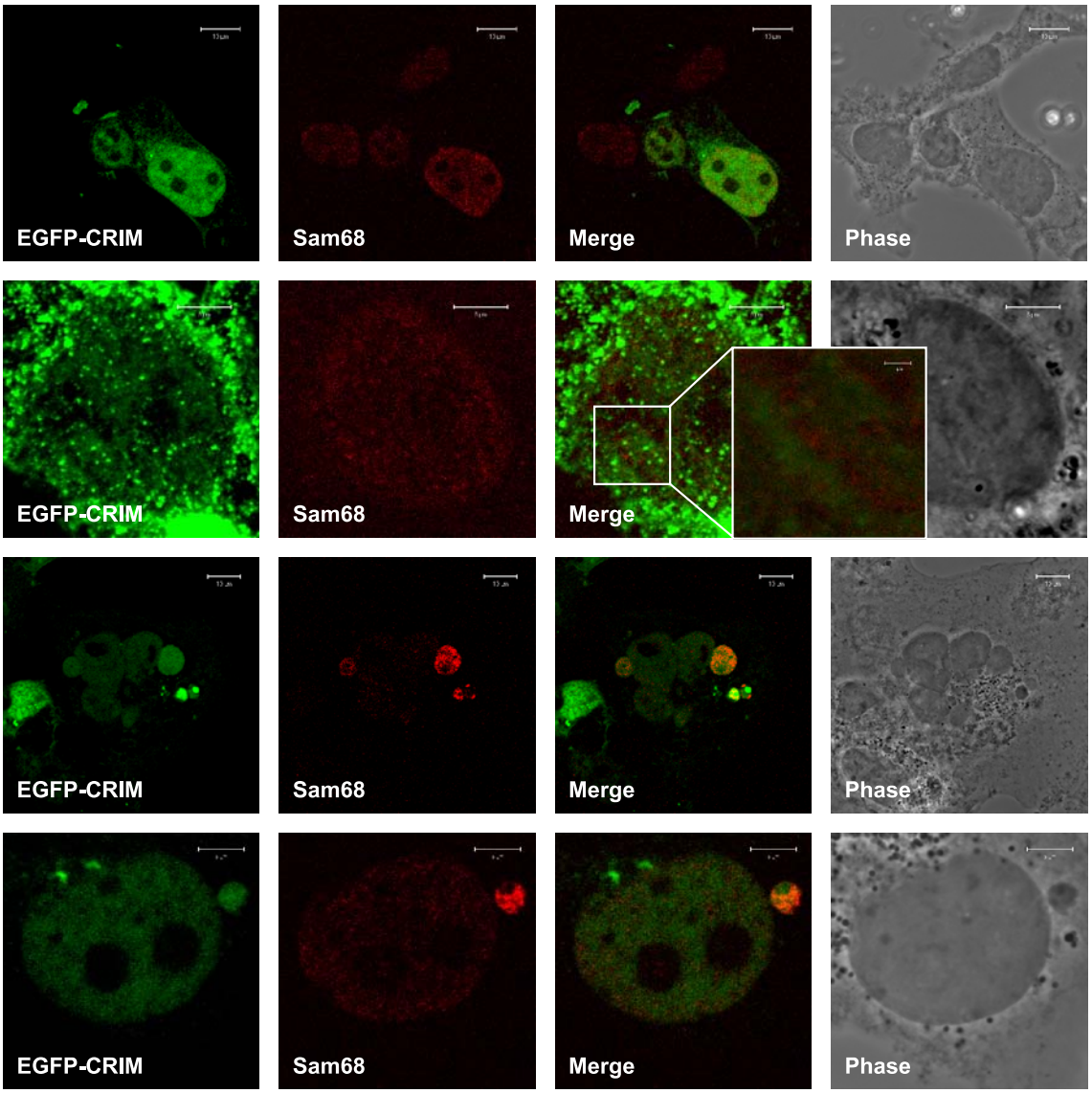


Figure 4.3.8. Endogenous Sam68 and EGFP-CRIM do not co-localize in healthy cells. Cos7 cells were transiently transfected with EGFP-CRIM, harvested after 24 h, and co-stained for endogenous Sam68. Cells with normal nuclei (top two panels) and abnormal nuclei (bottom two panels) were analysed by confocal microscopy. Despite both proteins localizing to the nucleus, partial co-localization between the two proteins was only apparent in cells bearing abnormal nuclei.

4.4 DISCUSSION

Over-expression of most Sin1 isoforms can induce apoptosis in HEK293 cells. Although the mechanisms of Sin1 induced apoptosis remain unclear, three pieces of evidence suggest that the nuclear localization of Sin1 is likely to be important for induction of apoptosis. Firstly, Sin1β was the only isoform unable to significantly induce apoptosis, and is also the only isoform that lacks appreciable nuclear localization in HEK293 cells (Chapter Two). Secondly, the only Sin1 domain significantly associated with apoptosis (the CRIM) localizes to the nucleus. Thirdly, N-terminally tagged EGFP Sin1 isoforms, partly inhibited in their ability to enter the nucleus (Chapter Two), less reliably induce apoptosis than their C-terminally tagged counterparts (large standard error; Figure 4.3.2A).

With regards to the biological significance of this effect, it is worth recalling that Sin1 interacts with the type I interferon (IFN) signalling receptor subunit, IFNAR2 (Wang and Roberts, 2004). Type I IFNs are mediators of apoptosis in virally-infected cells (Tanaka *et al.*, 1998), and *Sin1* first came to the attention of our group as a gene up-regulated during EBV infection (Schroder *et al.*, unpublished data), suggesting that Sin1 proteins could potentially play a role in modulating the apoptotic response to pathogens. This study is the first to describe a physiological effect of Sin1 over-expression in vertebrate cells, although whether the induction of apoptosis by Sin1 is an artifact of over-expression, or a reflection of the function of endogenous Sin1 is not yet clear.

This study is also the first to identify cellular stresses which affect the protein levels of endogenous Sin1. Full length Sin1 protein levels rapidly decrease in response to serum and media withdrawal in HEK293 cells. This is a context-specific response, as the same treatment in Cos7 and DG75 cells rapidly increases Sin1 protein expression. The decrease in Sin1 protein levels was also shown to require *de novo* synthesis of RNA and protein, as the inhibition of transcription or translation by actinomycin D or cycloheximide, respectively, blocked this effect. The p70 Sin1 isoform (thought to be Sin1β) also responds to hypoosmotic stress, rapidly and significantly increasing in response to media dilution. Interestingly though, neither of the major protein isoforms responded to hyperosmotic stress. It is possible that the hyperosmostic stress pathway has degenerated in most mammalian cells (Sbrissa *et al.*, 2004), which may explain the lack of response to this cellular insult, although it is also possible that Sin1 is actually responding to a dilution of an important nutrient rather than the change in osmolarity.

Sin1 proteins may be able to form dimers *in vivo*, mediated by the coiled-coil region in the N-terminal domain. The N-terminal domain appears to be post-translationally modified, as the apparent molecular mass of EGFP-NTD is approximately 13 kDa larger than its predicted molecular mass. Other constructs, including EGFP alone, have apparent molecular masses within 5 kDa of their predicted molecular masses (Chapter 2: Figure 2.3.6B). Additionally, bacterially-expressed GST-NTD also has an apparent molecular mass within 5 kDa of its predicted molecular mass, suggesting a type of post-translational modification specific to mammalian cells. This

modification may play a role in regulating Sin1 dimer formation, as bacterially expressed GST-NTD could bind Sin1 proteins but mammalian expressed EGFP-NTD appeared unable to do so. Supporting this, mammalian expressed EGFP-SCDII (containing the coiled-coil region) could also bind Sin1 proteins, and appears to lack substantial protein modifications. Further experiments are required to confirm post-translational modifications of the Sin1 proteins, and any role that these may play in regulating Sin1 dimer formation.

The ability of Sin1 to form dimers *in vivo* may go some way to resolving the issue of SAPK-binding sites in Sin1 proteins. As reviewed in Chapter One, the PH-like domain of *S. pombe* Sin1 is both sufficient and necessary for interaction with the SAPK Sty1. The function of this domain is also evolutionarily conserved, as the PH-like domain of chicken Sin1 (94% identical to human Sin1) restored function to a yeast strain where the PH-like domain of *S. pombe* Sin1 was deleted (Wilkinson *et al.*, 1999). In contrast, Schroder *et al.* (2005) found that the PH-like domain was dispensable for SAPK binding, as Sin1α (lacking both the Ras binding domain and the PH-like domain) was able to immunoprecipitate the Sty1-orthologue JNK. If Sin1α can form dimers with Sin1 isoforms containing intact PH-like domains, then this provides a plausible explanation that can account for both sets of data.

Collectively, the results presented in this chapter represent an advance in our knowledge of Sin1 function. The final chapter in this thesis discusses the function and evolution of the Sin1 proteins, and suggests a direction for future research.

CHAPTER FIVE

General discussion

5.1 General Discussion

The inception of bioinformatics has provided the easiest method for the prediction of the biological functions of an uncharacterized protein through the sophisticated matching to homologous proteins with known functions. For human Sin1, there is an unfortunate lack of paralogues within the genome, and the biological function of the orthologous Sin1 proteins is largely unknown. Despite these challenges, Sin1 is an interesting target for study, as it is ubiquitously expressed in mammalian tissues, the gene copy number is strictly conserved from yeast to humans, and the protein interacts with a diverse range of important signalling molecules (reviewed in Chapter One). Collectively, this suggests that human Sin1 has important cellular roles, and the overall aim of this study has focussed on discovering what some of these roles might be. A variety of approaches were used, including studies in structure and localization (Chapter Two), investigating the genomic regulation of human *Sin1* gene expression (Chapter Three), as well as examining the biochemical properties and functions of the human Sin1 protein (Chapter Four). Thus, these studies have contributed significantly to our understanding of this protein.

5.1.1 The protein domains and functions of Sin1

Identifying the domains within a protein is an important step in understanding the function of a protein. However, the sparsity of proteins with significant sequence homology to Sin1 limited the use of traditional sequence-based approaches to domain

detection. By predicting the interdomain regions and using secondary structure homology, the four major domains within human Sin1 could be deduced, and the contributions of these domains to the localization and function of Sin1 studied (Chapter Two). The N-terminal domain (NTD) is likely to be responsible for the nucleo-cytoplasmic shuttling of Sin1 proteins, as it contains functional nuclear localization and nuclear export signals. This domain also contains a functional coiled-coil region, which mediates Sin1 dimer formation (Chapter Four). Interestingly, the coiled-coil region (I80-Q107) overlaps with the NLS (I81-R97), and this may provide a mechanism for the spatial regulation of Sin1 isoforms. For example, dimerization of Sin1 proteins may prevent nuclear import proteins binding to the NLS, resulting in a cytoplasmic location. Similarly, Sin1 dimerization may regulate Sin1 function by distorting the tertiary conformation of the protein, thereby changing the accessibility of Sin1 to its protein partners. Coiled-coil regions can be predicted in all orthologues of Sin1, except RIP3 (*Dictyostelium*), although the location of these regions within the protein is not conserved. Instead, RIP3 contains several poly-glutamine tracts (Chapter Two: Figure 2.3.3), which can also function as self-association domains (Scherzinger *et al.*, 1997; Perutz, 1999), suggesting that self-dimerization may be a molecular property conserved in all Sin1 orthologues.

The conserved region in the middle (CRIM) may be a ubiquitin-like β-grasp protein binding domain. This domain is the key feature of the Sin1 protein family, being the only domain conserved in all Sin1 orthologues, and one of two domains preserved in

all confirmed human *Sin1* splice variants (Chapter Three). Sin1-induced apoptosis is mediated through this domain (Chapter Four), and therefore this may be a function relevant in lower eukaryotes as well. The induction of apoptosis is particularly intriguing, given the interaction of Sin1 with IFNAR2 (Wang and Roberts, 2004) and the apparent co-regulation of *Sin1* with genes involved with the host response to pathogen invasions (Chapter Three). IFNAR2 is a major subunit of the type I interferon (IFN) receptor. Type I IFNs have a dual role in the response to viral infections, inducing apoptosis in infected cells, and eliciting a protective anti-viral state in uninfected cells (Tanaka *et al.*, 1998). This response may be mediated through regulation of IFNAR2, as the ratio between IFNAR2 splice variants (both endogenously and exogenously expressed) influences the susceptibility of cells to type I IFN induced apoptosis (Gazziola *et al.*, 2005). As *Sin1* was originally identified by our laboratory as a gene potentially upregulated in response to Epstein Barr Virus (EBV) infection (Schroder *et al*., unpublished data), and *Sin1* appears to be co-regulated with genes involved in the response to pathogen infection (Chapter Three), it seems reasonable to speculate that IFNAR2 may signal through Sin1 mediated pathways to induce apoptosis in response to viral infection. Curiously, the PI3K/mTOR pathway also seems crucial for type I IFN induced apoptosis (Thyrell *et al.*, 2004). As discussed in Chapter One, Sin1 orthologues in lower eukaryotes have been shown to bind TOR (Loewith *et al.*, 2002; Wedaman *et al.*, 2003; Lee *et al.*, 2005), and an interaction between mammalian Sin1 and mammalian TOR (mTOR) has been suspected but not yet confirmed (Loewith *et al.*,

2002). Identifying the binding partners of the CRIM will undoubtedly shed light on the mechanisms of Sin1-induced apoptosis.

This study also identified a Ras binding domain (RBD) in the human Sin1 protein, consistent with RBDs previously identified in the *S. cerevisiae* and *Dictyostelium* orthologues. Recent biochemical evidence supports this prediction: over-expressed human Sin1 binds preferentially to activated Ras *in vitro*; over-expressed human Sin1 co-localizes with over-expressed H-Ras and K-Ras in BHK cells; and over-expression of bacterially-expressed Sin1 deletion constructs cannot pull down GFP-Ras from cell lysates if the RBD is not present (Schroder *et al*, manuscript submitted). Together, these data suggest that the RBD of human Sin1 may bind activated Ras, but further experiments using endogenous proteins are required to confirm the relevance of this interaction in the mammalian cell. The *Dictyostelium* Sin1 orthologue RIP3 also binds preferentially to activated Ras (Lee *et al.*, 1999), and the residue that appears critical for Ras binding in RIP3 (R681) is conserved in human Sin1 (R312) (Lee *et al.*, 2005). The Sin1/Ras interaction is probably not conserved in insects, as *A. gambiae* and *D. melanogaster* both appear to lack the RBD (Chapter Two). The human Sin1 isoforms Sin1α and Sin1β both have disrupted RBDs (Figure 2.3.2), suggesting that while an interaction with Ras might be an important function of Sin1, it is not a universal function of the Sin1 protein family.

The final Sin1 domain identified in this study was a PH-like domain in the C-terminal end of the protein, and it shares the greatest homology with family of

phosphotyrosine binding (PTB) domains. Although the ability of this protein to bind to phosphorylated tyrosine motifs has not been confirmed biochemically, it is interesting that two proteins known to bind Sin1 in this region, IFNAR2 in *O. aries* (Wang and Roberts, 2004) and Sty1 in *S. pombe* (Wilkinson *et al.*, 1999), are both tyrosine phosphorylated proteins (Ihle and Kerr, 1995; Shieh *et al.*, 1998). The PH-like domain plays an important role in the cellular localization of Sin1, as it contains a PIP binding motif that is likely to target Sin1 to the plasma membrane and to lipid rafts *in vivo* (Chapter Two). Indeed, this motif is deleted in the human splice variant Sin1γ, which does not localize at the plasma membrane.

The affinity of the Sin1 PH-like domain has not been thoroughly examined in this study; however it does appear to bind most strongly to PtdIns(3)P and PtdIns(3,5)P_2 *in vitro* (Chapter Two: Figure 2.3.7). Determining the affinities of the PH-like domain for the different phosphatidylinositol derivatives should reveal something about the function of Sin1. If Sin1 is found to bind many phosphoinositides with low specificity, a feature common to many PH domains (Kavran *et al.*, 1998), then the *in vivo* signalling through Sin1 may be dependent on the most abundant phosphoinositide species within the cell. Alternatively, if Sin1 does bind particular phosphoinositides with high specificity, than the localization and function of these specific phosphoinositides is likely to influence Sin1 function. Over-expression of the EGFP tagged PH-like domain revealed predominant plasma membrane localization (Chapter Two), suggesting that this fusion protein may be binding PtdIns(4,5) *in vivo* (Balla and Varnai, 2002; Meyer

and Teruel, 2003), however this does not necessarily reflect the endogenous function of this domain. The preliminary results presented in Chapter Two suggest specific binding may occur with PtdIns(3)P and PtdIns(3,5)P_2. PtdIns(3)P is constitutively produced in both yeast and mammalian cells (Stenmark and Aasland, 1999), and localizes to early endosomes and nucleoli (Gillooly *et al.*, 2000). PtdIns(3,5)P_2 is present in a wide range of eukaryotes, and localizes to late endosomes (Birkeland and Stenmark, 2004). Its relative abundance is rare, comprising only 0.8% of all seven phosphoinositide derivatives (Ikonomov *et al.*, 2001; Rudge *et al.*, 2004). Interestingly though, PtdIns(3,5)P_2 levels appear to rise in response to osmotic stresses in yeast (Dove *et al.*, 1997; McEwen *et al.*, 1999; Bonangelino *et al.*, 2002), plants (Meijer *et al.*, 2001; Zonia and Munnik, 2004), and mammalian adipocytes (Sbrissa and Shisheva, 2005). As Sin1 protein levels were also found to change following osmotic stress (Chapter Four), a more thorough investigation of Sin1 lipid affinities is warranted.

Studying the human Sin1 isoforms has helped predict important protein sequence motifs and domains within the protein. Subsequently, understanding the function of Sin1 domains may allow predictions regarding the function of Sin1 isoforms. It is possible that the C-terminally truncated Sin1α isoform may act as a dominant negative protein inhibiting full length Sin1 function, since Sin1α does not contain RBD or PH-like domains. Sin1β and Sin1γ are also likely to have different biological roles within the cell, potentially unable to respond to Ras and PtdIns signalling respectively. Some evidence for the different cellular roles of Sin1 isoforms

was presented in Chapter Four. Full length Sin1 protein levels decreased in HEK293 cells following serum and media deprivation whereas no significant effect was observed for the p70 isoform. Similarly, the p70 isoform was found to increase following hypoosmotic stress, with no apparent effect on the full length protein. It should be noted that the identity of the p70 isoform (currently thought to represent Sin1β; Schroder *et al*., manuscript submitted) has not been confirmed. As this band was detected using a polyclonal antibody, caution should be used when interpreting these results.

5.1.2 The evolution of Sin1

Any consideration regarding the biological significance of Sin1 needs to place the protein in its proper evolutionary context. The evolutionary conservation of Sin1 function has been assumed but not demonstrated, and is largely based on the conservation of the TORC2 complex in *S. cerevisiae* and *Dictyostelium* (Loewith *et al.*, 2002; Lee *et al.*, 2005). It is important to remember, however, that human Sin1 and yeast Sin1 orthologues are separated by great evolutionary distance (Chapter Two; Wang and Roberts, 2005), and that specific molecular functions for these proteins (such as interactions with other proteins) may have diverged over time. Despite this, there are significant similarities between Sin1 orthologues, not only in a sequence-based context, but in a genomic context as well.

Sin1 exists as a single gene in every genome in which it is found, and there are no identifiable paralogues or pseudogenes (Schroder *et al.*, 2004; Wang and Roberts,

2005). Given that gene duplication is a common evolutionary strategy for genetic diversification (Saccone *et al.*, 2002), and that a large proportion of mammalian genes belong to multicopy gene families (Zhang, 2003), there seems to be significant evolutionary pressure against the duplication of *Sin1*. This study has shown that, in mammalian cell systems, over-expression of Sin1 can induce apoptosis or mitotic catastrophe (Chapter Four). This effect may explain why there are no paralogues or pseudogenes of *Sin1*. It may also explain why the human *Sin1* gene may be more heavily regulated than an average human gene (Chapter Three). Almost 5% of the non-coding, non-repetitive genomic *Sin1* sequence may be under negative selection pressure (where mutations in the sequence reduce the organism's fitness), compared with the predicted average of 3.5% (Smith *et al.*, 2004).

Even so, there must also be significant advantages for organisms to retain *Sin1*, or else the cost of maintaining this gene would outweigh the benefit, and the gene would have been deleted from the genome. Certainly, there are significant benefits for *S. cerevisiae*, where AVO1 is a gene essential for life (Loewith *et al.*, 2002). *S. pombe* and *D. discoideum* also benefit from Sin1 and RIP3, respectively, which are both proteins required for sexual reproduction (Lee *et al.*, 1999; Wilkinson *et al.*, 1999). It is currently unknown whether Sin1 is essential for mammalian cells, as to date no mammalian Sin1 knockout experiments have been performed. Although not specifically examined, no lethal phenotype was reported for the only Sin1 knockdown experiments that has been done in mammalian cell lines (Cheng *et al.*, 2005). It is possible that Sin1 may not be

required for normal proliferation of mammalian cells in culture, but still may be required for survival of cells under stress. If not crucial for cell survival, Sin1 is important in vertebrates, as the conservation within vertebrate Sin1 proteins (approximately 85%) is comparable with the cell cycle protein CDK1 (Wang and Roberts, 2005). Additionally, in more than 600 single nucleotide polymorphisms detected for the human *Sin1* gene (GeneCards), only one results in an amino acid substitution (conservative: K to Q, residue 267 of the human sequence). It is therefore likely that Sin1 has evolved an important and highly conserved function in vertebrates.

This function may relate to the IFN signalling pathways that first arose in vertebrates (Hughes, 1995; Roberts *et al.*, 1998), or it may be due to a role for Sin1 in vertebrate specific embryonic development. Rhombomeres are morphologically distinct sections of the developing neural crest present in vertebrates but not invertebrates. During development, large-scale apoptosis occurs in rhombomeres R3 and R5 allowing physical segregation of the neighbouring neural crest cell lines (Graham *et al.*, 1993; Knabe *et al.*, 2004). In the developing chicken hindbrain, *Sin1* mRNA is specifically expressed in R3 and R5 (Christiansen *et al.*, 2001), and given the purported role for Sin1 in apoptosis (Chapter Four), it is possible that Sin1 is involved in this specific morphogenic feature. Sin1 may also be involved in muscle differentiation (myogenesis), as full length Sin1 mRNA is highly expressed in the heart and skeletal muscle (Loewith *et al.*, 2002; Schroder *et al.*, 2004; Cheng *et al.*, 2005). Intriguingly, myogenesis is a process that also involves IFN signalling (Clemens and McNurlan, 1985; Birnbaum *et*

al., 1990). Other cell types in which *Sin1* mRNA is highly expressed are also linked to IFN signalling. GNF Symatlas (http://symatlas.gnf.org/SymAtlas/; Su *et al.*, 2002) reports that a probe corresponding to part of exons 10 and 11 of human *Sin1* is highly expressed (more than three times the median for that gene) in PB-BDCA4+ dendritic cells and 721 B-lymphoblasts. 721 B lymphoblasts are an EBV-infected B cell line (Kavathas *et al.*, 1980). PB-BDCA4+ cells are plasmacytoid dendritic cells (Dzionek *et al.*, 2000) which can produce very high amounts of type I IFN in response to pathogen infection, and provide a link between the innate and adaptive immune systems (reviewed in McKenna *et al.*, 2005). Symatlas also reported extremely high expression of *Sin1* in mouse oocysts and fertilized eggs. Here however, the two probes used were not specific to mouse *Sin1*, both containing simple repeats. Blast searches of the mouse genome revealed more than 10,000 high specificity hits for each probe, suggesting that the expression results obtained are not likely to be specific for *Sin1* mRNA. Collectively, this data suggests that vertebrate Sin1 proteins have an important and highly conserved function that is likely to involve IFN signalling.

Another important consideration is the notable absence of Sin1 orthologues from the plant kingdom, when orthologues can be found in all other eukaryotic kingdoms. While plants lack IFN signalling and vertebrate patterning, fungi and amoeba also lack these systems but retain Sin1 as a necessary component of signal transduction, interacting with the small GTPase Ras and MAPKs (Lee *et al.*, 1999; Wilkinson *et al.*, 1999). As plants have small GTPases (Jekely, 2003) and utilize MAPK signalling

pathways (Asai *et al.*, 2002; Jonak *et al.*, 2002), it seems plausible that Sin1 orthologues may be present in plants, but are so divergent that sequence-based searches are unable to detect them reliably. Although highly conserved in vertebrates, Sin1 orthologues are only weakly conserved between the lower eukaryotes (approximately 25% sequence identity; Wang and Roberts, 2005), supporting the notion of sequence diversity as a limitation to detection. Of course, plant physiology may mean that Sin1 orthologues are unnecessary, or convergent evolution may have provided another solution. Any attempt to define Sin1 orthologues in the plant kingdom would necessarily include a detailed analysis of secondary structure homology.

5.2 Future directions

Several important contributions to the overall knowledge of human Sin1 function have resulted from the work presented here, the most important of which relates to the ability of Sin1 to induce apoptosis and its link to innate immunity and development. It is critical to find out if the observed apoptosis is a reflection of the function of endogenous Sin1, or whether it is an artifact of over-expression of the protein. Circumstantial evidence supports the former hypothesis: Sin1 genes are very tightly regulated, participate in pathways that lead to apoptosis, and are up-regulated in tissues that undergo substantial apoptosis. If true, Sin1 could be investigated as a drug target for the treatment of malignant tumours, either alone or augmenting the anti-cancer effects of existing IFN treatments.

To accomplish this, more needs to be understood about the biochemical pathways through which Sin1 signals to induce apoptosis. Key questions include: does Sin1 modulate the apoptotic response to IFN treatment? Does Sin1 knockdown abrogate IFN induced signalling? Is Sin1 post-translationally modified, and if so how does this effect the regulation of Sin1 induced apoptosis? Which caspases are activated in apoptosis induced by Sin1 over-expression? As human Sin1 proteins contain a cleavage site for caspase 1 (cutting after residue 399 of the human sequence, after the PIP binding motif, removing most of the PH-like domain), does this allow the transport of Sin1 binding proteins to the nucleus to signal apoptosis? Identifying the binding partners of the CRIM will be essential, as will understanding the regulation of Sin1. Investigating the dynamics of Sin1 containing protein complexes (including the temporal and spatial arrangement of components) will be particularly revealing. However, caution should be used when interpreting data based solely on the over-expression of proteins. Where possible, functional observations should be confirmed using endogenous proteins.

In conclusion, the structure and sub-cellular localization of Sin1, the biochemical functions of Sin1, and the regulatory regions of Sin1 have been investigated. The summation of these results suggests a potential biological role: Sin1 might modulate an apoptotic response to viral infection through the IFN signalling pathway. Although further studies are needed to confirm this biological function, the Sin1 protein family could represent a new family of targets for anti-cancer therapies.

APPENDIX I

Multiple sequence alignments of Sin1 protein domains

Alignment of the N-terminal domain (NTD):

```
H.sapiens  : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKIHPPSMPGDS-----  :  50
P.troglody : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKIHPPSMPGDS-----  :  50
P.pygmaeus : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKIHPPSMPGDS-----  :  50
M.musculus : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKTHPPSVPGDS-----  :  50
R.norvegic : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKTHPPSVPGDS-----  :  50
B.taurus   : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKIHPPSMPGDS-----  :  50
O.aries    : MGFLDNPTIILAHIRQSHVTSDDIG--MCEVVLIDHD---VDLEKIHPPSMPGDS-----  :  50
C.familiar : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKIHPPSMPGDS-----  :  50
S.scrofa   : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKIHPPSVPGDS-----  :  50
G.gallus   : MAFLDNPTIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKFNPSSTYGDS-----  :  50
X.laevis   : MAFLDNPLIILAHIRQSHVTSDDIG--MCEMVLIDHD---VDLEKLYQSSVSGEN-----  :  50
F.rubripes : MAFLDNPGIILAHIRQSHVTSDDIG--MCEMVLIDQD---VDLEKCQVALVPGSSNYMST  :  55
A.gambiae  : MATYNNAHWLLSHIRNSFISTDDIG--VSETVMVMED---IPLQLAKRFRESQARVAETT  :  55
D.melanoga : MATYSNQHWLLSHIRNSFISTDDIG--MCETVMLSDD---MPKHYLRKFGNSGAGGDHYH  :  55
C.elegans  : MGHVDRE-DLLNVIRHELRLEDDDGPGLGSRLLLNPD(7)LPLDFRLKNGDLMDDGDAGF  :  56

H.sapiens  : ------------------------GSEIQGSNGETQGYVY----------------AQSV  :  70
P.troglody : ------------------------GSEIQGSNGETQGYVY----------------AQSV  :  70
P.pygmaeus : ------------------------GSEIQGSNGETQGYVY----------------AQSV  :  70
M.musculus : ------------------------GSEVQGSSGETQGYIY----------------AQSV  :  70
R.norvegic : ------------------------GSEVQGSSGETQGYIY----------------AQSV  :  70
B.taurus   : ------------------------GSEIQGSNGETQGYVY----------------AQSV  :  70
O.aries    : ------------------------GSEIQGSNGETQGYVY----------------AQSV  :  70
C.familiar : ------------------------GSEIQGSNGETQGYVY----------------AQSV  :  70
S.scrofa   : ------------------------GSEIQGSNGETQGYVY----------------AQSV  :  70
G.gallus   : ------------------------ASETQGSNGETQGYVY----------------SQSV  :  70
X.laevis   : ------------------------STQMQSNGGETQGYVY----------------SQSV  :  70
F.rubripes : GS----------------------GSFNEGGNNVTDPHACDL--------------SQSM  :  79
A.gambiae  : NQPQPQPETTAPDEDSDGPGIDPIKQQMQARFGVDYEYFFSYPGLDDTDEEDADTLSQSY  : 115
D.melanoga : WR---RAHKTPP----TGGGTTPERNTRHPDAPLQEVDFICYPGLDLSDDE-EDMSTHSF  : 107
C.elegans  : DDIYEVRRG--------FWKKCFPELKKKALSAFSAKYFIYFLFKKNIFSLKTHFLPKIF  : 108

H.sapiens  : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
P.troglody : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
P.pygmaeus : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
M.musculus : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
R.norvegic : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
B.taurus   : DIT----SSWDFGIRR----RSNTAQRIERLRKERQNQIKCKNIQWKER  : 111
O.aries    : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
C.familiar : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
S.scrofa   : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNIQWKER  : 111
G.gallus   : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIKCKNVQWKDR  : 111
X.laevis   : DIT----SSWDFGIRR----RSNTAQKLERLRKERQNQIKCKNVQWKDR  : 111
F.rubripes : DIT----SSWDFGIRR----RSNTAQRLERLRKERQNQIRCKNIQWKDR  : 120
A.gambiae  : EIQ----MDQESGFHRQ---RSNTAQKLEKLEIARRKASQIRSVKLDDS  : 157
D.melanoga : DIQ----MYPEVGAHRF---RSNTAQKLEKLDIAKRRAARIKSVNYQEE  : 149
C.elegans  : TIL(25)FSNKKGIKK(24)PIKSIPLYEEPIHTRTLINDSLALRTAEK  : 149
```

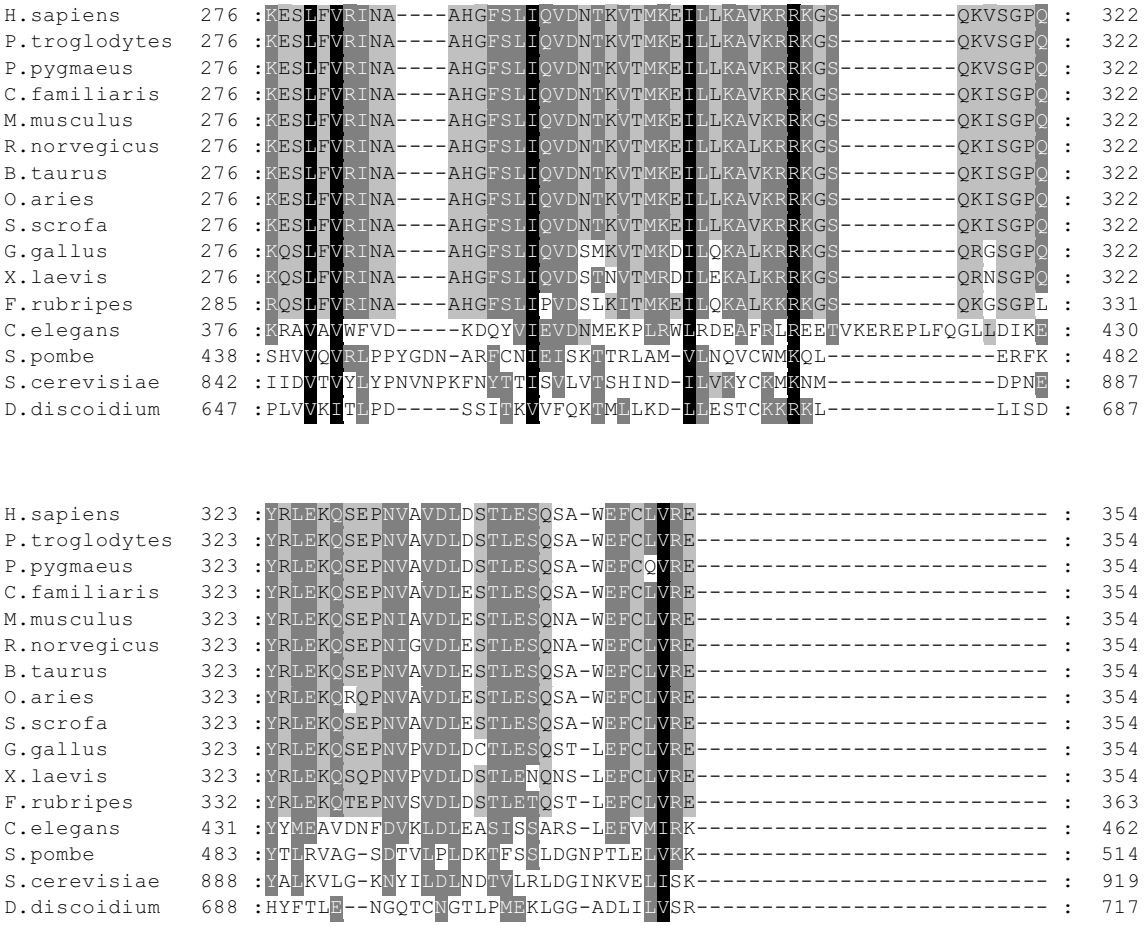
Alignment of the conserved region in the middle (CRIM):

```
H.sapiens        140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
P.troglodytes    140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
P.pygmaeus       140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
C.familiaris     140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
M.musculus       140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
R.norvegicus     140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
B.taurus         140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
O.aries          140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
S.scrofa         140 :SILSVRLEQC---PLQLNNPL-YEYSKSDGKGHVGTTATKKIDVYLPLHSSQDRLL-PMT :  194
G.gallus         140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLPLHANQDKLQ-PMT :  194
X.laevis         140 :SILSVRLEQC---PLQLNNPF-NEYSKFDGKGHVGTTATKKIDVYLSMQTSQDKLL-PMT :  194
F.rubripes       149 :STLSLRLEQC---PQQLNNPF-NEYSKFDGKGHIGTTATKKIDVYLSMQMTQEKLH-PMT :  203
A.gambiae        190 :STLSEQLSTL---PALPQNQF-LEYAKFDGTGHIE-VPVETFKVFVPALPEEQRAF-PLP :  243
D.melanogaster   190 :SQLLEQLAKS---PKQAQNRF-IEFARFDGTSQVG-MQTKRINVFLNMLPEPDRNY-PLK :  243
C.elegans        242 :SAIERYLEEN---SANLNNPL-GEYSKFAATTTDP---SRQIEIIIPMSCDEEIGFKTLK :  294
S.pombe          256 :SALRALLEHK--ENSSQNGPLAENFATPSGHAESN---ALRLNIYFPSSESPSK---PLF :  307
S.cerevisiae     638 :SQLSNMFNKKKKRTNTNSVDVLEYFSFVC-GDKVPNYESMGLEIYIQASKKYKRN---SF :  693
D.discoidium     402 :SLLLRLVKPNSEEA---------EYGDIVP--------PPGMGLTLSIHTGNTGAG-QLK :  443

H.sapiens        195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
P.troglodytes    195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
P.pygmaeus       195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
C.familiaris     195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
M.musculus       195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
R.norvegicus     195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
B.taurus         195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
O.aries          195 :VVTMASARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
S.scrofa         195 :VVTMDSARVQDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
G.gallus         195 :VVTIANAKVHDLIGLICWQYTSEGREPKLN----------DNVSAYCLHIAEDDGE-VDT :  243
X.laevis         195 :VVTIANAKVHDLIGLICWQYTTEGREPKLN----------DNVDAFCLHIAEDDGE-VDT :  243
F.rubripes       204 :VVTIANARVHDLIGLICWQYTSEGREPKLN----------ENVNAYCLHIAEDDGE-VDT :  252
A.gambiae        244 :VCVLATAKIQEFIGLICYKCTIANPAVEIR----------PVRHYGLYMTPEDGE-LDL :  291
D.melanogaster   244 :ICVVATAKIQEVIGFVCYRTSLQYPDVPLK----------SLQHYALYMTEDNDD-ME- :  290
C.elegans        295 :IEVLTTARIREVIGYCLLQYYLTFDSYLPG----------EVDDYQFYLAEEDGE-IEH :  342
S.pombe          308 :VELRKNVLVSEAIGYILLQYVNQQLVPPIEDEA-------QNPNYWNLRIVEDDGE-LDE :  359
S.cerevisiae     694 :TTKVRKSSTIFEVIGFALFLYSTEKKPDNFEEDGLTVEDISNPNNFSLKIVDEDGEPFED :  753
D.discoidium     444 :VRVLEKATIIQTIFATLKLHHNNGGTGLLP----------DPKAYNLRIADSNGR-IDQ :  491

H.sapiens        244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
P.troglodytes    244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
P.pygmaeus       244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
C.familiaris     244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
M.musculus       244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
R.norvegicus     244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
B.taurus         244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
O.aries          244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
S.scrofa         244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
G.gallus         244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
X.laevis         244 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  267
F.rubripes       253 :DFPPLDSNEPIHKFGFSTLALVEK---------------------------------- :  276
A.gambiae        292 :DFPPLDVNEPCSKFRFTHLVMAER---------------------------------- :  315
D.melanogaster   291 :DFPPLDNREPCSKFGFSQLTLAER---------------------------------- :  314
C.elegans        343 :ELPPLDSSKLVGQVGFTCLGLVSR---------------------------------- :  366
S.pombe          360 :DFFALDRVGPLSKFGFDAFALVKA---------------------------------- :  383
S.cerevisiae     754 :NFGKLDRKSTIQSIS-DSEVVLCK---------------------------------- :  776
D.discoidium     492 :DFPPLDPNQYITKFKDEVLVLCPN---------------------------------- :  515
```

Alignment of the Ras binding domain (RBD):



REFERENCES

Antequera, F., and Bird, A. (1993) *Number of CpG islands and genes in human and mouse*. **Proc Natl Acad Sci U S A**, 90(24), 11995-9.

Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002) *MAP kinase signalling cascade in Arabidopsis innate immunity*. **Nature**, 415(6875), 977-83.

Balla, T., and Varnai, P. (2002) *Visualizing cellular phosphoinositide pools with GFP-fused protein-modules*. **Sci STKE**, 2002(125), PL3.

Belloc, F., Dumain, P., Boisseau, M.R., Jalloustre, C., Reiffers, J., Bernard, P., and Lacombe, F. (1994) *A flow cytometric method using Hoechst 33342 and propidium iodide for simultaneous cell cycle analysis and apoptosis determination in unfixed cells*. **Cytometry**, 17(1), 59-65.

Birkeland, H.C., and Stenmark, H. (2004) *Protein targeting to endosomes and phagosomes via FYVE and PX domains*. **Curr Top Microbiol Immunol**, 282, 89-115.

Birnbaum, M., Trink, B., Shainberg, A., and Salzberg, S. (1990) *Activation of the interferon system during myogenesis in vitro*. **Differentiation**, 45(2), 138-45.

Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000) *Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements)*. **Crit Rev Eukaryot Gene Expr**, 10(1), 73-90.

Bode, J., Schlake, T., Rios-Ramirez, M., Mielke, C., Stengert, M., Kay, V., and Klehr-Wirth, D. (1995) *Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci*. **Int Rev Cytol**, 162A, 389-454.

Bompard, G., Martin, M., Roy, C., Vignon, F., and Freiss, G. (2003) *Membrane targeting of protein tyrosine phosphatase PTPL1 through its FERM domain via binding to phosphatidylinositol 4,5-biphosphate*. **J Cell Sci**, 116(Pt 12), 2519-30.

Bonangelino, C.J., Nau, J.J., Duex, J.E., Brinkman, M., Wurmser, A.E., Gary, J.D., Emr, S.D., and Weisman, L.S. (2002) *Osmotic stress-induced increase of phosphatidylinositol 3,5-bisphosphate requires Vac14p, an activator of the lipid kinase Fab1p*. **J Cell Biol**, 156(6), 1015-28.

Bos, J.L. (1989) *ras oncogenes in human cancer: a review*. **Cancer Res**, 49(17), 4682-9.

Boulikas, T. (1995) *Chromatin domains and prediction of MAR sequences*. **Int Rev Cytol**, 162A, 279-388.

Burack, W.R., and Shaw, A.S. (2000) *Signal transduction: hanging on a scaffold*. **Curr Opin Cell Biol**, 12(2), 211-6.

Burkhard, P., Stetefeld, J., and Strelkov, S.V. (2001) *Coiled coils: a highly versatile protein folding motif*. **Trends Cell Biol**, 11(2), 82-8.

Challacombe, J.F., Rechtsteiner, A., Gottardo, R., Rocha, L.M., Browne, E.P., Shenk, T., Altherr, M.R., and Brettin, T.S. (2004) *Evaluation of the host transcriptional response to human cytomegalovirus infection*. **Physiol Genomics**, 18(1), 51-62.

Chen, C.Y., and Shyu, A.B. (1994) *Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements*. **Mol Cell Biol**, 14(12), 8471-82.

Cheng, J., Zhang, D., Kim, K., Zhao, Y., and Su, B. (2005) *Mip1, an MEKK2-interacting protein, controls MEKK2 dimerization and activation*. **Mol Cell Biol**, 25(14), 5955-64.

Christiansen, J.H., Coles, E.G., Robinson, V., Pasini, A., and Wilkinson, D.G. (2001) *Screening from a subtracted embryonic chick hindbrain cDNA library: identification of genes expressed during hindbrain, midbrain and cranial neural crest development*. **Mech Dev**, 102(1-2), 119-33.

Clemens, M.J., and McNurlan, M.A. (1985) *Regulation of cell proliferation and differentiation by interferons*. **Biochem J**, 226(2), 345-60.

Colicelli, J., Nicolette, C., Birchmeier, C., Rodgers, L., Riggs, M., and Wigler, M. (1991) *Expression of three mammalian cDNAs that interfere with RAS function in Saccharomyces cerevisiae*. **Proc Natl Acad Sci U S A**, 88(7), 2913-7.

Doualla-Bell, F., and Koromilas, A.E. (2001) *Induction of PG G/H synthase-2 in bovine myometrial cells by interferon-tau requires the activation of the p38 MAPK pathway*. **Endocrinology**, 142(12), 5107-15.

Dove, S.K., Cooke, F.T., Douglas, M.R., Sayers, L.G., Parker, P.J., and Michell, R.H. (1997) *Osmotic stress activates phosphatidylinositol-3,5-bisphosphate synthesis*. **Nature**, 390(6656), 187-92.

Dyson, H.J., and Wright, P.E. (2002) *Coupling of folding and binding for unstructured proteins*. **Curr Opin Struct Biol**, 12(1), 54-60.

Dzionek, A., Fuchs, A., Schmidt, P., Cremer, S., Zysk, M., Miltenyi, S., Buck, D.W., and Schmitz, J. (2000) *BDCA-2, BDCA-3, and BDCA-4: three markers for distinct subsets of dendritic cells in human peripheral blood*. **J Immunol**, 165(11), 6037-46.

Ferrell, J.E., Jr. (2000) *What do scaffold proteins really do?* **Sci STKE**, 2000(52), PE1.

Frisch, M., Frech, K., Klingenhoff, A., Cartharius, K., Liebich, I., and Werner, T. (2002) *In silico prediction of scaffold/matrix attachment regions in large genomic sequences*. **Genome Res**, 12(2), 349-54.

Gardiner-Garden, M., and Frommer, M. (1987) *CpG islands in vertebrate genomes*. **J Mol Biol**, 196(2), 261-82.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) *Functional organization of the yeast proteome by systematic analysis of protein complexes*. **Nature**, 415(6868), 141-7.

Gazziola, C., Cordani, N., Carta, S., De Lorenzo, E., Colombatti, A., and Perris, R. (2005) *The relative endogenous expression levels of the IFNAR2 isoforms influence the cytostatic and pro-apoptotic effect of IFNalpha on pleomorphic sarcoma cells*. **Int J Oncol**, 26(1), 129-40.

Gillooly, D.J., Morrow, I.C., Lindsay, M., Gould, R., Bryant, N.J., Gaullier, J.M., Parton, R.G., and Stenmark, H. (2000) *Localization of phosphatidylinositol 3-phosphate in yeast and mammalian cells*. **EMBO J**, 19(17), 4577-88.

Goh, K.C., Haque, S.J., and Williams, B.R. (1999) *p38 MAP kinase is required for STAT1 serine phosphorylation and transcriptional activation induced by interferons*. **EMBO J**, 18(20), 5601-8.

Graham, A., Heyman, I., and Lumsden, A. (1993) *Even-numbered rhombomeres control the apoptotic elimination of neural crest cells from odd-numbered rhombomeres in the chick hindbrain*. **Development**, 119(1), 233-45.

Greenwell-Wild, T., Vazquez, N., Sim, D., Schito, M., Chatterjee, D., Orenstein, J.M., and Wahl, S.M. (2002) *Mycobacterium avium infection and modulation of human macrophage gene expression*. **J Immunol**, 169(11), 6286-97.

Hentges, K., Thompson, K., and Peterson, A. (1999) *The flat-top gene is required for the expansion and regionalization of the telencephalic primordium*. **Development**, 126(8), 1601-9.

Ho, H.L., Shiau, Y.S., and Chen, M.Y. (2005) *Saccharomyces cerevisiaeTSC11/AVO3 participates in regulating cell integrity and functionally interacts with components of the Tor2 complex*. **Curr Genet**, 47(5), 273-88.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W., Figeys, D., and Tyers, M. (2002) *Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry*. **Nature**, 415(6868), 180-3.

Holcik, M., Sonenberg, N., and Korneluk, R.G. (2000) *Internal ribosome initiation of translation and the control of cell death*. **Trends Genet**, 16(10), 469-73.

Holm, P.S., Bergmann, S., Jurchott, K., Lage, H., Brand, K., Ladhoff, A., Mantwill, K., Curiel, D.T., Dobbelstein, M., Dietel, M., Gansbacher, B., and Royer, H.D. (2002) *YB-1 relocates to the nucleus in adenovirus-infected cells and facilitates viral replication by inducing E2 gene expression through the E2 late promoter*. **J Biol Chem**, 277(12), 10427-34.

Hope, H.R., and Pike, L.J. (1996) *Phosphoinositides and phosphoinositide-utilizing enzymes in detergent-insoluble lipid domains*. **Mol Biol Cell**, 7(6), 843-51.

Hughes, A.L. (1995) *The evolution of the type I interferon gene family in mammals*. **J Mol Evol**, 41(5), 539-48.

Ihle, J.N., and Kerr, I.M. (1995) *Jaks and Stats in signaling by the cytokine receptor superfamily*. **Trends Genet**, 11(2), 69-74.

Ikonomov, O.C., Sbrissa, D., and Shisheva, A. (2001) *Mammalian cell morphology and endocytic membrane homeostasis require enzymatically active phosphoinositide 5-kinase PIKfyve*. **J Biol Chem**, 276(28), 26141-7.

Isakoff, S.J., Cardozo, T., Andreev, J., Li, Z., Ferguson, K.M., Abagyan, R., Lemmon, M.A., Aronheim, A., and Skolnik, E.Y. (1998) *Identification and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel in vivo assay in yeast*. **EMBO J**, 17(18), 5374-87.

Jacinto, E., Loewith, R., Schmidt, A., Lin, S., Ruegg, M.A., Hall, A., and Hall, M.N. (2004) *Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive*. **Nat Cell Biol**, 6(11), 1122-8.

Jekely, G. (2003) *Small GTPases and the evolution of the eukaryotic cell*. **Bioessays**, 25(11), 1129-38.

Jonak, C., Okresz, L., Bogre, L., and Hirt, H. (2002) *Complexity, cross talk and integration of plant MAP kinase signalling*. **Curr Opin Plant Biol**, 5(5), 415-24.

Kavathas, P., Bach, F.H., and DeMars, R. (1980) *Gamma ray-induced loss of expression of HLA and glyoxalase I alleles in lymphoblastoid cells*. **Proc Natl Acad Sci U S A**, 77(7), 4251-5.

Kavran, J.M., Klein, D.E., Lee, A., Falasca, M., Isakoff, S.J., Skolnik, E.Y., and Lemmon, M.A. (1998) *Specificity and promiscuity in phosphoinositide binding by pleckstrin homology domains*. **J Biol Chem**, 273(46), 30497-508.

Knabe, W., Washausen, S., Brunnett, G., and Kuhn, H.J. (2004) *Rhombomere-specific patterns of apoptosis in the tree shrew Tupaia belangeri*. **Cell Tissue Res**, 316(1), 1-13.

Kozak, M. (1986) *Influences of mRNA secondary structure on initiation by eukaryotic ribosomes*. **Proc Natl Acad Sci U S A**, 83(9), 2850-4.

Kozak, M. (1992) *A consideration of alternative models for the initiation of translation in eukaryotes*. **Crit Rev Biochem Mol Biol**, 27(4-5), 385-402.

Laemmli, U.K. (1970) *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. **Nature**, 227(5259), 680-5.

Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J., and Pavlidis, P. (2004) *Coexpression analysis of human genes across many microarray data sets*. **Genome Res**, 14(6), 1085-94.

Lee, S., Comer, F.I., Sasaki, A., McLeod, I.X., Duong, Y., Okumura, K., Yates, J.R., 3rd, Parent, C.A., and Firtel, R.A. (2005) *TOR Complex 2 Integrates Cell Movement during Chemotaxis and Signal Relay in Dictyostelium*. **Mol Biol Cell**, 16(10), 4572-83.

Lee, S., Parent, C.A., Insall, R., and Firtel, R.A. (1999) *A novel Ras-interacting protein required for chemotaxis and cyclic adenosine monophosphate signal relay in Dictyostelium*. **Mol Biol Cell**, 10(9), 2829-45.

Li, Q.J., Vaingankar, S., Sladek, F.M., and Martins-Green, M. (2000) *Novel nuclear target for thrombin: activation of the Elk1 transcription factor leads to chemokine gene expression*. **Blood**, 96(12), 3696-706.

Loddick, S.A., Wong, M.L., Bongiorno, P.B., Gold, P.W., Licinio, J., and Rothwell, N.J. (1997) *Endogenous interleukin-1 receptor antagonist is neuroprotective*. **Biochem Biophys Res Commun**, 234(1), 211-5.

Loewith, R., Jacinto, E., Wullschleger, S., Lorberg, A., Crespo, J.L., Bonenfant, D., Oppliger, W., Jenoe, P., and Hall, M.N. (2002) *Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control*. **Mol Cell**, 10(3), 457-68.

Martin-Serrano, J., Eastman, S.W., Chung, W., and Bieniasz, P.D. (2005) *HECT ubiquitin ligases link viral and cellular PPXY motifs to the vacuolar protein-sorting pathway*. **J Cell Biol**, 168(1), 89-101.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999) *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure*. **J Mol Biol**, 288(5), 911-40.

McEwen, R.K., Dove, S.K., Cooke, F.T., Painter, G.F., Holmes, A.B., Shisheva, A., Ohya, Y., Parker, P.J., and Michell, R.H. (1999) *Complementation analysis in PtdInsP kinase-deficient yeast mutants demonstrates that Schizosaccharomyces pombe and murine Fab1p homologues are phosphatidylinositol 3-phosphate 5-kinases*. **J Biol Chem**, 274(48), 33905-12.

McGarvey, J.A., Wagner, D., and Bermudez, L.E. (2004) *Differential gene expression in mononuclear phagocytes infected with pathogenic and non-pathogenic mycobacteria*. **Clin Exp Immunol**, 136(3), 490-500.

McKenna, K., Beignon, A.S., and Bhardwaj, N. (2005) *Plasmacytoid dendritic cells: linking innate and adaptive immunity*. **J Virol**, 79(1), 17-27.

Meijer, H.J., Berrie, C.P., Iurisci, C., Divecha, N., Musgrave, A., and Munnik, T. (2001) *Identification of a new polyphosphoinositide in plants, phosphatidylinositol 5-monophosphate (PtdIns5P), and its accumulation upon osmotic stress*. **Biochem J**, 360(Pt 2), 491-8.

Meyer, T., and Teruel, M.N. (2003) *Fluorescence imaging of signaling networks*. **Trends Cell Biol**, 13(2), 101-6.

Millar, J.B. (1999) *Stress-activated MAP kinase (mitogen-activated protein kinase) pathways of budding and fission yeasts*. **Biochem Soc Symp**, 64, 49-62.

Miranda, E.I., Santana, C., Rojas, E., Hernandez, S., Ostrosky-Wegman, P., and Garcia-Carranca, A. (1996) *Induced mitotic death of HeLa cells by abnormal expression of c-H-ras*. **Mutat Res**, 349(2), 173-82.

Najib, S., Martin-Romero, C., Gonzalez-Yanes, C., and Sanchez-Margalet, V. (2005) *Role of Sam68 as an adaptor protein in signal transduction*. **Cell Mol Life Sci**, 62(1), 36-43.

Nakayama, K. (1997) *Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins*. **Biochem J**, 327 (Pt 3), 625-35.

Nathan, C., and Shiloh, M.U. (2000) *Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens*. **Proc Natl Acad Sci U S A**, 97(16), 8841-8.

Nevins, T.A., Harder, Z.M., Korneluk, R.G., and Holcik, M. (2003) *Distinct regulation of internal ribosome entry site-mediated translation following cellular stress is mediated by apoptotic fragments of eIF4G translation initiation factor family members eIF4GI and p97/DAP5/NAT1*. **J Biol Chem**, 278(6), 3572-9.

Ohsako, S., and Elkon, K.B. (1999) *Apoptosis in the effector phase of autoimmune diabetes, multiple sclerosis and thyroiditis*. **Cell Death Differ**, 6(1), 13-21.

Olson, M.O. (2004) *Sensing cellular stress: another new function for the nucleolus?* **Sci STKE**, 2004(224), pe10.

Olson, M.O., Hingorani, K., and Szebeni, A. (2002) *Conventional and nonconventional roles of the nucleolus*. **Int Rev Cytol**, 219, 199-266.

Orengo, C.A., Jones, D.T., and Thornton, J.M. (1994) *Protein superfamilies and domain superfolds*. **Nature**, 372(6507), 631-4.

Pederson, T. (1998) *The plurifunctional nucleolus*. **Nucleic Acids Res**, 26(17), 3871-6.

Perutz, M.F. (1999) *Glutamine repeats and neurodegenerative diseases: molecular aspects*. **Trends Biochem Sci**, 24(2), 58-63.

Pesole, G., Liuni, S., Grillo, G., Ippedico, M., Larizza, A., Makalowski, W., and Saccone, C. (1999) *UTRdb: a specialized database of 5' and 3' untranslated regions of eukaryotic mRNAs*. **Nucleic Acids Res**, 27(1), 188-91.

Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) *MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data*. **Nucleic Acids Res**, 23(23), 4878-84.

Rice, P., Longden, I., and Bleasby, A. (2000) *EMBOSS: the European Molecular Biology Open Software Suite*. **Trends Genet**, 16(6), 276-7.

Roberts, R.M., Liu, L., Guo, Q., Leaman, D., and Bixby, J. (1998) *The evolution of the type I interferons*. **J Interferon Cytokine Res**, 18(10), 805-16.

Rodenhuis, S. (1992) *ras and human tumors*. **Semin Cancer Biol**, 3(4), 241-7.

Rudge, S.A., Anderson, D.M., and Emr, S.D. (2004) *Vacuole size control: regulation of PtdIns(3,5)P2 levels by the vacuole-associated Vac14-Fig4 complex, a PtdIns(3,5)P2-specific phosphatase*. **Mol Biol Cell**, 15(1), 24-36.

Saccone, C., Barome, P.O., D'Erchia, A.M., D'Errico, I., Pesole, G., Sbisa, E., and Tullo, A. (2002) *Molecular strategies in Metazoan genomic evolution*. **Gene**, 300(1-2), 195-201.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. **Cold Spring Harbor Laboratory**, Cold Spring Harbor, NY.

Samson, S.L., and Wong, N.C. (2002) *Role of Sp1 in insulin regulation of gene expression*. **J Mol Endocrinol**, 29(3), 265-79.

Sarbassov, D.D., Ali, S.M., Kim, D.H., Guertin, D.A., Latek, R.R., Erdjument-Bromage, H., Tempst, P., and Sabatini, D.M. (2004) *Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton*. **Curr Biol**, 14(14), 1296-302.

Sasai, Y., Kageyama, R., Tagawa, Y., Shigemoto, R., and Nakanishi, S. (1992) *Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split*. **Genes Dev**, 6(12B), 2620-34.

Sbrissa, D., Ikonomov, O.C., Strakova, J., Dondapati, R., Mlak, K., Deeb, R., Silver, R., and Shisheva, A. (2004) *A mammalian ortholog of Saccharomyces cerevisiae Vac14 that associates with and up-regulates PIKfyve phosphoinositide 5-kinase activity*. **Mol Cell Biol**, 24(23), 10437-47.

Sbrissa, D., and Shisheva, A. (2005) *Acquisition of unprecedented phosphatidylinositol 3,5-bisphosphate rise in hyperosmotically stressed 3T3-L1 adipocytes, mediated by ArPIKfyve-PIKfyve pathway*. **J Biol Chem**, 280(9), 7883-9.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G.P., Davies, S.W., Lehrach, H., and Wanker, E.E. (1997) *Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo*. **Cell**, 90(3), 549-58.

Schroder, W., Bushell, G., and Sculley, T. (2005) *The human stress-activated protein kinase-interacting 1 gene encodes JNK-binding proteins*. **Cell Signal**, 17(6), 761-7.

Schroder, W., Cloonan, N., Bushell, G., and Sculley, T. (2004) *Alternative polyadenylation and splicing of mRNAs transcribed from the human Sin1 gene*. **Gene**, 339, 17-23.

Shaw, P.E., and Saxton, J. (2003) *Ternary complex factors: prime nuclear targets for mitogen-activated protein kinases*. **Int J Biochem Cell Biol**, 35(8), 1210-26.

Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S., and Mizumoto, K. (1992) *mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylytransferase subunit from Saccharomyces cerevisiae*. **J Biol Chem**, 267(14), 9521-8.

Shieh, J.C., Martin, H., and Millar, J.B. (1998) *Evidence for a novel MAPKKK-independent pathway controlling the stress activated Sty1/Spc1 MAP kinase in fission yeast*. **J Cell Sci**, 111 (Pt 18), 2799-807.

Sibson, N.R., Blamire, A.M., Perry, V.H., Gauldie, J., Styles, P., and Anthony, D.C. (2002) *TNF-alpha reduces cerebral blood volume and disrupts tissue homeostasis via an endothelin- and TNFR2-dependent pathway*. **Brain**, 125(Pt 11), 2446-59.

Singh, G.B., Kramer, J.A., and Krawetz, S.A. (1997) *Mathematical model to predict regions of chromatin attachment to the nuclear matrix*. **Nucleic Acids Res**, 25(7), 1419-25.

Smith, N.G., Brandstrom, M., and Ellegren, H. (2004) *Evidence for turnover of functional noncoding DNA in mammalian genome evolution*. **Genomics**, 84(5), 806-13.

Spaargaren, M., Bischoff, J.R., and McCormick, F. (1995) *Signal transduction by Ras-like GTPases: a potential target for anticancer drugs*. **Gene Expr**, 4(6), 345-56.

Stenmark, H., and Aasland, R. (1999) *FYVE-finger proteins--effectors of an inositol lipid*. **J Cell Sci**, 112 (Pt 23), 4175-83.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., Patapoutian, A., Hampton, G.M., Schultz, P.G., and Hogenesch, J.B. (2002) *Large-scale analysis of the human and mouse transcriptomes*. **Proc Natl Acad Sci U S A**, 99(7), 4465-70.

Takaoka, A., Mitani, Y., Suemori, H., Sato, M., Yokochi, T., Noguchi, S., Tanaka, N., and Taniguchi, T. (2000) *Cross talk between interferon-gamma and -alpha/beta signaling components in caveolar membrane domains*. **Science**, 288(5475), 2357-60.

Tanaka, N., Sato, M., Lamphier, M.S., Nozawa, H., Oda, E., Noguchi, S., Schreiber, R.D., Tsujimoto, Y., and Taniguchi, T. (1998) *Type I interferons are essential mediators of apoptotic death in virally infected cells*. **Genes Cells**, 3(1), 29-37.

Tanaka, T., Kuroda, Y., and Yokoyama, S. (2003) *Characteristics and prediction of domain linker sequences in multi-domain proteins*. **J Struct Funct Genomics**, 4(2-3), 79-85.

Thiel, G., and Cibelli, G. (2002) *Regulation of life and death by the zinc finger transcription factor Egr-1*. **J Cell Physiol**, 193(3), 287-92.

Thyrell, L., Hjortsberg, L., Arulampalam, V., Panaretakis, T., Uhles, S., Dagnell, M., Zhivotovsky, B., Leibiger, I., Grander, D., and Pokrovskaja, K. (2004) *Interferon alpha-induced apoptosis in tumor cells is mediated through the phosphoinositide 3-kinase/mammalian target of rapamycin signaling pathway*. **J Biol Chem**, 279(23), 24152-62.

Uddin, S., Majchrzak, B., Woodson, J., Arunkumar, P., Alsayed, Y., Pine, R., Young, P.R., Fish, E.N., and Platanias, L.C. (1999) *Activation of the p38 mitogen-activated protein kinase by type I interferons*. **J Biol Chem**, 274(42), 30127-31.

Venema, J., Vos, H.R., Faber, A.W., van Venrooij, W.J., and Raue, H.A. (2000) *Yeast Rrp9p is an evolutionarily conserved U3 snoRNP protein essential for early pre-rRNA processing cleavages and requires box C for its association*. **RNA**, 6(11), 1660-71.

Visintin, R., and Amon, A. (2000) *The nucleolus: the magician's hat for cell cycle tricks*. **Curr Opin Cell Biol**, 12(3), 372-7.

Wang, B.M., Weiner, N.D., Takada, A., and Schacht, J. (1984) *Characterization of aminoglycoside-lipid interactions and development of a refined model for ototoxicity testing*. **Biochem Pharmacol**, 33(20), 3257-62.

Wang, S.Z., and Roberts, R.M. (2004) *Interaction of stress-activated protein kinase-interacting protein-1 with the interferon receptor subunit IFNAR2 in uterine endometrium*. **Endocrinology**, 145(12), 5820-31.

Wang, S.Z., and Roberts, R.M. (2005) *The evolution of the Sin1 gene product, a little known protein implicated in stress responses and type I interferon signaling in vertebrates*. **BMC Evol Biol**, 5(1), 13.

Wedaman, K.P., Reinke, A., Anderson, S., Yates, J., 3rd, McCaffery, J.M., and Powers, T. (2003) *Tor kinases are in distinct membrane-associated protein complexes in Saccharomyces cerevisiae*. **Mol Biol Cell**, 14(3), 1204-20.

Welch, W.J., and Feramisco, J.R. (1984) *Nuclear and nucleolar localization of the 72,000-dalton heat shock protein in heat-shocked mammalian cells*. **J Biol Chem**, 259(7), 4501-13.

Wilkinson, M.G., Pino, T.S., Tournier, S., Buck, V., Martin, H., Christiansen, J., Wilkinson, D.G., and Millar, J.B. (1999) *Sin1: an evolutionarily conserved component of the eukaryotic SAPK pathway*. **EMBO J**, 18(15), 4210-21.

Williams, S.H., Mouchel, N., and Harris, A. (2003) *A comparative genomic analysis of the cow, pig, and human CFTR genes identifies potential intronic regulatory elements*. **Genomics**, 81(6), 628-39.

Wilusz, C.J., Wormington, M., and Peltz, S.W. (2001) *The cap-to-tail guide to mRNA turnover*. **Nat Rev Mol Cell Biol**, 2(4), 237-46.

Wootton, J.C., and Drummond, M.H. (1989) *The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins*. **Protein Eng**, 2(7), 535-43.

Wullschleger, S., Loewith, R., Oppliger, W., and Hall, M.N. (2005) *Molecular organization of target of rapamycin complex 2*. **J Biol Chem**, 280(35), 30697-704.

Yan, W., Lee, H., Yi, E.C., Reiss, D., Shannon, P., Kwieciszewski, B.K., Coito, C., Li, X.J., Keller, A., Eng, J., Galitski, T., Goodlett, D.R., Aebersold, R., and Katze, M.G. (2004) *System-based proteomic analysis of the interferon response in human liver cells*. **Genome Biol**, 5(8), R54.

Yim, E.K., Meoyng, J., Namakoong, S.E., Um, S.J., and Park, J.S. (2004) *Genomic and proteomic expression patterns in HPV-16 E6 gene transfected stable human carcinoma cell lines*. **DNA Cell Biol**, 23(12), 826-35.

Zehntner, S.P., Mackay-Sim, A., and Bushell, G.R. (1998) *Differentiation in an olfactory cell line. Analysis via differential display*. **Ann N Y Acad Sci**, 855, 235-9.

Zhang, J. (2003) *Evolution by gene duplication: an update*. **Trends Ecol Evol**, 18(6), 292-298.

Zhang, X., Shu, L., Hosoi, H., Murti, K.G., and Houghton, P.J. (2002) *Predominant nuclear localization of mammalian target of rapamycin in normal and malignant cells in culture*. **J Biol Chem**, 277(31), 28127-34.

Zonia, L., and Munnik, T. (2004) *Osmotically induced cell swelling versus cell shrinking elicits specific changes in phospholipid signals in tobacco pollen tubes*. **Plant Physiol**, 134(2), 813-23.

Zuker, M. (2003) *Mfold web server for nucleic acid folding and hybridization prediction*. **Nucleic Acids Res**, 31(13), 3406-15.